Published online 23 September 2021

Nucleic Acids Research, 2021, Vol. 49, No. 22 ***e131***
https://doi.org/10.1093/nar/gkab838

A simple and rapid method for enzymatic synthesis of CRISPR-Cas9 sgRNA libraries

Joshua D. Yates[1], Robert C. Russell[1], Nathaniel J. Barton[1], H. Joseph Yost[2] and Jonathon T. Hill[1,*]

[1]Department of Cell Biology and Physiology, Brigham Young University, Provo, UT, USA and [2]Molecular Medicine Program and Department of Neurobiology, University of Utah, Salt Lake City, UT, USA

Received January 29, 2021; Revised September 03, 2021; Editorial Decision September 06, 2021; Accepted September 09, 2021

ABSTRACT

CRISPR-Cas9 sgRNA libraries have transformed functional genetic screening and have enabled several innovative methods that rely on simultaneously targeting numerous genetic loci. Such libraries could be used in a vast number of biological systems and in the development of new technologies, but library generation is hindered by the cost, time, and sequence data required for sgRNA library synthesis. Here, we describe a rapid enzymatic method for generating robust, variant-matched libraries from any source of cDNA in under 3 h. This method, which we have named SLALOM, utilizes a custom sgRNA scaffold sequence and a novel method for detaching oligonucleotides from solid supports by a strand displacing polymerase. With this method, we constructed libraries targeting the *E. coli* genome and the transcriptome of developing zebrafish hearts, demonstrating its ability to expand the reach of CRISPR technology and facilitate methods requiring custom libraries.

INTRODUCTION

Although originally harnessed for targeting single genomic loci (1,2), the *Streptococcus pyogenes* CRISPR-Cas9 system has since been expanded to include a number of high-throughput molecular techniques that utilize oligonucleotide libraries encoding tens of thousands of unique single-guide RNAs (sgRNAs) to target a large number of genetic loci (3,4). These genome-wide libraries have typically been cloned into lentiviral vectors for forward-genetic screening of cells in culture (3,5–8). However, more recent studies have also adapted them for novel techniques, such as chromosome painting (9,10) and gene regulatory analysis (11), expanding the potential for sgRNA library-based methods.

A major hurdle in the application of the CRISPR sgRNA libraries is in generating the library itself. Designing and chemically synthesizing a custom sgRNA library using microarray-based or similar technologies (12) can be expensive and time consuming and requires detailed sequence information and specialized bioinformatics tools. Because they are based on a reference sequence, the targeting sequences in chemically synthesized libraries can also differ significantly from the actual genome sequences in the recipient organisms, limiting their application to a few model research species with well-assembled genomes and low rates of DNA sequence polymorphism. To address these issues, several approaches have been developed to enzymatically generate libraries from various DNA inputs (10,13–15). Using an enzymatic approach eliminates the need for subject-specific genomic sequence data, greatly diminishes the number of sequence mismatches between the library and subject, and can be synthesized at a fraction of the cost and time of chemically synthesized libraries. However, these methods can be difficult to carry out, require large amounts of DNA, or result in libraries where most of the sgRNAs are non-functional.

Here, we describe a streamlined enzymatic sgRNA library generation method that produces high quality sgRNA libraries from small amounts of input material. By designing a custom sgRNA sequence containing a restriction endonuclease recognition sequence within the repeat:anti-repeat duplex, we were able to develop a method comprising just two consecutive sets of restriction digests and ligations. Additionally, our method constructs the library on the surface of magnetic beads—minimizing the loss of material and simplifying purification between steps. This method, which we have named SLALOM (sgRNA Library Assembly by Ligation On Magnetic beads), can be carried out in a few hours and with <1 μg of DNA or cDNA as input. Using SLALOM, we have generated sgRNA libraries from *Escherichia coli* genomic DNA and normalized cDNA from developing zebrafish hearts. We also show that libraries generated by this method are effective *in vitro* and *in vivo*. Together, our data demonstrate that SLALOM is a

*To whom correspondence should be addressed. Tel: +1 801 422 8970; Fax: +1 801 422 0044; Email: jhill@byu.edu

simple, rapid, cost-effective method for generating sgRNA libraries to be used in a wide range of biological systems and techniques.

MATERIALS AND METHODS

Oligonucleotides and bead preparation

In vitro transcription of sgRNAs was accomplished by annealing and extending pairs of oligos (see Supplemental Table S1 for oligos ordered from Integrated DNA Technologies (Coralville, USA) or Eurofins (Louisville, KY)) by incubating at 66°C for 20 min to produce double-stranded DNA templates containing a T7 promoter sequence, spacer sequence, and sgRNA scaffold sequence followed by column purification. Either the MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific) or the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, Ipswich, USA) was used to transcribe the templates. In either case, 210–300 ng of the template was used and incubated at 37°C for 1–2 h. DNase I was then added, and the reaction incubated for an additional 15 min at 37°C. The sgRNA was then purified using either the RNA Clean & Concentrator-5 Kit (Zymo Research, Irvine, USA) or by phenol-chloroform extraction and ethanol precipitation.

Capture beads were prepared by resuspending 50 μl of streptavidin magnetic beads (New England Biolabs, Ipswich, USA) in 20 μl of containing 100 pmol of the biotinylated oligo and incubating at room temperature for 15 min. The beads were then washed twice by resuspending the beads in 50–100 μl of 1× Cutsmart® buffer. The ability of the biotinylated oligo to bind to the magnetic beads was similar in both the recommended binding buffer as well as CutSmart® buffer.

In vitro digestion with Cas9

In vitro digestion of DNA fragments was accomplished by adding 200 ng of DNA, 70 pmol Cas9 Nuclease, S. pyogenes (1000 nM) (New England Biolabs, Ipswich, USA), and 100 ng of the sgRNA in NEBuffer 3.1 (New England Biolabs, Ipswich, USA) for 20 min at 37°C and 10 min at 65°C, followed by adding 1 μl of proteinase K (800 units/ml) (NEB) and incubating at room temperature for 10 min before being run on a 1.7% agarose gel with a 100 bp ladder for 35 min.

DNA adapters were prepared by resuspending complimentary oligos at a final concentration of 10 μM of each in CutSmart® buffer (New England Biolabs, Ipswich, USA). The reaction was then heated to 98°C for 2 min then set to ramp from 85°C to 65°C for 1 h and finally from 65°C to 8°C for 30 min.

Zebrafish embryo injections

Zebrafish (*Danio rerio*) embryos were collected and injected at the single-cell stage with Cas9:sgRNA Ribonucleoprotein (RNP) by incubating Cas9 (Integrated DNA Technologies, Coralville, USA) with sgRNA in a 300 mM KCl solution for 5 min before injecting ∼1 nl of solution into each zygote. Injected embryos were kept at 28.5°C and after 2 days were visually examined for the extent of pigmentation. Photos were taken using an Olympus SZX16 microscope.

SLALOM library construction method

A complete protocol for the SLALOM method can be found in Supplementary Note 1. Briefly, DNA containing about 10 pmol of recognition sites for HpaII (CCGG) was added to a 50 μl reaction with a final concentration of 1× CutSmart® buffer and 10 units of HpaII. The reaction was incubated at 37°C for 20 min and heat inactivated at 80°C for 20 min. 2000 units of T4 DNA ligase, 10 pmol sgRNA adapter, and 50% PEG 6000 (w/v) were added to the reaction at room temperature to bring the reaction to a total volume of 75 μl and 7.5% PEG. The reaction was then incubated at room temperature for 20 min. The reaction was mixed with capture beads for 15 min and the beads washed twice by exchanging the buffer for 50 μl 1× CutSmart® buffer. The beads were then resuspended in 50 μl of 2 units MmeI and 50 μM *S*-adenosylmethionine (SAM) in 1× CutSmart® buffer and incubated at room temperature for 20 min while keeping the beads in solution by occasionally pipetting. Beads were washed as described above and resuspended in 50 μl of 1× CutSmart® containing 2000 units of T4 and 30 pmol of the T7 adapter in 1× CutSmart®. The reaction was incubated for 20 min at room temperature. The beads were washed and resuspended in 50 μl of 1× CutSmart® containing 10 units of DNA Polymerase 1 and 200 μM dNTPs in 1× CutSmart® and incubated for 5–10 min at RT. A DNA Clean and Concentrate column (Zymo Research, Irvine, USA) was used to purify the final library.

Genome-wide *in silico* human SLALOM library and analysis

The *in silico* SLALOM library was generated using the protein-coding transcripts as annotated by the Gencode protein-coding transcript database, release 35 (GRCH38.p13) (https://www.gencodegenes.org/human/). For genes with multiple coding sequences, the longest transcript sequence was used. Each CCGG site was annotated, and two guides were predicted per restriction site. In total, 380 811 sgRNAs were predicted. Next, we compared the predicted Slalom library efficiency to the GeCKOv1 (3), GeCKOv2 (16) and Brunello libraries (17) by using the Rule Set 2 scoring metric. The Rule Set 2 scoring metric requires not only the protospacer, but also four nucleotides immediately downstream, the PAM site, and the three nucleotides immediately upstream of the PAM site. To find the genomic context of the GeCKOv1 and GeCKOv2 libraries, we searched the Gencode gene sequences associated with the guides' target genes. For protospacers that did not find a match we subsequently searched the entire human genome as provided by Ensembl (ftp://ftp.ensembl.org/pub/release-101/fasta/homo_sapiens/dna/). Protospacers for which no genomic context could be found were dropped from further analysis. Likewise, we dropped the control sgRNAs and sgRNAs targeting miRNAs from the GeCKOv2 library. The Brunello library annotation included the genomic context, so it was not necessary to search for their genomic location.

Once the genomic context for every sgRNA in each library had been identified, each sgRNA was scored using the Rule Set 2 scoring metric. The distribution of scores for each library was plotted as a boxplot with the whiskers showing

the 5th and 95th percentile. To more accurately compare the Slalom library to the GeCKO and Brunello libraries, they were also compared after filtering for the top 3 highest scoring sgRNAs per gene.

To compare the off-target effects of the Slalom library compared to the GeCKOv1, GeCKOv2 and Brunello libraries, we first predicted the potential off-target sites using Cas-OFFinder (18). For each guide in each library, we utilized the Cas-OFFinder algorithm to find all sequences with one to four mismatches in the Gencode protein-coding transcript database sequences, release 35 (GRCH38.p13) (https://www.gencodegenes.org/human/). We chose to use the Gencode protein-coding transcript sequences to focus only on potential off-targets located in sequences coding for proteins. The Cas-OFFinder algorithm predicted 172 065 off-targets for the GeCKOv1 library, 409 294 off-targets for the GeCKOv2 library, 271 713 off-targets for the Brunello library, and 33 744 871 off-targets for the Slalom library. We then calculated the cutting frequency determination (CFD) score for each off-target site (17). The number of off-target sites receiving a CFD score >0.2 was determined for each sgRNA, and the cumulative percentage calculated and plotted. To more accurately compare the Slalom library to the GeCKO and Brunello libraries, we next filtered each library to only include the three sgRNAs with the least number of off-target sites receiving a CFD score >0.2 per gene.

Lambda-phage genome digestion

Lambda-phage genomic DNA was obtained from New England Biolabs (Ipswich, MA). An sgRNA library was then created using the standard SLALOM protocol described above and transcribed using the HiScribe™ T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, Ipswich, USA). 20 pmol of sgRNA library was then mixed with 20 pmol of Cas9, incubated at 25°C for 10 min, and added to 360 ng of phage DNA (containing 2 pmol of HpaII cut sites). The reaction was mixed and incubated at 37°C for 30 min. After digestion, the reaction was stopped by adding 1.5 ul Proteinase K, and the digested DNA was purified using the Zymo Clean and Concentrate-5 kit (Zymo Research, Irvine, CA). A sequencing library of the digested fragments was prepared using the Ligation Sequencing Kit and sequenced using the Flongle Flow Cell on a MinION sequencer (Oxford Nanopore, Oxford, UK).

Sequencing reads were mapped to the lambda genome using the Bowtie2 (v. 2.4.4) software package using the default settings, resulting in 110 315 aligned reads containing 159 069 568 bases (~3280× coverage). Because the Ligation Sequencing Kit simply ligates adapters to the ends of all fragments and the MinION creates full sequence reads, almost all the sequencing-read ends should correspond to the sites digested by Cas9. However, due to basecalling and alignment noise at the ends of the reads, a 6 bp bin surrounding each expected Cas9 cut site was used to calculate the number of sequencing-read ends mapping to each location. A null distribution was also created by quantifying the number of sequencing-read ends aligning to randomly selected 6 bp bins outside of the expected cut locations. Enrichment was determined by comparing the number of sequencing-read ends at each expected cut site to the median background counts, as was done previously (4).

Fluorescent knockout screening

GFP-LC3 HeLa cells stably expressing GFP were cultured in D10 medium (Dulbecco's modified Eagle's medium (DMEM), Fetal bovine serum (10%), L-glutamine, penicillin, Streptomycin) at 37°C with 5% CO_2 and passaged every 3 days to maintain growth conditions. Genomic DNA from these cells was isolated by collecting cells by centrifugation at 1000 g for 5 min and resuspending in 1× RIPA buffer containing Pronase at 37°C for 10–15 min followed by phenol chloroform extraction and ethanol precipitation. Genomic DNA was used as a template to amplify a 716 bp fragment of GFP with 4 HpaII (CCGG) sites. This fragment was used as a substrate for SLALOM, and the resulting library was cloned into the lentiCRISPRv2 plasmid (Addgene, cat. no. 52961). The library was digested with Esp3I and purified using a DNA Clean & Concentrator-5 kit (Zymo Research, Irvine, USA). The plasmid was digested with Esp3I and NheI and the resulting 12 895 bp fragment was gel extracted. Ligation using T4 DNA ligase of the two fragments was followed by transformation into NEB® Stable Competent *E. coli* cells (New England Biolabs, Ipswich, USA) where 100 μl was plated and 900 μl was used for an overnight culture in LB with 100 mg/ml Ampicillin. The overnight culture was used to inoculate 500 ml LB with the antibiotic, and the plasmid DNA was isolated using NucleoBond Xtra Midi EF (Macherey-Nagel, Dueren, Germany). Sequences flanking the spacer sequences were amplified and sequenced. Following validation of the library, it was packaged into lentivirus by VectorBuilder (Chicago, USA).

To calculate the multiplicity of infection (MOI), two six-well dishes were seeded at ~20% confluency and allowed to attach to the plates. Cells from the first dish were transduced with increasing amounts of the virus in media containing 6 μg/ml Polybrene. After 2 days of growth, media was replaced with fresh media containing 8 μg/ml Puromycin for two additional days until all cells in the control well had died and the surviving cells were counted, and the ratio was used to calculate a curve. Cells infected at an estimated MOI of 0.3 and sub-cultured to obtain enough cells for sorting. After 10 days, cells were sorted based on fluorescence using a FACS Aria Fusion and the sorted cells were cultured separately. Genomic DNA from these two populations as well as from uninfected cells was isolated and the GFP coding region was sequenced. The spacers from these two populations were also sequenced.

E. coli genomic DNA isolation

Genomic DNA from *E. coli* MG1655 was isolated using the PowerLyzer® UltraClean® Microbial DNA Isolation Kit from Qiagen (Hilden, Germany). Bacterial cells were harvested and lysed using glass microbeads. The final product was run on a gel and a clear band of high molecular weight was observed. The genomic DNA was then used as a substrate for SLALOM, and the resulting library was sequenced.

Developing zebrafish heart mRNA extraction

Zebrafish hearts were isolated as published previously (19,20). Briefly, about 200 Tg(myl7:GFP) zebrafish embryos were collected at 48 h post-fertilization and placed in media containing tricaine. Embryos were resuspended in L-15 media with 10% FBS, and the tissue of the embryos was disrupted by passing them through a 19G needle. Intact hearts were isolated under a fluorescent microscope with a pipette and, after being washed, were resuspended and homogenized in TRI-Reagent® (Zymo Research, Irvine, USA). Zebrafish heart mRNA was then purified using the Direct-zol™ RNA Kit (Zymogen Research, Irvine, USA).

cDNA library synthesis and normalization

The SMART® cDNA Library Construction Kit (Clontech) was used to reverse transcribe the zebrafish heart RNA. An alternative oligonucleotide that lacked an HpaII binding site was used in place of the 5' oligo provided by the kit to prevent gRNA creation to the adapters in the cDNA library (Supplementary Table S1). The cDNA library was then normalized using the Trimmer-2 cDNA normalization kit (Evrogen, Moscow, Russia).

High-throughput sequencing and data analysis of the *E. coli* and zebrafish libraries

High throughput sequencing of the libraries was carried out at the Huntsman Cancer Institute High-Throughput Genomics and Bioinformatics Analysis Shared Resource on an Illumina HiSeq 2500. PhiX DNA was added to each library before sequencing to allow for more diversity during the initial reading process. Custom scripts were generated to extract the targeting region of each sgRNA template and analyze the data. Because the length of the library was 128 bp, but the read length was only 50 bp, about half of the reads were discarded because they did not include the spacer. The extracted spacer sequences were then aligned to either the Escherichia coli MG1655 genome or the zebrafish GRCz11 genome. Statistical analysis was then used to determine the coverage of each library.

Confirmation of the GFP mutation rates was conducted by high-throughput sequencing using the Genewiz Amplicon-EZ service (South Plainfield, USA) and analyzed using custom scripts in R based on the Cris.py pipeline (21). Briefly, reads containing complete PCR products were selected and searched for a 30 bp sequence spanning each HpaII site. Reads were categorized by whether they contained wild type sequences at each site. Reads with identified mutations at each site were then tabulated to find the mutation frequency at the site and the most common mutations. Reads with large deletions that spanned multiple HpaII sites were categorized separately as 'Multiple'.

RESULTS

Method overview

The SLALOM method, described here, relies on two major innovations that improve the speed and efficiency of enzymatic sgRNA library synthesis over previous methods. The first innovation is the design and creation of a dual-role adapter that can function as a binding site for MmeI and as a template for transcription of the sgRNA scaffold. The second innovation simplifies purification and handling of the library during construction by incorporating magnetic bead purification into the workflow. These two factors allowed us to greatly decrease the number of steps in the SLALOM protocol (outlined in Figure 1, complete protocol in Supplemental Note 1). In this protocol, the restriction enzyme HpaII (CCGG), which recognizes a DNA sequence containing a PAM, is first used to fragment the DNA (Step 1). HpaII was selected because its short, palindromic recognition sequence ensures high coverage and cuts adjacent to protospacer adjacent motifs (PAMs) on both the forward and reverse strands (10). Next, an adapter is ligated to the ends of the digested fragments (Step 2). This adapter contains a custom sgRNA scaffold sequence modified to incorporate an MmeI recognition sequence and a long 5′ overhang at the 3′ end of the scaffold sequence. After ligation, the products are immobilized on magnetic beads by hybridization of the long single-stranded overhang to a complementary sequence affixed to the beads. After washing, the ligated DNA is digested by MmeI, which cleaves at a fixed distance from its recognition sequence (22), leaving 18 or 19 bp of the ligated DNA sequence to form the sgRNA spacer (Step 3). After an additional wash step, the fragments are ligated with a second adapter containing a promoter region (e.g. T7 for *in vitro* transcription or U6 for *in vivo* transcription) (Step 4). Finally, the beads are washed, and the completed library detached and repaired using a strand-displacing polymerase (Step 5).

Bifunctional scaffold adapter

Creating an adapter that could be used as both a binding site for a long reach restriction enzyme and as a template for transcription of the sgRNA requires modification of the Cas9 binding region of the sgRNA without disrupting endonuclease activity. A previously reported crystal structure of Cas9 in complex with the sgRNA revealed that the REC1 domain of Cas9 interacts with the repeat:anti-repeat duplex of the sgRNA (23). Functional experiments in this same study showed that some mutations to the sequence of the duplex are permitted, while others can diminish or completely inhibit the Cas9 activity. It was therefore unclear if an MmeI recognition sequence could be incorporated into the repeat sequence without impairing the function of the Cas9 complex.

Based on this information, we incorporated the MmeI recognition sequence (TCCRAC) into the repeat sequence of the sgRNA at position 21 (directly after the spacer region) by making two mutations, U24G and U25G, in the repeat sequence. We also made corresponding mutations, A46C and A47C, in the anti-repeat sequence to maintain the repeat:antirepeat stem loop structure (Figure 2A). Cas9 digestion of a 1000 bp DNA fragment *in vitro* using the wild type or modified sgRNA showed cleavage efficiencies similar to or better than the unmodified sgRNA with both an 18 and 20 bp spacer, indicating that endonuclease activity of the protein was retained (Figure 2B, lanes 4 and 5).

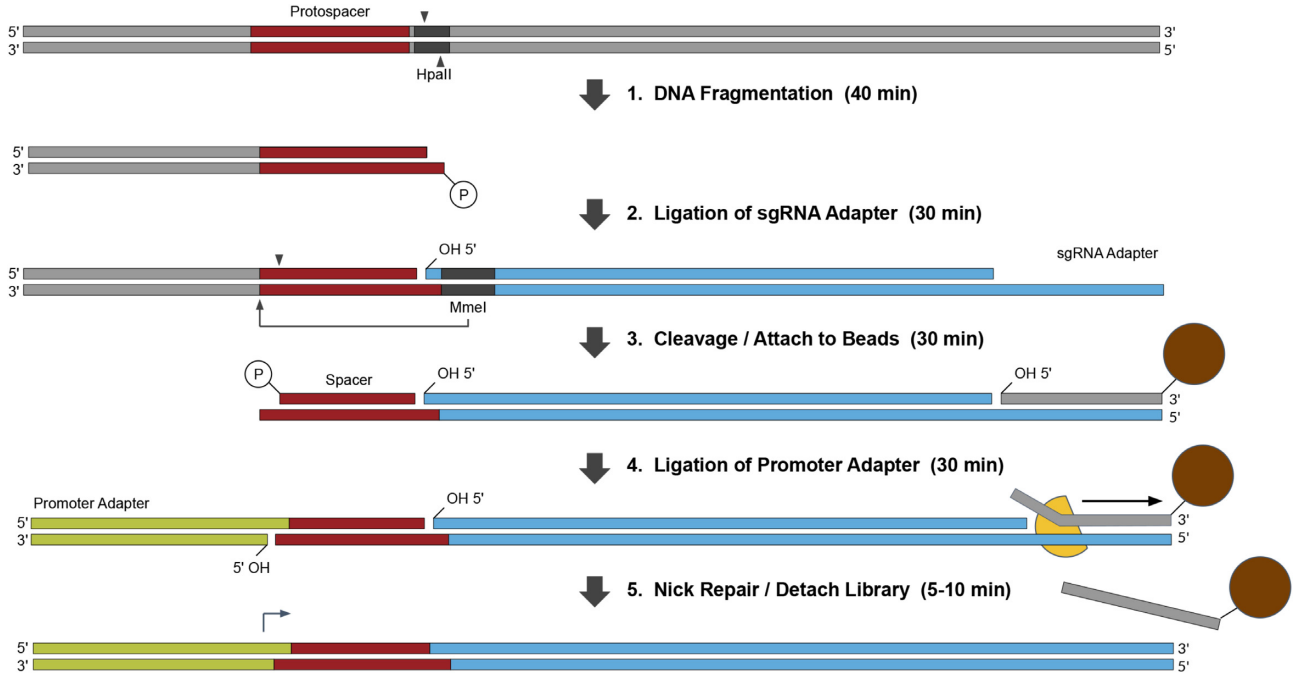

Figure 1. Schematic of the sgRNA Library Synthesis Method. Each step of the process is represented by a black arrow. Restriction enzyme binding sites are shown in black. Magnetic beads are represented as orange circles. (Step 1) DNA from an arbitrary source is fragmented by a restriction enzyme such as HpaII (CCGG) that contains the canonical PAM (NGG) in its recognition sequence. (Step 2) An adapter (blue) encoding a modified sgRNA sequence is ligated to the DNA fragments. The adapter is unphosphorylated, contains an MmeI recognition sequence, has a two base-pair, single-stranded overhang compatible with the fragmented DNA, and includes a long single-stranded overhang capable of hybridizing to a single-stranded oligo. (Step 3) A single-stranded capture oligonucleotide attached to a magnetic bead at the 3′ end is used to immobilize the adapter to magnetic beads through hybridization and excess DNA is washed away before digestion with MmeI to capture the spacer sequence. (Step 4) An unphosphorylated adapter containing the T7 promoter sequence (green) and a 3′ overhang containing all possible dinucleotides is ligated to the cleaved spacer sequence. (Step 5) The nicked library is detached from the beads by a strand displacing polymerase (yellow) and nicks are simultaneously repaired by the same mechanism. The detached double stranded oligonucleotides constitute an sgRNA library containing spacers targeting the DNA source and can be transcribed *in vitro* or cloned into plasmids for downstream applications.

The HpaII enzyme used to fragment the DNA prior to ligation with this adapter also produces short 3′ overhangs not compatible with the wildtype Cas9 sgRNA sequence. Therefore, we next engineered additional mutations in the scaffold to allow the sgRNA adapter sequence to hybridize with the fragmented ends, eliminating the need for removal of the overhangs and increasing ligation efficiency over blunt-end ligation. This was accomplished by shifting the MmeI recognition sequence to position 23 and modifying the sgRNA sequence by making mutations G21C, T22G, T23G and A26G in the repeat sequence and T45C, A48C and A49C in the anti-repeat sequence. Moving the MmeI site to this position resulted in guide sequences of 18 and 19 bp (Figure 2A) instead of the standard 20 bp. However, digestion of the 1000 bp DNA fragment *in vitro* using an sgRNA with these modifications again showed cleavage at the predicted location with similar efficiency to that of the unmodified sgRNA using either a 20 or 18 bp protospacer (Figure 2B, lanes 6 and 7).

To confirm that sgRNAs with these changes function *in vivo*, we tested an sgRNA with these modifications and a spacer sequence targeting the zebrafish pigment gene *gol* (slc24a5) (24). Injection of single-cell zebrafish embryos with the modified sgRNA showed levels of pigment loss at 48 hpf comparable to an unmodified sgRNA targeting the same spacer (Figure 2C). Together, these results demonstrate that modified sgRNAs do not disrupt Cas9 nuclease function and can be used to guide Cas9 editing *in vitro* and *in vivo*.

Finally, to further verify that the sgRNA modifications were not detrimental to the activity of the Cas9–sgRNA complex when used in a large pool, we designed a high-throughput experiment to test the activity of a complete SLALOM library. We first used lambda phage genomic DNA to create a SLALOM library with 654 sgRNAs. The library was then transcribed and used to digest the intact lambda-phage genome *in vitro*. The resulting DNA fragments were ligated to adapters and sequenced using an Oxford Nanopore MinION sequencer. As the MinION creates full-length sequencing reads, the ends of each read can be used to determine the locations and relative frequency of Cas9 cutting, similar to past Sanger sequencing approaches for characterizing restriction enzyme activity (22). Analysis of the results showed that 88.1% of the expected cut sites (median coverage = 16) were enriched for sequencing-read ends over the background (median coverage = 2, Supplemental Figure S1). Thus, a large majority of the guides made using SLALOM were active *in vitro*, even when used as part of a complex sgRNA pool. This result is also consistent with a previous study where 87.5% of sgRNAs in a synthesized library were enriched over the median background level in human cells (25), indicating that our library is of

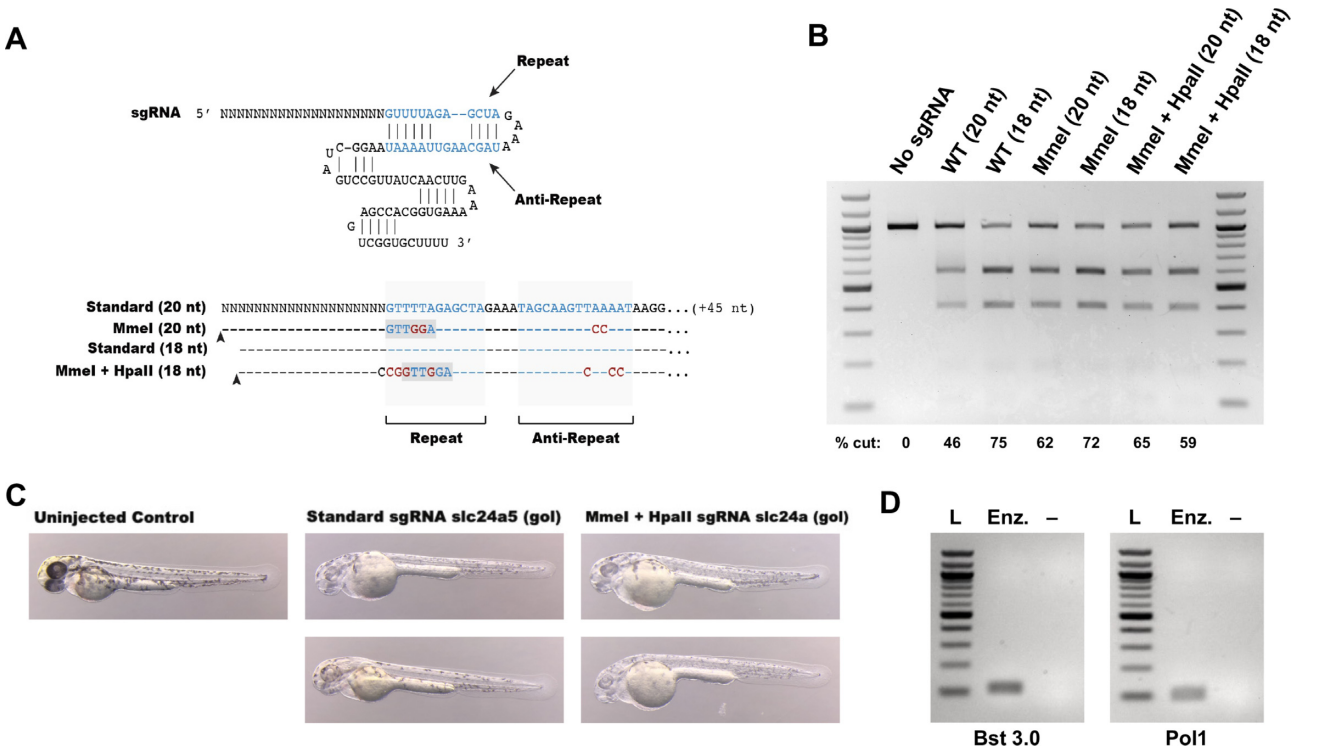

Figure 2. Incorporation of the MmeI Recognition Sequence into the Repeat:Anti-Repeat Duplex and Elution with a Strand Displacing Polymerase. (**A**) Modifications were made to the Repeat:Anti-Repeat Duplex of the sgRNA sequence to insert a single MmeI site that cleaves 18-20 bp upstream of the sgRNA recognition sequence. Modifications are shown in red. The MmeI binding sites are shown in dark gray boxes. The MmeI cut site is represented with a triangle. (**B**) *in vitro* Cas9 digestion of a 1000 bp DNA fragment with each sgRNA variant. (**C**) Injection of zebrafish embryos with Cas9 and the sgRNA variant targeting the *gol* gene compared to an uninjected control (minimum n per pool = 150). (**D**) Detachment of the sgRNA scaffold from magnetic beads using Bst 3.0 or Pol1.

similar quality to previously published sgRNA libraries, and that the sgRNA modifications used in this technique were not detrimental to the activity of the Cas9–sgRNA complex.

Detaching oligonucleotides by polymerase strand displacement

Constructing the library on the surface of magnetic beads allows for rapid purification between steps without disrupting the function of restriction enzyme or T4 DNA ligase activity (26). To immobilize the modified sgRNA adapter on the beads, a biotinylated ssDNA oligo was designed to hybridize with the 3′ end of the modified sgRNA adapter and a corresponding 19 base overhang was designed into the 3′ end of the adapter. The ssDNA oligos were then attached to streptavidin-coated paramagnetic beads during the SLALOM protocol. Elution was accomplished by incubating the beads with a strand-displacing polymerase, which elongates the 3' end of an oligo while displacing the pre-existing strand, detaching the final library product from the beads (See Figure 1). Two strand displacing polymerases were tested: Bst 3.0, which is most efficient at temperatures near the melting temperature of the bead-attached oligo, and Pol I, which is active at room temperature and contains 5′-3′ exonuclease activity, which completely digests the oligo being displaced. Oligo hybridization and elution assays showed that both enzymes were highly efficient (Figure 2D), thus providing a simple technique to irreversibly detach oligonucleotides from solid supports while simultaneously repairing the nicks caused by using unphosphorylated substrates in the earlier ligation steps.

Genome-wide library comparisons

Several genome-wide libraries have been designed with the aid of scoring systems that attempt to predict sgRNA targets with relatively high specificity and activity in order to increase effective gene coverage as multiple targets for each gene facilitates the identification of true positive hits. Using the annotated coding regions for the human genome we modelled a human genome-wide SLALOM library *in silico* and compared it to the routinely used GeCKOv1 (3), GeCKOv2 (16) and Brunello (17) libraries.

The number of genes targeted by each library was similar, targeting between 92.0 and 94.1% of the annotated genome (Figure 3A). The number of guides per gene in the SLALOM library was higher than that of the other libraries, with an average of 20.21 guides per gene in the SLALOM library, 5.95 in GeCKOv2, 4.0 in the Brunello, and 3.44 in GeCKOv1. Increasing the coverage of a library necessarily decreases its overall efficiency, which resulted in the SLALOM library having a lower overall efficiency than the Brunello library, which was specifically designed to select sgRNAs with high Rule Set 2 scores, and slightly lower than the GeCKOv1 and GeCKOv2 libraries. (Figure 3B).

A

Library	Total Guides	Genes Targeted (%)	Guides per Gene
Slalom	380811	18840 (92.7%)	20.21
GeCKOv1	64209	18690 (92.0%)	3.44
GeCKOv2	113272	19047 (93.8%)	5.95
Brunello	76441	19114 (94.1%)	4.0

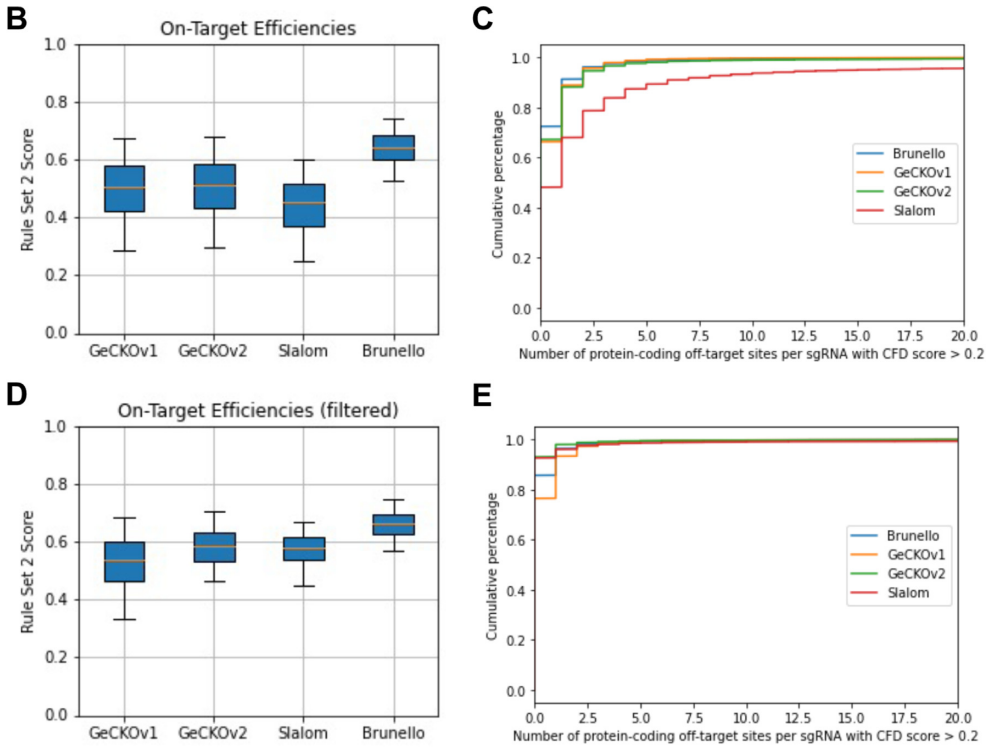

Figure 3. Comparison of genome-wide SLOLAM library to the GeCKOv1, GeCKOv2, and Brunello sgRNA Libraries. (**A**) Comparison of key library statistics. (**B**) Boxplot comparison of the Rule Set 2 predicted efficiency scores for each library. (**C**) Comparison of the CFD scores for each library. (**D**) Boxplot comparison of the Rule Set 2 predicted efficiency scores of the top 3 sgRNAs per gene. (**E**) Comparison of the CFD scores for the top 3 sgRNAs per gene. In both boxplots, horizontal lines in the boxes indicate the 25th, 50th and 75th percentile, and the whiskers indicate the 5th and 95th percentile.

Similarly, the predicted off-target score distribution for the SLALOM library was slightly lower than that of the other libraries (Figure 3C).

The higher coverage of the SLALOM library could make it easier to distinguish false negative and false positive hits, as multiple hits will be expected in any gene identified. In addition, while it is not possible to select particular guides from the library, it is possible to filter the post screen data to tune a library to a particular scoring system. We filtered the SLALOM library to include only the top 3 guides per gene according to the Rule Set 2 scoring system and showed that this improved the overall score to be similar between all of the libraries (Figure 3D and E). Thus, even though poor guides are not actively excluded during the SLOLAM library, overall quality is comparable to synthesized libraries, and, if desired, the higher coverage in the SLOLAM library makes it possible to filter the targets during the post-screen data analysis to only include guides that meet a predetermined set of parameters.

GFP library and fluorescent knockout screen

To assess functionality of a SLALOM-generated library in cell culture, we used SLALOM to create a library targeting the GFP coding region in GFP-LC3 Hela cells (27). The GFP construct in these cells contains four HpaII sites (CCGG) (Figure 4A). We PCR amplified the GFP coding region (716 bp) to use as the SLALOM DNA input. In this iteration, the sgRNA adapter was also modified to use a longer stem-loop, which has been shown to potentially increase sgRNA efficiency (9), and both adapters incorporated Esp3I restriction sites for cloning into the Lenti-CRISPRv2 plasmid (16). Library sequencing confirmed complete coverage of the 8 expected sgRNAs (Supplemental Figure S2). The cloned library was then packaged into lentivirus and used to transfect HeLa GFP-LC3 cells. Sorting of the transfected and untransfected cells by FACS showed that ∼15% of the transfected cells lost GFP expression (Figure 4B, Supplemental Figure S3), comparable to previous studies looking at CRISPR efficiency (28).

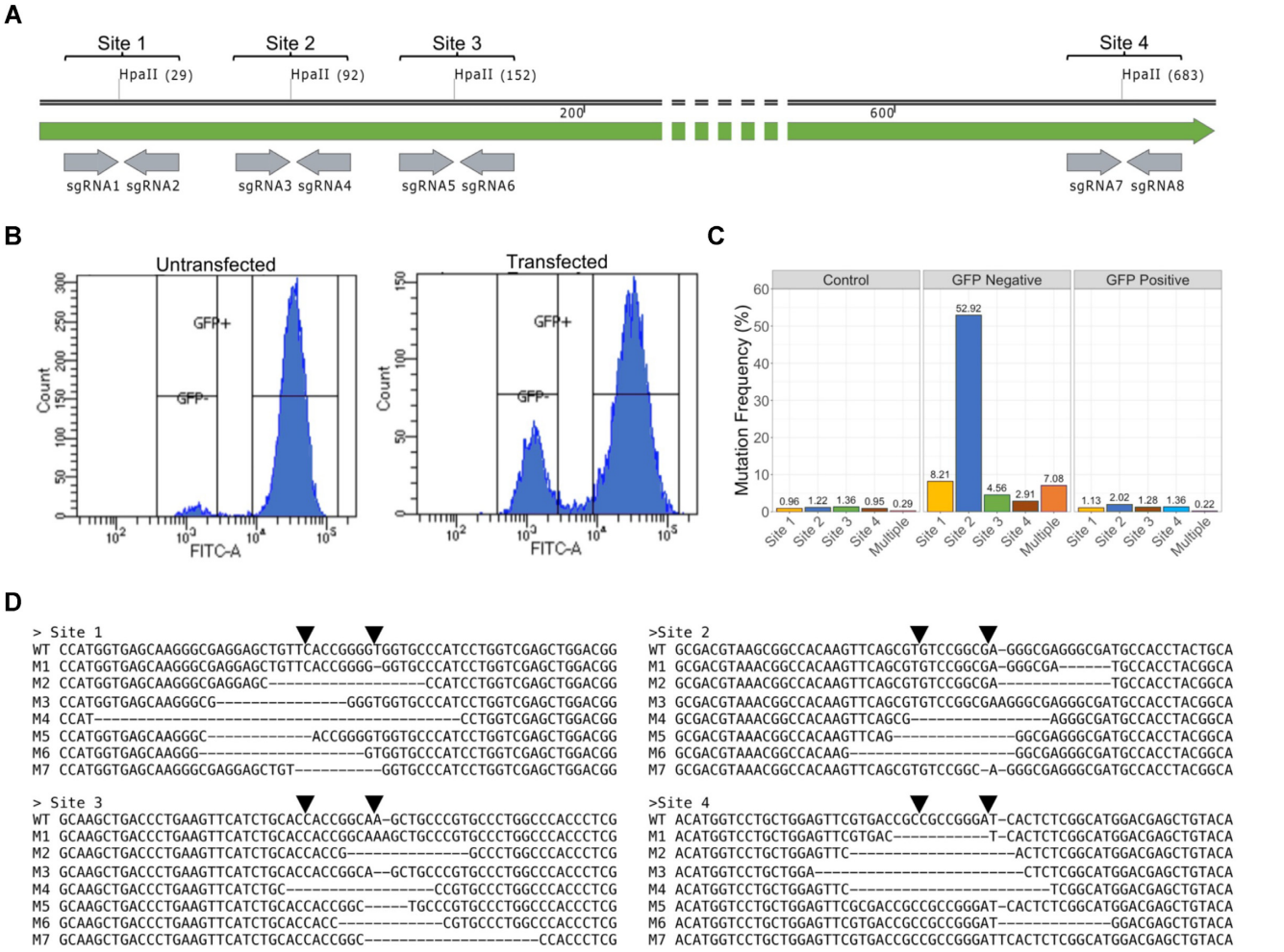

Figure 4. GFP Knockout Screen. (**A**) Genomic DNA was isolated from HeLa cells stably expressing GFP-LC3. The four HpaII sites within the GFP ORF and the predicted sgRNAs are labeled. (**B**) FACS sorting of transfected and untransfected cells. GFP fluorescence is on the x-axis and counts are on the y-axis. (**C**) Mutation frequency at each of the HpaII sites based on high-throughput sequencing of the GFP locus. (**D**) The most common variants for each of the four HpaII sites.

Sequencing of the incorporated spacers and GFP mutations in each pool showed that all eight potential spacers could be found in the GFP-negative pool and mutations occurred at all four sites (Figure 4C). However, it was not possible to identify the mutation rate of each individual sgRNA, as many of the common mutations deleted both cut sites within a pair (Figure 4D).

After establishing that all of the sgRNAs were active, we next analyzed their relative mutation efficiency. Relative enrichment of each sgRNA in the GFP minus pool was determined by normalizing the fractional counts from the sgRNA spacer sequencing results of the GFP minus pool to their fractional counts in the GFP positive pool (Supplemental Figure S4A). It should be noted that this analysis is not completely analogous to enrichment analyses in genome-wide screens, as there is no background gene set for normalization, but does show the relative efficiency of each sgRNA in this library. These data showed that one sgRNA, a 19 bp sequence targeting the negative strand of site 2, was more active than the others. There were also two sequences that appeared to be significantly less effective than the average, an 18 bp sgRNA targeting site 3 and an 18 bp sgRNA targeting site 4.

We initially hypothesized that this variance may be due to the fact that SLALOM does not select for sgRNAs based on predicted efficiency. However, enrichment in the GFP-negative pool did not show strong correlation with predicted efficiency score (29) (Supplemental Figure S4B, $r^2 = 0.003$), position in the gene (Supplemental Figure S4C, $r^2 = 0.077$) or length of the match (Supplemental Figure S4D, $r^2 = 0.213$). Overall, these results show that the library can effectively target multiple sites and generate a measurable enrichment of mutations causing a specific phenotype of interest, despite not using bioinformatics analysis during the design the library. Furthermore, it highlights the benefits of targeting each gene a large number of times in light of the clear limitations of current prediction algorithms.

E. coli and zebrafish heart libraries

To test the method on a more complex DNA substrate, we isolated genomic DNA from *E. coli* MG1655 and created an

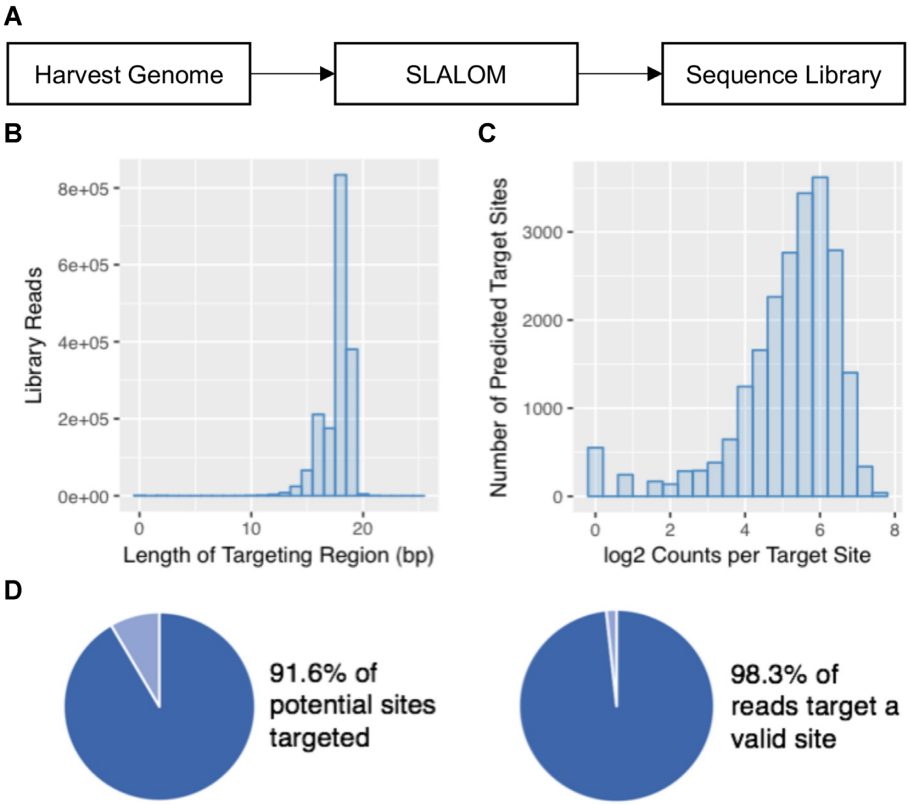

Figure 5. High-throughput sequencing of a SLALOM-generated sgRNA library to the *E. coli* MG1655 genome. (**A**) Schematic diagram of the library generation process. (**B**) Histogram showing the distribution of spacer lengths within the library. (**C**) Histogram showing representation of each sgRNA target in the library. (**D**) Analysis showing the coverage and purity of the library.

sgRNA library targeting the entire genome using SLALOM (Figure 5A). The *E. coli* genome contains 24 311 HpaII sites. Because SLALOM produces two sgRNAs, one on the plus strand and one on the minus strand, for each site, we predicted a full-coverage library would contain 48 622 sgRNA templates. High-throughput sequencing of the spacers was then conducted to characterize the resulting sgRNA library. Sequencing results showed that the most common spacer lengths were 18 and 19 bases (Figure 5B) and were evenly represented in the library (Figure 5C). Sites in the library covered 91.6% of the predicted HpaII sites and targeted 92.7% of the genes at least once (median 8 guides per gene). The library was also very clean, as 98.3% of the reads represented a PAM-adjacent spacer (Figure 5D), indicating that SLALOM can be used to create high quality sgRNA libraries from a high molecular weight DNA input.

A clear advantage of SLALOM over chemically synthesized libraries is the ability to create tissue or cell-type specific libraries without prior information on the genes expressed in those cells. For example, cDNA libraries from a tissue of interest could be used to create an sgRNA library that targets only the coding regions of actively expressed genes. This reduces the overall library size while allowing the researcher to generate and focus on mutations in the genomic regions most likely to provide phenotypes in a particular tissue or organ of interest. Thus, we next tested the ability of SLALOM to create an sgRNA library targeting all of—but only—the genes expressed in the developing zebrafish heart (Figure 6A). Zebrafish hearts were isolated from 48 hours post-fertilization (hpf) embryos for mRNA isolation. Because genes are expressed at various levels, we next created normalized zebrafish heart cDNA by subtractive hybridization (30). The normalized cDNA library was confirmed by qPCR (Figure 6B) and used as input for SLALOM.

Similar to the *E. coli* genome library, sequencing validation of the zebrafish heart library found spacers that were predominately 18 or 19 bp long (Figure 6C). As sgRNAs in this library were expected to only target expressed genes, we compared the library to a set of genes detected in an existing zebrafish heart RNA-seq dataset (20). These data showed the library targets at least 11 217 of the genes expressed in the heart (Figure 6D). Representation in the library is not correlated with gene expression levels in the heart, confirming that cDNA normalization was effective (Figure 6E). Thus, SLALOM can be used to target an sgRNA library to a specific tissue.

DISCUSSION

Generation of sgRNA libraries by chemical synthesis can be expensive and slow, and it requires significant bioinformatic analysis of a known genome to identify suitable spacers. Because of these limitations, synthesized libraries have been largely limited to a few well-established model organisms with low DNA sequence polymorphism rates (such as

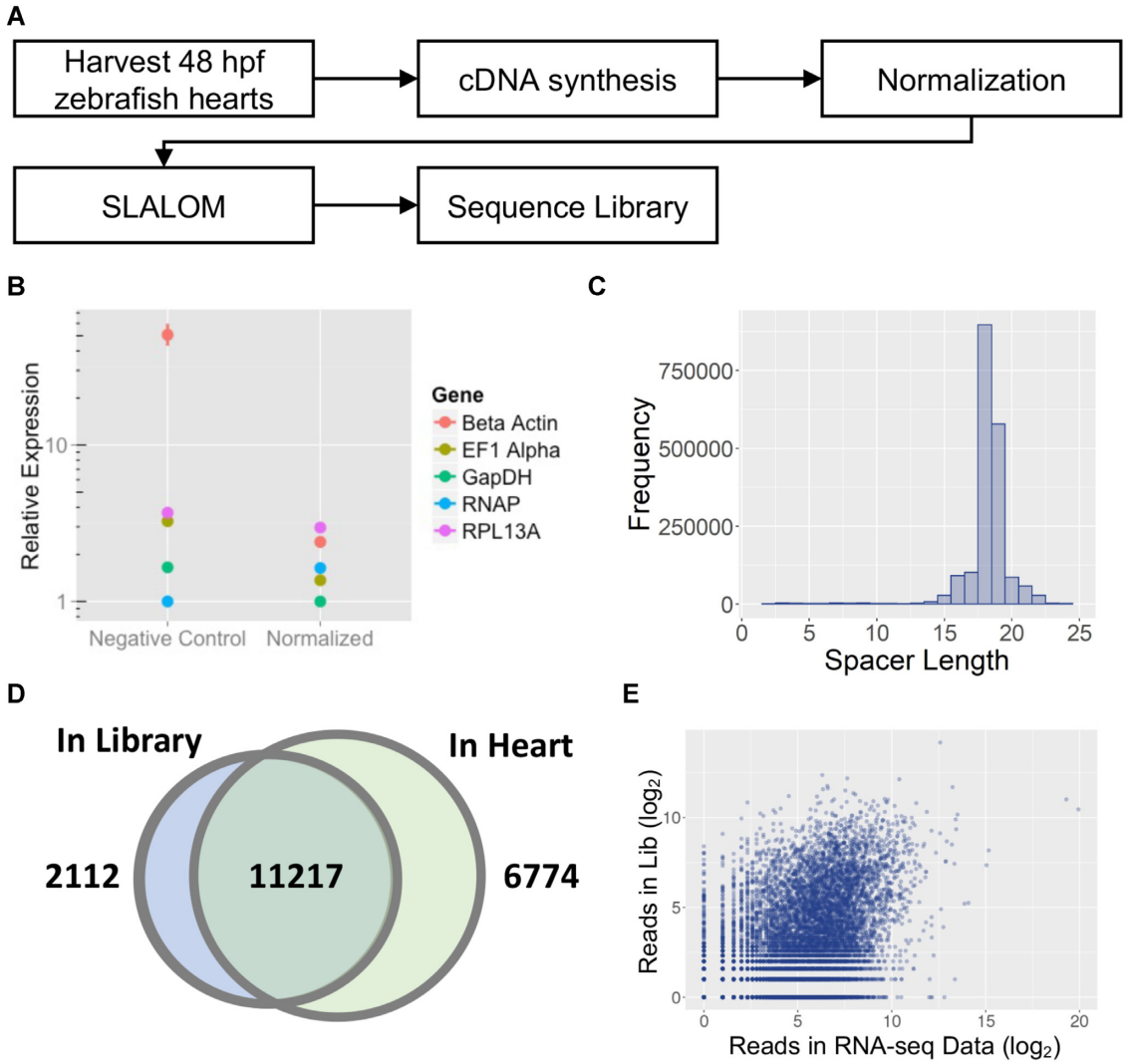

Figure 6. High-throughput sequencing of a SLALOM-generated sgRNA library to normalized cDNA extracted from 48 hpf zebrafish hearts. (**A**) Schematic diagram of the library generation process. B) Results of qPCR analysis on five genes with varying expression levels in the endogenous tissue before and after cDNA normalization. (**C**) Histogram showing the distribution of spacer lengths within the library. (**D**) Venn diagram of the genes targeted by at least one spacer in the library compared to genes detected in the 48 hpf zebrafish heart by RNA-seq. A threshold of 1 read per gene was used to designate both genes and spacers as 'detected'. (**E**) Comparison of the total number of spacer reads vs. expression level in an RNA-seq dataset for 48 hpf zebrafish hearts.

inbred mice and human cell lines). They have also been generally designed to mutate complete genomes, as customization to the genes expressed in a particular tissue or cell type would require additional genomic data, bioinformatic analysis, and custom synthesis, significantly raising costs even further. However, the resulting libraries cover only a small fraction of the biological systems where they could potentially be applied. Wider adoption of CRISPR library methods in non-traditional organisms and the development of new CRISPR-based technologies dependent on custom subsets of the genome require the development of more attainable library generation methods.

One alternative to chemical synthesis is enzymatic processing of DNA inputs into sgRNAs. Several such methods have been proposed. Lane *et al.* (10) digested DNA with restriction enzymes containing a PAM in their recognition sequences and, after ligation of a temporary adapter, used the restriction enzyme MmeI to capture the spacer before removing this adapter and attaching an adapter containing the sgRNA sequence. This method resulted in an *E. coli* genome library where 44% of the sequenced reads represented functional guides and 51% of the predicted spacers were included. In contrast, the *E. coli* library presented here achieved much better results of 98.3% and 91.6%, respectively, for these same analyses. Arakawa, *et al.* (13) also used a restriction enzyme that cleaves outside of its recognition sequence to capture the spacer but selected for PAMs by creating a cDNA library using a semi-random hexamer containing the dinucleotide CC. This resulted in a library where 77.6% of the spacers were adjacent to the PAM. In contrast, Cheng *et al.* (14) and Köferle *et al.* (15) did not select for PAMs but produced libraries from completely random fragments of a DNA sample. As a result, they produced dense libraries that target practically all valid PAMs,

but the vast majority of spacers were not functional. Thus, there is a wide range of issues in the previous approaches, including complexity, reliability, difficulty, time requirements, input amounts, library quality, and density.

SLALOM has several advantages over these methods, as it requires fewer steps and smaller amounts of starting material while yielding sgRNA libraries in which almost all of the final sgRNAs contains functional guide sequences. Similar to the Lane *et al*. method described above, SLALOM takes advantage of a restriction enzyme to determine the PAM, as this method makes it possible to use any source of DNA input (e.g. genomic DNA, cDNA, or PCR products) and to tune the density of the library by choosing enzymes with different recognition sequence lengths and/or combining multiple restriction enzymes. However, SLALOM incorporates a modified sgRNA sequence to improve the efficiency and speed of library generation. For example, unlike the Lane *et al*. method, SLALOM uses a two-base overhang incorporated into the adapter sequence, eliminating the need for mung-bean nuclease blunting of the digested fragments. Also, unlike several previously reported methods (10,13,14) that temporarily ligate an intermediate adapter containing a recognition sequence for a restriction enzyme that cleaves outside of this sequence to capture ∼20 bp of the fragment before removing the adapter, SALOM incorporates this restriction enzyme site directly into the sgRNA template, completely eliminating the need for an intermediate adapter. Removing the intermediate adapter also allows the reaction steps to occur while attached to a magnetic bead, making purification between steps fast and simple and reducing material loss.

While SLALOM can already make high quality libraries, it also serves as a platform to expand and diversify gene-targeting methods. For example, other restriction enzymes could be used to create different sets of sgRNAs or in combination to increase library density. There might also be ways to improve the mutation frequency, such as using a Type IIS enzyme with a longer reach to generate sgRNAs that are 20–21 instead of 18–19 bases long. Based on previous studies, this would likely increase cutting efficiency but could also decrease specificity (31,32). Finally, we made libraries using genomic DNA and normalized cDNA, but other sources of the DNA input, such as immunoprecipitations to obtain DNA fragments bound by various proteins, could be used as starting material. For example, conducting an RNA-PolII pulldown could make the representation of genes actively transcribed in the tissue of interest more consistent than normalized cDNA while still limiting the library to active genes, and pulldowns of other DNA-binding proteins could also be used to examine subsets of enhancers or other non-coding regions, especially if combined with repressor- or activator-bound Cas9 variants (5,33–35).

One potential application of SLALOM-generated sgRNA libraries is in forward-genetic screening. CRISPR screening in cell culture is already well established and has been shown to be more effective than chemical mutagenesis or siRNA-based methods (8,36). However, current sgRNA libraries are generally synthesized to target all of the genes in the genome, which reduces the number of times each gene can be targeted and increases the number of cells without mutations in active genes, limiting the rate of gene discovery. However, enzymatic sgRNA generation allows the library to be tailored to the genes expressed in the tissue and at the time point of interest. For example, the zebrafish heart library reported here contains only sgRNAs targeting the exons of genes expressed during heart looping morphogenesis. Thus, this method will restrict mutations to genes involved in heart development, greatly improving the rate of gene discovery and potentially expanding the number of biological systems that can be screened using CRISPR technology beyond cell culture. Similar applications of SLALOM to novel experimental designs and organisms will greatly increase the reach of CRISPR-Cas9 technology in a wide range of fields.

DATA AVAILABILITY

All high-throughput sequencing data used in this study are available through the Sequence Read Archive (Project ID: PRJNA642300).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The Hela cells used here were a generous gift from Joshua L. Anderson. We would also like to thank Marc Hansen, Brent Nielsen, and Jeffery Barrow for their critical reading of the manuscript.

IACUC APPROVAL

All animal studies were approved by the Brigham Young University Institutional Animal Care and Use Committee under protocol number 18-0704.

FUNDING

NIH [1R15HD098969 to J.T.H.; UM1 HL098160 to H.J.Y., in part]; sequencing was supported by a core facilities support grant to CCHCM [U01HL131003], from the National Heart, Lung, and Blood Institute. Funding for open access charge: NIH [1R15HD098969]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
Conflict of interest statement. J.H. and J.Y. are inventors of US Patent No. 10,669,539 and are co-founders of Pioneer Biolabs, LLC.

REFERENCES

1. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816–821.
2. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).*, **339**, 819–823.
3. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.

4. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
5. Sanson,K.R., Hanna,R.E., Hegde,M., Donovan,K.F., Strand,C., Sullender,M.E., Vaimberg,E.W., Goodale,A., Root,D.E., Piccioni,F. *et al.* (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.*, **9**, 5416.
6. Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 192.
7. Viswanatha,R., Li,Z., Hu,Y. and Perrimon,N. (2018) Pooled genome-wide CRISPR screening for basal and context-specific fitness gene essentiality in Drosophila cells. *Elife*, **7**, e36333.
8. Koike-Yusa,H., Li,Y., Tan,E.P., Velasco-Herrera,M.D.C. and Yusa,K. (2014) Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.*, **32**, 267–273.
9. Chen,B., Gilbert,L.A., Cimini,B.A., Schnitzbauer,J., Zhang,W., Li,G.W., Park,J., Blackburn,E.H., Weissman,J.S., Qi,L.S. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, **155**, 1479–1491.
10. Lane,A.B., Strzelecka,M., Ettinger,A., Grenfell,A.W., Wittmann,T. and Heald,R. (2015) Enzymatically generated CRISPR libraries for genome labeling and screening. *Dev. Cell*, **34**, 373–378.
11. Klann,T.S., Black,J.B., Chellappan,M., Safi,A., Song,L., Hilton,I.B., Crawford,G.E., Reddy,T.E. and Gersbach,C.A. (2017) CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.*, **35**, 561–568.
12. Kosuri,S. and Church,G.M. (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods*, **11**, 499–507.
13. Arakawa,H. (2016) A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism. *Sci. Adv.*, **2**, e1600699.
14. Cheng,J., Roden,C.A., Pan,W., Zhu,S., Baccei,A., Pan,X., Jiang,T., Kluger,Y., Weissman,S.M., Guo,S. *et al.* (2016) A Molecular Chipper technology for CRISPR sgRNA library generation and functional mapping of noncoding regions. *Nat. Commun.*, **7**, 11178.
15. Köferle,A., Worf,K., Breunig,C., Baumann,V., Herrero,J., Wiesbeck,M., Hutter,L.H., Götz,M., Fuchs,C., Beck,S. *et al.* (2016) CORALINA: a universal method for the generation of gRNA libraries for CRISPR-based screening. *BMC Genomics*, **17**, 917.
16. Sanjana,N.E., Shalem,O. and Zhang,F. (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods*, **11**, 783–784.
17. Doench,J.G., Fusi,N., Sullender,M., Hegde,M., Vaimberg,E.W., Donovan,K.F., Smith,I., Tothova,Z., Wilen,C., Orchard,R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
18. Sangsu,B, Jeongbin,P. and Kim,J.-S. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473–1475.
19. Burns,C.G. and MacRae,C.A. (2006) Purification of hearts from zebrafish embryos. *Biotechniques*, **40**, 274–282.
20. Hill,J.T., Demarest,B., Gorsi,B., Smith,M., Yost,H.J., Gorsi,B., Yost,H.J., Smith,M. and Yost,H.J. (2017) Heart morphogenesis gene regulatory networks revealed by temporal expression analysis. *Development*, **144**, 3487–3498.
21. Connelly,J.P. and Pruett-Miller,S.M. (2019) CRIS.py: a versatile and High-throughput analysis program for CRISPR-based Genome Editing. *Sci. Rep.*, **9**, 4194.
22. Morgan,R.D., Dwinell,E.A., Bhatia,T.K., Lang,E.M. and Luyten,Y.A. (2009) The MmeI family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res.*, **37**, 5208–5221.
23. Nishimasu,H., Ran,F.A., Hsu,P.D., Konermann,S., Shehata,S.I., Dohmae,N., Ishitani,R., Zhang,F. and Nureki,O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935–949.
24. Burger,A., Lindsay,H., Felker,A., Hess,C., Anders,C., Chiavacci,E., Zaugg,J., Weber,L.M., Catena,R., Jinek,M. *et al.* (2016) Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development*, **143**, 2025–2037.
25. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
26. Pengpumkiat,S., Koesdjojo,M., Rowley,E.R., Mockler,T.C. and Remcho,V.T. (2016) Rapid synthesis of a long double-stranded oligonucleotide from a single-stranded nucleotide using magnetic beads and an oligo library. *PLoS One*, **11**, e0149774.
27. Mizushima,N., Yoshimori,T. and Levine,B. (2010) Methods in mammalian autophagy research. *Cell*, **140**, 313–326.
28. Jin,J., Xu,Y., Huo,L., Ma,L., Scott,A.W., Pizzi,M.P., Li,Y., Wang,Y., Yao,X., Song,S. *et al.* (2020) An improved strategy for CRISPR/Cas9 gene knockout and subsequent wildtype and mutant gene rescue. *PLoS One*, **15**, e0228910.
29. Kim,H.K., Kim,Y., Lee,S., Min,S., Bae,J.Y., Choi,J.W., Park,J., Jung,D., Yoon,S. and Kim,H.H. (2019) SpCas9 activity prediction by DeepSpCas9, a deep learning–based model with high generalization performance. *Sci. Adv.*, **5**, eaax9249.
30. Diatchenko,L., Lau,Y.F.C., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. *et al.* (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 6025–6030.
31. Dagdas,Y.S., Chen,J.S., Sternberg,S.H., Doudna,J.A. and Yildiz,A. (2017) A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.*, **3**, eaao0027.
32. Fu,Y., Sander,J.D., Reyon,D., Cascio,V.M. and Joung,J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279–284.
33. Horlbeck,M.A., Gilbert,L.A., Villalta,J.E., Adamson,B., Pak,R.A., Chen,Y., Fields,A.P., Park,C.Y., Corn,J.E., Kampmann,M. *et al.* (2016) Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife*, **5**, e19760.
34. Gilbert,L.A., Horlbeck,M.A., Adamson,B., Villalta,J.E., Chen,Y., Whitehead,E.H., Guimaraes,C., Panning,B., Ploegh,H.L., Bassik,M.C. *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell*, **159**, 647–661.
35. Konermann,S., Brigham,M.D., Trevino,A.E., Joung,J., Abudayyeh,O.O., Barcena,C., Hsu,P.D., Habib,N., Gootenberg,J.S., Nishimasu,H. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583–588.
36. Chen,S., Sanjana,N.E., Zheng,K., Shalem,O., Lee,K., Shi,X., Scott,D.A., Song,J., Pan,J.Q., Weissleder,R. *et al.* (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, **160**, 1246–1260.