

A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism

Hiroshi Arakawa

The clustered regularly interspersed palindromic repeats (CRISPR)/Cas9 (CRISPR-associated protein 9) system is a powerful tool for genome editing that can be used to construct a guide RNA (gRNA) library for genetic screening. For gRNA design, one must know the sequence of the 20-mer flanking the protospacer adjacent motif (PAM), which seriously impedes experimentally making gRNA. I describe a method to construct a gRNA library via molecular biology techniques without relying on bioinformatics. Briefly, one synthesizes complementary DNA from the mRNA sequence using a semi-random primer containing a PAM complementary sequence and then cuts out the 20-mer adjacent to the PAM using type IIS and type III restriction enzymes to create a gRNA library. The described approach does not require prior knowledge about the target DNA sequences, making it applicable to any species.

INTRODUCTION

The clustered regularly interspersed palindromic repeats (CRISPR) system is responsible for the acquired immunity of bacteria (1), which is shared among 40% of eubacteria and 90% of archaea (2). When bacteria are attacked by infectious agents, such as phages or plasmids, a subpopulation of the bacteria incorporates segments of the infectious DNA into a CRISPR locus as a memory of the bacterial adaptive immune system (1). If the bacteria are infected with the same pathogen, short RNA transcribed from the CRISPR locus is integrated into CRISPR-associated protein 9 (Cas9), which acts as a sequence-specific endonuclease and eliminates the infectious pathogen (3).

CRISPR/Cas9 is available as a sequence-specific endonuclease (4, 5) that can cleave any locus of the genome if a guide RNA (gRNA) is provided. Indels on the genomic loci generated by nonhomologous end joining (NHEJ) can knock out the corresponding gene (4, 5). By designing gRNA for the gene of interest, individual genes can be knocked out one by one (reverse genetics); however, this strategy is not helpful when the gene responsible for the phenomenon of interest is not identified. If a proper readout and selection method is available, phenotype screening (forward genetics) is an attractive alternative.

Recently, genome-scale pooled gRNA libraries have been applied for forward genetics screening in mammals (6–9). Whereas phenotypic screening depends on the experimental setup, the most straightforward method is screening based on the viability of mutant cell lines that are combined with either positive or negative selection. Negative selection screens for human gRNA libraries have identified essential gene sets involved in fundamental processes (6–8). Screens for resistance to nucleotide analogs or anticancer drugs successfully identified previously validated genes as well as novel targets (6–8). Thus, Cas9/gRNA screening is a powerful tool for systematic genetic analysis in mammalian cells.

The gRNA for *Streptococcus pyogenes* Cas9 can be designed as a 20-base pair (bp) sequence that is adjacent to the protospacer adjacent motif (PAM) NGG (4, 5). This sequence can usually be identified from the coding sequence or locus of interest by bioinformatics techniques, but this approach is difficult for species with poorly annotated genetic information. Despite current advances in genome bioinformatics,

annotation of the genetic information is incomplete in most species, except for well-established model organisms such as human, mouse, or yeast. Although the diversity of species represents a diversity of special biological abilities, according to the organism, many of the genes encoding special abilities in a variety of species are left untouched, leaving an untapped gold mine of genetic information. Nevertheless, species-specific abilities are certainly beneficial because of possible transplantation in humans or applications for medical research.

If one wants to convert the mRNA into gRNA without prior knowledge of the target DNA sequences, the major challenges are to find the sequences flanking the PAM and to cut out the 20-bp fragment. Here, I describe molecular biology techniques to convert mRNA into a gRNA library. This method does not rely on bioinformatics and opens a path for forward genetics screening of any species, independent of their genetic characterization.

RESULTS

A strategy to convert mRNA to guide sequences

How does one find the sequences flanking the PAM? A random primer is commonly used for complementary DNA (cDNA) synthesis; instead, I reasoned that a semi-random primer containing a PAM complementary sequence could be used as the cDNA synthesis primer instead of a random primer (Fig. 1A).

How does one cut out the 20-bp fragment? Type IIS or type III restriction enzymes cleave sequences separated from their recognition sequences. The type III restriction enzyme Eco P15I cleaves 25/27 bp away from its recognition site but requires a pair of inversely oriented recognition sites for efficient cleavage (10). The type IIS restriction enzyme Acu I cleaves 13/15 bp away from its recognition site. I have developed an approach that allows to cut out a 20-mer by carefully arranging the positions of these restriction sites (Fig. 1B).

gRNA library construction via molecular biology techniques

Using a semi-random primer (NCCNNN) that contained the PAM complementary CCN, I reverse-transcribed cDNA from poly(A) RNA of the chicken B cell line DT40^{Cre1} (Fig. 1C) (11, 12). At that time, the 5' SMART tag sequence containing the Eco P15I site was

2016 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC). 10.1126/sciadv.1600699

IFOM–FIRC Institute of Molecular Oncology Foundation, 20139 Milano, Italy. Email: hiroshi.arakawa@ifom.eu

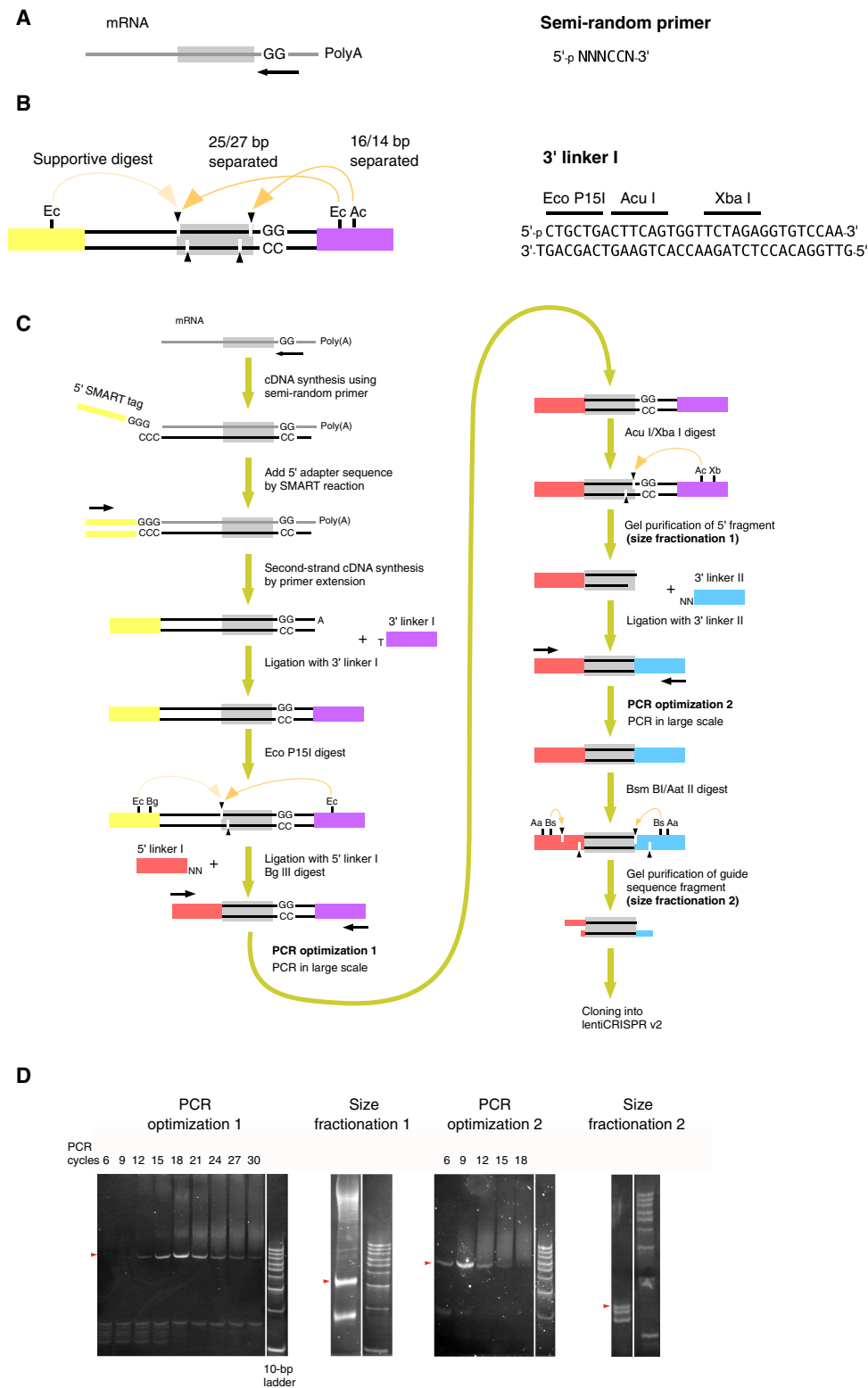


Fig. 1. gRNA library construction using a semi-random primer. (A) Semi-random primer. Poly(A), polyadenylate. (B) Type III and type IIS restriction sites to cut out the 20-bp guide sequence. Ec, Eco P15I; Ac, Acu I. (C) Scheme of gRNA library construction. Bg, Bgl II; Xb, Xba I; Bs, Bsm BI; Aa, Aat II. PCR, polymerase chain reaction; lentiCRISPR v2, lentiCRISPR version 2. (D) Short-range PCR for PCR cycle optimization and size fractionation of the guide sequence. PCR products were run on 20% polyacrylamide gels. A 10-bp ladder was used as the size marker. Bands of the expected sizes are marked by triangles.

added onto the 5' side by the switching mechanism at RNA transcript (SMART) method (13). The second strand of cDNA was synthesized by primer extension using a primer that annealed at the 5' SMART tag sequence with Advantage 2 PCR Polymerase, which generated A-overhang at the 3' terminus. This A-overhang was ligated with 3' linker I, which contains the Eco P15I and Acu I sites to cut out the guide sequence afterward. The double-stranded cDNA (ds cDNA) was digested with Eco P15I to remove the 5' SMART tag sequence and ligated with 5' linker I, which included a Bsm BI site, a cloning site for the gRNA expression vector. The DNA was then digested with Bgl II to destroy the 5' SMART tag backbone.

The gRNA library was amplified by PCR at this stage. To determine the optimal number of PCR cycles, I performed a titration between 6 and 30 cycles (Fig. 1D, PCR optimization 1). The expected PCR product, approximately 80 bp, was visible after 12 cycles; however, as the number of cycles increased, a larger, nonspecific smear appeared. Additionally, unnecessary cycle number increases may reduce the complexity of the library. Thus, PCR amplification was repeated on a large scale using the optimal PCR cycle number of around 17 cycles. The PCR product was subsequently digested with Acu I and Xba I and examined using 20% polyacrylamide gel electrophoresis. The 45-bp fragment was purified (Fig. 1D, size fractionation 1), ligated with the 3' linker II that included a Bsm BI cloning site, and used for the next PCR.

To determine the optimal PCR cycle number, additionally, I performed a titration between 6 and 18 PCR cycles (Fig. 1D, PCR optimization 2). PCR amplification was repeated on a large scale with the optimal number of nine PCR cycles. The PCR product was then digested with Bsm BI and Aat II. The restriction digest generated a 25-bp fragment, as well as 24- and 23-bp fragments (Fig. 1D, size fractionation 2), which were likely generated because of the inaccurate breakpoints of type IIS and type III restriction enzymes (14); careful purification of the 25-bp fragment minimized the possible problems with those artifacts. The guide sequence insert library, generated as described above, was finally cloned into a Bsm BI-digested lentiCRISPR v2 (15) vector and then electroporated into Stbl4 electrocompetent cells.

Guide sequences in the gRNA library

Plasmid DNA was purified from the generated gRNA library by maxiprep. Initially, the DNA was sequenced as a mixed plasmid population. A highly complexed and heterogeneous sequence was observed in the lentiCRISPR v2 cloning site between the U6 promoter and the gRNA scaffold (Fig. 2A), indicating that (i) no-insert clones are rare, (ii) cloned guide sequences are highly complexed, and (iii) most of the guide sequences are 20 bp long. After retransformation of the library in bacteria, a total of 236 bacterial clones were randomly picked and used for plasmid miniprep and sequencing.

As shown in the example of sequencing for 12 random clones (Fig. 2B), the cloned guide sequences were heterogeneous; these guide sequences were subsequently analyzed using the National Center for Biotechnology Information Basic Local Alignment Search Tool (BLAST). As shown in Fig. 2C, typically one gene was hit by each guide sequence. A PAM was identified adjacent to the guide sequence. For more than three-quarters of the guide sequences, the original genes from which those guides were generated were identified using BLAST. Most of these guide sequences were derived from single genes.

Notably, three of the guide sequences among the 236 plasmid clones were derived from different positions adjacent to the PAMs

on the *immunoglobulin (Ig) heavy chain C μ* gene (Fig. 2D). Thus, multiple guide sequences were generated from the same gene. Unexpectedly, the reversed-orientation guide sequences, like C μ guide 3 (Fig. 2D), were also observed at a relatively low frequency (~10%) (table S1). However, most of these were accompanied by a PAM (table S1). PAM priming might have worked even from the first-strand cDNA and not only from the mRNA. These reversed guide sequences are expected to work in genome cleavage, contributing to the knockout library.

The cloning of the guide sequences was efficient (100%), and most guide sequences (89%) were 20 bp long (Fig. 2E and table S1). Whereas 66% of the insert sequences were derived from mRNA, 11% of the insert sequences were derived from ribosomal RNA (rRNA), and 23% were from unknown origins, possibly derived from unannotated genes (Fig. 2E). Ninety-one percent of the guide sequences with identified origins were accompanied by PAMs, confirming that PAM priming using the semi-random primer functioned as intended. In addition, PAMs were also found near most of the remaining guide sequences (7%), but separated by 1 bp (Fig. 2E). This is most likely due to the inaccurate breakpoints of Acu I, because the length of those guide sequences was often 19 bp.

Functional validation of guide sequences

Three guide sequences specific to C μ (Fig. 2D) were further tested to functionally validate the guide sequences in the library. These lentiviral clones were transduced into the AID^{-/-} DT40 cell line, which constitutively expresses cell surface IgM (sIgM) because of the absence of Ig gene conversion (12). C μ guides 1, 2, and 3 generated 5.9, 11.7, and 9.2% sIgM (-) populations 2 weeks after transduction, as estimated by flow cytometry analysis (Fig. 3, upper panels), and these sIgM (-) populations were further isolated by fluorescence-activated cell sorting (FACS). Because the Ig heavy chain genomic locus is poorly characterized and only the rearranged VDJ allele is transcribed, its cDNA, rather than its genomic locus, was analyzed by Sanger sequencing. Sequencing analysis of about 30 IgM cDNA-containing plasmid clones for each sorted sIgM (-) population clarified the insertions, deletions, and mutations on the locus (Fig. 3, bottom). Most of the indels were focused around the guide sequences. Relatively large deletions observed on the cDNA sequence indicate that the clones in the library can sometimes cause even large functional deletions in the corresponding transcripts.

Deep characterization of the gRNA library

To characterize the complexity of the gRNA library, the library was deep-sequenced using Illumina MiSeq and analyzed by an RNA sequencing (RNA-seq) protocol using the Ensembl chicken genome database (16) as a reference. For example, approximately 500,000 of the guide sequences were mapped to chromosome 1, suggesting robust generation of guide sequences from various loci in the genome. Although the Ensembl database includes 15,916 chicken genes, the number of annotated chicken genes appears to be at least 4000 less than those in other established genetic model vertebrates, such as humans, mice, and zebrafish (16). Among the 5,209,083 sequence reads, 4,052,174 reads (77.8%) were mapped to chicken genes, and most of those sequences were accompanied by PAM (Fig. 4B). Nevertheless, one-quarter of the unmapped reads could be due to the relatively poor genetic annotation of the chicken genome, which again emphasizes the limitations of bioinformatics approaches for specific species. The

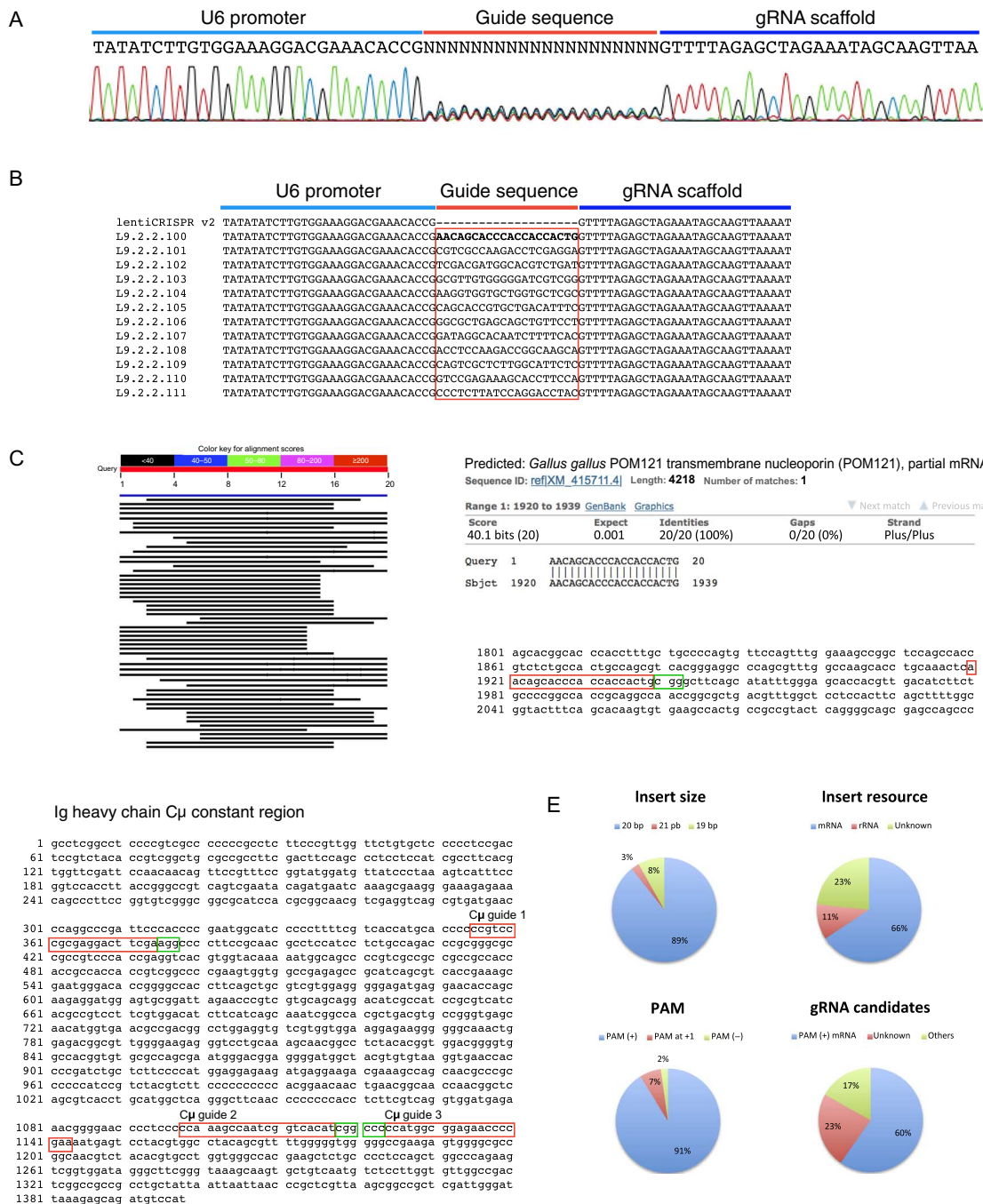


Fig. 2. Guide sequences in the gRNA library. (A) Mass sequencing of the gRNA library. (B) An example of sequencing for 12 random clones. (C) An example of the BLAST search analysis of a guide sequence. The first guide sequence clone in Fig. 2A is shown as an example. A 20-bp guide sequence (red frame) is accompanied by a PAM (green frame). (D) Three different guide sequences derived from the same gene, the *Ig heavy chain Cμ* gene. (E) Features of the gRNA library. Percentages in the PAM graph were calculated among the guide sequences where their origins were identified. "Others" in the gRNA candidate graph indicates the sum of guide sequences of rRNA and PAM (-) mRNA.

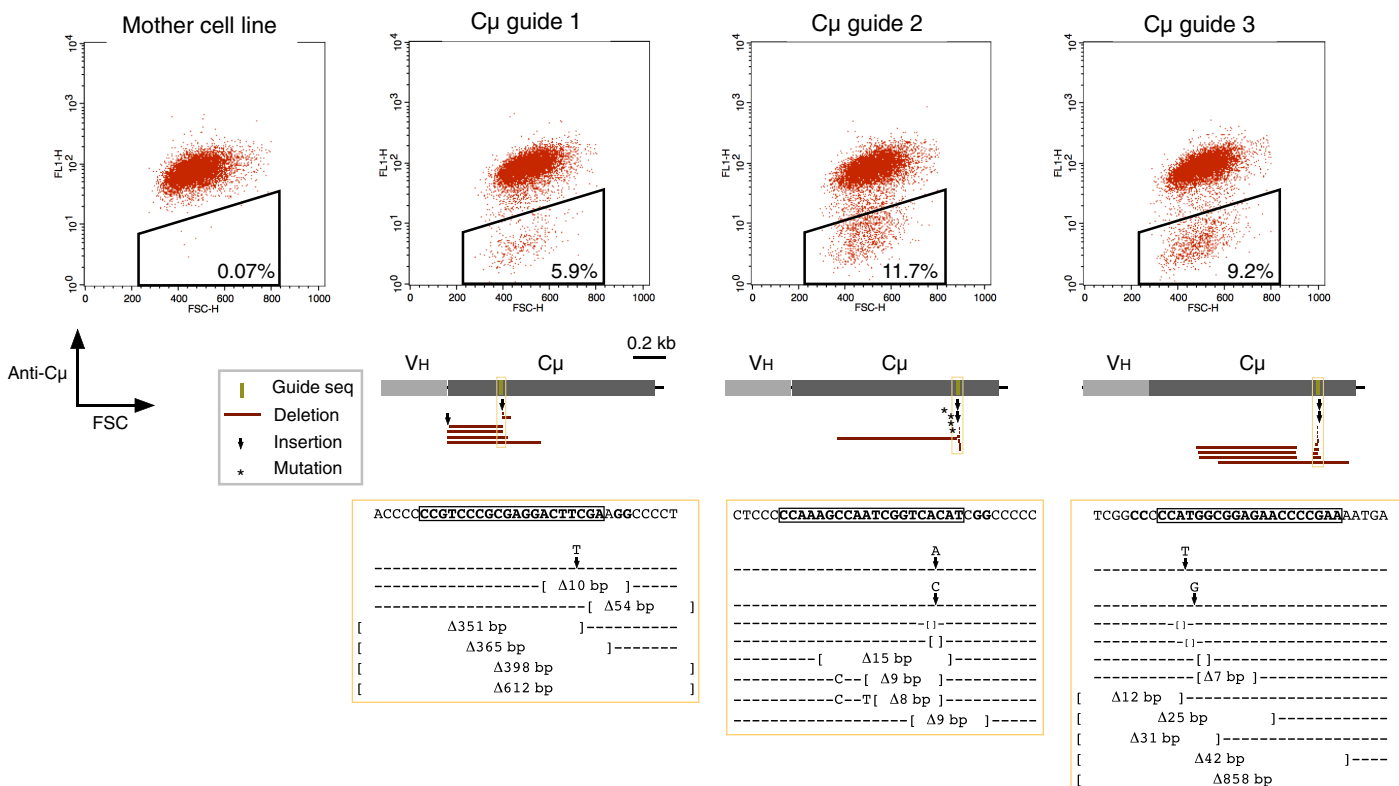


Fig. 3. Functional validation of guide sequences. Three lentiviral clones specific to C_μ (C_μ guides 1, 2, and 3 in Fig. 2D) were transduced into the AID^{-/-} sIgM (-) DT40 cell line. FACS profiles 2 weeks after transduction are shown with sIgM (-) gatings, which were used for FACS sorting (upper panels). The cDNA of the *IgM* gene from the sorted sIgM (-) cells is mapped together with the position of guide sequences, insertions, deletions, and mutations (lower panels). Detailed cDNA sequences around the guide sequences are shown at the bottom.

average length of guide sequence reads was 19.9 bp. Although 2.0% of the guide sequences that mapped to exon/exon junctions appeared as nonfunctional, 3,936,069 (75.6%) of the guide sequences, including 2,626,362 different guide sequences, were considered as functional. Guide sequences were generated even from genes with low-expression levels, covering 91.8% of the annotated genes (14,617 of 15,916) (Fig. 4B, heatmap). Whereas two or more unique guide sequences were identified for 97.8% of those genes, more than 100 different guide sequence species were identified for 46.0% of these genes (Fig. 4B, circle graph). Thus, the gRNA library appeared to have sufficient diversity for genetic screening.

Functional validation of the gRNA library

The transduction of the library into the AID^{-/-} DT40 cell line induced a significant sIgM (-) population (0.3%) (Fig. 4C, left) compared to the mother cell line (Fig. 3, left). This sIgM (-) population was further enriched 100-fold by FACS sorting, and their guide sequences were analyzed by deep sequencing. Unexpectedly, contaminated sIgM (+) cells appeared to expand more rapidly than sIgM (-) cells, possibly because of B cell receptor signaling, leading to incomplete enrichment of sIgM (-) cells. Nevertheless, because IgM-specific guide sequences achieved the second highest score of sequence reads in the sorted sIgM (-) population (Fig. 4C, right), IgM-specific guide sequences were obviously enriched after sIgM (-) sorting (Fig. 4D, left). Whereas 224 of the unique guide sequences specific to IgM were identified in the plas-

mid library, a few of these guide sequences were highly increased in the sorted sIgM (-) population (Fig. 4D, right). Sanger sequencing of 29 plasmid clones of the IgM cDNA from the sorted sIgM (-) population independently identified four deletions and one mutation (Fig. 4E). Three large deletions were likely generated by alternative NHEJ via microhomology, and one appeared to be generated by missplicing, possibly because of the indels around splicing signals. Therefore, the library can be used to screen knockout clones once the proper screening method is available.

DISCUSSION

Together, a diverse and functional gRNA library was successfully generated using the described method. The generated gRNA library is a specialized short cDNA library and is, therefore, also useful as a customized gRNA library specific to organs or cell lines.

Recently, the construction of a gRNA library using molecular biology techniques has been reported by other groups. Cheng *et al.* (17) developed a Molecular Chipper technology to generate dense gRNA libraries for genomic regions of interest using a type III restriction enzyme, and they identified novel cis-regulatory domains for microRNA-142 biogenesis in a proof-of-principle screen. Lane *et al.* (18) developed an elegant approach using PAM-like restriction enzymes to generate guide libraries, which can label chromosomal loci in

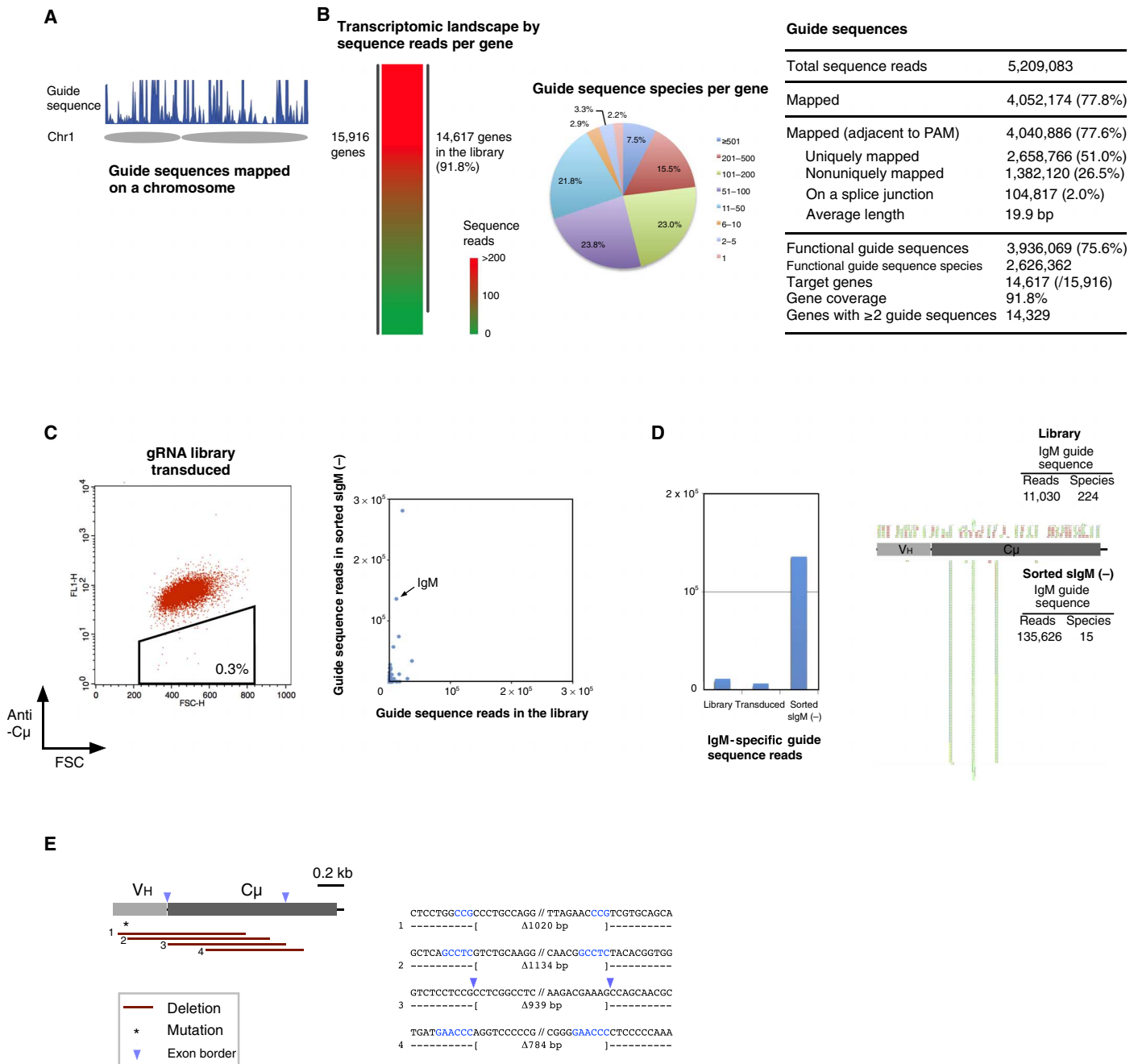


Fig. 4. Characterization and functional validation of the gRNA library. (A) Distribution of guide sequences on a chromosome. (B) Diversity of the gRNA library. Sequence reads per gene reflecting the transcriptomic landscape of the guide sequences (heatmap; shown with a scale bar). Guide sequence species per gene (circle graph). (C) Lentiviral transduction of gRNA library. Left: A FACS profile 2 weeks after transduction is shown with the slgM (-) gating, which was used for FACS sorting. Right: The graph shows the total sequence reads in the library versus those in the sorted slgM (-). Each dot represents a different gene. (D) IgM-specific guide sequences. Sequence reads specific to IgM (graph). Guide sequences mapped on IgM cDNA (map). Green and red bars represent guide sequences with forward and reverse orientations, respectively. (E) Deletions in the IgM cDNA in sorted slgM (-). Left: The cDNA of the IgM gene from sorted slgM (-) cells is shown, as well as the position of guide sequences, deletions, mutations, and exon borders. Right: Detailed sequences around breakpoints. Microhomologies in the reference sequences are highlighted in blue.

Xenopus egg extracts or can target the *Escherichia coli* genome at high frequency.

Here, I generated a gRNA library for a higher eukaryotic transcriptome using molecular biology techniques. To my knowledge, this is the first gRNA library created from mRNA and the first library created from a poorly genetically characterized species. The semi-random primer can potentially target any NGG on mRNA, generating a highly complexed gRNA library that covers more than 90% of the annotated genes (Fig. 4B). Furthermore, the method described here could be applied to CRISPR systems in organisms other than *S. pyogenes* by customizing the semi-random primer.

Multiple guide sequences were efficiently generated from the same gene (Figs. 2D and 4, B and D), like the native CRISPR system in bacteria (1); this is an important advantage of the developed method. Although each guide sequence may differ in genome cleavage efficiency for each target gene, relatively more efficient guide sequences for each gene are included in the library (Fig. 4D).

Because the gRNA library created here is on a B cell transcriptomic scale rather than a genome scale, guide sequences will not be generated from nontranscribed genes. Guide sequences were more frequently generated from abundantly transcribed mRNAs but less frequently generated from rare mRNAs (Fig. 4B). By combining the techniques of a normalized library, in which one normalizes the amount of mRNA for each gene, it is possible to increase the frequency of guide sequences generated from rare mRNA (19). If the promoters in the lentiCRISPR v2 for Cas9 or gRNA expression are replaced with optimal promoters for each cell type or species, this will further improve the transduction or knockout efficiency of the gRNA library.

Guide sequences can be generated not only from the coding sequence but also from the 5' and 3' untranslated regions (UTRs). Because gRNA from UTRs will not cause indels within the coding sequence, gRNAs are not usually designed on UTRs to knock out genes; however, because several key features, such as mRNA stability or translation control, are determined by regulatory sequences located in the UTRs, indels occurring in these areas can lead to the unexpected elucidation of the gene's function. In this regard, this method can be also usefully applied for species such as human, whose large-scale gRNA libraries are already constructed (6–8). It can be also useful to make personalized human gRNA libraries, which represent collections of single-nucleotide polymorphisms from different exons. These personalized human gRNA libraries could be used to study allelic variations and their phenotypes, leading to better characterizations of rare diseases.

Approximately 23% of the guide sequences were derived from unknown origins (Figs. 2E and 4B). These sequences may be, at least partly, derived from mRNA with insufficient genetic annotation. This is the greatest advantage of the developed method: the sum of these “unknown” sequences and PAM⁺ mRNA cover 83% of the library and are expected guide sequence candidates available for genetic screening (Fig. 2E). Because this method is not based on bioinformatics, it is possible to create guide sequences even from unknown genetic information. This bioinformatics-independent approach is obviously advantageous to species with insufficient genetic analysis.

Some cell type- or species-specific biological properties may be driven by uncharacterized or unannotated genes. For example, I suspect that these unknown genes may play a key role in Ig gene conversion (20) or hypertargeted integration (21) in chicken B cells. Moreover, many “minor” organisms exist that have not been used as genetic models despite their unique biological characteristics, for example,

planaria with extraordinary regeneration ability (22), naked mole rats with cancer resistance (23), and red sea urchins with their 200-year life span (24). Knockout libraries can be important genetic tools to shed light on genetic backgrounds with unique biological properties. Using this technique, it is possible to create a gRNA library, even from species with poorly annotated genetic information; some “forgotten” species may be converted into attractive genetic models by this technology.

MATERIALS AND METHODS

Preparation of RNA

Total RNA was prepared from DT40^{Cre1} cells (11, 12) using TRIzol reagent (Invitrogen). Poly(A) RNA was prepared from DT40^{Cre1} total RNA using the Oligotex mRNA Mini Kit (Qiagen). To enrich mRNA, hybridization of poly(A)⁺ RNA and washing with buffer OBB (from the Oligotex kit) were repeated twice, according to the stringent wash protocol from the manufacturer's recommendations.

Oligonucleotides

The following oligonucleotides were used: semi-random primer, p NNNCCN; 5' SMART tag, TGGTCAAGCTTCAGCAGATCTACACGGACGTCGrGrGrG; 5' SMART PCR primer, TGGTCAAGCTTCAGCAGATCTACACG; 3' linker I forward, p CTGCTGACTTCAGTGGTTCTAGAGGTGTCCAA; 3' linker I reverse, GTTGACACCTCTAGAACCCTGAAGTCAGCAGT; 5' linker I forward, GCATATAAGCTTGACGTCTCTCACC; 5' linker I reverse, p NNCGGTGAGAGACGTCAAGCTTATATGC; 3' linker II forward, p GTTTGGAGACGTCTTCTAGATCAGCG; 3' linker II reverse, CGCTGATCTAGAAGACGTCTCCAAACNN; 3' linker I PCR primer, GTTGGACACCTCTAGAACCCTGAAGTCAGCAGTNNCC; 3' linker II PCR primer, CGCTGATCTAGAAGACGTCTCCAAAC; sequencing primer, TTTTCGGGTTTATTACAGGGACGACAG; lentiCRISPR forward, CTTGGCTTTATATCTTTGTGGAAAGGACG; lentiCRISPR reverse, CGGACTAGCCTTATTTAACTTGCTATTTCTAG; universal forward, AGCGATAACAATTTACACAGGA; universal reverse, CGCCAGGGTTTTCCCAGTCACGAC; Ig heavy chain 1, CCGCAACCAAGCTTATGAGCCCACTCGTCTCCTCCCTCC; Ig heavy chain 2, CGTCCATCTAGAATGGACATCTGCTCTTTAATCCCAATC-GAG; Ig heavy chain 3, GCTGAACAACCTCAGGGCTGAGGACACC; Ig heavy chain 4, AGCAACGCCCGCCCCCATCCGTCTACGTCTT.

Linker preparation

The following reagents were combined in a 1.5-ml microcentrifuge tube: 10 µl of 100 µM linker forward oligo, 10 µl of 100 µM linker reverse oligo, and 2.2 µl of 10× T4 DNA Ligase Buffer [New England Biolabs (NEB)]. The tubes were placed in a water bath containing 2 liters of boiled water and were incubated as the water cooled naturally. The annealed oligos were diluted with 77.8 µl of TE buffer (pH 8.0) and used as 10 µM linkers.

gRNA library construction

First-strand cDNA synthesis. The following reagents were combined in a 0.2-ml PCR tube: 200 ng of DT40^{Cre1} poly(A) RNA, 0.6 µl of 25 µM semi-random primer, and ribonuclease (RNase)-free water

in a 4.75- μ l volume. The tube was incubated at 72°C in a hot-lid thermal cycler for 3 min, cooled on ice for 2 min, and further incubated at 25°C for 10 min. The temperature was then increased to 42°C, and a 5.25- μ l mixture containing the following reagents was added: 0.5 μ l of 25 μ M 5' SMART tag, 2 μ l of 5 \times SMARTScribe buffer, 0.25 μ l of 100 mM dithiothreitol, 1 μ l of 10 mM deoxynucleotide triphosphate (dNTP) mix, 0.5 μ l of RNaseOUT (Invitrogen), and 1 μ l of SMARTScribe Reverse Transcriptase (100 U) (Clontech). The first-strand cDNA reaction mixture was incubated at 42°C for 90 min and then at 68°C for 10 min. To degrade RNA, 1 μ l of RNase H (Invitrogen) was added to the mixture, and then the mixture was incubated at 37°C for 20 min.

ds cDNA synthesis by primer extension. Eleven microliters of prepared first-strand poly(A) cDNA was mixed with 74 μ l of Milli-Q water, 10 μ l of 10 \times Advantage 2 PCR Buffer, 2 μ l of 10 mM dNTP mix, 1 μ l of 25 μ M 5' SMART PCR primer, and 2 μ l of 50 \times Advantage 2 Polymerase Mix (Clontech). A 100- μ l volume of the reaction mixture for primer extension was incubated at 95°C for 1 min, 68°C for 20 min, and then 70°C for 10 min. The prepared ds cDNA was purified using QIAquick PCR Purification Kit (Qiagen) and was eluted with 40 μ l of TE buffer (pH 8.0).

3' linker I ligation. DT40^{Cre1} double-stranded poly(A) cDNA was mixed with 0.5 μ l of 10 μ M 3' linker I and 1 μ l of Quick T4 DNA Ligase (NEB) in 1 \times Quick Ligation Buffer. The ligation reaction mixture was incubated at room temperature for 15 min, then purified using QIAquick PCR Purification Kit, and eluted with 80 μ l of TE buffer.

Eco P15I digestion. The 3' linker I-ligated DNA was digested with 1 μ l of Eco P15I (10 U/ μ l; NEB) in 1 \times NEBuffer 3.1 containing 1 \times adenosine 5'-triphosphate in a 100- μ l volume at 37°C overnight. The Eco P15I-digested DNA was purified using QIAquick PCR Purification Kit and eluted with 40 μ l of TE buffer.

5' linker I ligation and Bgl II digestion. The digested DNA was mixed with 0.5 μ l of 10 μ M 5' linker I and 1 μ l of Quick T4 DNA Ligase (NEB) in 1 \times Quick Ligation Buffer. The ligation reaction mixture was incubated at room temperature for 15 min, purified using QIAquick PCR Purification Kit, and eluted with 80 μ l of TE buffer. The DNA was further digested with 1 μ l of Bgl II (10 U/ μ l; NEB) in 1 \times NEBuffer 3.1 in a 100- μ l volume at 37°C for 3 hours. The Eco P15I/Bgl II-digested DNA was purified using QIAquick PCR Purification Kit and eluted with 50 μ l of TE buffer.

First PCR optimization. To determine the optimal number of PCR cycles, a 0.2-ml PCR tube was prepared containing 5 μ l of ds cDNA ligated with 5' linker I/3' linker I, 0.5 μ l of 25 μ M 5' linker I forward primer, 0.5 μ l of 25 μ M 3' linker I PCR primer, 5 μ l of 1 \times Advantage 2 PCR Buffer, 1 μ l of 10 mM dNTP mix, 1 μ l of 50 \times Advantage 2 Polymerase Mix, and Milli-Q water in a 50- μ l volume. PCR was carried out with the following cycling parameters: six cycles at 98°C for 10 s and 68°C for 10 s. After the six cycles, 5 μ l of the reaction was transferred to a clean microcentrifuge tube. The rest of the PCR reaction mixture underwent three additional cycles at 98°C for 10 s and 68°C for 10 s. After these additional three cycles, 5 μ l was transferred to a clean microcentrifuge tube. In the same way, additional PCR was repeated until reaching 30 total cycles. Thus, a series of PCRs of 6, 9, 12, 15, 18, 21, 24, 27, and 30 cycles were prepared and analyzed by 20% polyacrylamide gel electrophoresis to compare the band patterns. The optimal number of PCR cycles was determined as the minimal number of PCR cycles yielding the greatest quantity of the 84-bp product (typically around 17 cycles). Two 50- μ l PCRs were repeated with the optimal number of PCR cycles. The PCR product

was purified using the QIAquick PCR Purification Kit and eluted with 50 μ l of TE buffer.

Acu I/Xba I digestion. The PCR product was digested with 2 μ l of Acu I (5 U/ μ l; NEB) and 2 μ l of Xba I (20 U/ μ l; NEB) in 1 \times CutSmart Buffer containing 40 μ M S-adenosylmethionine in a 60- μ l volume at 37°C overnight. The Acu I/Xba I-digested DNA was run on a 20% polyacrylamide gel. The 45-bp fragment was cut out of the gel, purified by the crush and soak procedure, and dissolved into 20 μ l of TE buffer.

3' linker II ligation. The digested DNA was mixed with 2 μ l of 10 μ M 3' linker II and 1 μ l of Quick T4 DNA Ligase (NEB) in 1 \times Quick Ligation Buffer. The ligation reaction mixture was incubated at room temperature for 15 min, purified using QIAquick PCR Purification Kit, and eluted with 100 μ l of TE buffer.

Second PCR optimization. To determine the optimal number of PCR cycles, a 0.2-ml PCR tube was prepared, containing 5 μ l of ds cDNA ligated with 5' linker I/3' linker II, 0.5 μ l of 25 μ M 5' linker I forward primer, 0.5 μ l of 25 μ M 3' linker II PCR primer, 5 μ l of 1 \times Advantage 2 PCR Buffer, 1 μ l of 10 mM dNTP mix, 1 μ l of 50 \times Advantage 2 Polymerase Mix, and Milli-Q water in a 50- μ l volume. PCR was carried out with the following cycling parameters: six cycles at 98°C for 10 s and 68°C for 10 s. After the six cycles, 5 μ l of the reaction was transferred to a clean microcentrifuge tube. The rest of the PCR reaction mixture underwent an additional three cycles at 98°C for 10 s and 68°C for 10 s. After these additional three cycles, 5 μ l of the reaction was transferred to a clean microcentrifuge tube. In the same way, additional PCR cycles were repeated until 18 total cycles were reached. Thus, a series of PCR reactions of 6, 9, 12, 15, and 18 cycles were prepared and analyzed by 20% polyacrylamide gel electrophoresis to compare the band patterns. The optimal number of PCR cycles was determined as the minimal number of PCR cycles yielding the greatest quantity of the 72-bp product (typically around nine cycles). Five PCRs, each containing 50 μ l, were repeated with the optimal number of PCR cycles. The PCR product was purified using QIAquick PCR Purification Kit and eluted with 100 μ l of TE buffer.

Bsm BI/Aat II digestion. The PCR product was digested with 10 μ l of Bsm BI (10 U/ μ l; NEB) in 1 \times NEBuffer 3.1 in a 100- μ l volume at 55°C for 6 hours, and then 5 μ l of Aat II (20 U/ μ l; NEB) was added to the solution, which was left at 37°C overnight. The Bsm BI/Aat II-digested DNA was run on a 20% polyacrylamide gel. Typically, three bands, corresponding to 25, 24, and 23 bp, were visible. The 25-bp fragment was cut out of the gel, purified by the crush and soak procedure, and dissolved into 50 μ l of TE buffer. The concentration of the purified DNA was measured by Qubit dsDNA HS Assay Kit (Life Technologies).

Cloning. The lentiCRISPR v2 (Addgene) (15) was digested with Bsm BI, treated with calf intestine phosphatase, extracted with phenol/chloroform, and purified by ethanol precipitation. Five nanograms of the purified 25-bp guide sequence fragment was mixed with 3 μ g of lentiCRISPR v2 and 1 μ l of Quick T4 DNA Ligase (NEB) in 1 \times Quick Ligation Buffer in a 40- μ l volume. The ligation reaction mixture was incubated at room temperature for 15 min and then purified by ethanol precipitation. The prepared gRNA library was electroporated into Stbl4 electrocompetent cells (Invitrogen) using the following electroporator conditions: 1200 V, 25 μ F, and 200 ohms.

Sequencing and sequence analysis

Plasmid DNA was purified using the Wizard Plus SV Minipreps DNA Purification System (Promega) from 236 of randomly selected

clones from the gRNA library, in accordance with the manufacturer's protocol. The guide sequence clones were sequenced with the sequencing primer using Model 373 Automated DNA Sequencer (Applied Biosystems). The cloned guide sequences were compared with the GenBank database using BLAST.

Tips to avoid background noise in the gRNA library

During the setup of the methodology for gRNA library construction, rRNA contamination was observed in poly(A) RNA that was purified using an oligo_{dT} column, and rRNA-originated guide sequences sometimes occupied 40 to 50% of the total original library. Because rRNA occupies more than 90% of the intracellular RNA, generally speaking, it is hard to avoid having some rRNA contamination. The stringent wash protocol for poly(A) RNA purification successfully reduced the rRNA-derived guide sequences to around 10%. PCR artifacts amplifying the linker sequences were also observed during the setup of the methodology. For this reason, the linker sequence was designed with additional restriction sites, namely, Bgl II for the 5' SMART tag, Xba I for the 3' linker I, and Aat II for the 5' linker I and 3' linker II. By cutting with these additional restriction enzymes, it was possible to remove most of the PCR artifacts amplifying the linker sequences. The Bsm BI restriction digest of the final PCR generated the right size of the DNA fragment (25 bp) in addition to 1- or 2-bp shorter, unexpected DNA fragments. These shorter DNA fragments probably resulted from the inaccuracy of the cleavage position of type III and type IIS restriction enzymes. After Bsm BI cleavage, it was possible to minimize shorter DNA artifacts by carefully purifying the 25-bp fragment with a 20% polyacrylamide gel.

Lentiviral vectors

LentiCRISPR v2 (15) was a gift from F. Zhang (Addgene plasmid #52961). pCMV-VSV-G (25) was a gift from B. Weinberg (Addgene plasmid #8454). psPAX2 was a gift from D. Trono (Addgene plasmid #12260).

Lentiviral packaging

To produce lentivirus, a T-225 flask of human embryonic kidney 293T cells was seeded at ~40% confluence the day before transfection in D10 medium (Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum). One hour before transfection, the medium was removed and 13 ml of prewarmed Opti-MEM Reduced Serum Medium (Life Technologies) was added to the flask. Transfection was performed using Lipofectamine 2000 (Life Technologies). Twenty micrograms of gRNA plasmid library, 10 µg of pCMV-VSV-G (Addgene) (25), and 15 µg of psPAX2 (Addgene) were mixed with 4 ml of Opti-MEM (Life Technologies). One hundred microliters of Lipofectamine 2000 was diluted in 4 ml of Opti-MEM, and this solution was, after 5 min, added to the mixture of DNA. The complete mixture was incubated for 20 min before being added to the cells. After overnight incubation, the medium was changed to 30 ml of D10. After 2 days, the medium was removed and centrifuged at 3000 rpm at 4°C for 10 min to pellet cell debris. The supernatant was filtered through a 0.45-µm low-protein binding membrane (Millipore Steriflip HV/polyvinylidene difluoride). The gRNA library virus was further enriched 100-fold by polyethylene glycol (PEG) precipitation.

Lentiviral vectors containing C_μ guide sequences were packaged as described above except for the following modifications. Five micrograms of C_μ guide lentiviral vectors was used instead of 20 µg of

the gRNA library. The experiment was done in a quarter scale concerning solutions or culture medium without changing incubation times. Plates (100 mm) were used for lentiviral packaging instead of a T-225 flask. C_μ gRNA virus was directly used for transduction without enrichment by PEG precipitation.

Lentiviral transduction

Cells were transduced with the gRNA library via spinfection. Briefly, 2×10^6 cells per well were plated into a 12-well plate in DT40 culture medium supplemented with polybrene (8 µg/ml; Sigma). Each well received either 1 ml of C_μ gRNA virus or 100 µl of 100-fold enriched gRNA library virus along with a no-transduction control. The 12-well plate was centrifuged at 2000 rpm for 2 hours at 37°C. Cells were incubated overnight, transferred to culture flasks containing DT40 culture medium, and then selected with puromycin (1 µg/ml).

Sorting of sIgM (-) population

The AID^{-/-} sIgM (+) cell line with or without lentiviral transduction was first stained with a monoclonal antibody to chicken C_μ (M1) (SouthernBiotech) and then with polyclonal fluorescein isothiocyanate-conjugated goat antibodies to mouse IgG (Fab)₂ (Sigma). The sIgM (-) population was sorted using FACSAria (BD Biosciences).

Cloning and sequencing of the Ig heavy chain gene

The sorted sIgM (-) cells were further expanded and used for total RNA and genomic DNA preparation. Total RNA was purified using TRIzol reagent (Invitrogen). Total RNA was reverse-transcribed using SuperScript III Reverse Transcriptase (Invitrogen) with oligo_{dT} primer according to the manufacturer's instructions. The *IgM heavy chain* gene was amplified from the total cDNA of the sorted sIgM (-) population with Ig heavy chain primers 1 and 2. PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB) with the following cycling parameters: 30 s of initial incubation at 98°C, 35 cycles at 98°C for 10 s and at 72°C for 2 min, and a final elongation step of 2 min at 72°C. The PCR product was purified using QIAquick Gel Extraction Kit (Qiagen), digested with Hind III (NEB) and Xba I (NEB), and cloned into the pUC119 plasmid vector. Approximately 30 plasmid clones for each sorted sIgM (-) population were sequenced using universal forward, reverse, and Ig heavy chain primers 3 and 4.

Deep sequencing

Genomic DNA of the transduced cell library or sorted sIgM (-) cells was purified using the Easy-DNA Kit (Invitrogen). Either 100 ng of lentiviral plasmid library or 1 µg of genomic DNA was used as the PCR template. The guide sequences were amplified with lentiCRISPR forward and reverse primers using Advantage 2 Polymerase (Clontech). PCR was carried out with the following cycling parameters: 15 cycles at 98°C for 10 s and 68°C for 10 s for plasmid DNA and 27 cycles at 98°C for 10 s and 68°C for 10 s for genomic DNA. The 100-bp PCR fragment containing the guide sequence was purified using QIAquick Gel Extraction Kit (Qiagen). The deep sequencing library was prepared using TruSeq Nano DNA Library Preparation Kit (Illumina) and deep-sequenced using MiSeq (Illumina).

Bioinformatics

FASTQ files demultiplexed by Illumina MiSeq were analyzed using the CLC Genomics Workbench (Qiagen). Briefly, the sequence reads

were trimmed to exclude vector backbone sequences and added with the PAM sequence NGG. The sequence reads before or after adding NGG were aligned with the Ensembl chicken genome database (16) using the RNA-seq analysis toolbox with the read mapping parameters optimized for comprehensive analysis. After alignment, duplicates were removed from the mapped sequence reads to identify different guide sequence species. Afterward, the guide sequence reads and species per gene were calculated from the numbers of sequence reads mapped on the annotated genes. Because *Ig* genes were not annotated in the Ensembl database, the cDNA sequence of the *IgM* gene of the AID knockout DT40 cell line was used as a reference for the mapping of guide sequences specific to *IgM*.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/2/8/e1600699/DC1>
table S1. Guide sequences.

REFERENCES AND NOTES

- Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Grissa, G. Vergnaud, C. Pourcel, The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172 (2007).
- J. E. Garneau, M.-. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán, S. Moineau, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
- L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
- T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
- Y. Zhou, S. Zhu, C. Cai, P. Yuan, C. Li, Y. Huang, W. Wei, High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
- H. Koike-Yusa, Y. Li, E.-P. Tan, M. Del Castillo Velasco-Herrera, K. Yusa, Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
- A. Meisel, T. A. Bickle, D. H. Krüger, C. Schroeder, Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**, 467–469 (1992).
- H. Arakawa, D. Lodygin, J.-M. Buerstedde, Mutant loxP vectors for selectable marker recycle and conditional knock-outs. *BMC Biotechnol.* **1**, 7 (2001).
- H. Arakawa, J. Hauschild, J.-M. Buerstedde, Requirement of the activation-induced deaminase (*AID*) gene for immunoglobulin gene conversion. *Science* **295**, 1301–1306 (2002).
- Y. Y. Zhu, E. M. Machleder, A. Chenchik, R. Li, P. D. Siebert, Reverse transcriptase template switching: A SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897 (2001).
- S. Lundin, A. Jemt, F. Terje-Hegge, N. Foam, E. Pettersson, M. Käller, V. Wirta, P. Lexow, J. Lundeberg, Endonuclease specificity and sequence dependence of type IIS restriction enzymes. *PLoS One* **10**, e0117059 (2015).
- N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
- International Chicken Genome Sequencing Consortium, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
- J. Cheng, C. A. Roden, W. Pan, S. Zhu, A. Baccei, X. Pan, T. Jiang, Y. Kluger, S. M. Weissman, S. Guo, R. A. Flavell, Y. Ding, J. Lu, A. Molecular Chipping technology for CRISPR sgRNA library generation and functional mapping of noncoding regions. *Nat. Commun.* **7**, 11178 (2016).
- A. B. Lane, M. Strzelecka, A. Ettinger, A. W. Grenfell, T. Wittmann, R. Heald, Enzymatically Generated CRISPR Libraries for Genome Labeling and Screening. *Dev. Cell* **34**, 373–378 (2015).
- S. R. Patanjali, S. Parimoo, S. M. Weissman, Construction of a uniform-abundance (normalized) cDNA library. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1943–1947 (1991).
- C.-A. Reynaud, V. Anquez, H. Grimal, J.-C. Weill, A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* **48**, 379–388 (1987).
- J.-M. Buerstedde, S. Takeda, Increased ratio of targeted to random integration after transfection of chicken B cell lines. *Cell* **67**, 179–188 (1991).
- Y. Umesono, J. Tasaki, Y. Nishimura, M. Hrouda, E. Kawaguchi, S. Yazawa, O. Nishimura, K. Hosoda, T. Inoue, K. Agata, The molecular logic for planarian regeneration along the anterior–posterior axis. *Nature* **500**, 73–76 (2013).
- X. Tian, J. Azpurua, C. Hine, A. Vaidya, M. Myakishev-Rempel, J. Ablaeva, Z. Mao, E. Nevo, V. Gorbunova, A. Seluanov, High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature* **499**, 346–349 (2013).
- T. A. Ebert, J. R. Southon, Red sea urchins (*Strongylocentrotus franciscanus*) can live over 100 years: Confirmation with A-bomb 14carbon. *Fish. Bull.* **101**, 915–922 (2003).
- S. A. Stewart, D. M. Dykxhoorn, D. Palliser, H. Mizuno, E. Y. Yu, D. S. An, D. M. Sabatini, I. S. Chen, W. C. Hahn, P. A. Sharp, R. A. Weinberg, C. D. Novina, Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).

Acknowledgments: I am grateful to I. Psakhye for critically reading the manuscript, to S. Minardi and M. Riboni for deep sequencing, to H. Kajihio for advice on the lentiviral experiments, to W. Carotenuto for helpful discussions regarding the bioinformatics, and to M. Foiani for his generous mentorship. **Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. **Author contributions:** H.A. conceived the project, designed the experiments, executed the experiments, analyzed the results, conducted the computational analysis, and wrote the manuscript. **Competing interests:** The author declares that he has no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the author.

Submitted 2 April 2016
Accepted 29 July 2016
Published 24 August 2016
10.1126/sciadv.1600699

Citation: H. Arakawa, A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism. *Sci. Adv.* **2**, e1600699 (2016).