

# SCIENTIFIC REPORTS

OPEN

## Synthesis of an arrayed sgRNA library targeting the human genome

Tobias Schmidt\*, Jonathan L. Schmid-Burgk\* &amp; Veit Hornung

Received: 08 March 2015

Accepted: 21 August 2015

Published: 08 October 2015

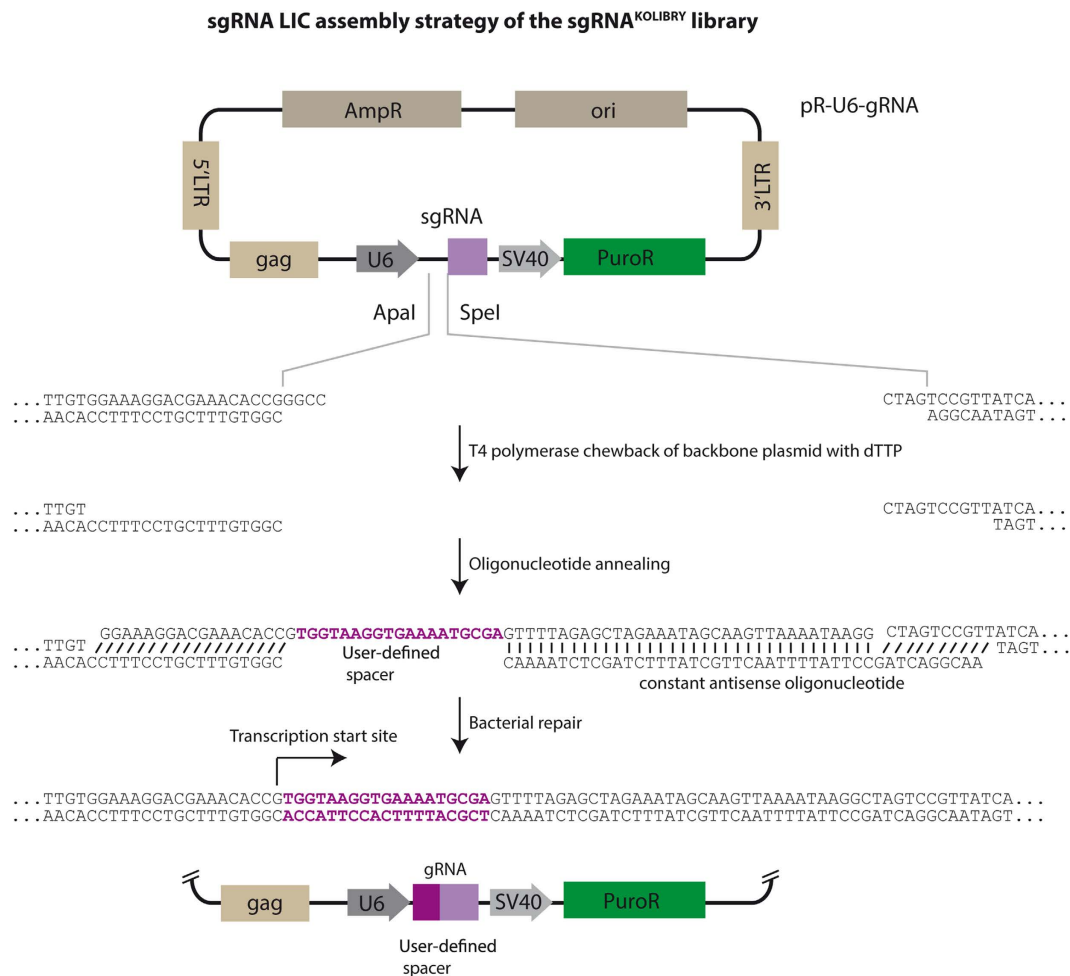
Clustered regularly interspaced short palindromic repeats (CRISPR) in conjunction with CRISPR-associated proteins (Cas) can be employed to introduce double strand breaks into mammalian genomes at user-defined loci. The endonuclease activity of the Cas complex can be targeted to a specific genomic region using a single guide RNA (sgRNA). We developed a ligation-independent cloning (LIC) assembly method for efficient and bias-free generation of large sgRNA libraries. Using this system, we performed an iterative shotgun cloning approach to generate an arrayed sgRNA library that targets one critical exon of almost every protein-coding human gene. An orthogonal mixing and deconvolution approach was used to obtain 19,506 unique sequence-validated sgRNAs (91.4% coverage). As tested in HEK 293T cells, constructs of this library have a median genome editing activity of 54.6% and employing sgRNAs of this library to generate knockout cells was successful for 19 out of 19 genes tested.

Clustered regularly interspaced short palindromic repeats (CRISPR) in conjunction with CRISPR-associated proteins (Cas) provide an adaptive immune system to bacteria and archaea targeting foreign genetic material<sup>1,2</sup>. Cas9, a member of the type II CRISPR-Cas system, requires RNA molecules to be directed as a sequence-specific endonuclease to its target sequence<sup>3</sup>. The native crRNA:tracrRNA complex can be mimicked by a chimeric, single guide RNA molecule (sgRNA), in which the invariant base pairing region is replaced by an intramolecular stem loop structure<sup>3</sup>. The target-specific protospacer region of the sgRNA usually consists of 20 nucleotides, in which the highest stringency in base pairing with a target site is demanded within the PAM adjacent region<sup>3</sup>. The CRISPR/Cas9 system can be reprogrammed in a user-defined fashion to induce double-strand breaks (DSBs) in prokaryotic<sup>3,4</sup> and eukaryotic genomes<sup>5–10</sup>.

To reveal novel components in biological pathways, the CRISPR/Cas9 system has recently been used for polyclonal forward genetic screens, in which CRISPR libraries were stably transduced followed by a phenotypical selection step<sup>11–13</sup>. Deep sequencing then allowed the identification of sgRNA sequences that were enriched in selected populations. In these studies PCR- and Gibson-assembly-based approaches were employed to create large polyclonal pools of sgRNA vectors from chip-synthesized oligonucleotides covering the entire human or mouse transcriptome.

Besides polyclonal screening efforts, CRISPR/Cas9 constitutes a promising tool for the fast and efficient generation of large numbers of knockout cells, e.g. comprising targets from entire pathways. To this end, arrayed sgRNA libraries encompassing the entire protein coding genome will constitute a powerful resource. Since the synthesis of sgRNAs using individual oligonucleotides is laborious and costly, the approach of generating arrayed libraries from chip-synthesized oligo pools is currently the most attractive choice. To this end, a method would be desirable that allows singling out clonal sgRNA constructs from polyclonal source material with minimal sequence bias. At the same time, an efficient sequence validation method is needed to allocate individual sgRNA clones in an arrayed library. To address these challenges, we developed an amplification-free LIC-based library assembly method combined with an orthogonal deep sequencing validation strategy to facilitate the synthesis of large, arrayed sgRNA libraries

Institute of Molecular Medicine, University Hospital, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to V.H. (email: veit.hornung@uni-bonn.de)



**Figure 1. LIC based assembly strategy.** A schematic view of the LIC-based sgRNA cloning system is depicted. In brief, a backbone plasmid containing constant portions of a hybrid sgRNA downstream the U6 promoter is first linearized using *ApaI* and *SpeI* restriction enzymes. Subsequently, a T4 DNA polymerase chewback reaction in the presence of dTTP sets free long 5' overhangs required for LIC. After annealing of the backbone plasmid to one constant as well as to one target-specific oligonucleotide, a 19 bp stretch remains single-stranded. Upon transformation into *E. coli*, a perfectly repaired, amplification-competent circular plasmid is obtained, containing the desired hybrid sgRNA sequence.

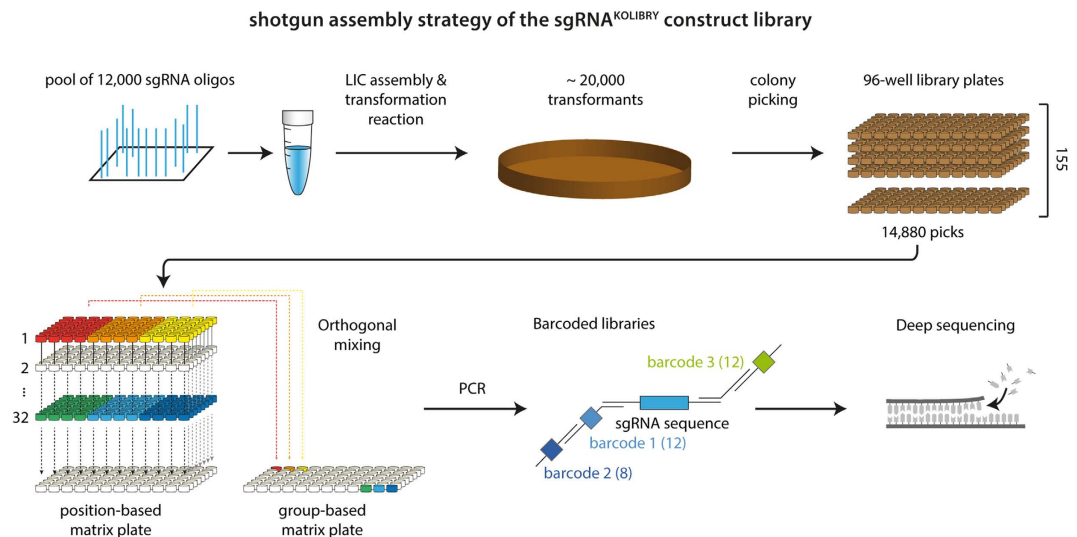
from polyclonal oligonucleotide pools. In four iterative rounds of shotgun assembly and sequencing we were able to generate an arrayed genome-spanning library comprising clonal sgRNAs for targeting most genes in the human genome.

## Results

**Ligation-independent assembly of sgRNA vectors.** Apart from traditional restriction-ligation cloning, LIC provides an efficient assembly route for DNA cassettes bearing complementary single-stranded overhangs through making use of covalent ligation by bacterial ligases<sup>14–16</sup>. Long overhangs ( $\geq 10$  nt) can be generated by T4 DNA polymerase, whose proof reading activity removes bases from dsDNA 3' ends until reaching a sequence position where a specific dNTP supplied in the reaction can be incorporated by the enzyme's polymerase activity. Based on LIC, we developed a cloning strategy that allows determining the spacer sequence of an sgRNA expression plasmid only by a single ssDNA oligonucleotide, thus enabling low-quantity chip-synthesized oligo pools to be used as input DNA without prior PCR amplification (Fig. 1). For that, we prepared a number of LIC-compatible plasmids, in which sgRNA expression is driven by a U6 promoter (Supplementary Fig. S1).

For testing the method, we mixed 96 individual sgRNA oligonucleotides with a master mix containing LIC-ready pR-U6-gRNA backbone DNA as well as a PAGE-purified constant antisense oligonucleotide (Supplementary Table S1). Upon chemical transformation into *E. coli*, we observed 87.5% of resulting clones to be correct by Sanger sequencing. 71 of the assembled CRISPR constructs targeting human genes were tested for their genome editing activity by co-transfection with a Cas9 expression plasmid into HEK293T cells (Fig. 2a). Subsequent deep sequencing revealed a mean genome editing activity of





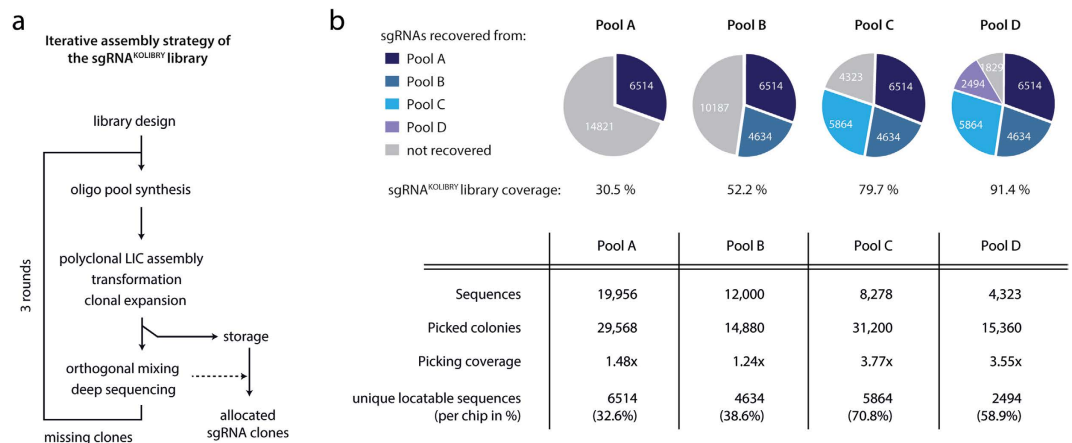
**Figure 3. Shotgun sgRNA assembly strategy.** Schematic diagram of the shotgun sgRNA assembly strategy using chip-synthesized oligo pools and of the orthogonal mixing strategy. The oligo pool was cloned by LIC and subsequently transformed into bacteria. Once colonies had grown on LB agar plates, clones were picked into single wells of 96-well plates (clone plates). Each set of 32 clone plates (each containing  $32 \times 96$  clones) was mixed both position-wise and group-wise, wherein a group corresponds to the equally colored part of an individual clone plate. Two resulting matrix plates per set are used as templates for a dual-PCR. The first PCR amplifies the sgRNA sequence and adds one of 12 possible barcodes specific for the individual matrix plates (6 for the position-based matrix plates and 6 for the group-based matrix plates = barcode 1). The second PCR amplifies the first PCR product, introduces 96 barcode combinations ( $8 \times 12 =$  barcodes 2 and 3) that are specific for every position on the 96-well matrix plates, as well as Illumina linker sequences. Finally, the PCRs are subjected to deep sequencing.

linker sequences as well as three barcodes to identify all 32 members of each pool by deep sequencing (Fig. 3). The design of the barcoding scheme allowed differentiation of up to 1,152 individual pools, translating to 18,432 individual clones being possibly analyzed in a single deep sequencing run.

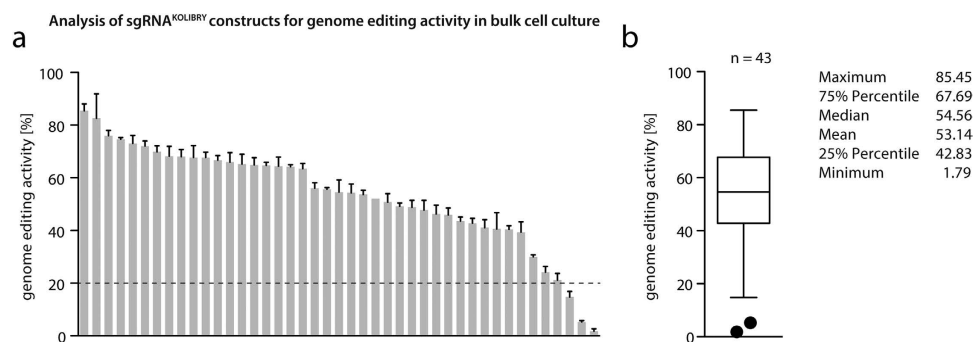
From the first sequencing round, 6,514 unique, error-free library members could be unambiguously allocated to a defined position on the clone plates. Individually barcoded deep sequencing confirmed all position calls from the orthogonal mixing to be correct (Supplementary Fig. S5a). 54.2% of the clones contained a correct sgRNA insert, whereas mutations were due to the oligo chip synthesis process and not the LIC assembly technology. However, we observed a non-random overrepresentation of sequence duplicates on clone plates picked from a single agar plate (Supplementary Fig. S5a), likely due to clonal amplification of bacteria before agar plate streaking. Clonal duplicates inevitably result in a reduced coverage of the picked library and also hamper sequence calling by orthogonal mixing (Supplementary Notes). To this end, for subsequent sgRNA assemblies, we spared the usual 1 h incubation step of the bacteria prior to plating to eliminate the problem of clonal duplicates (Supplementary Fig. S5b). Consequently, the coverage distribution of individual sequences matched the expected random distribution of a non-redundant library (Supplementary Fig. S5c).

Based on this improved protocol, we went on to complete the library with three subsequent rounds of synthesis, clone picking, and sequencing as described above (pool B, C and D, Fig. 4). Pool B contained 12,000 sgRNA library constructs and with a picking coverage of 1.24 fold (14,880 colonies) we identified 4,635 additional library constructs. From pool C, comprising of 8,278 sgRNA library members, we picked 31,200 colonies (3.77 fold picking coverage) to allocate an additional set of 5,864 correct library sgRNAs. Pool D contained the remaining 4,323 sgRNAs and we picked 15,360 colonies (3.55 picking-fold) from this pool to successfully allocate 2,494 additional clones. Altogether, we obtained an overall library coverage of 91.4% (Supplementary Table S4).

**Validation of library performance.** To validate the genome editing performance of the constructed library we picked 43 sgRNA constructs targeting 43 genes associated with cytosolic nucleic acid sensing pathways (Supplementary Table S5). These sgRNA constructs were co-transfected with a Cas9 expression plasmid into HEK293T cells. Two days later genomic DNA was extracted to perform locus-specific deep sequencing. Analyzing these data revealed a mean genome editing activity of 53.14%, with 93% of all sgRNAs (40 of 43) having an activity higher than 20% (Fig. 5). Of note, the genome editing activity of this set of sgRNA constructs was not significantly higher when compared to the previous panel of 71 randomly designed sgRNAs side by side (data not shown). However, a formal quantitative



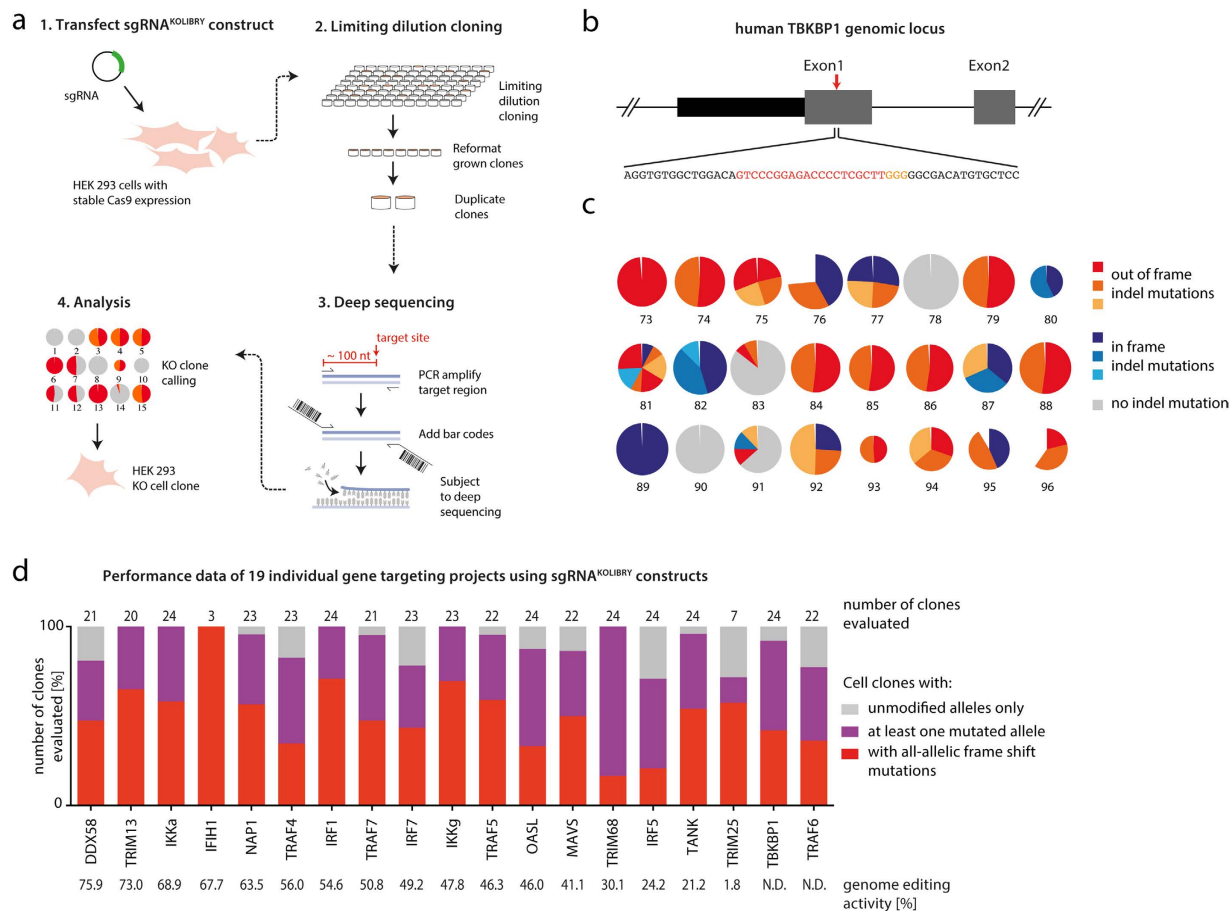
**Figure 4. Iterative assembly of the sgRNA<sup>KOLIBRY</sup> library.** (a) Iterative assembly strategy to generate large sgRNA libraries. (b) Pie charts and table summarizing the results and performance of the iterative assembly rounds of the sgRNA<sup>KOLIBRY</sup> library.



**Figure 5. Analysis of sgRNA<sup>KOLIBRY</sup> library constructs for genome editing activity.** Shown are the genome editing activities of 43 picked sgRNA<sup>KOLIBRY</sup> library constructs. The constructs were transfected together with a Cas9 expression plasmid into HEK 293T cells and genome editing activity was then determined by deep sequencing. (a) The activity of each sgRNA construct is depicted as mean values + SEM from three independent experiments. The dashed line indicates a threshold of 20% genome editing activity. (b) The box plot summarizes the genome editing activity of 43 sgRNAs, whereas Tukey outliers are depicted as individual dots.

comparison cannot be performed given the fact that different backbones were used for these studies (5.5 kb pR-U6-gRNA vs. 7.3 kb pL-U6-gRNA).

Finally, we went on to verify that the library is suitable for the generation of knockout cell lines. To do so, we selected 22 sgRNA constructs and transfected them into HEK293T cells stably expressing Cas9 (Fig. 6a). Two days following sgRNA transfection, cells were re-plated under limiting dilution conditions. After two weeks 24 single cell clones per target gene were harvested and expanded for subsequent genotype analysis. Using a workflow that allows the multiplexed analysis of several cell clones from different gene targeting experiments in parallel<sup>19</sup>, the clones were analyzed for indel events at the genomic locus of interest. One exemplary result is depicted for the target gene TBKBP1 (Fig. 6b). In total, sequencing data of 19 individual genes targeted provided enough read numbers for subsequent analysis. All of these yielded at least three cell clones harboring all-allelic out-of-frame indel mutations in the target gene, whereby knockout efficiencies of up to 100% were achieved (Fig. 6c and Supplementary Fig. S6). We additionally validated the protein expression of six different target genes for which three individual knockout cell clones were available. These experiments revealed that all knockout clones lacked the expression of the respective target gene, as observed by immunoblotting (Supplementary Fig. S7). Similar results were obtained when employing constructs of this library to generate genetic knockouts in other cell lines and when using different means of transgene delivery, e.g. by electroporation of THP1 cells<sup>16,20–22</sup>. We furthermore tested the genome editing activity of two sgRNA constructs that were delivered by lentiviral transduction. Doing so, we observed that virally transduced sgRNAs induced indel mutations at the desired locus, whereas the overall genome editing activity was tightly correlated with the amount of cells being transduced (Supplementary



**Figure 6. Generation of KO cell lines utilizing the sgRNA<sup>KOLIBRY</sup> library.** (a) Schematic diagram of knockout cell line generation. (b) Schematic view of the genetic locus of the human TBKBP1 gene. The red arrow indicates the location of the target site of the sgRNA. The red sequence highlights the target site whereby the orange sequence indicates the PAM. (c) OutKnocker was used to analyze deep sequencing data obtained from gene targeting experiments. Each pie chart represents the deep sequencing data from an individual cell clone. Each pie piece sums up all reads that have the same indel mutation, with its size representing the proportion of total reads analyzed. Grey pie charts show reads without indel mutations, blue colors indicate indel mutations that are in frame and red colors indicate indel mutation reads that are out of frame, whereas the different shades of blue and red represent individual indel mutations in one pie chart. Depicted is an exemplary output for the generation of TBKBP1 knockout HEK cells. (d) Shown is a histogram summarizing the results of 19 individual gene targeting projects. Grey bars show the number of clones that carry only unmodified alleles, while purple bars indicate clones with at least one mutated allele and red bars clones with all-allelic frame shift mutations.

Fig. S8). All in all these data confirmed that constructs from our sgRNA library are well suited to generate knockout cell lines.

## Discussion

We here present a LIC-based assembly method and an orthogonal sequence validation method for CRISPR sgRNA constructs that facilitate rapid iterative assembly and validation of large sgRNA libraries. Using this approach, we were able to recover 19,506 unique sequence-validated sgRNAs (91.4% library coverage) in four iterative clone picking rounds. To fulfill this task, we made use of the LIC assembly technology that only relies on a single fluid mixing step and no enzymatic amplification, yet stands out with high specificity and efficiency. In the absence of amplification steps, we obtained an evenly distributed polyclonal sgRNA pool, which could subsequently be used to generate an arrayed, monoclonal library with reasonable effort. Of note, an even distribution of library sequences is not only required for an optimal coverage, but also constitutes an important requirement to apply orthogonal mixing and deconvolution, since redundancy dramatically impacts on reliable allocation of sequences (Supplementary Notes).

Testing a randomly selected set of 71 sgRNA constructs for genome editing activity in human cells, we in general observed high genome editing activities. The mean activity level was ranging around 55%, with a low interquartile range of about 23%. Of note, in this panel only five sgRNAs displayed genome editing activities below 20% and two of these constructs encoded for four consecutive uridines, which is known to terminate RNA polymerase III driven expression<sup>23</sup>. To reveal possible targeting rules, we analyzed the genome editing activity of sgRNA sequences as a function of the position-wise C/G content of the sgRNA molecule rather than taking all individual base identities into account. We chose this reduced classification given the limited number of data points available, but also given the fact that this categorization represents a well-proven, simple measure of the local stability of a double stranded nucleic acid molecule. In this regard it is interesting to note that the recently published structures of RNA:DNA hybrids bound to Cas9 imply that the Cas9 protein binds to the sgRNA and the protospacer region in a largely sequence independent manner<sup>24–26</sup>. This suggests that the impact of the sgRNA sequence on Cas9 activity is rather due to thermodynamic properties of the respective sgRNA molecule or the sgRNA:protospacer complex than on Cas9 binding. Analyzing our data, we observed a trend towards C/G rich regions being favored at positions 17 and 18 of the protospacer region, whereas higher genome editing activity was observed for protospacer regions harboring A/T at positions 14 and 15. In line with our results, another study, in which data from pooled sgRNA screens were analyzed to predict possible targeting rules, suggested that C/G at positions 17 and 18 of the sgRNA molecule were associated with higher genome editing activity, whereas being less favorable at positions 14 and 15 of the sgRNA molecule<sup>11</sup>. Moreover, another group reported similar results for positions 14 and 15 of the sgRNA, while obtaining differing results for other positions<sup>27</sup>. All in all, we speculate that these results reflect the existence of a favorable thermodynamic profile of the protospacer region, the sgRNA:protospacer complex or the sgRNA molecule itself to facilitate functional enzyme activity, yet additional experiments are required to substantiate this hypothesis. Indeed, studying the genome editing activity of sgRNA constructs that were designed according to these rules did not display higher activities. However, in this regard it has to be noted that we switched to a considerably larger plasmid backbone to generate our genome-wide sgRNA library.

Designing our genome-wide sgRNA library, we favored sgRNA constructs abiding to the aforementioned position-based rules and we additionally premised an overall GC content of >35%, a constraint which has been experimentally validated by other groups<sup>11,27,28</sup>. Due to the fact that almost all sgRNAs show genome editing activity we decided to generate one sgRNA per gene, which we consider sufficient for the application of this library to generate knockout cells. However for arrayed screening purposes a higher coverage would be advantageous.

To render a target gene dysfunctional at the level of protein expression, we chose to place sgRNAs within the earliest possible coding exon, but at least within the first half of the coding region of a gene. An indel mutation in an early coding exon can shift the reading frame of the downstream coding region. This can result in premature translation-termination codon (PTC) and as such lead to a C-terminally truncated protein product. At the same time, the presence of a PTC can provoke the degradation of the mRNA due to nonsense-mediate mRNA decay, a mechanism that is most effective if the PTC is located in the first half of the open reading frame<sup>29</sup>. In line with these considerations, recent studies reported increased functional knock out rates for sgRNAs targeting early coding exons<sup>11,27</sup>.

At the time of designing our library, we opted for a stringent measure to avoid sgRNAs with putative off-target activities by assessing the PAM-proximal 13-mer sgRNA seed sequence to be unique in the genome<sup>5</sup>. Since then, several groups have reported on significant off-target mutagenesis by the CRISPR/Cas9 system<sup>30–38</sup>, suggesting that our specificity measures were probably not sufficient to eliminate all possible off-target effects of our sgRNAs constructs. Of note, while we have not experimentally assessed the frequency of putative off-target effects of our sgRNA construct design, we speculate that a large sample size would be required to obtain informative results that would allow reliable predictions for the entire collection. Nonetheless, recent studies analyzing single cell clones for CRISPR/Cas9 mediated off-target effects at a genome-wide level revealed a low incidence of such events. This suggests that off-target indels might not be a major concern when generating clonal knockout cell lines<sup>39,40</sup>. Indeed, it appears that when generating knockout cell lines by single cell cloning, clonal heterogeneity obtained irrespective of exogenous manipulations may indeed introduce a bigger variability than putative off target effects themselves<sup>39</sup>. Given the fact that our library was primarily designed for the generation of knockout cell lines, we therefore assume that off-target effects should not be of primary concern when applying the here-described sgRNA constructs for this purpose.

When employing constructs of our sgRNA library to generate knockout cells, we observed very high on-target activity, thus confirming the usefulness of our collection. Interestingly, we did not observe that knockout efficiencies correlated with the measured polyclonal activities of sgRNAs. In contrast, even sgRNA constructs with low transient genome editing activities yielded high amounts of knockout cell clones. The high knockout frequencies might be due to stable Cas9 expression in the cell line used and at the same time it is possible that the prolonged sgRNA and Cas9 expression during the cloning procedure leads to a high, cumulative genome editing frequency.

Taken together, we present an arrayed genome-wide sgRNA library designed to target human protein coding genes for loss of protein expression. Constructs of this library display high genome editing activity in polyclonal cell culture and are highly functional in generating knockout cell clones. This library will be a useful tool for dedicated, large-scale knock out cell generation projects or even the conduction

of pathway- or genome-wide genetic screens. The sgRNA KnockOut LIBRARY (sgRNA<sup>KOLIBRY</sup>) will be freely available as a resource to the research community.

## Methods

**CRISPR sgRNA plasmid assembly.** 10  $\mu$ l of the sgRNA entry vector (200 ng/ $\mu$ L in H<sub>2</sub>O) were digested with ApaI, SpeI, and EcoRV (FastDigest, Fermentas) at 37 °C for 2 h. The digestion mix was separated on an agarose gel (1%), the cleaved product was purified using an innuPrep gel extraction kit (Analytik Jena) and eluted in 15  $\mu$ l H<sub>2</sub>O (final concentration: 70 ng/ $\mu$ l). The linearized vector was chewed using T4 DNA polymerase (Enzymatics) for 5 min at 27 °C in the presence of dTTP as a stopping nucleotide. After heat-inactivation, the reaction was diluted 10-fold in 2 $\times$  NEB2 buffer in the presence of 0.5  $\mu$ M short universal reverse strand oligo (PAGE purified, 100  $\mu$ M, IDT). Gene-specific sgRNA oligos (desalted, 100  $\mu$ M, IDT) were diluted 1:400 in H<sub>2</sub>O and then mixed at equal volume with the vector pre-dilution and annealed in a PCR cycler (C100, Biorad). For a detailed protocol please refer to the Supplementary Methods section.

**Arrayed transformation and bacterial culture.** 5  $\mu$ l of the assembly mix were transformed using 50  $\mu$ L of chemo competent DH10b *E. coli* at 42 °C for 45 sec. Next, the bacteria were incubated with 450  $\mu$ L LB medium at 37 °C, 900 rpm for 1 h. The bacteria were streaked out on LB-Agar plates supplemented with 100  $\mu$ g/ml Ampicillin and incubated at 37 °C over night. On the next day LB medium containing 100  $\mu$ g/ml Ampicillin was inoculated with a single colony and incubated at 37 °C, 900 rpm over night. The following day DNA was purified using standard silica spin column kits according to the manufacturer's protocol.

**CRISPR/Cas9 transfection.** HEK 293T or HEK 293-CAS9 cells were plated at a density of 2  $\times$  10<sup>4</sup> cells per 96-well. On the next day, CRISPR plasmids were transfected using GeneJuice transfection reagent (Merk Millipore) according to the manufacturer's protocol. pRZ-mCherry-Cas9 and sgRNA constructs were transfected at a ratio of 150 ng to 50 ng per well. All vector sequences can be found in Supplementary Table S7.

**Cell lysis.** The medium was removed and the cells were lysed in 30  $\mu$ l of the following lysis buffer: 0.2 mg/ml proteinase K, 1 mM CaCl<sub>2</sub>, 3 mM MgCl<sub>2</sub>, 1 mM EDTA, 1% Triton X 100, 10 mM Tris pH 7.5. The reactions were incubated at 65 °C for 10 min and at 95 °C for 15 min.

**Dual PCR barcoding.** First-level PCR reactions are performed using 2  $\mu$ l PCR-compatible lysate as a template and a locus specific primer for a 12.5  $\mu$ l Phusion PCR reaction (Thermo Scientific) according to the manufacturer's protocol (annealing temperature: 60 °C, elongation time: 15 sec, 18 cycles). Of this reaction, 1  $\mu$ l is transferred to a second-level PCR using the same cycling conditions and a second set of primers that amplifies the first PCR, while adding 96 individual barcode combinations and an Illumina specific sequencing linker. For all primer sequences see Supplementary Table S6.

**Deep sequencing.** Crude PCR products were pooled and size-separated using a 1% agarose gel run at 100 V. After visualization with ethidium bromide under UV light, DNA bands from 350 bp to 500 bp were cut out and purified using Jena Analytik innuPrep gel extraction kit according to the manufacturer's protocol. Eluted DNA was precipitated by adding 0.1 volumes of 3 M NaOAc (pH 5.2) and 1.1 volumes of isopropanol. After centrifugation for 15 minutes at 4 °C, the resulting pellets were washed once in 70% EtOH and air-dried. 35  $\mu$ l water was added, non-soluble fractions were spun down and removed, and the DNA concentration was quantified using a NanoDrop spectrophotometer (Thermo scientific). Deep sequencing was performed according to the manufacturer's protocol using the MiSeq (Illumina) bench top sequencing system and a 300-cycle-V2 cassette. Data were obtained in FASTQ format.

**Library assembly.** Chip-synthesized oligo pools were obtained from CustomArray. The oligo pools were diluted to a final concentration of 0.04 ng/ $\mu$ l and incubated with the LIC-ready vector pre dilution as described for arrayed assemblies.

**Deep sequencing evaluation of polyclonal genome editing activity.** Using a modified version of the OutKnocker software<sup>19</sup>, sequencing reads were compared to a given reference sequence. Reads with indel mutations spanning or touching CRISPR target sites were counted as mutated reads. To calculate the mutation frequency, the number of mutated reads was divided by the total number of reads.

**Orthogonal mixing and deconvolution.** Bacterial liquid cultures were mixed using a Biomek FXp liquid handling station (Beckman Coulter) into group-based matrix plates and position-based matrix plates as indicated in Fig. 3. The first PCR amplifies the sgRNA sequence and adds one of 12 possible barcodes specific for the individual pool plates using the primers gRNA\_fwd\_1-gRNA\_fwd\_12 and gRNA\_rev (barcode 1 in Fig. 3). The second PCR amplifies the first PCR, while adding 96 individual barcode combinations and an Illumina specific sequencing linker (barcode 2 and 3 in Fig. 3). For all



primer sequences see Supplementary Table S6. Deconvolution was performed using a custom-written javascript program.

**Generation of knock out cell lines.** HEK 293 cells stably expressing mCherry-CAS9 (HEK 293-CAS9) are HEK-Blue™ IFN- $\alpha/\beta$  Cells (purchased from Invivogen) that were stably transduced with a mCherry-CAS9 and cloned by limiting dilution cloning. HEK 293-CAS9 cells were plated at a density of  $2 \times 10^4$  cells per 96-well. Cells were transfected with 200 ng of sgRNA expression plasmid per well using GeneJuice (Merck Millipore). Two days after transfection, cells were trypsinized, counted, and monoclonal clones were seeded at a density of 0.8 cells per 96-well in three 96-well plates. After two weeks, 24 monoclonal clones per targeted gene were picked, expanded, and genotyped as described<sup>19</sup>. Only monoclonal clones with at least 50 alignable reads were evaluated. No mycoplasma contamination was detected in regular screenings of our cell lines.

**Viral transduction.** HEK293T cells were seeded at a density of  $3 \times 10^5$  cells/ml in a six well dish the day before transfection. A six well was transfected with the library plasmid containing the sgRNA, a viral packaging plasmid expressing pCMV $\Delta$ 8.91 and a pseudotyping plasmid expressing VSV-G in a ratio of 10:10:1 with Lipofectamine 2000 according to the manufacturer's instruction. One day after transfection the medium was changed to a standard medium containing 30% FCS. 48 h later the virus was harvested and filtered. HEK 293-CAS9 were seeded at a density of  $3 \times 10^5$  cells/ml the day before transduction in a 96 well plate. On the next day cells were transduced with virus. A titrating from 100  $\mu$ l to 50  $\mu$ l, 25  $\mu$ l, 12.5  $\mu$ l and finally 0  $\mu$ l per well was performed. The wells were filled up to 100  $\mu$ l with standard medium. Two days after transduction the cells were fixed with paraformaldehyde and analyzed by FACS. Lysis and analysis by deep sequencing was done as described above.

**Western blotting.** Laemmli lysates of knockout cell clones were subjected to SDS-PAGE and subsequent immunoblotting. Membranes were probed for  $\beta$ -actin, IKK $\alpha$ , IRF1, MAVS, Sintbad, TANK and TRAF6. Antibodies were obtained from Cell Signaling (anti-IKK $\alpha$  #2682, anti-IRF1 #8478, anti-Sintbad #8605, anti-TANK #2141 and anti-TRAF6 #8028) and Santa Cruz Biotechnology (anti-MAVS sc-166583, anti- $\beta$ -actin-HRP sc-47778 and secondary antibody anti-rabbit-HRP sc-2004).

**Statistical analysis.** Outliers (Fig. 2b or Fig. 5b) were identified using the Tukey's outlier filter. A two-tailed, unpaired t-test with or without Welch's correction (an F test was performed to identify unequal variances) was used to compare genome editing activities of the pool of sgRNAs with either A/T or C/G at every position of the target region (Fig. 2c and Supplementary Fig. S2).

**Availability.** All functionally validated constructs are available upon request.

## References

- Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170, doi: 10.1126/science.1179555 (2010).
- Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338, doi: 10.1038/nature10886 (2012).
- Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821, doi: 10.1126/science.1225829 (2012).
- Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**, 233–239, doi: 10.1038/nbt.2508 (2013).
- Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826, doi: 10.1126/science.1232033 (2013).
- Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823, doi: 10.1126/science.1231143 (2013).
- Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227–229, doi: 10.1038/nbt.2501 (2013).
- Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi: 10.7554/eLife.00471 (2013).
- Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230–232, doi: 10.1038/nbt.2507 (2013).
- Chang, N. *et al.* Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res* **23**, 465–472, doi: 10.1038/cr.2013.45 (2013).
- Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84, doi: 10.1126/science.1246981 (2014).
- Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87, doi: 10.1126/science.1247005 (2014).
- Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267–273, doi: 10.1038/nbt.2800 (2014).
- Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**, 6069–6074 (1990).
- Schmid-Burgk, J. L. *et al.* Rapid hierarchical assembly of medium-size DNA cassettes. *Nucleic Acids Res* **40**, e92, doi: 10.1093/nar/gks236 (2012).
- Schmid-Burgk, J. L., Schmidt, T., Kaiser, V., Honing, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* **31**, 76–81, doi: 10.1038/nbt.2460 (2013).
- Erlich, Y. *et al.* DNA Sudoku--harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome research* **19**, 1243–1253, doi: 10.1101/gr.092957.109 (2009).
- Burckstummer, T. *et al.* A reversible gene trap collection empowers haploid genetics in human cells. *Nature methods* **10**, 965–971, doi: 10.1038/nmeth.2609 (2013).

19. Schmid-Burgk, J. L. *et al.* OutKnocker: a web tool for rapid and simple genotyping of designer nuclease edited cell lines. *Genome research* doi: 10.1101/gr.176701.114 (2014).
20. Pelka, K. *et al.* Cutting Edge: The UNC93B1 Tyrosine-Based Motif Regulates Trafficking and TLR Responses via Separate Mechanisms. *J Immunol* **193**, 3257–3261, doi: 10.4049/jimmunol.1301886 (2014).
21. Mankan, A. K. *et al.* Cytosolic RNA:DNA hybrids activate the cGAS-STING axis. *Embo J* **33**, 2937–2946, doi: 10.15252/emboj.201488726 (2014).
22. Wassermann, R. *et al.* Mycobacterium tuberculosis Differentially Activates cGAS- and Inflammasome-Dependent Intracellular Immune Responses through ESX-1. *Cell host & microbe* **17**, 799–810, doi: 10.1016/j.chom.2015.05.003 (2015).
23. Das, G., Henning, D., Wright, D. & Reddy, R. Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *Embo J* **7**, 503–512 (1988).
24. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573, doi: 10.1038/nature13579 (2014).
25. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949, doi: 10.1016/j.cell.2014.02.001 (2014).
26. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997, doi: 10.1126/science.1247997 (2014).
27. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* doi: 10.1038/nbt.3026 (2014).
28. Gagnon, J. A. *et al.* Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, e98186, doi: 10.1371/journal.pone.0098186 (2014).
29. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat Struct Mol Biol* **16**, 107–113, doi: 10.1038/nsmb.1550 (2009).
30. Cradick, T. J., Fine, E. J., Antico, C. J. & Bao, G. CRISPR/Cas9 systems targeting beta-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res* **41**, 9584–9592, doi: 10.1093/nar/gkt714 (2013).
31. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839–843, doi: 10.1038/nbt.2673 (2013).
32. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822–826, doi: 10.1038/nbt.2623 (2013).
33. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827–832, doi: 10.1038/nbt.2647 (2013).
34. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677–683, doi: 10.1038/nbt.2916 (2014).
35. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670–676, doi: 10.1038/nbt.2889 (2014).
36. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research* **24**, 132–141, doi: 10.1101/gr.162339.113 (2014).
37. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179–186, doi: 10.1038/nbt.3101 (2015).
38. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187–197, doi: 10.1038/nbt.3117 (2015).
39. Veres, A. *et al.* Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**, 27–30, doi: 10.1016/j.stem.2014.04.020 (2014).
40. Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12–13, doi: 10.1016/j.stem.2014.06.011 (2014).

## Acknowledgements

J.L.S.-B. is supported by a scholarship of the Studienstiftung des Deutschen Volkes. V.H. is member of the excellence cluster ImmunoSensation and supported by grants from the German Research Foundation (SFB704 and SFB670) and the European Research Council (ERC-2009-StG 243046). We kindly thank H. Schilling, C. Wallerath, K. Höning and J.-P. Stümpel for great technical support.

## Author Contributions

T.S., J.L.S.-B. and V.H. developed the methodology, designed experiments, analyzed the data and wrote the manuscript. T.S. and J.L.S.-B. performed the experiments. V.H. conceived and supervised the study.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Schmidt, T. *et al.* Synthesis of an arrayed sgRNA library targeting the human genome. *Sci. Rep.* **5**, 14987; doi: 10.1038/srep14987 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>