Circular synthesized CRISPR/Cas gRNAs for functional interrogations in the coding and noncoding genome

Martin Wegner1,#, Valentina Diehl1,#, Verena Bittl1,3, Rahel de Bruyn1, Svenja Wiechmann1,4, Yves Matthes1, Marie Hebel1, Michael G. B. Hayes2, Simone Schaubeck1, Christopher Benner2, Sven Heinz2, Anja Bremm1,3, Ivan Dikic1,3,5,6, Andreas Ernst1,4, Manuel Kaulich1,5,6,*

1 Institute of Biochemistry II, Goethe University Frankfurt - Medical Faculty, University Hospital, Frankfurt am Main, Germany.
2 Department of Medicine, University of California, San Diego, USA.
3 Buchmann Institute for Molecular Life Sciences, Goethe University, Frankfurt am Main, Germany.
4 Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine & Pharmacology TMP, Frankfurt am Main, Germany.
5 Frankfurt Cancer Institute (FCI), Frankfurt am Main, Germany.
6 Cardio-Pulmonary Institute (CPI), Frankfurt am Main, Germany.
# These authors contributed equally to this work.
*Correspondence should be addressed to Manuel Kaulich (kaulich@em.uni-frankfurt.de).
ABSTRACT

Current technologies to generate CRISPR/Cas gene perturbation reagents are labor intense and require multiple ligation and cloning steps. Furthermore, increasing gRNA sequence diversity negatively affects gRNA distribution, leading to libraries of heterogeneous quality. Here, we present a rapid and cloning-free mutagenesis technology to efficiently generate covalently-closed-circular-synthesized (3Cs) CRISPR/Cas gRNA reagents that uncouples sequence diversity from sequence distribution. We demonstrate fidelity and performance of 3Cs reagents by tailored targeting of all human deubiquitinating enzymes (DUBs) and identify their essentiality for cell fitness. To explore high-content screening, we aimed at generating the up-to-date largest gRNA library to simultaneously interrogate the coding and noncoding human genome and identify genes, predicted promoter flanking regions, transcription factor and CTCF binding sites linked to doxorubicin resistance. Our 3Cs technology enables fast and robust generation of bias-free gene perturbation libraries with yet unmatched diversities and should be considered an alternative to established technologies.

INTRODUCTION

CRISPR/Cas has rapidly become the gold standard for unbiased high-throughput experiments, outperforming preexisting technologies such as RNAi (Evers et al., 2016; Morgens et al., 2016). A fundamentally important aspect for high-fidelity CRISPR/Cas screening is the quality of the gRNA library, with its diversity and distribution primarily influencing downstream experimental scales (Sanson et al., 2018). Conventionally used methods to generate gRNA libraries in pooled or arrayed formats include T4 ligase or homology-based cloning techniques and further require PCR-based amplification of gRNA-encoding oligonucleotides as well as the presence of open plasmid DNA for successful gRNA sequence cloning (Arakawa, 2016; Koike-Yusa et al., 2014; Ong et al., 2017; Schmidt et al., 2015; Shalem et al., 2014; Vidigal and Ventura, 2015; Wang et al., 2014). Due to these technical constraints, conventional libraries contain an unwanted PCR and cloning-dependent bias in gRNA distribution that influences the experimental scale required for statistically significant hit calling (Shalem et al., 2014; Wang et al., 2014). CRISPR libraries have become
ubiquitously used in functional genomics efforts, underscoring relevance and utility for new PCR-
and cloning-free technologies.

The rod-shaped filamentous phage M13 differs from other bacteriophages in that its genome
packaging capacity is variable and that it is present as single-stranded (ss) DNA. Kunkel
mutagenesis utilizes the ease of ssDNA purification from M13 phage and enables rapid site-
specific mutagenesis to construct high-quality phage display libraries (Handa and Varshney, 1998;
Huang et al., 2012; Kunkel, 1985, 2001a). In combination with M13’s malleable coat, Kunkel
mutagenesis has significantly contributed to the great success of phage display technologies
(Ernst et al., 2013; Sidhu, 2001).

Here, we demonstrate the applicability of Kunkel mutagenesis for the generation of high-quality
and high-fidelity CRISPR/Cas and RNAi gene perturbation reagents. In more detail, we developed
a highly reproducible improved Kunkel mutagenesis technology designed to robustly generate 3Cs
CRISPR/Cas gRNA libraries over a broad range of gRNA diversities. We demonstrate the high
fidelity of 3Cs gRNA libraries by targeting all human DUBs and determine their proliferative
depletion phenotype, confirming previously known and discovering hitherto unknown DUB
phenotypes. In an effort to enable unbiased screening within coding and noncoding regions, we
encoded SpCas9 nucleotide preferences into a degenerated oligonucleotide and generated a
highly complex CRISPR/Cas gRNA library (truly genome-wide, TGW). Doxorubicin positive
selection screens with the TGW library in unperturbed human telomerase-immortalized retinal
pigmented epithelial cells (hTERT-RPE1) identifies coding and noncoding regions, emphasizing
the relevance of noncoding sequence elements in drug resistance mechanisms. To enable high-
content functional interrogations in a truly genome-wide manner, we introduce an optimized
version of this library (oTGW). In summary, we establish the 3Cs technology as a robust alternative
method to generate high-quality CRISPR/Cas gene perturbation libraries.

RESULTS

Circular synthesized gRNAs are high-quality CRISPR/Cas reagents
In classical Kunkel mutagenesis (Kunkel, 1985, 2001a), circular ssDNA isolated from filamentous phage is hybridized with a complementary oligonucleotide that is extended and ligated to obtain a double-stranded DNA plasmid. Since Kunkel mutagenesis is performed on ssDNA, we anticipated it to be insensitive to secondary DNA structures of viral sequence elements and therefore should enable the PCR and cloning-free generation of lentiviral gene perturbation reagents (Huang et al., 2012; Kunkel, 1985). We therefore hypothesized that the generation of lentiviral CRISPR/Cas gRNA libraries using circular ssDNA and Kunkel mutagenesis would reduce coupling of gRNA diversity to gRNA distribution and generate reagents with high quality (Figure 1A).

To demonstrate its general applicability on lentiviral CRISPR/Cas plasmids, we transformed Escherichia coli (E. coli) CJ236 bacteria with the commonly used pLentiGuide and pLentiCRISPRv2 plasmids (Sanjana et al., 2014), both of which contain a U6 promoter-controlled non-human targeting (NHT) placeholder gRNA followed by a SpCas9 tracrRNA sequence (J G Doench et al., 2014; Sanjana et al., 2014). Importantly, F-factor containing CJ236 bacteria lack dUTPase (dut) and uracil-glycosylase (ung) thereby tolerating the presence of deoxyuridine (dU) in genomic and plasmid DNA (Kim et al., 2012). Superinfection of single-colony CJ236 culture with M13KO7 bacteriophage (10^8 cfu/mL) facilitated the generation of >30 µg of dU-containing circular ssDNA. Although circular ssDNA is identical in length to dsDNA, circular ssDNA of lentiviral CRISPR/Cas plasmids migrated faster and appeared as a single band in gel electrophoresis (Figure 1B). Circular dU-ssDNA was hybridized with a gRNA-encoding complementary oligonucleotide that contained sequence homology regions (3Cs homology) at its 5' and 3' ends, and extended and ligated with T7 polymerase and T4 ligase, respectively (Figure 1A). This resulted in heteroduplexed 3Cs DNA (3Cs-dsDNA), composed of dU-template ssDNA and deoxythymidine-containing newly synthesized complementary DNA that also includes the gRNA-encoding oligonucleotide (Figure 1A) (Huang et al., 2012; Kunkel, 1985, 2001a). To gain insights into oligonucleotide requirements and kinetics of 3Cs reactions, we tested different 3Cs homology lengths of 10, 13, 15, and 18 nucleotides and performed a 3Cs reaction time series and identified 18 nucleotides of homology (above 45°C annealing temperature) and 2 hours of 3Cs reaction time to be sufficient (Figure 1-figure supplement 1A-C) (Kunkel, 2001b).
Using rule set 2 (RS2) (Doench et al., 2016; John G. Doench et al., 2014), we designed six GFP-targeting gRNA sequences and extended them by 5' and 3' 3Cs homology. Synthesized gRNA-encoding oligonucleotides were hand-pooled in equimolar ratios, phosphorylated and used in a 1:5 ratio (2 µg ssDNA to 60 ng oligonucleotide) to generate heteroduplex dU-3Cs-sDNA (Figure 1C).

To remove NHT/dU-containing template and amplification of the gRNA-encoding complementary strand, 3Cs products were column-purified and transformed in dut'/ung' bacteria. Bacterial clones were grown and their plasmid DNA SANGER sequenced, revealing that 81% of pLentiGuide and 82% of plentiCRISPRv2 contained GFP-targeting gRNAs (Figure 1D and Figure 1-figure supplement 1D). To test if dU supplementation reduces the amount of NHT-containing template plasmid by improving dU-incorporation during ssDNA production, CJ236 culture media was supplemented with 2.5 µM dU. In addition, the gRNA placeholder sequence of pLentiGuide and plentiCRISPRv2 was changed to contain an I-SceI restriction enzyme recognition site. While the effect of increased dU concentrations was negligible, I-SceI-mediated removal of wildtype plasmid reduced their level to below our detection limit (Figure 1D and Figure 1-figure supplement 1D-F). We performed next-generation sequencing (NGS) on the plentiCRISPRv2 sample with an average read count of 1.15 million per GFP sequence and identified a wildtype rate below 0.3% in the absence of any apparent sequence bias (Figure 1E and Supplementary File 1). A one-sided Chi-squared test for goodness of fit identified a uniform distribution (p=0.1) of all six gRNA sequences. The uniform gRNA distribution was supported by a low coefficient of variation (CV) of 33.18% and an area under the curve (AUC, Lorenz curve) of only 0.56 (Figure 1E-F and Figure 1-figure supplement 1G).

To test for cellular functionality of 3Cs gRNAs, we used the plentiCRISPRv2 GFP-targeting 3Cs gRNA constructs to generate infectious lentiviral particles and transduced GFP-positive human telomerase-immortalized retinal pigmented epithelial (hTERT-RPE1) cells. Seven days post-transduction, we performed a T7 Endonuclease I assay and observed robust GFP gene editing both by a single GFP-targeting 3Cs gRNA (3Cs-gRNA) and by the pool of six 3Cs-gRNAs, while un-transduced (-) and an NHT control gRNA failed to do so (Figure 1G). GFP gene editing translated to a lentiviral dose-dependent loss of GFP protein when analyzed by fluorescence flow-
cytometry and immunoblotting (**Figure 1H-I** and **Figure 1-figure supplement 1H**). Taken together, we demonstrate that the 3Cs technology enables the rapid and cloning-free generation of high-quality single and pooled CRISPR/Cas gRNAs.

**3Cs uncouples sequence diversity from sequence distribution**

The absence of PCR amplification and cloning steps, in combination with the uniform distribution of the six GFP-targeting 3Cs-gRNAs, led us to reason that 3Cs may uncouple sequence diversity from sequence distribution during gRNA library generation. To test this hypothesis, we designed six degenerated 3Cs oligonucleotides with increasing numbers of randomized nucleotides to mimic gRNA sequence pools with diversities ranging from 256 to 262,144 individual sequences (**Figure 2A**). The six pools were applied in parallel 3Cs syntheses on dU-ssDNA template of pLentiCRISPRv2 (**Figure 2B** and **Figure 2-figure supplement 2A**). Independent of an oligonucleotide’s diversity, NGS and computational analyses identified all individual sequences and uniform distributions with area under the curve values between 0.6 and 0.73 (**Figure 2-figure supplement 2B** and **Supplementary File 2**). Despite the uniform distribution, we observed a prominent cytosine (C) bias in the randomized libraries, with C contents of above 40% within the top 10% of most abundant gRNAs (**Figure 2C**). We reasoned that the C bias is likely due to incomplete phosphoramidite mixing during oligonucleotide synthesis and should therefore be absent from gRNA libraries containing nonrandom gRNA sequences (Ellington and Pollard, 2009).

To test this hypothesis, we designed and generated a nonrandom 3Cs-gRNA library targeting all 105 human DUBs, each with three gRNAs (DUB library). NGS and nucleotide content analysis confirmed our hypothesis and revealed the absence of C bias from the nonrandom DUB library (**Figure 2C and Supplementary File 3**). To correct the randomized libraries for the C bias, we determined the individual nucleotide frequency at every randomized position and used them to normalize the original read counts, leading to improved AUC values and sequence distributions (**Figure 2D and Supplementary File 2**), further confirming the uncoupling of sequence diversity and distribution in 3Cs reactions. Taken together, we conclude that 3Cs is a robust technology that uncouples sequence distribution from sequence diversity and, therefore, is a powerful alternative technology to conventional gRNA cloning methods to generate gRNA libraries.
3Cs-gRNA libraries are of high fidelity: proliferative essentiality of human DUBs

Next, we aimed at investigating the performance of 3Cs-gRNA reagents in cellular screenings. To do so, we generated infectious lentiviral particles of the 3Cs-gRNA DUB library and applied them to a proliferation screen in non-transformed hTERT-RPE1 cells in biological duplicates (MOI 0.2, coverage 1,000). Two days after lentiviral transduction, cells were either collected (day 0, reference time point) or selected by puromycin and kept in culture for 11 days (day 11) or 21 days (day 21) in cycling conditions representing at least a 1,000-fold library coverage (Figure 3A). Cells collected at day 0, 11, or 21 were subject to genomic DNA extraction and amplicon-based NGS library preparation, as was reported previously (Doench et al., 2016; Koike-Yusa et al., 2014). We performed single-read sequencing on an Illumina NextSeq500 with an averaged read count per gRNA of above 35,000 across all biological samples and replicates (Supplementary File 4). Similar to previously reported CRISPR analysis algorithms and to enable a comparison of individual time points, we summed all individual gRNA read counts per gene and normalized each gene read count per sample to the total number of read counts within that sample (Supplementary File 4) (Li et al., 2014; Spahn et al., 2017). In line with reports of high experimental reproducibility of CRISPR/Cas screenings (Evers et al., 2016; Morgens et al., 2016), we determined $R^2$ values of 0.95, 0.88, and 0.90 for time points day 0, 11, and 21, respectively (Figure 3-figure supplement 3A-C). Spearman correlation and Shapiro–Wilk confidence test revealed correlations of above 0.88 and p-values of below 0.001, respectively, all of which demonstrate a high experimental on the level of gRNA representation (Figure 3B and Figure 3-figure supplement 3A-C) (Shapiro and Wilk, 1965). To analyze the reproducibility of our screen on the level of gene phenotypes, we applied MAGeCK and PinAPL-Py, two established algorithms for the analysis of CRISPR/Cas screens, to raw gRNA read counts of both replicates and calculated aggregated positive and negative proliferation phenotypes by means of log2 fold changes with associated p-values (Figure 3C-E and Figure 3-figure supplement 3D-H and Supplementary File 5-6) (Li et al., 2014; Spahn et al., 2017). Consistent over both time points, cells depleted of USP28 or BRCC3 proliferated faster than cells harboring non-human target sequences (NHTs), identifying both as negative regulators of hTERT-RPE1 proliferation (Figure 3C-D and Figure 3F and Supplementary File 4-7).
In contrast, cells depleted of PSMD14, USP7 or COPS6 proliferated slower than cells harboring non-human target sequences (NHTs), identifying them as positive regulators of hTERT-RPE1 proliferation (*Figure 3C-D and Figure 3F and Supplementary File 4-5*).

CRISPR/Cas drop out screens are performed with varying experimental durations, ranging from 5 to 15 days (Joung et al., 2017; Potting et al., 2017). However, recent work demonstrates that CRISPR/Cas induces a G1 phase arrest in p53 proficient hTERT-RPE1 cells impacting hit calling (Haapaniemi et al., 2018), suggesting later screening time points to be beneficial for hit calling. Indeed, when comparing normalized gene ranks we observed a trend of increased phenotype resolution among negative gene ranks over time, while this effect was largely absent from positive gene ranks (*Figure 3-figure supplement 3I and Figure 3- Supplementary File 4-5*). This effect was reported previously and can potentially be explained by the disproportional assay window of positive and negative cell proliferation phenotypes, leading to a higher phenotypic resolution among negative proliferative effects (Shalem et al., 2014; Wang et al., 2014).

Stable and robust proliferative phenotypes are time-independent and as such the phenotype resolution enhances over time (*Figure 3-figure supplement 3I*). However, multiple mechanisms and cellular backgrounds can influence phenotype strength, timing and orientation. To identify time-dependent phenotypes, we analyzed MAGeCK-derived log2 fold changes with associated p-values of time points day 11 and day 21 and identified genes whose deletion phenotype significantly changed between day 11 and day 21 (*Figure 3E-F*). Depletion of USP28 and USP46 induced the strongest positive change while deletion of USP22, USP48 or TNFAIP3 induced the most significant negative change in phenotype between day 11 and day 21, suggesting a time-dependent absence of compensatory mechanisms to accommodate an early loss-of-function phenotype (*Figure 3E-F and Figure 3-figure supplement 3I*). In order to validate our findings, we chose two positive and negative proliferation-inducing DUBs, generated lentiviral supernatant to deliver shRNA sequences targeting the selected DUBs, and transduced hTERT-RPE1 cells. Over the course of 2 weeks after transduction, we measured cell numbers by AlamarBlue staining. When compared to negative (Luciferase) and positive (Plk1) control shRNA sequences, depletion of USP28 and BRCC3 induced a rapid positive proliferation effect (*Figure 3G*). In contrast, depletion of USP7 and COPS6 induced an instant and strong negative proliferation effect (*Figure 8*).
3Cs is versatile and generates arrayed and pooled 3Cs-shRNA reagents

Due to its versatility, the CRISPR/Cas technology has become the method of choice for gene perturbation experiments, yet classical short hairpin RNAs (shRNA, RNAi) remain widely used. However, shRNA oligonucleotides contain complementary sequences that form stable secondary structures that render the generation of shRNA reagents inefficient (McIntyre and Fanning, 2006). A crucial step in our improved Kunkel mutagenesis technology is the denaturation of the gRNA-encoding oligonucleotides and their subsequent annealing to template ssDNA (see Materials and Methods section). Due to this denaturation and annealing step, we anticipated that improved Kunkel mutagenesis would circumvent shRNA oligonucleotide secondary structures and enable the generation of shRNA reagents. We chose pLKO.1 (Stewart et al., 2003), one of the most widely used lentiviral and shRNA-expressing plasmids to generate circular dU-ssDNA. Similar to circular ssDNA of CRISPR/Cas plasmids, the circular ssDNA of pLKO.1 migrated as a single band in agarose gel electrophoresis (Figure 3-figure supplement 4A). Next, we designed a GFP-targeting 3Cs shRNA (3Cs-shRNA) primer consisting of 5' and 3' 3Cs homology and two complementary GFP-shRNA sequences separated by a six-nucleotide hairpin sequence (Figure 3-figure supplement 4B). Performing two parallel 3Cs reactions using 60 ng and 120 ng of ssDNA, both reactions yielded the characteristic 3Cs-DNA band pattern with no major difference in bacterial transformation efficiency between the two tested scales (Figure 3-figure supplement 4C). To demonstrate the successful integration of the GFP-shRNA sequence into pLKO.1, we amplified single bacterial clones carrying 3Cs-DNA of shRNA reactions and analyzed their plasmid DNA by SANGER sequencing, resulting in the expected GFP-shRNA sequence and the absence of adjacent nucleotide changes (Figure 3-figure supplement 4D), from which we concluded that 3Cs is a versatile technology that generates high-quality gRNA and shRNA reagents. To demonstrate 3Cs-shRNA fidelity, we generated infectious lentiviral particles of the GFP-targeting 3Cs-shRNA and transduced GFP-positive hTERT-RPE1 cells. Strikingly, 96 hours after lentiviral transduction, we observed a reduction in GFP-fluorescence, confirming the functionality of the
3Cs-shRNAs in cells (Figure 3-figure supplement 4E). Moreover, we investigated the performance of improved Kunkel mutagenesis in generating 3Cs-shRNA libraries. Based on the above described principles, we designed a 3Cs-shRNA library targeting all human ubiquitin-conjugating E2 enzymes (E2s) each with two shRNAs (Supplementary File 7). To generate the library, individually synthesized oligonucleotides were pooled in equimolar ratios and applied to a pooled 3Cs reaction. The resulting products were resolved by gel electrophoresis (Figure 3-figure supplement 5A). Similar to the I-SceI-mediated depletion of wildtype remnants from CRISPR/Cas 3Cs-gRNA constructs, a Bsu36I restriction enzyme clean-up step removed pLKO.1 wildtype remnants and SANGER sequencing of the final E2 3Cs-shRNA library (E2.2) confirmed a randomization of forward and reverse-complement shRNA sequences (Figure 3-figure supplement 5B-C). To determine the E2 3Cs-shRNA distribution more accurately, we performed NGS sequencing with an average shRNA read count of >8,300 and determined a wildtype remnant level of 0.04%, a CV of 37.9% and an AUC of 0.68 demonstrating an almost uniform distribution (Figure 3-figure supplement 5D-E). To correlate 3Cs-shRNA abundance and distribution of the E2 3Cs-shRNA libraries before and after Bsu36I enzyme digest, we determined the ratios of their respective normalized read counts. Importantly, all ratios were close to one and lined up at the respective diagonal with a linear regression $R^2$ of 0.9687 (Figure 3-figure supplement 5F), demonstrating a high correlation of individual data points and no influence of the Bsu36I digest on 3Cs-shRNA sequence distribution. In summary, this demonstrates that our 3Cs technology can be adapted to generate high-quality shRNA reagents in single and pooled formats.

A partially randomized 3Cs gRNA library to target the coding and noncoding genome simultaneously

The 3Cs method does not require PCR-amplification of gRNA-encoding oligonucleotides, is free of conventional cloning steps and uncouples sequence diversity from sequence distribution. Thus we hypothesized that 3Cs gRNA library diversity is mostly limited by the number of distinguishable oligonucleotides within a 3Cs reaction and the subsequent bacterial electroporation efficiencies. Limitations in electroporation efficiencies can be overcome by accumulating the individual
efficiencies of multiple parallel reactions, as routinely performed to amplify phage libraries with diversities beyond $10^9$ (Smith and Scott, 1993). The number of distinguishable oligonucleotides is limited by the capacity of synthetic oligonucleotide synthesis, rendering truly genome-wide gene perturbation libraries unfeasible.

SpCas9 nucleotide preferences had previously been identified and revealed a preference for 3’ pyridine bases while thymidine nucleotides are disfavored (Doench et al., 2016; John G. Doench et al., 2014). In an exploratory effort, we translated SpCas9 gRNA nucleotide preferences into a degenerated oligonucleotide sequence (truly genome-wide, TGW) of 20 nucleotides, representing a theoretical diversity of $7.3 \times 10^{10}$ (**Figure 4A**), maximally targeting $1.65 \times 10^7$ sites in the human coding and noncoding genome (**Figure 4A and Supplementary File 8**). As mentioned above, randomized positions in DNA oligonucleotides can contain strong single nucleotide cytosine bias, we therefore used hand-mixed phosphoramidite pools to generate this oligonucleotide pool. In eight parallel large-scale 3Cs reactions (each 20 µg ssDNA and 600 ng oligonucleotide), we applied this oligonucleotide to dU-ssDNA of pLentiGuide and pLentiCRISPRv2 and resolved the 3Cs products by gel electrophoresis (**Figure 4B**). Importantly, we note that the amplification of this degenerated library is limited by the number of transformed bacteria. Since complete generation of this reagent is currently unfeasible, we limited our efforts to eight parallel electroporation reactions and achieved a cumulative transformation efficiency of $1.2 \times 10^{10}$, accounting for ~16% of TGW sequences, assuming a stringent uniform sequence distribution. In order to approximate the gRNA distribution, we generated 14.4 million NGS reads and found 94.19% to be unique (**Figure 4C and Supplementary File 9**). We went on to extract the nucleotide frequencies for each gRNA position from TGW NGS reads, translated them to IUPAC nomenclature, and identified the identical degenerated sequence that we initially applied in the form of the degenerated oligonucleotide pool (**Figure 4D**). Furthermore, the distribution of TGW read counts had a CV of 0.26% and an AUC of 0.52, suggesting a nearly uniform distribution of sequenced and represented gRNA sequences (**Figure 4E-F**) (Makowski and Soares, 2003).

The 3Cs technology enables the generation of gRNA libraries with sequence diversities exceeding our current abilities for coverage-based screenings. Being aware of the coverage limitations, we explored a TGW library screen in the context of a strong positive selection pressure. In hTERT-
Wegner et al, 3Cs-gRNAs

RPE1 cells, doxorubicin induces a robust, irreversible and dose-dependent reduction of cell viability within 4 days (**Figure 5-figure supplement 6A**). We therefore generated 5.5x10^8 infectious lentiviral particles of the TGW library and applied them to screen for resistance to doxorubicin (**Figure 5A-B**). In three biologically independent experiments, we transduced a total of 5.5x10^8 hTERT-RPE1 cells with a MOI of 1, added 1 µM doxorubicin seven days post transduction and replaced the media every seven days for 21 consecutive days. Cells surviving the treatment were harvested and their genomic DNA was extracted for NGS (**Figure 5B**). Although the experimental reproducibility was low (0.004%), we identified an experimental overlap of 4,232 gRNAs, with associated Spearman ranking and Pearson correlations of above 0.75 (**Figure 5C and Figure 5-figure supplement 6B**). To validate these sequences, we designed and generated a new 3Cs-gRNA validation library consisting of the identified 4,232 gRNAs and repeated the doxorubicin resistance screen with established experimental parameters (coverage of 1,000 and MOI of 0.2) (**Figure 5-figure supplement 7A-B**). As a result, we reidentified 2,716 gRNAs of which 795 were more two-fold enriched after 21 days of doxorubicin treatment, when compared to the untreated control (**Figure 5D and Supplementary File 11**). In order to map the 795 gRNA sequences to a location within the human genome, we applied Cas-OFFinder and used Ensembl, ENCODE, Roadmap Epigenomics and Blueprint databases for sequence annotation (Bernstein et al., 2010; Consortium et al., 2012; Fernandez et al., 2016; Zerbino et al., 2018). We identified seven gRNAs to target five genes (PDE8B, AVPR2, CYSLTR2, IL3RA, and POLE2), of which PDE8B and AVPR2 were targeted by two gRNAs, and a single gRNA sequence matched a noncoding location within chromosome 8 (chr8:93022800) (**Figure 5E and Figure 5G and Supplementary File 12-13**). Among the coding hits we identified CysLTR2, a Leukotriene C4 G-protein-coupled eicosanoid receptor that was recently reported to induce doxorubicin resistance by abolishing the accumulation of reactive oxygen species (Dvash et al., 2015). To validate CysLTR2 as a doxorubicin-resistance inducing hit, we chemically inhibited CysLTR2 with increasing concentrations of Bay-CysLT2 or Bay-u9773 in the presence of doxorubicin and quantified cell viability by AlamarBlue staining. Importantly, both drugs reverted the doxorubicin-induced toxicity in a dose-dependent manner (**Figure 5F**), suggesting CysLTR2 loss to cause doxorubicin resistance.
To account for sequence differences between hTERT-RPE1 cells and the reference genome (GRCh38.86), as well as SpCas9 off-target activity (Cho et al., 2014; Hsu et al., 2013; Pattanayak et al., 2013), we extended our computational analysis by allowing up to two mismatches during Cas-OFFinder-based target sequence identification. As expected, the number of gRNAs that could be mapped to the reference genome increased to 192 and 222 for coding and noncoding target sites, respectively, accounting for 50.3% of the 795 gRNAs (Figure 5E and Supplementary File 12). Interestingly, when mismatches are allowed, we identified three gRNAs to target two different coding positions within the AKAP6 gene (chr14:32671632, chr14:32784395), as well as five different gRNAs to target the exact same coding position within the ASPA2 gene (chr2:9229295) (Figure 5E and Supplementary File 12). Within the noncoding gRNA target sites, we identified four gRNAs to target four different positions on chromosome X (56546543, 57766898, 63133046, 63245878), all of which are in close proximity to the SPIN2A gene (Figure 5G and Supplementary File 13), suggesting a doxorubicin tolerance-inducing function in this locus.

In order to reveal whether the identified set of coding genes correlated with reported phenotypes or gene ontologies, we performed a molecular signature analysis on the 178 coding target regions and identified 25 genes to match the UV_RESPONSE_VIA_ERCC3 (downregulated in mutant ERCC3-expressing fibroblasts) as the most significant hit (p-value of 6.11E-17) (Figure 5H-I and Supplementary File 14). Importantly, doxorubicin-induced interstrand crosslinks are repaired by ERCC3-dependent nucleotide excision repair (NER) and NER-deficient cells have been shown to display higher tolerance to adduct-forming anthracycline treatment, connecting these 25 genes to an increased doxorubicin tolerance (Bret et al., 2013; Spencer et al., 2008; van Brabant et al., 2000). Furthermore, mutations in noncoding sequences have been linked to the misregulation of adjacent genes by disrupting cis-regulatory elements (Hnisz et al., 2016; Katainen et al., 2015; Weinhold et al., 2014). Therefore, we collected available biotypes associated with the identified noncoding target regions and could identify target sites matching ‘predicted promoter’ (12.6%), ‘lincRNAs’ (11.3%) as well as ‘processed pseudogenes’ (4.5%) and ‘CTCF binding sites’ (3.6%) annotations (Figure 5J and Supplementary File 13). However, for 52.7% of the noncoding gRNAs no biotype or genomic annotation was available, we therefore collected the nearest 5’ and 3’ located genes and used them to perform a molecular signature analysis (Figure 5K and 13
Supplementary File 15). Among the 4 most enriched molecular signatures, we identified genes regulated by transcription factors FOXO4, KLF1, and NFAT, of which FOXO4 and NFAT downregulation has previously been reported to increased doxorubicin tolerance (Figure 5K and Supplementary File 15).

Taken together, we explored the possibility to generate a partially degenerated SpCas9-gRNA library and its application in positive selection screens. Despite limitations attributed to the generation of such a reagent and its applicability in cellular screens, we identified previously known and unknown genes presumably linked to doxorubicin resistance. In addition, we identified noncoding sequence regions and their neighboring genes for which gene set enrichment analyses revealed an enrichment for transcription factors connected to increased doxorubicin tolerance.

An optimized truly genome-wide 3Cs gRNA library

A library’s sequence diversity and distribution directly dictates the experimental scale for positive and negative selection screens. Therefore, reducing the size of the TGW library to enable coverage-based screens is highly desirable. In line with this, gRNAs truncated to 17 nucleotides have been demonstrated to maintain on-target efficiencies while reducing off-target effects (Fu et al., 2014; Wyvekens et al., 2015). We therefore truncated the degenerated TGW oligonucleotide sequence to 17 nucleotides (optimized TGA, oTGW), resembling a diversity of 1.5x10^9 (Figure 6A). Importantly, the oTGW sequence diversity is 50-times smaller than the TGW sequence diversity, while the 1.65x10^7 unique target sequences in the human genome remain identical (Figure 6A and Supplementary File 8). Similar to the TGW oligonucleotide, we used hand-mixed phosphoramidite pools to synthesize the oTGW oligonucleotide and performed 3Cs reactions by combining it with ssDNA dU-template of the three conventionally used lentiviral CRISPR/Cas plasmids pLentiGuide, pLentiCRISPRv2(Puro) and pLentiCRISPR(GFP-Puro) and determined successful 3Cs reactions by gel electrophoresis (Figure 6B). Subsequent to bacterial amplification, an I-SceI clean-up step was performed before the three oTGW libraries were analyzed by NGS with an average of 28.4 million reads per library (Figure 6C-E and Supplementary File 8).

Importantly, extracted gRNA-position nucleotide frequencies were extracted and translated to IUPAC nomenclature revealing the initial oTGW degenerated oligonucleotide sequence (Figure 14.
Furthermore, an average wildtype remnant rate of 0.2% was determined and AUC values were 0.54 or below (Figure 6-figure supplement 8), suggesting a uniform distribution of represented gRNA sequences in all three oTGW libraries (Makowski and Soares, 2003). Thus, these oTGW libraries represent the first of their kind and have the potential to elevate functional genomics approaches and will be made available to the scientific community by the Goethe University Depository: (http://www.innovectis.de/INNOVECTIS-Frankfurt/Technologieangebote/Depository).

DISCUSSION

In the present study, we describe the 3Cs technology, an improved Kunkel mutagenesis protocol, to facilitate the one-step and cloning-free generation of high-fidelity CRISPR/Cas and RNAi gene perturbation reagents. 3Cs uncouples sequence diversity from sequence distribution, making it useful for CRISPR/Cas gRNA library generation of arbitrary sequence diversities.

The 3Cs technology has several unique features; first, the bacteriophage-mediated generation of ssDNA makes the technology applicable to all plasmids containing a f1-origin of replication.

Second, ssDNA-mutagenic oligonucleotides are annealed to ssDNA of template DNA, thereby circumventing the need for two oligonucleotides per gRNA and amplification by PCR, reducing associated costs and sequence bias. In line with this, T7 DNA polymerase, used in the 3Cs reaction, has an error rate of approximately $15 \times 10^{-6}$, resulting in as little as 0.0015% of mutated heteroduplex 3Cs product, assuming 2 µg ssDNA of a 10kb plasmid (Kong et al., 1993). Third, the presence of a gRNA placeholder sequence enables the near-complete removal of wildtype plasmid remnants. Lastly, we demonstrate 3Cs applicability and performance on the example of lentiviral plasmids. However, we foresee the plasmid range to be expanded to recombinant Adeno-Associated Virus (rAAV) plasmids, adenoviruses, as well as coding sequences for protein mutagenesis that enable in cell and in vivo functional screenings.

We demonstrate fidelity and performance of 3Cs reagents by identifying the proliferative phenotype of human DUBs, validating previously known and uncovering hitherto unknown DUB phenotypes. As such, we identify USP28 and BRCC3 depletion to induce positive proliferation phenotypes.
suggesting tumor suppressive functions. In line with this, USP28 was recently identified to prevent p53 elevation in response to centrosome loss due to Plk4 inhibition, thereby preventing growth arrest in response to prolonged mitosis (Meitinger et al., 2016). On the other hand, we identify DUB enzymes whose depletion reduces cell fitness dramatically. Among them are COPS6 and USP7, both of which have been implicated in DNA damage response and enhanced p53 stability, leading to a prolonged G1 phase cell cycle arrest (Li et al., 2004). In line with this, recent work demonstrates a direct connection between CRISPR/Cas gene editing and a p53-dependent DNA damage response and associated G1 cell cycle arrest (Haapaniemi et al., 2018), suggesting the usage of DNA damage-associated DUBs (e.g. COPS6 or USP7) as p53-alternatives to control for DNA-damage induced cell cycle arrest and negative hit calling in CRISPR/Cas functional genomic screens.

CRISPR/Cas functional genomic screens are widely used to interrogate protein coding regions, while only few studies have investigated the noncoding genome through CRISPR/Cas gene editing (Canver et al., 2015; Diao et al., 2016; Korkmaz et al., 2016; Sanjana et al., 2016). Although these studies have the potential to open a new area of functional genomics, their general applicability is limited to a predefined set of gRNAs and therefore to a small subset of genomic regions. A unique feature of the 3Cs technology is the uncoupling of gRNA sequence diversity from gRNA sequence distribution, facilitating the generation of partially randomized gRNA libraries. A fully randomized library with oligonucleotides of length 20 (20N) resembles the entire space of possible gRNA sequences and comprises $4^{20} = 1.1 \times 10^{12}$ different sequences. Although an oligonucleotide pool covering this huge sequence space could be theoretically synthesized, there are at least two major reasons that render the experimental application unfeasible. The first reason is that the fraction of gRNA sequences that have a target site in the genome of interest would be very low. The ~300 million SpCas9 target sites in the human genome would be covered by a fully randomized library, but would represent only 0.027 % of the library. Consequently, the second reason is that the experimental scale would have to be extremely high to cover all naturally occurring target sites with sufficient coverage, including all non-human targeting gRNAs that also need to be included.

Screening the 20N library with a coverage of 100 and a MOI of 0.5 would require $2.2 \times 10^{14}$ cells, a cell number that is clearly not feasible in current experimental setups. By focusing on SpCas9
nucleotide preferences, we introduce the partially randomized TGW library that preferentially
targets active gRNA sequences in the entire genome, including both coding and noncoding
regions. The size of the TGW library is dramatically reduced in comparison to a fully randomized
library but still comprises $7.3 \times 10^{10}$ different gRNA sequences. We sought to explore the
experimental application of such a large library and chose to screen for resistance to the cytotoxic
agent doxorubicin in hTERT-RPE1 cells. Although insufficient TGW coverage lead to low biological
reproducibility, we were still able to retrieve gRNA overlap of three experiments. We identified
protein-coding genes that have previously been associated with doxorubicin resistance, as, for
example, CysLTR2, whose inhibition by two small chemical compounds reverted the doxorubicin-
induced toxicity. Interestingly, about half of the gRNAs for which we were able to identify a
matching sequence in the human genome map to regions in the noncoding genome. Noncoding
mutations have been shown to play pivotal roles in tumorigenesis by disrupting the function of cis-
regulatory elements (e.g. promoters, enhancer, or transcription factor binding sites) as well as
topologically associating domains (TADs), thereby directly affecting the transcriptional regulation of
adjacently located genes (Katainen et al., 2015; Weinhold et al., 2014). We annotated noncoding
regions associated with hits from our TGW screen by mapping the corresponding gRNA
sequences against the human reference genome. By allowing up to two mismatches we attempted
to account not only for exact matches but also mismatched target sites. This approach yielded a
number of potential hits which should be interpreted with caution because cut sites might be called
incorrectly and more stringent validation criteria are necessary. However, we found a small set of
gRNAs that presumably target predicted promoter sequences as well as CTCF and transcription
factor binding sites, though further validation is necessary to gain mechanistic insights on how
these sequences are link to doxorubicin resistance.

Furthermore, our computational approach for coding and noncoding hit calling is sensitive to
incorrect calling of cut sites that can lead to false positive target regions. We therefore suggest to
use the actual genome sequence of the cell line of interest to limit false positive hit calling. We
used the human reference genome (GRCh38.86), which might differ from the hTERT-RPE1
genome potentially giving misleading conclusions in terms of hit calling. Another strategy to
increase the rate of true positive hits is to increase library coverage in the experiment. However, it
is currently not feasible to screen either the 20N or the TGW library with sufficient coverage, at least not in adherent cell lines. To enable coverage-based truly-genome wide screenings, we reduced the TGW library diversity by truncating the library to 17-mers, yielding optimized TGW (oTGW) libraries that are more suited for high-throughput experiments with suspension cells. We propose the oTGW CRISPR/Cas gRNA libraries to be suited for broad biological screenings with the highest genetic target complexity. Such screens can be followed by targeted validation screens using newly synthesized libraries tailored for initially identified cut sites, which will enable rapid functional validation of such complex genetic experiments.

**Materials and Methods**

**Key Resources Table**

<table>
<thead>
<tr>
<th>Reagent type (species) or resource</th>
<th>Designation</th>
<th>Source or reference</th>
<th>RRID identifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell line (human)</td>
<td>HEK293T</td>
<td>ATCC</td>
<td>RRID:CVCL_0063</td>
</tr>
<tr>
<td>cell line (human)</td>
<td>hTERT-RPE1</td>
<td>ATCC</td>
<td>RRID:CVCL_4388</td>
</tr>
<tr>
<td>cell line (human)</td>
<td>RPE1</td>
<td>Ian Cheeseman</td>
<td></td>
</tr>
<tr>
<td>antibody</td>
<td>anti-GFP (B-2)</td>
<td>Santa Cruz Biotechnology</td>
<td>RRID:AB_627695</td>
</tr>
<tr>
<td>antibody</td>
<td>anti-alpha Tubulin</td>
<td>DSHB</td>
<td>RRID:AB_2315509</td>
</tr>
<tr>
<td>antibody</td>
<td>Goat anti-Mouse IgG (H+L) Secondary Antibody</td>
<td>Thermo Fisher Scientific</td>
<td>RRID:AB_228307</td>
</tr>
<tr>
<td>antibody</td>
<td>Goat anti-Rabbit IgG (H+L) Secondary Antibody</td>
<td>Thermo Fisher Scientific</td>
<td>RRID:AB_228341</td>
</tr>
<tr>
<td>bacteria (E. coli)</td>
<td>K12 CJ236</td>
<td>NEB (E4141)</td>
<td></td>
</tr>
<tr>
<td>bacteria (E. coli)</td>
<td>10 beta</td>
<td>NEB (C3020K)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pLentiGuide</td>
<td>Addgene (52963)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pLentiCRISPRv2</td>
<td>Addgene (52961)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>PLKO.1</td>
<td>Addgene (8453)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pPax2</td>
<td>Addgene (12260)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td>pMD2.G</td>
<td>Addgene (12259)</td>
<td></td>
</tr>
<tr>
<td>commercial kit</td>
<td>E.Z.N.A. M13 DNA Mini Kit</td>
<td>Omega Bio-Tek (D69001-01)</td>
<td></td>
</tr>
<tr>
<td>commercial kit</td>
<td>GeneJET Gel extraction kit</td>
<td>Thermo Fisher (K0692)</td>
<td></td>
</tr>
<tr>
<td>commercial kit</td>
<td>Plasmid Maxi Kit</td>
<td>Qiagen (12163)</td>
<td></td>
</tr>
<tr>
<td>commercial kit</td>
<td>PureLink Genomic DNA Mini Kit</td>
<td>Invitrogen (K1820-01)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>ampicillin</td>
<td>Roth (K029.2)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>chloramphenicol</td>
<td>Roth (3886.1)</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>kanamycin</td>
<td>Roth (T832.3)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>NaCl</td>
<td>Roth (31434)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>ATP</td>
<td>NEB (756)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>DTT</td>
<td>Cell Signaling Technology Europe (7016)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>dNTP mix</td>
<td>Roth (0178.1/2)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>Penicillin-streptomycin</td>
<td>Sigma-Aldrich (P4333)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>hygromycin</td>
<td>Capricorn Scientific (HYG-H)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>Lipofectamin 2000</td>
<td>Thermo Fisher (11668019)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>polybrene</td>
<td>Sigma Aldrich (H9268)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>doxorubicin</td>
<td>Selleckchem (S1208)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>Bay-CysLT2</td>
<td>Cayman Chemical (10532)</td>
<td></td>
</tr>
<tr>
<td>chemical compound, drug</td>
<td>Bay-u9773</td>
<td>Tocris Bioscience (3138)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>T4 DNA ligase</td>
<td>NEB (M0202)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>T7 DNA polymerase (unmodified)</td>
<td>NEB (M0274)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>2mm electroporation cuvette</td>
<td>BTX (45-0125)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>Gene Pulser electroporation system</td>
<td>BioRad (164-2076)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>I-SceI</td>
<td>NEB (R0694)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>DMEM</td>
<td>Thermo Fisher (41965-039)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>DMEM/F12</td>
<td>Thermo Fisher (11320-074)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>FBS</td>
<td>Thermo Fisher (10270)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>M13KO7 helper phage</td>
<td>NEB (N0315)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>polyethylene glycol</td>
<td>Roth (263.2)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>SOC outgrowth medium</td>
<td>Thermo Fisher (15544034)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>2YT medium</td>
<td>Roth (6676.2)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>T4 polynucleotide kinase</td>
<td>NEB (M0201)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>T7 endonuclease</td>
<td>NEB (M0302)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>OneTaq DNA polymerase</td>
<td>NEB (M0480)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>Next High-Fidelity 2x PCR Master Mix</td>
<td>NEB (M0541)</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>NextSeq 500</td>
<td>Illumina</td>
<td></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>bcl2fastq</td>
<td>Illumina RRID:SCR_015058</td>
<td></td>
</tr>
<tr>
<td>software, algorithm</td>
<td>cutadapt v.1.15</td>
<td>Martin et al., 2011 RRID:SCR_011841</td>
<td></td>
</tr>
</tbody>
</table>
Cloning of 3Cs template plasmids

The NHT and I-SceI gRNA sequences (see DNA oligonucleotides) were annealed and cloned into pLentiGuide (Addgene 52963) and pLentiCRISPRv2 (Addgene 52961) via BsmBI restriction enzyme digest (NEB, R0580) and subsequent ligation with T4 ligase (NEB, M0202). Correct clones were identified by SANGER sequencing at Microsynth SeqLab, Switzerland, using a U6 primer (see DNA oligonucleotides).

3Cs oligonucleotide design

All 3Cs oligonucleotides that were used in experiments are listed in section DNA oligonucleotides. DNA oligonucleotides were purchased from Sigma-Aldrich and Integrated DNA Technologies (IDT) as single or pooled oligonucleotides, and from Twist Bioscience or CustomArray Inc. as oligonucleotide pools. The 3Cs oligonucleotides were designed with two homology regions flanking the intended 20 nucleotide gRNA sequence. The homology regions were at least 15 nucleotides in length (\(T_m\) above 50 °C) and matched the 3’ end of the U6 promoter region and the 5’ start of the gRNA scaffold in the template plasmids. The TGW and the oTGW 3Cs oligonucleotides were designed based on a pattern of nucleotide preferences as previously determined (Doench et al., 2016; John G. Doench et al., 2014). The observed nucleotide preferences were translated into a degenerated 17-nucleotide DNA sequence (oTGW, see DNA Oligonucleotides). The randomized oligonucleotides for the six libraries of increasing diversity each had stretches of an increasing number of fully randomized nucleotides (see DNA oligonucleotides). The oligonucleotide with four randomized positions was designed to contain the stretch of four consecutive Ns beginning at position 30 of the oligonucleotide. Oligonucleotides with increasing randomization were designed by extending the randomized pattern in an alternating fashion left and right by one randomized position each. The randomized segments and the flanking constant regions were designed to
replace the I-Scel recognition site in the template plasmid to enable the clean-up digestion step. In general, every gRNA was designed to avoid occurrence of the I-Scel recognition site.

Overview of reagents and equipment needed for the 3Cs synthesis

Equipment: Desktop microcentrifuge, shaking incubator at 37 °C, collection tubes 1.5 ml, filtered, sterile pipette tips, thermoblocks at 90 °C and 50 °C (e.g., Thermo Fisher, 88870004), ultracentrifuge capable of spinning 50 ml falcon tubes at 10,000 rpm - Beckman Coulter Avanti J-30 I ultracentifuge and a Beckman JA-12 fixed angle rotor, falcon tubes - polypropylene, 50 ml (Corning 352070), Bio-Rad Gene Pulser electroporation system (BioRad 164-2076), electroporation cuvettes Plus (2mm, Model no. 620 (BTX)), gel electrophoresis chamber, erlenmeyer flasks (glass, 100 ml).

dU-ssDNA template amplification: KCM Transformation: 5x KCM buffer (0.5M KCl, 0.15M CaCl₂, 0.25M MgCl₂), escherichia coli strain K12 CJ236 (NEB, E4141), SOC outgrowth medium (ThermoFisher Scientific, 15544034), LB-agar plates supplemented with 100 µg/ml ampicillin (Roth, K029.2).

Phage amplification and ssDNA purification: 2YT media (Roth, 6676.2), M13KO7 helper phage (NEB, N0315), Ampicillin (Roth, K029.2), chloramphenicol (Roth, 3886.1), Kanamycin (Roth, T832.3), uridine (Sigma-Aldrich, U3750), 20% PEG/NaCl (20% polyethylene glycol [Roth, 0263.2], 2.5 M NaCl [Roth, 31434]), Dulbecco’s phosphate buffered saline (PBS, Sigma, D8662), E.Z.N.A. M13 DNA Mini Kit (Omega Bio-Tek, D69001-01), store purified phage in PBS at 4 °C.

3Cs-DNA synthesis: 10x TM buffer (0.1 M MgCl₂, 0.5 M Tris-HCl, pH 7.5), 10 mM ATP (NEB, 0756), 100 mM DTT (Cell Signaling Technology Europe, 7016), T4 polynucleotide kinase (NEB, M0201), 100 mM dNTP mix (Roth, 0178.1/2), T4 DNA ligase (NEB, M0202), T7 DNA polymerase (unmodified) (NEB, M0274), Thermo Fisher Scientific GeneJET Gel Extraction Kit (Thermo Fisher, K0692), 3M sodium acetate (Sigma-Aldrich, 71196).

Electroporation and I-Scel clean-up digest: 2 mm cuvette (BTX, 45-0125), electrocompetent E. coli (10-beta, NEB, C3020K), SOC outgrowth medium (Thermo Fisher, 15544034), LB-media (Roth, X964.3) supplemented with 100 µg/ml ampicillin, Qiagen Plasmid Maxi Kit (Qiagen, 12163), I-Scel
Wegner et al, 3Cs-gRNAs

(NEB, R0694), NEB CutSmart buffer (NEB, B7204), 0.5% TAE/agarose gel, Thermo Fisher Scientific GeneJET Gel Extraction Kit.

dU-ssDNA template amplification

Bacteria (*Escherichia coli* strain K12 CJ236, NEB, E4141) were transformed with 500 ng of template plasmid according to the following protocol: DNA was mixed with 2 µl of 5x KCM buffer (0.5M KCl, 0.15M CaCl₂, 0.25M MgCl₂) set to 10 µl with water and chilled on ice for 10 minutes. An equal volume of CJ236 bacteria was added to the DNA/KCM mixture, gently mixed, and incubated on ice for 15 minutes. The bacteria/DNA mixture was then incubated at room temperature for 10 minutes, and subsequently inoculated into 200 µl of prewarmed SOC media (ThermoFisher Scientific, 15544034). Bacteria were incubated at 37 °C and 200 rpm for 1 hour and then selected with ampicillin (100 µg/ml) on LB-agar plates overnight at 37 °C.

The next morning, a single colony of transformed CJ236 was picked into 1 ml of 2YT media (Roth, 6676.2) supplemented with M13KO7 helper phage (NEB, N0315) to a final concentration of 1x10⁸ pfu/ml, chloramphenicol (final concentration 35 µg/ml), and ampicillin (final concentration 100 µg/ml) to maintain the host F’ episome and the phagemid, respectively. Supplementation of uridine (Sigma-Aldrich, U3750) was set to 2.5 µM. After 2 hours of shaking at 200 rpm and 37 °C, kanamycin was added to a final concentration of 25 µg/ml to select for bacteria that have been infected with M13KO7 helper phage. Bacteria were kept at 200 rpm and 37 °C for additional 6 hours before the culture was transferred to 30 ml of 2YT media supplemented with ampicillin (final concentration 100 µg/ml) and kanamycin (final concentration 25 µg/ml). After 20 hours of shaking at 200 rpm and 37 °C, the bacterial culture was centrifuged for 10 minutes at 10,000 rpm and 4 °C in a Beckman JA-12 fixed angle rotor. To precipitate phage particles, the supernatant was transferred to 6 ml (1/5 of culture volume) PEG/NaCl (20% polyethylene glycol 8,000, 2.5 M NaCl), incubated for 1 hour at room temperature and subsequently centrifuged for 10 minutes at 10,000 rpm and 4 °C in a Beckman JA-12 fixed angle rotor. The phage pellet was resuspended in 1 ml Dulbecco’s phosphate buffered saline (PBS, Sigma, D8662), centrifuged at 13,000 rpm for 5 min, and the phage-containing supernatant was stored at 4 °C. Circular ssDNA was purified from the
resuspended phages with the E.Z.N.A. M13 DNA Mini Kit (Omega Bio-Tek, D69001-01) according to the manufacturer’s protocol and purified ssDNA was stored at 4 °C.

3Cs-DNA synthesis

Oligonucleotides that were used for 3Cs reactions and the suppliers are listed separately (see section DNA Oligonucleotides). 3Cs oligonucleotides for specific pools were mixed in equimolar ratios. 600 ng of pooled oligonucleotides were phosphorylated by mixing them with 2 µl 10x TM buffer (0.1 M MgCl₂, 0.5 M Tris-HCl, pH 7.5), 2 µl 10 mM ATP (NEB, 0756), 1 µl 100 mM DTT (Cell Signaling Technology Europe, 7016), 20 units of T4 polynucleotide kinase (NEB, M0201) and water to a total volume of 20 µl. The mixture was incubated for 1 h at 37 °C.

Phosphorylated oligonucleotides were annealed to the circular dU-ssDNA template by adding 20 µl of phosphorylation product to 25 µl 10x TM buffer, 20 µg of dU-ssDNA template, and water to a total volume of 250 µl. The mixture was denatured for 3 min at 90 °C, annealed 5 min at 50 °C, and cooled down for 5 min at room temperature.

3Cs-DNA was generated by adding 10 µl of 10 mM ATP, 10 µl of 100 mM dNTP mix (Roth, 0178.1/2), 15 µl of 100 mM DTT, 2,000 ligation units of T4 DNA ligase (NEB, M0202), and 30 units of T7 DNA polymerase (NEB, M0274) to the annealed oligonucleotide/ssDNA mixture. The 3Cs synthesis mix was incubated for 12 hours (overnight) at room temperature. The 3Cs synthesis product was then affinity purified and desalted using a Thermo Fisher Scientific GeneJET Gel Extraction Kit (Thermo Fisher, K0692) according to the following protocol: 600 µl of binding buffer and 5 µl 3M sodium acetate (Sigma-Aldrich, 71196) were added to the synthesis product, mixed and applied to two purification columns which were centrifuged for 3 min at 460 g (2,500 rpm in a Sigma-Aldrich 1-14 table top centrifuge). The flow-through was applied a second time to the same purification column to maximize yield and centrifuged for 3 min at 460 g. DNA was eluted in 40 µl warm water. The 3Cs reaction product was analyzed by gel electrophoresis alongside the dU-ssDNA template on a 0.8% TAE/agarose gel (100 V, 30 min). 3Cs-shRNA libraries were synthesized according to the above described protocol with the following modifications: in two setups, either 60 ng or 120 ng of circular template dU-ssDNA of pLKO.1 (Addgene: 1864) was used.
Electroporations and I-SceI clean-up digest

To generate pool 1 (P1) of a library, 3Cs-DNA constructs were electroporated with a cold 2 mm cuvette (BTX, 45-0125) into electrocompetent *E. coli* (10-beta, NEB, C3020K) using a Bio-Rad Gene Pulser with the following settings: resistance 200 Ohm, capacity 25 F, voltage 2.5 kV. 2 µg of DNA were mixed with 400 µl of freshly thawed cells. Electroporated cells were rescued in 25 ml of pre-warmed SOC media and incubated for 30 min at 37 °C and 200 rpm.

After 30 min of incubation, a dilution series was performed to determine the transformation efficiency and the number of transformed bacteria. 10 µl of culture were diluted to $10^{-1}$ to $10^{-12}$ and plated on LB agar plates supplemented with 100 µg/ml ampicillin. Remaining culture was added to 200 ml LB-media (Roth, X964.3) supplemented with 100 µg/ml ampicillin. Plates were incubated overnight at 37 °C, the liquid cultures were incubated overnight at 37 °C and 200 rpm. The next day the electroporation efficiency and the number of transformed bacteria were determined.

Plasmid DNA of overnight liquid cultures was purified using a Qiagen Plasmid Maxi Kit (Qiagen, 12163) according to the manufacturer’s protocol.

To generate the final pool 2 (P2) of a library, 10 µg of purified P1 DNA were digested with 50 units I-SceI (NEB, R0694) and 5 µl NEB CutSmart buffer (NEB, B7204) in a reaction volume of 50 µl for 1.5 h at 37 °C. The digestion reaction was subjected to gel electrophoresis on a 0.5% TAE/agarose gel (100 V, 30 min) to separate the undigested 3Cs synthesis product from linearized template plasmid. The band resembling the undigested correct 3Cs synthesis product was purified using a Thermo Fisher Scientific GeneJET Gel Extraction Kit according to the manufacturer’s protocol. In a second step, the purified 3Cs synthesis product was electroporated according to the electroporation protocol described above. The final P2 library preparation was purified from liquid culture using a Qiagen Plasmid Maxi Kit according to the manufacturer’s protocol and quality controlled with analytical restriction enzyme digests. 3Cs-shRNA pools were generated according to the above protocol with the following modifications. Instead of using I-SceI for the clean-up digestion, we used Bsu36I to digest template plasmid remnants in the first DNA pool (P1). P1 was electroporated using the settings described above to yield the final pool (P2). Both 3Cs-shRNA pools were purified from liquid culture using a Qiagen Plasmid Maxi Kit according to the
manufacturer’s protocol and were quality controlled with analytical restriction enzyme digests and SANGER sequencing.

**Cell culture**

HEK293T cells (ATCC, CRL-3216) were maintained in Dulbecco’s Modified Eagle’s Medium (DMEM, Thermo Fisher Scientific, 41965-039) and hTERT-RPE1 cells (ATCC, CRL-4000 and Ian Cheeseman’s) in DMEM: Nutrient Mixture F-12 (DMEM/F12, Thermo Fisher Scientific, 11320-074), each supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific, 10270) and 1% penicillin-streptomycin (Sigma-Aldrich, P4333) at 37 °C with 5% CO₂. Additionally, hTERT-RPE1 cells were supplemented with 0.01 mg/ml hygromycin B (Capricorn Scientific, HYG-H). hTERT-RPE1 cells were obtained from ATCC/LGC (CRL-4000) and Ian Cheeseman. No method to ensure the state of authentication has been applied. Mycoplasma contamination testing has been performed immediately after arrival of cells and multiple times during the course of the experiments.

**Cell extracts and antibodies**

Preparation of lysates and immunoblot analyses were performed as described previously using Tris lysis buffer (50 mM Tris–HCl pH 7.8, 150 mM NaCl, 1% IGEPAL CA-630) containing 20 mM NaF, 20 mM β-glycerophosphate, 0.3 mM Na-vanadate, 20 μg/ml RNase A, 20 μg/ml DNase and 1/300 protease inhibitor cocktail (Sigma-Aldrich, P8340) and phosphatase inhibitor cocktail #2 (Sigma-Aldrich, P5726) (Kaulich et al., 2015). Antibodies used in this study were purchased from the following sources: mouse anti-GFP (GFP (B-2): sc-9996, 1:2,000, Santa Cruz Biotechnology, Inc.), mouse anti-Tubulin (clone 12G10, 1:1,000, Developmental Studies Hybridoma Bank, University of Iowa). Secondary antibodies used for western blot analysis were goat anti-mouse (Thermo Scientific, 31430) and, goat anti-rabbit (Thermo Scientific, 31460). Mouse anti-Tubulin hybridoma cell line (clone #12G10) was developed by J. Frankel and E.M. Nelson under the auspices of the NICHD and maintained by the Developmental Studies Hybridoma Bank. Protein levels were visualized with Pierce ECL Western Blotting Substrate on a BioRad ChemiDoc MP imaging system and analyzed with Bio-Rad Image Lab software (version 4.1 build 16).
**Generation and quantification of lentiviral particles**

The day before transfection, HEK293T cells were seeded to $5 \times 10^5$ cells/ml. To transflect HEK293T cells, transfection media containing 1/10 of culture volume Opti-MEM I (Thermo Fisher Scientific, 31985-047), 10.5 µl Lipofectamin 2000 (Thermo Fisher Scientific, 11668019), 1.65 µg/ml transfer vector, 1.35 µg/ml pPAX2 (Addgene, 12260) and 0.5/µg pMD2.G (Addgene, 12259) was prepared. The mixture was incubated for 30 minutes at room temperature and added drop-wise to the media. The next morning, transfection media was replaced with fresh media to remove the transfection reagent. Lentiviral supernatant was harvested at 24 h and 48 h after transfection, pooled and stored at -80 °C.

To determine the lentiviral titer, hTERT-RPE1 cells were plated in a 24 well plate with 20,000 cells per well. The following day, cells were transduced using 8 µg/ml polybrene (Sigma, H9268) and a series of 0.5, 1, 5, and 10 µl of viral supernatant. After 3 days of incubation at 37 °C, the percentage of fluorescence-positive cells was determined by flow cytometry. The following formula was used to calculate the viral titer:

$$\text{Virustiter (transducing units} / \text{ml}) = \frac{20,000 \text{ target cells} \times \% \text{ of GFP positive cells}}{\text{volume of supernatant} (\text{ml}) \times 100}.$$ 

Alternatively, lentiviral titers were determined by colony formation titering assay for lentivirus.

**Flow cytometry**

All samples were analyzed on a FACSCanto II flow cytometer (BD Biosciences), and data were processed by FlowJo software (FlowJo, LLC). Gating was done based on viable and single cells that were identified on the basis of their scatter morphology.

**Lentiviral Transduction**

hTERT-RPE1 cells were seeded at an appropriate density for each experiment with a maximal confluency of 60-70% in DMEM/F12, supplemented with 10% FBS, 0.02 µg/ml hygromycin, and 1% penicillin-streptomycin. At the day of transduction, polybrene was added to the media to a final concentration of 8 µg/ml. The volume of lentiviral supernatant was calculated based on the...
diversity of the respective library, and the desired coverage and multiplicity of infection (MOI) of the experiment. The number of cells that were transduced at the beginning of an experiment was calculated by multiplying the diversity of the library with the desired coverage and the desired MOI. For example, the parameters for the DUB library screen were set at a coverage of 1.000 and an MOI of 0.2, i.e., one lentiviral particle per five cells. The total number of cells transduced, thus, was calculated as follows: 363 * 1,000 * 5 = 1,815,000. The next morning, the media was replaced with fresh media and the cells were subjected to antibiotic selection or experimental analysis.

**Homology arm lengths and 3Cs reaction times**

To test different homology arm lengths, four 3Cs reactions were performed using four different oligonucleotides with increasing lengths of homology to the pLentiGuide NHT, according to the 3Cs synthesis protocol described above. The reaction products were analyzed by gel electrophoresis. To monitor the 3Cs synthesis process over time, we annealed the TGW oligonucleotide to the pLentiGuide NHT and generated 3Cs-dsDNA. 2 µl of the reaction were sampled from the reaction tube and transferred to -20 °C at different timepoints starting from 0 h to 20 h. All samples were analyzed together by agarose gel electrophoresis. To visualize the kinetics of 3Cs reactions, 3Cs-dsDNA band intensities were determined and normalized to time point 0 before plotting against the time of their harvest using the Bio-Rad Image Lab software (version 4.1 build 16).

**eGFP gene editing and T7 Endonuclease I assay**

The efficiency of eGFP gene editing was analyzed by transducing eGFP-expressing hTERT-RPE1 cells with 3Cs gRNA constructs based on pLentiCRISPRv2, a subsequent T7 Endonuclease I assay, and immunoblotting. The experiment was performed in triplicates using a control gRNA (NHT), a single GFP-targeting 3Cs-gRNA (GFP#1) or a pool of six GFP-targeting 3Cs-gRNAs (GFP#1-6), respectively. After seven days of incubation at 37 °C without antibiotic selection, cells were trypsinized and the genomic DNA was purified using a PureLink™ Genomic DNA Mini Kit (Invitrogen, K1820-01) according to the manufacturer's protocol. To assess genome targeting efficiency of the 3Cs reagents, we analyzed the four cell populations that were transduced with either the NHT-gRNA, the GFP#1 gRNA, the GFP#1-6 pool, or were not
transduced at all. We PCR-amplified the GFP locus with OneTaq DNA polymerase (NEB, M0480) using 1 µg of genomic DNA, 40 µM dNTPs (final concentration), 0.2 µM of each forward and reverse amplification primers (see DNA oligonucleotides eGFP T7 forward and eGFP T7 reverse), 10x OneTaq standard buffer, and 2.5 units OneTaq DNA polymerase. The cycles were set up as follows: initial denaturation at 94 °C for 3 min, 39 cycles of denaturation at 94 °C for 20 sec, annealing at 55 °C for 30 sec, strand extension at 68 °C for 2 min, and final strand extension at 68 °C for 5 min. The PCR products were analyzed on a 0.8% TAE/agarose gel (100 V, 30 min) and purified using a Thermo Fisher Scientific GeneJET Gel Extraction Kit according to the manufacturer’s protocol. The T7 endonuclease I digestion was assembled with 6 µg of purified PCR product, 10x NEBuffer 2 water to 48 µl, denatured at 95 °C for 5 minutes, and annealed in two steps from 95-85 °C with -2 °C/second, and from 85-25 °C with -0.1 °C/second. To the annealed PCR product, 7 µl of T7 Endonuclease I (NEB, M0302) was added and incubated for 15 minutes at 37 °C. The fragmented PCR products were analyzed on a 0.8% TAE/agarose gel (100 V, 30 min) and band intensities were determined using the Bio-Rad Image Lab software (version 4.1 build 16).

**DUB proliferation screen**

The DUB proliferation screen was performed in biological duplicates. hTERT-RPE1 cells were transduced with lentiviral supernatant with a MOI of 0.2 and a library coverage of 1,000. For each replicate and time point, 2.5 million cells were seeded. Cells corresponding to the control time point were harvested two days post-transduction. All remaining cells were kept in growing and library-diversity maintaining conditions in the presence of 10 µg/ml puromycin. After 11 and 21 days, cells were harvested and their genomic DNA purified and processed for NGS. Validation of DUB screen hit candidates was performed in hTERT-RPE1 cells with 3Cs-shRNA-mediated target gene knockdown and subsequent assessment of cell proliferation using an AlamarBlue assay (Bio-Rad, BUF012A).

**Doxorubicin resistance screen**
hTERT-RPE1 cells were treated with increasing concentrations of doxorubicin, ranging from 0 to 1,000 nM, for four consecutive days. After four days, the treatment was stopped by changing media to doxorubicin-free media and cells were cultivated for another four days. After a total of 8 days, cell viability was determined and quantified with an AlamarBlue assay (Bio-Rad, BUF012A).

To screen for doxorubicin resistance, the TGW library was delivered in triplicates to a total of 5.5x10^8 hTERT-RPE1 cells with doxycycline inducible Cas9 expression via lentiviral transduction at a MOI of 1. Transduced cells were cultured for 7 days in standard media supplemented with 1 µM doxycycline (Sigma-Aldrich) and 10 µg/ml puromycin. At day 7, media was changed to selection media containing 1 µM doxorubicin (Selleckchem, S1208). After 3 weeks of selection (fresh doxorubicin every 4 days), surviving cells were harvested and processed for NGS.

**NGS of plasmid and genomic DNA**

To purify genomic DNA, surviving cells were trypsinized and pelleted. Genomic DNA was extracted using the PureLink™ Genomic DNA Mini Kit according to the manufacturer’s protocol. For NGS library preparation, 100 ng of plasmid or up to 2 µg of genomic DNA per reaction were used in a 50 µl PCR reaction using Next High-Fidelity 2x PCR Master Mix (NEB, M0541) according to the manufacturer’s protocol and 1 µl of 10 µM primers each of forward and reverse primers. Primer sequences are listed separately (see section DNA Oligonucleotides). The sequencing primers contained an 8-nucleotide long barcode sequence, enabling multiplexing of several samples in a single sequencing run and Illumina adapter sequences. Thermal cycler parameters were set as follows: initial denaturation at 98 °C for 5 min, 19 cycles of denaturation at 98 °C for 30 seconds, annealing at 55 °C for 30 seconds, extension at 72 °C for 1 minute, and final extension at 72 °C for 5 minutes. PCR products were purified from a 0.5% TAE/agarose gel using a Thermo Fisher Scientific GeneJet Gel Extraction Kit according to the manufacturer’s protocol. The purified PCR product was prepared to a final concentration of 2.4 pM in a total volume of 2.2 ml and loaded on a NextSeq 500 sequencer (Illumina), according to the manufacturer’s protocol. Sequencing was performed with single end reads, 75 cycles and 8 cycles of single index reading.

**Data processing and analysis**
All data obtained from NGS were demultiplexed using the Illumina command line tool bcl2fastq, v2.17. gRNA representation of all libraries was assessed using cutadapt v1.15 (Martin, 2011) and custom Python scripts. In brief, 3’ sequencing adapters were trimmed using a prefix of the 3’ homology sequence; trimmed reads were further trimmed by keeping only the last 20 nucleotides for all libraries except the oTGW for which the last 17 nucleotides were kept. Only reads with no ambiguously sequenced nucleotides were considered for further analyses. For the TGW and the oTGW, the resulting sequences were compared to the TGW or oTGW DNA sequence pattern, respectively, using Python and regular expressions. The reads obtained from sequencing the six randomized libraries with diversities ranging from 256 to 262,144 gRNAs were processed similarly by comparing the trimmed reads with the gRNA pattern of the respective library. For the GFP and DUB libraries, the reads were aligned to the respective sequence library. Matching sequences were counted to determine the read count distribution of a sample. Read counts of individual gRNAs for a sample were normalized by the total number of read counts that could be assigned to the respective library. The same was done for screening samples after treatment of the cells. To determine the dispersion of the read counts, the coefficient of variation was computed by dividing the standard deviation of the normalized read counts by the mean of the normalized read count $\bar{x}$, $CV = \frac{s}{\bar{x}}$. To assess the uniformity of each library distribution, we generated Lorenz curves of gRNA representation. The Lorenz curves of gRNA representation rank gRNAs by abundance scaled to 1 and show the fraction of total sequencing reads that are represented by the sum of gRNA read counts. The area under the curve (AUC) was computed in GraphPad Prism 5.0b for Mac (GraphPad Software, La Jolla California USA, www.graphpad.com) or with a custom Python script using Numpy 1.14.2 (Oliphant, 2010). Heat maps were generated by accumulating the nucleotide frequency at each position of the sequenced reads and normalized by the total number of read counts.

To correct the read counts of the six randomized libraries with diversities ranging between 256 and 262,144 gRNAs for C bias, we determined the nucleotide frequencies for each sequence position of the trimmed and final reads and normalized the observed frequencies to the expected nucleotide frequency of 25%. Each read was then scored by summing the normalized frequencies for all
reads individually. The observed read count per gRNA was then multiplied with this score, divided by the sum of all read counts that matched the respective gRNA pattern, and normalized to the sum of all corrected and normalized read counts. Lorenz curves were generated based on the corrected and normalized read counts.

Read count data from the DUB screen was analyzed by summing all individual gRNA read counts per gene and normalizing each gene read count per sample to the total number of read counts within that sample. Spearman correlation and Shapiro-Wilk confidence tests were performed to assess reproducibility of DUB screen replicates. MAGeCK and PinAPL-Py were used to analyze read counts of both replicates and calculate aggregated positive and negative proliferation phenotypes by means of log2 fold changes with associated p-values.

**Analysis of gRNA on- and off-target locations**

To determine the on and off-targets of the 4,232 hits from the doxorubicin resistance screen, CasOFFinder (v2.4) was applied to search the human genome (GRCh38.86) for gRNA target sites with up to 2 mismatches (Bae et al., 2014). The genomic positions of each on- and off-target were annotated with Ensembl genome assembly GRCh38.86, using SnpEff 4.3T (Cingolani et al., 2012) and custom Python scripts. Multiple annotations for a location were collapsed onto a single gene type and the corresponding gene name, if available. Genomic locations associated with an intergenic region were considered to be not annotated. Additional noncoding, regulatory, and pseudogene information was annotated using the Ensembl regulatory and motif features from release 91 and the Gencode consensus pseudogenes dataset from release 27 (GRCh38.p10). Additional standard annotation data from Gencode, release 27, were also included. Spearman rank and Pearson correlation were computed with NumPy (1.14.2). To determine the putative effect of gRNA off-targets on previously identified on-target locations, we mapped the gene names that were associated with off-targets back to the genes that were associated with on-target hits.

**Validation of TGW doxorubicin resistance screening hits**

To validate CysLTR2 as a doxorubicin-resistance inducing gene, we applied the CysLT2 receptor antagonists Bay-CysLT2 (Cayman Chemical, 10532) and Bay-u9773 (Tocris Bioscience, 3138). In
two different triplicates, cells were treated with each inhibitor with increasing concentrations (0 µM, 0.01 µM, 0.05 µM, 0.1 µM, 0.25 µM, 0.5 µM, 1 µM) as well as with increasing doxorubicin concentrations (0 µM, 1 µM, and 10 µM). After four days, cell survival was assessed with an AlamarBlue assay. 10% AlamarBlue was added to the cultured cells, incubated for 2 h at 37 °C and fluorescence was measured with an excitation wavelength of 560 nm and a fluorescence emission of 590 nm on a BioTek Synergy H1 microplate reader. The given measured fluorescence emissions were averaged over all replications for each experiment.

The 4,232 hits that were found in the TGW doxorubicin screen were compiled into an individual 3Cs library (validation library). The validation library was generated according to the 3Cs DNA synthesis protocol described above. We seeded 1x10^6 hTERT-RPE1 in T175 cell culture flasks in DMEM/F12 and transduced them the next day with lentivirus of the validation library using 8 µg/ml polybrene with an MOI of 0.1 and an experimental coverage of 1,000. After three days, the control cells were harvested. The screen was conducted with 2.5 µg/ml (final concentration) puromycin selection and 1 µM doxorubicin treatment. Media was changed every third day to maintain constant puromycin and doxorubicin concentrations. After three weeks of selection, surviving cells were harvested and processed for NGS according to the procedures described above.

**Molecular signatures of coding and noncoding hits**

Hits for targets with zero, one and two mismatches were merged and divided in two subsets according to the annotation, consisting of protein coding hits and noncoding hits, respectively. For the coding hits gene set, a set of 159 non-redundant genes was created from all hits with target sites in protein-coding genes. The frequency of each gene was determined. For the remaining hits, the 5 closest genes upstream and the 5 closest genes downstream to the target site were determined using GRCh.93 Ensembl gene data. The starting position of a gene and the starting position of a target site was taken as measure for proximity. A noncoding hits gene set of 1,805 non-redundant genes was created and the frequency of each gene was determined. Overlaps between both gene sets and all gene sets in Molecular Signatures Database (MSigDB) were computed using MiSigDB Web Application to Investigate Gene Sets and FDR q-value below 0.05. Heatmaps with top 4 overlapping gene sets were created using the resulting excel tables (Figure 32)
To collect interaction data, we searched the 25 genes that were associated with 'down in ERCC3 mutated cells' in the String 10.5 database by choosing up to 5 interactors in first and second shell (Szklarczyk et al., 2017) (Figure 5I).

**Data availability**

NGS data are provided as raw read count tables as Supplementary Files 1, and 17-23. Please note, raw read count tables associated with TGW and the three oTGW libraries are available from Dryad, https://doi.org/10.5061/dryad.rs432pr. Plasmids encoding oTGW 3Cs-gRNA libraries will be available through the Goethe University Depository (http://www.innovectis.de/INNOVECTIS-Frankfurt/Technologieangebote/Depository).

**Code availability**

Custom software is publicly available from GitHub, https://github.com/GEG-IBC2/3Cs.

**DNA Oligonucleotides**

DNA oligonucleotides were purchased from Sigma-Aldrich and Integrated DNA Technologies (IDT) as single or pooled oligonucleotides, and from Twist Bioscience or CustomArray Inc. as oligonucleotide pools. A detailed list of all oligonucleotides can be found as supplementary information (Supplementary File 16).

**ACKNOWLEDGEMENTS**

We thank Ian Cheeseman and Kara Lavidge McKinley for providing Cas9-inducible hTERT-RPE1 cells. We thank Tony Gutschner and Christian Münch for critical reading and commenting on the manuscript. We thank Tobias Schmidt for valuable advice and support with NGS. The hybridoma (12G10, alpha-tubulin) developed by the University of Iowa was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. This work was supported by Hessian Ministry for
Science and the Arts (HMWK, LOEWE-CGT, IIIL5-518/17.004), the German Research Foundation (DFG; CEF-MC - EXC115/2; ECCPS - EXC147/2) and in part by the LOEWE Center Frankfurt Cancer Institute (FCI) funded by the Hessen State Ministry for Higher Education, Research and the Arts [IIIL5-519/03/03.001-(0015)].

REFERENCES

Arakawa H. 2016. A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism. Sci Adv 2:e1600699–e1600699. doi:10.1126/sciadv.1600699


doi:10.4161/fly.19695

Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB,
Lassmann T, Li Q, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJSL, Reddy TE,
Xi HS, Yip KY, Zhuang J, Bernstein BE, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon
GE, Dekker J, Elnitski L, Farnham PJ, Giddings MC, Gingeras TR, Guigó R, Hubbard TJ,
Kent WJ, Lieb JD, Margulies EH, Myers RM, Stamatoyannopoulos JA, Tenenbaum SA, Weng
Z, White KP, Wold B, Yu Y, Wrobel J, Risk BA, Gunawardena HP, Kuiper HC, Maier CW, Xie
Williams BA, Zaleski C, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I,
Baer MT, Batut P, Bell I, Bell K, Chakrabortty S, Chrest J, Curado J, Derrien T, Drenkow J,
Sammeth M, Sandhu KS, Schaeffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi
H, Tilgner H, Trout D, Walters N, Wang HH, Hayashizaki Y, Reymond A, Antonarakis SE,
Hannon GJ, Ruan Y, Carninci P, Sloan CA, Learned K, Malladi VS, Wong MC, Barber GP,
Cline MS, Dreszer TR, Heitner SG, Karolchik D, Kirkup VM, Meyer LR, Long JC, Maddren M,
Raney BJ, Grasfeder LL, Giresi PG, Battenhouse A, Sheffield NC, Showers KA, London D,
Bhinge AA, Shestak C, Schaner MR, Kim SK, Zhang ZZZZ, Mieczkowski PA, Mieczkowska
C, Bansal A, Pepke S, Jain P, Amrhein H, Bowling KM, Anaya M, Cross MK, Muratet MA,
Wegner et al, 3Cs-gRNAs

Newberry KM, McCue K, Nesmith AS, Fisher-Aylor KL, Pusey B, DeSalvo G, Parker SCJSL,
Balasubramanian SS, Davis NS, Meadows SK, Eggleston T, Newberry JS, Levy SE, Absher
D, Barson G, Berry A, Bignell A, Boychenko V, Bussotti G, Davidson C, Despacio-Reyes G,
Diekhans M, Ezkurdia I, Frankish A, Gilbert J, Gonzalez JM, Griffiths E, Harte R, Hendrix DA,
Hunt T, Jungreis I, Kay M, Khurana E, Leng J, Lin MF, Loveland J, Lu Z, Manthravadi D,
RP, Auerbach RK, Balasubramanian SS, Bettinger K, Bhardwaj N, Boyle AP, Cao AR,
L, Xu X, Yan KK, Yang X, Struhl K, Weissman SM, Penalva LO, Karmakar S, Bhanvadia RR,
DL, Byron R, Canfield TK, Diegel MJ, Dunn D, Ebersol AK, Frum T, Garg K, Gist E, Hansen
RS, Boatman L, Haugen E, Humbert R, Johnson AK, Johnson EM, Kutyavin T V., Lee K,
Lotakis D, Maurano MT, Nephi SJ, Neri F V., Nguyen ED, Qu H, Reynolds AP, Roach V,
Rynes E, Sanchez ME, Sandstrom RS, Shafer AO, Stengachis AB, Thomas S, Vernot B,
Vierstra J, Vong S, Wang HH, Weaver MA, Yan Y, Zhang M, Akey JM, Bender M, Dorschner
P, Johnson N, Lukk M, Luscombe NM, Sobral D, Vaquerizas JM, Batzoglou S, Sidow A,
Hussami N, Kyriazopoulou-Panagiotopoulou S, Libbrecht MW, Schaub MA, Miller W, Bickel
PJ, Banfai B, Boley NP, Huang H, Li JJ, Noble WS, Bilmes JA, Buske OJ, Sahu AD,
of DNA elements in the human genome. Nature 489:57–74. doi:10.1038/nature11247


Wegner et al, 3Cs-gRNAs


Wegner et al, 3Cs-gRNAs


**FIGURE AND TABLE LEGENDS**

**Figure 1**

The 3Cs technology - Covalently-closed-circular-synthesized (3Cs) CRISPR/Cas gRNA reagents.

(A) Displayed is the general 3Cs workflow. Between individual steps of the protocol (grey arrows) time requirements (on top of arrow) and used or expected DNA yields (below arrow) are highlighted. Time requirements are separated by total versus hands-on time (grey scaled bars).

Please note that the protocol contains two possible break points (red stop signs) at which purified phages can be stored at 4°C (break point #1) or bacterial pellets/purified plasmid DNA can be stored at -20°C (break point #2). In more detail, f1-origin containing double-stranded CRISPR/Cas plasmids are converted to dU-containing circular ssDNA. Guide RNA sequence (orange triangle)
containing oligonucleotides (orange arrows) are annealed to ssDNA, extended and ligated by T7 DNA polymerase and T4 ligase. Heteroduplex dU-3Cs-DNA is transformed into base-excision repair-sufficient bacteria to deplete template DNA (grey strand) and selectively amplify the newly synthesized DNA (orange). (B) Lentiviral CRISPR/Cas plasmids (pLentiGuide, pLentiCRISPRv2) and the mammalian cDNA expression plasmid pcDNA3 (positive control) were converted to dU-containing circular ssDNA and analyzed by gel electrophoresis. Although identical in size, circular ssDNA appears as a single band and migrates faster than the corresponding dsDNA form. (C) Lentiviral circular ssDNA of (B) was annealed with a pool of six oligonucleotides, encoding for six GFP-targeting gRNAs, to generate a pool of 3Cs-dsDNA and analyzed by gel electrophoresis. A successful 3Cs in vitro reaction is indicated by three distinct 3Cs-dsDNA product bands (Huang et al., 2012). (D) Bar graph visualizing the degree of template remnants in final 3Cs products in presence and absence of additional Uridine in phage culture media as well as an I-SceI clean-up step. gRNA libraries from (C), were sequenced by NGS before and after I-SceI restriction enzyme digest. While the effect of Uridine is marginal, an enzymatic digest with I-SceI removes template plasmid remnants. (E, F) gRNA distribution displayed as raw read count data points (E) and normalized values in box plot format (F). The coefficient of variation was calculated by dividing the standard deviation through the mean of the library’s read counts and displayed as percentage above the box plot (F). Data were derived from NGS data of (D). The final GFP-targeting 3Cs-gRNA library is free of sequence bias, demonstrated by the low coefficient of variation of 33.18%, and the uniform sequence distribution (E, also see Figure 1-figure supplement 1H). (G) GFP-expressing hTERT-RPE1 cells were transduced with lentiviral 3Cs-gRNA constructs (non-targeting control gRNA (NHT), a single GFP-targeting 3Cs-gRNA (GFP#1) or a pool of six GFP-targeting 3Cs-gRNAs (GFP#1-6)), selected with puromycin before GFP gene editing was analyzed by T7 endonuclease I assay (Guschin et al., 2010). Individual band intensities were quantified (black numbers), empty control (-) served as reference. (H) A dose-dependent reduction of GFP fluorescence was determined by flow cytometry of GFP-expressing hTERT-RPE1 cells, transduced with increasing volumes of lentiviral supernatant, containing a pool of six GFP-targeting 3Cs-gRNAs (GFP#1-6). Error bars represent standard deviation (SD) over three biological replicates (n=3). (I) Immunoblot analysis of hTERT-REP1 cells treated as in (G) demonstrates that GFP-
targeting 3Cs-gRNAs induce a 3- to 4-fold reduction in total GFP protein levels over three biological replicates (n=3, for quantification see also **Figure 1-figure supplement 1I**).

**Figure 2**

3Cs is a robust technology and uncouples sequence diversity from sequence distribution.

(A) To determine the sequence distribution of 3Cs-gRNA libraries with increasing gRNA diversity, an increasing number of randomized nucleotides (orange) were incorporated into 3Cs oligonucleotides to mimic gRNA diversities ranging from 256 to 262,144 sequences (4-9N libraries). A range of four to nine randomized nucleotides (orange) were introduced into an NHT gRNA sequence. Randomization of the central nucleotides ensures the replacement of the template I-SceI restriction site to prevent digestion of correctly synthesized 3Cs synthesis products.

(B) The 3Cs synthesis products of the combination of randomized primers and pLentiGuide were resolved by gel electrophoresis. (C) The Scatter plot displays ranked gRNA abundances per library against the gRNA cytosine (C) content. Shown are gRNA libraries derived from (A) and (B) and the library with gRNAs targeted against DUBs (DUBs library). All libraries were processed by I-SceI-dependent removal of template plasmid remnants and subjected to NGS and computational analysis. Importantly, all gRNA libraries were complete, irrespective of their individual gRNA diversity. However, the partially randomized gRNA libraries displayed a strong C bias within the most abundant gRNA sequences. In fact, the top 5% of most abundant gRNAs had a C content of above 60%. The DUBs library did not show this C bias, strongly suggesting incomplete phosphoramidite mixing during oligonucleotide synthesis as the main cause of the C bias. (D) Lorenz curves displaying the cumulative fraction of represented NGS reads versus the gRNAs ranked by abundance of each partially randomized (4-9N) and nonrandomized (DUB) library revealed a uniform distribution of gRNA sequences. Area under the curve values (AUC, number next to library name) confirm the uniform gRNA distribution of these libraries and demonstrate that 3Cs uncouples sequence diversity from sequence distribution.

**Figure 3**

3Cs reagents are of high fidelity - Essentiality of human DUBs for cell fitness.
Schematic of the performed CRISPR screen. Highlighted are the experimental conditions under which the screen was performed (MOI of 0.2, library coverage of 1,000). In brief, hTERT-RPE1 cells were transduced with lentivirus for 48 hours in duplicates, after which the cells of one duplicate were harvested (day 0) to extrapolate the baseline gRNA distribution. Simultaneously, cells of the second duplicate were subject to puromycin selection for 11 days at which time all cells were harvested (day 11), counted and plated back in low density to the original library representation of 1,000-fold coverage. Plated cells remained in cycling conditions until day 21 at which all cells were collected (day 21). After harvesting the cells, their genomic DNA was extracted and processed for gRNA NGS. (B) Graph showing the distribution of individual sgRNAs. Highlighted are means ± standard deviation. (C-E) Volcano plots visualizing log2 fold changes of gene phenotypes and their associated p-values. Data derived from MAGeCK analyses, corresponding to day 11 (C), day 21 (D) and day 11/21 (E). The dashed red line shows p=0.05 with points above the line having p<0.05 and points below the line having p>0.05. Data points with p>0.05 are displayed transparent. Genes of interest are highlighted. (F) Venn diagram of significantly enriched (blue) or depleted (orange) DUBs. The time point overlap visualizes DUB genes with time independent (overlap of three) and time dependent (overlap of two) proliferation phenotypes. (G) Fold increase in cell number after shRNA-mediated depletion of target genes. Data are means of duplicates.

Figure 4

A truly genome-wide (TGW) CRISPR/Cas 3Cs-gRNA library to interrogate the coding and noncoding genome. (A) Previously reported SpCas9 nucleotide preferences were translated into a degenerated oligonucleotide sequence (TGW) representing a total sequence diversity of $7.3 \times 10^{10}$ (John G. Doench et al., 2014). (B) The TGW oligonucleotide (A) was used in a 3Cs reaction on template ssDNA derived from pLentiGuide and plentiCRISPRv2 plasmids to generate 3Cs-dsDNA, analyzed by gel electrophoresis. (C) Scatter plot visualizing TGW library NGS data of 14,448,469 total reads. Displayed are the log10 values of gRNA abundance (reads) against the log10 of the respective percentage of identified TGW gRNAs. 94.2% of all identified gRNAs were found only
once (see also Supplementary File 9). (D) High throughput sequencing data from (C) were used to compute the nucleotide frequency at each gRNA nucleotide position to determine the nucleotide profile of the TGW library. The identified nucleotide frequencies closely resemble the pattern of the degenerated TGW oligonucleotide from (A). Color code represents nucleotide frequency. (E) Box plot of TGW gRNA distribution with data derived from (C). The coefficient of variation of 0.26% suggests a uniform distribution of represented sequences. (F) The gRNA distribution of the TGW library as derived from (C) plotted as a Lorenz curve. TGW NGS data derived from (C). The area under the curve (AUC) of 0.52 suggests a uniform distribution of gRNA sequences.

Figure 5

TGW-based identification of coding and noncoding doxorubicin resistance associated sequences.

(A) The scheme illustrates the workflow to generate pooled lentivirus of the TGW library. DNA of 8 independent TGW 3Cs syntheses was pooled and used to transfect HEK293T cells to produce 5.5x10^8 infectious lentiviral particles. (B) Experimental workflow of the doxorubicin screen in hTERT-RPE1 cells. hTERT-RPE1 cells were transduced with TGW lentivirus with an MOI of 1, selected with puromycin for 7 days, and treated with 1 µM doxorubicin. After three weeks of continuous doxorubicin treatment, all surviving cells were collected and processed for further analysis. (C) Genomic DNA derived from three independent experiments (n=3), performed according to the scheme in (B), was used to perform NGS and gRNA sequence identification. Computational analysis identified an experimental overlap of 4,232 gRNAs (see also Supplementary File 10). (D) A 3Cs library containing the experimental overlap of 4,232 gRNAs (validation library), was generated and screened with an experimental coverage of 1,000 and an MOI of 0.1 similar to (B) (see also Figure 5-figure supplement 7). NGS of all surviving cells and computational analysis identified 795 gRNAs that were enriched more than two-fold (orange) when compared to an untreated control. (E) Pie chart visualizing the distribution of coding target regions with respect to SpCas9 off-target rate (0 to 2 mismatches). A total of 192 gRNAs (22.38% of 795 gRNAs) could be mapped to coding regions. Color code represents degree of nucleotide mismatch. (F) Chemical inhibition of CysLTR2 rescued cells from doxorubicin-mediated toxicity. hTERT-RPE1 cells were treated for 4 days with increasing concentrations of doxorubicin and two
chemical inhibitors of CysLTR2 (Bay-CysLT2 and Bay-u9773) before cellular viability was determined by AlamarBlue assays. Displayed are averaged values over three biological replicates (n=3) in arbitrary units (AU). (G) Pie chart visualizing the distribution of noncoding target regions with respect to SpCas9 off-target rate (0 to 2 mismatches). A total of 222 gRNAs (27.92% of 795 gRNAs) could be mapped to noncoding regions. Color code represents degree of nucleotide mismatch. (H) Molecular signature analysis of coding gRNA target sites identifies a set of genes, downregulated in cells expressing mutant forms of ERCC3, as top hit. 25 genes, out of the 178-coding gRNA target site-associated genes, are part of the ERCC3 group (total of 855 genes) with high confidence (p=3.7e-17). (I) List of ‘down in ERCC3 mutated cells’ 25 genes (light green), as well as their known first- and second-degree interacting genes (grey), identifies cytokinesis (DOCK1/4 genes) and vesicle transport (SEC24B/TRAPPC8 genes) gene interactions. Interaction data adapted from String 10.5. (J) Pie chart visualizing the distribution of noncoding gRNA target site annotations, including their frequency (as % of total noncoding hits). Please note: for 52.7% of all noncoding gRNA target sites, no annotation is available. (K) Molecular signature analysis of noncoding gRNA target sites, using adjacently located genes (one for each, 5’ and 3’). 33 genes, out of the 211 genes analyzed, are part of the ‘Biosynthetic process’ group (total of 1805 genes) with high confidence (p=3.4e-10).

**Figure 6**

Optimized TGW (oTGW) libraries for functional interrogations in the coding and noncoding genome. (A) oTGW oligonucleotide sequence, based on reported SpCas9 nucleotide preferences. Truncation of three 5’ nucleotides results in 17-mer gRNAs with a total oligonucleotide diversity of 1.5x10⁹. (B) oTGW 3Cs-dsDNA was synthesized on ssDNA-template of pLentiGuide, pLentiCRISPRv2-Puro and pLentiCRISPRv2-GFP/Puro. 3Cs products are analyzed by gel electrophoresis on a 0.8% TAE/agarose gel. (C-E) Removal of template plasmid remnants with I-SceI restriction enzyme digest. oTGW 3Cs-dsDNA was electroporated with efficiencies above 6.31x10⁹ and amplified for DNA purification (P1). A subsequent I-SceI restriction enzyme digest and an electroporation of P1 yielded the final 3Cs libraries containing no detectable template plasmid (P2). An analytical restriction enzyme digest with I-SceI and EcoRV removes a 2.5 kb DNA
fragment from the template plasmid (empty) and to a minor degree from P1 DNA pools. No 2.5 kb
fragment could be observed in the final P2 DNA library pools, demonstrating high purity of final
libraries (see also Figure 6-figure supplement 8). (F-H) High throughput sequencing data derived
from (C-E) were used to compute the nucleotide frequency of each gRNA nucleotide position,
visualized as heat maps. The identified nucleotide frequencies closely resemble the pattern of the
degenerated oTGW oligonucleotide (A). Color code represents nucleotide frequency (0% in blue to
50% in red).

Supplementary Figure 1

(A) Lentiviral circular ssDNA of pLentiGuide was annealed with gRNA-encoding oligonucleotides of
increasing homology (5’ and 3’, 10 to 18 nucleotides) and annealing temperatures (28°C to 45°C)
to produce 3Cs-dsDNA, analyzed by gel electrophoresis. (B, C) 3Cs-dsDNA was generated
efficiently within a few hours of incubation. Lentiviral circular ssDNA of pLentiGuide was annealed
with a single oligonucleotide (18 nts, 45°C) to produce 3Cs-dsDNA. Individual time points were
collected by removing 2 µl of the reaction mixture and transferring them to -20°C before all
samples were analyzed by gel electrophoresis (B). To visualize the kinetics of 3Cs reactions, 3Cs-
dsDNA band intensities were determined and normalized to time point 0 before plotting against the
time of their harvest (C), revealing a time-dependent 3Cs-dsDNA generation. While overnight 3Cs
reactions have been performed routinely, 2 hours of incubation time are sufficient. (D) 3Cs-dsDNA
generated on templates pLentiGuide and plentiCRISPRv2-Puro with a pool of six GFP-targeting
gRNA oligonucleotides was transformed and individual colonies were analyzed by SANGER
sequencing to identify template remnants (NHT) and gRNA-containing (GFP#1-6) plasmids. (E-F)
Removal of template remnants with an I-SceI restriction digest. Circular ssDNA of pLentiGuide
containing an I-SceI restriction site as a gRNA placeholder sequence was annealed with a pool of
six GFP-targeting gRNA oligonucleotides for 3Cs-dsDNA generation (E). Template remnant
removal by I-SceI digest (F). (G) NGS data derived from Figure 1D were used to generate Lorenz
curves and to determine the area under the curve (AUC). The AUC value of 0.56 indicates a
uniform distribution of GFP-targeting gRNA sequences. (H) Bar graph shows the reduction of
immunoblot GFP-protein intensities related to Figure 1I. Error bars represent standard deviation over two biological replicates (n=2).

**Supplementary Figure 2**

(A) Gel electrophoresis of P1 3Cs libraries, generated with randomized nucleotide positions (related to Figure 2A). Template pLentiCRISPRv2 is linearized by I-Scel digests, whereas only P1 libraries are partially I-Scel digested. P2 libraries are unaffected by I-Scel digests, demonstrating their high purity. (B) The distribution of the randomized nucleotide libraries, derived from (A), visualized with Lorenz curves. The AUC values indicate that 3Cs uncouples sequence distribution from sequence diversity.

**Supplementary Figure 3**

(A-C) 3Cs DUB gRNA screens are highly reproducible, visualized by scatter plots comparing biological replicates (Exp#1 / Exp#2) of normalized gRNA read counts per time point day 0 (A), days 11 (B) and day 21 (C). High linear regression (R² between 0.8842 and 0.9552), spearman correlation (SC between 0.8863 and 0.9731), and Shapiro–Wilk confidence (W between 0.4437 and 0.9416), including low statistical testing (p<0.0001), demonstrate high biological reproducibility. (D-F) gRNA read count tables of 3Cs DUB gRNAs were used as input for PinAPL-Py to perform positive and negative computational analyses. Experimental reproducibility is visualized by individual scatter plot per time point, day 0 (D), day 11 (E), and day 21 (F). Please note, PinAPL-Py computes Pearson and Spearman correlations, both of which values are above 0.8 over all time points and biological replicates, demonstrating high experimental reproducibility. (G-H) Scatter plot of MAGeCK-derived positive and negative gene ranks, corresponding to data of Figure 3E. Genes of interest are highlighted. (I) Scatter plot shows normalized gene ranks of day 11 versus day 21. Data corresponding to Figure 3E.
Supplementary Figure 4

(A) Gel electrophoresis of ssDNA of the conventional shRNA expressing lentiviral plasmid pLKO.1.

(B) 3Cs-shRNA oligonucleotide design. Regions with homology to pLKO.1 are shown in light grey, complementary shRNA sequences are displayed in dark grey. 3Cs-shRNAs have a length of 78 nucleotides. (C) Two different amounts of pLKO.1 circular ssDNAs (60 ng and 120 ng) were annealed with a single 3Cs-shRNA oligonucleotide, encoding for a GFP-targeting shRNA, to produce 3Cs-dsDNA, analyzed by gel electrophoresis. (D) 3Cs-dsDNA derived of (C) was amplified in E. coli. SANGER-sequencing of the amplified plasmid DNA confirmed the successful integration of the GFP-targeting shRNA sequences (highlighted in red). (E) GFP-expressing hTERT-RPE1 cells were transduced with either a scrambled or a GFP-targeting shRNA and analyzed by flow-cytometry in biological replicates (n=2). Please note, the GFP intensity was reduced by more than 60%, indicating functionality of 3Cs-shRNA reagents.

Supplementary Figure 5

(A) A 3Cs synthesis with ssDNA template of pLKO.1 and a pool of 3Cs oligonucleotides, encoding for shRNAs targeting all human ubiquitin E2 ligases (each E2 with two shRNAs), yielded 3Cs-dsDNA that is resolved by gel electrophoresis. (B) Similar to 3Cs-gRNA reagents, 3Cs-shRNA reagents are subjected to the removal of template remnants by performing an initial (E2.1) and analytical (E2.2) Bsu36I restriction enzyme digest. Please note the absence of the 3 kb DNA fragment from the E2.2 library. (C) SANGER sequencing of the final E2 3Cs-shRNA library (E2.2) confirmed a randomization of forward and reverse-complement shRNA sequences. (D) Initial and final E2 3Cs-shRNA libraries were subject to high throughput sequencing. Individual shRNA distributions were displayed as a box plot with their respective coefficients of variation (CV, percent number above box plot). (E) The distribution of both pools visualized as Lorenz curves, comparing E2.1 and E2.2. Data derived from (D). Importantly, while effectively removing template plasmid remnants, the Bsu36I digest does not affect the distribution of 3Cs-shRNAs. (F) Scatter plot of normalized read counts of data derived from (D), visualizing individual shRNA abundance before (-) and after (+) Bsu36I digest. The Pearson correlation coefficient of $R^2$ is 0.9687, demonstrating
high correlation of individual shRNA abundances and no effect of the clean-up digestion on the final distribution.

**Supplementary Figure 6**

(A) Doxorubicin toxicity in hTERT-RPE1 cells. Cells were incubated for four consecutive days with increasing concentrations of doxorubicin before cellular viability was assessed by AlamarBlue assay (orange). To quantify cellular recovery after 4 days of doxorubicin treatment, cells were exposed to doxorubicin with increasing concentrations ranging from 0 to 1,000 nM for 4 consecutive days, after which the treatment was stopped and the cells were cultivated for another four days without doxorubicin. After a total of 8 days, cellular viability was determined with an AlamarBlue assay (green). Concentrations of doxorubicin above 250 nM induced robust and irreversible toxicity in hTERT-RPE1 cells. Error bars represent standard deviation (SD) over two biological replicates (n=2). (B) Scatter plot correlating the abundance of the identified TGW gRNAs between biological replicates of the TGW validation screen with 4,232 gRNAs. gRNA read counts of each experiment were normalized towards the sum of all reads per sample and ranked according to their experimental abundance. Spearman correlations of ranked gRNA abundances between experimental replicates were computed. Colors indicate the correlation between experiments: orange (Exp. 1:2), blue (Exp. 1:3), and green (Exp. 2:3) and numbers indicate the respective Spearman correlation efficient ($R^2$ values between 0.75 and 0.82).

**Supplementary Figure 7**

(A) Gel electrophoresis visualizing the quality of P1 and P2 preparations of the TGW validation library (4,232 gRNAs). Please note the absence of the 3 kb DNA fragment in the final (P2) validation library. (B) The distributions of TGW validation library P1 and P2 preparations, visualized as Lorenz curves. Pre-I-SceI digested library (P1) and post-I-SceI digested library (P2) have similar gRNA distributions. The low area under the curve (AUC) values of 0.65 (P1) and 0.64 (P2) indicate that the I-SceI clean-up digestion does not affect the distribution of gRNAs in the final product.
**Supplementary Figure 8**

(A) Bar graph visualizing the low rate of template plasmid remnants in the final oTGW gRNA libraries. The rate of template plasmid remnants is below 0.3 % in all libraries. NGS data derived from Figure 6F-H. (B-D) The distribution of each oTGW library visualized as Lorenz curves. Please note the low AUC values of between 0.5 and 0.54, demonstrating uniform gRNA distributions of represented gRNAs.

**Supplementary File 1**

3Cs-gRNA GFP library - NGS analysis including total read counts.

**Supplementary File 2**

List of 4-9N 3Cs libraries - NGS analysis including total read counts per library.

**Supplementary File 3**

3Cs-gRNA DUB library - NGS analysis including total read counts of DUB library

**Supplementary File 4**

3Cs-gRNA DUB screen - NGS analysis including total read counts and normalizations.

**Supplementary File 5**

3Cs-gRNA DUB screen - NGS analysis using MAGeCK.

**Supplementary File 6**

3Cs-gRNA DUB screen - NGS analysis using PinAPL-Py.

**Supplementary File 7**

List of 3Cs-shRNA E2 library sequences and total NGS read counts.
Supplementary File 8
Number of TGW and oTGW target sequences per human chromosome and total oTGW NGS read counts per library.

Supplementary File 9
3Cs TGW library – List of total NGS read counts.

Supplementary File 10
List of 4232 3Cs TGW gRNA sequences derived from doxorubicin screen.

Supplementary File 11
3Cs-gRNA TGW validation screen - NGS analysis including raw read counts, normalizations and ratios.

Supplementary File 12
Annotation list of validated coding TGW hits.

Supplementary File 13
Annotation list of validated noncoding TGW hits.

Supplementary File 14
List of coding hits - molecular signature analysis.

Supplementary File 15
List of noncoding hits - molecular signature analysis.

Supplementary File 16
List of DNA oligonucleotides.

**Supplementary File 17**
Raw sequencing counts of the six randomized libraries.

**Supplementary File 18**
DUB library raw sequencing counts.

**Supplementary File 19**
DUB screen raw sequencing read counts, includes day 0, day 11, day 21 for two replicates.

**Supplementary File 20**
Raw sequencing counts of the E2-shRNA library.

**Supplementary File 21**
Raw sequencing reads of the TGS screen triplicates.

**Supplementary File 22**
Raw sequencing reads of the TGW screen validation library.

**Supplementary File 23**
Raw sequencing reads of CTRL and Treatment samples of the TGW validation screen.

**COMPETING INTERESTS**
The Goethe University Frankfurt has filed a patent related to this work on which M.K., A.E., M.W., V.D., R.D.B., and S.W. are inventors (WO2017EP84625). M.K. and I.D. are co-founders and shareholders of Vivlion GmbH. All other authors declare no competing interests.
Figure 1, Wegner et al.

A

Break point #1

- f1-origin plasmid into dut/ung bacteria
- pick colony, add M13K07(10^7)
- expand culture
- ssDNA to PO4-oligos, add T7 DNA polymerase
- electroporate into dut/ung bacteria

Break point #2

- purify dsDNA, perform QC

<table>
<thead>
<tr>
<th>Total vs. hands-on time</th>
<th>2d / 1hr</th>
<th>1d / 5hr</th>
<th>1d / 4hr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1kb Plus</td>
<td>&gt;30ug</td>
<td>&gt;2ug:60ng</td>
<td>&gt;3ug:60ng</td>
</tr>
</tbody>
</table>

B

pcDNA3 Guide CRISPRv2

C

GFP #1-6 gRNA

D

GFP gRNA pair copy number (log10)

E

GFP gRNA pair copy number (log10)

F

gRNA pair copy number (reads to median normalized, log10)

G

+ T7 Endonuclease I

H

% GFP-positive cells

I

Tub GFP

NHT GFP

GFP #1-6
Figure 2, Wegner et al.

<table>
<thead>
<tr>
<th>name</th>
<th>sequence (5’ - 3’)</th>
<th>diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>wt</td>
<td>gacgaaacacctaggatataacaggtataaatgttttaga</td>
<td></td>
</tr>
<tr>
<td>4N</td>
<td>CTGCTTTGctcgagtgNNNNNagttgacCAAATCT</td>
<td>256</td>
</tr>
<tr>
<td>5N</td>
<td>CTGCTTTGctcgagtgNNNNNagttgacCAAATCT</td>
<td>1,024</td>
</tr>
<tr>
<td>6N</td>
<td>CTGCTTTGctcgagtgNNNNNagttgacCAAATCT</td>
<td>4,096</td>
</tr>
<tr>
<td>7N</td>
<td>CTGCTTTGctcgagtgNNNNNagttgacCAAATCT</td>
<td>16,384</td>
</tr>
<tr>
<td>8N</td>
<td>CTGCTTTGctcgagNNNNNNNagttgacCAAATCT</td>
<td>65,536</td>
</tr>
<tr>
<td>9N</td>
<td>CTGCTTTGctcgagNNNNNNNagttgacCAAATCT</td>
<td>262,144</td>
</tr>
</tbody>
</table>

A B

C

D

C content (%)

gRNAs ranked by abundance

DsDNA + + + + + +
ssDNA + + + + + +
3Cs-dsDNA + + + + + +

DUBs, 0.60

gRNAs ranked by abundance

cumulative fraction of NGS reads represented

4N, 0.57
5N, 0.58
6N, 0.61
7N, 0.70
8N, 0.71
9N, 0.70
DUBs, 0.60
Figure 3, Wegner et al.

A

3Cs DUB gRNA library

MOI 0.2 Cov 1,000
day 0 harvest cells, gDNA extraction
day 11 harvest cells, gDNA extraction
day 21 harvest cells, gDNA extraction

C D E

C

positive
negative

log2 fold change (d0/d11) log2 fold change (d11/21)
p-value (log10)

F

enriched DUBs

day 11
USP17L7
USP26

BRCC3

USP9Y

USP28

USP7

USP17L12
USP17L26
USP17L25
USP9X

day 21

G

Fold change in cell no.

Time (days)

Luciferase

USP28

BRCC3

BAP1

USP7

COPS6

0 5 10 15

0 10 20
Figure 4, Wegner et al.

A

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' - 3')</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGW</td>
<td>NNDNNNNHNNNHDHVVR</td>
<td>7.3x10^10</td>
</tr>
</tbody>
</table>

B

C

D

E

F

0.26%
A. Pooling lentivirus

- HEK293T transfection
- Lentiviral supernatant
- Titer determination

B. TGW gRNA library

C. HEK293T transfection

- 8 x TGW 3Cs-reactions (each 1.5E9 electrop. eff.)
- Pooling lentivirus, 55 mL with >1E7 titer (5.5E8 different TGW 3Cs-gRNAs)

D. HEK293T transfection

- MOI 1
- harvest cells, gDNA extraction
- Doxorubicin (1µM)

E. TGW gRNA library

F. Doxorubicin (µM)

G. Coding Hits

- Down in ERCC3 Catalytic Activity
- Regulation Unknown

H. Coding Hits

- 25/855
- 29/1518
- 31/1791
- 30/1890

I. Coding Hits

- Predicted promoter (12.6%)
- lincRNAs (11.3%)
- Open chromatin (6.8%)
- Antisense (5.0%)
- Processed pseudogene (4.5%)
- CTCF binding site (5.6%)
- TF binding site (3.1%)
- Predicted enhancer (0.4%)
- No annotation available (52.7%)

J. Biosynthetic process

- 33/1805
- 33/1890

K. Biosynthetic process

- 1.9e-14
- 6.8e-14
- 3.1e-13
- 1.4e-12
- 3.9e-12
- 1.9e-11
- 2.8e-11
Figure 6, Wegner et al.

A

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence (5' - 3')</th>
<th>Diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>oTGW</td>
<td>--NNNNHNNNHNNHNVDHVNR</td>
<td>1.5x10^6</td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th>pLentiCRISPRv2-GFP/Puro</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
</tbody>
</table>

D

<table>
<thead>
<tr>
<th>pLentiGuide</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
</tbody>
</table>

E

<table>
<thead>
<tr>
<th>pLentiCRISPRv2-Puro</th>
</tr>
</thead>
<tbody>
<tr>
<td>empty</td>
</tr>
<tr>
<td>P1</td>
</tr>
<tr>
<td>P2</td>
</tr>
</tbody>
</table>
Supplementary Figure 1, Wegner et al.

A. pLentiGuide

<table>
<thead>
<tr>
<th>dsDNA +</th>
<th>ssDNA +</th>
<th>3Cs incubation time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>180</td>
</tr>
<tr>
<td></td>
<td></td>
<td>240</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1,260</td>
</tr>
</tbody>
</table>

B. dsDNA + ssDNA + 3Cs incubation time (min)

C. kb: 12

D. NHT NHT

E. pLentiCRISPRv2-Puro

F. pLentiCRISPRv2-Puro (wildtype plasmid)

G. Fold Reduction GFP (AU)

H. gRNAs ranked by abundance AUC, 0.56
Supplementary Figure 2, Wegner et al.

A

<table>
<thead>
<tr>
<th></th>
<th>4N</th>
<th>5N</th>
<th>6N</th>
<th>7N</th>
<th>8N</th>
<th>9N</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>P2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

B

- gRNAs ranked by abundance
- Cumulative fraction of NGS reads represented
- 4N, 0.60
- 5N, 0.60
- 6N, 0.64
- 7N, 0.73
- 8N, 0.74
- 9N, 0.73
Supplementary Figure 3, Wegner et al.

A. Day 0

Day 11

Day 21

D. Correlation Control_1 Control_2

E. Correlation T11_1 T11_2

F. Correlation T21_1 T21_2

G. day 11

H. day 21

I. day 11 (norm. gene rank)
Supplementary Figure 4, Wegner et al.

A

B

C

D

E

Supplementary Figure 4, Wegner et al.
### Supplementary Figure 5, Wegner et al.

#### A

<table>
<thead>
<tr>
<th></th>
<th>pLKO.1 WT</th>
<th>E2.1</th>
<th>E2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>dsDNA</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ssDNA</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3Cs-dsDNA</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

#### B

![Gel images showing dsDNA, ssDNA, and 3Cs-dsDNA with Bsu36I digestion](image)

#### C

<table>
<thead>
<tr>
<th>pLKO.1</th>
<th>CGGTCCTAAGGTTAAGTCGCCCTCGACTCGAGCGAGGGCGACTTAACCTTAGGTTTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3Cs-shRNA</td>
<td>CGGTCNGNNNNNNNNNNNNNNNNNNNCTCGAGNNNNNNNNNNNNNNNNNGGTTTT</td>
</tr>
</tbody>
</table>

#### D

**ShRNA Pair Copy Number**

E2.1: 39.9%

E2.2: 37.9%

#### E

**Cumulative Fraction of NGS Reads Represented**

E2.1 (AUC 0.66)

E2.2 (AUC 0.68)

#### F

**Norm. Read Count (E2.1) vs. Norm. Read Count (E2.2)**

- Bsu36I: $R^2 = 0.9687$
Supplementary Figure 6, Wegner et al.

(A) Viability (A.U.) as a function of Doxorubicin concentration over time.

(B) Scatter plot showing Spearman rank correlation coefficients for different ratios.
Supplementary Figure 7, Wegner et al.

A

Validation library of 4,232 gRNAs

<table>
<thead>
<tr>
<th>pLentiCRISPRv2-Puro/GFP</th>
<th>WT</th>
<th>P1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1kb Plus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-SceI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-SceI+EcoRV</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B

cumulative fraction of NGS reads represented

gRNAs ranked by abundance

P1(0.65)
P2(0.64)
Supplementary Figure 8, Wegner et al.

A

Assigned reads (%)

99.8
99.6
99.4
99.2
99.0

pLentiCRISPRv2-GFP/Puro

pLentiGuide

pLentiCRISPRv2-Puro

oTGW WT

B

C

D

GnRNAs ranked by abundance

cumulative fraction of NGS reads represented

AUC, 0.54

AUC, 0.50

AUC, 0.51

0.0 0.2 0.4 0.6 0.8 1.0

0.0 0.2 0.4 0.6 0.8 1.0

0.0 0.2 0.4 0.6 0.8 1.0