Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells

Ophir Shalem#1,2, Neville E. Sanjana#1,2, Ella Hartenian1, Xi Shi1,3, David A. Scott1,2, Tarjei Mikkelson1, Dirk Heckl4, Benjamin L. Ebert4, David E. Root1, John G. Doench1, and Feng Zhang1,2,†

1Broad Institute of MIT and Harvard, 7 Cambridge Center, MA 02142, USA
2McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
3Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA
4Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

# These authors contributed equally to this work.

Abstract

The simplicity of programming the CRISPR-associated nuclease Cas9 to modify specific genomic loci suggests a new way to interrogate gene function on a genome-wide scale. We show that lentiviral delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeting 18,080 genes with 64,751 unique guide sequences enables both negative and positive selection screening in human cells. First, we used the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, we screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Our highest-ranking candidates include previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. We observe a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, demonstrating the promise of genome-scale screening with Cas9.

A major goal since the completion of the Human Genome Project is the functional characterization of all annotated genetic elements in normal biological processes and disease (1). Genome-scale loss-of-function screens have provided a wealth of information in diverse model systems (2–5). In mammalian cells, RNA interference (RNAi) is the predominant method for genome-wide loss-of-function screening (2, 3), but its utility is limited by the
inherent incompleteness of protein depletion by RNAi and confounding off-target effects (6, 7).

The RNA-guided CRISPR (clustered regularly interspaced short palindrome repeats)-associated nuclease Cas9 provides an effective means of introducing targeted loss-of-function mutations at specific sites in the genome (8, 9). Cas9 can be programmed to induce DNA double strand breaks (DSBs) at specific genomic loci (8, 9) through a synthetic single guide RNA (sgRNA) (10), which when targeted to coding regions of genes can create frame shift indel mutations that result in a loss-of-function allele. Because the targeting specificity of Cas9 is conferred by short guide sequences, which can be easily generated at large scale by array-based oligonucleotide library synthesis (11), we sought to explore the potential of Cas9 for pooled genome-scale functional screening.

Lentiviral vectors are commonly used for delivery of pooled short hairpin RNAs (shRNAs) in RNAi since they can be easily titrated to control transgene copy number, and are stably maintained as genomic integrants during subsequent cell replication (2, 12, 13). Therefore we designed a single lentiviral vector to deliver Cas9, a sgRNA, and a puromycin selection marker into target cells (lentiCRISPR) (Fig. 1A). The ability to simultaneously deliver Cas9 and sgRNA through a single vector enables application to any cell type of interest, without the need to first generate cell lines that express Cas9.

To determine the efficacy of gene knockout by lentiCRISPR transduction, we tested six sgRNAs targeting enhanced green fluorescent protein (EGFP) in a HEK293T cell line containing a single copy of EGFP (fig. S1). After transduction at a low multiplicity of infection (MOI = 0.3) followed by selection with puromycin, lentiCRISPRs abolished EGFP fluorescence in 93 ± 8% (mean ± s.d.) of cells after 11 days (Fig. 1B). Deep sequencing of the EGFP locus revealed a 92 ± 9% indel frequency (n ≥ 10^4 sequencing reads per condition) (fig. S2). In contrast, transduction of cells with lentiviral vectors expressing EGFP-targeting shRNA led to incomplete knockdown of EGFP fluorescence (Fig. 1C).

Given the high efficacy of gene knockout by lentiCRISPR transduction, we tested the feasibility of conducting genome-scale CRISPR-Cas9 knockout (GeCKO) screening with a pooled lentiCRISPR library. We designed a library of sgRNAs targeting 5′ constitutive exons (Fig. 2A) of 18,080 genes in the human genome with an average coverage of 3-4 sgRNAs per gene (table S1), and each target site was selected to minimize off-target modification (14) (supplementary discussion).

To test the efficacy of the full GeCKO library at achieving knock out of endogenous gene targets, we conducted a negative selection screen by profiling the depletion of sgRNAs targeting essential survival genes (Fig. 2A). We transduced the human melanoma cell line A375 and the human stem cell line HUES62 with the GeCKO library at a MOI of 0.3. As expected, deep sequencing (figs. S3 and S4) 14 days post-transduction revealed a significant reduction in the diversity of sgRNAs in the surviving A375 and HUES62 cells (Fig. 2, B and C) (Wilcoxon rank sum test, P < 10^-10 for both cell types). Gene set enrichment analysis (GSEA) (15) indicated that most of the depleted sgRNAs targeted essential genes such as ribosomal structural constituents (Fig. 2, D and E, and tables S2 and S3). The overlap in
highly depleted genes and functional gene categories between the two cell lines (fig. S5) indicates that GeCKO can identify essential genes and that enrichment analysis of depleted sgRNAs can pinpoint gene targets in negative selection screens. To test the efficacy of GeCKO for positive selection, we sought to identify gene knockouts that result in resistance to the BRAF protein kinase inhibitor vemurafenib (PLX) in melanoma (16) (Fig. 3A). Exposure to PLX resulted in growth arrest of transduced A375 cells, which harbor the V600E gain-of-function BRAF mutation (17) (Fig. 3B), therefore enabling the enrichment of a small group of cells that were rendered drug-resistant by Cas9:sgRNA-mediated modification. After 14 days of PLX treatment, the sgRNA distribution was significantly different when compared with vehicle-treated cells (Fig. 3C) (Wilcoxon rank-sum test, \( P < 10^{-10} \)) and clustered separately from all other conditions (Fig. 3D and fig. S6).

For a subset of genes, we found enrichment of multiple sgRNAs that target each gene after 14 days of PLX treatment (Fig. 3E), suggesting that loss of these particular genes contributes to PLX resistance. We used the RNAi Gene Enrichment Ranking (RIGER) algorithm to rank screening hits by the consistent enrichment among multiple sgRNAs targeting the same gene (Fig. 3F and table S4) (12). Our highest ranking genes included previously reported candidates NF1 and MED12 (18, 19) and also several genes not previously implicated in PLX resistance, including neurofibromin 2 (NF2), Cullin 3 E3 ligase (CUL3), and members of the STAGA histone acetyltransferase complex (TADA1 and TADA2B). These candidates yield new testable hypotheses regarding PLX resistance mechanisms (supplementary discussion). For example, NF1 and NF2, although unrelated in sequence, are each mutated in similar but distinct forms of neurofibromatosis (20). In addition, epigenetic dysregulation resulting from mutations in the mechanistically related STAGA and Mediator complexes (21) may have a role in acquired drug resistance. All of these hits were also identified through a second independent transduction (figs. S7 and S8, and tables S5 and S6).

A similar screen to identify PLX drug resistance in A375 cells was previously conducted using a pooled library of 90,000 shRNAs (19). To compare the efficacy and reliability of genome-scale shRNA screening with GeCKO, we used several methods to evaluate the degree of consistency among the sgRNAs or shRNAs targeting the top candidate genes. First, we calculated the aggregate \( P \) value distribution for the top 100 hits using either RIGER (Fig. 4A) or RSA (fig. S9) scoring. Lower \( P \) values for the GeCKO versus shRNA screen indicate better scoring consistency among sgRNAs. Second, for the top 10 RIGER hit genes, 78 ± 27% of sgRNAs targeting each gene ranked among the top 5% of enriched sgRNAs, whereas 20 ± 12% of shRNAs targeting each gene ranked among the top 5% of enriched shRNAs (Fig. 4B).

We validated top ranking genes from the GeCKO screen individually using 3-5 sgRNAs (Fig. 4, C to E, and figs. S10 and S11). For NF2, we found that 4/5 sgRNAs resulted in >98% allele modification 7 days post-transduction, and all 5 sgRNAs showed >99% allele modification 14 days post-transduction (Fig. 4C). We compared sgRNA and shRNA-mediated protein depletion and PLX resistance using Western blot (Fig. 4D) and cell growth assays (Fig. 4E). Interestingly, while all five sgRNAs conferred resistance to PLX, only the best shRNA achieved sufficient knockdown to increase PLX resistance (Fig. 4E), suggesting...
that even low levels of NF2 are sufficient to retain sensitivity to PLX. Additionally, sgRNAs targeting NF1, MED12, CUL3, TADA1, and TADA2B led to a decrease in protein expression and increased resistance to PLX (figs. S10 and S11). Deep sequencing confirmed a high rate of mutagenesis at targeted loci (figs. S12 and S13), with a small subset of off-target sites exhibiting indels (figs. S14 to S16), which may be alleviated using an offset nicking approach (22, 23) that was recently shown to reduce off-target modifications (22).

GeCKO screening provides a mechanistically distinct method to RNAi for systematic perturbation of gene function. Whereas RNAi reduces protein expression by targeting RNA, GeCKO introduces loss-of-function mutations into genomic DNA. While some indel mutations are expected to maintain the reading frame, homozygous knockout yields high screening sensitivity, which is especially important in cases where incomplete knockdown retains gene function. In addition, RNAi is limited to transcripts, whereas Cas9:sgRNAs can target elements across the entire genome, including promoters, enhancers, introns, and intergenic regions. Furthermore, catalytically inactive mutants of Cas9 can be tethered to different functional domains (23–27) to broaden the repertoire of perturbation modalities, including genome-scale gain-of-function screening using Cas9-activators and epigenetic modifiers. In the GeCKO screens presented here, the efficiency of complete knockout, the consistency of distinct sgRNAs, and the validation rate for top screen hits demonstrate the potential of Cas9:sgRNA-based technology to transform functional genomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank G. Cowley, W. Harrington, J. Wright, E. Hodis, S. Whittaker, J. Merkin, C. Burge, L. P. Club, and the entire Zhang laboratory for technical support and critical discussions. O.S. is a Klarman Fellow, N.S. is a Simons Fellow, D.A.S. is an NSF Fellow, and J.D. is a Merkin Institute Fellow. F.Z. is supported by an NIH Director’s Pioneer Award (1DP1-MH100706), a NIH Transformative R01 grant (1R01-DK097768), the Keck, McKnight, Merkin, Vallee, Damon Runyon, Searle Scholars, Klingenstein, and Simons Foundations, Bob Metcalfe, and Jane

References and Notes


Fig. 1. Lentiviral delivery of Cas9 and sgRNA provides efficient depletion of target genes

(A) Lentiviral expression vector for Cas9 and sgRNA (lentiCRISPR). Puromycin selection
rev marker (puro), psi packaging signal (psi+), response element (RRE), central polypurine
tract (cPPT), elongation factor-1α short promoter (EFS), 2A self-cleaving peptide (P2A),
and posttranscriptional regulatory element (WPRE).

(B) Distribution of fluorescence from
293T-EGFP cells transduced by EGFP-targeting lentiCRISPR (sgRNAs 1-6, outlined peaks)
and Cas9-only (green-shaded peak) vectors, and non-fluorescent 293T cells (gray shaded
peak).

(C) Distribution of fluorescence from 293T-EGFP cells transduced by EGFP-
targeting shRNA (shRNAs 1-4, outlined peaks) and control shRNA (green-shaded peak)
vectors, and non-fluorescent 293T cells (gray shaded peak).
Fig. 2. GeCKO library design and application for genome-scale negative selection screening
(A) Design of sgRNA library for genome-scale knockout of coding sequences in human cells (supplementary discussion). (B and C) Cumulative frequency of sgRNAs 3 and 14 days post transduction in A375 and hES cells respectively. Shift in the 14 day curve represents the depletion in a subset of sgRNAs. (D and E) Five most significantly depleted gene sets in A375 cells \( p < 10^{-5} \), FDR-corrected \( q < 10^{-5} \) and HUES62 cells (nominal \( p < 10^{-5} \), FDR-corrected \( q < 10^{-3} \) indentified by Gene Set Enrichment Analysis (DSEA) (15).
Fig. 3. GeCKO screen in A375 melanoma cells reveals genes whose loss confers vemurafenib (PLX) resistance

(A) Timeline of PLX resistance screen in A375 melanoma cells. (B) Growth of A375 cells when treated with DMSO or PLX over 14 days. (C) Boxplot showing the distribution of sgRNA frequencies at different time points, with and without PLX treatment (vehicle = DMSO). The box extends from the first to the third quartile with the whiskers denoting 1.5 times the interquartile range. Enrichment of specific sgRNAs: 7 days of PLX treatment, 1 sgRNA greater than 10-fold enrichment; 14 days of PLX treatment, 379 and 49 sgRNAs greater than 10-fold and 100-fold enrichment respectively. (D) Rank correlation of normalized sgRNA read count between biological replicates and treatment conditions. (E) Scatterplot showing enrichment of specific sgRNAs after PLX treatment. (F) Identification of top candidate genes using the RNAi Gene Enrichment Ranking (RIGER) P value analysis.
Fig. 4. Comparison of GeCKO and shRNA screens and validation of neurofibromin 2 (NF2)

(A) RIGER p values for the top 100 hits from GeCKO and shRNA (19) screens for genes whose loss results in PLX resistance. Analysis using the Redundant siRNA Activity (RSA) algorithm shows a similar trend (fig. S9). (B) For the top 10 RIGER hits, the percent of unique sgRNAs (top) or shRNAs (bottom) targeting each gene that are in top 5% of all enriched sgRNAs or shRNAs. (C) Deep sequencing analysis of lentiCRISPR-mediated indel at the NF2 locus. (D) A375 cells transduced with NF2-targeting lentiCRISPR and shRNA vectors both show a decrease in NF2 protein levels. (E) Dose response curves for A375 cells transduced with individual NF2-targeting lentiCRISPR or shRNA vectors. Controls were EGFP-targeting lentiCRISPR or null hairpin shRNA vectors. Cells transduced with NF2-targeting lentiCRISPRs show a significant increase ($F_{1,8} = 30.3$, $p < 0.001$, $n = 4$ replicates) in the half maximal effective concentration ($EC_{50}$) whereas cells transduced with NF2-targeting shRNA vectors do not ($F_{1,8} = 0.47$, $p = 0.51$, $n = 4$ replicates).