Sciencexpress Reports

Genetic Screens in Human Cells Using the CRISPR/Cas9 System

Tim Wang,[1,2,3,4] Jenny J. Wei,[1,2] David M. Sabatini,[1,2,3,4,5]*† Eric S. Lander[1,3,6]*†

[1]Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge MA 02139, USA. [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA. [3]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA. [4]David H. Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA 02139, USA. [5]Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [6]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA.

*Corresponding author. E-mail: sabatini@wi.mit.edu (D.M.S.); lander@broadinstitute.org (E.S.L.)

†These authors contributed equally to this work.

The bacterial CRISPR/Cas9 system for genome editing has greatly expanded the toolbox for mammalian genetics, enabling the rapid generation of isogenic cell lines and mice with modified alleles. Here, we describe a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library. sgRNA expression cassettes were stably integrated into the genome, which enabled a complex mutant pool to be tracked by massively parallel sequencing. We used a library containing 73,000 sgRNAs to generate knockout collections and performed screens in two human cell lines. A screen for resistance to the nucleotide analog 6-thioguanine identified all expected members of the DNA mismatch repair pathway, while another for the DNA topoisomerase II (*TOP2A*) poison etoposide identified *TOP2A*, as expected, and also cyclin-dependent kinase 6, *CDK6.* A negative selection screen for essential genes identified numerous gene sets corresponding to fundamental processes. Finally, we show that sgRNA efficiency is associated with specific sequence motifs, enabling the prediction of more effective sgRNAs. Collectively, these results establish Cas9/sgRNA screens as a powerful tool for systematic genetic analysis in mammalian cells.

A critical need in biology is the ability to efficiently identify the set of genes underlying a cellular process. In microorganisms, powerful methods allow systematic loss-of-function genetic screening (*1*, *2*). In mammalian cells, however, current screening methods fall short – primarily because of the difficulty of inactivating both copies of a gene in a diploid mammalian cell. Insertional mutagenesis screens in cell lines that are near-haploid or carry *Blm* mutations, that cause frequent somatic crossing-over, have proven powerful but are not applicable to most cell lines and suffer from integration biases of the insertion vectors (*3*, *4*). The primary solution has been to target mRNAs with RNA interference (RNAi) (*5–9*). However, this approach is also imperfect as it only partially suppresses target gene levels and can have off-target effects on other mRNAs – resulting in false negative and false positive results (*10–12*). Thus, there remains an unmet need for an efficient, large-scale, loss of function screening method in mammalian cells

Recently, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) pathway, which functions as an adaptive immune system in bacteria (*13*), has been co-opted to engineer mammalian genomes in an efficient manner (*14–16*). In this two-component system, a single guide RNA (sgRNA) directs the Cas9 nuclease to cause double-stranded cleavage of matching target DNA sequences (*17*). In contrast to previous genome-editing techniques, such as zinc-finger nucleases and TALENs, the target specificity of CRISPR/Cas9 is dictated by a 20-base pair sequence at the 5′-end of the sgRNA, allowing for much greater ease of construction of knockout reagents. Mutant cells lines and mice bearing multiple modified alleles can be generated with this technology (*18*, *19*).We set out to explore the feasibility of using the CRISPR/Cas9 system to perform large-scale, loss-of-function screens in mammalian cells. The idea was to use a pool of sgRNA-expressing lentivirus to generate a library of knockout cells that could be screened under both positive and negative selection. Each sgRNA would serve as a distinct DNA barcode that can be used to count the number of cells carrying it using high-throughput sequencing (Fig. 1A). Pooled screening requires that single-copy sgRNA integrants are sufficient to induce efficient cleavage of both copies of a targeted locus. This contrasts with the high expression of sgRNAs achieved by transfection that is typically used to engineer a specific genomic change using the CRISPR/Cas9 system.

We first tested the concept in the near-haploid, human KBM7 CML cell line, by creating a clonal derivative expressing the Cas9 nuclease (with a FLAG-tag at its N terminus) under a doxycycline-inducible promoter (Fig. 1B). Transduction of these cells at low multiplicity of infection (MOI) with a lentivirus expressing a sgRNA targeting the endogenous AAVS1 locus revealed substantial cleavage at the AAVS1 locus 48 hours after infection (Fig. 1C). Moreover, because the sgRNA was stably expressed, genomic cleavage continued to increase over the course of the experiment. Deep sequencing of the locus revealed that repair of Cas9-induced double-strand breaks resulted in small deletions (<20 bp) in the target sequence, with tiny insertions or substitutions (<3 bp) occurring at a lower frequency (Fig. 1D). The vast majority of the lesions, occurring in a protein-coding region, would be predicted to give rise to a non-functional protein product, indicating that CRISPR/Cas9 is an efficient means of generating loss-of-function alleles.

We also analyzed off-target activity of CRISPR/Cas9. Although the specificity of CRISPR/Cas9 has been extensively characterized in transfection-based settings (*20–22*), we wanted to examine its off-target behavior in our system, where Cas9 and a single guide RNA targeting AAVS1 (sgAAVS1) were stably expressed for two weeks. We compared the level of cleavage observed at the target locus (97%) to levels at 13 potential off-target cleavage sites in the genome (defined as sites differing by up to 3 bp from sgAAVS1). Minimal cleavage (<2.5%) was observed at all sites with one exception, which was the only site that had perfect complementarity in the 'seed' region (terminal 8 bp). On average, sgRNAs have ~2.2 such sites in the genome, almost always (as in this case) occurring in non-coding DNA and thus less likely to affect gene function (Note S1).

To test the ability to simultaneously screen tens of thousands of sgRNAs, we designed a sgRNA library with 73,151 members, consisting of multiple sgRNAs targeting 7,114 genes and 100 non-targeting controls (Fig. 1E, table S1, and supplementary materials, materials and methods). sgRNAs were designed against constitutive coding exons near the beginning of each gene and filtered for potential off-target effects

Downloaded from www.sciencemag.org on December 13, 2013

based on sequence similarity to the rest of the human genome (Fig. 1, F and G). The library included 10 sgRNAs for each of 7033 genes and all possible sgRNAs for each of the 84 genes encoding ribosomal proteins (Fig. 1H). To assess the effective representation of our microarray synthesized library, we sequenced sgRNA barcodes from KBM7 cells 24 hours after infection with the entire lentiviral pool and were able to detect the overwhelming majority (>99%) of our sgRNAs, with high uniformity across constructs (only 6-fold increase in abundance between the 10^{th} and 90^{th} percentiles) (fig. S2A).

As an initial test of our approach, we screened the library for genes that function in DNA mismatch repair (MMR). In the presence of the nucleotide analog 6-thioguianine (6-TG), MMR-proficient cells are unable to repair 6-TG-induced lesions and arrest at the G2-M cell-cycle checkpoint, while MMR-defective cells do not recognize the lesions and continue to divide (*23*). We infected Cas9-KBM7 cells with the entire sgRNA library, cultured the cells in a concentration of 6-TG that is lethal to wild-type KBM7 cells, and sequenced the sgRNA barcodes in the final population. sgRNAs targeting the genes encoding the four components of the MMR pathway (*MSH2*, *MSH6*, *MLH1* and *PMS2*) (*24*) were dramatically enriched in the 6-TG-treated cells. At least four independent sgRNAs for each gene showed very strong enrichment and barcodes corresponding to these genes made up >30% of all barcodes (Fig. 2, A and B). Strikingly, each of the twenty most abundant sgRNAs targeted one of these four genes. The fact that few of the other 73,000 sgRNAs scored highly in this assay suggests a low frequency of off-target effects.

We next addressed the challenge of loss of function screening in diploid cells, which require bi-allelic inactivation of a target gene. We therefore generated an inducible Cas9 derivative of the HL60 pseudodiploid human leukemic cell line. In both HL60 and KBM7 cells, we screened for genes whose loss conferred resistance to etoposide, a chemotherapeutic agent that poisons DNA topoisomerase IIA (*TOP2A*). To identify hit genes, we calculated the difference in abundance between the treated and untreated populations for each sgRNA, calculated a score for each gene by using a Kolmogorov-Smirnov test to compare the sgRNAs targeting the gene against the non-targeting control sgRNAs, and corrected for multiple hypothesis testing (Fig. 2, C to E, Table S2). Identical genes were detected in both screens, with significance levels exceeding all other genes by more than 100-fold. As expected, loss of *TOP2A* itself conferred strong protection to etoposide (*25*). The screen also revealed a role for *CDK6*, a G1 cyclin-dependent kinase, in mediating etoposide-induced cytotoxicity. Notably, every one of the 20 sgRNAs in the library targeting *TOP2A* or *CDK6* was strongly enriched ($>90^{th}$ percentile) in both screens, indicating that the effective coverage of our libraries is very high. We generated isogenic HL60 cell lines with individual sgRNAs against *TOP2A* and *CDK6* and, consistent with the screen results, these lines were much more resistant to etoposide than parental or sgAAVS1-modified HL60 cells (Fig. 2, F and G). Thus, our Cas9/sgRNA system enables large-scale positive selection loss-of-function screens.

To identify genes required for cellular proliferation we screened for genes whose loss conferred a selective disadvantage on cells. Such a screen requires accurate identification of sgRNAs that are depleted from the final cell population. Importantly, a sgRNA will show depletion only if cleavage of the target gene occurs in the majority of cells carrying the construct.

As an initial test, we screened KBM7 cells with a small library containing sgRNAs targeting the *BCR* and *ABL1* genes (Table S3). The survival of KBM7 cells depends on the fusion protein produced by the BCR-ABL translocation (*26*). As expected, depletion was seen only for sgRNAs targeting the exons of *BCR* and *ABL*1 that encode the fusion protein, but not for those targeting the other exons of *BCR* and *ABL1* (Fig. 3A).

We then infected Cas9-HL60, Cas9-KBM7, and WT KBM7 cells with the entire 73,000-member sgRNA library and used deep sequencing of the sgRNA barcodes to monitor the change in abundance of each sgRNA between the initial seeding and a final population obtained after twelve cell doublings (fig. S2, A and B).

We began by analyzing ribosomal proteins genes, for which the library contained all possible sgRNAs. We observed strong Cas9-dependent depletion of sgRNAs targeting genes encoding ribosomal proteins, with good concordance between the sets of ribosomal protein genes essential for cell proliferation in the HL60 and KBM7 screens (the median sgRNA fold-change in abundance was used as a measure of gene essentiality) (Fig. 3, B and C). Interestingly, a few ribosomal protein genes were not found to be essential. These were two genes encoded on chromosome Y (*RPS4Y2*, which is testes-specific (*27*), and *RPS4Y1*, which is expressed at low levels compared to its homolog *RPS4X* on chromosome X (*28*)), and 'ribosome-like' proteins, which may be required only in select tissues (*27*) and generally are lowly expressed in KBM7 cells (fig. S3A).

We then turned our attention to other genes within our dataset, for which ten sgRNAs were designed. As for the ribosomal genes, the essentiality scores of these genes were also strongly correlated between the two cells lines (fig. S3B and table S4). For the twenty highest scoring genes, we found independent evidence for essentiality, based primarily on data from large-scale functional studies in model organisms (table S5).

To evaluate the results at a global level, we tested 4722 gene sets to see if they showed strong signatures of essentiality, using Gene Set Enrichment Analysis (*29*). Gene sets related to fundamental biological processes – including DNA replication, gene transcription, and protein degradation – showed strong depletion, consistent with their essentiality (Fig. 3D and table S6).

Finally, we sought to understand the features underlying sgRNA efficacy. Although the vast majority of sgRNAs against ribosomal protein genes showed depletion, detailed comparison of sgRNAs targeting the same gene revealed substantial variation in the precise amounts of depletion. These differences are unlikely to be caused by local accessibility to the Cas9/sgRNA complex inasmuch as comparable variability was observed even among sgRNAs targeting neighboring target sites of a given gene (fig. S4A). Given that our library includes all possible sgRNAs against each of the 84 ribosomal genes, the data allowed us to search for factors that might explain the differential efficacy of sgRNAs. Because the majority of ribosomal proteins genes are essential, we reasoned that the level of depletion of a given ribosomal protein-targeting sgRNA could serve as a proxy for its cleavage efficiency. Applying this approach, we found several trends related to sgRNA efficacy: (i) Single guide sequences with very high or low GC content were less effective against their targets. (ii) sgRNAs targeting the last coding exon were less effective than those targeting earlier exons, consistent with the notion that disruption of the terminal exon would be expected to have less impact on gene function. (iii) sgRNAs targeting the transcribed strand were less effective than those targeting the non-transcribed strand (Fig. 3E). Although these trends were statistically significant, they explained only a small proportion of differences in sgRNA efficacy (table S7).

We hypothesized that differences in sgRNA efficacy might also result from sequence features governing interactions with Cas9. To test this, we developed a method to profile the sgRNAs directly bound to Cas9 in a highly parallel manner (Methods). By comparing the abundance of sgRNAs bound to Cas9 relative to the abundance of their corresponding genomic integrants, we found that the nucleotide composition near the 3′- end of the spacer sequence was the most important determinant of Cas9 loading (Fig. 3F). Specifically, Cas9 preferentially bound sgRNAs containing purines in the last 4 nucleotides of the spacer sequence whereas pyrimidines were disfavored. A similar pattern emerged when we examined depletion of ribosomal protein-targeting sgRNAs (*r*

= 0.81), suggesting that, in significant part, the cleavage efficiency of a sgRNA was determined by its affinity for Cas9 (table S7).

We then sought to build an algorithm to discriminate between strong and weak sgRNAs (Fig. 3G). We trained a support-vector-machine classifier based on the target sequences and depletion scores of ribosomal protein-targeting single guide RNAs. As an independent test, we used the classifier to predict the efficacy of sgRNAs targeting the 400 top scoring (i.e., essential) non-ribosomal genes. The top two-thirds of our predictions exhibited 3-fold higher efficacy than the remaining fraction, confirming the accuracy of the algorithm.

Using this algorithm, we designed a whole-genome sgRNA library consisting of sequences predicted to have higher efficacy (table S8). As with the sgRNA pool used in our screens, this new collection was also filtered for potential off-target matches. This reference set of sgRNAs may be useful both for targeting single genes as well as large-scale sgRNA screening.

Taken together, these results demonstrate the utility of CRISPR/Cas9 for conducting large-scale genetic screens in mammalian cells. Based on our initial experiments, this system appears to offer several powerful features that together provide significant advantages over current functional screening methods.

First, CRISPR/Cas9 inactivates genes at the DNA level, making it possible to study phenotypes that require a complete loss of gene function to be elicited. In addition, the system should also enable functional interrogation of non-transcribed elements, which are inaccessible by RNAi.

Second, a large proportion of sgRNAs successfully generate mutations at their target sites. While this parameter is difficult to directly assess in pooled screens, we can obtain an estimate by examining the 'hit rate' at known genes. Applying a z-score analysis of our positive selection screens, we find that over 75% (46/60) of sgRNAs score at a significance threshold that perfectly separates true and false positives on a gene level (fig. S5, A to D). Together these results show that the effective coverage of our library is very high and that the rate of false negatives should be low even in a large-scale screen.

Third, off-target effects do not appear to seriously hamper our screens, based on several lines of evidence. Direct sequencing of potential off-target loci detected minimal cleavage at secondary sites, which typically reside in non-coding regions and do not impact gene function. Moreover, in the 6-TG screens, the twenty most abundant sgRNAs all targeted one of the four members of the MMR pathway. In total, they represented over 30% of the final pool, a fraction greater the next 400 sgRNAs combined. In the etoposide screen, the two top genes scored far above background levels (*P* values 100-fold smaller than the next best gene), enabling clear discrimination between true and false positive hits. Lastly, new versions of the CRISPR/Cas9 system have recently been developed that substantially decrease off-target activity (*30*, *31*).

Although we limited our investigation to proliferation-based phenotypes, our approach can be applied to a much wider range of biological phenomena. With appropriate sgRNA libraries, the method should enable genetic analyses of mammalian cells to be conducted with a degree of rigor and completeness currently possible only in the study of microorganisms.

References and Notes

1. G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K. D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C. Y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, M. Johnston, Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387–391 (2002). doi:10.1038/nature00935 Medline
2. M. Costanzo, A. Baryshnikova, J. Bellay, Y. Kim, E. D. Spear, C. S. Sevier, H. Ding, J. L. Koh, K. Toufighi, S. Mostafavi, J. Prinz, R. P. St Onge, B. VanderSluis, T. Makhnevych, F. J. Vizeacoumar, S. Alizadeh, S. Bahr, R. L. Brost, Y. Chen, M. Cokol, R. Deshpande, Z. Li, Z. Y. Lin, W. Liang, M. Marback, J. Paw, B. J. San Luis, E. Shuteriqi, A. H. Tong, N. van Dyk, I. M. Wallace, J. A. Whitney, M. T. Weirauch, G. Zhong, H. Zhu, W. A. Houry, M. Brudno, S. Ragibizadeh, B. Papp, C. Pál, F. P. Roth, G. Giaever, C. Nislow, O. G. Troyanskaya, H. Bussey, G. D. Bader, A. C. Gingras, Q. D. Morris, P. M. Kim, C. A. Kaiser, C. L. Myers, B. J. Andrews, C. Boone, The genetic landscape of a cell. *Science* **327**, 425–431 (2010). doi:10.1126/science.1180823 Medline
3. J. E. Carette, C. P. Guimaraes, M. Varadarajan, A. S. Park, I. Wuethrich, A. Godarova, M. Kotecki, B. H. Cochran, E. Spooner, H. L. Ploegh, T. R. Brummelkamp, Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009). doi:10.1126/science.1178955 Medline
4. G. Guo, W. Wang, A. Bradley, Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* **429**, 891–895 (2004). doi:10.1038/nature02653 Medline
5. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998). doi:10.1038/35888 Medline
6. T. R. Brummelkamp, R. Bernards, R. Agami, A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002). doi:10.1126/science.1068999 Medline
7. J. Moffat, D. A. Grueneberg, X. Yang, S. Y. Kim, A. M. Kloepfer, G. Hinkle, B. Piqani, T. M. Eisenhaure, B. Luo, J. K. Grenier, A. E. Carpenter, S. Y. Foo, S. A. Stewart, B. R. Stockwell, N. Hacohen, W. C. Hahn, E. S. Lander, D. M. Sabatini, D. E. Root, A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006). doi:10.1016/j.cell.2006.01.040 Medline
8. V. N. Ngo, R. E. Davis, L. Lamy, X. Yu, H. Zhao, G. Lenz, L. T. Lam, S. Dave, L. Yang, J. Powell, L. M. Staudt, A loss-of-function RNA interference screen for molecular targets in cancer. *Nature* **441**, 106–110 (2006). doi:10.1038/nature04687 Medline
9. H. W. Cheung, G. S. Cowley, B. A. Weir, J. S. Boehm, S. Rusin, J. A. Scott, A. East, L. D. Ali, P. H. Lizotte, T. C. Wong, G. Jiang, J. Hsiao, C. H. Mermel, G. Getz, J. Barretina, S. Gopal, P. Tamayo, J. Gould, A. Tsherniak, N. Stransky, B. Luo, Y. Ren, R. Drapkin, S. N. Bhatia, J. P. Mesirov, L. A. Garraway, M. Meyerson, E. S. Lander, D. E. Root, W. C. Hahn, Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 12372–12377 (2011). doi:10.1073/pnas.1109363108 Medline
10. C. J. Echeverri, P. A. Beachy, B. Baum, M. Boutros, F. Buchholz, S. K. Chanda, J. Downward, J. Ellenberg, A. G. Fraser, N. Hacohen, W. C. Hahn, A. L. Jackson, A. Kiger, P. S. Linsley, L. Lum, Y. Ma, B. Mathey-Prévôt, D. E. Root, D. M. Sabatini, J. Taipale, N. Perrimon, R. Bernards, Minimizing the risk of reporting false positives in large-scale RNAi screens. *Nat. Methods* **3**, 777–779 (2006). doi:10.1038/nmeth1006-777 Medline
11. M. Booker, A. A. Samsonova, Y. Kwon, I. Flockhart, S. E. Mohr, N. Perrimon, False negative rates in Drosophila cell-based RNAi screens: a case study. *BMC Genomics* **12**, 50 (2011). doi:10.1186/1471-2164-12-50 Medline
12. W. G. Kaelin, Jr., Use and Abuse of RNAi to Study Mammalian Gene Function, Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science* **337**, 421–422 (2012). doi:10.1126/science.1225787 Medline

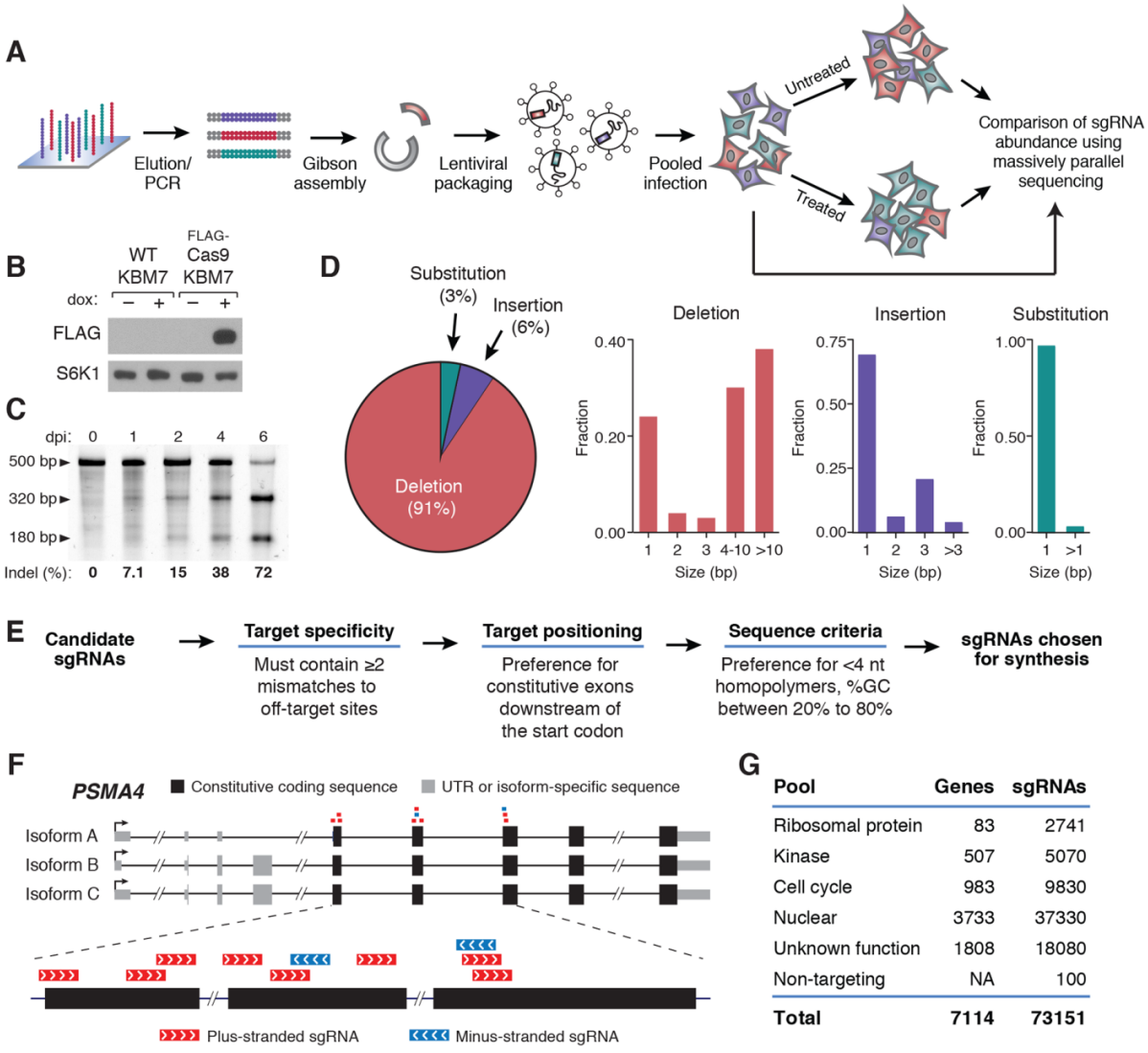
13. R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, P. Horvath, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007). doi:10.1126/science.1138140 Medline
14. L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013). doi:10.1126/science.1231143 Medline
15. P. Mali, L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville, G. M. Church, RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013). doi:10.1126/science.1232033 Medline
16. W. Y. Hwang, Y. Fu, D. Reyon, M. L. Maeder, S. Q. Tsai, J. D. Sander, R. T. Peterson, J. R. Yeh, J. K. Joung, Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013). doi:10.1038/nbt.2501 Medline
17. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012). doi:10.1126/science.1225829 Medline
18. T. Horii, D. Tamura, S. Morita, M. Kimura, I. Hatada, Generation of an ICF syndrome model by efficient genome editing of human induced pluripotent stem cells using the CRISPR system. *Int. J. Mol. Sci.* **14**, 19774–19781 (2013). doi:10.3390/ijms141019774 Medline
19. H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang, R. Jaenisch, One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013). doi:10.1016/j.cell.2013.04.025 Medline
20. P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, F. Zhang, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013). doi:10.1038/nbt.2647 Medline
21. Y. Fu, J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, J. K. Joung, J. D. Sander, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013). doi:10.1038/nbt.2623 Medline
22. V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, D. R. Liu, High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **31**, 839–843 (2013). doi:10.1038/nbt.2673 Medline
23. T. Yan, S. E. Berry, A. B. Desai, T. J. Kinsella, DNA mismatch repair (MMR) mediates 6-thioguanine genotoxicity by introducing single-strand breaks to signal a G2-M arrest in MMR-proficient RKO cells. *Clin. Cancer Res.* **9**, 2327–2334 (2003). Medline
24. R. D. Kolodner, G. T. Marsischky, Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* **9**, 89–96 (1999). doi:10.1016/S0959-437X(99)80013-6 Medline
25. D. J. Burgess, J. Doles, L. Zender, W. Xue, B. Ma, W. R. McCombie, G. J. Hannon, S. W. Lowe, M. T. Hemann, Topoisomerase levels determine chemotherapy response in vitro and in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9053–9058 (2008). doi:10.1073/pnas.0803513105 Medline
26. B. Scappini, S. Gatto, F. Onida, C. Ricci, V. Divoky, W. G. Wierda, M. Andreeff, L. Dong, K. Hayes, S. Verstovsek, H. M. Kantarjian, M. Beran, Changes associated with the development of resistance to imatinib (STI571) in two leukemia cell lines expressing p210 Bcr/Abl protein. *Cancer* **100**, 1459–1471 (2004). doi:10.1002/cncr.20131 Medline
27. S. Xue, M. Barna, Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat. Rev. Mol. Cell Biol.* **13**, 355–369 (2012). doi:10.1038/nrm3359 Medline
28. C. M. Johnston, F. L. Lovell, D. A. Leongamornlert, B. E. Stranger, E. T. Dermitzakis, M. T. Ross, Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genet.* **4**, e9 (2008). doi:10.1371/journal.pgen.0040009 Medline
29. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi:10.1073/pnas.0506580102 Medline
30. F. A. Ran, P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang, F. Zhang, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013). doi:10.1016/j.cell.2013.08.021 Medline
31. P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, G. M. Church, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013). doi:10.1038/nbt.2675 Medline
32. K. Yoshimoto, M. Mizoguchi, N. Hata, H. Murata, R. Hatae, T. Amano, A. Nakamizo, T. Sasaki, Complex DNA repair pathways as possible therapeutic targets to overcome temozolomide resistance in glioblastoma. *Front. Oncol.* **2**, 186 (2012). doi:10.3389/fonc.2012.00186 Medline
33. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi:10.1038/nmeth.1923 Medline
34. J. M. Engreitz, A. Pandya-Jones, P. McDonel, A. Shishkin, K. Sirokman, C. Surka, S. Kadri, J. Xing, A. Goren, E. S. Lander, K. Plath, M. Guttman, The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013). doi:10.1126/science.1237973 Medline
35. S. S. Liu, H. X. Zheng, H. D. Jiang, J. He, Y. Yu, Y. P. Qu, L. Yue, Y. Zhang, Y. Li, Identification and characterization of a novel gene, c1orf109, encoding a CK2 substrate that is involved in cancer cell proliferation. *J. Biomed. Sci.* **19**, 49 (2012). doi:10.1186/1423-0127-19-49 Medline
36. R. Renella, N. A. Roberts, J. M. Brown, M. De Gobbi, L. E. Bird, T. Hassanali, J. A. Sharpe, J. Sloane-Stanley, D. J. Ferguson, J. Cordell, V. J. Buckle, D. R. Higgs, W. G. Wood, Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1α localization in erythroblasts. *Blood* **117**, 6928–6938 (2011). doi:10.1182/blood-2010-09-308478 Medline
37. S. H. Chen, P. S. Wu, C. H. Chou, Y. T. Yan, H. Liu, S. Y. Weng, H. F. Yang-Yen, A knockout mouse approach reveals that TCTP functions as an essential factor for cell proliferation and survival in a tissue- or cell type-specific manner. *Mol. Biol. Cell* **18**, 2525–2532 (2007). doi:10.1091/mbc.E07-02-0188 Medline
38. C. Cayrol, C. Lacroix, C. Mathe, V. Ecochard, M. Ceribelli, E. Loreau, V. Lazar, P. Dessen, R. Mantovani, L. Aguilar, J. P. Girard, The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. *Blood* **109**, 584–594 (2007). doi:10.1182/blood-2006-03-012013 Medline
39. B. Sönnichsen, L. B. Koski, A. Walsh, P. Marschall, B. Neumann, M. Brehm, A. M. Alleaume, J. Artelt, P. Bettencourt, E. Cassin, M. Hewitson, C. Holz, M. Khan, S. Lazik, C. Martin, B. Nitzsche, M. Ruer, J. Stamford, M. Winzi, R. Heinkel, M. Röder, J. Finell, H. Häntsch, S. J. Jones, M. Jones, F. Piano, K. C. Gunsalus, K. Oegema, P. Gönczy, A. Coulson, A. A. Hyman, C. J. Echeverri, Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* **434**, 462–469 (2005). doi:10.1038/nature03353 Medline
40. J.-F. Rual, J. Ceron, J. Koreth, T. Hao, A. S. Nicot, T. Hirozane-Kishikawa, J. Vandenhaute, S. H. Orkin, D. E. Hill, S. van den Heuvel, M. Vidal, Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res.* **14**, (10B), 2162–2168 (2004). doi:10.1101/gr.2505604 Medline
41. J. L. Mummery-Widmer, M. Yamazaki, T. Stoeger, M. Novatchkova, S. Bhalerao, D. Chen, G. Dietzl, B. J. Dickson, J. A. Knoblich, Genome-wide analysis of Notch signalling in Drosophila by transgenic RNAi. *Nature* **458**, 987–992 (2009). doi:10.1038/nature07936 Medline
42. A. C. Spradling, D. Stern, A. Beaton, E. J. Rhem, T. Laverty, N. Mozden, S. Misra, G. M. Rubin, The Berkeley Drosophila Genome

Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135–177 (1999). Medline
43. A. Amsterdam, R. M. Nissen, Z. Sun, E. C. Swindell, S. Farrington, N. Hopkins, Identification of 315 genes essential for early zebrafish development. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12792–12797 (2004). doi:10.1073/pnas.0403929101 Medline

Acknowledgments: We thank all members of the Sabatini and Lander labs especially J. Engreitz, S. Schwartz, A. Shishkin, and Z. Tsun for protocols, reagents and advice; T. Mikkelsen for assistance with oligonucleotide synthesis; L. Gaffney for assistance with figures. This work was supported by the US National Institutes of Health (CA103866) (D.M.S.), National Human Genome Research Institute (2U54HG003067-10) (E.S.L.), the Broad Institute of MIT and Harvard (E.S.L.) and an award from the US National Science Foundation (T.W.). The composition of the sgRNA pools and screening data can be found in the Supplementary Materials. T.W., D.M.S., and E.S.L. are inventors on a patent application from the Broad Institute for functional genomics using CRISPR-Cas systems. Inducible Cas9 and sgRNA backbone lentiviral vectors and the genome-scale sgRNA plasmid pool are deposited in Addgene.

Supplementary Materials
www.sciencemag.org/content/science.1246981/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S8
References (*33*–*43*)

8 October 2013; accepted 2 December 2013
Published online 12 December 2013
10.1126/science.1246981



Pool	Genes	sgRNAs
Ribosomal protein	83	2741
Kinase	507	5070
Cell cycle	983	9830
Nuclear	3733	37330
Unknown function	1808	18080
Non-targeting	NA	100
Total	**7114**	**73151**

Fig. 1. A pooled approach for genetic screening in mammalian cells using a lentiviral CRISPR/Cas9 system (**A**) Outline of sgRNA library construction and genetic screening strategy (**B**) Immunoblot analysis of wild-type KBM7 cells and KBM7 cells transduced with a doxycycline inducible FLAG-Cas9 construct upon doxycycline induction. S6K1 was used as a loading control. (**C**) Sufficiency of single copy sgRNAs to induce genomic cleavage. Cas9-expressing KBM7 cells were transduced with AAVS1-targeting sgRNA lentivirus at low MOI. The SURVEYOR mutation detection assay was performed on cells at the indicated days post-infection (dpi). Briefly, mutations resulting from cleavage of the AAVS1 locus were detected through PCR amplification of a 500-bp amplicon flanking the target sequence, re-annealing of the PCR product and selective digestion of mismatched heteroduplex fragments. (**D**) Characterization of mutations induced by CRISPR/Cas9 as analyzed by high-throughput sequencing. (**E**) sgRNA library design pipeline. (**F**) Example of sgRNAs designed for *PSMA4*. sgRNAs targeting constitutive exonic coding sequences nearest to the start codon were chosen for construction. (**G**) Composition of genome-scale sgRNA library.

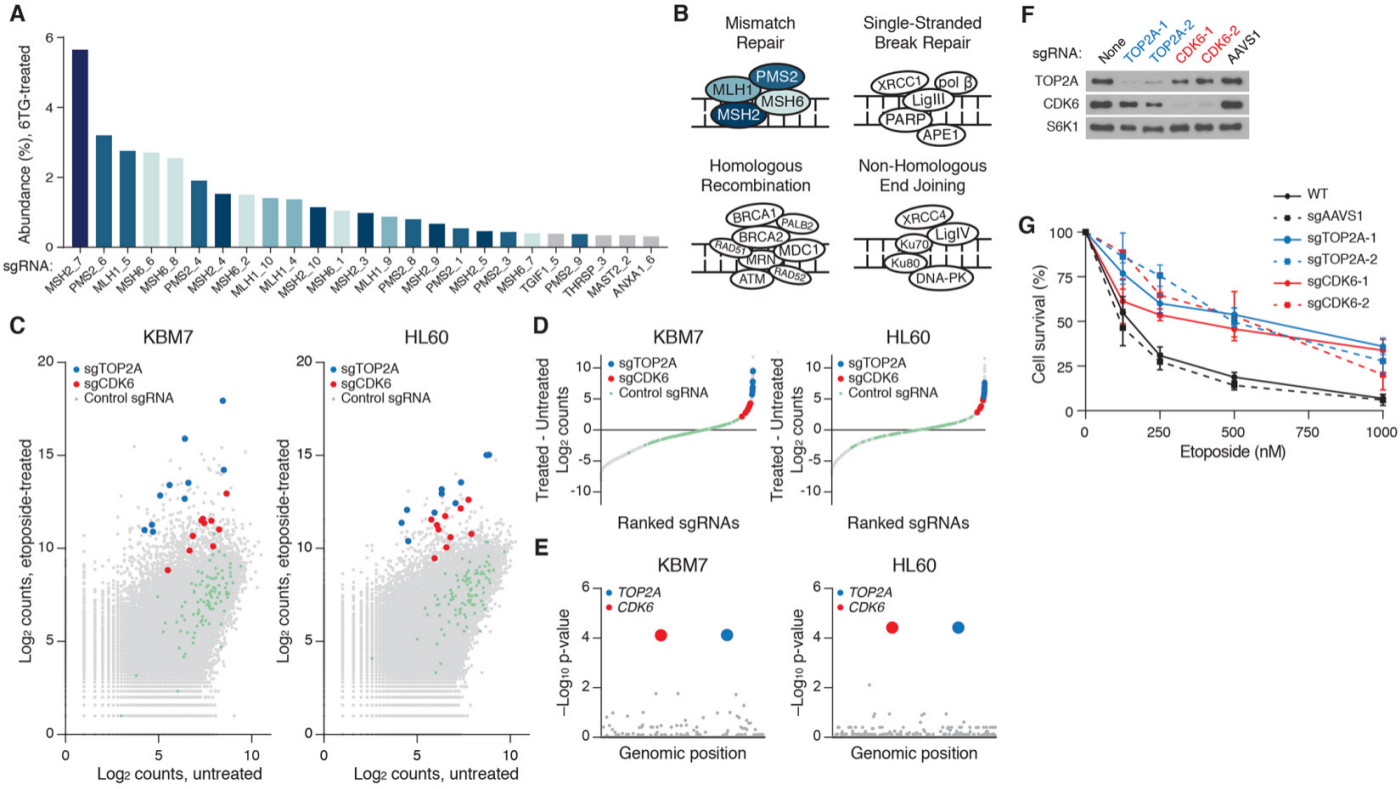


Fig. 2. Resistance screens using CRISPR/Cas9. (**A**) Raw abundance (%) of sgRNA barcodes after 12 days of selection with 6-thioguanine (6-TG). (**B**) Mismatch repair (MMR) deficiency confers resistance to 6-TG. Diagram depicts cellular DNA repair processes. Only sgRNAs targeting components of the DNA MMR pathway were enriched. Diagram modified and adapted from (*32*). (**C**) Primary etoposide screening data. The count for a sgRNA is defined as the number of reads that perfectly match the sgRNA target sequence. (**D**) sgRNAs from both screens were ranked by their differential abundance between the treated versus untreated populations. For clarity, sgRNAs with no change in abundance are omitted. (**E**) Gene hit identification by comparing differential abundances of all sgRNAs targeting a gene to differential abundances of non-targeting sgRNAs in a one-sided Kolmogorov-Smirnov test. p-values are corrected for multiple hypothesis testing. (**F**) Immunoblot analysis of WT and sgRNA-modified HL60 cells 1 week after infection. S6K1 was used as a loading control. (**G**) Viability, as measured by cellular ATP concentration, of WT and sgRNA-modified HL60 cells at indicated etoposide concentrations. Error bars denote standard deviation (n = 3).

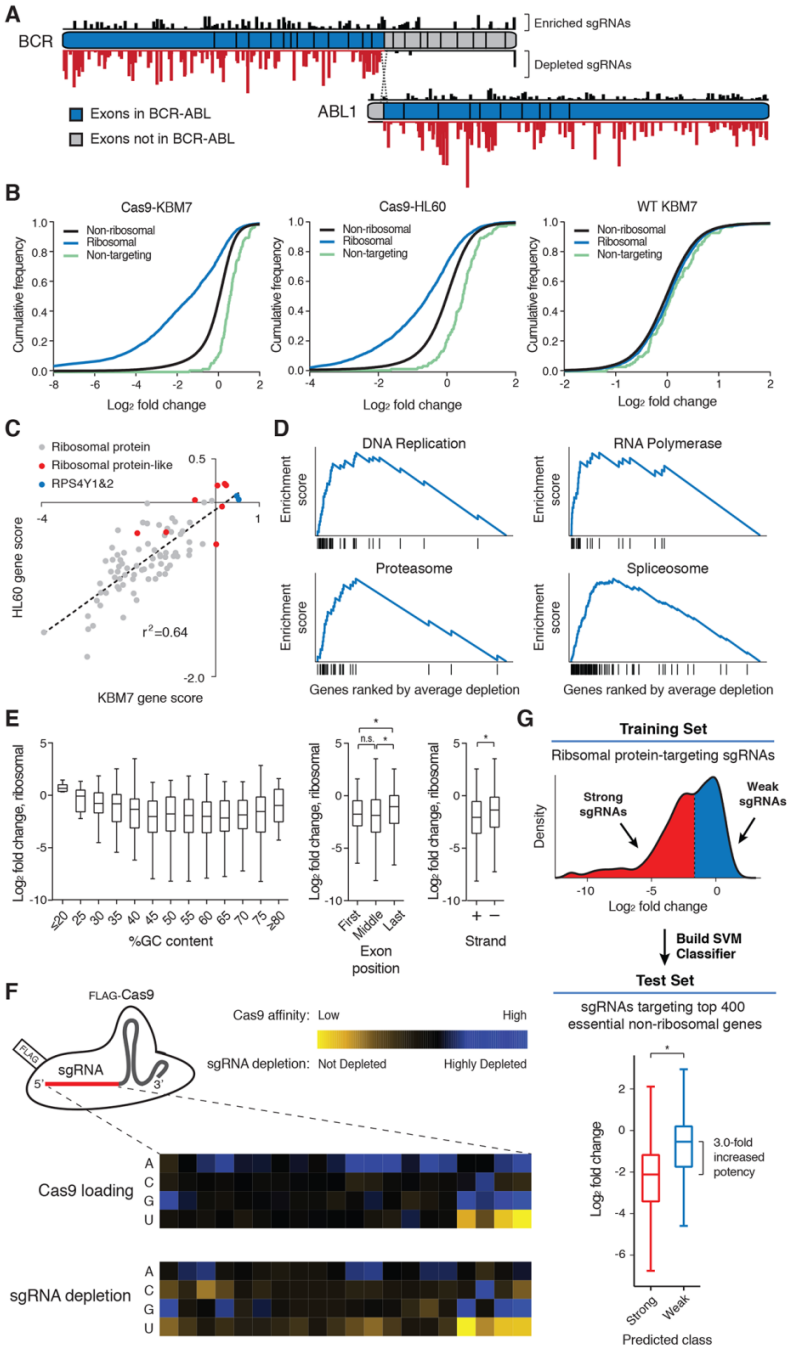


Fig. 3. Negative selection screens using CRISPR/Cas9 reveal rules governing sgRNA efficacy. (**A**) Selective depletion of sgRNAs targeting exons of *BCR* and *ABL1* present in the fusion protein. Individual sgRNAs are plotted according to their target sequence position along each gene and the height of each bar indicates the level of depletion observed. Boxes indicate individual exons. (**B**) Cas9-dependent depletion of sgRNAs targeting ribosomal proteins. Cumulative distribution function plots of $\log_2$ fold changes in sgRNA abundance before and after twelve cell doublings in Cas9-KBM7, Cas9-HL60 and WT-KBM7 cells. (**C**) Requirement of similar sets of ribosomal protein genes for proliferation in the HL60 and KBM7 cells. Gene scores are defined as the median $\log_2$ fold change of all sgRNAs targeting a gene. (**D**) Depleted sgRNAs target genes involved in fundamental biological processes. Gene Set Enrichment Analysis was performed on genes ranked by their combined depletion scores from screens in HL60 and KBM7 cells. Vertical lines underneath the x-axis denote members of the gene set analyzed. (**E**) Features influencing sgRNA efficacy. Depletion ($\log_2$ fold change) of sgRNAs targeting ribosomal protein genes was used as an indicator of sgRNA efficacy. Correlation between $\log_2$ fold changes and spacer %GC content (left), exon position targeted (middle) and strand targeted (right) are depicted. (*P < 0.05) (**F**) sgRNA target sequence preferences for Cas9 loading and cleavage efficiency. Position-specific nucleotide preferences for Cas9 loading are determined by counting sgRNAs bound to Cas9 normalized to the number of corresponding genomic integrations. Heatmaps depict sequence-dependent variation in Cas9 loading (top) and ribosomal protein gene-targeting sgRNA depletion (bottom). The color scale represents the median value (of Cas9 affinity or $\log_2$ fold-change) for all sgRNAs with the specified nucleotide at the specified position. (**G**) sgRNA efficacy prediction. Ribosomal protein gene-targeting sgRNAs were designated as 'weak' or 'strong' based their $\log_2$ fold change and used to train a support-vector-machine (SVM) classifier. As an independent test, the SVM was used to predict the efficacy of sgRNAs targeting 400 essential non-ribosomal genes. (*P < 0.05)