

grID: A CRISPR-Cas9 guide RNA Database and Resource for Genome-Editing

Vinod Jaskula-Ranga^{1*} and Donald J. Zack^{1, 2, 3, 4}

¹Department of Ophthalmology, Wilmer Eye Institute, The Johns Hopkins University School of Medicine, USA

²Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, USA

³Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University School of Medicine, USA

⁴Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, USA

*To whom correspondence should be addressed. E-mail: vinny1@jhmi.edu and dzack@jhmi.edu

ABSTRACT

CRISPR-Cas9 genome-editing is a revolutionary technology that is transforming biological research. The explosive growth and advances in CRISPR research over the last few years, coupled with the potential for clinical applications and therapeutics, is heralding a new era for genome engineering. To further support this technology platform and to provide a universal CRISPR annotation system, we introduce the grID database (<http://crispr.technology>), an extensive compilation of gRNA properties including sequence and variations, thermodynamic parameters, off-target analyses, and alternative PAM sites, among others. To aid in the design of optimal gRNAs, the website is integrated with other prominent databases, providing a wealth of additional resources to guide users from *in silico* analysis through experimental CRISPR targeting. Here, we make available all the tools, protocols, and plasmids that are needed for successful CRISPR-based genome targeting.

INTRODUCTION

CRISPR type II systems are naturally occurring prokaryotic immune defense mechanisms that have been repurposed for genome engineering (1-3), regulation (4-6), and other sequence-specific genomic manipulations (7). CRISPR targeting occurs through a ribonucleoprotein complex comprised of the Cas9 nuclease and the guide RNA (gRNA), a synthetic fusion between the bacterial crRNA and tracrRNA (8). The target site, formed from a spacer region followed by a Protospacer Adjacent Motif (PAM), is recognized by both components: Cas9 binds the short nucleotide PAM sequence, while the gRNA serves as the primary determinant in targeting through complementary base-pairing with the spacer region. The Cas9/gRNA complex from *S. pyogenes*, the most commonly used CRISPR system, recognizes a 20-nucleotide spacer region and the NGG PAM sequence, and therefore, the CRISPR targeting site can be represented as a contiguous sequence in the form N₂₀NGG (9).

A key feature of the CRISPR RNA-directed DNA cleavage system is that Cas9 target recognition is programmable through ribonucleotide sequence changes in the guide RNA (8). While the N₂₀NGG sequence forms the minimal recognition site, further constraints are generally imposed due to the pol III polymerase used to synthesize the gRNA (3,10), or other sequence-dependent factors such as T_m , GC content (11), or polymeric runs. Another important factor is the presence of highly similar sequences in the genome, which can lead to off-target mutagenesis. This effect, further compounded by the number and position of mismatches, can significantly affect CRISPR targeting. Thus, judicious gRNA selection remains the principal factor in successful targeting.

To aid in the design and selection of optimal gRNAs, we have created grID (gRNA Identification), a CRISPR-Cas9 guide RNA database and resource for genome editing (<http://crispr.technology/>). Additionally, we have integrated within grID a mechanism for members of the research community to contribute data on their success, or lack thereof, with particular gRNAs, general comments, and suggestions for database improvements. We hope that grID will serve as an evolving database of useful resources that will aid the genome editing research community by adding to the power and variety of already existing CRISPR/Cas9 design tools (see <http://omictools.com/crispr-cas9-c1268-p1.html> for a partial compilation of available resources).

MATERIALS AND METHODS

Identification of CRISPR sites

To determine the potential set of targeting sites in a genome for a given Cas9 ortholog, engineered variant, or alternative CRISPR system, a custom regular expression Perl script was used to search for the occurrences of that sequence, allowing presence in either orientation and for overlapping sites. For example, the 23-nucleotide sequence N₂₀NGG, was searched for SpCas9. Orthologous Cas9 searches were performed using N₂₃GRRT for SaCas9(12), N₁₈NNAGAAW for StCas9 (9), and N₂₄NNNNGATT for NmCas9 (13). To search for engineered Cas9 variants with altered PAM specificity, searches were performed using N₂₀NGCG for SpCas9 VRER, N₂₀NGAG for SpCas9 EQR, and N₂₀NGAN for SpCas9 VQR (14). The SpCas9 D1135E variant, and high-specificity/high-fidelity variants, eSpCas9 and SpCas9-HF1, are a more specific SpCas9 targeting Cas9 proteins(15,16). N₂₀NAG, considered an off-target for SpCas9, was also searched. For the alternative Cpf1 CRISPR system (AsCpf1 and LbCpf1), TTTNN₂₃ was searched (17). The reference genome sequences were constructed by concatenation into a single chromosome file obtained from the repeat-masked assembly sequence; sequences masked by N's were purposely excluded by the Perl regular expression, and thus excluded from the initial set of CRISPR sites. The reference genomes used were: hg19 for Human (NCBI build GRCh37), mm10 for Mouse (NCBI build GRCm38); rn5 for Rat (RGSC Rnor_5.0), danRer7 for Zebrafish (Sanger Institute Zv9), ce10 for *C. elegans* (WormBase v. WS220), and sacCer3 for *S. cerevisiae* (SGD). Subsequent text manipulations were performed using standard Unix, AWK, and/or Perl scripting. grID numbering is assigned incrementally starting with the first chromosome. The total set of CRISPR sites is available for download as either FASTA formatted or BED formatted files (<http://crispr.technology/data/>).

Thermodynamic Properties

ΔH° and ΔS° parameters for nearest neighbor DNA/RNA hybrid formation (18) were used to calculate the T_m from the free energy derived equation: $T_m = \frac{\Delta H^\circ}{\Delta S^\circ + RT \ln C_t}$. RNA secondary structure analysis of the spacer was performed using RNAFold (19) after converting the protospacer sequence to ribonucleotides, the dot-bracket notation and ΔG values were databased.

Off-target analysis

Potential genomic off-target sites were predicted using Bowtie (20), allowing up to three mismatches along the target site (-v = 3), while suppressing the output for any sites reporting more than twenty-five alignments in the genome (-m = 25).

Paired Nickase/Offset nicking

The offset is defined as the distance between the internal ends of the CRISPR targeting site, while the 5' overhang length is calculated as the distance between the two nick sites (21).

Data sources

Annotations for genomic, transcript, and protein sequences, as well as gene descriptions, were obtained from the NCBI RefSeq Database (<http://www.ncbi.nlm.nih.gov/refseq/>) (22), and through the UCSC genome browser (23). Sequences and annotations are locally databased and accessed using BioPerl (24). Restriction enzyme analysis is performed through Perl scripting using the target sites from REBASE (<http://rebase.neb.com/rebase/rebase.html>) (25). OMIM and disease data was obtained from OMIM (Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD, USA), 2013. World Wide Web URL: <http://omim.org/>) (26). dbSNP human Build 142 based on GRCh37/hg19 was used as the data source for SNPs and variation (27).

UCSC genome visualization

The grID data in the BED file format was converted to compressed binary indexed BigBed file using the bedToBigBed Kent utility (28). Visualization through the UCSC genome browser (<http://genome.ucsc.edu>).

Construction and implementation of the grID database

The grID data is stored in a MySQL relational database installed on a Unix-Apache web server. Server-side scripting is performed using Perl, while dynamic and client-side scripting is performed using JavaScript and jQuery. The front-end HTML, CSS, and JavaScript interface was developed on the Twitter Bootstrap framework and is accessible through desktop, tablet, and mobile devices.

Plasmid construction

Plasmids encoding the H1 promoter::AvrII::gRNA scaffold or U6 promoter::AvrII::gRNA scaffold were constructed by gBlock synthesis (IDT) and Gibson assembly (NEB) (29) into pUC19 digested with *NdeI* and *XbaI* restriction enzymes. For *in vitro* gRNA expression, T7, T3, or SP6 bacteriophage promoters were similarly constructed by Gibson assembly using gBlock synthesis cloned into pUC19 (*NdeI/XbaI*). The promoter::target::gRNA scaffold unit is flanked by an upstream 5' *ClaI* site and a 3' *NotI* site; the latter is used to linearize the vector generating a 5' overhang.

RESULTS AND DISCUSSION

grID system

In order to provide a universal CRISPR naming system, we introduce grID (**gRNA Identification**), a nomenclature system based on the systematic identification of CRISPR-Cas9 sites. CRISPR recognition sites in the genome are uniquely designated by a two-letter prefix to indicate the species (e.g. hs for *Homo sapiens*) followed by a sequentially assigned numerical digit (e.g. [hs025713228](#)). The grID system provides a simple and unambiguous framework for the identification of CRISPR sites in any genome. Furthermore, the identification system is flexible enough to accommodate PAM sequences from other existing orthologs (1,12,13), engineered or evolved Cas9 variants (14-16), and alternative CRISPR systems (17) (**Supplemental Table 1**).

We began by determining the potential CRISPR sites in the human genome for the *S. pyogenes* Cas9 by

searching for all occurrences of the sequence N₂₀NGG. Since for most genome-engineering applications targeting within repetitive regions is undesirable, we suppressed such loci by using the repeat-masked sequence for the initial genome reference. In total, there are 137,409,562 potential CRISPR sites that fit these criteria within the human genome ([hs000000001](#) - [hs137409562](#)), and their location and sequence are easily accessible through the grID database.

Off-target analysis is a major component of the database, comprising the computationally predicted off-targets for grID entries. The off-target parameters are pre-calculated and databased to avoid the computational time that users would otherwise expend while performing analysis; however, to balance virtually instantaneous data retrieval with database size, we limited the pre-calculated analysis to 0, 1, 2, or 3 mismatches in CRISPR sites totaling less than 25 predicted off-target loci. Sites with 25 or greater off-target loci are still indicated; however, users are generally advised against using such sites. For example, in the human genome we first performed a genome-wide alignment for the 137,409,562 sequences in the grID database allowing for up to 3 mismatches along the entire targeting sequence, which resulted in 10,172,371,379 potential loci. By limiting the results to sequences with 25 or less loci, the total number was reduced to 1,575,631,967 off-target sites. We find that this provides ample data for the vast majority of researchers, as users are unlikely to perform sequencing analysis and verification for a large number of loci. Additionally, those who are highly concerned with off-targets are likely to use more discriminating variants like the D1135E SpCas9 variant (14), or the high-specificity/high-fidelity variants (eSpCas9 and SpCas9-HF1) that exhibit low-to-undetectable off-target mutagenesis (15,16).

By querying the database against a small set of experimentally determined CRISPR sites (30), we found that the database performs accurately on all sites (**Supplemental Table 1**). Of the 6 total sites that were examined, 4 were reported to elicit a high-level of off-target mutagenesis; the two sites that were not reported to have off-target hits ([hs008507971](#) and [hs019839404](#)) also do not produce any user warnings from the database. 2 of the remaining 4 sites (VEGFA T2 and VEGFA T3) are suppressed because they intersect with repeat-masked regions, and desirably, these potentially confounding sites are absent from user queries. The remaining 2 sites produce user warnings for potential off-targets due to the presence of polymeric runs, dinucleotide repeats, and high GC content ([hs106331938](#)), or due to the presence of a trinucleotide repeat ([hs064758550](#)).

As dissociation kinetics governs Cas9 cleavage (31), we reasoned that relative target site uniqueness toward the high-affinity PAM proximal region could be used to modulate on- and off-rates. The grID Specificity is based on an approach we developed to classify gRNAs according to their relative uniqueness in the genome through iterative 5' truncations and genomic alignments. For example, with SpCas9 the target sequences (N₂₀NGG) were first aligned to the genome to identify unique sites. Next, the 5' nucleotide was removed and the sequence (now N₁₉NGG) was realigned. Those that failed to align uniquely were scored and removed from the next cycle. This process identifies more unique CRISPR targeting sites in the genome with a bias towards the seed region, which we score as “low,” “medium,” or “high” specificity. We also include other on-target scoring algorithms pre-computed for each gRNA, where applicable (32,33).

DATABASE

The grID database provides several query options. In addition to searching by grID (e.g. [hs025713228](#)), users can search by NCBI reference sequence (RefSeq) ID (e.g. [NM_002046](#)), Gene Symbol (e.g. [GAPDH](#)), or any valid 23-bp targeting sequence in the form N₂₀NGG (e.g. [GCATCTTCTTTTGCCTCGCCAGG](#)). User queries from Gene Symbols with multiple splice variants are directed to an isoform page that provides links to a unique RefSeq ID.

grID Page

Every grID page provides an extensive set of properties that are useful in gRNA selection and as reference information:

- (i) A Summary section provides an overview that contains the CRISPR target sequence, grID specificity (low, medium, high), and Cas9 species of origin (e.g. *S. pyogenes*) (**Figure 1A**).

- (ii) The Genomic Context indicates the exact genomic coordinate (chr:start-end) and strand (+/-) orientation for the target site. Additionally, the location is mapped against RefSeq coordinates, providing a descriptive location (exonic, intronic, or intergenic) (**Figure 1B**). The nearest genes subsection indicates the genes closest to the target site.
- (iii) The Browser section provides a simple graphical and interactive visualization tool. The embedded Jbrowse (34) frame found on each page automatically loads to the specific region of interest (**Figure 1C**).
- (iv) The Sequence Analysis section details the transcribed RNA sequence (gRNA spacer sequence), uniqueness status of the target (protospacer) sequence (unique or non-unique), the PAM motif, and the presence of either homopolymeric runs or pol III terminator sequences (yes or no) (**Figure 1D**).
- (v) A Thermodynamic Properties section includes GC-content, melting temperature (T_m), and RNAfold prediction and stability. These properties are qualified (e.g. “very low,” “low,” “moderate,” “high,” and “very high”) to provide the users with some guidance for potentially problematic targeting sequences (**Figure 1E**).
- (vi) Users are provided with a general summary of the mismatches and numbers of off-target sites, along with detailed alignments, and mismatch location: distal (position 20-16), distal seed (position 15-11), proximal seed (position 10-1), and PAM N mismatches (**Figure 1F**).
- (vii) A Paired Nickase/offset nicking section is available for users who prefer to use the Cas9 (D10A) nickase (5,21). The potential gRNA mates compatible with the Cas9 (D10A) mutant, along with the respective offset distance and 5' overhang length, are ranked by increasing distance.
- (viii) A Sequence section provides users with quick access to the flanking genomic sequence. 500bp of the reference sequence upstream and downstream of the target site is shown (**Supplemental Figure 1A**).
- (ix) The Screening Analysis section, which indicates the presence of restriction sites that fall within the targeting region, is available to aid in downstream CRISPR experimental analysis. Restriction sites, particularly those that overlap the cutsite, can be used as a convenient alternative to mismatch detection and quantification by the standard Surveyor or T7 endo I assay (35,36) (**Supplemental Figure 1B**).
- (x) The SNPs and Variations section provides a quick overview of dbSNP142 data within the target site that could lead to potential sequence differences (27). Minor sequence variations can explain recalcitrant targeting or be exploited for allele-specific targeting (**Supplemental Figure 1C**).
- (xi) The Notes section provides user alerts about gRNAs that are potentially problematic. Warnings for properties that can contribute towards a higher propensity of off-target mutagenesis are noted, such as non-unique targeting sequences, homopolymeric runs, high T_m , or high GC content. Additionally, potential downstream pitfalls like strong secondary structure within the spacer or the presence of a pol III terminator are also indicated (**Supplemental Figure 1D**).
- (xii) A Reference section is provided on each grID Page, allowing for the cataloging of citations publishing a target sequence. A deposit form will be available on the website, encouraging authors or general users to contribute references to the database. This in turn provides a valuable resource for the community, as well as a potential mechanism for author recognition through citations.

- (xiii) Finally, in an effort to promote scientific collaboration and to disseminate valuable information to the CRISPR community, we encourage use of a Comments section that is provided at the bottom of each grID page. Comments made on a specific site are only available through that grID Page, ensuring that only pertinent information is displayed to the user.

Other Resources, Plasmids and Protocols

gRNA selection is only one step in the process of genome editing. Since an important aspect of the grID database/website is to create an encompassing CRISPR ecosystem available through a single centralized location, and to facilitate CRISPR research by bridging *in silico* analysis with laboratory resources, we provide additional resources to help potential users.

Similar to the grID page, the website provides useful information from the context of the gene: (i) The Summary section provides a brief gene description and summary from RefSeq (22), as well as quick links to other databases; (ii) a Disease Information section with information from the OMIM database is provided (26); (iii) a Browser section that indicates a similar graphical and interactive visualization tool as the grID page; (iv) a Genomic Context section for genome coordinates; (v) Sequence information to provide quick access to genomic and protein sequences; (vi) an interactive search and filter to identify high scoring gRNAs targeting specific exons; (vii) Notes, which provide useful additional gRNA-specific information, (viii) References, and (ix) a Comments section analogous to that described for the grID page.

The advanced search page (<http://crispr.technology/search.html>) allows users to target specific regions in the genome that would be useful for a wide range of CRISPR experiments. For example, users can target exons, introns, 5' or 3' UTRs and promoter regions, and application-specific searches, like CRISPRi (37,38), are available as well. For users who prefer data visualization through a genome browser, grID data can be viewed through a full-page JBrowse genome browser (<http://crispr.technology/browser/>)(34), or via the UCSC genome browser (<http://crispr.technology/genomes.html>)(23). Both options allow users to interact with their own data. The database and resources are primarily geared towards human and mouse targeting, however other model genomes such as rat, zebrafish, *C. elegans*, and *S. cerevisiae* are also available.

Though our resource page users have access to extensively detailed protocols that we have tested and optimized for generating specific CRISPR targeting constructs. We generated several plasmids using common *in vivo* promoters (H1 and U6) and *in vitro* promoters (T7, T3 and SP6) to facilitate gRNA transcription, and all of these are freely available through Addgene (www.addgene.org) (**Supplemental Sequences**) (10). An oligonucleotide builder (http://crispr.technology/resources/oligo_form.html) is available to generate primers for cloning any gRNA in the database into various expression vectors. Construct generation for the user is simplified through the use of a common restriction enzyme (*AvrII*), cloning method (Gibson Assembly) (29), antibiotic selection marker (ampicillin), and sequencing primers (M13R). Optimized experimental protocols including for cell line transfection (<http://crispr.technology/resources/transfection.html>) and quantification of genome-editing efficiency (<http://crispr.technology/resources/quantification.html>), or tools useful for common genome-editing calculations (<http://crispr.technology/resources/calculations.html>) are also available.

Our goal is to maintain grID as a dynamic database serving the genome editing community, and efforts will be made to expand and modify the database as needed to keep pace with the rapid evolution of CRISPR-Cas9 technology.

ACKNOWLEDGEMENTS

FUNDING

This work was supported by a Knights Templar Career Starter Research Grant, grants from the National Institutes of Health (P30EY001765, EY009769), unrestricted funds from Research to Prevent Blindness, Inc., funds from Foundation Fighting Blindness, and generous gifts from the Guerrieri Family Foundation

and from Mr. and Mrs. Robert and Clarice Smith. Some of the computation work was performed on the Maryland Data Intensive Academic Grid (DIAG), which is supported by the National Science Foundation funded MRI-R2 project #DBI-0959894. Funding for open access charge: National Eye Institute, National Institutes of Health.

REFERENCES

1. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
2. Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) RNA-programmed genome editing in human cells. *eLife*, **2**, e00471.
3. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
4. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442-451.
5. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.*
6. Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, **517**, 583-588.
7. Mali, P., Esvelt, K.M. and Church, G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nature methods*, **10**, 957-963.
8. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
9. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733-740.
10. Ranganathan, V., Wahlin, K., Maruotti, J. and Zack, D.J. (2014) Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nature communications*, **5**, 4516.
11. Wang, T., Wei, J.J., Sabatini, D.M. and Lander, E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80-84.
12. Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S. *et al.* (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*.
13. Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.F., Sontheimer, E.J. and Thomson, J.A. (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc Natl Acad Sci U S A*.
14. Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R. *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481-485.
15. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490-495.
16. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X. and Zhang, F. (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84-88.
17. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759-771.
18. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M. and Sasaki, M. (1995) Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211-11216.
19. Lorenz, R., Bernhart, S.H., Honer Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB*, **6**, 26.
20. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, **10**, R25.

21. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell*.
22. Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic acids research*, **42**, D756-763.
23. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M. *et al.* (2015) The UCSC Genome Browser database: 2015 update. *Nucleic acids research*, **43**, D670-681.
24. Stajich, J.E., Block, D., Boulez, K., Brenner, S.E., Chervitz, S.A., Dagdigan, C., Fuellen, G., Gilbert, J.G., Korf, I., Lapp, H. *et al.* (2002) The Bioperl toolkit: Perl modules for the life sciences. *Genome research*, **12**, 1611-1618.
25. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2015) REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic acids research*, **43**, D298-299.
26. Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. and Hamosh, A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic acids research*, **43**, D789-798.
27. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, **29**, 308-311.
28. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. and Karolchik, D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204-2207.
29. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, **6**, 343-345.
30. Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*.
31. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, **507**, 62-67.
32. Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J. and Root, D.E. (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, **32**, 1262-1267.
33. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, **34**, 184-191.
34. Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. and Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome research*, **19**, 1630-1638.
35. Qiu, P., Shandilya, H., D'Alessio, J.M., O'Connor, K., Durocher, J. and Gerard, G.F. (2004) Mutation detection using Surveyor nuclease. *BioTechniques*, **36**, 702-707.
36. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome research*, **19**, 1279-1288.
37. Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S. and Qi, L.S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature protocols*, **8**, 2180-2196.
38. Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y. *et al.* (2016) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science*.

TABLE AND FIGURES LEGENDS

Figure 1. Overview of the grID page. A summary of the CRISPR target sequence and on-target score from various algorithms (**A**), genomic context of the target site (**B**), an embedded genome browser (**C**), sequence analysis (**D**), thermodynamic properties (**E**), and off-target analysis (**F**).

A Summary

grID:	hs025713230
Sequence (23-mer):	GATGCGGCTGACTGTGCAACAGG
grID Score:	high (700)
Doench (2014) Score:	0.09586
Doench (2016) Score:	0.46257
Cas9:	<i>S. pyogenes</i>

B Genomic Context

Genomic Location:	chr12:6643718-6643696
Strand:	-
gRNA location:	exon:refGene(NM_002046.exon0)

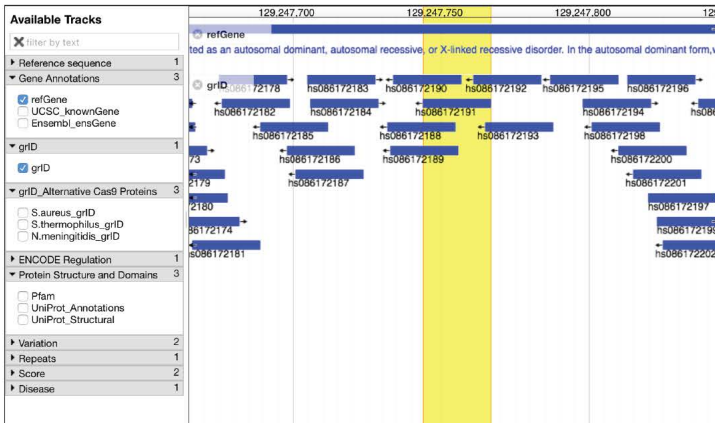
C Nearest Genes

refseq	gene	chrom	start	end	distance
NM_002046	GAPDH	chr12	6643584	6647537	0
NM_001256799	GAPDH	chr12	6644408	6647537	712
NM_014865	NCAPD2	chr12	6603297	6641132	2586
NM_001039670	IFFO1	chr12	6648893	6665249	4997
NM_001193457	IFFO1	chr12	6648893	6665249	4997

D DNase I hypersensitive sites from ENCODE

DNase I hypersensitivity region:	chr12:6643560-6643710
Numbers of Cell Types:	123
Cell Types:	8988t, A549, AG04449, AG04450, AG09309, AG09319, AG10803, Adult_CD4_Th0, AoAF, AoSMC, BE2_C, BJ, Caco-2, CD14+, CD20+, CD34+_Mobilized, CMK, Chorion, Cili, HSMM_emb, Fibrobl, Fibrop, GM06990, GM12864, GM12865, GM12878, Gliobla, Gm12891, Gm12892, Gm18507, Gm19238, Gm19239, Gm19240, H9es, HAEPiC, HAc, HA-h, HBMEC, HCF, HCFaa, HCM, HCFEpiC, HCT-116, HConF, HEEpiC, H1-hESC, HFF, HFF-Myc, HGF, HIPEpiC, HL-60, HMEC, HMF, HMVEC-LBI, HMVEC-Lly, HMVEC-dAd, HMVEC-dBi-Ad, HMVEC-dBi-Neo, HMVEC-dLy-Ad, HMVEC-dLy-Neo, HMVEC-dNeo, HNPCEpiC, HPAEC, HPAF, HPF, HPdLF, HRCEpiC, HRE, HRGEC, HRPEpiC, HSMM, HSMM_FSHD, HUVEC, HVMF, HeLa-S3, HeLa-S3s3fna4h, HepG2, Hepatocytes, HPDE6-E6E7, HTR8svn, Huh-7, Huh-7.5, Ips, Ishikawa, Ishikawa, Jurkat, K562, LNCap, LNCaP, MCF-7, MCF-7, Medullo, Melano, Myometr, NB4, NH-A, NHDF-Ad, NHDF-neo, NHEK, NHLF, NT2-D1, Osteobl, PanisletD, Panislets, Phte, PREC, Proglif, RPTEC, RWPE1, SAEC, SK-N-SH, SK-N-MC, SkMC, Stellate, T-47D, Urothelia, UrotheliaU1189, WERI-Rb-1, WI-38, WI-38_TAM, H1-hESC, Th1, Th2,

C Browser View



D Sequence Analysis

Target Sequence (RNA):	GAUGCGGCGUGACUGUGCAAC
Target Sequence Occurrence:	unique
PAM motif:	AGG
Homopolymeric Runs:	none
Pol III Terminator:	none

E Thermodynamic Properties

GC-content (%):	60
GC-score:	moderate
Tm (°C):	58.63
Tm score:	moderate
RNAfold (dot bracket):(((.....)))....
deltaG (Kcal/mol):	-4.20
Predicted Folding:	moderate secondary structure

F Offtarget Analysis

Overview Summary Details

Total Predicted Genomic Targets:	13
Total Genomic Sites With 0 Mismatches:	1
Total Genomic Sites With 1 Mismatches:	1
Total Genomic Sites With 2 Mismatches:	
Total Genomic Sites With 3 Mismatches:	11

Overview Summary Details

1	chr12:6643695-6643718	0	CCTGTTGACAGTCAGCCGCATC * * * * CCTGTTGACAGTCAGCCGCATC
2	chrX:39647430-39647453	1	GATGCGGCTGACTGTGCAACAGG * * * * GATGCGGCTGACTGTGCAACAGG
3	chr13:29880856-29880879	3	CCTGTTGACAGTCAGCCGCATC * * * * CCTGTTGACAGTCAGCCGCATG
4	chr15:64821905-64821928	3	GATGCGGCTGACTGTGCAACAGG * * * * * TATGTGCTGTCTGTGCAACAGG
5	chr6:80663763-80663786	3	GATGCGGCTGACTGTGCAACAGG * * * * GATGCACTGTCTGTCAAACAGG
6	chr5:159378540-159378563	3	GATGCGGCTGACTGTGCAACAGG * * * * GATGCACTGTCTGTGCAACGGG

5

Number of Mismatches:3

Location: chr6:80663763-80663786
Genomic Sequence: GATGCACTGTCTGTCAAACAGG
Strand: +
Genomic Region: **intergenic**
3' PAM Orientation: GATGCACTGTCTGTCAAACAGG

Alignment:
GATGCGGCTGACTGTGCAACAGG
|||||*|||||*|||||*|||||*
GATGCACTGTCTGTCAAACAGG

Mismatch:
.....A.....
.....T.....
.....A.....
Distal mismatch
Proximal seed mismatch
Proximal seed mismatch

6

Number of Mismatches:3

Location: chr5:159378540-159378563
Genomic Sequence: GATGCACTGTCTGTGCAACGGG
Strand: +
Genomic Region: **gene(ADRA1B)**
3' PAM Orientation: GATGCACTGTCTGTGCAACGGG

Alignment:
GATGCGGCTGACTGTGCAACAGG
|||||*|||||*|||||*|||||*
GATGCACTGTCTGTGCAACGGG

Mismatch:
.....A.....
.....T.....
.....G.....
Distal mismatch
Proximal seed mismatch
PAM N Offtarget