



Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation

Max A Horlbeck^{1,2,3,4}, Luke A Gilbert^{1,2,3,4}, Jacqueline E Villalta^{1,2,3,4†}, Britt Adamson^{1,2,3,4}, Ryan A Pak^{1,5}, Yuwen Chen^{1,2,3,4}, Alexander P Fields^{1,2,3,4}, Chong Yon Park^{1,5}, Jacob E Corn^{5,6}, Martin Kampmann^{1,2,3,4,7}, Jonathan S Weissman^{1,2,3,4*}

¹Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, United States; ²Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, United States; ³California Institute for Quantitative Biomedical Research, University of California, San Francisco, San Francisco, United States; ⁴Center for RNA Systems Biology, University of California, San Francisco, San Francisco, United States; ⁵Innovative Genomics Initiative, University of California, Berkeley, Berkeley, United States; ⁶Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, United States; ⁷Institute for Neurodegenerative Diseases, University of California, San Francisco, San Francisco, United States

Abstract We recently found that nucleosomes directly block access of CRISPR/Cas9 to DNA (*Horlbeck et al., 2016*). Here, we build on this observation with a comprehensive algorithm that incorporates chromatin, position, and sequence features to accurately predict highly effective single guide RNAs (sgRNAs) for targeting nuclease-dead Cas9-mediated transcriptional repression (CRISPRi) and activation (CRISPRa). We use this algorithm to design next-generation genome-scale CRISPRi and CRISPRa libraries targeting human and mouse genomes. A CRISPRi screen for essential genes in K562 cells demonstrates that the large majority of sgRNAs are highly active. We also find CRISPRi does not exhibit any detectable non-specific toxicity recently observed with CRISPR nuclease approaches. Precision-recall analysis shows that we detect over 90% of essential genes with minimal false positives using a compact 5 sgRNA/gene library. Our results establish CRISPRi and CRISPRa as premier tools for loss- or gain-of-function studies and provide a general strategy for identifying Cas9 target sites.

DOI: [10.7554/eLife.19760.001](https://doi.org/10.7554/eLife.19760.001)

*For correspondence: Jonathan Weissman@ucsf.edu

Present address: [†]Calico Life Sciences LLC, South San Francisco, United States

Competing interest: See [page 16](#)

Funding: See [page 16](#)

Received: 19 July 2016

Accepted: 22 September 2016

Published: 23 September 2016

Reviewing editor: Karen Adelman, Harvard Medical School, United States

© Copyright Horlbeck et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

Highly multiplexed pooled genetic screening methodologies have emerged as powerful and broadly accessible tools for systematically profiling gene function at the scale of mammalian genomes (*Paddison et al., 2004*). Recently, a number of pooled screening platforms have been developed that utilize the bacterial CRISPR (Clustered Regularly Interspaced Palindromic Repeats)-associated nuclease Cas9 paired with libraries of single-guide RNAs (sgRNAs) to disrupt targeted genes (reviewed in *Shalem et al., 2015*). We and others have developed tools based on nuclease-dead Cas9 (dCas9) (*Qi et al., 2013*) to programmably interfere with (CRISPRi) or activate (CRISPRa) transcription (*Gilbert et al., 2014, 2013*; *Konermann et al., 2015*; *Maeder et al., 2013*; *Perez-Pinera et al., 2013*; *Tanenbaum et al., 2014*), and used these to systematically manipulate gene

expression at genome scale (*Gilbert et al., 2014; Konermann et al., 2015*). Together, these screening platforms represent a powerful toolkit for unbiased forward gain-of-function and loss-of-function genetic screens in mammalian cells.

A key step in implementing CRISPR genetic screens is selecting sgRNAs that mediate high Cas9 activity. We and others recently found that nucleosomes provide a direct and profound impediment to Cas9 access to DNA (*Hinz et al., 2015; Horlbeck et al., 2016; Isaac et al., 2016*), an observation we expected to be particularly important for applications such as CRISPRi and CRISPRa, which require sustained binding of dCas9 to DNA. We found that nucleosome occupancy was predictive of Cas9 activity complementary to and independent of previously described sgRNA sequence features (*Chari et al., 2015; Doench et al., 2014; Xu et al., 2015*), adding an additional dimension to the set of parameters expected to influence Cas9 activity. These observations, along with the strong nucleosome-dependent phasing observed downstream of the FANTOM consortium-annotated transcription start site (TSS) (*FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014; Horlbeck et al., 2016*), suggested that a quantitative model incorporating all of these features could greatly enhance our ability to predict highly active sgRNAs for CRISPRi and CRISPRa.

To test this, we developed a comprehensive machine learning pipeline trained on data collected from 30 CRISPRi and 9 CRISPRa screens. We found that the resulting models were highly predictive of sgRNA efficacy and strongly weighted nucleosome positioning and specific sequence features. We used these models to design and generate CRISPRi and CRISPRa version 2 (v2) libraries, targeting human and mouse genomes, which are greatly enriched for sgRNAs with high predicted activity. These libraries include several additional improvements, including the option to screen with either 10 sgRNAs per gene or a compact half-library containing the top 5 predicted sgRNAs for each gene. To benchmark this new algorithm, we validated the human CRISPRi v2 (hCRISPRi-v2) library with a screen designed to identify genes essential for robust cell growth. In this experiment, essential genes represent a large class of expected positive controls. We identified over 2100 essential genes with high statistical confidence, significantly improving upon our CRISPRi v1 library (*Gilbert et al., 2014*), and precision-recall analysis showed increased discrimination of gold standard essential and non-essential genes with both 10 sgRNA/gene and 5 sgRNA/gene hCRISPRi-v2 libraries (*Hart et al., 2014*). A large majority of the hCRISPRi-v2 sgRNAs targeting known essential genes produced robust growth phenotypes, a key advance over previous CRISPRi libraries (*Evers et al., 2016*), and our algorithm can accurately predict sgRNA activity for data from screens performed with an independently designed library and in different cell types. Furthermore, we observed that CRISPRi lacked any detectable non-specific toxicity associated with genomic DNA breaks and repair, enabling sensitive detection of genes with subtle growth phenotypes. We also conducted a screen for genes that modify growth rates upon overexpression with our hCRISPRa-v2 and found this identified 60% more genes, with greater enrichment for previously established classes of hit genes, than our version 1 CRISPRa screen. Our results suggest that the CRISPRi and CRISPRa v2 libraries have numerous favorable properties relative to alternate approaches as a resource for targeted or genome-scale loss-of-function and gain-of-function studies in mammalian cells.

Results

An integrated machine learning approach predicts highly active sgRNAs for CRISPRi and CRISPRa

We sought to improve upon our first generation CRISPRi and CRISPRa libraries by taking a comprehensive approach that incorporated nucleosome positioning, sequence features, refinement of our original sgRNA design rules, and other potentially informative factors. In order to quantitatively model the contribution of these features to CRISPRi activity, we turned to our recently described CRISPRi activity dataset (*Horlbeck et al., 2016*) in which we integrated data from 30 CRISPRi screens to select 1,539 high-confidence hit genes, and normalized the phenotypes for sgRNAs targeting each gene to the strongest sgRNAs for that gene, resulting in 'activity scores' for 18,380 sgRNAs. We used this set as training data for elastic net linear regression (*Figure 1A*) (*Hui Zou, 2005*). As many of the features included in the model were nonlinear with activity, we first adapted each feature set according to its relationship with activity. Categorical and non-linear parameters were binned prior to linear regression. Because the relationship between CRISPRi activity and target

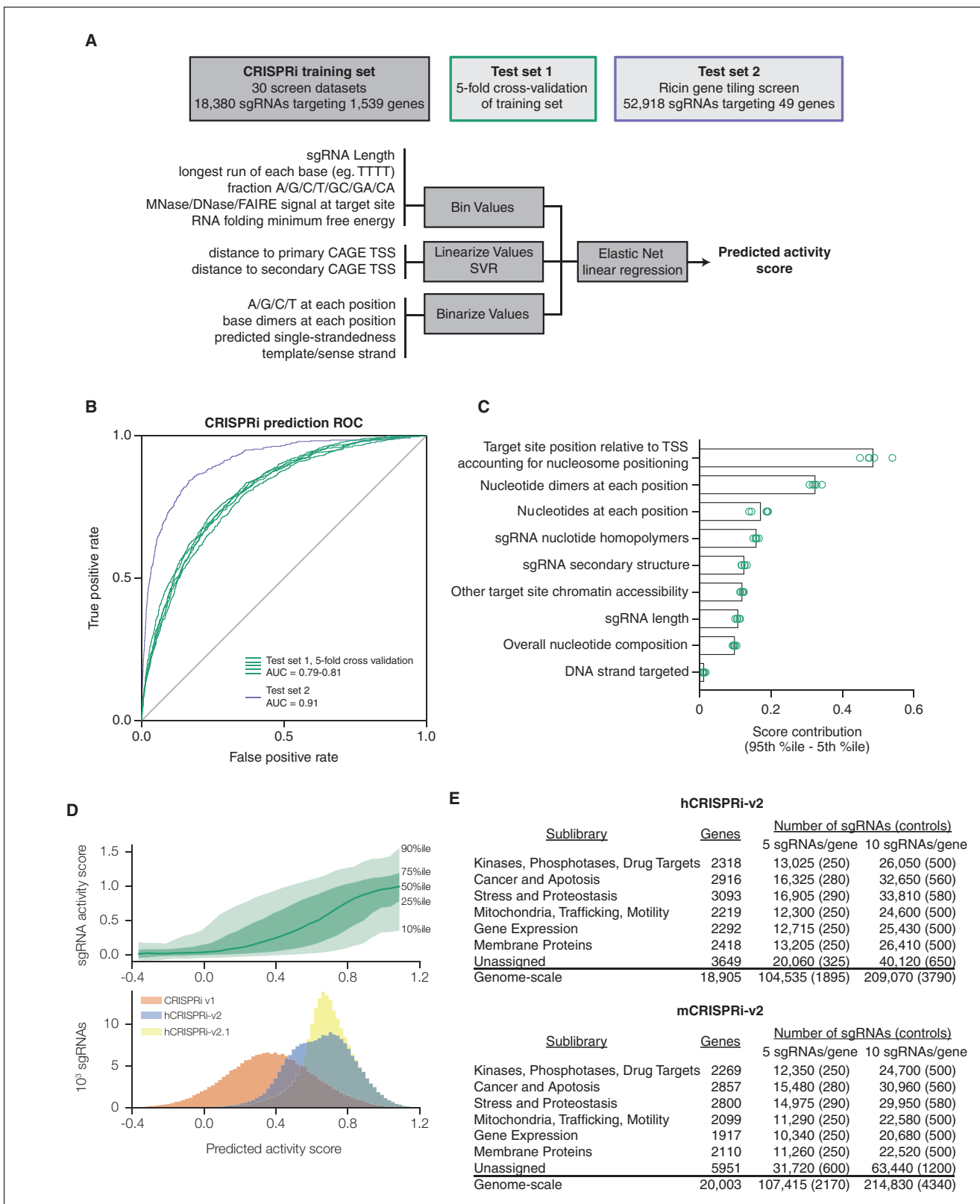


Figure 1. A machine learning approach for identifying highly active sgRNAs for CRISPRi. (A) Schematic of machine learning strategy and datasets. 808 features were calculated for each sgRNA, linearized as indicated, and z-standardized. A linear regression model was then generated using these features to fit to the activity scores of the CRISPRi training set (Horlbeck et al., 2016). 20% of the genes in the training set were reserved to test the predictive value of the resulting model. For display in Figures 1B,C, and Figure 1—figure supplement 2, five non-overlapping 20% datasets were used. Figure 1 continued on next page

Figure 1 continued

randomly selected and training was performed on the corresponding 80% sets. An orthogonal dataset, based on tiling of every possible sgRNA within 10 kb of the TSS of 49 genes known to modulate sensitivity to ricin (Bassik et al., 2013; Gilbert et al., 2014), was also used to assess the predictive value of this model. (B) ROC analysis of the ability of the machine learning approach in (A) to predict highly active sgRNAs. For test set 1, sgRNAs with an activity score greater than 0.75 were considered highly active. For test set 2, sgRNAs with a phenotype greater than 0.75 of the maximum phenotype for each gene were considered highly active. (C) Relative contribution of feature categories to the sgRNA predicted scores. The individual weighting of each feature assigned by the linear regression model (see Figure 2—figure supplement 2) was grouped by the indicated categories, and the summed weights for each sgRNA within the 20% test datasets was calculated. The scores of the 95th and 5th percentile sgRNAs were subtracted to compute the overall contribution of the feature category to the distribution of predicted activity scores. Bars indicate the mean of the contributions from five 20% datasets (green circles). The target site position includes both the distance to the TSS and the periodic relationship as fit by SVR (Figure 1—figure supplement 1). (D) Distribution of predicted activity scores in next-generation CRISPRi libraries. (Top) Predicted CRISPRi activity correlates with empirical activity scores. For the 80%/20% division used to predict sgRNAs for the hCRISPRi-v2.1 library, predicted scores for the 20% test set were plotted against the empirical activity score. Activity score percentiles are from all sgRNAs within 0.25 of the indicated activity score. Predicted activity was highly correlated with activity, with a Pearson R of 0.56 ($p < 10^{-296}$). (Bottom) Distribution of predicted activity scores for CRISPRi v1, hCRISPRi-v2, and hCRISPRi-v2.1, as calculated by the hCRISPRi-v2.1 regression model. (E) Composition of hCRISPRi-v2 and mCRISPRi-v2 sublibraries.

DOI: 10.7554/eLife.19760.002

The following figure supplements are available for figure 1:

Figure supplement 1. Relationship between CRISPRi activity and sgRNA position relative to the TSS as predicted by SVR.

DOI: 10.7554/eLife.19760.003

Figure supplement 2. Individual sgRNA feature contributions to predicted CRISPRi activity.

DOI: 10.7554/eLife.19760.004

site distance from the TSS was highly periodic and asymmetric, as we had recently shown (Horlbeck et al., 2016), we fit sgRNA positioning features using support vector regression (SVR) to predict a continuous function for any target site (Supplementary Figure 1). An important improvement was the use of FANTOM consortium annotations instead of Ensembl/Gencode to define the TSS (Cunningham et al., 2015; FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014; Harrow et al., 2012) (Supplementary file 2), a finding also recently reported by Radzishheuskaya and colleagues (Radzishheuskaya et al., 2016).

We first evaluated the performance of this algorithm using five-fold cross-validation. By performing regression on a training set of only 80% of the genes in the CRISPRi activity dataset, we found that the model was highly predictive of activity in the test set comprising the remaining 20% of the genes, with a receiver operating characteristic area under the curve (ROC-AUC) of 0.80 (Figure 1B). Importantly, this high predictive value was consistent across randomly selected training and test sets. As the sgRNAs in this dataset were pre-selected using our CRISPRi v1 design rules, we also tested our ability to predict the performance of sgRNAs in our previously published data set in which we tiled every target site around the TSS of 49 genes known to modulate resistance or sensitivity to the toxin ricin (Gilbert et al., 2014) and obtained an ROC-AUC of 0.91 (Figure 1B), indicating that we could identify active sgRNAs in the genome with high accuracy.

We next analyzed which features contributed most to CRISPRi activity in the predictive model (Figure 1—figure supplement 2). Overall, the predicted scores were most influenced by the position relative to the TSS, including both distance from the TSS and avoidance of canonical nucleosome-occupied regions (Figure 1C and Figure 1—figure supplement 1). In particular, the nucleosome-deprived region immediately downstream of the TSS yielded the strongest predicted activity by SVR, likely due in part to the contribution of dCas9 directly interfering with early transcriptional initiation or elongation (Gilbert et al., 2013; Qi et al., 2013). Sequence features also represented a large contribution to the model, and salient relationships included the disfavoring of guanine directly downstream of the protospacer adjacent motif (PAM) which recapitulated previous findings (Doench et al., 2014; Xu et al., 2015). Additional parameters that contributed to the prediction, included sgRNA secondary structure as predicted by ViennaRNA (Doench et al., 2016; Lorenz et al., 2011), sgRNA protospacer length, and chromatin accessibility features not accounted for by the nucleosome positioning relationship. The contribution of each individual parameter was also remarkably consistent across 80%/20% divisions of the training dataset, suggesting that the model was capturing underlying biological signal rather than overfitting the data.

Having established the robustness and accuracy of this approach, we used a version of our sgRNA predictions to design a CRISPRi genome-scale library targeting the human protein-coding

transcriptome (hCRISPRi-v2) (**Supplementary file 3**; an *in silico* library design based on the final version of our predictions, hCRISPRi-v2.1, is also available in **Supplementary file 3**). While the predicted scores for sgRNAs in our v1 library were broadly distributed, many sgRNAs of higher predicted activity were available in the genome, and by picking the top 10 predicted sgRNAs per gene we expected that we could greatly enrich the library for guides of high activity (**Figure 1D**). In constructing this library, we also incorporated empirical information for highly active sgRNAs where available, revised off-target filtering, and implemented changes to the sgRNA expression vector to facilitate the processing of screen samples for sequencing (see Materials and methods). In addition, we cloned the library as separate pools for the top 5 and next-best 5 predicted sgRNAs per gene to facilitate screens where a smaller library may be advantageous, and further divided the pools into 7 thematic sublibraries based on our previous divisions of shRNA libraries (**Kampmann et al., 2015**) (**Figure 1E**).

We then used the same approach to design a next-generation CRISPRa library. Due to the requirement for CRISPRa targeting to be upstream of the TSS for maximal activity (**Gilbert et al., 2014**), we collected an activity score dataset of 2,898 sgRNAs from 9 CRISPRa screens (Y.C. and M. K., personal communication) to train an independent predictive model (**Figure 2A**). While the input set was significantly smaller than for CRISPRi, the resulting linear regression still had good predictive value (ROC-AUC 0.70; **Figure 2B**) and generally shared features of the CRISPRi model (**Figure 2C** and **Figure 2—figure supplement 1–2**). We observed that periodic relationship between distance to the TSS and sgRNA activity was less pronounced for CRISPRa than for CRISPRi (**Figure 2—figure supplement 1**; **Figure 1—figure supplement 1**), a difference we attributed to the reduced dynamic range in the nucleosome-depleted region around the TSS, and the smaller number and relatively lower expression of the genes targeted in the CRISPRa training dataset. We used the top 10 predicted sgRNAs for each gene to construct a next-generation library, which significantly increased the predicted activity of the library over our v1 designs (**Figure 2D**). The hCRISPRa-v2 library was partitioned into 14 sublibraries as described for the v2 CRISPRi library above (**Figure 2E**). Importantly, while several strategies have been described for CRISPR-mediated activation (**Chavez et al., 2015**; **Gilbert et al., 2013**; **Hilton et al., 2015**; **Konermann et al., 2015**; **Maeder et al., 2013**; **Perez-Pinera et al., 2013**; **Tanenbaum et al., 2014**; **Zalatan et al., 2015**), a recent comparison of these strategies observed that sgRNA activity generally correlated across all approaches (**Chavez et al., 2016**). Our CRISPRa-v2 libraries are thus likely to serve as a valuable resource for effectively targeting most activator systems.

Finally, we applied the CRISPRi and CRISPRa models to predict highly active sgRNAs targeting the mouse protein-coding transcriptome and generated corresponding genome-scale libraries (mCRISPRi-v2 and mCRISPRa-v2) (**Figures 1E, 2E**). All four library designs are included as **Supplementary file 3–6**, sgRNA prediction and library design scripts are available online (see Materials and methods), and the cloned lentiviral libraries are available on Addgene.

The large majority of sgRNAs in the hCRISPRi-v2 library are effective

A central goal in developing the hCRISPRi-v2 library was to enrich for highly active sgRNAs, which would improve statistical confidence in hits and enable high sensitivity even with a compact 5 sgRNA/gene library. Underscoring the importance of this goal, Evers et al. generated a small-scale CRISPRi library targeting predetermined sets of essential and non-essential genes and conducted screens for growth phenotypes in the bladder carcinoma cell line RT-112, and found that ~50% of the CRISPRi sgRNAs targeting essential genes were inactive (**Evers et al., 2016**), similar to rates observed using our CRISPRi v1 library. To test whether our next-generation sgRNA prediction algorithms were able to identify these inactive sgRNAs, we evaluated the predicted activity of the sgRNAs targeting known essential genes in this screen. Despite the screen being performed with an independently designed library, using an sgRNA constant region we have found to underperform our current design (**Chen et al., 2013**), and in a cell type not evaluated in any of our training or test datasets, predicted activity correlated well with the sgRNA growth phenotypes observed in the screen (**Figure 3A**; Pearson $R = -0.58$, $p < 10^{-37}$). Over 40% of the sgRNAs in this library had a predicted activity score less than 0.4, a regime in which the vast majority (over 87%) of sgRNAs were inactive, as defined by z-score > -2 relative to non-essential gene-targeting sgRNAs, enabling a *priori* elimination of 60% of the inactive sgRNAs by simply applying this threshold. By contrast, of the sgRNAs in that library with predicted activity scores ≥ 0.6 , 77% are active. Use of the improved

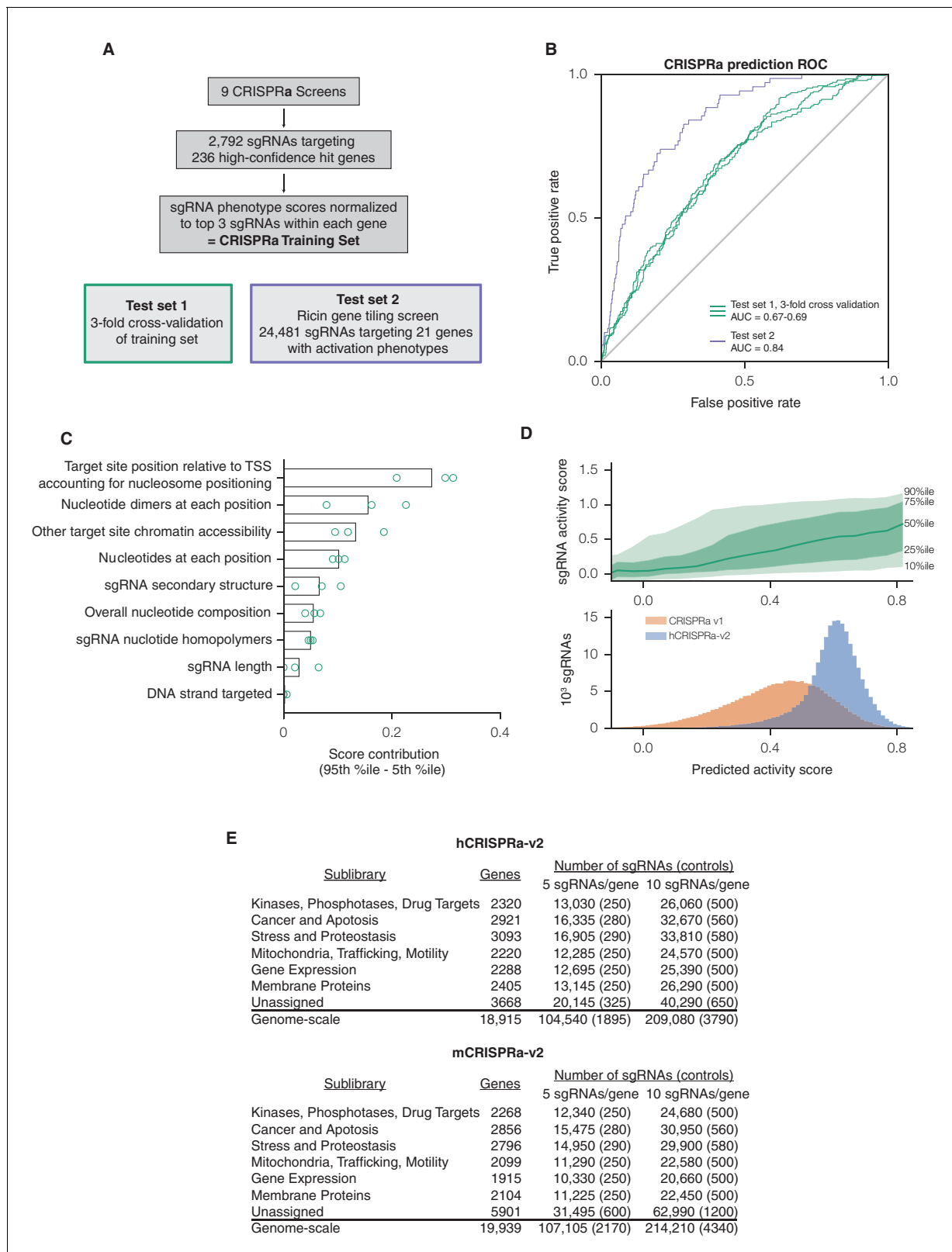


Figure 2. A machine learning approach for identifying highly active sgRNAs for CRISPRa. (A) Schematic of CRISPRa datasets. CRISPRa activity scores were generated from screen data and subjected to 3-fold cross-validation due to the smaller sample size. Ricin tiling data was limited to 21 genes that were previously shown to modulate sensitivity to ricin upon CRISPRa overexpression (Gilbert et al., 2014). (B) ROC analysis of machine learning approach using CRISPRa datasets, conducted as in Figure 1B. (C) Relative contribution of feature categories for CRISPRa, calculated as in Figure 1C. Figure 2 continued on next page

Figure 2 continued

(D) Distribution of predicted activity scores in next-generation CRISPRa libraries. (Top) Predicted CRISPRa activity correlates with empirical activity scores. For the 67%/33% division used to predict sgRNAs for the hCRISPRa-v2 library, predicted scores for the 33% test set were plotted against the empirical activity score. Activity score percentiles are from all sgRNAs within 0.25 of the indicated activity score. Predicted activity was highly correlated with activity, with a Pearson R of 0.41 ($p < 10^{-38}$). (Bottom) Distribution of predicted activity scores for CRISPRa v1 and hCRISPRa-v2 as calculated by the hCRISPRa-v2 regression model. (E) Composition of hCRISPRa-v2 and mCRISPRa-v2 sublibraries.

DOI: [10.7554/eLife.19760.005](https://doi.org/10.7554/eLife.19760.005)

The following figure supplements are available for figure 2:

Figure supplement 1. Relationship between CRISPRa activity and sgRNA position relative to the TSS as predicted by SVR.

DOI: [10.7554/eLife.19760.006](https://doi.org/10.7554/eLife.19760.006)

Figure supplement 2. Individual sgRNA feature contributions to predicted CRISPRa activity.

DOI: [10.7554/eLife.19760.007](https://doi.org/10.7554/eLife.19760.007)

sgRNA constant region would be expected to further increase this fraction (Chen et al., 2013; Dang et al., 2015).

In order to validate our hCRISPRi-v2 library design and directly compare its performance to our published v1 library screens (Gilbert et al., 2014), we conducted a screen for genes essential for robust growth in the chronic myeloid leukemia cell line K562. We calculated the growth phenotype (γ) for each sgRNA (Bassik et al., 2013; Kampmann et al., 2013) and averaged these values across two screen replicates (Figure 3—figure supplement 1 and Supplementary file 7). We found that the hCRISPRi-v2 sgRNA growth phenotypes targeting the Evers et al. essential gene sets correlated with predicted activity as above (Figure 3A; Pearson R = -0.42 , $p < 10^{-21}$), and therefore designing the libraries based on the top predicted scores selected for highly active sgRNAs. To quantify the fraction of active sgRNAs in our genome-scale libraries, we performed sgRNA-level ROC analysis, ranking sgRNAs by growth phenotype γ and classifying them as true or false positives if they targeted essential or non-essential genes, respectively (Evers et al., 2016). This analysis showed that hCRISPRi-v2 was greatly enriched for active sgRNAs (Figure 3B), and in particular the top 5 sgRNA/gene library contained 80% active sgRNAs at 5% false positives, comparable to the pilot nuclease library tested by Evers et al. This improvement was due to the significant reduction in the number of inactive sgRNAs rather than any difference in the noise as assessed by the background distribution of non-essential gene-targeting sgRNAs (Figure 3—figure supplement 1B). In some instances, as with the known essential gene VCP, the difference in sgRNA phenotypes between v1 and v2 libraries was likely attributable to the transition from Ensembl to FANTOM as the TSS annotation source (Figure 3—figure supplement 1C) (Cunningham et al., 2015; FANTOM Consortium and the RIKEN PMI and CLST (DGT) et al., 2014; Harrow et al., 2012). Taken together, the above observations indicate the lower fraction of effective sgRNAs in previous libraries was a result of the algorithm and the TSS annotation used rather than any intrinsic limitation of CRISPRi.

The hCRISPRi-v2 library robustly identifies essential genes in precision-recall analysis

We next sought to evaluate whether the enrichment for active sgRNAs in hCRISPRi-v2 over CRISPRi v1 resulted in improved accuracy and confidence for calling hit genes. We analyzed the screens with a consistent pipeline, scoring genes both by assigning a phenotype based on the mean of the top 3 sgRNAs targeting the gene (by absolute value) and by calculating the Mann-Whitney p-value of all 10 sgRNAs compared to the negative control sgRNAs. We visualized these gene scores as a volcano plot, with the phenotype effect size on the x-axis and p-value on the y-axis (Figure 3C and Supplementary file 8). Many hit genes exhibited much stronger p-values, including a substantial fraction that reached the optimal p-value obtainable in the Mann-Whitney test, indicating that the phenotypes of all targeting sgRNAs for these genes were more pronounced than any of the 3,790 non-targeting control sgRNAs. We also modeled noise and off-target effects in the system by generating a large set of ‘negative control genes’ composed of randomly selected sets of 10 non-targeting sgRNAs and scoring these genes as we did for true genes. In order to classify genes as hits in the screen, we used a score that integrated effect size and statistical confidence and applied the same threshold to both v1 and v2 screens. While in both screens fewer than 0.21% of these negative control genes scored as hits by these criteria, representing a ~2% empirically estimated false

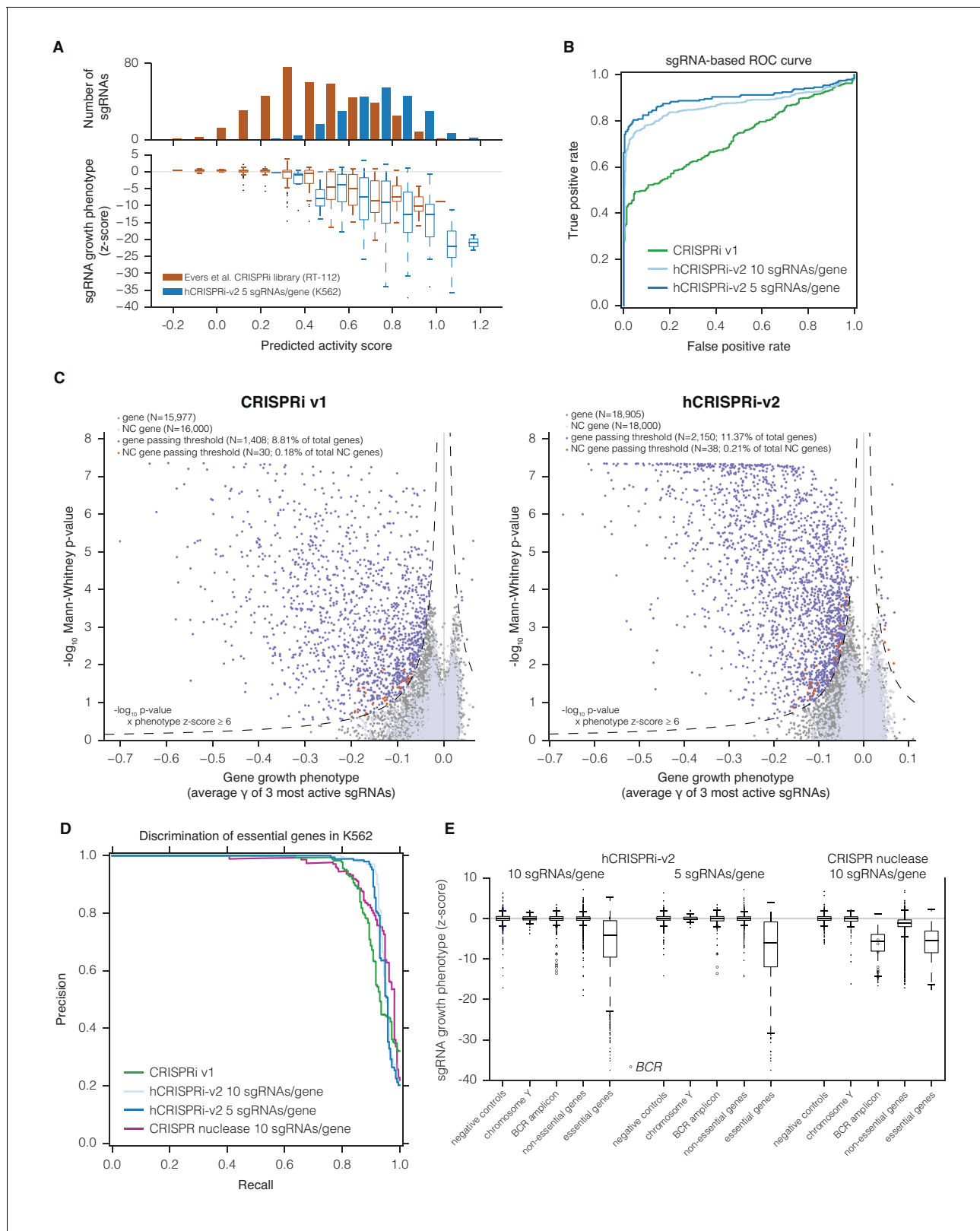


Figure 3. hCRISPRi-v2 outperforms CRISPRi v1 in screens for essential genes in K562. (A) Distribution and predicted scores for sgRNAs targeting essential genes. (Top) Predicted activity scores for sgRNAs from *Evers et al., 2016* or hCRISPRi-v2 targeting essential genes as defined by *Evers et al. (Evers et al., 2016)*, binned in increments of 0.1. (Bottom) sgRNA growth phenotypes of the sgRNAs in the above bins, z-standardized to the distribution of sgRNAs targeting *Evers et al.* non-essential genes. (B) ROC analysis of sgRNAs from CRISPRi v1 or hCRISPRi-v2 targeting essential and non-essential genes. (C) Scatter plots of $-\log_{10}$ Mann-Whitney p-value vs Gene growth phenotype (average γ of 3 most active sgRNAs) for CRISPRi v1 (left) and hCRISPRi-v2 (right). (D) Precision-Recall curve for the discrimination of essential genes in K562. (E) Box plots of sgRNA growth phenotype (z-score) for various gene categories. * BCR

Figure 3 continued

non-essential genes. sgRNAs were ranked by γ , and considered true or false positives if they targeted essential or non-essential genes, respectively, as defined by Evers et al. (C) Volcano plots of gene phenotypes and p-values for growth screens performed with CRISPRi v1 (Gilbert et al., 2014) and hCRISPRi-v2. For each screen, genes phenotypes were calculated by averaging the growth phenotype (γ) of the 3 sgRNAs with the strongest γ by absolute value, and gene p-values were calculated by performing the Mann-Whitney test comparing all sgRNAs targeting the gene to the full set of negative control sgRNAs. For genes with multiple TSSs targeted, sgRNAs were grouped by TSS and the TSS with the lowest p-value was used for downstream analysis. A comparable number of negative control (NC) genes were generated by randomly sampling 10 non-targeting sgRNAs (with replacement) and analyzed as true genes. Empirically derived thresholds (dashed lines) were calculated as shown, using the NC gene distribution to derive the background standard deviation for z-score. (D) Precision-recall analysis of essential gene screens performed in K562. Statistical precision and recall of essential and non-essential gene sets (Hart et al., 2014) were calculated for genes ranked by growth phenotype in K562. For both CRISPRi and CRISPR nuclease screens (Wang et al., 2015), gene-level phenotypes were calculated as the average \log_2 fold-change of all sgRNAs targeting the gene (termed CRISPR Scores in ref. [Wang et al., 2015]). (E) Boxplots of CRISPRi and CRISPR nuclease sgRNA phenotypes for several gene sets. sgRNA γ scores (CRISPRi) or \log_2 enrichments (nuclease) were z-standardized to the corresponding negative control set. Boxplots display the distribution of the negative control sgRNAs or sgRNAs targeting genes on Y chromosome (excluding pseudo-autosomal genes), within the BCR amplicon, or in the gold standard essential sets used in (D). Individual phenotypes for sgRNAs targeting BCR are overlaid with the corresponding boxplot.

DOI: 10.7554/eLife.19760.008

The following figure supplements are available for figure 3:

Figure supplement 1. sgRNA phenotypes from CRISPRi v1 and hCRISPRi-v2 growth screens.

DOI: 10.7554/eLife.19760.009

Figure supplement 2. Precision-recall analysis of second-generation CRISPR nuclease essential gene screens.

DOI: 10.7554/eLife.19760.010

discovery rate overall, we could confidently identify 2,150 essential gene hits in the v2 screen while our v1 screen identified 1,408 essential genes (Figure 3C).

In order to assess whether the stronger sgRNA- and gene-level growth phenotype γ scores produced by the hCRISPRi-v2 library resulted in improved discrimination of essential genes, we turned to precision-recall analysis of large 'gold standard' essential and non-essential gene sets introduced by Hart and colleagues (Hart et al., 2014). We ranked genes by their growth phenotype (γ) score and calculated at each phenotype threshold the trade-off between the recall of true essential genes and the avoidance of false positive non-essential genes, termed precision. In this analysis, the hCRISPRi-v2 library recalled over 91.2% of the gold standard essential genes at 95% precision compared to 81.5% in CRISPRi v1 (Figure 3D). We also found that the precision-recall of the top 5 sgRNA/gene half library was essentially identical to the full 10 sgRNA per gene library. Therefore, the 5 sgRNA/gene CRISPRi library represents a compelling tool for applications such as cell sorting-based screens (Liberati et al., 2015), or for *in vivo* screens where cell engraftment and library representation may represent a limiting factor (Braun et al., 2016; Chen et al., 2015).

Finally, we wanted to benchmark our hCRISPRi-v2 library against other recent genome-scale growth screens performed with nuclease-active Cas9 and novel second-generation libraries (Doench et al., 2016; Hart et al., 2015; Wang et al., 2015). Although these screens were performed in different labs and generally in different cell lines, precision-recall analysis offers a useful metric to compare these screens in an unbiased fashion (Hart et al., 2014). We found that our CRISPRi v2 library showed comparable or in many cases much greater discrimination (Figure 3—figure supplement 2). One published CRISPR nuclease screen was conducted in K562 (Wang et al., 2015) with a ~10 sgRNA/gene library, allowing for more direct comparison albeit still tempered by differences between labs and screening protocols. This screen recalled somewhat fewer (78.7% vs 90.8%) essential genes at 95% precision than our v2 library screen (Figure 3D). Together, these results indicate that our hCRISPRi-v2 library has a low false negative rate with few false positives, and the enrichment for highly active sgRNAs enables robust detection of phenotypes even with a compact library.

CRISPRi does not induce non-specific toxicity at amplified genomic loci

We were also intrigued by the observation by Wang and colleagues of K562-specific essentiality of many genes neighboring the BCR-ABL translocation, which they demonstrated to be mediated by non-specific toxicity of CRISPR-induced double-stranded breaks in the amplified locus (Wang et al., 2015; Wu et al., 1995). This toxicity appears to be pervasive as similar effects have been observed

across a range of cancer cell lines (Aguirre *et al.*, 2016; Munoz *et al.*, 2016). To test whether this toxicity could be caused by CRISPRi as well, we used our hCRISPRi-v2 screen as a representative dataset. When we compared the phenotypes of CRISPRi sgRNAs to CRISPR nuclease at the *BCR* amplicon, with all phenotypes standardized to the distribution of negative controls to facilitate comparison, we found that sgRNAs targeting *BCR* were strongly depleted in both screens, as expected based on the critical role of the BCR-ABL fusion in this cancer cell line (Naumann *et al.*, 2001), but few other CRISPRi sgRNAs in the region elicited growth defects (Figure 3E). We also found that CRISPR nuclease sgRNAs targeting the non-essential gene set were generally depleted relative to negative controls or chromosome Y-targeting sgRNAs, which should have no targets in the female-derived K562 cell line (Klein *et al.*, 1976), suggesting that in a CRISPR screen Cas9 nuclease activity can lead to measurable toxicity not related to the function of individual genes but instead due to the formation of on-target DNA double strand breaks, even with alleles present only at 2–3 copies (Naumann *et al.*, 2001). CRISPRi did not exhibit this generic toxicity at non-essential genes, allowing for detection of genes with subtle phenotypes relative to negative controls. Importantly, however, the vast majority of sgRNAs targeting essential genes showed a clear separation from the non-essential gene distribution (Figure 3E), demonstrating the high degree of sensitivity for detecting loss-of-function phenotypes with both CRISPR nuclease and CRISPRi screens.

The hCRISPRa-v2 library identifies more genes that modify robust growth rates upon overexpression

Finally, we sought to validate our hCRISPRa-v2 library design by conducting a screen for growth phenotypes in K562 cells expressing SunTag-VP64 constructs (Gilbert *et al.*, 2014; Tanenbaum *et al.*, 2014). We conducted two replicate growth screens (Figure 4—figure supplement 1A and Supplementary file 9) and analyzed the screen as with the hCRISPRi-v2 screen above to directly compare the results to our published CRISPRa screens (Gilbert *et al.*, 2014). Our hCRISPRa-v2 screen identified 540 genes to modify robust growth rates upon overexpression, 257 more than our previous CRISPRa screen (Figure 4A and Supplementary file 10). Beyond these additional hits, the v1 and v2 screens showed good agreement (Figure 4—figure supplement 1B), and the top categories in DAVID analysis of the v1 screen (Huang *et al.*, 2009), enrichment for transcription factor genes (in particular homeobox and forkhead box transcription factors), received ~3-fold greater enrichment scores in the hCRISPRa-v2 hits (Figure 4—figure supplement 1C), indicating the strong biological coherence of the additional genes. Analysis of the sgRNA growth phenotype (γ) scores for genes that were hits in both v1 and v2 screens showed that a greater fraction of sgRNAs were highly active (69.3% in hCRISPRa-v2 with 5 sgRNAs/gene versus 45.1% in CRISPRa v1; Figure 4C and Figure 4—figure supplement 1D), further validating improvements in the library design. In addition, as with our CRISPRi results, several genes identified in the hCRISPRi-v2 screen but not in v1, including hematopoietically-expressed homeobox *HHEX* and forkhead box C1 *FOXC1*, could be attributed to the use of the CAGE-based FANTOM5 TSS annotation (Figure 4D and Figure 4—figure supplement 1E). Finally, we compared the growth phenotypes from hCRISPRi-v2 to hCRISPRa-v2 and found that the two methods identified non-overlapping sets of genes that modify robust growth (Figure 4—figure supplement 1F), consistent with our previous results and highlighting the complementary information provided by these two approaches.

Discussion

Establishing design rules for effective reagents is critical to the implementation of genome-scale screening technologies. Our previous work established genome-scale CRISPRi and CRISPRa libraries as robust, specific, and complementary tools for dissecting biological pathways in human cells (Gilbert *et al.*, 2014). Here, we significantly improve upon this technology by developing a comprehensive predictive model to accurately identify highly active sgRNAs. This model includes both features specific for CRISPRi and CRISPRa, such as positioning relative to the TSS, as well as features like nucleosome occupancy that we expect to be generally important for most Cas9-mediated applications (Horlbeck *et al.*, 2016), and was able to accurately predict sgRNA activity in screen performed in a cell type it had not previously evaluated (Evers *et al.*, 2016). We used this prediction algorithm to design four new genome-scale libraries targeting human and mouse genomes. These

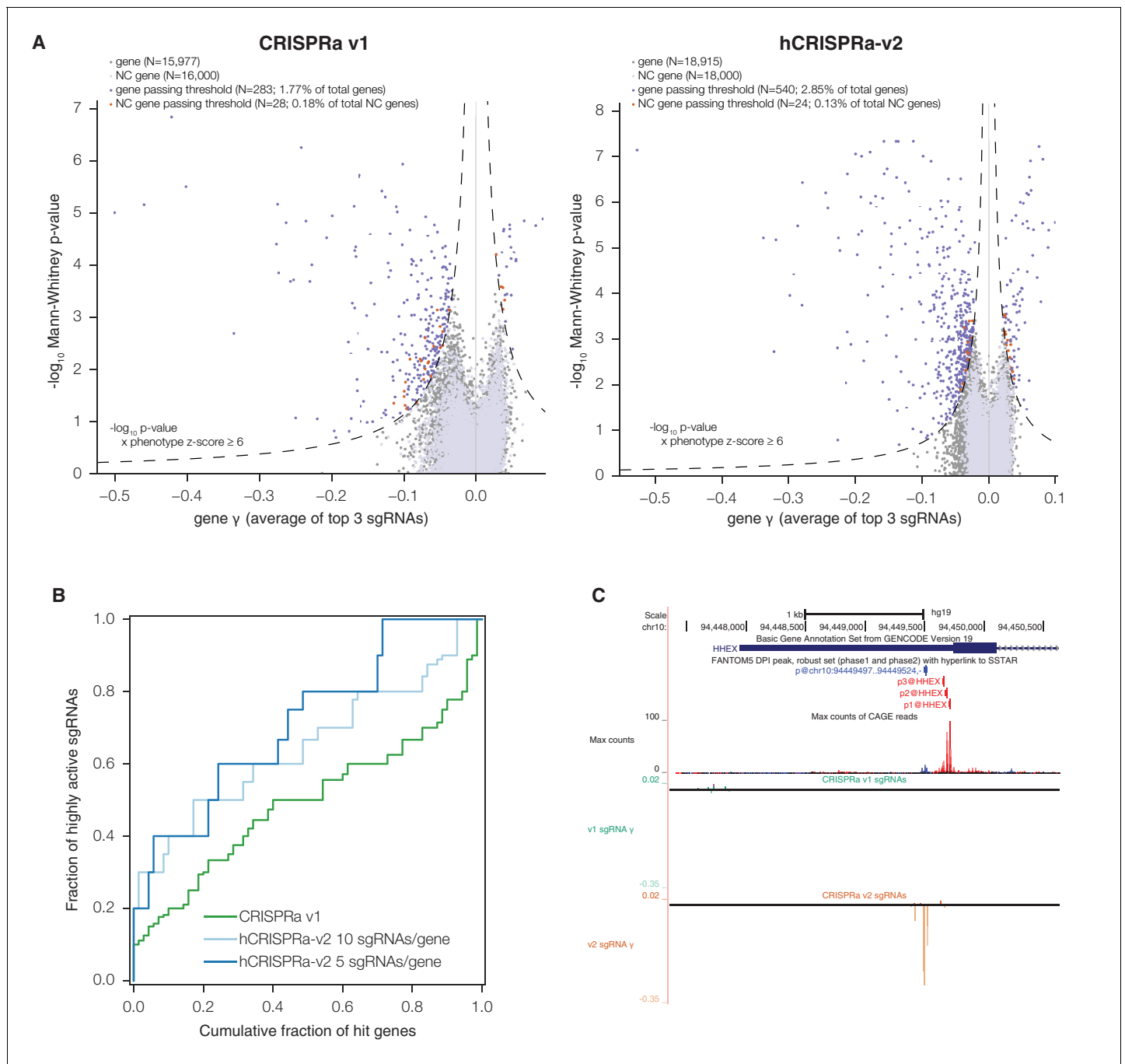


Figure 4. hCRISPRa-v2 outperforms CRISPRa v1 in screens for genes that modify growth rates upon overexpression. (A) Volcano plots of gene phenotypes and p-values for growth screens performed with CRISPRa v1 (Gilbert et al., 2014) and hCRISPRa-v2, presented as in Figure 3C. (B) Cumulative distributions of fraction of highly active sgRNAs targeting strong hit genes shared between CRISPRa v1 and hCRISPRa-v2 screens. Highly active sgRNAs for CRISPRa were defined as those with negative γ scores more than two standard deviations from the mean of non-targeting control sgRNAs (see Figure 4—figure supplement 1D). (C) UCSC Genome Browser tracks depicting TSS annotations and CRISPRa growth phenotypes for example gene HHEX.

DOI: 10.7554/eLife.19760.011

The following figure supplement is available for figure 4:

Figure supplement 1. sgRNA phenotypes and gene category enrichment scores from CRISPRa v1 and hCRISPRa-v2 growth screens.

DOI: 10.7554/eLife.19760.012

libraries are available on Addgene and *in silico* to facilitate design of focused libraries or targeted experiments (**Supplementary file 3–6**).

By performing CRISPRi and CRISPRa screens for genes that modify robust growth, we validate our sgRNA predictions and find that our hCRISPRi-v2 and hCRISPRa-v2 libraries represent a significant advance over our previous work in the fraction of highly active sgRNAs, the number of hits detected, and the statistical confidence of these hits. For CRISPRi, these improvements result in the accurate discrimination of essential genes by precision-recall analysis even with a compact 5 sgRNA/gene library. We believe that the greatly improved sgRNA prediction and lack of non-specific toxicity due to nuclease activity, combined with our previous findings that CRISPRi enables inducible, reversible, and homogenous manipulation of gene expression (**Gilbert et al., 2014; Mandegar et al., 2016; Qi et al., 2013**), make CRISPRi a state-of-the-art approach for loss-of-function studies.

While the libraries described here target genes annotated as protein-coding, CRISPRi and CRISPRa have been shown to be effective for repressing and activating transcription of non-coding genes as well (**Gilbert et al., 2014; Luo et al., 2016; Zhao et al., 2014**). Our v2 sgRNA prediction algorithms may enable the design of libraries to systematically manipulate expression of these transcripts as well. Furthermore, combining CRISPRi and CRISPRa with methods for robustly expressing multiple sgRNAs (**Kabadi et al., 2014; Wong et al., 2016; Zalatan et al., 2015**) will allow for simultaneous control of several genes, facilitating dissection of cellular pathways and systematic mapping of mammalian genetic interactions (**Bassik et al., 2013; Costanzo et al., 2010; Schuldiner et al., 2005**). Broadly, increased quantitative understanding of the factors dictating (d)Cas9 activity and specificity will greatly enhance the expanding set CRISPR-mediated technologies for controlling gene expression (**Hilton et al., 2015; Perez-Pinera et al., 2013; Vojta et al., 2016**), imaging targeted loci (**Chen et al., 2013; Shao et al., 2016**), or precisely editing the genome (**Komor et al., 2016; Tsai et al., 2014**).

Materials and methods

Machine learning for CRISPRi and CRISPRa sgRNA activity

Training and test datasets

The CRISPRi activity score dataset was obtained from **Horlbeck et al. (2016)**. CRISPRi and CRISPRa ricin tiling data was obtained from **Gilbert et al. (2014)**. CRISPRa activity scores were generated as previously described for the CRISPRi activity dataset, using data from 9 published and unpublished screening datasets. Hit genes were selected using the formula $|\text{effect size Z-score} \times \log_{10} \text{p-value}| \geq 20$ in any screen, and the phenotypes for sgRNAs targeting each gene were extracted from the screen in which the gene was a hit and normalized to the mean of the top 3 sgRNAs by absolute value. All datasets are included here as **Supplementary file 1**.

Generating TSS annotations

In order to leverage the high accuracy of the FANTOM TSS annotations but remain compatible with the comprehensive, systematic, and established Ensembl transcript models, a hybrid approach was taken. First, the full set of protein-coding genes and transcripts were selected as previously described (**Gilbert et al., 2014**), using Ensembl release 74 (corresponding to genome assemblies hg19 for human and mm10 for mouse; RRID:SCR_002344) and the APPRIS pipeline to identify the relevant functional transcripts and establish a preliminary set of TSS annotations (**Cunningham et al., 2015**). FANTOM CAGE peak BED files (RRID:SCR_002678; Human: http://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/TSS_classifier/TSS_human.bed.gz, accessed March 2, 2015; Mouse: http://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/TSS_classifier/TSS_mouse.bed.gz, accessed June 28, 2015 and lifted over to mm10 coordinates with UCSC liftOver tool) were then used to refine this annotation. Specifically, for each gene, the TSS was identified as the same-stranded peaks within 30 kb of any Ensembl TSS that matched the gene symbol and was labeled 'p1@gene' (here referred to as 'primary TSS') or 'p2@gene' ('secondary TSS'; included only where the peak passed the FANTOM robust threshold for TSS), and the annotation support was labeled as 'CAGE, matched peaks.' If multiple matches were found the closest TSS to a known Ensembl TSS was used. If no match was found, the gene could then be matched with any same-stranded CAGE peak within 500 bp labeled

as 'p1' or 'p2,' and annotation support was considered 'CAGE, primary peaks.' In the above cases, primary and secondary TSSs were targeted separately (i.e. by 10 sgRNAs each) if they were farther than 1 kb apart, or together as one 'P1P2' TSS. Where no matched or un-matched primary peaks were found, TSS annotations could be refined by any robust peaks or permissive peaks within 200 bp of the annotation (labeled 'CAGE, robust peak' or 'CAGE, permissive peak') or simply use the Ensembl/APPRIS annotation where no CAGE support was available ('Annotation'). This combined annotation is included as **Supplementary file 2**.

Calculating features

Position relative to the primary and secondary TSSs was calculated from the genomic coordinate of the 3'G of the PAM for each sgRNA to the upstream and downstream edge of each TSS range. Sequence features of the sgRNA and target sites were determined using custom scripts written in Python 2.7 (RRID:SCR_008394) with the Biopython module (RRID:SCR_007173) (Cock et al., 2009). The hCRISPRi-v2.1 algorithm included a change in these scripts to fix how certain sequence homopolymers were scored. All other libraries were generated with this improvement incorporated, and all analysis in this paper was performed with the v2.1 algorithm. RNA folding metrics were calculated using the ViennaRNA package (RRID:SCR_008550; version 2.2.5) with default parameters (Lorenz et al., 2011). Chromatin features at the target site were calculated as described previously (Horlbeck et al., 2016), averaging the signal at each base of the target site including the PAM. Custom Python scripts with the module bxpython (v0.5.0, <https://github.com/bxlab/bx-python>) to extract the processed continuous signal from the following BigWig files obtained from the ENCODE consortium: MNase-seq <https://www.encodeproject.org/files/ENCF000VNN/> (Michael Snyder lab, Stanford University), DNase-seq <https://www.encodeproject.org/files/ENCF000SVY/> (Gregory Crawford lab, Duke University), and FAIRE-seq <https://www.encodeproject.org/files/ENCF000TLU/> (Jason Lieb lab, University of North Carolina) (ENCODE Project Consortium, 2012). Beyond the nucleosome positioning information incorporated in the sgRNA positioning learning models, no chromatin data was used for predicting sgRNA activity for the mouse genome.

Machine learning

Training and test activity score sets were first divided into 80%/20% or 67%/33% sets of genes for CRISPRi or CRISPRa, respectively, as described in the results section. The training set parameters were then transformed according to their distribution as depicted in **Figure 1A**. For binning parameters, a fixed width was chosen for each feature and applied over the range of values, with the uppermost and lower-most bins collapsed with the neighboring bins if the number of data points at each edge were sparse. Each feature was then split into individual parameters for each bin and sgRNAs were assigned a 1 for the bin if the value fell within the bin or 0 if not. For linearizing sgRNA positioning parameters with continuous curves, sgRNA positions were fit to the activity score (individually for the distance to each TSS coordinate) using SVR with a radial basis function kernel and hyperparameters C and gamma determined using a grid search approach cross-validated within the training set. The fit score at each position was then used as the transformed linear parameter. Binary parameters were assigned a 1 if true or a 0 if false. All linearized parameters were z-standardized and fit with elastic net linear regression, with the l1/l2 ratio set by cross-validated grid search. All machine learning and downstream analysis was performed with custom Python scripts and the scikit-learn suite, version 0.15.0 (RRID:SCR_002577) (Pedregosa et al., 2011).

Design of next-generation CRISPRi and CRISPRa libraries

Prediction of sgRNA activity scores

All sequences within -25 and 500 bp (for CRISPRi) or -550 and -25 bp (for CRISPRa) of the upstream or the downstream edge of the primary or secondary TSS and containing 19 bp followed by an NGG PAM were extracted as potential sgRNAs for downstream scoring of predicted activity. All sequences were prepended with a 5' G to enable robust transcription from the U6 promoter, whether or not this base was present in the genomic sequence. Parameters were calculated for all sgRNAs as above, and transformed and scored using the CRISPRi or CRISPRa regression model from an arbitrarily chosen training set (test set ROC-AUC corresponding to these sets reported in results section).

Off-target scoring

Prediction of sgRNA off-target effects was performed using weighted Bowtie (v1.0.0, RRID:SCR_005476 [Ben Langmead et al., 2009]) alignment largely as previously described (Gilbert et al., 2014) with several adjustments. The '-tryhard' flag was added to the Bowtie command to increase sensitivity for mismatched sgRNA target sites. The hg19 and mm10 genomes used for alignment were masked at mitochondrial sequences and pseudoautosomal sequence to eliminate 'false positive' multiple alignments. Most importantly, as CRISPRi and CRISPRa have maximal effects proximal to the TSS, potential off-target alignments in these regions now were prioritized by creating a reference sequence corresponding to 1 kb windows around each TSS as defined above, along with the 5' end of every Ensembl transcript annotation. Reference sequences were generated using bedtools (v2.17.0, RRID:SCR_006646 [Quinlan and Hall, 2010]). In order to pass at the strictest threshold, sgRNAs were required to have no more than 1 alignment (the sgRNA target site itself) with 'mismatch score' (Gilbert et al., 2014) less than 31 proximal to the TSS and under 21 in the genome. (For hCRISPRi-v2, 96.6% of sgRNAs incorporated passed at this threshold.) In cases of difficult to target genes or close gene families, sgRNAs were allowed at relaxed thresholds. In descending order, these were: 1 alignment under 31 proximal to the TSS (no genomic threshold), 1 alignment under 21 in the genome, 2 alignments under 31 proximal to the TSS, and 3 alignments under 31 proximal to the TSS.

sgRNA selection

sgRNAs were chosen for inclusion into the genome-scale libraries based on predicted activity scores, empirical activity scores where available, off-target filtering, overlap with other sgRNAs already selected, and sequences with no restriction sites for enzymes used in cloning or sequencing sample processing (BstXI, BlnI, and SbfI). Empirical activity scores for CRISPRi/a v1 sgRNAs were generated as for the training sets above at a lower discriminant threshold of 7, and the corresponding sgRNAs were standardized to 19 bp with a 5' G prepended as above and subjected to the same revised off-target scoring procedure. For each TSS targeted by the library, up to 2 sgRNAs with the strongest empirical evidence were included first if the empirical activity score was at least 0.75, the sgRNA was less than 5 kb from the revised v2 TSS, the sgRNA passed the most stringent off-target filter, the sgRNA plus flanking cloning sequences did not contain extra restriction sites, and the sgRNA target site was at least 3 bp shifted from any previously selected target sequence. Once 2 empirically validated v1 sgRNAs were included, further sgRNAs fitting these criteria were not included but their predicted activity scores were increased by 0.2 to reflect the balance of information from the algorithm and empirical activity. Predicted sgRNAs were then sorted by best predicted activity score and included if the sgRNA passed the most stringent off-target filter, the sgRNA plus flanking cloning sequences did not contain extra restriction sites, the sgRNA target site was at least 3 bp shifted from any previously selected target sequence, and no more than 10 sgRNAs had been selected for the TSS. If fewer than 10 sgRNAs were selected by this algorithm, off-target stringency was iteratively relaxed as above and selection was continued to attain 10 sgRNAs. If 10 sgRNAs passing the most relaxed threshold could not be identified, the TSS was not targeted by the library.

sgRNA on-target and off-target prediction algorithms, library design scripts, and associated files are available at <https://github.com/mhorlbeck/CRISPRiDesign>.

Negative controls

For each library, the frequency of each DNA base at each position along the sgRNA protospacer sequence was calculated. Random sgRNA protospacer sequences weighted by these base frequencies were then generated to mirror the composition of the targeting sgRNAs. These were then filtered for sgRNAs with 0 alignments with a mismatch score less than 31 proximal to the TSS and 0 alignments under 21 in the genome as above.

Library cloning

Protospacer sequences were appended with cloning sequences and then unique PCR adapters corresponding to the designated sublibrary. The half libraries were determined from the first 5 and second 5 sgRNAs selected into the library for each TSS according to the algorithm above. Genes were then partitioned into one of the 13 sublibraries defined by (Kampmann et al., 2015), compressed

Daniel Ryan, Dr. Luke Fairbairn, and Dr. Peter Tsang of Agilent Technologies for their assistance on the design and synthesis of oligonucleotide pools. MAH, LAG, JEV, BA, YC, APF, and JSW were supported by the Howard Hughes Medical Institutes and the National Institutes of Health (P50 GM102706, U01 CA168370, R01 DA036858). LAG was supported by the Leukemia and Lymphoma Society. RAP, CYP, and JEC were supported by the Li Ka Shing Foundation. MK was supported by NIH/NIGMS DP2 GM119139.

Additional information

Competing interests

MAH: patent application related to CRISPRi and CRISPRa screening (PCT/US15/40449). JSW is a founder of, and MAH and LAG are consultants for, KSQ Therapeutics, a CRISPR functional genomics company. LAG, JSW: filed a patent application related to CRISPRi and CRISPRa screening (PCT/US15/40449). JSW is a founder of, and MAH and LAG are consultants for, KSQ Therapeutics, a CRISPR functional genomics company. MK: patent application related to CRISPRi and CRISPRa screening (PCT/US15/40449). The other authors declare that no competing interests exist.

Funding

Funder	Grant reference number	Author
Howard Hughes Medical Institute		Max A Horlbeck Luke A Gilbert Jacqueline E Villalta Britt Adamson Yuwen Chen Alexander P Fields
National Institutes of Health	P50 GM102706	Max A Horlbeck Luke A Gilbert Jacqueline E Villalta Britt Adamson Yuwen Chen Alexander P Fields
National Institutes of Health	U01 CA168370	Max A Horlbeck Luke A Gilbert Jacqueline E Villalta Britt Adamson Yuwen Chen Alexander P Fields
National Institutes of Health	R01 DA036858	Max A Horlbeck Luke A Gilbert Jacqueline E Villalta Britt Adamson Yuwen Chen Alexander P Fields
Leukemia and Lymphoma Society		Luke A Gilbert
National Cancer Institute	Pathway to Independence Award, K99 CA204602	Luke A Gilbert
Li Ka Shing Foundation		Ryan A Pak Chong Yon Park Jacob E Corn
National Institute of General Medical Sciences	DP2 GM119139	Martin Kampmann

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

MAH, Conceived of and conducted algorithm development and data analysis, contributed to genome-scale screening experiments, and wrote this report; LAG, Contributed to sgRNA library development and genome-scale screening experiments, and helped write this report; JEV,

Contributed to sgRNA library development, generated sgRNA libraries, and contributed technical assistance; BA, Conducted genome-scale screening experiments; RAP, Generated sgRNA libraries; YC, Contributed technical assistance; APF, Contributed to algorithm development; CYP, Supervised sgRNA library generation and conducted genome-scale screening experiments; JEC, Contributed to algorithm development and supervised sgRNA library generation; MK, Contributed to and supervised algorithm development; JSW, Conceived of and supervised the study, and helped write this report

Author ORCIDs

Max A Horlbeck, <http://orcid.org/0000-0002-3875-871X>

Ryan A Pak, <http://orcid.org/0000-0003-3507-3122>

Jacob E Corn, <http://orcid.org/0000-0002-7798-5309>

Martin Kampmann, <http://orcid.org/0000-0002-3819-7019>

Jonathan S Weissman, <http://orcid.org/0000-0003-2445-670X>

Additional files

Supplementary files

- Supplementary file 1. CRISPRi and CRISPRa activity score datasets.
DOI: [10.7554/eLife.19760.013](https://doi.org/10.7554/eLife.19760.013)
- Supplementary file 2. TSS annotations for hg19 and mm10 genomes.
DOI: [10.7554/eLife.19760.014](https://doi.org/10.7554/eLife.19760.014)
- Supplementary file 3. Library composition of hCRISPRi-v2 and hCRISPRi-v2.1.
DOI: [10.7554/eLife.19760.015](https://doi.org/10.7554/eLife.19760.015)
- Supplementary file 4. Library composition of mCRISPRi-v2.
DOI: [10.7554/eLife.19760.016](https://doi.org/10.7554/eLife.19760.016)
- Supplementary file 5. Library composition of hCRISPRa-v2.
DOI: [10.7554/eLife.19760.017](https://doi.org/10.7554/eLife.19760.017)
- Supplementary file 6. Library composition of mCRISPRa-v2.
DOI: [10.7554/eLife.19760.018](https://doi.org/10.7554/eLife.19760.018)
- Supplementary file 7. sgRNA read counts and growth phenotypes for hCRISPRi-v2 screens performed in K562.
DOI: [10.7554/eLife.19760.019](https://doi.org/10.7554/eLife.19760.019)
- Supplementary file 8. Gene growth phenotypes and p-values for hCRISPRi-v2 screens performed in K562.
DOI: [10.7554/eLife.19760.020](https://doi.org/10.7554/eLife.19760.020)
- Supplementary file 9. sgRNA read counts and growth phenotypes for hCRISPRa-v2 screens performed in K562.
DOI: [10.7554/eLife.19760.021](https://doi.org/10.7554/eLife.19760.021)
- Supplementary file 10. Gene growth phenotypes and p-values for hCRISPRa-v2 screens performed in K562.
DOI: [10.7554/eLife.19760.022](https://doi.org/10.7554/eLife.19760.022)

Major datasets

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
FANTOM Consortium	2014	A promoter-level mammalian expression atlas	http://fantom.gsc.riken.jp/5/datafiles/phase1.3/extra/TSS_classifier/	Publicly available at FANTOM (files TSS_human.bed.gz, TSS_mouse.bed.gz)

ENCODE Consortium/Snyder	2011	Determinants of nucleosome organization in primary human cells	https://www.encodeproject.org/files/ENCF000VNN	Publicly available at ENCODE (accession no. ENCF000VNN)
ENCODE Consortium	2012	An integrated encyclopedia of DNA elements in the human genome	https://www.encodeproject.org/files/ENCF000TLU	Publicly available at ENCODE (accession no. ENCF000TLU)

References

- Aguirre AJ**, Meyers RM, Weir BA, Vazquez F, Zhang CZ, Ben-David U, Cook A, Ha G, Harrington WF, Doshi MB, Kost-Alimova M, Gill S, Xu H, Ali LD, Jiang G, Pantel S, Lee Y, Goodale A, Cherniack AD, Oh C, et al. 2016. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discovery* **6**:914–929. doi: [10.1158/2159-8290.CD-16-0154](https://doi.org/10.1158/2159-8290.CD-16-0154), PMID: [27260156](https://pubmed.ncbi.nlm.nih.gov/27260156/)
- Bassik MC**, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, Weibezahn J, Horlbeck MA, Chen S, Mann M, Hyman AA, Leproust EM, McManus MT, Weissman JS. 2013. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152**:909–922. doi: [10.1016/j.cell.2013.01.030](https://doi.org/10.1016/j.cell.2013.01.030), PMID: [23394947](https://pubmed.ncbi.nlm.nih.gov/23394947/)
- Braun CJ**, Bruno PM, Horlbeck MA, Gilbert LA, Weissman JS, Hemann MT. 2016. Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. *PNAS* **113**:E3892–E3900. doi: [10.1073/pnas.1600582113](https://doi.org/10.1073/pnas.1600582113)
- Chari R**, Mali P, Moosburner M, Church GM. 2015. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nature Methods* **12**:823–826. doi: [10.1038/nmeth.3473](https://doi.org/10.1038/nmeth.3473), PMID: [26167643](https://pubmed.ncbi.nlm.nih.gov/26167643/)
- Chavez A**, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, Lin S, Kiani S, Guzman CD, Wiegand DJ, Ter-Ovanesyan D, Braff JL, Davidsohn N, Housden BE, Perrimon N, Weiss R, Aach J, Collins JJ, Church GM, Iyer PR, et al. 2015. Highly efficient Cas9-mediated transcriptional programming. *Nature Methods* **12**:326–328. doi: [10.1038/nmeth.3312](https://doi.org/10.1038/nmeth.3312), PMID: [25730490](https://pubmed.ncbi.nlm.nih.gov/25730490/)
- Chavez A**, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJ, Buchthal J, Housden BE, Perrimon N, Collins JJ, Church G, Haque D. 2016. Comparison of Cas9 activators in multiple species. *Nature Methods* **13**:563–567. doi: [10.1038/nmeth.3871](https://doi.org/10.1038/nmeth.3871), PMID: [27214048](https://pubmed.ncbi.nlm.nih.gov/27214048/)
- Chen B**, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**:1479–1491. doi: [10.1016/j.cell.2013.12.001](https://doi.org/10.1016/j.cell.2013.12.001), PMID: [24360272](https://pubmed.ncbi.nlm.nih.gov/24360272/)
- Chen S**, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. 2015. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**:1246–1260. doi: [10.1016/j.cell.2015.02.038](https://doi.org/10.1016/j.cell.2015.02.038), PMID: [25748654](https://pubmed.ncbi.nlm.nih.gov/25748654/)
- Cock PJ**, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**:1422–1423. doi: [10.1093/bioinformatics/btp163](https://doi.org/10.1093/bioinformatics/btp163), PMID: [19304878](https://pubmed.ncbi.nlm.nih.gov/19304878/)
- Costanzo M**, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, et al. 2010. The genetic landscape of a cell. *Science* **327**:425–431. doi: [10.1126/science.1180823](https://doi.org/10.1126/science.1180823), PMID: [20093466](https://pubmed.ncbi.nlm.nih.gov/20093466/)
- Cunningham F**, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, Hunt SE, Janacek SH, Johnson N, Juettemann T, Kähäri AK, Keenan S, et al. 2015. Ensembl 2015. *Nucleic Acids Research* **43**:D662–D669. doi: [10.1093/nar/gku1010](https://doi.org/10.1093/nar/gku1010), PMID: [25352552](https://pubmed.ncbi.nlm.nih.gov/25352552/)
- Dang Y**, Jia G, Choi J, Ma H, Anaya E, Ye C, Shankar P, Wu H. 2015. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology* **16**:631–633. doi: [10.1186/s13059-015-0846-3](https://doi.org/10.1186/s13059-015-0846-3)
- Doench JG**, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J, Root DE. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**:184–191. doi: [10.1038/nbt.3437](https://doi.org/10.1038/nbt.3437), PMID: [26780180](https://pubmed.ncbi.nlm.nih.gov/26780180/)
- Doench JG**, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nature Biotechnology* **32**:1262–U130. doi: [10.1038/nbt.3026](https://doi.org/10.1038/nbt.3026), PMID: [25184501](https://pubmed.ncbi.nlm.nih.gov/25184501/)
- ENCODE Project Consortium**. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57–74. doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247), PMID: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)
- Evers B**, Jastrzebski K, Heijmans JP, Grønrum W, Beijersbergen RL, Bernards R. 2016. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nature Biotechnology* **34**:631–3. doi: [10.1038/nbt.3536](https://doi.org/10.1038/nbt.3536), PMID: [27111720](https://pubmed.ncbi.nlm.nih.gov/27111720/)
- FANTOM Consortium and the RIKEN PMI and CLST (DGT)**, Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Itoh M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jørgensen M, Dimont E, Arner E, Schmid C, Schaefer U, et al. 2014. A promoter-level mammalian expression atlas. *Nature* **507**:462–470. doi: [10.1038/nature13182](https://doi.org/10.1038/nature13182), PMID: [24670764](https://pubmed.ncbi.nlm.nih.gov/24670764/)
- Gilbert LA**, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS. 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**:647–661. doi: [10.1016/j.cell.2014.09.029](https://doi.org/10.1016/j.cell.2014.09.029), PMID: [25307932](https://pubmed.ncbi.nlm.nih.gov/25307932/)

- Gilbert LA**, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**:442–451. doi: [10.1016/j.cell.2013.06.044](https://doi.org/10.1016/j.cell.2013.06.044), PMID: [23849981](https://pubmed.ncbi.nlm.nih.gov/23849981/)
- Harrow J**, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Research* **22**:1760–1774. doi: [10.1101/gr.135350.111](https://doi.org/10.1101/gr.135350.111), PMID: [22955987](https://pubmed.ncbi.nlm.nih.gov/22955987/)
- Hart T**, Brown KR, Sircoulomb F, Rottapel R, Moffat J. 2014. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Molecular Systems Biology* **10**:733. doi: [10.15252/msb.20145216](https://doi.org/10.15252/msb.20145216), PMID: [24987113](https://pubmed.ncbi.nlm.nih.gov/24987113/)
- Hart T**, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat J. 2015. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**:1515–1526. doi: [10.1016/j.cell.2015.11.015](https://doi.org/10.1016/j.cell.2015.11.015), PMID: [26627737](https://pubmed.ncbi.nlm.nih.gov/26627737/)
- Hilton IB**, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA. 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology* **33**:510–517. doi: [10.1038/nbt.3199](https://doi.org/10.1038/nbt.3199), PMID: [25849900](https://pubmed.ncbi.nlm.nih.gov/25849900/)
- Hinz JM**, Laughery MF, Wyrick JJ. 2015. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry* **54**:7063–7066. doi: [10.1021/acs.biochem.5b01108](https://doi.org/10.1021/acs.biochem.5b01108), PMID: [26579937](https://pubmed.ncbi.nlm.nih.gov/26579937/)
- Horlbeck MA**, Witkowski LB, Guglielmi B, Replogle JM, Gilbert LA, Villalta JE, Torigoe SE, Tjian R, Weissman JS. 2016. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **5**:e12677. doi: [10.7554/eLife.12677](https://doi.org/10.7554/eLife.12677), PMID: [26987018](https://pubmed.ncbi.nlm.nih.gov/26987018/)
- Huang da W**, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**:1–13. doi: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923), PMID: [19033363](https://pubmed.ncbi.nlm.nih.gov/19033363/)
- Zou H**, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B* **67**:301–320. doi: [10.1111/j.1467-9868.2005.00503.x](https://doi.org/10.1111/j.1467-9868.2005.00503.x)
- Isaac RS**, Jiang F, Doudna JA, Lim WA, Narlikar GJ, Almeida R. 2016. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **5**:e13450. doi: [10.7554/eLife.13450](https://doi.org/10.7554/eLife.13450), PMID: [27130520](https://pubmed.ncbi.nlm.nih.gov/27130520/)
- Kabadi AM**, Ousterout DG, Hilton IB, Gersbach CA. 2014. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Research* **42**:e147. doi: [10.1093/nar/gku749](https://doi.org/10.1093/nar/gku749), PMID: [25122746](https://pubmed.ncbi.nlm.nih.gov/25122746/)
- Kampmann M**, Bassik MC, Weissman JS. 2013. Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *PNAS* **110**:E2317–E2326. doi: [10.1073/pnas.1307002110](https://doi.org/10.1073/pnas.1307002110)
- Kampmann M**, Horlbeck MA, Chen Y, Tsai JC, Bassik MC, Gilbert LA, Villalta JE, Kwon SC, Chang H, Kim VN, Weissman JS. 2015. Next-generation libraries for robust RNA interference-based genome-wide screens. *PNAS* **112**:E3384–E3391. doi: [10.1073/pnas.1508821112](https://doi.org/10.1073/pnas.1508821112)
- Klein E**, Ben-Bassat H, Neumann H, Ralph P, Zeuthen J, Polliack A, Vánky F. 1976. Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *International Journal of Cancer* **18**:421–431. doi: [10.1002/ijc.2910180405](https://doi.org/10.1002/ijc.2910180405), PMID: [789258](https://pubmed.ncbi.nlm.nih.gov/789258/)
- Komor AC**, Kim YB, Packer MS, Zuris JA, Liu DR. 2016. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**:420–424. doi: [10.1038/nature17946](https://doi.org/10.1038/nature17946), PMID: [27096365](https://pubmed.ncbi.nlm.nih.gov/27096365/)
- Konermann S**, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**:583–588. doi: [10.1038/nature14136](https://doi.org/10.1038/nature14136), PMID: [25494202](https://pubmed.ncbi.nlm.nih.gov/25494202/)
- Langmead B**, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. doi: [10.1186/gb-2009-10-3-r25](https://doi.org/10.1186/gb-2009-10-3-r25), PMID: [19261174](https://pubmed.ncbi.nlm.nih.gov/19261174/)
- Liberali P**, Snijder B, Pelkmans L. 2015. Single-cell and multivariate approaches in genetic perturbation screens. *Nature Reviews Genetics* **16**:18–32. doi: [10.1038/nrg3768](https://doi.org/10.1038/nrg3768), PMID: [25446316](https://pubmed.ncbi.nlm.nih.gov/25446316/)
- Lorenz R**, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**:26. doi: [10.1186/1748-7188-6-26](https://doi.org/10.1186/1748-7188-6-26), PMID: [22115189](https://pubmed.ncbi.nlm.nih.gov/22115189/)
- Luo S**, Lu JY, Liu L, Yin Y, Chen C, Han X, Wu B, Xu R, Liu W, Yan P, Shao W, Lu Z, Li H, Na J, Tang F, Wang J, Zhang YE, Shen X. 2016. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* **18**:637–652. doi: [10.1016/j.stem.2016.01.024](https://doi.org/10.1016/j.stem.2016.01.024), PMID: [26996597](https://pubmed.ncbi.nlm.nih.gov/26996597/)
- Maeder ML**, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10**:977–979. doi: [10.1038/nmeth.2598](https://doi.org/10.1038/nmeth.2598), PMID: [23892898](https://pubmed.ncbi.nlm.nih.gov/23892898/)
- Mandegar MA**, Huebsch N, Frolov EB, Shin E, Truong A, Olvera MP, Chan AH, Miyaoka Y, Holmes K, Spencer CI, Judge LM, Gordon DE, Eskildsen TV, Villalta JE, Horlbeck MA, Gilbert LA, Krogan NJ, Sheikh SP, Weissman JS, Qi LS, et al. 2016. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**:541–553. doi: [10.1016/j.stem.2016.01.022](https://doi.org/10.1016/j.stem.2016.01.022), PMID: [26971820](https://pubmed.ncbi.nlm.nih.gov/26971820/)
- Munoz DM**, Cassiani PJ, Li L, Billy E, Korn JM, Jones MD, Golji J, Ruddy DA, Yu K, McAllister G, DeWeck A, Abramowski D, Wan J, Shirley MD, Neshat SY, Rakiec D, de Beaumont R, Weber O, Kauffmann A, McDonald ER, et al. 2016. CRISPR Screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discovery* **6**:900–913. doi: [10.1158/2159-8290.CD-16-0178](https://doi.org/10.1158/2159-8290.CD-16-0178), PMID: [27260157](https://pubmed.ncbi.nlm.nih.gov/27260157/)
- Naumann S**, Reutzel D, Speicher M, Decker HJ. 2001. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ

- hybridization, and comparative genomic hybridization. *Leukemia Research* **25**:313–322. doi: [10.1016/S0145-2126\(00\)00125-9](https://doi.org/10.1016/S0145-2126(00)00125-9), PMID: [11248328](https://pubmed.ncbi.nlm.nih.gov/11248328/)
- Paddison PJ**, Silva JM, Conklin DS, Schlabach M, Li M, Aruleba S, Balija V, O’Shaughnessy A, Gnoj L, Scobie K, Chang K, Westbrook T, Cleary M, Sachidanandam R, McCombie WR, Elledge SJ, Hannon GJ. 2004. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**:427–431. doi: [10.1038/nature02370](https://doi.org/10.1038/nature02370), PMID: [15042091](https://pubmed.ncbi.nlm.nih.gov/15042091/)
- Pedregosa F**, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V. 2011. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research* **12**:2825–2830.
- Perez-Pinera P**, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA, Thakore LR, Glass PI. 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature Methods* **10**:973–976. doi: [10.1038/nmeth.2600](https://doi.org/10.1038/nmeth.2600)
- Qi LS**, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**:1173–1183. doi: [10.1016/j.cell.2013.02.022](https://doi.org/10.1016/j.cell.2013.02.022), PMID: [23452860](https://pubmed.ncbi.nlm.nih.gov/23452860/)
- Quinlan AR**, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi: [10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033), PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
- Radziszewska A**, Shlyueva D, Müller I, Helin K. 2016. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Research*:gkw583. doi: [10.1093/nar/gkw583](https://doi.org/10.1093/nar/gkw583), PMID: [27353328](https://pubmed.ncbi.nlm.nih.gov/27353328/)
- Schuldiner M**, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ. 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**:507–519. doi: [10.1016/j.cell.2005.08.031](https://doi.org/10.1016/j.cell.2005.08.031), PMID: [16269340](https://pubmed.ncbi.nlm.nih.gov/16269340/)
- Shalem O**, Sanjana NE, Zhang F. 2015. High-throughput functional genomics using CRISPR-Cas9. *Nature Reviews Genetics* **16**:299–311. doi: [10.1038/nrg3899](https://doi.org/10.1038/nrg3899), PMID: [25854182](https://pubmed.ncbi.nlm.nih.gov/25854182/)
- Shao S**, Zhang W, Hu H, Xue B, Qin J, Sun C, Sun Y, Wei W, Sun Y. 2016. Long-term dual-color tracking of genomic loci by modified sgRNAs of the CRISPR/Cas9 system. *Nucleic Acids Research* **44**:e86. doi: [10.1093/nar/gkw066](https://doi.org/10.1093/nar/gkw066), PMID: [26850639](https://pubmed.ncbi.nlm.nih.gov/26850639/)
- Tanenbaum ME**, Gilbert LA, Qi LS, Weissman JS, Vale RD. 2014. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**:635–646. doi: [10.1016/j.cell.2014.09.039](https://doi.org/10.1016/j.cell.2014.09.039), PMID: [25307933](https://pubmed.ncbi.nlm.nih.gov/25307933/)
- Tsai SQ**, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology* **32**:569–576. doi: [10.1038/nbt.2908](https://doi.org/10.1038/nbt.2908), PMID: [24770325](https://pubmed.ncbi.nlm.nih.gov/24770325/)
- Vojta A**, Dobrinic P, Tadic V, Bočkor L, Korać P, Julg B, Klasić M, Zoldoš V. 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research* **44**:gkw159. doi: [10.1093/nar/gkw159](https://doi.org/10.1093/nar/gkw159), PMID: [26969735](https://pubmed.ncbi.nlm.nih.gov/26969735/)
- Wang T**, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM. 2015. Identification and characterization of essential genes in the human genome. *Science* **350**:1096–1101. doi: [10.1126/science.aac7041](https://doi.org/10.1126/science.aac7041), PMID: [26472758](https://pubmed.ncbi.nlm.nih.gov/26472758/)
- Wong ASL**, Choi GCG, Cui CH, Pregonig G, Milani P, Adam M, Perli SD, Kazer SW, Gaillard A, Hermann M, Shalek AK, Fraenkel E, Lu TK. 2016. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *PNAS* **113**:2544–2549. doi: [10.1073/pnas.1517883113](https://doi.org/10.1073/pnas.1517883113)
- Wu SQ**, Voelkerding KV, Sabatini L, Chen XR, Huang J, Meisner LF. 1995. Extensive amplification of bcr/abl fusion genes clustered on three marker chromosomes in human leukemic cell line K-562. *Leukemia* **9**:858–862. PMID: [7769849](https://pubmed.ncbi.nlm.nih.gov/7769849/)
- Xu H**, Xiao T, Chen CH, Li W, Meyer CA, Wu Q, Wu D, Cong L, Zhang F, Liu JS, Brown M, Liu XS. 2015. Sequence determinants of improved CRISPR sgRNA design. *Genome Research* **25**:1147–1157. doi: [10.1101/gr.191452.115](https://doi.org/10.1101/gr.191452.115), PMID: [26063738](https://pubmed.ncbi.nlm.nih.gov/26063738/)
- Zalatan JG**, Lee ME, Almeida R, Gilbert LA, Whitehead EH, La Russa M, Tsai JC, Weissman JS, Dueber JE, Qi LS, Lim WA. 2015. Engineering complex synthetic transcriptional programs with CRISPR RNA Scaffolds. *Cell* **160**:339–350. doi: [10.1016/j.cell.2014.11.052](https://doi.org/10.1016/j.cell.2014.11.052), PMID: [25533786](https://pubmed.ncbi.nlm.nih.gov/25533786/)
- Zhao Y**, Dai Z, Liang Y, Yin M, Ma K, He M, Ouyang H, Teng CB. 2014. Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Scientific Reports* **4**:3943. doi: [10.1038/srep03943](https://doi.org/10.1038/srep03943), PMID: [24487629](https://pubmed.ncbi.nlm.nih.gov/24487629/)