PORTLAND PRESS

Review Article

Common computational tools for analyzing CRISPR screens

Medina Colic[1,2] **and** ⓘ **Traver Hart**[1,3]

[1]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.; [2]UTHealth Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, U.S.A.; [3]Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, U.S.A.

Correspondence: Traver Hart (traver@hart-lab.org)

OPEN ACCESS

CRISPR–Cas technology offers a versatile toolbox for genome editing, with applications in various cancer-related fields such as functional genomics, immunotherapy, synthetic lethality and drug resistance, metastasis, genome regulation, chromatic accessibility and RNA-targeting. The variety of screening platforms and questions in which they are used have caused the development of a wide array of analytical methods for CRISPR analysis. In this review, we focus on the algorithms and frameworks used in the computational analysis of pooled CRISPR knockout (KO) screens and highlight some of the most significant target discoveries made using these methods. Lastly, we offer perspectives on the design and analysis of state-of-art multiplex screening for genetic interactions.

Introduction

The last decade highlights the new genome editing technology repurposed from a bacterial adaptive immune system [1] and its adaptation to mammalian genome engineering [2,3]. The CRISPR–Cas system utilizes the Cas nuclease, which is guided to the target sequence by a short guide RNA molecule (gRNA), where it introduces a double-strand break at the desired locus (Figure 1A). Activation of error-prone repair by nonhomologous end-joining pathways (NHEJ) results in a frameshift mutation creating a gene knockout (KO). When DNA damage is too great a burden on the model systems, alternative, engineered Cas approaches are available. Nuclease-inactivaed 'dead' Cas9 (dCas9) can be fused with transcriptional activation or repression domains and targeted to gene promoters to activate (CRISPRa) or repress/inhibit (CRISPRi) gene transcription (Figure 1B,C). dCas9 systems have been reviewed in a greater detail by Kazi and Biswas [4].

Gene KO is the most widely used tool in the CRISPR toolkit. CRISPR KO screens answer how essential, or how necessary a gene is for a cellular fitness, with genes showing moderate to severe fitness defects often called 'fitness genes' or 'essential genes'. Exceptional examples of genome-wide CRISPR KO screens are two large pan-cancer CRISPR–Cas9 studies performed by the Broad Institute and the Wellcome Sanger Institutes [5,6], in which over a thousand cancer cell lines were screened with genome-scale KO screens. In addition to individual efforts, these two institutes work collaboratively [7–9] with an aim of creating a comprehensive map of all the intracellular genetic dependencies and vulnerabilities of cancer, known as the Cancer Dependency Map (DepMap) project [10,11]. Such efforts hold a premise of providing a comprehensive representation of cancer heterogeneity and an avenue for developing new therapies.

The emergence of CRISPR-mediated genetic screens and continued improvement in CRISPR reagent design [12–14] has enable investigation of genome-wide and custom libraries gene-drug interaction in human cells [15–27]. These studies illustrated the power of chemogenetic screens (CRISPR + drug perturbation) (Figure 1D) in identifying new genetic vulnerabilities to PARP, ATR, BRAF, NAMPT inhibitors, and temozolomide, and shed a light on using such experimental set-ups for a discovery of novel therapeutic targets. A more comprehensive overview of chemogenetic screens in human cancer cell lines can be found in a similar review [28].
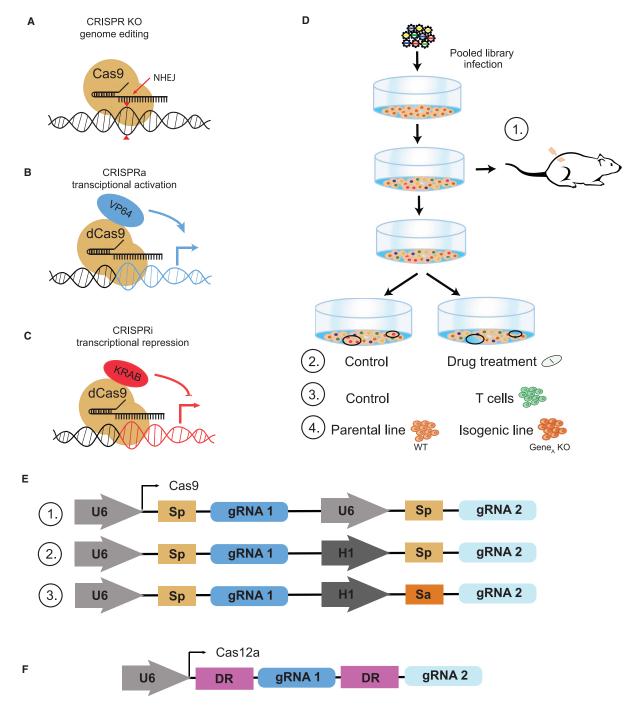
Figure 1. CRISPR Toolbox.
(**A**) CRISPR knockout. (**B**) CRISPR activation. (**C**) CRISPR interferance, (**D**) Pooled screens: (**1**) *In vivo*, (**2**) Chemogenetic, (**3**) Immuno-oncology, and (**4**) Isogenic screens. (**E**) Cas9 multiplex platforms: (**1**) Single Cas9 (e.g. S. pyogenes) system using two copies of the U6 promoter. (**2**) Single Cas9 system uses two different promoters. (**3**) A two Cas9, two different promoters system. (**F**) EnCas12a multiplex platform.

In comparison with cell culture, *in vivo* systems are preferred for translational cancer research (e.g. evaluating tumor progression and therapeutic response), as they provide a more clinically relevant environment for tumor modeling. CRISPR editing in *in vivo* model in conducted by creating the mutant cell population of interest in a dish and then implanting those into a mouse, often subcutaneously or intravenously (Figure 1D). In the last

few years, CRISPR technology has been used in living model organisms for studying various cancers and cancer specific processes [29–35], though the complexity of the approach limits these screens to targeted gene panels.

CRISPR screens are being used in immune-oncology studies as well, with the most common approach being to proliferate CRISPR-mutagenized cells in the presence or absence of T cells (or other immune system components) (Figure 1D). In the recent years, pooled CRISPR screens in tumor/immune co-culture systems have provided insights into tumor mechanisms that cause resistance to immunotherapies [36,37], genes involved in the immune synergistic interactions [36], and identification of novel targets for immune-oncology [38]. The studies described in the review focusing on interrogating immune cells and cancer with CRISPR–Cas9 [39] are the proof that the CRISPR screens are a powerful tool for investigating tumor–immune co-culture systems.

Though these approaches offer an enormous advantage over the prior state of the art, widespread genetic buffering imposes clear constraints on the ability of monogenic KO systems to provide saturating screens. These constraints have driven the development of multiplex targeting platforms via delivery of multiple sgRNAs per cell. This can be facilitated by using two Cas9 nucleases derived from different bacterial species, e. g. *S. pyogenes* and *S. aureus*, with species-specific gRNA expressed from different promoters (e.g. hU6 and mU6) (Figure 1E). Other systems use a single SpCas9 with two gRNA expressed from different promoters (Figure 1E). The research produced by these combined multiplex targeting systems has shown the potential to identify context-specific genetic interactions, candidate combinatorial drug treatments and potential drug targets [40–47], but library design and construction is highly complex.

Analysis of first-generation CRISPR screens was carried out with the help of methods primarily developed for the analysis of RNAi [48,49] and RNA-seq [50–53] data. However, it was noticeable that there was a need for computational methods which are considerate of sgRNA library size, design, and relevant controls, all of which are specific to CRISPR data. To address that need some groups have developed methods specific to the context of their studies, whereas other groups have developed more generic software and frameworks that are widely accepted and used by scientific community. In this review, we will provide an overview for the computational methods mainly for pooled CRISPR KO screens. Undoubtedly additional tools will be required as the experimental technology continues its rapid advances.

It is noticeable that there are common features across these methods such as the input data, which is typically a matrix populated with raw read counts, where rows are individual sgRNAs used in the library and headers are the samples and/or replicates in which the screens are carries out. An essential step in these analyses is the normalization of the raw read counts to prevent the comparisons of extreme values which could lead to the increase in false positives in the downstream analysis. The most commonly implemented normalization approaches in the methods mentioned in this review are based on mean, median and total read counts normalization factors. Another component that is accounted for in some of the methods is the number of replicates which directly relates to the modeling the distribution of sgRNAs, while some methods use the direct estimates of distribution parameters others use resampling approaches to smooth these estimates. Lastly, the hits in these screens are identified mainly by comparing the abundances of sgRNAs at the start versus the end of the experiment, or untreated versus treated samples, and where the corresponding analytical methods differ mostly is which modeling approach they use to quantify this difference in sgRNA abundance. An overview of computational methods widely used for the analysis of CRISPR-based pooled screens is presented in this section and a brief description for several others can be found in Table 1. The overview for each method includes the purpose of its design, brief explanation of its computational and mathematical design, as well as the most significant targets discovered or reconfirmed by it.

Analysis of individual screens
MAGeCK

The purpose of MAGeCK, the Model-based Analysis of Genome-wide CRISPR/Cas9 KOs [54], is to prioritize sgRNAs, genes and pathways in genome-scale CRISPR/Cas9 KO screens across different experimental conditions. The various MAGeCK algorithms share a few main steps which start with median normalizing the raw read counts corresponding to sgRNAs. Afterwards, modeling the mean-variance is used to capture the relationship of mean and variance in replicates and a negative binomial (NB) is used to test whether sgRNA abundance differs significantly between treatments (where treatments could be perturbagens or KO time points) and controls (untreated or initial state, T0). This modeling is similar to what has been used for evaluating

Table 1 Non web-based algorithms/methods for analysis of pooled CRISPR screens

Algorithm name	Description	Language
MAGeCK [54–56]	Negative binomial model — based analysis of genome-wide CRISPR–Cas9 KO screens for prioritizing sgRNAs, genes and pathways.	Python, R
HiTSelect [57]	Uses Poisson distribution to evaluate sgRNAs and stochastic multiobjective ranking method to generate gene-level statistics.	Matlab
ScreenBEAM [58]	Bayesian hierarchical (multilevel) model to directly assess gene-level activity from all relevant measurements.	R
STARS [12]	Gene-ranking algorithm for genetic perturbation screens — gene scores are computed using the probability mass function of a binomial distribution.	Python
BAGEL [59,60]	Bayesian analysis for identifying essential genes from pooled screens, based on core essential and nonessential gene sets.	Python
CaRpools [61]	A pipeline for end-to-end analysis of pooled CRISPR/Cas9 screening data. Including in-depth analysis of screening quality and sgRNA phenotypes.	R
CasTLE [62]	Maximum likelihood estimator and empirical Bayesian framework to account for multiple sources of variability, including reagent efficacy and off-target effects for the analysis of large-scale genomic perturbation screens.	Python
CERES [5]	A method to estimate gene dependency from essentiality screens while computationally correcting the copy number effect, therefore enabling unbiased interpretation of gene dependency at all levels of copy number.	R
ENCoRE [63]	Workflow for NGS to CRISPR gene results.	Java
PBNPA [64]	Permutation-based non-parametric analysis, which computes P-values at the gene level by permuting sgRNA labels, therefore avoids restrictive distributional assumptions.	R
CRISPhieRmix [65]	Broad-tailed null distribution is fit using negative control sgRNAs. Then, a mixture distribution is fit on all sgRNAs, ignoring gene identities. Lastly, using the mixture distribution the false discovery rate for each gene is calculated.	R
CB2 [66]	Beta-binomial model with a modified Student's t-test to measure differences in sgRNA levels, followed by Fisher's combined probability test to estimate the gene-level significance.	R
JACKS [67]	Bayesian method that jointly analyzes screens performed with the same library and assigns a gene P-value based on empirically derived null distribution based on essentiality scores in a known set of negative control genes.	Python
DrugZ [68]	Identifies synergistic and suppressor drug-gene interactions from CRISPR-based chemogenetic screens.	Python
Gscreend [69]	Mixture of a parametric null distribution is used to calculate P-value for every sgRNA, and robust rank aggregation (RRA) algorithm is used to aggregate and score the data on gene-level.	R
CRISPRcleanR [70]	Unsupervised copy number correction of gene-independent responses in genome wide CRISPR KO screens based on circular binary segmentation algorithm.	Python, R
CRISPy [71]	Supervised copy number correction of gene-independent effects, which uses Gaussian processes regression to model non-linear effects between the segment copy number ratio and CRISPR fold changes.	R

differential RNA-seq [50–53]. The sgRNAs-level statistics are calculated using the learned NB mean-variance model. The genes targeted by sgRNAs that are ranked consistently higher (by significance) using robust rank aggregation (RRA) [72], are considered the essential genes. The RRA algorithm assumes that if a gene has no effect on selection, then sgRNAs targeting that gene should be uniformly distributed across the ranked list of all the sgRNAs. To rank the genes, RRA compares the skew in rankings according to the uniform null model and prioritizes genes whose sgRNAs' rankings are consistently higher than expected. Lastly, to identify the enriched pathways the same RRA algorithm is applied to the ranked list of genes. In addition to the original MAGeCK algorithm, MAGeCK-VISPR [56] and MAGeCKFlute [55] are the more recent and updated versions of this algorithm. MAGeCK-VISPR is a workflow that includes a comprehensive quality control (QC), analysis, and visualization workflow for CRISPR screens. This workflow defines a set of QC measures to assess the

quality of an experiment and includes a maximum-likelihood algorithm (MLA) (in addition to the original MAGeCK that scores genes based on RRA algorithm) to call essential genes simultaneously under different experimental conditions. The MLA uses a generalized linear model to deconvolute different effects and employs expectation-maximization to iteratively estimate sgRNA KO efficiency and gene essentiality. MAGeCKFlute is an integration of MAGeCK and MAGeCK-VISPR algorithms, with the addition of some QC functions, batch effect removal, copy-number bias correction, and downstream functional enrichment analysis for CRISPR screens. Additionally, it is worth to note that the MAGeCK frameworks include the MAGeCKcount function, which maps the raw FASTQ data to reference library file and count the reads for each sgRNA, which are used for the downstream analysis. Upon its initial development in 2014, MAGeCK has aided in identifying known and novel biologically interesting essential genes and pathways, including EGFR in vemurafenib-treated A375 cells carrying a BRAF mutation, cell-type specific essential genes including BCR and ABL1 in KBM7 cells with a BCR–ABL fusion, and an insulin like growth factor IGF1R in HL-60 cells.

STARS

STARS [12] is an algorithm for ranking genes based on their essentiality in genetic perturbation screens. The algorithm takes a list of ranked perturbations (i.e. sgRNAs) as an input. The gene score is computed using the probability mass function of a binomial distribution of the sgRNAs. This is calculated for all sgRNAs that are above the user-defined threshold. The value of the least probable sgRNA for each gene is then assigned to the gene as the STARS score. It is required that at least two sgRNAs rank above the user-defined threshold when scoring a gene to avoid the single sgRNA hits. The null distribution is generated by performing permutation testing on the list of sgRNAs to facilitate the calculation of *P*-values, FDR (false discovery rate) and *q*-values for hit genes. STARS analysis has identified HPRT1 and NUDT5 genes to confer 6-thioguanine resistance in three different cell lines, previously validated genes driving vemurafenib resistance in A375 cells, and modifiers of interferon signaling in mouse BV2 cells.

BAGEL

BAGEL [59], or Bayesian Analysis of Gene EssentiaLity, is a supervised learning method for analyzing gene KO screens, trained with gold standard [73] essential and nonessential genes. The initial release of BAGEL (Hart 2016) offered significantly greater sensitivity than the other methods, while BAGEL2 [60] reduced runtime due to the computational optimizations. As an input BAGEL takes the tab-delimited file of sgRNA read counts and reference essential and nonessential gene lists. The first step of BAGEL algorithm is to estimate the distribution of fold changes of sgRNAs targeting all genes in the essential and nonessential training sets. Afterwards, it estimates the likelihood that the observed fold changes for sgRNAs targeting the test gene (withheld from the first step) are drawn from either the essential or nonessential training distributions. The final score for each gene is a log Bayes Factor (BF) which is calculated as the ratio of probabilities that the fold change of the tested gene is drawn from the distribution of essential training set or from the distribution of nonessential training set. The BF score can be calculated for each replicate or at each timepoint if the screen has multiple timepoints. To evaluate the screen performance BAGEL calculates the precision-recall (PR) curves, using the core essential and nonessential gene lists as the test set. The bootstrap resampling of genes in the training set is implemented for validation purposes. BAGEL2 [60] is an improved version of BAGEL, which includes a modeling approach that offers a greater dynamic range of BF, enabling detection of tumor suppressor genes, a multi-target correction that reduces false positives from off-target CRISPR sgRNA, and the implementation of a cross-validation strategy that improves performance ~10× in comparison with the previous bootstrap sampling approach. In BAGEL2, all genes in the experiment are resampled into training and test sets by 10-fold cross-validation or bootstrapping to calculate the BF. A kernel density estimation is used in each iteration of this sampling to generate the fold change distributions for essential and nonessential genes, using all sgRNAs targeting the control essential or nonessential genes in the resample. The log ratio of essential and nonessential distributions is then taken to generate the log BF on a sgRNA level. A variation of BAGEL, which normalizes gene-level BFs based on the number of gRNA targeting each gene, is used in the Project Score dataset [6].

casTLE

Despite the dominance of CRISPR-based screening platform, Morgens et al. developed a Cas9 high-throughput maximum likelihood estimator (casTLE) with the goal of leveraging the data from both RNAi and CRISPR–Cas9 screens. casTLE combines measurements from multiple targeting reagents (e.g. sgRNAs) and the

phenotypes of negative controls to estimate the maximum effect size effect for each gene and an associated log-likelihood ratio. In the original study, the casTLE score is presented as twice the log-likelihood ratio. casTLE can be applied on single replicates from different screen types, as well as on combined diverse data types while separately considering experimental noise and variability caused by heterogeneous reagents. The authors of casTLE report that combining the data from RNAi and CRISPR screens led to the improvement in performance in terms of the percent of gold standard essential genes (>85%) at a certain FDR threshold (in this case ~1% FDR) with an AUC of 0.98 and the identification of ~4500 nonessential genes. They suggest that this is explained by the RNAi and CRISPR screens revealing different aspects of biology, which is also shown in the low correlation of individual results from the two screens types. To validate casTLE's performance for RNAi and CRISPR screens, casTLE was used to re-analyze data from a few published screens. For each of those screens, casTLE produced results consistent with previous findings, additionally identifying the positive regulators of LPS-induced TNF expression in primary mouse dendritic cells [62].

PBNPA

Permutation-based non-parametric analysis (PBNPA) [64] is designed to analyze CRISPR data but can be used to analyze other genetic screens as well. PBNPA computes *P*-values on a gene level by permuting sgRNA labels, thus avoiding restrictive distributional assumptions. To achieve this end-result, PBNPA is designed to accomplish several steps which start with normalizing the raw read counts by multiplying a factor of mean, which makes total read counts in each condition equal. For each sgRNA the natural logarithm gold change is calculated and the median of those is then used as R score for each gene. Gene labels are randomly permuted to form a null distribution of R, which is used to calculate a *P*-value for a gene if it negatively or positively selected gene. After getting the *P*-values for all genes, the small set of genes with *P*-values smaller than a threshold is removed, and the prior permutation process is repeated to get the null distribution with the significant genes removed. Then, the *P*-values are updated for each gene and lastly, Benjamini–Hochberg procedure is used to calculate the FDR values. The authors highlight that using the median log fold change of sgRNAs targeting a gene as the R score for that gene make it more robust against any outliers and influences from potential off-target effects. The authors of PBNPA have used Wang et al. [74] study to validate the performance of their algorithm showing that PNBPA has fewer falsely identifies genes compared with MAGeCK.

Joint analysis of a panel of screens
CERES

CERES is a method to estimate gene dependency from CRISPR KO screens while computationally correcting the copy number effect, therefore enabling unbiased interpretation of gene dependency at all levels of copy number (originally developed based on screens carried out in 342 cancer cell lines) [4]. CERES models the measured sgRNA phenotypic effect or depletion as a sum of gene KO and copy number effects. Additionally, given that the CERES was developed with regards to the 342 cell lines, the total number of cell lines is another determining constant in CERES modeling in addition to the previously three mentioned effects. In CERES, the effect of copy number is modeled by a linear spline, while the gene KO effect is a sum of cell line dependent and independent effects. The sum of these effects is then scaled by a guide activity score, restricted to values between 0 and 1, with the purpose of isolating and mitigating the influence of low-quality sgRNAs. The term accounting for noise, quantification of sgRNA abundance in the reference pool, is also included in the inference model. To infer the mentioned effects and terms, least squares regression is used to model the observed data. CERES gene-level scores are the inferred gene KO effects which are scaled per cell line such that scores of 0 and −1 represent the median effects or nonessential and core essential genes, respectively.

JACKS

The Joint Analysis of CRISPR Knockout Screens (JACKS) is a Bayesian method that analyzes panels of screens performed with the same guide RNA library [67]. It is designed to address the signal variability from different sgRNAs that target the same gene, which confounds gene effect estimation and dictates large experiment sizes. JACKS defines the observed $\log_2$ fold change of an sgRNA as the mean across median-normalized replicate measurements, which is modeled as a Gaussian distribution based on condition-dependent gene effect, condition-independent sgRNA efficacy and the precision parameter of $\log_2$ fold change which uses a non-parametric approach to assign an empirical Bayes prior that accounts for mean-dependent variability of the

log$_2$ fold change within the replicate measurements. To overcome the poor direct estimates of these variances in usual cases where the CRISPR screens are carried out with 2–3 replicates, JACKS computes a smoothed mean-dependent estimate of this empirical variance based on all sgRNAs in each condition and then assigns the priors on the precision parameter. To infer the posterior distributions of the three mentioned parameters (condition-dependent gene effect, condition-independent guide effect, and precision) variational inference is used. *P*-values on gene-level are derived from a non-parametric distribution of condition-independent guide effects computed by running JACKS inference on a number of sgRNAs (from the negative controls set, i.e. guide targeting non-essential genes or cutting only in noncoding regions) targeting randomly selected test genes.

Conclusions and future directions

As illustrated above, a variety of computational tools are available for analysis of CRISPR screens. These tools use a variety of approaches to identify genes undergoing negative selection in a pooled library screen. Only a subset, however, take into account the quality of the underlying data. The reagents and enzymes behind CRISPR screens have evolved tremendously over the last half-decade, and high signal to noise is now the expectation for a modern CRISPR library in a tractable model. With high data quality, choice of analytical approach is less important, as most approaches will give highly concordant results and marginal hits are likely less interesting because of their lower KO fitness defects. Quality control remains a critical first step.

Copy number effect influences the measurement of the gene deletion effect on cell survival and it can be detected even among low-level copy number amplifications and deletions [5]. Therefore, correcting for copy number [75] is an important step in processing and analyzing CRISPR screens. CRISPRcleanR [70] and CRISPy [71] are two single sample-based methods for correcting the copy number effect, contrary to CERES which performs correction across multiple samples. CRISPRclearR corrects the gene-independent responses in genome-wide CRISPR KO screens in an unsupervised manner, whereas CRISPy corrects it in a supervised way using segment level copy number ratios. CRISPRcleanR is library agnostic, and as results outputs sgRNA fold changes and normalized read counts, which makes it compatible with downstream analysis tools, such as BAGEL or MAGeCK, for hit calling and interpretation of gene KO phenotypic effects.

Going forward, CRISPR editing will continue to be a fast-moving field, with new perturbation modalities demanding development of new analytical approaches. In this context, it is worth highlighting the presence of Cas12a-based screening platforms. The Cas12a nuclease, which can process multiple gRNA from a single poly-cistronic transcript, offers an attractive alternative to Cas9 for multiplex screening, which facilitates the large-scale investigation of genetic interactions in mammalian cells (Figure 1F). In the DeWeirdt et al. study, which optimized libraries for the engineered enCas12a variant, the group also screened for synthetic lethality in two cancer cell lines (OVCAR8 and A375) and discovered previously unreported interaction between MARCH5 and WSB2 [76]. Dede et al. has utilized the enCas12a platform to investigate the functional buffering among ~400 candidate paralog pairs in three cell lines. The authors observed 24 synthetic lethal paralog pairs that were previously undetected by monogenic KO screens. The Moffat group took advantage of the Cas12a system in a different fashion — combing Cas9 and Cas12a to create a hybrid Cas platform, CHyMErA, to evaluate a set of 672 human paralog pairs, and explore chemogenetic interactions in the mTOR pathway [77]. These and other similar studies indicate the need for robust methods specifically designed to analyze and quantify the mutual phenotypic effects of simultaneous digenic or higher-order KOs. There are several computational methods for scoring digenic KOs CRISPR screens [40,41,78–81], however only two of them [80,81] are generalized and available as R packages.

Summary

- The advent of CRISPR technology has revolutionized the cancer biology by providing insights about cancer treatment strategies through the discovery of essential genes for drug targets, identification of metastatic regulators, drug resistance mechanisms, immunotherapy targets, and synthetic lethality.

- There exists a number of bioinformatics tools and methods for the analysis of CRISPR screens based on different mathematical models and being continuously optimized.

- As CRISPR versatility expands and technologies continue to develop, we foresee the need for the development of methods specifically designed to robustly quantify these novelties and provide the functional characterizations for genetic systems of interest.

Abbreviations
BF, Bayes Factor; FDR, false discovery rate; JACKS, Joint Analysis of CRISPR Knockout Screens; KO, knockout; MLA, maximum-likelihood algorithm; NB, negative binomial; PBNPA, Permutation-based non-parametric analysis; QC, quality control; RRA, robust rank aggregation.

References

1 Wright, A.V., Nuñez, J.K. and Doudna, J.A. (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29–44 https://doi.org/10.1016/j.cell.2015.12.035

2 Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E. et al. (2013) RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 https://doi.org/10.1126/science.1232033

3 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 https://doi.org/10.1126/science.1231143

4 Kazi, T.A. and Biswas, S.R. (2021) CRISPR/dcas system as the modulator of gene expression. *Prog. Mol. Biol. Transl. Sci.* **178**, 99–122 https://doi.org/10.1016/bs.pmbts.2020.12.002

5 Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H. et al. (2017) Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 https://doi.org/10.1038/ng.3984

6 Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G. et al. (2019) Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nature* **568**, 511–516 https://doi.org/10.1038/s41586-019-1103-9

7 Dempster, J.M., Pacini, C., Pantel, S., Behan, F.M., Green, T., Krill-Burger, J. et al. (2019) Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. *Nat. Commun.* **10**, 5817 https://doi.org/10.1038/s41467-019-13805-y

8 Pacini, C., Dempster, J.M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E. et al. (2021) Integrated cross-study datasets of genetic dependencies in cancer. *Nat. Commun.* **12**, 1661 https://doi.org/10.1038/s41467-021-21898-7

9 Boehm, J.S., Garnett, M.J., Adams, D.J., Francies, H.E., Golub, T.R., Hahn, W.C. et al. (2021) Cancer research needs a better map. *Nature* **589**, 514–516 https://doi.org/10.1038/d41586-021-00182-0

10 Broad Institute. (2019) DepMap: The Cancer Dependency Map. Available from: https://depmap.org/portal/

11 Sanger Institute. (2019) Cancer Dependency Map. Available from: https://depmap.sanger.ac.uk

12 Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F. et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 https://doi.org/10.1038/nbt.3437

13 Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M. et al. (2017) Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3(Bethesda)* **7**, 2719–2727 https://doi.org/10.1534/g3.117.041277

14 Gonçalves, E., Thomas, M., Behan, F.M., Picco, G., Pacini, C., Allen, F. et al. (2021) Minimal genome-wide human CRISPR-Cas9 library. *Genome Biol.* **22**, 40 https://doi.org/10.1186/s13059-021-02268-4

15 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 https://doi.org/10.1126/science.1246981

16 Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S. et al. (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 https://doi.org/10.1126/science.1247005

17 Deans, R.M., Morgens, D.W., Ökesli, A., Pillay, S., Horlbeck, M.A., Kampmann, M. et al. (2016) Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. *Nat. Chem. Biol.* **12**, 361–366 https://doi.org/10.1038/nchembio.2050

18 Estoppey, D., Lee, C.M., Janoschke, M., Lee, B.H., Wan, K.F., Dong, H. et al. (2017) The natural product cavinafungin selectively interferes with zika and dengue virus replication by inhibition of the host signal peptidase. *Cell Rep.* **19**, 451–460 https://doi.org/10.1016/j.celrep.2017.03.071

19 Estoppey, D., Hewett, J.W., Guy, C.T., Harrington, E., Thomas, J.R., Schirle, M. et al. (2017) Identification of a novel NAMPT inhibitor by CRISPR/Cas9 chemogenomic profiling in mammalian cells. *Sci. Rep.* **7**, 42728 https://doi.org/10.1038/srep42728

20 Zimmermann, M., Murina, O., Reijns, M.A.M., Agathanggelou, A., Challis, R., Tarnauskaitė, Ž et al. (2018) CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature* **559**, 285–289 https://doi.org/10.1038/s41586-018-0291-z

21 Noordermeer, S.M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S.J., Ling, A.K. et al. (2018) The shieldin complex mediates 53BP1-dependent DNA repair. *Nature* **560**, 117–121 https://doi.org/10.1038/s41586-018-0340-7

22 Hustedt, N., Álvarez-Quilón, A., McEwan, A., Yuan, J.Y., Cho, T., Koob, L. et al. (2019) A consensus set of genetic vulnerabilities to ATR inhibition. *Open Biol.* **9**, 190156 https://doi.org/10.1098/rsob.190156

23 Olivieri, M., Cho, T., Álvarez-Quilón, A., Li, K., Schellenberg, M.J., Zimmermann, M. et al. (2020) A genetic map of the response to DNA damage in human cells. *Cell* **182**, 481–496.e21 https://doi.org/10.1016/j.cell.2020.05.040

24 MacLeod, G., Bozek, D.A., Rajakulendran, N., Monteiro, V., Ahmadi, M., Steinhart, Z. et al. (2019) Genome-wide CRISPR-Cas9 screens expose genetic vulnerabilities and mechanisms of temozolomide sensitivity in glioblastoma stem cells. *Cell Rep.* **27**, 971–986.e9 https://doi.org/10.1016/j.celrep.2019.03.047

25 Yoshimoto, K., Mizoguchi, M., Hata, N., Murata, H., Hatae, R., Amano, T. et al. (2012) Complex DNA repair pathways as possible therapeutic targets to overcome temozolomide resistance in glioblastoma. *Front. Oncol.* **2**, 186 https://doi.org/10.3389/fonc.2012.00186

26 Wang, C., Wang, G., Feng, X., Shepherd, P., Zhang, J., Tang, M. et al. (2018) Genome-wide CRISPR screens reveal synthetic lethality of RNASEH2 deficiency and ATR inhibition. *Oncogene* **38**, 2451–2463 https://doi.org/10.1038/s41388-018-0606-4

27 Su, D., Feng, X., Colic, M., Wang, Y., Zhang, C., Wang, C. et al. (2020) CRISPR/CAS9-based DNA damage response screens reveal gene-drug interactions. *DNA Repair* **87**, 102803 https://doi.org/10.1016/j.dnarep.2020.102803

28 Colic, M. and Hart, T. (2019) Chemogenetic interactions in human cancer cells. *Comput. Struct. Biotechnol. J.* **17**, 1318–1325 https://doi.org/10.1016/j.csbj.2019.09.006

29 Dai, M., Yan, G., Wang, N., Daliah, G., Edick, A.M., Poulet, S. et al. (2021) *In vivo* genome-wide CRISPR screen reveals breast cancer vulnerabilities and synergistic mTOR/Hippo targeted combination therapy. *Nat. Commun.* **12**, 3055 https://doi.org/10.1038/s41467-021-23316-4

30 Griffin, G.K., Wu, J., Iracheta-Vellve, A., Patti, J.C., Hsu, J., Davis, T. et al. (2021) Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. *Nature* **595**, 309–314 https://doi.org/10.1038/s41586-021-03520-4

31 van der Weyden, L., Harle, V., Turner, G., Offord, V., Iyer, V., Droop, A. et al. (2021) CRISPR activation screen in mice identifies novel membrane proteins enhancing pulmonary metastatic colonisation. *Commun. Biol.* **4**, 395 https://doi.org/10.1038/s42003-021-01912-w

32 Gautron, A., Bachelot, L., Aubry, M., Leclerc, D., Quéméner, A.M., Corre, S. et al. (2021) CRISPR screens identify tumor-promoting genes conferring melanoma cell plasticity and resistance. *EMBO Mol. Med.* **13**, e13466 https://doi.org/10.15252/emmm.202013466

33 Manguso, R.T., Pope, H.W., Zimmer, M.D., Brown, F.D., Yates, K.B., Miller, B.C. et al. (2017) *In vivo* CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413–418 https://doi.org/10.1038/nature23270

34 Dong, M.B., Wang, G., Chow, R.D., Ye, L., Zhu, L., Dai, X. et al. (2019) Systematic immunotherapy target discovery using genome-scale *in vivo* CRISPR screens in CD8T cells. *Cell* **178**, 1189–1204.e23 https://doi.org/10.1016/j.cell.2019.07.044

35 Bajaj, J., Hamilton, M., Shima, Y., Chambers, K., Spinler, K., Van Nostrand, E.L. et al. (2020) An *in vivo* genome-wide CRISPR screen identifies the RNA-binding protein Staufen2 as a key regulator of myeloid leukemia. *Nat. Cancer* **1**, 410–422 https://doi.org/10.1038/s43018-020-0054-2

36 Lawson, K.A., Sousa, C.M., Zhang, X., Kim, E., Akthar, R., Caumanns, J.J. et al. (2020) Functional genomic landscape of cancer-intrinsic evasion of killing by T cells. *Nature* **586**, 120–126 https://doi.org/10.1038/s41586-020-2746-2

37 Hou, J., Wang, Y., Shi, L., Chen, Y., Xu, C., Saeedi, A. et al. (2021) Integrating genome-wide CRISPR immune screen with multi-omic clinical data reveals distinct classes of tumor intrinsic immune regulators. *J. Immunother. Cancer* **9**, e001819 https://doi.org/10.1136/jitc-2020-001819

38 Mair, B., Aldridge, P.M., Atwal, R.S., Philpott, D., Zhang, M., Masud, S.N. et al. (2019) High-throughput genome-wide phenotypic screening via immunomagnetic cell sorting. *Nat. Biomed. Eng.* **3**, 796–805 https://doi.org/10.1038/s41551-019-0454-8

39 Buquicchio, F.A. and Satpathy, A.T. (2021) Interrogating immune cells and cancer with CRISPR-Cas9. *Trends Immunol.* **42**, 432–446 https://doi.org/10.1016/j.it.2021.03.003

40 Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A. et al. (2017) Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions. *Nat. Methods* **14**, 573–576 https://doi.org/10.1038/nmeth.4225

41 Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W. et al. (2018) Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189 https://doi.org/10.1038/nbt.4048

42 Boettcher, M., Tian, R., Blau, J.A., Markegard, E., Wagner, R.T., Wu, D. et al. (2018) Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* **36**, 170–178 https://doi.org/10.1038/nbt.4062

43 DeWeirdt, P.C., Sangree, A.K., Hanna, R.E. et al. (2020) Genetic screens in isogenic mammalian cell lines without single cell cloning. *Nat. Commun.* **11**, 752 https://doi.org/10.1038/s41467-020-14620-6

44 Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M. et al. (2016) Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. *Proc. Natl Acad. Sci. U.S.A.* **113**, 2544–2549 https://doi.org/10.1073/pnas.1517883113

45 Diehl, V., Wegner, M., Grumati, P., Husnjak, K., Schaubeck, S., Gubas, A. et al. (2021) Minimized combinatorial CRISPR screens identify genetic interactions in autophagy. *Nucleic Acids Res.* **49**, 5684–5704 https://doi.org/10.1093/nar/gkab309

46 Thompson, N.A., Ranzani, M., van der Weyden, L., Iyer, V., Offord, V., Droop, A. et al. (2021 Dec) Combinatorial CRISPR screen identifies fitness effects of gene paralogues. *Nat. Commun.* **12**, 1302 https://doi.org/10.1038/s41467-021-21478-9

47 Phoebe C.R. Parrish, James D. Thomas, Austin M. Gabel, Shriya Kamlapurkar, Robert K. Bradley, Alice H. Berger (2021) Discovery of synthetic lethal and tumor suppressor paralog pairs in the human genome. *Cell Reports*, **36**, https://doi.org/10.1016/j.celrep.2021.109597

48 König, R., Chiang, C., Tu, B.P., Yan, S.F., DeJesus, P.D., Romero, A. et al. (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* **4**, 847–849 https://doi.org/10.1038/nmeth1089

49 Luo, B., Cheung, H.W., Subramanian, A., Sharifnia, T., Okamoto, M., Yang, X. et al. (2008) Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. U.S.A.* **105**, 20380–5 https://doi.org/10.1073/pnas.0810485105

50 Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) Edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 https://doi.org/10.1093/bioinformatics/btp616

51 Dai, Z., Sheridan, J.M., Gearing, L.J., Moore, D.L., Su, S., Wormald, S. et al. (2014) Edger: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res.* **3**, 95 https://doi.org/10.12688/f1000research.3928.2

52 McCarthy, D.J., Chen, Y. and Smyth, G.K. (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 https://doi.org/10.1093/nar/gks042

53 Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 https://doi.org/10.1186/s13059-014-0550-8

54 Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F. et al. (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 https://doi.org/10.1186/s13059-014-0554-4

55 Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.-H., Wu, A. et al. (2019) Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. *Nat. Protoc.* **14**, 756–780 https://doi.org/10.1038/s41596-018-0113-7

56 Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J.S. et al. (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* **16**, 281 https://doi.org/10.1186/s13059-015-0843-6

57 Diaz, A.A., Qin, H., Ramalho-Santos, M. and Song, J.S. (2015) HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res.* **43**, e16–e16 https://doi.org/10.1093/nar/gku1197

58 Yu Jiyang, Silva Jose, Califano Andrea (2016) ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics*, **32**, 260–267 https://doi.org/10.1093/bioinformatics/btv556

59 Hart, T. and Moffat, J. (2016) BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164 https://doi.org/10.1186/s12859-016-1015-8

60 Kim, E. and Hart, T. (2021) Improved analysis of CRISPR fitness screens and reduced off-target effects with the BAGEL2 gene essentiality classifier. *Genome Med.* **13**, 2 https://doi.org/10.1186/s13073-020-00809-3

61 Winter, J., Breinig, M., Heigwer, F., Brügemann, D., Leible, S., Pelz, O. et al. (2016) Carpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. *Bioinformatics* **32**, 632–634 https://doi.org/10.1093/bioinformatics/btv617

62 Morgens, D.W., Deans, R.M., Li, A. and Bassik, M.C. (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 https://doi.org/10.1038/nbt.3567

63 Trümbach, D., Pfeiffer, S., Poppe, M., Scherb, H., Doll, S., Wurst, W. et al. (2017) ENCoRE: an efficient software for CRISPR screens identifies new players in extrinsic apoptosis. *BMC Genomics* **18**, 905 https://doi.org/10.1186/s12864-017-4285-2

64 Jia, G., Wang, X. and Xiao, G. (2017) A permutation-based non-parametric analysis of CRISPR screen data. *BMC Genomics* **18**, 545 https://doi.org/10.1186/s12864-017-3938-5

65 Daley, T.P., Lin, Z., Lin, X., Liu, Y., Wong, W.H. and Qi, L.S. (2018) CRISPhiermix: a hierarchical mixture model for CRISPR pooled screens. *Genome Biol.* **19**, 159 https://doi.org/10.1186/s13059-018-1538-6

66 Jeong, H.-H., Kim, S.Y., Rousseaux, M.W.C., Zoghbi, H.Y. and Liu, Z. (2019) Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives. *Genome Res.* **29**, 999–1008 https://doi.org/10.1101/gr.245571.118

67 Allen, F., Behan, F., Khodak, A., Iorio, F., Yusa, K., Garnett, M. et al. (2019) JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome Res.* **29**, 464–471 https://doi.org/10.1101/gr.238923.118

68 Colic, M., Wang, G., Zimmermann, M., Mascall, K., McLaughlin, M., Bertolet, L. et al. (2019) Identifying chemogenetic interactions from CRISPR screens with drugZ. *Genome Med.* **11**, 52 https://doi.org/10.1186/s13073-019-0665-3

69 Imkeller, K., Ambrosi, G., Boutros, M. et al. (2020) gscreend: modelling asymmetric count ratios in CRISPR screens to decrease experiment size and improve phenotype detection. *Genome Biol* **21**, 53 https://doi.org/10.1186/s13059-020-1939-1

70 Iorio, F., Behan, F.M., Gonçalves, E., Bhosle, S.G., Chen, E., Shepherd, R. et al. (2018) Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. *BMC Genomics* **19**, 604 https://doi.org/10.1186/s12864-018-4989-y

71 Gonçalves, E., Behan, F.M., Louzada, S., Arnol, D., Stronach, E.A., Yang, F. et al. (2019) Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biol.* **20**, 27 https://doi.org/10.1186/s13059-019-1637-z

72 Kolde, R., Laur, S., Adler, P. and Vilo, J. (2012) Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573–580 https://doi.org/10.1093/bioinformatics/btr709

73 Hart, T., Brown, K.R., Sircoulomb, F., Rottapel, R. and Moffat, J. (2014) Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733 https://doi.org/10.15252/msb.20145216

74 Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J. et al. (2015) Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 https://doi.org/10.1126/science.aac7041

75 Shen, J.P. and Ideker, T. (2017) Correcting CRISPR for copy number. *Nat. Genet.* **49**, 1674–1675 https://doi.org/10.1038/ng.3994

76 DeWeirdt, P.C., Sanson, K.R., Sangree, A.K., Hegde, M., Hanna, R.E., Feeley, M.N. et al. (2021) Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat. Biotechnol.* **39**, 94–104 https://doi.org/10.1038/s41587-020-0600-6

77 Gonatopoulos-Pournatzis, T., Aregger, M., Brown, K.R., Farhangmehr, S., Braunschweig, U., Ward, H.N. et al. (2020) Genetic interaction mapping and exon-resolution functional genomics with a hybrid Cas9–Cas12a platform. *Nat. Biotechnol.* **38**, 638–648 https://doi.org/10.1038/s41587-020-0437-z

78 Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A. and Bassik, M.C. (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* **35**, 463–474 https://doi.org/10.1038/nbt.3834

79 Horlbeck, M.A., Xu, A., Wang, M., Bennett, N.K., Park, C.Y., Bogdanoff, D. et al. (2018) Mapping the genetic landscape of human cells. *Cell* **174**, 953–967.e22 https://doi.org/10.1016/j.cell.2018.06.010

80 Zamanighomi, M., Jain, S.S., Ito, T., Pal, D., Daley, T.P. and Sellers, W.R. (2019) GEMINI: a variational Bayesian approach to identify genetic interactions from combinatorial CRISPR screens. *Genome Biol.* **20**, 137 https://doi.org/10.1186/s13059-019-1745-9

81 Ward, H.N., Aregger, M., Gonatopoulos-Pournatzis, T., Billmann, M., Ohsumi, T.K., Brown, K.R. et al. (2021) Analysis of combinatorial CRISPR screens with the Orthrus scoring pipeline. *Nat. Protoc.* **16**, 4766–4798 https://doi.org/10.1038/s41596-021-00596-0