Trends in Biotechnology

CellPress
REVIEWS

Review

RNAi/CRISPR Screens: from a Pool to a Valid Hit

Anne Schuster,[1] Hélène Erasimus,[1] Sabrina Fritah,[1] Petr V. Nazarov,[2] Eric van Dyck,[1] Simone P. Niclou ,[1,3,4,*] and Anna Golebiewska[1,4,*,@]

High-throughput genetic screens interfering with gene expression are invaluable tools to identify gene function and phenotype-to-genotype interactions. Implementing such screens in the laboratory is challenging, and the choice between currently available technologies based on RNAi and CRISPR/Cas9 (CRISPR-associated protein 9) is not trivial. Identifying reliable candidate hits requires a streamlined experimental setup adjusted to the specific biological question. Here, we provide a critical assessment of the various RNAi/CRISPR approaches to pooled screens and discuss their advantages and pitfalls. We specify a set of best practices for key parameters enabling a reproducible screen and provide a detailed overview of analysis methods and repositories for identifying the best candidate gene hits.

Highlights

Pooled genetic screens based on RNAi and CRISPR technologies are a powerful approach for high-throughput interrogation of loss- or gain-of-function and phenotype-to-genotype correlations.

Several CRISPR technologies are applicable for pooled screens, allowing for a wide range of genetic perturbations and mutagenesis beyond classical RNAi-based gene knockdown.

Stringent experimental design, appropriate controls and careful library selection are essential to identify valid hits. Appropriate library representation throughout the screening procedure is key to avoid false positives/negatives.

Different bioinformatics pipelines can be applied to data analysis, and their combination may lead to increased specificity of selected hits.

Principle of High-Throughput Genetic Screening

Gene perturbation technologies such as RNAi and CRISPR/Cas9 (**CRISPR-associated protein 9**; see Glossary) have evolved as powerful tools to interrogate gene functions. The optimization of such technologies paved the way for large-scale screens, addressing functional phenotypes in a high-throughput setting. Direct gene knockout [CRISPR knockout (CRISPRko)] or knockdown [RNAi, CRISPR interference (CRISPRi)] elucidates gene function based on interference with gene expression (Box 1). Essential genes are defined by the loss of fitness to produce a phenotype of interest or overall viability, which can be addressed by **loss-of-function screens** [1]. Although, by definition, true essentiality requires total gene inactivation, a gene knockdown or heterozygous knockout can also reveal important functions, including when the null phenotype is lethal. **Gain-of-function screens**, although primarily performed by exogenous gene overexpression [open reading frame (ORF) screens], nowadays rely mostly on the more recent CRISPR activation (CRISPRa) technology. RNAi technology is mediated by **siRNAs**, frequently produced from **small hairpin RNA** (shRNA) precursors, whereas CRISPR technologies rely on **single guide RNAs** (sgRNAs) (Box 1).

Genetic pooled screens are based on the simultaneous targeting of a large number of genes or **regulatory elements** in a pooled, single batch manner, under conditions of one gene perturbation per cell. The entire cellular population is then subjected to selective pressure, the nature of which is dictated by the underlying biological question, leading to the enrichment or depletion of cells in response to the applied environment. High-throughput genetic screens allow the unbiased dissection of direct phenotype-to-genotype relationships in a systematic manner and the identification of key dependencies of a cell or organism. Genetic screens not only complement studies of rare gene variants in the human and gene knockout studies in mice [1], but also allow the discovery of novel druggable targets.

[1]NorLux Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health, Luxembourg
[2]Genomics and Proteomics Research Unit, Department of Oncology, Luxembourg Institute of Health, Luxembourg
[3]KG Jebsen Brain Tumour Research Center, Department of Biomedicine, University of Bergen, Bergen, Norway
[4]Co-senior authors
[@]Twitters: @LIH_Luxembourg, @AnGolebiewska

*Correspondence:
simone.niclou@lih.lu (S.P. Niclou) and anna.golebiewska@lih.lu (A. Golebiewska).

 https://doi.org/10.1016/j.tibtech.2018.08.002

Designing and performing a genetic screen requires detailed knowledge in molecular biology, sequencing technology as well as biostatistics and high-throughput data analysis, which can be challenging for non-experts. Here, we provide a comprehensive review for pooled screens, focusing on the advantages and pitfalls of available technologies, and highlighting crucial aspects of study design, protocol steps, and analysis. Considering specific biological applications, we conclude with the perspectives and challenges of this fast-developing field.

Types of Screens

Genetic screens are designed in two ways: (i) 'plate' ('arrayed'; 'single well') format where siRNAs, shRNAs or sgRNAs are separately introduced to cells; that is, in distinct culture wells; or (ii) 'pooled' or 'barcode' format, where a mixture of sh/sgRNAs is applied simultaneously, resulting in the specific barcoding of individual cells. Pooled screens are generally advantageous over the plate format based on lower costs, and they are less labor intensive with no requirement for high-throughput robotics. Although initially low throughput, the plate format (not reviewed here) can now also be applied to high-throughput imaging [2] to screen for complex phenotypes such as cellular shape or subcellular localization. These screens may be key to identification of genes involved in cell–cell communication and paracrine signaling, which may not be detected in pooled formats where perturbed cells may still receive missing molecules from surrounding unperturbed cells.

Contrary to the plate format, pooled screens require the use of a **library** of shRNA or sgRNA molecules targeting multiple genes. Designing such a library and fine tuning the parameters ensuring one genetic perturbation per cell are critical to the success of a pooled screen. Although most screens aim at one perturbation per cell, combinatorial screens are possible [3–8], where the functional relationship between two genes is examined; for example, in the synthetic lethality context. Although commonly carried out *in vitro*, screens can also be performed *ex vivo* and *in vivo*. *Ex vivo* screens are performed on primary cells or organoids isolated from an organism [9], whereas *in vivo* screens involve either injection of viral particles into animals [10,11] or transduction of cells *in vitro/ex vivo* before implantation [12,13].

Pooled screens rely on the physical separation of cells into subpopulations either enriched or depleted for the phenotype of interest (Figure 1). Such a separation is based on a relevant biological property, for example: protein expression identified by flow cytometry [14]; selection based on gene reporter activity [15]; or physical separation based on cell activity (e.g., migration [16]). In the case of viability screens, which aim to reveal genes involved in cell survival and proliferation, the impact of the selective pressure is monitored over time. If required by the experimental setup, for instance when screening for genes sensitizing cells to cytotoxic molecules, the cells are divided into control (vehicle-treated cells) and experimental (treated with cytotoxin) arms before applying the selective pressure.

The genes driving the phenotype of interest are identified by comparing the relative abundance of corresponding sh/sgRNAs in both fractions. Negative and positive selection can be considered, for instance, through loss of fitness (e.g., viability or drug sensitivity, inability to invade) or gain of fitness (e.g., drug resistance, tumorigenic potential), respectively. Generally, screen readout is based on identification of sh/sgRNA barcodes in a pooled cell fraction. The identification of screen **hits** is determined statistically by assessing the relative depletion ('dropout') or enrichment of sh/sgRNAs. Alternative CRISPR-based approaches allow for detection of hits in single cells, where the readout is based on single cell RNA-seq [17–20] or cell-specific unique molecular identifiers (UMIs) [21].

Glossary

Base editor: an enzyme with base-specific deaminase activity that is used in genome-editing systems as fusions to CRISPR/dCas9, to induce specific base pair substitutions (e.g., transitions and conversions) at a target site to which they are recruited.
CRISPR-associated protein 9 (Cas9): an RNA-guided DNA endonuclease that cleaves DNA to generate DSBs in a targeted manner, applied in combination with sgRNA for genome editing.
CRISPR-associated protein 13 (Cas13): RNA-guided ribonuclease that cleaves single-stranded RNA in a targeted manner.
Functional titer: number of infectious viral particles in a preparation, expressed as infectious units per ml (IFU/ml) or transducing units per ml (TU/ml). The functional titer (TU) is determined based on the following formulas: (a) TU/ml for colony formation assay = Number of colonies/Virus volume in ml; (b) TU/ml for flow cytometric assessment = (Number of cells transduced x % fluorescent cells)/Virus volume in ml.
Gain-of-function screen: type of a functional screen involving or resulting from the activation of gene/noncoding element expression intended to rescue genetic defects or perturb cell homeostasis.
Hit: gene that is identified by the top differentially represented sh/sgRNAs.
Indel (Insertion/deletion): short insertion or deletion of nucleotides introduced during the processing of CRISPR/Cas9-induced DSBs by the NHEJ machinery, leading to amino acid insertions or deletions, or frameshifts and the possible introduction of a premature stop codon within the ORF.
Library: collection of sequences designed to target a defined number of genes, present in form of oligonucleotides, plasmids, viral particles or transduced cells. Its size is defined as the number of sh/sgRNAs targeting the genes/sequences represented in the library.
Loss-of-function screen: type of functional screen involving or resulting from the inactivation or downregulation of gene/noncoding element expression.
Multiplicity of infection (MOI): ratio between the number of functional viral particles and the number of host cells in an infection.

Gene Modulation and Editing Technologies for Pooled Screens

While genetic screens were originally based on RNAi technology, further developments of CRISPR/Cas9 systems have broadened the range of applications. Although the choice of screening technology is dictated by the underlying biological question, several technologies are often equally applicable, making it difficult to choose (Table 1). Here, we summarize the current technologies available for pooled screening, while highlighting their possible applications, advantages, and pitfalls.

RNAi

RNAi has been widely used to perform pooled screens and many results are centrally available through data resources such as Project Achilles [22] and GenomeRNAi [23]. RNAi technology is based on degradation of target protein-coding transcripts due to base-paring between RNA and siRNA (Box 1). As pooled screening requires barcoding and tracking genetic perturbations, only virus-based shRNA technologies are applied (Box 2). Although somewhat outdated, RNAi has numerous advantages for specific applications ([24], Table 1). Since the RNAi machinery is mostly cytoplasmic and acts on RNA, gene knockdowns are not biased by cell ploidy, chromatin conformation, and locus accessibility. In addition, RNAi does not require co-delivery of exogenous proteins and annotated **transcriptional start sites** (TSSs), in contrast to CRISPR-based technologies. Also, shRNA screens are based on the endogenous RNAi pathway, eliminating the need of introducing foreign sequences encoding large proteins (e.g., endonucleases, transcriptional repressors and activators, and base editors) and clonal selection before infection with the shRNA library. This simplifies gene engineering, which may be particularly advantageous for difficult cellular models.

Unfortunately, RNAi screens suffer from limited validation or overlap across studies, which have been mainly attributed to off-target effects and differences in knockdown efficiencies. In addition, they show limitations when the aim is to detect essential genes expressed at moderate or low levels [25]. Furthermore, siRNA binding can be nonspecific, leading to repression of several transcripts, where phenotypes derived from off-target effects might be dominant over the intended target [26]. Although efforts were made to optimize existing libraries such as minimizing sequence-specific off-targets by a well-balanced GC content and a low match with endogenous miRNAs and low internal complementarity, robust screens still require a significant number of shRNAs per gene, and shRNAs harboring common miRNA seed sequences need to be analyzed postscreen [24,27,28]. Importantly, shRNAs are produced in lower concentrations than exogenously applied siRNAs, which limits the concentration-dependent saturation of the endogenous miRNA pathway observed with synthetic siRNAs.

CRISPRko

Despite its relatively recent development, many CRISPRko-based pooled screens are described in the literature and are accessible via the GenomeCRISPR database [29] and project Achilles [22], for example. Pooled screens using CRISPRko take advantage of DNA cleavage mechanisms mediated by sgRNA-guided Cas9 that are subsequently repaired by nonhomologous end joining (NHEJ). Introduced **indels** can lead to a coding frameshift or introduction of a premature stop codon (reviewed in [30], Box 1). The ease of sgRNA design and production has led to the fast and cost-effective development of screening libraries (Box 2). CRISPRko also circumvents inherent problems of RNAi, as the knockout efficiency is based on a one-time editing event (Table 1). CRISPR screens are more sensitive and specific in detecting essential genes, particularly those with moderate expression levels [25]. By directly modulating DNA, indels can be introduced in all noncoding DNA sequences, including promoters, enhancers [31,32], and miRNA sequences [33].

Nickase Cas9 (nCas9): mutated Cas9 whose cleavage activity is restricted to one DNA strand only; for example, Cas9 D10A mutant only cleaves the DNA strand complementary to sgRNA.
Nuclease-deficient Cas9 (dCas9): Cas9 enzyme with point mutations in the nuclease domains, leading to a dead Cas9 that lacks DNA cleavage activity but retains the ability to bind target DNA based on the sgRNA target sequence.
Protospacer adjacent motif (PAM): 2-6 nucleotidic DNA sequence recognized and bound by a Cas enzyme. The target DNA of sgRNAs should be in close proximity to PAM to enable on-target nuclease activity. Each Cas enzyme recognizes a specific PAM sequence.
Regulatory element: regions of noncoding DNA that regulate the transcription of associated genes. These include promoters, enhancers, and silencers.
Single guide RNA (sgRNA): short chimeric RNA molecule composed of a variable portion called crRNA that defines the target DNA and a constant portion called transactivating crRNA (tracrRNA) that serves as a scaffold for the recruitment of Cas9.
siRNA: small duplex RNA molecule that directs gene silencing. siRNAs are naturally generated through DICER-mediated cleavage of longer duplex RNA precursors. When used as a tool, they are produced by chemical synthesis for delivery into cells.
Small hairpin RNA (shRNA): vector-based effector of RNAi, where the duplex RNA molecule is linked via a hairpin loop. shRNAs are generated through processing of pre-shRNA transcripts originating from an expression vector introduced into cells, before being associated with the RISC complex to direct gene silencing.
RNA polymerase (RNA-Pol): ribonucleic acid polymerase involved in RNA transcription. RNA-Pol II synthesizes precursors of mRNA, most small nuclear ribonucleic acid RNAs, and miRNAs. RNA-Pol III synthesizes tRNAs, rRNA 5S, and other small RNAs.
Transcriptional start site (TSS): nucleotide corresponding to the start of transcription at the 5′-end of a gene sequence.

Box 1. Pooled Screens – Biological Principles and Technical Aspects

RNAi is a crucial cellular process that mediates the post-transcriptional silencing of target genes (Figure IA). Integration into genome and exogenous expression of shRNAs is possible via viral particles carrying shRNA-encoding vectors (Figure IA). shRNA sequences are transcribed and processed to siRNAs by the ribonuclease DICER, subsequently entering the RNA-induced silencing complex (RISC) silencing machinery. RISCs expose one strand that directs formation of a specific siRNA/target RNA complex and RISC-mediated cleavage of the target mRNA. CRISPR/Cas9 technology is based on exogenous expression of CRISPR components in the eukaryotic cells. CRISPR/Cas9 involves the cooperation between a gene-specific sgRNA targeting Cas9 to the target DNA. Upon binding to the adjacent protospacer adjacent motif (PAM), Cas9 generates DNA DSBs that trigger either error-prone DNA repair by NHEJ or error-free repair by HDR, allowing irreversible knockout-inducing indels (CRISPRko, Figure IB) or more controlled editing at the targeted site. Newer variations of CRISPR technologies such as CRISPRi (Figure IC) and CRISPRa (Figure ID) allow further reduction or increase of endogenous gene expression respectively without genome editing. Both are based on sgRNA-guided localization of dCAS9 to endogenous gene promoters. dCAS9 binding to the TSS is sufficient to regulate gene expression of the downstream genes in a reversible manner and can be combined with transcriptional repressors (e.g., KRAB) or activators (e.g., VP64), respectively. dCas9 can also be linked to epigenetic modifiers (e.g., p300 or DNMT3A [93]), leading to the modulation of gene expression via specific epigenetic marks. CRISPRi-mediated repression can be reversed if an inducible dCas9 is utilized [44]. CRISPR base editing technology (Figure IE) can be applied to create point mutations or stop codons, as an alternative to CRISPRko [58]. The technology is based on dCas9 or nCas9 fused with a base-specific deaminase (APOBEC1 cytidine deaminase [57] or the B cell-specific AID [59,60]) that inserts point mutations. The conversion of cytidine to uridine results in U–G mismatches that can lead to nucleotide conversion following DNA replication or error-prone DNA repair. Precise nucleotide conversions are possible (C–T, G–A) in the presence of a uracyl glycosylase inhibitor.

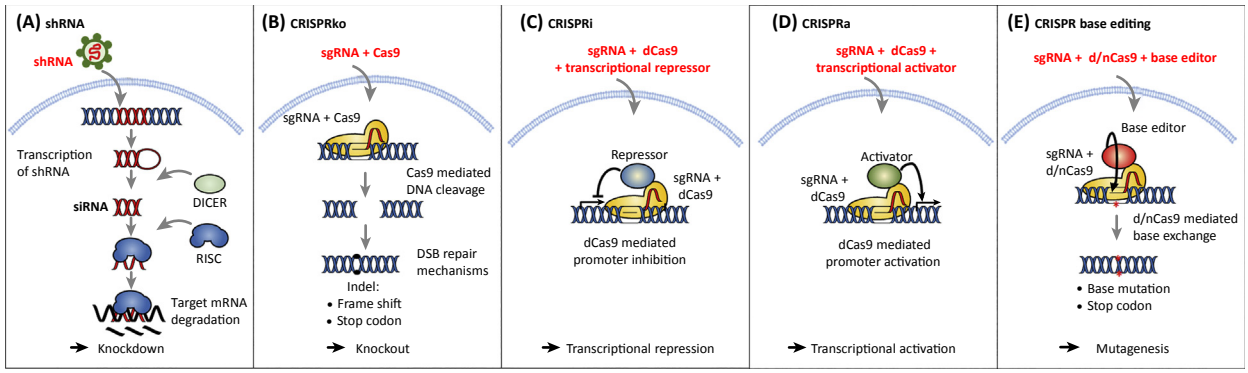
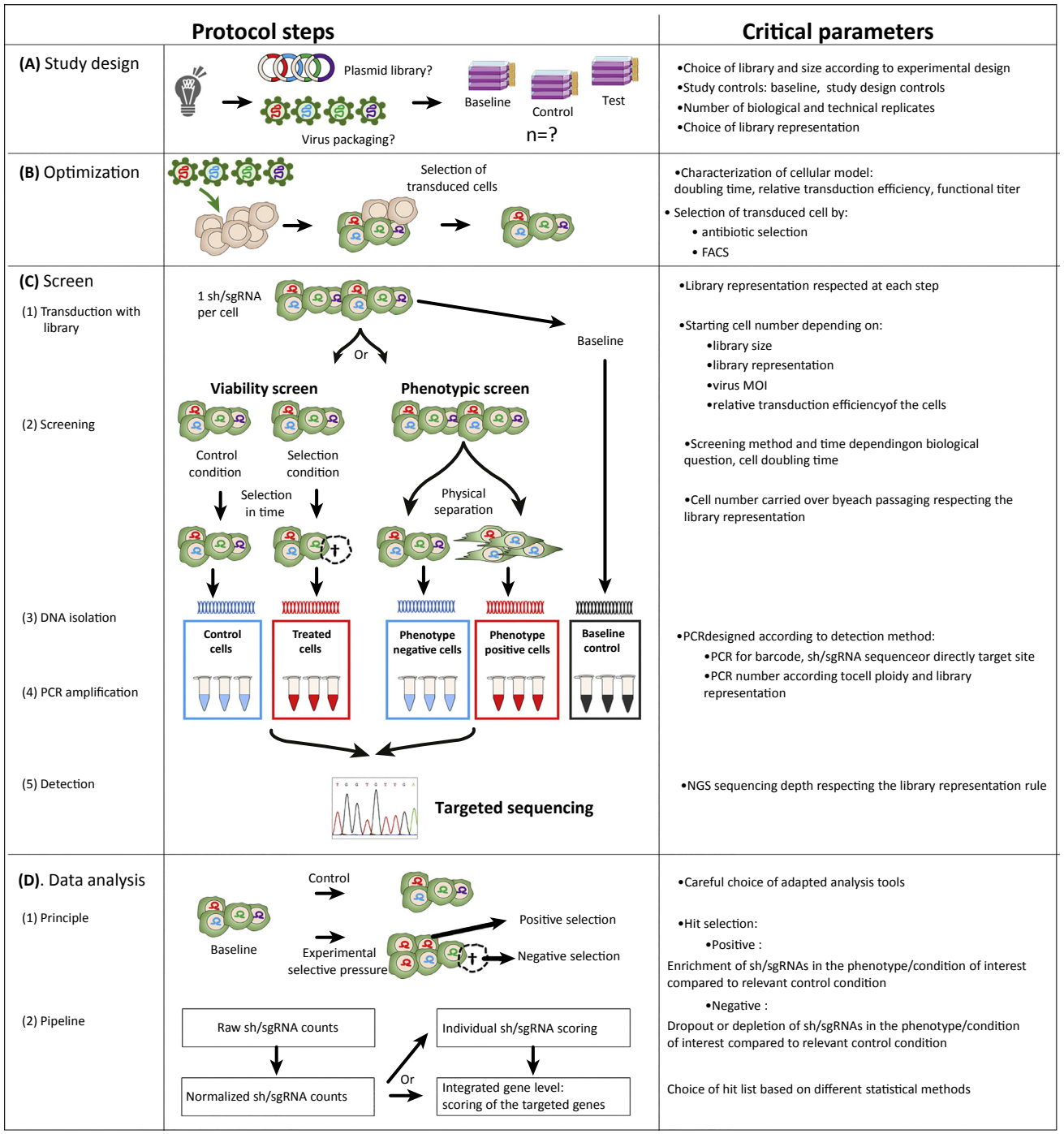


Figure I. Gene-Silencing Mechanisms. (A) Expression of virally-encoded shRNA leads RISC-mediated degradation of target RNA. (B) CRISPRko involves gene editing at the DNA level, leading to small insertions/deletions (indels) causing gene knockout. Alternative outcomes (not shown) are mutation correction (when donor DNA sequences are provided that serve as templates for DSB repair) or large deletions (using paired sgRNAs). (C) CRISPRi represses gene expression via gene promoter blockage. This is achieved by dCas9 binding to the promoter of interest and can be further amplified by fusion of dCas9 with a transcriptional repressor. (D) Activation of gene expression by CRISPRa is achieved via dCas9 associated with a transcriptional activator bound to the promoter sequence of interest. (E) CRISPR base editing is possible via directing dCas9 or nCas9 together with a base editor (deaminase enzyme) to perform nucleotide conversion to a target site. Abbreviations: Cas9, CRISPR-associated protein 9; CRISPRa, CRISPR activation; CRISPRi, CRISPR interference; CRISPRko, CRISPR knockout; dCas9, nuclease-deficient Cas9; DSB, double-strand break; sgRNA, single guide RNA.

Although highly specific, CRISPRko suffers from a low cutting efficiency and off-target effects. The target sequences are classically designed in the N-terminal coding exons, although targeting functional protein domains may be preferred to ensure gene inactivation [34]. The obtained genotypes include true knockouts with heterogeneous indels, heterozygote knockouts, or wild-type cells; and not all mutations introduce a frameshift or a premature termination codon. The biallelic functional gene mutations depend on sgRNA potency [35] and are estimated to be 30–70% [14], which leads to heterogeneity of clones with the same sgRNA and limits the achievable depletion screen readout. This bias can be partially resolved in pooled screens with readouts based on UMIs, where perturbed cells receiving a unique sgRNA molecule can be traced with the sequence-specific barcode [21]. Although more tedious, sequencing-based readouts can be adapted to directly detect the generated indels, rather than sgRNA sequences [10]. The ploidy and DNA repair status of the target cells influence the efficiency of CRISPRko, and the chromatin structure can affect the interaction of Cas9 with its



Trends in Biotechnology

Figure 1. An Overview of a Pooled Genetic Screening Protocol. (A) Study design involves the choice of an appropriate library, controls, number of replicates and library representation. (B) Optimization includes the determination of viral titer and cell transduction parameters. (C) Steps of a pooled screen: (1) Generation of knockdown/knockouts by cell transduction and selection. A baseline control is collected in the absence of any experimental selective pressure. (2) Screening process. The transduced cell population is physically separated based on the phenotype of interest. Viability screens involve selection over time and, if needed, contain an additional control arm without selection pressure. (3) DNA is isolated from controls, selected populations and baseline samples. (4) sh/sgRNA sequences, barcodes or

(Figure legend continued on the bottom of the next page.)

DNA target [36]. Off-target activity remains a challenge that can only be partially controlled [37–39]. Moreover, the induction of double-strand breaks (DSBs) induces a phenotypic response to DNA damage stress, resulting in impaired fitness and increased toxicity. This is particularly perceptible in amplified genomic regions, where false-positive results can occur due to the generation of multiple DSBs [34,40]. The unspecific toxicity of CRISPRko and heterogeneous indels may also hamper combinatorial screens with more than one target per cell. In these specific cases, gene knockdown strategies, such as RNAi and CRISPRi, may be more advantageous to reduce false positive hits.

CRISPRi

Inhibition of gene transcription can be achieved by the CRISPR system with **nuclease-deficient Cas9 (dCas9)** targeted to the gene promoter or other regulatory elements near the TSS. The inhibition can be further improved by fusing dCas9 to various repressor domains (reviewed in [30], Box 1), an approach that is commonly applied in pooled screens.

Although the rules for designing sgRNAs remain similar to CRISPRko, CRISPRi requires target sequences close to the TSS (−50 to +300, Table 1), with sgRNAs designed based on TSSs annotated in the FANTOM database appearing the most active [41]. CRISPRi overcomes several drawbacks of RNAi and CRISPRko technologies. By regulating endogenous transcription, CRISPRi (and also CRISPRa) allows the functional assessment of regulatory elements [15,42] and nuclear-retained noncoding RNAs [43]. The latter cannot be easily targeted by RNAi and are not always correctly perturbed by CRISPRko-introduced indels, which usually require large indels to perturb their functions. As CRISPRi does not produce indels, the associated off-target effects, such as indel heterogeneity or unspecific toxicity is minimized. CRISPRi activity is highly sensitive to base pair mismatches, leading to decreased off-target effects [44], but making it vulnerable to DNA polymorphism, for example, in regulatory regions. CRISPRi knockdown efficiencies are better than with RNAi, although clonal variations and variability across different sgRNAs targeting the same gene are still observed [44].

Still, CRISPRi carries several limitations. As observed for Cas9, the structure and dynamics of chromatin can affect dCas9 binding [36], which may alter its ability to affect gene expression. The outcome of CRISPRi may vary for genes surrounded by different cell type-specific epigenetic modifications and/or genes differentially regulated across cell cycle phases. Thus, cell-type specific libraries may be more precise in this case [43]. Since CRISPRi involves the targeting of dCas9 to TSSs, CRISPRi is suboptimal for genes regulated by more than one TSS and RNAi remains the standard for analyzing the function of gene splice variants; except if variants are regulated by different TSSs. CRISPRi can also lead to undesired effects when the target TSS regulates more than one transcript or in the case of bidirectional promoters, as single sgRNA may repress more than one gene [45]. Although not common, this can be revised during data analysis by assessing the genomic location of candidate genes.

CRISPRa

Although gain-of-function screens involving the ectopic expression of cloned ORFs have been carried out [46], these had numerous drawbacks including library design, representation, and cost. CRISPRa activates gene expression via dCas9 fused with transcriptional activators

target sites are amplified by PCR. (5) PCR-amplified sequences are detected by targeted sequencing. (D) Data analysis of a genetic screen: (1) selected conditions are compared to the relevant controls. Negative selection screens identify sh/sgRNAs depleted in the population of interest, whereas positive selection screens uncover sh/sgRNAs enriched in the phenotype of interest. (2) In the analysis pipeline, normalized counts can either be processed to score sh/sgRNAs based on individual sh/sgRNA representation or analyzed at the integrated gene level by scoring each targeted gene. Abbreviations: FACS, fluorescence activated cell sorting; MOI, multiplicity of infection; sgRNA, single guide RNA; shRNA, small hairpin RNA.

Table 1. Summary of Technologies Available for Pooled Screening

Technology	Principle	Mechanism of action	Reversibility of perturbation	Targeting possibilities	Targeting splice variants	Engineering required	On-target efficacy	Off-target effects	Refs
RNAi	Genome manipulation: loss-of-function	Knockdown: transcript degradation or inhibition of translation	Reversible	Protein-coding sequences, miRNAs, cytoplasmic long noncoding RNAs	Yes	shRNA	Depends on the target sequence and basal gene expression level	High: partial mismatch and repression of additional mRNAs, saturation of RNAi machinery with high siRNA levels	[13,16, 22–24,26,69]
CRISPRko	Genome modulation: loss-of-function	Knockout: frameshift or stop codon (indel) in DNA sequence	Irreversible	Protein-coding sequences, noncoding regions	Only if specific sequence targeted	sgRNA, Cas9	Depends on the sgRNA potency, chromatin structure, cell ploidy and indel specificity	Low: additional indels in the genome, increased toxicity while targeting multiple sites simultaneously	[22,25,29, 31–34,36, 40,68]
CRISPRi	Genome manipulation: loss-of-function	Knockdown: inhibition of transcriptional initiation	Reversible	Protein-coding sequences, noncoding regions, long noncoding RNAs	Only if TSS different	sgRNA, dCas9 + transcriptional repressor(s)	Depends on the sgRNA potency, TSS annotation, chromatin structure and basal gene expression level	Very low: repression of additional genes from a common TSS, repression via unspecific dCas9 binding	[15,42–45]
CRISPRa	Genome manipulation: gain-of-function	Overexpression: activation of transcriptional initiation	Reversible	Protein-coding sequences, noncoding regions, long noncoding RNAs	Only if TSS different	sgRNA, dCas9 + transcriptional activator(s)	Depends on the sgRNA potency, TSS annotation, chromatin structure and basal gene expression level	Very low: expression of additional genes from a common TSS	[15,44,47, 48,51,53,56]
CRISPR/base editors	Genome modulation: loss/ gain-of-function	Mutagenesis: nucleotide change in DNA sequence	Irreversible	Protein-coding sequences, noncoding regions	Only if specific sequence targeted	sgRNA, d/ nCas9 + base editor	Depends on the sgRNA design, enzyme specificity, chromatin structure	Unprecise base exchange on target gene(s), mutagenesis of unspecific sites	[57–61]

CellPress REVIEWS

(reviewed in [30], Box 1). Gene transcription is induced at the endogenous locus and can recreate splicing isoform diversity. CRISPRa can interrogate the function of protein-coding genes, noncoding RNA genes [47,48] and gene regulatory elements [15] (Table 1). The targeted sequences promoting efficient CRISPRa are located in the proximity of TSSs, but distinct from those required for CRISPRi (−300 to 0).

Although initial CRISPRa systems resulted only in moderate gene activation and were not applicable for genome-wide screening, improvements have been achieved by applying scaffolds of antibody epitopes recruiting VP64 copies (SunTag arrays) [44,49] or by tethering multiple activators to dCas9 (VPR system) [50]. Settings leading to synergistic activation [Synergistic Activation Mediator (SAM) system] through the recruitment of multiple activators by modified sgRNAs have also been efficient in genome-wide screens [48,51]. Modified scaffolds containing different RNA-binding motifs also allow for simultaneous CRISPRi and CRISPRa [7]. These systems achieve robust increases in gene expression, although variability between sgRNAs is observed [52,53]. Alternatively, genes may be activated by modifying the epigenetic status of the DNA sequences [15,54].

The pitfalls of CRISPRa are similar to those of CRISPRi due to its dependency on TSS annotation and DNA structure. Multiple transcripts expressed from the same TSS cannot be controlled, and ORFs are still advantageous for interrogating specific splice variants or mutated proteins [55]. CRISPRa may also activate intronic noncoding RNAs as an off-target effect. Finally, efficient activation strategies require the use of separate cloning vectors in addition to dCAs9 [56], complicating the screening protocol.

Box 2. Library Set-up

Pooled screens require libraries made of vectors that integrate into the genome and can be detected by barcoded sequences. Duplex DNA oligonucleotides encoding each sh/sgRNA of the library are cloned into viral vectors to generate a pooled plasmid DNA library, which is then converted into a virus library that can be used for the screen (Figure I). Transduced cells should contain all sh/sgRNAs in the desired representation; that is, absolute number of cells per each sh/sgRNA molecule. Libraries are typically based on lentiviral vectors applicable to dividing and nondividing cells. Adenoviruses are particularly efficient *in vivo*, but they do not integrate into the genome and require detection via sequencing of targeted sites [10,11].

RNAi and CRISPR libraries are constantly optimized to enhance sh/sgRNA efficacy and appropriate coverage, as well as with novel plasmid packaging options and transduction methods. RNAi libraries are available as plasmids or as ready-to-use viral particles. Plasmid-based libraries allow higher production of viral particles in house, but quality of homemade libraries should be verified before use. Distinct vectors are available that use different types of **RNA polymerase** (RNA-Pol) to drive shRNA expression. RNA-Pol III-based vectors are typically used, though RNA-Pol II-based vectors allow simultaneous expression of the selection cassette and the shRNA [94]. CRISPR library components are supplied in various formats, including DNA oligonucleotides, plasmids, and viral particles. The CRISPR/Cas9 vector in the one-plasmid system contains both the Cas9 and the sgRNA sequences. The two-plasmid system uses one plasmid to generate Cas9-expressing cells and a second one to produce the sgRNA; both containing a selection marker. The two-plasmid system results generally in higher sgRNA titers and allows for selection of a control Cas9-expressing clone. dCas9 fused with repressors or activators is generally introduced as a separate plasmid. Certain technologies, such as the CRISPRa SAM system, require additional plasmids for delivering multiple CRISPR components. sgRNAs transcription is classically driven by RNA-Pol III, although RNA-Pol II versions are also available. CRISPR screening in single cells requires combined expression of functional sgRNAs via RNA-Pol III and sgRNA sequence-containing artificial mRNAs, detectable by RNA-seq, via RNA-Pol II [17–20]. UMI-based screens imply the use of sgRNAs combined with unique barcode sequences identifying each sgRNA molecule, which allow the discrimination of clones arising from unique sgRNA-barcode pairs representing UMIs [21]. For multiplex screens sh/sgRNAs can be transcribed from separate promoters or as a single transcript further cleaved for independent activities [6,8]. Inducible and cell type-specific vectors are particularly advantageous for *in vivo* screens, where temporal and spatial control of sh/sgRNA expression is required [31]. The use of Cre-dependent Cas9 transgenic mice [95] enables faster *in vivo/ex vivo* screens on difficult primary cells, tissues and whole organism.

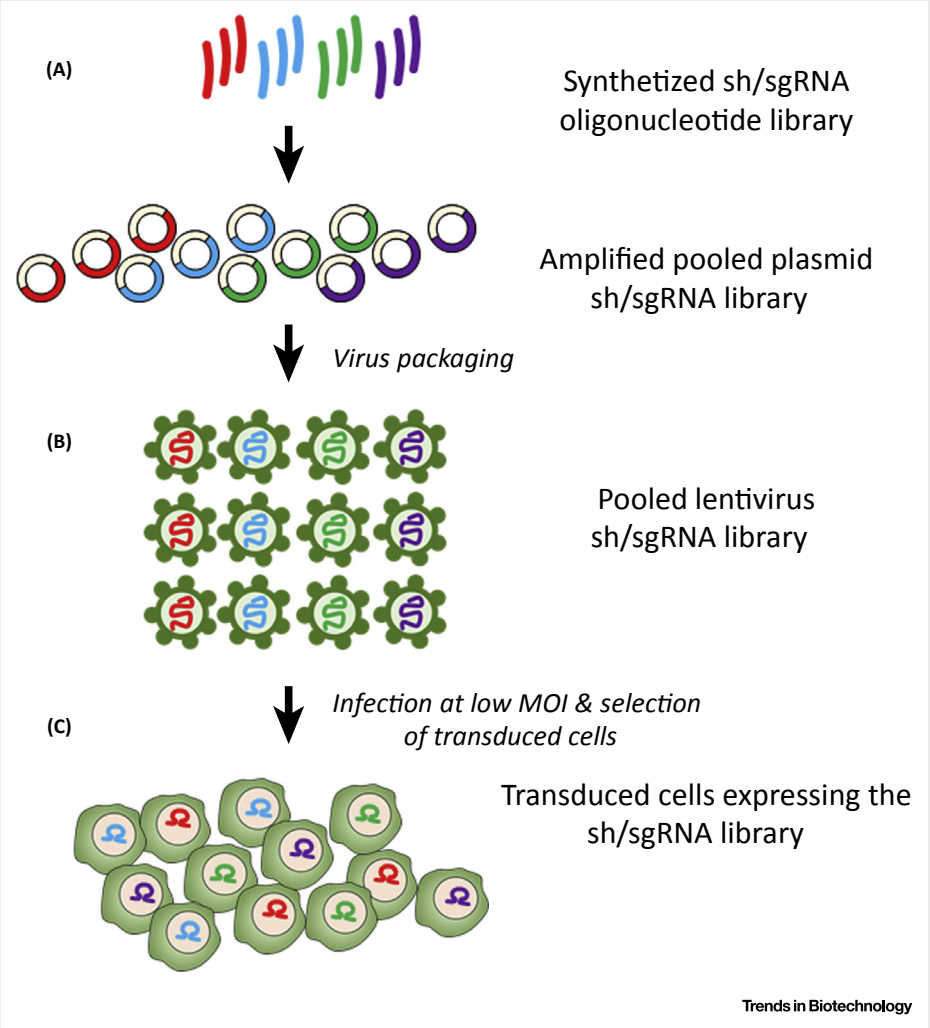


Figure I. Library Representation in the Pooled Screen. (A) *In silico* designed sh/sgRNA oligonucleotides are first synthesized and cloned in the plasmid of choice constituting the so-called plasmid library. The library should contain sufficient and equal amounts of each sh/sgRNA sequence represented in plasmid constructs. (B) After amplification in bacteria, plasmids are further incorporated into lentiviral particles with similar representation of sh/sgRNA sequences. (C) A lentivirus library is applied to the cells at the onset of a screen, ensuring best conditions for creating a cell library with one viral particle per cell. The desired representation needs to be maintained over the entire course of the screen, including cell-passaging steps. Abbreviations: sgRNA, single guide RNA; shRNA, small hairpin RNA.

Second-Generation CRISPR Base Editing Systems

Systems that combine dCas9 or **nickase Cas9** (nCas9) with **base editors** such as apolipoprotein B mRNA editing enzyme catalytic polypeptide 1 (APOBEC1) cytidine deaminase [57] or B cell-specific activation-induced adenosine deaminase (AID) (reviewed in [30], Box 1), allow the conversion of single bases at the target sites without causing DSBs, thus offering a novel approach for interrogating the role of specific point mutations. So far, base editing appears more efficient than homology-directed recombination (HDR)-based CRISPRko, so it should be particularly valuable for recapitulating/identifying functional variants and oncogenic mutations associated with drug resistance (Table 1).

Two approaches are applied for pooled screening: (i) targeting multiple loci by saturating mutagenesis triggered by sequence-specific sgRNAs; and (ii) designing multiple sgRNAs tiled across the same target region [57–60]. So far, dCAS9–AID fusion has been reported to generate more diverse point mutations than dCAS9–APOBEC1, whereas nCAS9–AID appears more suitable for generating precise single-nucleotide conversions than the panels of variations required for screening [61]. Importantly, deaminase and sgRNA-specific off-target effects are reported to be rare. However, these base-editing technologies are currently halted by the limited number of options available for generating point mutations. Mutations are not introduced precisely but rather within a window of several base pairs and do not occur at equal frequencies. Also, the base editors are even bigger than Cas9, limiting the delivery options. The use of mutant cytidine deaminase with a narrower editing window [62] or tRNA adenosine deaminase for A to G conversion [63] may enable more precise and diverse mutation options.

Key Parameters of Experimental Design and Protocol

The choice of screen technology and the protocol details are based on not only the biological question but also numerous technical aspects such as the properties of the cellular model, the available technical know-how, equipment, and the time frame of the experiment. Each screen requires a dedicated experimental design and appropriate control conditions. Detailed protocols describing RNAi [64,65] and CRISPR [56,66] screens are available elsewhere. Here, we present a summary of the key parameters essential for a robust and reproducible experimental set up (Figure 1), intended to serve as guide in designing a screen.

Library Design

A great variety of genome-wide and targeted libraries are available from academic centers; for example, via Addgene repository, and from commercial providers (Box 2 and Figure 1A). Typically, pooled libraries contain several sh/sgRNAs targeting the same gene or noncoding regions, that is, library coverage or redundancy. The size of the library is defined by the number of different sh/sgRNAs targeting a set number of genes/regions. Computational tools help to design sh/sgRNAs with high on-target specificity, efficiency, and low off-target effects (reviewed in [38,67]). To counteract the biases introduced by differences in target RNA depletion efficiency and off-target effects, the library should include at least four independent sh/sgRNAs per gene. Adding additional sgRNAs to the screen identified $<5\%$ more hits per sgRNA added [68], whereas a high number of shRNAs per gene was particularly beneficial to eliminate off-target effects [69]. Screens assessing noncoding regions target promoter functional elements (e.g., transcription factor binding sites), or putative regulatory sequences of genes of interest. For the latter, multiple sgRNAs are tiled across the corresponding genomic sequence [70]. In some instances, the library size will be dictated by the experimental setup; for example, low cell number of primary cultures and organoids. *In vivo* experiments are limited by the amount and accessibility of endogenous cells, by the number of cells that can be implanted, and the engraftment efficiency. In these cases, CRISPR-based technologies may be more advantageous over RNAi, as they allow lower library coverage, thus lower starting cell number, whilst achieving similar specificity.

Library Representation

The library representation indicates the absolute number of molecules of each sh/sgRNA carried across all steps of the screen (Box 2). High library representation has a greater impact on robustness of the screen than the number of biological replicates [71]. When too low, representations introduce a high background noise, resulting in false positives due to random variations of sh/sgRNAs present in low numbers at the onset of the screen. A high representation is key for a negative selection to avoid depletion by random chance. If small changes are

anticipated, a 500–1000-fold representation is indispensable [56,71]. Certain screening conditions, for example, tumor engraftment *or* differentiation protocols in organoids or primary cells, can create a natural bottleneck leading to unspecific loss of the library representation. The bottleneck may be desired in positive selection screens to identify the strongest phenotypes or in UMI-based screens to select the clones identified by unique sgRNA molecules [21]. For *in vivo* screens, which show higher background, a smaller pool size with higher representation can improve the signal-to-noise ratio. Also, multiple animals can be grouped as single biological replicates to provide a robust representation *in vivo* [13].

Screen Optimization

Important cell type-specific parameters need to be addressed at the onset of the screen (Figure 1B), including the transduction efficiency (assessed by **functional titer**), the applicable load of viral particles, and available methods for the removal of uninfected cells [i.e., antibiotic resistance or fluorescence activated cell sorting (FACS)]. Optimization with nontargeting sh/sgRNAs is recommended before embarking on the final experiment to identify variations introduced by the selection procedure to be considered for the final screen. For CRISPR-based screens, (d)Cas9 and other CRISPR components can be delivered beforehand or simultaneously with the sgRNA library. If the cellular model enables clonal expansion of single cells, the experiments may involve selection of the best (d)Cas9-expressing cell clones before introduction of the sgRNA library, increasing viral titer and reducing the risk of high variability in (d)Cas9 expression level and drift in (d)Cas9 expression over time. This is recommended, for example, for successful knockouts via less potent sgRNAs [35]. (d)Cas9 expression should be monitored regularly both in polyclonal and monoclonal populations. The polyclonal (d)Cas9-expressing population or a single plasmid system with simultaneous Cas9 and sgRNA expression may be more adapted for difficult cellular systems; for example, for primary cells, where genetic or phenotypic heterogeneity cannot be recreated from a single clone. One-plasmid systems or inducible Cas9 expression may be advantageous for cells not tolerating long-term, constitutive Cas9 expression.

Starting Cell Number and Controls

The starting cell number depends on the library size, representation, and viral **multiplicity of infection** (MOI) (Figure 1C). The applied MOI should be low (0.1–0.3) to increase the chance of delivering one sh/sgRNA molecule per cell. Although not possible in the classical screen readout, cells receiving more than one sgRNA molecule can be identified in UMI-based screens with single cell lineage tracing [21] or single cell RNA-Seq readout [17–20]. When the available cell number is restricted, limiting the library size is recommended rather than compromising the representation. Cells are ready for screening directly after appropriate selection. Although the CRISPRko timeline may depend on sgRNA potency [35], a time lapse of 7 days after sgRNA transduction was reported to give most efficient knockouts; whereas 4 days were best for CRISPRa SAM overexpression [56]. A baseline must be collected from sh/sgRNA-positive cells before the onset of selection (Figure 1C), which will serve as a threshold for each sh/sgRNA representation in the initial pool during analysis. Viability screens under a particular selective pressure should include an additional control to exclude genes essential for baseline cell survival and growth. Likewise, naïve cells (not edited by CRISPR) are needed as a control to evaluate mutagenesis/indel events at the targeted loci. At least two or three independent biological replicates are required for statistical power [68]. This is especially important when the selection could lead to a genetic or phenotypic drift; for example, with genotoxic drugs.

Screening Timeline

Screen endpoints are based on the selection strategy. The separation may be performed within hours or days; for example, when screening for biomarker-positive cells via reporter-based

assays or flow cytometry. The population doubling time dictates the duration of the viability screens, where a certain number of population doublings is indispensable for the required phenotype to be observable. Thus, the timeline should allow sufficient time for selection to occur, while minimizing random noise and drift. In most cases, 5–16 population doublings are adequate [22,72]. If the optimal duration of the screen is difficult to predict, cells can be collected at different times to determine the best time point experimentally, based on the signal-to-noise ratio.

DNA Amplification

Detection of sh/sgRNA sequences or barcodes integrated into the genome relies on PCR-based amplification of genomic DNA isolated from pooled barcoded cells. Generally, PCR amplification is performed directly on genomic DNA, although enriching for DNA sequences containing sh/sgRNA sequences can be envisaged by introducing specific enzymatic digestion of DNA flanking the introduced sequence [21]. To avoid self-annealing of shRNAs, leading to unequal recovery via PCR, shRNAs are typically detected by amplification of half of their sequence or association with barcodes [65]. Optionally, a mutation in the shRNA loop [69] or addition of a DNA relaxing agent prevents secondary structures [73]. The PCR format must reflect the predetermined library representation and depends on cell ploidy and number of collected cells [71].

DNA Detection

PCR-amplified sh/sgRNA sequences and/or barcodes are identified via targeted DNA sequencing. Next-generation sequencing (NGS) outperforms microarrays, providing higher flexibility, scalability, and dynamic range. Sequencing sg/shRNAs with additional barcodes allows for a more precise detection and quantification of introduced perturbations. Classical barcodes can be associated with sh/sgRNA sequences [4,65]; alternatively, variable barcodes can be introduced with each sgRNA molecule to create unique UMIs for each cell [21]. Because NGS counts are used as a proxy for cell clone abundance, the sequencing depth must reflect the library size and representation. Classical detection techniques are performed on bulk populations and do not allow further molecular readouts. In contrast, the development of artificially transcribed sgRNA-based barcodes enables detection of the introduced sgRNA in each individual cell, with a readout based on relative abundance of sgRNA-transcribing cells, rather than sgRNA sequencing reads [17–20]. These techniques allow simultaneous single-cell transcriptome readouts, revealing the direct impact of each perturbation on gene expression. Importantly, however, a robust analysis requires 50–200 single cells per sgRNA, limiting the screening to smaller libraries. CRISPRko screens may involve sequencing of targeted regions, enabling direct quantification of gene perturbations. Sequencing of indels can be a readout if nonintegrating vectors are applied or in *in vivo* screens where no baseline control is available [10]. Targeted sequencing represents a major readout for saturating mutagenesis screens due to variability of resulting mutations [59,61].

Data Analysis

Data analysis aims to translate sequencing reads to the relative abundance of the corresponding sh/sgRNAs to identify the most prominent hits (Box 3 and Figure 1D). The high-throughput nature of screens introduces significant technical noise, where data analysis suffers from the pitfalls of the underlying technique. Gene ranking based on results from all sh/sgRNA molecules per gene/regulatory element is key to visualizing the effect of induced perturbations in the pool context. Currently, no standard analysis pipeline exists and several bioinformatics tools are usually applied simultaneously (Box 3). Considering the intersection of hits given by individual tools can help to reduce false-positive hits and to narrow down the gene list for further

Box 3. Main Analytical Tools Available for Hit Selection

The main tools available for pooled screen analysis, including the input data and prerequisite computational knowledge are summarized in Table I. Data analysis is based on the comparison of sh/sgRNA representation between phenotype-positive and negative cells. In viability screens these represent the control and selective arms. Data analysis requires uploading of the sh/sgRNA read counts with corresponding IDs and target gene annotation. The sh/sgRNA representation should be curated using the baseline and sh/sgRNAs with initial low representation should be removed. The analysis steps involve quality control, read alignment and normalization, statistical analysis of differential sh/sgRNA representation (i.e., sh/sgRNA scoring), and integration of multiple sh/sgRNAs targeting each gene (i.e., gene ranking). This enables target ranking and filtering to reveal the most prominent hits that qualify for further experimental validation.

The analysis of screens with single cell RNA-seq readout addresses the relative abundance of single cells per each perturbation, rather than the relative abundance of sgRNA sequences measured by bulk sequencing [17–20]. Single cell RNA-seq further provides the transcriptomic profile in surviving cells, including a direct validation of the target gene expression level. The analysis of pooled screens based on single cell RNA-seq [17–20] or UMI-based quantification of unique sgRNA clones [21] allows the identification of cells with more than one perturbation introduced. Cell barcoding further helps to distinguish false-positive and negative hits by identifying the outliers as displaying strong selection from the remaining clones undergoing mild selection.

In the case of CRISPRko, the analysis can also be based on the relative abundance of indels present at the target sites, reflecting the actual sgRNA efficiency and indel heterogeneity. Saturating mutagenesis screens require custom scripts to identify the frequencies of mutagenesis in the targeted sequences [59,61]. For tiling screens, sgRNAs are assigned to neighboring sequences, to identify the 'window' with the highest score indicative of a functional sequence [42,70]. CRISPRko screening performed in aneuploid cells requires verification of genome locus of the hits, as multiple DSBs in amplified regions lead to unspecific toxicity [34,40]. Similarly, the genome locus identification can reveal multiple genes targeted by a single gRNA via CRISPRi/a [45].

Classical nontargeting controls and controls that target known nonessential genes or loci can be considered [96]. The latter reflect nonspecific perturbations linked to DNA cleavage and DNA damage in CRISPRko screens, whereas nontargeting controls give indications about the background perturbations induced by activation of the CRISPR or RNAi machineries. Positive controls are experiment specific and can include essential genes (e.g., housekeeping genes) when screening for coding genes or exonic sequences in the case of noncoding sequences [31].

Each tool applies different algorithms for score calculation and hit ranking. To reduce false positives and generate a stringent hit list, several analysis tools can be combined. Novel interfaces such as CRISPRAnalyzeR [97] caRpools [98], or PinAPL-Py [99] help to perform several analyses in one workflow in a user-friendly mode. Some packages perform downstream functional and network analyses or specific algorithms for, for example, incorporation of gene copy alterations (CERES, [100]), single cell transcriptome readout (MIMOSCA [17]), UMI identification (CRISPR_UMI [21]), or target sequence/indel mutation verification (CRISPResso, [101]).

Table I. Summary of data analysis methods

Algorithm	Complete name	Internal gold standard genes required	Computational knowledge (implementation, programming language)	Input content	Normalization of sh/sgRNA counts performed by the tool	License and restriction to use	Refs
JACKS	Joint Analysis of CRISPR/Cas9 Knock-out Screens	No	Yes (Python/R)	sh/sgRNA ID, gene ID, and their respective raw counts	Yes	Massachusetts Institute of Technology (MIT) License	[102]
PBNPA	Permutation Based Non-Parametric Analysis of CRISPR Screen Data	No	Yes (R)	sh/sgRNA ID, gene ID, and the counts for 2 single-replicate samples (control read counts and treatment read counts)	Yes	GNU General Public License version 3.0.	[103]
ENCoRE	Easy NGS-to-Gene CRISPR REsults	No	No (stand-alone Java program)	fastq files	Yes	GNU General Public License version 3.0.	[104]
CasTLE	Cas9 high-Throughput maximum Likelihood Estimator	Yes	Yes (Python)	sh/sgRNA ID and the counts for each sample, non-targeting sh/sgRNAs list	Yes	Redistribution and use in source and binary forms, with or without modification, are permitted provided that some conditions are met (https://bitbucket.org/dmorgens/castle)	[75]

Table I. (continued)

Algorithm	Complete name	Internal gold standard genes required	Computational knowledge (implementation, programming language)	Input content	Normalization of sh/sgRNA counts performed by the tool	License and restriction to use	Refs
BAGEL	Bayesian Analysis of Gene EssentiaLity	Yes	Yes (Python)	gene ID, sh/sgRNA ID, and their respective raw counts or log2 FC, sh/sgRNAs targeting essential and non-essential gene list ("gold standard")	Yes	Creative Commons Attribution 4.0 International Public License	[105]
ScreenBEAM	Screening Bayesian Evaluation and Analysis Method	No	Yes (R)	gene ID, sh/sgRNA ID and their respective counts	Yes	Massachusetts Institute of Technology (MIT) License	[106]
MAGeCK	Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout	No	Yes (Python)	gene ID, sh/sgRNA ID and their respective counts, or fastq files	Yes	Berkeley Software Distribution (BSD) License	[107]
HiTSelect	High Throughput Screen Deconvolution	No	Not necessarily (friendly interface available)	gene ID, sh/sgRNA ID and their respective counts	Yes	GNU General Public License	[108]
sgRSEA	single-guide RNA Set Enrichment Analysis	No	Yes (R)	sh/sgRNAs ID, gene ID, and raw sgRNA read counts under treatment and control condition	No	GNU General Public License version 2.0	(https://rdrr.io/cran/sgRSEA/)
CERES	Computational correction of copy-number effect in CRISPR-Cas9 essentiality screens	No	Yes (R)	gene ID, preprocessed logFC for sgRNA, copy number data	No	Berkeley Software Distribution (BSD) License (BSD 3-Clause)	[100]
RIGER	RNAi gene enrichment ranking	No	no (GENE-E software module)	gene ID, sh/sgRNA ID and their respective normalized counts	No	Free of charge by academic and other non-profit researchers. Commercial users have to contact the Broad Institute for licensing terms.	[109]
RSA	Redundant siRNA activity	No	Not necessarily (friendly interface available. Updated version in Python, Perl or R)	gene ID, sh/sgRNA ID and their respective activities (FC)	No	The Apache License, version 2.0	[110]
DESeq2/EdgeR	–	No	Yes (R)	gene ID, sh/sgRNA ID, and their respective raw counts	Yes	DESeq2 & EdgeR: GNU Lesser General Public License (LGPL)	[111,112]

validation. The continuous evolution of screen technologies will require novel analysis tools, considering complex kinetics or additional readouts.

Screen Validation

Lead candidates identified in genetic screens require experimental validation to exclude false positives and off-target effects. This generally includes verifying several sh/sgRNAs targeting the hit gene and validating the phenotype in multiple cellular models. Nevertheless, results may be cell type specific and certain phenotype-to-genotype associations may not be confirmed in another model; for example, tumor cells with different genetic backgrounds [74]. Rescreening with a different library type can be applied, where the intersection between RNAi, CRISPRko and/or CRISPRi screens can be considered [75]. If the initial screen is genome-wide, the validation screen can target restricted genes with increased sh/sgRNA coverage to enhance specificity and eliminate false positives. In such settings, novel cellular models may be useful; for example, primary cells or *in vivo* conditions. Screen reversal (e.g., the validation of loss-of-function using gain-of-function screens) may reveal opposite or complementary phenotypes [44,46]. Finally, rescue experiments remain the gold standard to demonstrate the specificity of sh/sgRNAs; for example, via a silencing-resistant version of the gene of interest.

Concluding Remarks and Future Perspectives

Pooled screens are labor intensive, yet, when properly designed and stringently analyzed, they represent a powerful tool for the identification of gene function. With a certain level of specificity inherent to each technology, RNAi, CRISPRi, and CRISPRko approaches are often complementary [75,76], although for some biological applications one approach may be preferable over another. For example, CRISPR-based screens are considered more specific to identify gene essentiality [34,77,78], while gene knockdown technologies, such as CRISPRi and RNAi, may be better suited to perform synthetic lethality and epistasis experiments. It remains to be seen to what extent the different outcomes of gene knockdown versus knockout screens relate to technical artefacts or true biological effects [25] (see Outstanding Questions). Indeed, the full inhibition of many molecular pathways by simultaneous indels is likely to be lethal. Knockdown technologies can also be better adapted to genes that do not tolerate a complete knockout and genes showing varying phenotypes at different expression levels. Combining data curation and comprehensive meta-analysis of screens based on different technologies will strengthen the reliability of the data and further our understanding of genes essential in multiple biological processes [22,23,29,79–81]. Also, the comparison of pooled screens with human population analysis based on gene variants and mouse knockout studies will be crucial to reveal gene essentiality under a panel of key *in vivo* conditions. So far, the essential genes identified in population studies correlate to the most severe *in vitro* phenotypes uncovered by CRISPRko screens [1].

The continuous refinement of technologies will certainly bring novel, more-efficient, and specific screening strategies. Investigating efficacy and off-target effects of individual sh/sgRNAs will also improve libraries and analysis pipelines [24,82,83]. Although a challenge in the genome-wide context, adapting the readout to the mutagenesis status induced by CRISPRko and base editing may further increase screen specificity and open the technology to more efficient nonviral delivery systems. Single cell barcode-based approaches help to decrease the noise arising from variable indels introduced by the same sgRNAs [21], yet, so far, they cannot directly assess the indel sequences. Although CRISPRko screens mainly rely on NHEJ-based mutations, novel HDR-based approaches open new possibilities for precise pooled screening [84,85]. Nevertheless, the low efficacy of HDR remains an important challenge. The

Outstanding Questions

Are the differences in outcome of knockout versus knockdown screens related to technical limitations or based on biological effects of differential gene expression level?

Can more precise editing mechanisms and enhanced readouts of the editing outcomes increase the specificity of pooled screen analysis?

How can pooled screens further evolve towards multiplexing for improved functional molecular and cellular readouts, while retaining a practical library size and regular experimental flow?

development of more accurate Cas9 and related editing tools will extend the screening applications [57,86]. In particular, screens based on **CRISPR-associated protein 13 (Cas13)** could allow direct targeting and editing of cytoplasmic and nuclear RNAs [87,88], which may have higher specificity and less off-target effects than RNAi.

Where individual gene perturbations may not reveal complex phenotypes, multiplexing technologies should facilitate the interrogation of the functional relationships between genes; for example, in synthetic lethality interactions and drug–target combinations. Although pairwise combinations have been reported [3–8], they require elaborate preparation of large paired libraries and increase the risk of toxicity and indel heterogeneity. Because Cpf1 (Cas12a) supports multiplex gene regulation [89], it could potentially be applied in a screen context following CRISPR RNA (crRNA) sequence optimization and development of new specific libraries. Alternatively, dual Cas9 systems allow multiplex editing not only for screening synthetic lethality interactions, but also for simultaneous repression and activation of different genes [90–92]. These could in principle be combined with any other CRISPR-based technology. Another promising development may arise from multiplexing pooled screens with various molecular readouts, such as transcriptome responses [17–20], opening exciting opportunities for the functional evaluation of perturbed cells.

Acknowledgments

This work was supported by the Luxembourg Institute of Health (LIH), Fondation Cancer of Luxembourg (projects INVGBM and INCOMING), Télévie-FNRS (Grant 7.6533.16 and 7.4592.14) and the Luxembourg National Research Fund (FNR; CORE Junior C17/BM/11664971/DEMICS).

References

1. Bartha, I. *et al.* (2018) Human gene essentiality. *Nat. Rev. Genet.* 19, 51–62
2. Boutros, M.F. *et al.* (2015) Microscopy-based high-content screening. *Cell* 163, 1314–1325
3. Horn, T. *et al.* (2011) Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods* 8, 341–346
4. Du, D. *et al.* (2017) Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat. Methods* 14, 577–580
5. Shen, J.P. *et al.* (2017) Combinatorial CRISPR-Cas9 screens for *de novo* mapping of genetic interactions. *Nat. Methods* 14, 573–576
6. Han, K. *et al.* (2017) Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.* 35, 463–474
7. Zalatan, J.G. *et al.* (2015) Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350
8. Horlbeck, M.A. *et al.* (2018) Mapping the genetic landscape of human cells. *Cell.* 174, 953–967.e22
9. Parnas, O. *et al.* (2015) A Genome-wide CRISPR screen in primary immune cells to dissect regulatory networks. *Cell* 162, 675–686
10. Wang, G. *et al.* (2018) Mapping a functional cancer genome atlas of tumor suppressors in mouse liver using AAV-CRISPR-mediated direct *in vivo* screening. *Sci. Adv.* 4, eaao5508
11. Chow, R.D. *et al.* (2017) AAV-mediated direct *in vivo* CRISPR screen identifies functional suppressors in glioblastoma. *Nat. Neurosci.* 20, 1329–1341
12. Katigbak, A. *et al.* (2016) A CRISPR/Cas9 Functional screen identifies rare tumor suppressors. *Sci. Rep.* 6, 38968
13. Miller, T.E. *et al.* (2017) Transcription elongation factors represent *in vivo* cancer dependencies in glioblastoma. *Nature* 547, 355–359
14. Doench, J.G. *et al.* (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* 32, 1262–1267
15. Klann, T.S. *et al.* (2017) CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* 35, 561–568
16. Seo, M. *et al.* (2014) RNAi-based functional selection identifies novel cell migration determinants dependent on PI3K and AKT pathways. *Nat. Commun.* 5, 5217
17. Dixit, A. *et al.* (2016) Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* 167, 1853–1866 e17
18. Adamson, B. *et al.* (2016) A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* 167, 1867–1882 e21
19. Jaitin, D.A. *et al.* (2016) Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq. *Cell* 167, 1883–1896 e15
20. Datlinger, P. *et al.* (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* 14, 297–301
21. Michlits, G. *et al.* (2017) CRISPR-UMI: single-cell lineage tracing of pooled CRISPR-Cas9 screens. *Nat. Methods* 14, 1191–1197
22. Cowley, G.S. *et al.* (2014) Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data* 1, 140035
23. Schmidt, E.E. *et al.* (2013) GenomeRNAi: a database for cell-based and *in vivo* RNAi phenotypes, 2013 update. *Nucleic Acids Res.* 41, D1021–1026
24. Kampmann, M. *et al.* (2015) Next-generation libraries for robust RNA interference-based genome-wide screens. *Proc. Natl. Acad. Sci. U. S. A.* 112, E3384-3391
25. Hart, T. *et al.* (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163, 1515–1526

26. Franceschini, A. *et al.* (2014) Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4548–4553
27. Tafer, H. (2014) Bioinformatics of siRNA design. *Methods Mol. Biol.* 1097, 477–490
28. Fellmann, C. and Lowe, S.W. (2014) Stable RNA interference rules for silencing. *Nat. Cell Biol.* 16, 10–18
29. Rauscher, B. *et al.* (2017) GenomeCRISPR – a database for high-throughput CRISPR/Cas9 screens. *Nucleic Acids Res.* 45, D679–D686
30. Mitsunobu, H. *et al.* (2017) Beyond native Cas9: manipulating genomic information and function. *Trends Biotechnol.* 35, 983–996
31. Canver, M.C. *et al.* (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* 527, 192–197
32. Korkmaz, G. *et al.* (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.* 34, 192–198
33. Chen, S. *et al.* (2015) Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* 160, 1246–1260
34. Munoz, D.M. *et al.* (2016) CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions. *Cancer Discov.* 6, 900–913
35. Yuen, G. *et al.* (2017) CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. *Nucleic Acids Res.* 45, 12039–12053
36. Kuscu, C. *et al.* (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* 32, 677–683
37. Tycko, J. *et al.* (2016) methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* 63, 355–370
38. Cui, Y. *et al.* (in press) Review of CRISPR/Cas9 sgRNA design tools. *Interdiscip. Sci.*
39. Kosicki, M.K. *et al.* (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771
40. Aguirre, A.J. *et al.* (2016) Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 6, 914–929
41. Radzisheuskaya, A. *et al.* (2016) Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. *Nucleic Acids Res.* 44, e141
42. Fulco, C.P. *et al.* (2016) Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 354, 769–773
43. Liu, S.J. *et al.* (2017) CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355
44. Gilbert, L.A. *et al.* (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661
45. Rosenbluh, J. *et al.* (2017) Complementary information derived from CRISPR Cas9 mediated gene deletion and suppression. *Nat. Commun.* 8, 15403
46. Sack, L.M. *et al.* (2018) Profound tissue specificity in proliferation control underlies cancer drivers and aneuploidy patterns. *Cell* 173, 499–514 e23
47. Joung, J. *et al.* (2017) Genome-scale activation screen identifies a lncRNA locus regulating a gene neighbourhood. *Nature* 548, 343–346
48. Bester, A.C. *et al.* (2018) An integrated genome-wide CRISPRa approach to functionalize lncRNAs in drug resistance. *Cell* 173, 649-664 e20
49. Tanenbaum, M.E. *et al.* (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
50. Chavez, A. *et al.* (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* 12, 326–328
51. Konermann, S. *et al.* (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588
52. Chavez, A. *et al.* (2016) Comparison of Cas9 activators in multiple species. *Nat. Methods* 13, 563–567
53. Heaton, B.E. *et al.* (2017) A CRISPR activation screen identifies a pan-avian influenza virus inhibitory host factor. *Cell Rep.* 20, 1503–1512
54. Morita, S. *et al.* (2016) Targeted DNA demethylation *in vivo* using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* 34, 1060–1065
55. Berger, A.H. *et al.* (2016) High-throughput phenotyping of lung cancer somatic mutations. *Cancer Cell* 30, 214–228
56. Joung, J. *et al.* (2017) Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nat. Protoc.* 12, 828–863
57. Komor, A.C. *et al.* (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424
58. Kuscu, C. *et al.* (2017) CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* 14, 710–712
59. Hess, G.T. *et al.* (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13, 1036–1042
60. Kuscu, C. and Adli, M. (2016) CRISPR-Cas9-AID base editor is a powerful gain-of-function screening tool. *Nat. Methods* 13, 983–984
61. Ma, Y. *et al.* (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13, 1029–1035
62. Kim, Y.B. *et al.* (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* 35, 371–376
63. Gaudelli, N.M. *et al.* (2017) Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* 551, 464–471
64. Boettcher, M. *et al.* (2010) Decoding pooled RNAi screens by means of barcode tiling arrays. *BMC Genomics* 11, 7
65. Boettcher, M. and Hoheisel, J.D. (2010) Pooled RNAi Screens – technical and biological aspects. *Curr. Genomics* 11, 162–167
66. Doench, J.G. (2018) Am I ready for CRISPR? A user's guide to genetic screens. *Nat. Rev. Genet.* 19, 67–80
67. Mohr, S.E. *et al.* (2014) RNAi screening comes of age: improved techniques and complementary approaches. *Nat. Rev. Mol. Cell Biol.* 15, 591–600
68. Hart, T. *et al.* (2017) Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* 7, 2719–2727
69. Kampmann, M. *et al.* (2015) Next-generation libraries for robust RNA interference-based genome-wide screens. *Proc. Natl. Acad. Sci.* 112, E3384–E3391
70. Sanjana, N.E. *et al.* (2016) High-resolution interrogation of functional elements in the noncoding genome. *Science* 353, 1545–1549
71. Strezoska, Z. *et al.* (2012) Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens. *PLoS One* 7, e42341
72. Herkert, B. *et al.* (2016) Maximizing the efficacy of MAPK-targeted treatment in PTENLOF/BRAFMUT melanoma through PI3K and IGF1R inhibition. *Cancer Res.* 76, 390–402
73. Taxman, D.J. *et al.* (2006) Criteria for effective design, construction, and gene knockdown by shRNA vectors. *BMC Biotechnol.* 6, 7
74. Wang, T. *et al.* (2017) Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic Ras. *Cell* 168, 890–903 e15

75. Morgens, D.W. *et al.* (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* 34, 634–636
76. Rosenbluh, J. *et al.* (2016) Genetic and proteomic interrogation of lower confidence candidate genes reveals signaling networks in beta-catenin-active cancers. *Cell Syst.* 3, 302-316 e4
77. Shalem, O. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343, 84–87
78. Evers, B. *et al.* (2016) CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* 34, 631–633
79. Tsherniak, A. *et al.* (2017) Defining a cancer dependency map. *Cell* 170, 564–576 e16
80. Pan, J. *et al.* (2018) Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. *Cell Syst.* 6, 555–568 e7
81. Rauscher, B. *et al.* (2018) Toward an integrated map of genetic interactions in cancer cells. *Mol. Syst. Biol.* 14, e7656
82. Hart, T. *et al.* (2017) Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* 7, 2719–2727
83. Zhong, R. *et al.* (2014) Computational detection and suppression of sequence-specific off-target phenotypes from whole genome RNAi screens. *Nucleic Acids Res.* 42, 8214–8222
84. Roy, K.R. *et al.* (2018) Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.* 36, 512–520
85. Bao, Z. *et al.* (2018) Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nat. Biotechnol.* 36, 505–518
86. Chen, J.S. *et al.* (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550, 407–410
87. Abudayyeh, O.O. *et al.* (2017) RNA targeting with CRISPR-Cas13. *Nature* 550, 280–284
88. Cox, D.B.T. *et al.* (2017) RNA editing with CRISPR-Cas13. *Science* 358, 1019–1027
89. Tak, Y.E. *et al.* (2017) Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat. Methods* 14, 1163–1166
90. Kweon, J. *et al.* (2017) Fusion guide RNAs for orthogonal gene manipulation with Cas9 and Cpf1. *Nat. Commun.* 8, 1723
91. Najm, F.J. *et al.* (2018) Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* 36, 179–189
92. Boettcher, M. *et al.* (2018) Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat. Biotechnol.* 36, 170–178
93. Galonska, C. *et al.* (2018) Genome-wide tracking of dCas9-methyltransferase footprints. *Nat. Commun.* 9, 597
94. Hu, G. and Luo, J. (2012) A primer on using pooled shRNA libraries for functional genomic screens. *Acta Biochim. Biophys. Sin. (Shanghai)* 44, 103–112
95. Platt, R.J. *et al.* (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159, 440–455
96. Hart, T. *et al.* (2014) Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* 10, 733
97. Winter, J. *et al.* (2017) CRISPRAnalyzeR: Interactive analysis, annotation and documentation of pooled CRISPR screens. *bioRxiv* Published online February 20, 2018. http://dx.doi.org/10.1101/109967
98. Winter, J. *et al.* (2016) caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. *Bioinformatics* 32, 632–634
99. Spahn, P.N. *et al.* (2017) PinAPL-Py: a comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci. Rep.* 7, 15854
100. Meyers, R.M. *et al.* (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* 49, 1779–1784
101. Canver, M.C. *et al.* (2018) Integrated design: execution, and analysis of arrayed and pooled CRISPR genome-editing experiments. *Nat. Protoc.* 13, 946–986
102. Allen, F. *et al.* (2018) JACKS: joint analysis of CRISPR/Cas9 knock-out screens. *bioRxiv*
103. Jia, G., Wang, X. and Xiao, G. (2017) A permutation-based non-parametric analysis of CRISPR screen data. *BMC Genomics* 18, 545
104. Trumbach, D. *et al.* (2017) ENCoRE: an efficient software for CRISPR screens identifies new players in extrinsic apoptosis. *BMC Genomics* 18, 905
105. Hart, T. and Moffat, J. (2016) BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* 17, 164
106. Yu, J. *et al.* (2016) ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics* 32, 260–267
107. Li, W. *et al.* (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.* 16, 281
108. Diaz, A.A. *et al.* (2015) HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res.* 43, e16
109. Luo, B. *et al.* (2008) Highly parallel identification of essential genes in cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 20380–20385
110. Konig, R. *et al.* (2007) A probability-based approach for the analysis of large-scale RNAi screens. *Nat. Methods* 4, 847–849
111. 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
112. Dai, Z. *et al.* (2014) edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res* 3, 95