Functional Genomics via CRISPR–Cas

Kyle Ford[1,†], Daniella McDonald[2,†] and Prashant Mali[1]

1 - Department of Bioengineering, University of California, San Diego, San Diego, CA 92093, USA
2 - Biomedical Sciences Graduate Program, University of California, San Diego, San Diego, CA 92093, USA

Correspondence to Prashant Mali: pmali@ucsd.edu
https://doi.org/10.1016/j.jmb.2018.06.034
Edited by Stanley Qi

Abstract

RNA-guided CRISPR (clustered regularly interspaced short palindromic repeat)-associated Cas proteins have recently emerged as versatile tools to investigate and engineer the genome. The programmability of CRISPR–Cas has proven especially useful for probing genomic function in high-throughput. Facile single-guide RNA library synthesis allows CRISPR–Cas screening to rapidly investigate the functional consequences of genomic, transcriptomic, and epigenomic perturbations. Furthermore, by combining CRISPR–Cas perturbations with downstream single-cell analyses (flow cytometry, expression profiling, etc.), forward screens can generate robust data sets linking genotypes to complex cellular phenotypes. In the following review, we highlight recent advances in CRISPR–Cas genomic screening while outlining protocols and pitfalls associated with screen implementation. Finally, we describe current challenges limiting the utility of CRISPR–Cas screening as well as future research needed to resolve these impediments. As CRISPR–Cas technologies develop, so too will their clinical applications. Looking ahead, patient centric functional screening in primary cells will likely play a greater role in disease management and therapeutic development.

Introduction

An ongoing challenge in biology is comprehensively mapping genotype–phenotype relationships. With this objective in mind, functional genomics makes use of data from all levels of biology (genome, transcriptome, epigenome, proteome, metabolome, etc.) to better define genetic and protein functions and interactions. In this way, researching functional genomics is essential for better understanding the human genome and its intricate interactions in healthy, as well as pathophysiologic states. Characterizing the functional consequences of genomic variation is crucial for many aspects of biomedical research including cancer screening methodologies, drug–drug interactions, drug sensitivity and resistance, gene therapy, regenerative medicine applications, infectious disease, and general understanding of human physiology. (See Fig. 1.)

It has become increasingly clear that the volume and complexity of genomic information necessitates rapid screening methodologies. Utilizing large-scale and high-throughput assays, researchers can more quickly map the function of a multitude of genes and/or proteins in parallel. To this end, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins have been utilized to help interrogate and realize functional outputs based on targeted editing strategies. CRISPR–Cas systems are powerful tools for targeted genome editing that have dramatically impacted genomic research and screens since their first mammalian applications in 2013 [1, 2]. This technology has revolutionized the field with its ease, speed, and targeting versatility, allowing for facile genetic perturbations and resulting functional output analysis in a multiplexed fashion. It has allowed for a large number of high-throughput functional genomic screens to be performed which have, in turn, identified key genes involved in a broad range of human health and disease including cancers, infections, immune regulators and responses, and metabolic diseases [3].
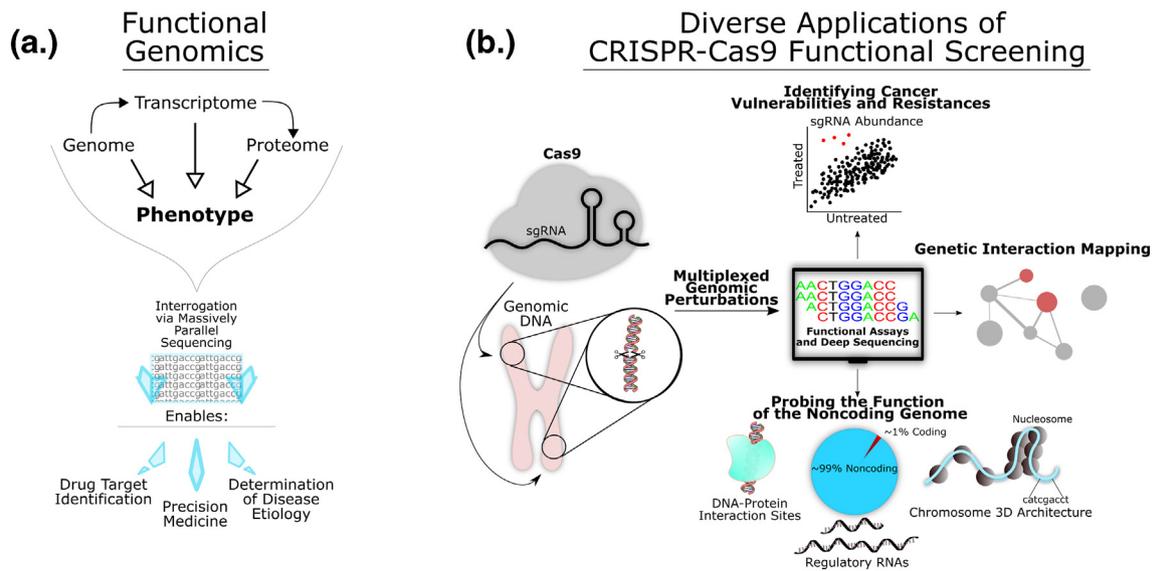
J Mol Biol (2018) **xx**, xxxxxx

Please cite this article as: K. Ford, et al., Functional Genomics via CRISPR–Cas, J. Mol. Biol. (2018), https://doi.org/10.1016/j.jmb.2018.06.034

Fig. 1. Functional Genomics and CRISPR–Cas. (a) The goal of functional genomics is to better understand how the genome informs diverse biological phenotypes. To this end, functional genomics makes use of mass data sets spanning the genome, the transcriptome, and the proteome. The declining cost of massively parallel sequencing platforms has made genome-wide functional screens broadly achievable and economically viable for academic labs of all sizes. (b) CRISPR–Cas9 has made multiplexed functional screening with single-cell resolution more robust than ever before. The ease of sgRNA design has led to accelerated functional mapping of the genome with extensive consequences for medicine and biotechnology. Because sgRNA targeting almost any region of the genome can be designed *in silico*, CRISPR–Cas screens can be rapidly designed and executed. Functional screens using Cas9 have been used for a wide variety of applications, such as identifying novel cancer therapeutics and vulnerabilities, quantifying genetic interactions, and exploring the function of the non-coding genome.

CRISPR–Cas Toolsets

CRISPR–Cas systems are divided into different classes, types, and subtypes. Class 1 utilizes multi-protein effector complexes, and class 2 utilizes single-protein effectors. Class 1 includes types I, III, and IV. Class 2 includes types II, V, and VI. There are a further 19 subtypes and this will likely continue to expand as new CRISPR–Cas systems are identified [4, 5].

The most common Cas protein used in functional screening is a type II single-protein effector derived from *Streptococcus pyogenes* (SpCas9). The SpCas9 uses a guide RNA to assist in effectively cleaving the target gene. Once Cas9 successfully finds a target sequence with proper pairing of the complement guide RNA and an appropriate protospacer adjacent motif (PAM), the endonuclease will cleave the phosphodiester bonds upstream of the PAM forming a double-strand break [6]. When the double-strand break occurs, non-homologous end joining (NHEJ) or homology-directed repair (HDR) will attempt to repair the damage. NHEJ often results in a small insertion–deletion mutation (indel). If targeted to a gene, this may result in a knockout due to generation of a frameshift resulting in a premature stop codon and nonsense-mediated decay of the transcript. NHEJ is often the repair process of choice for mutagenesis. HDR, however, is a templated repair process most commonly recognized for its natural use in the body during gamete formation allowing for genetic recombination. Its use in the cell is restricted to the S and G2 phase [7]. Due to its high fidelity, HDR can be utilized to insert a new custom region into the genome creating knock-ins or specific gene mutations (or corrections) if desired [8]. Increasing the efficiency and utility of HDR is still necessary to fully apply its uses for CRISPR–Cas systems.

Over the last several years, the versatility of CRISPR–Cas systems has increased dramatically. There currently exist Cas9 effector fusions with the ability to modify specific histones, edit particular DNA base pairs, activate or inhibit the transcription of certain genes (CRISPRa/CRISPRi), or effect DNA methylation/demethylation at user determined loci [9]. This wide array of effector functions enables a variety of genomic elements to be probed systematically in a high-throughput fashion.

The dominant method of generating Cas9 variants with novel functions consists of fusing a catalytically inactive Cas9 (dCas9) protein to an effector moiety [10]. In this way, the dCas9 serves only as a DNA targeting platform, which guides the effector moiety to the location of interest in the human genome. The benefit of this design strategy is that it enables rapid development of new dCas9 functionalities due to its

Table 1. Cas9 perturbation options for functional screens

Perturbation choice	Effect on the genome	Mechanism	References
wtCas9	Loss of function and deletions	Double-stranded DNA cleavage at the target locus	[11, 12]
CRISPRa	Transcriptional activation	Fusion of dCas9 to various activating domains (e.g., VP64 or the p65 subunit of nuclear factor kappa B (NF-κB))	[10, 13–15]
CRISPRi	Transcriptional repression	Fusion of dCas9 to domains which inhibit transcription [e.g., Krüppel-associated box (KRAB)]	[10, 14, 16]
Base editors	Catalyze a nucleotide base pair substitution without DNA cleavage	Fusion of dCas9 to enzymes which catalyze nucleobase conversion (e.g., activation-induced cytidine deaminase (AID) for C→T edits)	[17–22]
DNA methylation and demethylation	Cas9-guided DNA methylation and demethylation modifies chromosome structure and subsequent gene transcription	Fusion of dCas9 to DNA (cytosine-5)-methyltransferase 3A (DNMT3A) and ten-eleven translocation (TET) proteins, respectively	[23, 24]
Histone modification	Cas9-guided control of histone acetylation and methylation	Fusion of dCas9 to histone modifying enzymes (e.g., Histone deacetylase 3 (HDAC3), p300 acetyltransferase, or lysine-specific histone demethylase 1A (KDM1A/LSD1)	[25–28]

modularity. However, optimizing the efficacy and off-target effects of novel Cas9 fusions is a laborious undertaking which increases rapidly as the protein engineering search space is expanded. Furthermore, because the effector moiety is fused permanently to dCas9, orthogonal parallel perturbations require the co-delivery of multiple fusion constructs to the cells of interest. This, coupled with the large size of dCas9 fusions, imposes significant delivery challenges limiting their use in functional screens. Nevertheless, dCas9 fusions represent a robust set of tools with which to probe genome function.

The choice of appropriate Cas9 variant will depend heavily on what functionality is being investigated. The broad array of available Cas9-based perturbation systems is summarized in Table 1. While Table 1 includes the most common Cas9-based perturbation choices, it is far from exhaustive.

The wild type Cas9 protein functions as a targeted endonuclease, catalyzing DNA double-stranded breaks [6]. These double-stranded breaks often lead to indels via the error prone NHEJ. Frameshifts resulting from these mutations can knockout the function of protein coding genes, making wtCas9 ideal for loss of function studies [11, 12]. Knockout studies are often used to determine the *essentiality* of genes in high throughput, and simplifies downstream validation and data analysis due to the binary nature of the perturbation. However, this simplification in some ways limits the translational relevance of knockout screening. Although knockouts can inform our understanding of what genes are essential for specific biological processes *in vitro,* there is no guarantee that small-molecule or protein-mediated inhibition *in vivo* will have the same effect. Furthermore, knockout studies fail to recapitulate

gain-of-function mutations and transcriptional dysregulation which play a key role in many pathologies [29–31]. In this way, Cas9 knockout experiments should not be considered a surrogate for drug studies, but rather a parallel set of tools with which interrogate the user's model. For these reasons, knockout screening requires extensive downstream target validation before any significant conclusions can be drawn.

As an alternative to knockout experiments, CRISPRa/i systems use enhancer/repressor proteins fused to dCas9 as a way of modulating gene transcription at particular loci [10, 13, 32]. Because CRISPRa/i functions at the transcriptional level, it enables investigation of genome function without permanently modifying genomic structure. Unlike wtCas9, activation of target genes by CRISPRa can facilitate complex gain-of-function screening from endogenous genomic loci. In addition, CRISPRi can perform loss of function screening without the confounding effects of off target nuclease activity [32]. For even more robust genetic studies, the combination of CRISPR effector functions can generate complementary data sets with which researchers can generate conclusions with greater confidence [14]. As a recent example, by co-delivering both CRISPRa and wtCas9, researchers were able to interrogate the directionality of genetic interactions in high-throughput [33]. However, CRISPRa/i experiments suffer from their own set of limitations. First and foremost is the limited correlation between mRNA levels and protein expression [34]. While CRISPRa/i can reduce or increase the levels of a particular mRNA transcript, protein expression is subject to post-transcriptional regulation which has the potential to obfuscate the perturbations' actual effect [34]. In addition, the CRISPRa/i systems require sgRNAs

targeting the promoter region or transcriptional start site of the gene of interest [15, 16]. Promoter regions and transcriptional start sites can be rendered inaccessible to sgRNA due to chromatin structure or may not have an appropriate PAM sequence nearby, limiting the pool of genes for which CRISPRa/i is effective. In addition, some genes are controlled by multiple functional promoters, further confounding screens using CRISPRa/i. Ideally, these limitations ought to inform the experimental design of CRISPRa/i genomic screens to ensure output data are reproducible and conclusions justifiable.

Functional studies using DNA and Histone-modifying Cas9 fusion constructs operate in a similar fashion to CRISPRa/i [25]. By modifying the structure of DNA/Histones (via acetylation or methylation), these Cas9 fusions vary gene accessibility to transcriptional machinery and consequently gene expression [23, 24, 26]. A key difference is the mechanism underlying these structural perturbations. Whereas CRISPRa/i can modulate gene expression without leaving a scar on the target site, DNA/Histone modifications affect gene expression via lasting structural changes. The choice of perturbation is largely dependent on the nature of the biological question being asked. For probing the function of protein coding genes, CRISPRa/i and CRISPR knockout are well-validated systems with a spectrum of reagents available commercially, enabling a powerful toolset for genome-wide screening. However, if the goal of the experiment is mapping chromosomal structure–function relationships, the DNA/Histone epigenetic modifiers may be a more fitting choice. Several groups have used these DNA/Histone-modifying Cas9 variants to probe how chromosomal chemical structure and 3D architecture controls gene regulation through diverse mechanisms of action [27, 28]. Nevertheless, DNA/Histone-modifying Cas9 variants are not the only way to perturb chromosomal structure. Deletions and chromosomal rearrangements induced by wtCas9 have also been used to explore how structural variation in the human genome impacts nearby gene function [35].

In contrast with wtCas9, CRISPRa/i, and Cas9-based structural modifiers, CRISPR base editing constructs have recently been developed as novel tools for functional genomic screens. CRISPR base editors work by modifying individual nucleic acid base pairs within the target genes in a precise, or pseudo-random manner [17]. These systems function by fusing a cytidine deaminase or an adenosine deaminase to dCas9 to effect C→T mutations or A→G mutations, respectively [18, 19]. These novel systems represent a versatile avenue with which to model gain or loss-of-function mutations in an endogenous context [20–22].

Engineered sgRNAs have also been explored as an alternative way to impart novel function to the Cas9 system [36]. By incorporating protein binding RNA aptamers (PP7, MS2, etc.) into the sgRNA structure, Cas9 can recruit orthogonal proteins with a variety of functionalities. Because the perturbation choice is encoded in the sgRNA itself, multiple perturbation types can be explored in the same pooled screen using unmodified dCas9. This system has been used to effect multiplexed gene activation and interference in parallel (via sgRNA modified to recruit vp64 and KRAB, respectively) as well as perform multiplexed fluorescent labeling of specific genomic loci [37, 38].

Genomics Screens

The use of the CRISPR–Cas systems has many implications for functional genomics and has been the topic of much excitement. Functional screens, in turn, are typically performed in an arrayed or pooled format, and rely equally on three integral ingredients: a perturbation, a model, and an assay. In an arrayed screen, the reagents are added into a multi-well plate so that one reagent or a small pool is added to each well allowing for a single perturbation per well. Because each well will contain a population of cells with identical genomic perturbations, a wider array of phenotypic data can be assayed simultaneously (proteomics data, functional assays, tissue-level phenotypes, etc.) without limitation to growth phenotypes. Furthermore, arrayed screening precludes any paracrine-mediated cell–cell interactions which may obscure the effects of individual perturbations. Unfortunately, this arrayed format is significantly more expensive to perform and lower throughput [11]. Arrayed library screening often requires specialized automation for cell culture due to the need to culture large quantities of cells in isolation from one another [39]. These challenges have typically limited the widespread adoption of high-throughput arrayed screening to the biopharmaceutical industry. Because of this, pooled screening has rapidly become a key method of probing genome elements using Cas9. Pooled screens involve testing thousands of genetic perturbations in a single assay and have become increasingly popular over the past decade. Pooled screens allow for massive libraries of gene targets to be investigated in a single-cell culture dish, accelerating the process of functional screening. However, pooled screens are somewhat limited in the output data they can reliably produce. Because each cell in the dish will have a unique sgRNA delivered to it, only measurements with single-cell resolution [next-generation sequencing (NGS), fluorescence-activated cell sorting (FACS), etc.] can be used to quantitate the effect of the perturbations. Harnessing CRISPR–Cas systems effectively allows for a library of perturbations (sgRNA targeting a particular locus) to be performed in a cell population either in the arrayed or pooled format via typically lentiviral transduction. Cells successfully transduced

with the perturbation must then be selected for by some means (e.g., drug resistance, FACS). Follow-up assays are then performed to help delineate which perturbations caused which functional phenotypic changes. This can be done through multiple means either by high-content imaging (HCI) or through NGS [40–43]. HCI is beneficial for arrayed screens, allowing for quantification of spatially or temporally resolved images. This allows for a large output of phenotypic measurements while visualizing the biology. NGS is the high-throughput sequencing of DNA and RNA that performs quicker and cheaper than Sanger sequencing with the ability to quantitate reads. Massively parallel sequencing has helped revolutionize the study of functional genomics and molecular biology. In earlier years, identifying the causal mutations that led to functional changes would have been costly and labor intensive. With the advent of NGS platforms, mapping such mutations can be achieved quickly and with less costly streamlined protocols. Because of this, NGS has helped fuel pooled screens at a rapid pace. NGS enables single-molecule DNA quantitation and readout of library population dynamics. Thus, a quantification can be made on the proportion of uniquely integrated library constructs in the population of cells while assessing cell viability to determine which genes after being perturbed are enriched and/or depleted. To ensure the screen results are reproducible, it is critical to validate the top hits identified from the pooled screen using an arrayed screen, preferably selecting additional sgRNAs targeting similar genes. Further biological assays should also be performed to confirm top candidates [44, 45]

Although there are many diverse CRISPR tools, their use in genome-scale functional screening is relatively conserved. Rather than isolating a trait and investigating what in the genome causes that phenotype, Cas9 screens function by perturbing the genome and measuring the subsequent change in a phenotype of interest. A common example of the former would be The Cancer Genome Atlas (https://cancergenome.nih.gov/). This massive research effort attempts to determine the genomic etiology of cancer through mass sequencing of patient cancer samples (phenotype→genotype). Cas9 genetic screening inverts this protocol. By purposefully introducing a genomic perturbation with Cas9, the resulting trait can be recorded and genotype–phenotype relationships mapped.

The primary benefit of screening with Cas9 (or other CRISPR–Cas effectors) is the throughput. Rapid screening with Cas9 is made possible by the ability to perturb multiple parallel targets in the genome via a library of sgRNA. The declining cost of DNA synthesis (<1 cent/nucleotide) has enabled academic labs to construct these genome scale sgRNA libraries at low costs and with relatively low error rates, spurring Cas9's widespread adoption [46–48].

Cas9 genetic screening has most frequently been applied to screening various cancer cell lines (https://portals.broadinstitute.org/achilles) [12, 49]. Cancer cell lines have several features which make them ideal for Cas9 screening. Unlike many primary cells, cancer cell lines grow well *in vitro* and can be expanded to large numbers. This is necessary to effectively screen large genome scale libraries with proper coverage [9]. Furthermore, immortalized cancer cell lines can be genetically modified to constitutively express Cas9 from a stable location in their genome, obviating the challenge of delivering the Cas9 protein in the screen. Because Cas9 is expressed in every cell being screened, only the much smaller sgRNA constructs need to be delivered. Consequently, constitutive Cas9 expression enables simplified delivery of the sgRNA library resulting in typically higher perturbation efficiencies (albeit with greater off-target rates) [50]. However, this workaround is not feasible when studying primary cells, which require the co-delivery of Cas9 and sgRNA. In addition to providing many procedural benefits, screening in cancer cell lines is often performed to identify cancer specific genetic vulnerabilities. Mapping how genomic perturbations affect cell fitness can be used to circumvent drug resistances, as well as understand underlying genetic polymorphisms driving cancer growth [12, 49, 51].

However, CRISPR–Cas screening is not limited to just cancer research. The diversity of Cas9-based tools and the ease of sgRNA cloning has enabled the interrogation of genomic function across many disparate areas of biology. In principle, the genetic basis of any biological phenotype can be investigated using CRISPR–Cas perturbation screening, provided the phenotype of interest can quantitated. For instance, screening with Cas9 has shown great utility in the study of infectious diseases [52, 53]. By perturbing the target cells with libraries of sgRNA before infection with the pathogen of interest, researchers can identify genes regulating susceptibility and resistance to an infectious disease. Alternatively, the genome of the pathogen itself can be the target of CRISPR–Cas perturbations to identify essential genes controlling pathogenesis. In this way, functional screening with CRISPR–Cas can provide key information on the critical role host and pathogen genetics play in disease progression. These data can then be used to help determine new molecular targets for drug development and better understand the genetic basis of divergent responses to existing therapeutics [52]. For example, several groups have recently applied Cas9 functional screening to the study of HIV, malaria, and tuberculosis, identifying critical genetic host factors as well as essential genes regulating infection within the genomes of pathogenic viruses and bacteria [54–56]. These are only a small set of potential screening applications, and future work will assuredly involve

expanding the use of CRISPR–Cas to a greater number of novel biological problems.

Library Design and Synthesis

The first step in developing a genomic screen using Cas9 is identifying what genomic loci to perturb. Genome-wide Cas9 screens are increasingly popular due to their relatively unbiased interrogation of genome function. That being said, the choice of which genomic targets to perturb is primarily determined by the researcher's own personal interest. Regardless of what genes are perturbed, there are several key library design considerations that are universally relevant. (See Fig. 2.)

Nearly every gene (and non-coding region) can be considered a potential target, although the endonuclease activity of Cas9 is limited to sequences with an adjacent PAM motif (NGG for SpCas9). However, recent efforts to engineer Cas9 variants which tolerate expanded PAM sequences indicate that this barrier will not be a long-term impediment [57]. Many *in silico* tools are available to facilitate rapid guide RNA design, enabling large libraries of guide RNA to be designed efficiently [58].

Targeting a large library of sequences enables higher-throughput interrogation of genomic elements, while a small library of genomic perturbations will lend results greater accuracy due to better library coverage [9]. The theoretical max library size is limited by several factors. DNA synthesis is an inherently error prone process itself, increasing the likelihood of inaccurate synthesis at high library size [46]. Furthermore, researchers are limited by the amount of DNA they can effectively introduce to both bacterial and mammalian cells. While libraries of greater than 10^7 sgRNAs can be easily transformed and maintained in bacteria for DNA production, the sheer number of mammalian of cells required to screen such a large library serves as a practical limit to the library search space [9, 59]. Because of this, libraries greater than ~100,000 sgRNAs often require cells to be grown in large-scale cell culture setups or bio-reactors.

After choosing what genomic elements to study and how to perturb them, the library of sgRNA needs to be synthesized. There currently are a wide variety of premade sgRNA libraries available for purchase, ranging from genome-wide libraries with ~10^5 sgRNAs, to more targeted libraries focused on single pathways on gene families [12, 49, 60]. This is often the simplest option for many labs, but limits researchers to preselected gene targets which may be irrelevant to their study. Alternatively, custom sgRNA libraries can also be generated via commercial chip based DNA synthesis [48]. This allows researchers to preselect a curated library of genomic elements for perturbation, facilitating the development of more precise experiments.

Delivery Systems

Choice of delivery of the CRISPR–Cas reagents is key for high editing efficiencies, proper cell uptake, reduced off-target effects, and large cargo capacities. The advantages and challenges of these different methods are outlined in Table 2.

The choice of delivery method is important and should be catered to the unique needs of the experimental screen being run dependent on if it is an arrayed or pooled screen, cells being used, and cargo size. Standard delivery for most screening applications is viral, specifically lentivirus [61–64]. There are many advantages to utilizing lentivirus. It is a retrovirus with the ability to integrate into dividing and non-dividing cells thus, creating stable transductions that can later be read via NGS. This ability also makes lentiviral transduction ideal for delivery to primary cells that are notorious for being difficult to transfect. Lentivirus is also beneficial for large gene or multiple gene cassette deliveries with its large cargo capacity [65]. One study utilized a lentiviral vector library in human cells to identify the key genes that contribute to the intoxication of cells by anthrax and diphtheria toxins [64]. The benefits of being able to stably transduce a variety of cell types easily and quickly have ensured the continued use of lentivirus in screens.

A few studies have more recently looked at utilizing viruses for screens that do not integrate into the host genome such as the adeno-associated virus (AAV). The idea to use AAVs for functional screens is novel and somewhat limited, but could allow for functional screening of tissue-level phenotypes *in vivo*. This is of great value because much of the data sets obtained from *in vitro* screens need to be taken with some amount of skepticism. There is not true physiologic representation in a dish, meaning that the results of *in vitro* screens require rigorous validation. *In vivo* screening could help circumvent some of these issues, obtaining phenotypic outputs from a screen that was performed in live animals. One such study utilized the AAV to develop a unique *in vivo* CRISPR screen in conditional-Cas9 mice [66]. This study screened 49 genes known to be tumor suppressing with 5 sgRNAs for each gene. These guides were engineered into AAVs to allow for direct *in vivo* delivery into the lateral ventricle of immunocompetent living mice. Mice grew glioblastomas over time and whole brains were then homogenized to perform downstream analyses at the DNA, RNA, and protein level. The largest obstacle to overcome with this study was sequencing which tumors received which gene knockouts as the AAVs do not integrate into the host genome. This study designed probes to target-capture the predicted sequences of interest where expected gene knockouts would occur. This complex capture sequencing technique successfully could determine which tumors received which gene knockouts and follow up with multiple phenotypic metrics.
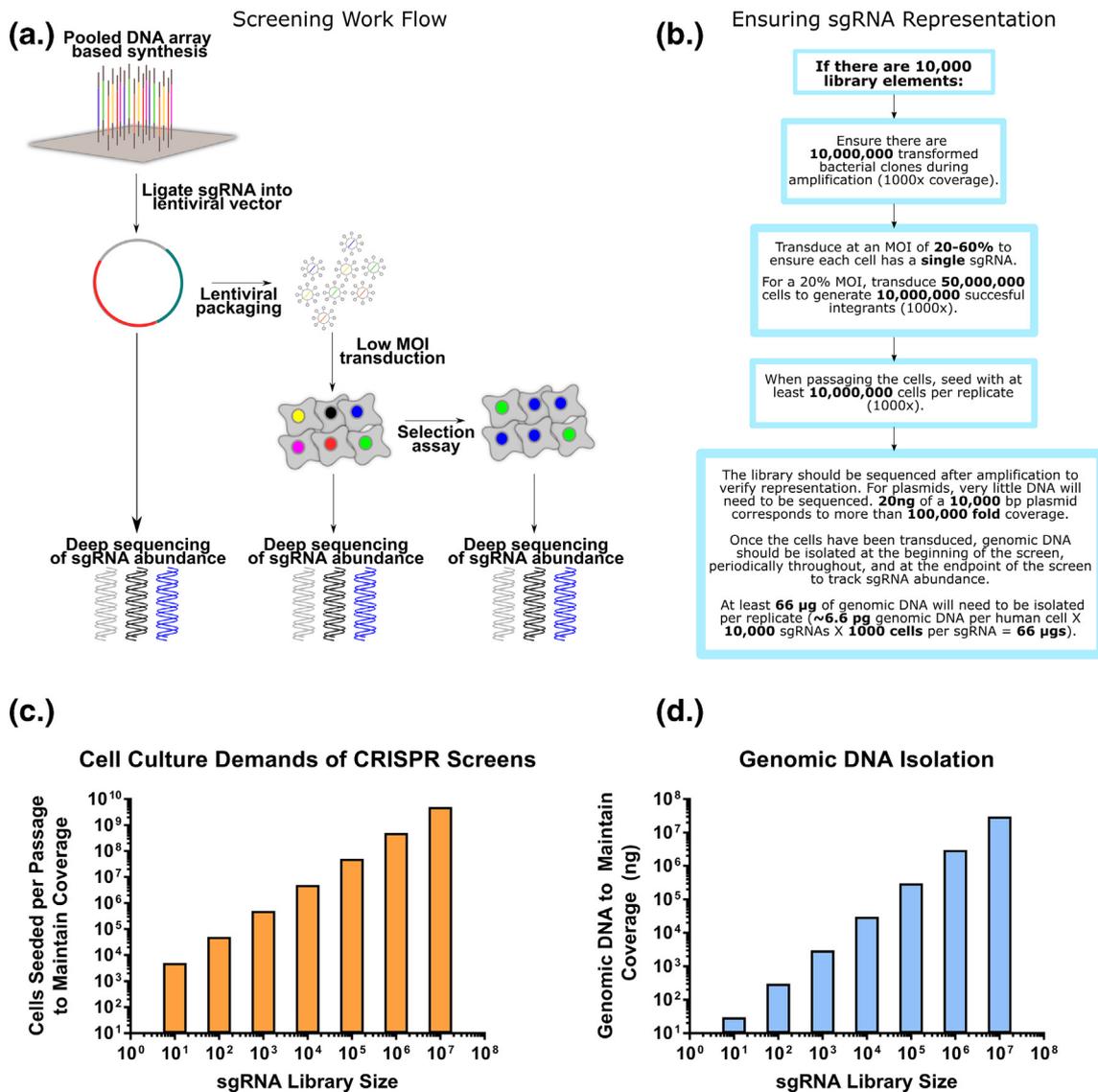
Fig. 2. Mechanics of CRISPR–Cas screens. Panels a and b show the key steps in performing a CRISPR screen in mammalian cells. Initially, the sgRNA library is ordered as a pooled tube of DNA oligonucleotides, typically synthesized commercially via chip-based DNA synthesis. The library is then amplified via PCR and cloned into an appropriate lentiviral vector, ensuring library coverage is maintained throughout. If the library is obtained in plasmid form (e.g., pooled sgRNA libraries available from Addgene), the library simply needs to be transformed into bacteria, expanded, and sequenced to confirm sgRNA representation. Once the library is in a suitable lentiviral vector, the next step is packaging the DNA into lentivirus. Standard lentiviral packaging protocols will suffice, so long as coverage is maintained throughout the packaging. After packaging the lentivirus, a test transduction should be performed to quantify the functional titer (i.e., the actual number of cells transduced per lentiviral particle delivered). This can then be used to determine the amount of lentivirus needed to achieve an MOI of 20%–60%. The transduced cells are then passaged with at least 500- to 1000-fold coverage of the library at each step to ensure accurate sgRNA quantitation. As the cells are passaged, it also is beneficial to store freeze and store aliquots of the library for subsequent massively parallel sequencing. At the end of the functional assay, the library is sequenced a final time to determine the relative enrichment and depletion of specific sgRNA, corresponding to target gene fitness. (c and d) Maintaining library coverage throughout the protocol is essential for insuring statistical confidence and preventing arbitrary library skewing. However, maintaining high coverage of the library imposes significant practical challenges for researchers attempting to implement a CRISPR–Cas screen. The figures above highlight the technical challenges of large library screening and can serve as a reference for future screen design (bar plots calculated assuming 500 fold coverage of the library). As the number of sgRNA in the library increases, the scale of the experiment may outpace available resources and become untenable. Correspondingly, when planning a CRISPR–Cas genetic screen, it is important to determine if the screen is executable in terms of lab equipment, reagents, and manpower. Once the screen has been started, the same mindfulness needs to be directed at ensuring there are no library bottlenecking points which could artificially influence the results of the assay.

Table 2. Advantages and disadvantages of different CRISPR–Cas delivery systems

Delivery method	Advantages	Disadvantages	References
Lentivirus	- Stable gene expression - High transduction efficiency - Good for difficult-to-transfect cells (primary cells) - Large cargo capacity	- Not ideal for *in vivo* delivery	[61–65]
AAV	- High transduction efficiency - Low cytotoxicity - Relevant for *in vivo* screens	- Limited cargo capacity (4.7 kb) - Expensive	[66, 67]
Electroporation	- High transfection efficiency - Good for difficult-to-transfect cells (primary cells) - Beneficial for RNP delivery	- High cytotoxicity - Limited to arrayed screens	[65, 68–70]
Lipid nanoparticles	- Low cost - Easy handling - Beneficial for RNP delivery	- Low transfection efficiency - Highly dependent on cell type - Limited to arrayed screens	[65, 68]
piggyBac transposon	- Stable gene expression	- Potential for off-target effects - Limited scalability in pooled formats	[71, 72]
Gold nanoparticles	- High transfection efficiency - Large cargo capacity - Less off-target effects - Beneficial for RNP delivery	- Limited to arrayed screens	[68, 73, 74]

More studies like this need to be emphasized in future research to truly recapitulate physiologic conditions during a screen. AAVs, however, cannot be utilized in *in vitro* screens because as cells divide the AAV will be diluted out and NGS studies that rely on genome integration could not be performed. Using clever tactics like targeted-capture sequencing as mentioned prior or reading the viral episome are possible strategies to help circumvent some of these issues for *in vivo* screening methodologies specifically. Another barrier with AAV usage is their limited cargo capacity. The cargo must be less than 4.7 kb and SpCas9 alone is encoded by a 4.2-kb sequence [67]. Utilizing conditional-Cas9 animals would be key for *in vivo* screening applications with AAVs. Other studies have performed *in vivo* screens utilizing lentiviral transduction of cancer cells *in vitro*, followed by transplantation into a mouse [75]. This simplifies downstream NGS analysis due to the integrated guides in the genomes of cell transplants.

There are also many non-viral delivery methods in place that are not frequently used, but could be useful for arrayed screens performed in multi-well plates. For non-viral delivery, because the sgRNA is not stably integrated into the target cells, an arrayed format is necessary to track which cells received which sgRNA. These methods often deliver the reagents either as mRNA or as ribonucleoprotein (RNP) complexes via electroporation or lipid nanoparticles [65, 68]. RNPs specifically have become a powerful perturbation modality and an important tool for arrayed screening especially in primary cells. One group engineered CD4(+) T cells via electroporation using Cas9 RNPs [69]. Forty percent of their cells were successfully engineered to lack the high-expressing cell surface receptor CXCR4 which is a known co-receptor involved in HIV entry into CD4(+) T cells. They

further combined this technology with HDR to perform successful knock-ins at an efficiency of 20%. This group also more recently used Cas9 RNPs to disrupt the programmed cell death protein 1 (PD-1) in chimeric antigen receptor (CAR) T cells enhancing anti-tumor efficacy [70]. Another effective way to introduce Cas9 and/or sgRNA into cells, and of particular benefit to functional pooled screens, is utilizing a piggyBac transposon system [71]. The piggyBac transposon system is a "cut and paste" mechanism, and during transposition, the PB transposase will recognize inverted terminal repeat sequences flanking the end of a transposon vector and then move those contents and integrate them into TTAA sites on the host's DNA. This allows for creating stable cell lines. One study effectively used the piggyBac system to perform an *in vivo* CRISPR library screen utilizing PB sgRNAs in mice looking at tumorigenesis [72]. Creating an inducible Cas9 cell line with this system would be beneficial for screens and then subsequently add the pooled sgRNA library of choice. Cas9 can then be selectively turned on via doxycycline to limit off-target effects. There have also been further developments in novel ways to introduce CRISPR–Cas reagents into cell types to improve efficiency, reduce off-target effects, and increase cargo capacities such as the use of gold nanoparticles [68, 73, 74]. However, additional benchmarking of these non-viral delivery methods is needed to determine what screening application they are most suited for.

Library Transduction and Maintenance

Due to the size of the Cas9 protein as well as the need to co-deliver sgRNAs, a large amount of payload must be delivered to cells to effectively perturb them.

In response to these delivery challenges, lentiviral gene delivery has emerged as the primary method for delivering the sgRNA library to cells, facilitated by the virus's high genetic capacity and broad tropism [9, 11, 76].

After identifying target genes and synthesizing the library of sgRNAs, the next step is ligating them into an appropriate lentiviral vector.

This ligation is the first of many potential bottlenecks where it is important to maintain coverage of the library (typically 500–1000× or more) [9]. To effectively screen a large library of gene targets with confidence, adequate representation of the library elements is key. After packaging the library of sgRNAs into lentivirus, the target cells are then transduced at a low multiplicity of infection (MOI), typically 20%–60% [9, 11]. The transduction is carried out at a low MOI to ensure each cell in the screen receives a single sgRNA. The cells are then routinely passaged, ensuring at least 500–1000× library representation each passage. This high coverage is used to limit false positives and negatives due to erroneous library skewing [9]. As they grow, the cells are then assayed to physically isolate cells displaying the phenotype of interest.

Data Outputs

The simplest form of output data obtainable from a CRISPR screen comes from cell growth and viability assays. Because the sgRNA is genetically encoded into the cell via lentiviral transduction, NGS enables analysis of the library population dynamics. In this way, the sgRNA a cell receives both causes the genetic perturbation and functions as a unique barcode to determine through sequencing how the population is evolving in response to the screen conditions. This method of determining perturbation effects vis-à-vis sgRNA abundance is especially suited for investigating cancer cell fitness and gene essentiality. For example, in a CRISPR knockout fitness, screen-enriched sgRNAs indicate that their target genes are nonessential or antithetical to growth. In the same way, sgRNAs that are depleted at the end of the screen indicate that their target genes are essential for cell growth under the assay conditions. Using this protocol, groups have mapped novel synthetically lethal genetic interactions, investigated how particular genes affect cancer cell drug resistance, and explored how key genes impact the efficacy of immune checkpoint blockers [15, 75, 77].

While fitness-based screening assays (to probe drug resistance or otherwise) are the simplest Cas9 screens to perform, there exist creative workarounds to probe diverse cell phenotypes independent of growth rate in a pooled format. Using an engineered fluorescent reporter system, one group utilized CRISPR screening to investigate the unfolded protein response. This pooled screen used an mCherry transcriptional reporter of IRE1α activation to facilitate cytometric isolation of cells with an activated unfolded protein response, thus enabling the enrichment of a unique phenotype separate from growth rate [78]. Utilizing similar methods, researchers have been able to quantitate how genomic perturbations affect diverse cellular processes such as protein stability and the innate immune response [79, 80]. However, FACS analysis is limited to predetermined targets that have fluorescently labeled antibodies commercially available, or to genetically encoded fluorescent reporter systems.

After isolating cells with the phenotype of interest in a pooled screen, the data output from CRISPR screens is not limited to simply measuring sgRNA abundance. Advancements in single-cell RNA sequencing have made it possible to analyze the transcriptome of thousands of single cells utilizing a unique barcoding strategy [81]. By associating a unique barcode with each cell's transcriptome, CRISPR perturbations can be tracked and associated with transcriptomic signatures [82–84]. This enables researchers to identify (on a cell-by-cell basis) the effect of unique perturbations on the gene expression profile of a cell, and determine clusters of perturbations that may function through similar mechanisms. Unfortunately, the throughput of single-cell RNA sequencing is currently not amenable for large genome-scale libraries. As the cost per cell of single-cell RNA sequencing decreases, this method will likely become more ubiquitous.

In contrast, when performing an arrayed screen, the user is not limited to data outputs with single-cell resolution. Since each unique sgRNA is physically separated from the onset of the screen, traditional RNA sequencing (using cDNA isolated from many cells) can be performed to analyze the effect of a given perturbation on gene expression. Furthermore, in an arrayed format, HCI can be used to examine the impact of a perturbation on cell morphology, cellular processes, and tissue-level phenotypes [40]. This gives arrayed screening a much wider set of phenotypes which can be examined, albeit at much lower throughputs.

Bioinformatic Analysis of Screening Results

At the conclusion of a standard pooled CRISPR screen, the user will have a set of sequencing data representing sgRNA abundances. These raw sequencing data correspond to which genetic perturbations are enriched or depleted for the phenotype of interest. Fortunately, there are many well-validated bioinformatics tools with which to analyze this sequencing data and generate relevant conclusions. Before getting involved in design packages and computational pipelines, it is wise to perform some manual examination to identify possible outliers or mislabeled samples. This vital information could be lost if a cut-and-paste data dump into a statistical tool is performed too quickly. In addition, the user should

manually average the effect of multiple sgRNAs targeting one gene to compile a preliminary list of top hits. If multiple sgRNAs targeting the same gene rank highly, that gene can be listed as a hit.

After these initial steps have been taken, the user can perform a more complete in-depth analysis using a wide array of design packages. Picking the proper statistical package for the user's needs is key. Many factors must be accounted for in addition to identifying sgRNAs that are significant. Most screens typically have little to no replicates which can be a potential setback when trying to estimate the variance of reads in addition to statistical significance between treatments and controls. In addition, researchers must utilize a computational tool that takes sgRNA variability into account in terms of specificities and efficiencies. Finally, knockout screens often result in only a few sgRNAs that tend to dominate the reads in positive selection. A successful algorithm will require robust read normalization. Some older algorithms such as baySeq, DESeq, edgeR, and NBPSeq have been used with some success [85–88]. They are commonly used algorithms for RNA-seq analysis, but limited to the sgRNA level in terms of statistical significance of hits.

Some of the more common tools for pooled screens that show robust results are MAGeCK, caRpools, and CRISPRcloud [89–91]. In brief, MAGeCK robustly identifies positively and negatively selected sgRNAs and genes simultaneously in genome-scale CRISPR–Cas9 knockout screens. Its four steps include read count normalization, mean-variance modeling, sgRNA ranking, and finally gene ranking. Interestingly, MAGeCK can assess relevant biological pathways by reporting positively and negatively selected pathways based on gene rankings in the pathway. This algorithm has been shown to outperform existing methods with its high sensitivity and low false discovery rate [89]. In addition, there is now MAGeCK-VISPR, which was developed for quality control and visualization of CRISPR screens [92]. CaRpools is a user-friendly R package that does not require prior programming knowledge. CaRpools provides the user with biological information for every hit with external links to databases. This package incorporates screening documentation into the analysis process to generate a comprehensive report. CRISPRcloud uniquely allows the user to deposit sequencing files confidentially and analyze them in a cloud-based online system.

Arrayed screens analyze more advanced phenotypes than simply growth and thus often utilize HCI. The vendors for many of these HCI platforms provide their own statistical packages for analysis. The largest challenge with these packages is they require extensive user interaction and can often lack statistical power as the data return from HCI is rich. Many packages are available and have been reviewed [93]. A few common open-source ones are CellProfiler and EBImage [94, 95]. Commercial software is available as well such as Columbus or MetaXpress. After features have been measured and collected with imaging software, these data must be analyzed for statistical significance. Statistical packages for R are commonly used such as *cytominer* (https://github.com/CellProfiler/cytominer/) to assess morphological cell features.

When looking at combinatorial screens, the user must assess the phenotypic effect when a combination of sgRNAs targets the same cell. The initial combinatorial studies were performed in yeast in mass arrays known as synthetic genetic arrays where a gene deletion could be crossed systematically with a deletion mutant array that contains all possible knockout ORFs in the genome [96]. More recently groups have scaled up this technology utilizing CRISPR–Cas for *de novo* mapping of genetic interactions in mammalian cells [77, 97]. This requires additional statistical packages such as the dual CRISPR software pipeline constructed from Python, R, and Jupyter Notebooks (http://ideker.ucsd.edu/papers/rsasik2017/) [98]. Other tools are also available such as TOPS, which is another open-source package to analyze and visualize data from functional genomic gene–gene and gene–drug interaction screens [99].

Single-cell screens have benefited greatly from the Seurat pipeline (http://satijalab.org/seurat/) [100]. Seurat is an R package designed to analyze single-cell RNA-seq data. This package uses canonical correlation analysis to determine shared correlation structures across data sets. After alignment, cells are transposed on a 2D plot (i.e., t-SNE) into clusters with shared transcriptomic reads. Clustering can identify cell types across conditions looking at shifts and cell-specific transcriptomic responses. Seurat allows users to identify and interpret sources of heterogeneity at the single-cell transcriptomic level.

Validating Results

CRISPR–Cas genome-wide screening is valuable because it provides an unbiased way to probe genome function, but the screen is only the first step in identifying functional genomic elements. After identifying potential genes of interest via a perturbation screen and subsequent bioinformatics analysis, significant work must be done to validate these targets. In this way, CRISPR–Cas genome-wide screening can be thought of as hypothesis generating experiments, which guide future genomic characterization efforts.

Initial validation is focused on ensuring the effects of the perturbations are consistent and reproducible. To this end, CRISPR screens often utilize multiple sgRNAs targeting each genomic element [60, 101]. Ideally, one would expect all sgRNAs targeting the same gene to have similar phenotypic effects. This

redundancy provides researchers with a way to ensure that the hits identified from the screen are due to the intended sgRNA mediated genetic perturbation, rather than off-target effects or random noise. Beyond that, potential hits can be sub-screened in a smaller more focused library [51]. This step provides researchers with greater confidence in their results, and helps narrow down target genes for further biological analysis. New sgRNAs targeting potential genes of interest can also be designed and used to verify reproducibility [102]. Furthermore, it can be informative to analyze data sets with different perturbational technologies (CRISPR, CRISPRi, RNAi) to ensure the data are reproducible across multiple systems [102]. However, each of these perturbations will have their own unique biases and limitations which may affect the reproducibility of data across different systems [9].

After several top hits have been established, a key validation step is checking the effects of the sgRNA of interest individually, outside of the context of the pooled screen, to remove any confounding paracrine effects. At the same time, if the gene of interest is protein coding, a Western blot can be used to ensure the gene is completely knocked out by its cognate sgRNA [102]. To generate further confidence in top hits, Cas9 can also be used to generate a clonal population of cells with identical genetic perturbations. Genotyping of this clonal population should then be performed to ensure the gene of interest is effectively knocked out via frame shifts or the introduction of stop codons. After establishing the clonal cell line, robust phenotypic data can be collected to fully interrogate the functional role of the gene of interest. The ultimate step in verifying the effect a gene has on cell phenotype is to restore gene function in the knockout cell line via delivery of cDNA encoding the gene of interest [103]. If the gene of interest is truly the cause of the phenotypic change, cDNA delivery should restore the wild type phenotype to the knockout cell line. If necessary, researchers can also begin testing the perturbation in multiple cell types. While genotype–phenotype relationships may not be consistent across multiple cell types, this step can provide a way to better understand the biology underlying the phenotypic effect of the genetic perturbation [9]. As well, small molecules or monoclonal antibodies targeting the gene(s) of interest can serve to verify the biological mechanism underlying the effect of the perturbation.

Challenges and Limitations

Although Cas9-based genetic screening is a rapidly maturing technology, there are still many technical challenges that have yet to be resolved. One large obstacle when it comes to performing pooled library screens in a dish are the potential effects of paracrine

signaling. In a pooled format, it is difficult to assess and eliminate cross-talk between neighboring cells in a dish that may all have unique genomic knockouts. Because of this, the importance of certain genes can be easily missed if the gene function can be rescued by nearby cells. For example, if a growth factor is knocked out in a specific cell, its neighbor may continue to release the growth factor, preventing a true knockout phenotype from appearing. In this way, a pooled genome-wide screen may still not identify all genes that are vital for a given phenotype.

Another issue with pooled approaches is the limit to phenotypic outputs that can be read. The researcher is typically restricted to measuring cell proliferation or survival. In addition, there can be efforts to look at phenotypes that FACS can select and sort through such as fluorescence or cell surface markers. More complex phenotypes will be difficult to measure in a pooled screen with reliability. In the future, cheaper robotics that can perform arrayed screens with unique perturbations in each well of multi-well plates will likely allow for more complex tissue-level phenotypes to be assayed. In addition, this sort of high-throughput arrayed screening would remove many of the paracrine effects that may confound results as mentioned previously. If a gene that is being studied is known to be essential for cell viability, it cannot be studied in a complete CRISPR knockout screen when assessing for additional phenotypes. Performing a knockdown study utilizing dCas9 would be more appropriate. In addition, genes that retain their function at low expression levels may easily be missed in knockdown studies and be better performed with a complete knockout screen.

Other issues may arise with false positives and false negatives. In particular, although uncommon, an in-frame repair could occur during a standard positive selection knockout screen resulting in a gain-of-function mutation [104, 105]. This issue is rare enough to not cause vast concern, but something to still be mindful of. More commonly false positives can occur with genes that have a high copy number such as oncogenes. When performing a standard Cas9 knockout screen, these genes will consistently be cleaved leading to multiple double-strand breaks and eventually too many will cause cells to apoptose thus, mistakenly assuming that gene was essential for cell fitness. A gene that may not truly have much of an effect on fitness can falsely appear to if the target site is in one of these amplified regions with a high gene copy number thus, inducing many more double-strand breaks by Cas9 than is typical [106–108]. This can be problematic when performing cancer screens. Many groups have looked at this in detail looking at several cancer cell lines, genes, and sgRNAs for analysis of this amplification effect [106–108]. Aneuploid cell lines produced false positives that mapped to amplified regions of the genome. CRISPR-mediated lethality of cells was independent of transcriptional halting, thus

showing this is due to double-strand breaks and not gene knockout. Previous studies have shown similar discoveries such as targeting the oncogenic BCR-ABL gene fusion that is present in high copy number in K562 cells and notorious for making up the Philadelphia chromosome in chronic myelogenous leukemia. Cas9 targeting resulted in decreased cell viability independent of the target genes function themselves [109]. Ways to prevent these false positives would be to use CRISPRi which do not cut the genome and only offer transcriptional repression. However, even with CRISPRi, other errors can occur especially when dealing with bidirectional promoters causing silencing of multiple genes instead of just the gene of interest. Attempts can be made to remove sgRNAs with massive off-target effects or exclude them from analysis [110]. Utilizing an inducible Cas9 can also be an effective solution to select specifically when to turn on Cas9 with the use of doxycycline.

False negatives come with their own share of complications. If an sgRNA has relatively low activity, it can inadvertently be read as a negative result in a screen. Machine learning approaches can help circumvent some of these issues to design and include only sgRNAs with high activity which has been actively utilized by groups [60, 111, 112]. However, *in silico* sgRNA design has its own share of challenges. When utilizing available online tools, the researcher needs to be aware of the underlying rules to limit off-target effects and increase effectiveness applied by the tool developers. There are also constant updates to gene annotations that need to be ensured for their accuracy and quality. In addition to using computational tools to predict guide efficacy, efforts can also be made to modify the sgRNA scaffold itself to improve activity [113].

One of the large concerns with the use of CRISPR–Cas systems for screens is the possibility of off-target effects. Because sgRNA libraries can contain more than 10^5 different guides, comprehensive individual sgRNA validation and testing is not possible. Multiple studies have shown that Cas9 can tolerate some mismatches between the sgRNA and target sequence allowing for targeting of the wrong gene [1, 114–116]. The farther these mismatches are from the PAM sequence, the more likely these mismatches will be tolerated [117]. It has also been shown that small insertions and deletions are somewhat tolerated as well leading to bulging of the sgRNA or target sequence [116]. Predictive scores have been developed to help the researcher in picking appropriate sgRNAs [118]. Additional Cas9 options are the high-fidelity Cas9 (SpCas9-HF1) or the enhanced specificity Cas9 (eSpCas9) [119, 120]. Many benefits have been shown by delivering Cas9 as a protein instead of a gene in a plasmid as the protein will act immediately and then be quickly degraded which eliminates the constant peaks in expression from a promoter [121]. One strategy to ensure a positive is

true and not from an off-target effect is through validation and ensuring that other reagents targeting that same gene have that same phenotype. However, when performing large pooled screens, there will be multiple sgRNAs targeting the same gene or noncoding region. Effects of a single sgRNA will be less problematic when multiple sgRNAs are targeting that region allowing for some consistency and realization of an off-target effect.

Another challenge is working with PAM sequence restrictions. SpCas9 has a PAM sequence that is more abundant in exons and thus coding regions of the genome which tend to be more GC rich. Other nucleases such as Cpf1 have a PAM sequence that is more abundant in introns which are more AT rich [122]. This is an important factor to keep in mind when selecting a nuclease for screening applications. Performing noncoding functional screens utilizing CRISPR–Cas systems to tile sgRNAs may benefit more from a nuclease such as Cpf1 than SpCas9. One group effectively engineered SpCas9 to recognize different PAM sequences [57]. This can increase specificity and reduce off-target effects while selecting a PAM that is appropriate and unique for the researcher's screening needs.

One often untapped tool for CRISPR–Cas screening is harnessing HDR to insert exogenous genes of interest into the host genome. With HDR's relatively low efficiency compared to NHEJ, it has proven to be difficult to benefit from this technology and perform large knock-in screens at endogenous loci. Knock-in screens can provide valuable information when assessing the roles of knocked-in promoters or repressors on gene function or knocking in mutated genes to mimic disease states. As well, knock-in screens using HDR would preclude the possibility of random lentiviral integration causing confounding effects on cell phenotype. Because of this, more research should be done on pushing the cell to favor HDR over NHEJ. One such study used blocking mutations to increase HDR efficiency [123]. They introduced silent mutations in either the PAM or sgRNA target sequence of the donor strand. These mutations prevented Cas9 from recutting the target sequence once the desired donor was introduced. Greatest efficiency of this is achieved when the mutation is closest to the cut site. This distance can also be optimized to focus on either a homozygous edit or heterozygous edit in the cell depending on the researcher's specific needs (homozygous edits are more likely when the mutation is closest to the cut site and heterozygous edits are more likely when further). Utilizing this blocking method, another study successfully performed a large screen utilizing HDR and saturation mutagenesis to determine function of regulatory elements [124]. They utilized a library of all possible 6-bp combinations to insert into exon 18 of the breast cancer susceptibility gene BRCA1 to measure transcript abundance. They had a similar approach for the lariat debranching

enzyme gene DBR1 to measure the relative effects on growth and function. Interestingly, HDR could also be harnessed to create a knock-in pooled library of sgRNAs in place of typical lentiviral delivery creating cells with stably integrated guides [125]. This could circumvent issues with off-target effects from lentivirus and avoid gene shuffling. Highlighting the potential of HDR-based screening approaches, one group recently performed a large-scale multiplexed HDR CRISPR screen in yeast, utilizing a fusion protein to enhance HDR efficiency [126]. They increased editing efficiency more than 5-fold with use of the fork head protein homolog 1 transcription factor (Fkh1p) fused with the DNA binding protein LexA creating a LexA–Fkh1p fusion protein. This fusion protein recruits donor DNA to the double-strand break site. Utilizing HDR, they incorporated unique barcodes into cells. In addition, they performed saturation editing of a gene encoding for the phospholipid transfer protein SEC14. They incorporated all possible amino acid combinations to identify amino acids critical for chemical inhibition of lipid signaling. Ideally, combining multiple strategies will improve HDR at the greatest efficiency when performing knock-in functional screens. In addition, a researcher could use base-editing techniques to perform a targeted knock-in screen instead of HDR. CRISPR base-editing techniques can modify individual nucleic acid base pairs within the target genes. This is especially beneficial to edit single-nucleotide polymorphisms. Groups have used this technique to identify novel mutations in drug resistance [20, 21]. Overall, screening from endogenous loci using HDR or base editors, although limited to unique screening needs, has significant unexplored potential for investigating genomic function.

Another challenge lies in the large reliability researchers place on cell lines to perform many of these pooled functional screens. Many of these cell lines may not adequately model human disease and functional genomics. In addition, unless kept at a low passage number, cells can begin to change over time with varying mutations, epigenetic changes, and chromosomal changes. Ideally primary cells, human tissues, or *in vivo* screens should be the gold standard. Validating findings in multiple model systems with different techniques is critical. However, with this is the caveat that obtaining different results in different cell lines is permissible if it further explains a critical phenotype unique to the biology of these different systems. In addition, plating cells with the correct growth medium and environmental parameters can be a challenge or whether they even properly plate in 2D. Studies have shown that many human cell types change their physiology in 2D or cannot be cultured at all. For instance, pancreatic cells are notorious for being difficult to culture in 2D and have lasted at most a mere week before huge losses in cell viability [127]. More efforts need to be placed in 3D culture systems and biomimetic environments to ideally model true physiology.

Future Directions

As technical challenges limiting Cas9-based genomic screens are resolved, their ability to inform our understanding of disease progression and treatment will rapidly evolve. By utilizing the expanding toolbox of genetic perturbations and better integrating multiomics data for downstream validation, screens will be able to identify functional elements in the genome more rapidly and accurately. At the same time, expanding screens to patient-derived cell types (iPSCs, tumor biopsies, etc.) will better model human pathologies while providing a potential way to identify patient specific disease vulnerabilities.

Because the majority of human diseases are polygenic (rather than mendelian), there is a clear need for screens which investigate multigene interactions [128, 129]. Toward this end, investigators have recently developed dual knockout Cas9 vectors which deliver two unique sgRNA to identify synthetically lethal genetic interactions in cancer cell lines [77, 130]. In parallel, other researchers have developed alternative dual knockout systems, using a combination of orthogonal Cas9 variants from different bacteria. By utilizing both SpCas9 and SaCas9 (each with their own cognate sgRNAs), they effectively reduce interference between delivered sgRNAs in a dual knockout screen [131]. Moving forward, characterizing a greater number of gene combinations will generate an improved understanding of the genetic basis of non-mendelian diseases. In addition, expanding combination gene perturbations beyond knockouts will provide scientists with a better understanding of directional genetic interactions. In order to characterize these directional interactions, researchers have recently implemented a dual knockout and activation screen in cancer cells to better understand therapeutically relevant genetic interactions networks [33]. Looking forward, integrating multiple different perturbation types in combination has the potential to generate unique datasets with which to probe genomic interactions. For example, integrating inducible Cas9/sgRNA constructs with pooled screening could elucidate temporal dependencies underlying dynamic genetic interactions [132].

Beyond probing exon function, there is an increasing understanding that the noncoding region of the human genome plays a significant role in disease progression across a wide variety of pathologies [133]. In order to better understand this relationship, there have recently been several parallel efforts to map the function of the noncoding portion of the genome using Cas9 [8]. While wtCas9 is ideal for inducing frameshift mutations in the coding regions of exons, probing the noncoding portion of the genome is

more challenging because insertions and deletions are less likely to impact structure and function. To overcome this challenge, CRISPR pooled screening of noncoding loci has primarily focused on using multiple tiled sgRNA to create indels across entire noncoding regulatory sections of the genome to determine functional hotspots. These strategies have identified critical components of endogenous enhancers, as well as novel regulatory elements in unannotated regions of the genome [134–136]. Combining this approach with novel downstream single-cell assays (single-cell RNA seq, etc.) should further aid in rapidly characterizing the structure–function relationship of the noncoding genome. Furthermore, screens utilizing the full CRISPR perturbation tool box will provide researchers with even more novel data sets with which to assay the noncoding genome.

While Cas9 genetic screening has enabled systematic characterization of a broad range of cancer cell lines (via the Broad Institute's Project Achilles among other work), screening primary cells is still in its infancy. Although there is a wealth of information to be gained from screening cancer cell lines, as discussed above, they are not ideal models for healthy cells or diseases other than cancer. Screening in primary cells would better model the *in vivo* genetic and epigenetic profile of the cells of interest, while simultaneously allowing for patient-specific screening strategies to be developed. Because primary cells can be obtained from individuals (or mice) afflicted with nearly any disease, a broader range of disease-specific screening strategies can be developed. In addition, screening in primary cells would allow scientists to unravel the genomic mechanisms underlying the function of various healthy cell types. Primary cell screening has so far been limited to immune cell types which grow sufficiently *in vitro*. As a proof of principal, two groups have recently described a protocol for lentiviral knockout CRISPR screens in mouse primary immune cells, identifying key regulators of the innate immune response and plasma cell differentiation [79, 137]. To push this technology forward, the editing efficiency of Cas9 in primary cells needs to be further optimized to allow for large library screening in many primary cell types. In parallel, improving *in vitro* primary cell culture techniques will drastically improve the ease of primary cell screening protocols. Looking ahead, transitioning this technology toward screening iPSCs could provide a novel method to understand biological development and patient-specific pathological phenotypes. Although iPSC CRISPR screens are still in their infancy, one group recently published a method using Cas9-mediated homologous recombination to fluorescently tag endogenous proteins in developing iPSCs [138]. This method would allow researchers to track the temporal expression and localization of diverse cellular proteins over the course of iPSC differentiation.

As an alternative way to more accurately model cell phenotypes, several groups have independently developed *in vivo* CRISPR screening protocols. *In vivo* CRISPR screening typically involves delivering a library of sgRNAs to a tumor cell line *ex vivo*, implanting the cells into a mouse model, and then tracking which sgRNAs are enriched or depleted as the tumor grows. This method has been used to effectively identify genetic vulnerabilities to immune checkpoint blockers, as well as track genetic drivers of metastasis [75, 139]. These *in vivo* screening methods represent a more robust contextual model with which to analyze cell function, and warrant additional investigation. Other efforts to screen cells in a context that better matches their native environment have utilized 3D culture systems and organoid models. While 3D and organoid models necessitate arrayed screening due to their multicellular architecture, the ability to investigate tissue-level phenotypes has immense implications for functional screens. In 2015, one study described a small-scale CRISPR knockout screen in an organoid model, investigating genetic elements controlling the differentiation of unpolarized basal progenitors into airway epithelium [140]. Although screens involving 3D culture models will certainly be restricted to small libraries of perturbations, their ability to dissect tissue-level phenotypes guarantees their utility to the biomedical community.

As CRISPR screens become more commonplace, it is necessary to stress the importance of using diverse output data to validate results. While sgRNA abundance provides valuable information on which genes are essential for a cellular phenotype, it provides little to no mechanistic data with which to understand gene function. To better understand the biology underlying CRISPR screen results, future research needs to be done on how to best integrate multiomics data with pooled CRISPR screens. Utilizing advances in proteomic and metabolomic measurements has great potential to complement next-generation DNA and RNA sequencing technologies already common place in CRISPR screens. As mass spectrometry pushes closer toward single-cell resolutions, these data will only become more robust, opening up new avenues for understanding the results of pooled screens [141, 142].

Although CRISPR knockout screening via the NHEJ repair pathway has seen widespread adoption, knock-in screening via the HDR templated repair mechanism has been less utilized due to its relatively low efficiency. Many parallel efforts are currently underway to improve the efficacy of HDR mediated gene editing, paving the way for library scale knock-in screening [126, 143, 144]. Knock-in screening using HDR to scarlessly insert a mutagenized DNA sequence at its endogenous locus has many unexplored applications. In the future, researchers could use HDR to perform site-directed mutagenesis of complex mammalian proteins in their endogenous loci, enabling the engineering of post-translationally modified proteins which may not be

amenable to production in yeast or bacteria. This same method could also be used to engineer mammalian cell lines with novel metabolic pathways for use in biopharmaceutical production.

The past half-decade has seen rapid development of novel CRISPR–Cas based tools with which to investigate genomic function. At the same time, *de novo* DNA synthesis and *in silico* sgRNA design tools have quickly become mature technologies, resolving many of the technical challenges preventing the widespread adoption of CRISPR–Cas genetic screens. Consequently, CRISPR–Cas genetic screening has transitioned from exciting new academic research, to a ubiquitous technology with few barriers to use. Looking forward, it now seems plausible that the many functional screens ongoing in immortalized cancer cell lines will lead to a complete mapping of cancer specific gene function and genetic interactions. While this research has great potential to inform our understanding of cancer etiology and drug candidate efficacy, the immense genetic variation in patient cancer samples limits the translational relevance of cell line-based genetic screening. In addition, conclusions drawn from screens performed in cancer cell lines may have limited relevance to other disease phenotypes. This genetic variation between patients and cancer cell lines necessitates the development of patient-specific CRISPR–Cas screening protocols. Building off existing cancer mapping initiatives, CRISPR–Cas functional screening efforts in patient-derived cells should one day help oncologists predict treatment efficacy and inform drug choice. In parallel, future screens in patient derived iPSCs will allow researchers to expand the range of disease phenotypes CRISPR–Cas functional screening can investigate. In this way, CRISPR–Cas screening can contribute to a growing body of research underlying precision medicine and personalized therapeutics.

Abbreviations used:
CRISPR, clustered regularly interspaced short palindromic repeat; Cas, CRISPR-associated; PAM, protospacer adjacent motif; NHEJ, non-homologous end joining; HDR, homology-directed repair; NGS, next-generation sequencing; FACS, fluorescence-activated cell sorting; HCI, high-content imaging; AAV, adeno-associated virus; RNP, ribonucleoprotein.

References

[1] P. Mali, et al., RNA-guided human genome engineering via Cas9, Science (80-.) 339 (2013) 823–826.

[2] L. Cong, et al., Multiplex genome engineering using CRISPR/Cas systems, Science (80-.) 339 (2013) 819–823.

[3] N.E. Sanjana, Genome-scale CRISPR pooled screens, Anal. Biochem. 532 (2017) 95–99.

[4] R. Barrangou, C.A. Gersbach, Expanding the CRISPR toolbox: targeting RNA with Cas13b, Mol. Cell 65 (2017) 582–584.

[5] E.V. Koonin, K.S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR–Cas systems, Curr. Opin. Microbiol. 37 (2017) 67–78.

[6] S.H. Sternberg, S. Redding, M. Jinek, E.C. Greene, J.A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9, Nature 507 (2014) 62–67.

[7] W.-D. Heyer, K.T. Ehmsen, J. Liu, Regulation of homologous recombination in eukaryotes, Annu. Rev. Genet. 44 (2010) 113–139.

[8] A. Montalbano, M.C. Canver, N.E. Sanjana, High-throughput approaches to pinpoint function within the noncoding genome, Mol. Cell 68 (2017) 44–59.

[9] J.G. Doench, Am I ready for CRISPR? A user's guide to genetic screens, Nat. Rev. Genet. (2017) https://doi.org/10.1038/nrg.2017.97.

[10] A.A. Dominguez, W.A. Lim, L.S. Qi, Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation, Nat. Rev. Mol. Cell Biol. 17 (2016) 5–15.

[11] O. Shalem, N.E. Sanjana, F. Zhang, High-throughput functional genomics using CRISPR–Cas9, Nat. Rev. Genet. 16 (2015) 299–311.

[12] T. Wang, et al., Genetic screens in human cells using the CRISPR/Cas9 system, Science 80 (2014) 80–85.

[13] L.R. Polstein, et al., Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators, Genome Res. 25 (2015) 1158–1169.

[14] M. Kampmann, CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine, ACS Chem. Biol. (2017) https://doi.org/10.1021/acschembio.7b00657.

[15] S. Konermann, et al., Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex, Nature 517 (2015) 583–588.

[16] L.A. Gilbert, et al., Genome-scale CRISPR-mediated control of gene repression and activation, Cell 159 (2014) 647–661.

[17] G.T. Hess, J. Tycko, D. Yao, M.C. Bassik, Methods and applications of CRISPR-mediated base editing in eukaryotic genomes, Mol. Cell 68 (2017) 26–43.

[18] N.M. Gaudelli, et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature 551 (2017) 464–471.

[19] A.C. Komor, Y.B. Kim, M.S. Packer, J.A. Zuris, D.R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, Nature 533 (2016) 420–424.

[20] G.T. Hess, et al., Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells, Nat. Methods 13 (2016) 1036–1042.

[21] Y. Ma, et al., Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells, Nat. Methods 13 (2016) 1029–1035.

[22] C. Kuscu, et al., CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations, Nat. Methods 14 (2017) 710–712.

[23] A. Vojta, et al., Repurposing the CRISPR–Cas9 system for targeted DNA methylation, Nucleic Acids Res. 44 (2016) 5615–5628.

[24] X.S. Liu, et al., Editing DNA methylation in the mammalian genome, Cell 167 (2016) 233–247.e17.

[25] B.I. Laufer, S.M. Singh, Strategies for precision modulation of gene expression by epigenome editing: an overview, Epigenetics Chromatin 8 (2015) 1–12.

[26] I.B. Hilton, et al., Epigenome editing by a CRISPR–Cas9-based acetyltransferase activates genes from promoters and enhancers, Nat. Biotechnol. 33 (2015) 510–517.

[27] D.Y. Kwon, Y.T. Zhao, J.M. Lamonica, Z. Zhou, Locus-specific histone deacetylation using a synthetic CRISPR–Cas9-based HDAC, Nat. Commun. 8 (2017) 1–8.

[28] N.A. Kearns, et al., Functional annotation of native enhancers with a Cas9-histone demethylase fusion, Nat. Methods 12 (2015) 401–403.

[29] M. Oren, V. Rotter, Mutant p53 gain-of-function in cancer, Cold Spring Harb. Perspect. Biol. 2 (2010) 1–15.

[30] P.R. Albert, B. Le François, A.M. Millar, Transcriptional dysregulation of 5-HT1A autoreceptors in mental illness, Mol. Brain 4 (2011) 1–14.

[31] T.J. Gonda, R.G. Ramsay, Directly targeting transcriptional dysregulation in cancer, Nat. Rev. Cancer 15 (2015) 686–694.

[32] L.A. Gilbert, et al., CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, Cell 154 (2013) 442–451.

[33] M. Boettcher, et al., Dual gene activation and knockout screen reveals directional dependencies in genetic networks, Nat. Biotechnol. 36 (2018) 170–178.

[34] C. Vogel, E.M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses, Nat. Rev. Genet. 13 (2012) 227–232.

[35] D.G. Lupiáñez, et al., Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions, Cell 161 (2015) 1012–1025.

[36] C.M. Nowak, S. Lawson, M. Zerez, L. Bleris, Guide RNA engineering for versatile Cas9 functionality, Nucleic Acids Res. 44 (2016) 9555–9564.

[37] J.G. Zalatan, et al., Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds, Cell 160 (2015) 339–350.

[38] H. Ma, et al., Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow, Nat. Biotechnol. 34 (2016) 528–530.

[39] A. Agrotis, R. Ketteler, A new age in functional genomics using CRISPR/Cas9 in arrayed library screening, Front. Genet. 6 (2015) 1–15.

[40] M. Boutros, F. Heigwer, C. Laufer, Microscopy-Based High-Content Screening, Cell 163 (2015) 1314–1325.

[41] B. Neumann, et al., High-throughput RNAi screening by time-lapse imaging of live human cells, Nat. Methods 3 (2006) 385–390.

[42] J. Moffat, et al., A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen, Cell 124 (2006) 1283–1298.

[43] K. Schneeberger, Using next-generation sequencing to isolate mutant genes from forward genetic screens, Nat. Rev. Genet. 15 (2014) 662–676.

[44] M.C. Canver, et al., Integrated design, execution, and analysis of arrayed and pooled CRISPR genome editing experiments, 2017 https://doi.org/10.1101/125245.

[45] J. Joung, et al., Genome-scale CRISPR–Cas9 knockout and transcriptional activation screening, Nat. Protoc. 12 (2017) 828–863.

[46] E.M. LeProust, et al., Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process, Nucleic Acids Res. 38 (2010) 2522–2540.

[47] S. Kosuri, et al., A scalable gene synthesis platform using high-fidelity DNA microchips, Nat Biotechnol 28 (2011) 1295–1299.

[48] S. Kosuri, G.M. Church, Large-scale de novo DNA synthesis: technologies and applications, Nat. Methods 11 (2014) 499–507.

[49] O. Shalem, et al., Genome-scale CRISPR–Cas9 knockout screening in human cells, Science (80-.) 343 (2014) 84–88.

[50] V. Pattanayak, et al., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity, Nat. Biotechnol. 31 (2013) 839–843.

[51] T. Hart, et al., High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities, Cell 163 (2015) 1515–1526.

[52] M. Doerflinger, W. Forsyth, G. Ebert, M. Pellegrini, M.J. Herold, CRISPR/Cas9—the ultimate weapon to battle infectious diseases? Cell. Microbiol. 19 (2017) 1–10.

[53] A.S. Puschnik, K. Majzoub, Y.S. Ooi, J.E. Carette, A CRISPR toolbox to study virus-host interactions, Nat. Rev. Microbiol. 15 (2017) 351–364.

[54] R.J. Park, et al., A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors, 49 (2017) 193–203.

[55] E.S. Egan, Beyond hemoglobin: screening for malaria host factors, Trends Genet. 34 (2017) 133–141.

[56] A.K. Singh, et al., Investigating essential gene function in *Mycobacterium tuberculosis* using an efficient CRISPR interference system, Nucleic Acids Res. 44 (2016), e143–e143.

[57] B.P. Kleinstiver, et al., Engineered CRISPR–Cas9 nucleases with altered PAM specificities, Nature 523 (2015) 481–485.

[58] G. hui Chuai, Q.L. Wang, Q. Liu, In silico meets in vivo: towards computational CRISPR-based sgRNA design, Trends Biotechnol. 35 (2017) 12–21.

[59] N. Wu, K. Matand, B. Kebede, G. Acquaah, S. Williams, Enhancing DNA electrotransformation efficiency in *Escherichia coli* DH10B electrocompetent cells, Electron. J. Biotechnol. 13 (2010) 1–9.

[60] J.G. Doench, et al., Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR–Cas9, Nat. Biotechnol. 34 (2016) 184–191.

[61] J.R. McDade, N.C. Waxmonsky, L.E. Swanson, M. Fan, Practical considerations for using pooled lentiviral CRISPR libraries, Curr. Protoc. Mol. Biol. 2016 (2016) 31.5.1–31.5.13.

[62] H. Koike-Yusa, Y. Li, E.P. Tan, M.D.C. Velasco-Herrera, K. Yusa, Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library, Nat. Biotechnol. 32 (2014) 267–273.

[63] T. Wang, E.S. Lander, D.M. Sabatini, Viral packaging and cell culture for CRISPR-based screens, Cold Spring Harb Protoc 2016 (2016) 289–296.

[64] Y. Zhou, et al., High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells, Nature 509 (2014) 487–491.

[65] M. Song, The CRISPR/Cas9 system: their delivery, in vivo and ex vivo applications and clinical development by startups, Biotechnol. Prog. 33 (2017) 1035–1045.

[66] R.D. Chow, et al., AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma, Nat. Neurosci. 20 (2017) 1329–1341.

[67] J.C. Grieger, R.J. Samulski, Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps, J. Virol. 79 (2005) 9933–9944.

[68] C. Liu, L. Zhang, H. Liu, K. Cheng, Delivery strategies of the CRISPR–Cas9 gene-editing system for therapeutic applications, J. Control. Release 266 (2017) 17–26.

[69] K. Schumann, et al., Generation of knock-in primary human T cells using Cas9 ribonucleoproteins, Proc. Natl. Acad. Sci. 112 (2015) 10437–10442.

[70] L.J. Rupp, et al., CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells, Sci. Rep. 7 (2017) 1–10.

[71] J.E. Vargas, et al., Retroviral vectors and transposons for stable gene therapy: Advances, current challenges and perspectives, J. Transl. Med. 14 (2016) 1–15.

[72] C. Xu, et al., *piggyBac* mediates efficient in vivo CRISPR library screening for tumorigenesis in mice, Proc. Natl. Acad. Sci. 114 (2017) 722–727.

[73] P. Wang, et al., Genome editing for cancer therapy: delivery of Cas9 protein/sgRNA plasmid via a gold nanocluster/lipid core–shell nanocarrier, Adv. Sci. 4 (2017).

[74] R. Mout, et al., Direct cytosolic delivery of CRISPR/Cas9-ribonucleoprotein for efficient gene editing, ACS Nano 11 (2017) 2452–2458.

[75] R.T. Manguso, et al., In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target, Nature 547 (2017) 413–418.

[76] L. Naldini, et al., In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector, Science (80-.) 272 (1996) 263–267.

[77] J.P. Shen, et al., Combinatorial CRISPR–Cas9 screens for de novo mapping of genetic interactions, Nat. Methods 14 (2017) 573–576.

[78] B. Adamson, et al., A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response, Cell 167 (2016) 1867–1882.e21.

[79] O. Parnas, et al., A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks, Cell 162 (2015) 675–686.

[80] J. Wu, et al., Generation of human organs in pigs via interspecies blastocyst complementation, Reprod. Domest. Anim. 51 (2016) 18–24.

[81] E.Z. Macosko, et al., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets, Cell 161 (2015) 1202–1214.

[82] A. Dixit, et al., Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens, Cell 167 (2016) 1853–1866.e17.

[83] D.A. Jaitin, et al., Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-Seq, Cell 167 (2016) 1883–1896.e15.

[84] P. Datlinger, et al., Pooled CRISPR screening with single-cell transcriptome readout, Nat. Methods 14 (2017) 297–301.

[85] T. Hardcastle, K. Kelly, Empirical Bayesian Methods for Differential Expression in Count Data, 2009.

[86] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2009) 139–140.

[87] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.

[88] Y. Di, D.W. Schafer, J.S. Cumbie, J.H. Chang, The NBP negative binomial model for assessing differential gene expression from RNA-Seq, Stat. Appl. Genet. Mol. Biol. 10 (2011).

[89] W. Li, et al., MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens, Genome Biol. 15 (2014) 554.

[90] J. Winter, et al., CaRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens, Bioinformatics 32 (2015) 632–634.

[91] H.H. Jeong, S.Y. Kim, M.W.C. Rousseaux, H.Y. Zoghbi, Z. Liu, CRISPRcloud: a secure cloud-based pipeline for CRISPR pooled screen deconvolution, Bioinformatics 33 (2017) 2963–2965.

[92] W. Li, et al., Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR, Genome Biol. 16 (2015) 1–13.

[93] J.C. Caicedo, et al., Data-analysis strategies for image-based cell profiling, Nat. Methods 14 (2017) 849–863.

[94] A.E. Carpenter, et al., CellProfiler: image analysis software for identifying and quantifying cell phenotypes, Genome Biol. (2006)https://doi.org/10.1186/gb-2006-7-10-r100.

[95] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, W. Huber, EBImage—an R package for image processing with applications to cellular phenotypes, Bioinformatics 26 (2010) 979–981.

[96] A.H.Y. Tong, et al., Systematic genetic analysis with ordered arrays of yeast deletion mutants, Science (80-.) 294 (2001) 2364–2368.

[97] D. Zhao, et al., Combinatorial CRISPR–Cas9 metabolic screens reveal critical redox control points dependent on the KEAP1–NRF2 regulatory axis, Mol. Cell 69 (2018) 648–663.e7.

[98] D. Zhao, J.P. Shen, R. Sasik, T. Ideker, P. Mali, Combinatorial CRISPR–Cas9 knockout screen, Protoc. Exch. (2017) 3–8, https://doi.org/10.1038/protex.2017.063.

[99] M.K. Muellner, et al., TOPS: a versatile software tool for statistical analysis and visualization of combinatorial gene–gene and gene–drug interaction screens, BMC Bioinf. 15 (2014) 1–12.

[100] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (2018).

[101] N.E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening, Nat. Methods 11 (2014) 783–784.

[102] L.A. Miles, R.J. Garippa, J.T. Poirier, Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens, FEBS J. 283 (2016) 3170–3180.

[103] Rodenburg, R. J. The functional genomics laboratory: functional validation of genetic variants. 297–307 (2018).

[104] G. Chuai, et al., Deciphering relationship between micro-homology and in-frame mutation occurrence in human CRISPR-based gene knockout, Mol. Ther.–Nucleic Acids 5 (2016), e323.

[105] J.J. Ipsaro, et al., Rapid generation of drug-resistance alleles at endogenous loci using CRISPR–Cas9 indel mutagenesis, PLoS One 12 (2017) 1–16.

[106] A. Sheel, W. Xue, Genomic amplifications cause false positives in CRISPR screens, Cancer Discov. 6 (2016) 824–826.

[107] D.M. Munoz, et al., CRISPR screens provide a comprehensive assessment of cancer vulnerabilities but generate false-positive hits for highly amplified genomic regions, Cancer Discov. 6 (2016) 900–913.

[108] A.J. Aguirre, et al., Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting, Cancer Discov. 6 (2016) 914–929.

[109] T. Wang, et al., Identification and characterization of essential genes in the human genome, Science (80-.) 350 (2015) 1096–1101.

[110] R.M. Meyers, et al., Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells, Nat. Genet. 49 (2017) 1779–1784.

[111] J.G. Doench, et al., Rational design of highly active sgRNAs for CRISPR–Cas9-mediated gene inactivation, Nat. Biotechnol. 32 (2014) 1262–1267.

[112] S.E. Mohr, et al., CRISPR guide RNA design for research applications, FEBS J. 283 (2016) 3232–3238.

[113] B.C.S. Cross, et al., Increasing the performance of pooled CRISPR–Cas9 drop-out screening, Sci. Rep. 6 (2016) 1–8.

[114] P.D. Hsu, et al., DNA targeting specificity of RNA-guided Cas9 nucleases, Nat. Biotechnol. 31 (2013) 827–832.

[115] X.H. Zhang, L.Y. Tee, X.G. Wang, Q.S. Huang, S.H. Yang, Off-target effects in CRISPR/Cas9-mediated genome engineering, Mol. Ther. - Nucleic Acids 4 (2015).

[116] Y. Lin, et al., CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences, Nucleic Acids Res. 42 (2014) 7473–7485.

[117] P.D. Hsu, E.S. Lander, F. Zhang, Development and applications of CRISPR–Cas9 for genome engineering, Cell 157 (2014) 1262–1278.

[118] J. Tycko, V.E. Myer, P.D. Hsu, Methods for optimizing CRISPR–Cas9 genome editing specificity, Mol. Cell 63 (2016) 355–370.

[119] B.P. Kleinstiver, et al., High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects, Nature 529 (2016) 490–495.

[120] I.M. Slaymaker, et al., Rationally engineered Cas9 nucleases with improved specificity, Science (80-.) 351 (2016) 84–88.

[121] S. Kim, D. Kim, S.W. Cho, J. Kim, J. Kim, Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins, Genom Res (2014) 1012–1019, https://doi.org/10.1101/gr.171322.113.

[122] M.C. Canver, et al., Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci, Nat. Genet. 49 (2017) 625–634.

[123] D. Paquet, et al., Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9, Nature 533 (2016) 125–129.

[124] G.M. Findlay, E.A. Boyle, R.J. Hause, J.C. Klein, J. Shendure, Saturation editing of genomic regions by multiplex homology-directed repair, Nature 513 (2014) 120–123.

[125] N. Rajagopal, et al., High-throughput mapping of regulatory DNA, Nat. Biotechnol. 34 (2016) 167–174.

[126] K.R. Roy, et al., Multiplexed precision genome editing with trackable genomic barcodes in yeast, Nat. Publ. Group (2018) https://doi.org/10.1038/nbt.4137.

[127] E.F. Jacobson, E.S. Tzanakakis, Human pluripotent stem cell differentiation to functional pancreatic cells for diabetes therapies: innovations, challenges and future directions, J. Biol. Eng. 11 (21) (2017).

[128] P.D.P. Pharoah, et al., Polygenic susceptibility to breast cancer and implications for prevention, Nat. Genet. 31 (2002) 33–36.

[129] D. Lvovs, O.O. Favorova, A.V. Favorov, A polygenic approach to the study of polygenic diseases, Acta Naturae 4 (2012) 59–71.

[130] A.S.L. Wong, et al., Multiplexed barcoded CRISPR–Cas9 screening enabled by CombiGEM, Proc. Natl. Acad. Sci. 113 (2016) 2544–2549.

[131] F.J. Najm, et al., Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens, Nat. Biotechnol. (2017) https://doi.org/10.1038/nbt.4048.

[132] B.J. Aubrey, et al., An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo, Cell Rep. 10 (2015) 1422–1432.

[133] F. Zhang, J.R. Lupski, Non-coding genetic variants in human disease, Hum. Mol. Genet. 24 (2015) R102–R110.

[134] Y. Diao, et al., A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells, Nat. Methods 14 (2017) 629–635.

[135] Y. Diao, et al., A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9 mediated genetic screening, Genome Res. 1–9 (2016) https://doi.org/10.1101/gr.197152.115.

[136] M.C. Canver, et al., BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Nature 527 (2015) 192–197.

[137] V.T. Chu, et al., Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line, Proc. Natl. Acad. Sci. 113 (2016) 12514–12519.

[138] A. Sharma, et al., CRISPR/Cas9-mediated fluorescent tagging of endogenous proteins in human pluripotent stem cells, Curr Protoc Hum Genet (2018) 21.11.1–21.11.20, https://doi.org/10.1002/CPHG.52.

[139] S. Chen, et al., Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis, Cell 160 (2015) 1246–1260.

[140] X. Gao, A.S. Bali, S.H. Randell, B.L.M. Hogan, GRHL2 coordinates regeneration of a polarized mucociliary epithelium from basal stem cells, J. Cell Biol. 211 (2015) 669–682.

[141] B. Budnik, E. Levy, G. Harmange, N. Slavov, Mass-Spectrometry of Single Mammalian Cells Quantifies Proteome Heterogeneity During Cell Differentiation, *bioRxiv* 102681 2018 https://doi.org/10.1101/102681.

[142] K.J. Boggio, et al., Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact, Expert Rev. Proteomics 8 (2011) 591–604.

[143] X. Liang, J. Potter, S. Kumar, N. Ravinder, J.D. Chesnut, Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA, J. Biotechnol. 241 (2017) 136–146.

[144] V.T. Chu, et al., Increasing the efficiency of homology-directed repair for CRISPR–Cas9-induced precise gene editing in mammalian cells, Nat. Biotechnol. 33 (2015) 543–548.