High-throughput functional genomics using CRISPR–Cas9

Ophir Shalem, Neville E. Sanjana and Feng Zhang

Abstract | Forward genetic screens are powerful tools for the discovery and functional annotation of genetic elements. Recently, the RNA-guided CRISPR (clustered regularly interspaced short palindromic repeat)-associated Cas9 nuclease has been combined with genome-scale guide RNA libraries for unbiased, phenotypic screening. In this Review, we describe recent advances using Cas9 for genome-scale screens, including knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity. We discuss practical aspects of screen design, provide comparisons with RNA interference (RNAi) screening, and outline future applications and challenges.

Small interfering RNA (siRNA). RNA molecules that are 21–23 nucleotides long and that are processed from long double-stranded RNAs; they are functional components of the RNA-induced silencing complex (RISC). siRNAs typically target and silence mRNAs by binding perfectly complementary sequences in the mRNA and causing their degradation and/or translational inhibition.

Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA; McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, and Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.
Correspondence to O.S. and F.Z. e-mails: ophir.shalem@gmail.com; zhang@broadinstitute.org
doi:10.1038/nrg3899
Published online 9 April 2015

A key goal in genetic analysis is to identify which genes contribute to specific biological phenotypes and diseases. Hypothesis-driven, reverse genetic methods take a 'genotype-to-phenotype' approach by using prior knowledge to test the causal role of specific genetic perturbations. By contrast, forward genetic screens are 'phenotype-to-genotype' approaches that involve modifying or modulating the expression of many genes, selecting for the cells or organisms with a phenotype of interest, and then characterizing the mutations that result in those phenotypic changes.

Initial forward genetic experiments carried out on model organisms such as yeast, flies, plants, zebrafish, nematodes and rodents[1–9] relied on the use of chemical DNA mutagens followed by the isolation of individuals with an aberrant phenotype. These screens have uncovered many basic biological mechanisms, such as RAS and NOTCH signalling pathways[10], as well as molecular mechanisms of embryonic patterning[11,12] and development[13,14].

A major shortcoming of DNA-mutagen-based screens is that the causal mutations in the selected clones are initially unknown. Identifying the causal mutations can be costly and labour intensive, requiring linkage analysis through crosses with characterized lines. These challenges can now be more easily addressed by mapping mutations using next-generation sequencing (NGS)[15] and by replacing chemical mutagens with viruses and transposons, which use defined insertion sequences that are amenable to sequencing-based analysis[16–18]. An additional limitation of random mutagenesis approaches is that the resulting mutants are typically heterozygotes, which can mask recessive phenotypes. In model organisms, homozygosity can be achieved by intercrossing progeny derived from the initial heterozygous mutant. In mammalian cell culture, recessive screens have been limited to near-haploid cell lines[19,20] or to cell lines that are deficient in Bloom helicase (BLM), which have an increased rate of mitotic recombination[21].

Over the past decade, forward genetic screens have been revolutionized by the development of tools that use the RNA interference (RNAi) pathway for gene knockdown. RNAi is a conserved endogenous pathway in which mRNA molecules are targeted for degradation on the basis of sequence complementarity[22,23], thus facilitating design and scalability of the tools. Several RNAi reagents have been developed, including long double-stranded RNA (dsRNA)[24], synthetic small interfering RNA (siRNA)[25], short hairpin RNA (shRNA)[26] and shRNAs embedded in microRNA (miRNA) precursors (shRNAmirs)[27,28]. Screens using RNAi tools have provided a wealth of information on gene function[1,26,29–32], but their utility has been hindered by incomplete gene knockdown and extensive off-target activity, making it difficult to interpret phenotypic changes[33–35].

Sequence-specific programmable nucleases have emerged as an exciting new genetic perturbation system that enables the targeted modification of the DNA sequence itself. In particular, the RNA-guided endonuclease Cas9 (REFS 36–41) from the microbial adaptive immune system CRISPR (clustered regularly interspaced short palindromic repeat) provides a convenient system for achieving targeted mutagenesis in eukaryotic cells[42,43]. Cas9 is targeted to specific genomic loci via a guide RNA, which recognizes the target DNA through

© 2015 Macmillan Publishers Limited. All rights reserved

Watson–Crick base pairing. Therefore, Cas9 combines the permanently mutagenic nature of classical mutagens with the programmability of RNAi.

In this Review, we discuss recent Cas9-based functional genetic screening tools, including genome-wide knockout approaches and related strategies using modified forms of Cas9 to cause gene knockdown or transcriptional activation in a non-mutagenic manner[44–49]. We discuss how these newer approaches compare with and complement existing RNAi-based screening technologies. We also present some practical considerations for designing Cas9-based screens and potential future directions for targeted screening technology development.

Mechanisms of perturbation

Loss-of-function perturbations mediated by Cas9 and RNAi. Cas9 nuclease is a component of the type II CRISPR bacterial adaptive immune system that has recently been adapted for genome editing in many eukaryotic models (reviewed in REFS 50,51). Targeted genome engineering with Cas9 and other nucleases exploits endogenous DNA double-strand break (DSB) repair pathways to create mutations at specific locations in the genome. Although there is a large diversity of DSB repair mechanisms, genome editing in mammalian cells primarily relies on homology-directed repair (HDR), in which an exogenous DNA template can facilitate precise repair, as well as non-homologous end-joining (NHEJ), which is an error-prone repair mechanism that introduces indel mutations at the repair site[52]. To induce DSBs, Cas9 can be targeted to specific locations in the genome by specifying a short single guide RNA (sgRNA)[41] to complement the target DNA. For the commonly used *Streptococcus pyogenes* Cas9, the sgRNA contains a 20-bp guide sequence. The target DNA needs to contain the 20-bp target sequence followed by a 3-bp protospacer-adjacent motif (PAM), although some mismatches can be tolerated (see below).

Loss-of-function mutations mediated by Cas9 nuclease are achieved by targeting a DSB to a constitutively spliced coding exon. When a DSB is repaired by NHEJ, it can introduce an indel mutation. This frequently causes a coding frameshift, resulting in a premature stop codon and the initiation of nonsense-mediated decay (NMD) of the transcript (FIG. 1). NMD might not be active for all genes and is not necessarily required for Cas9-mediated knockout, as an early frameshift mutation or large indels might be sufficient to produce a non-functional protein. Early exons are preferred for targeting, as indels in these exons have a higher probability of introducing an early stop codon or a frameshift of a larger portion of the protein[53]. As DSB induction and NHEJ-mediated repair occur independently at each allele in diploid cells, targeting by Cas9 results in a range of biallelic and heterozygous target gene lesions in different cells. We and others[44–47] have used the simple, RNA-mediated programmability of Cas9 and its nuclease function to conduct genome-scale knockout screens in mammalian cell cultures. These initial screens uncovered both known and novel insights into gene essentiality and resistance to drugs and toxins. Most importantly, Cas9-based screens displayed high reagent consistency, strong phenotypic effects and high validation rates, demonstrating the promise of this approach.

Although the application of Cas9 to targeted screening is relatively recent, similar approaches based on RNAi technologies have been extensively used over the past decade in mammalian cell culture and *in vivo*[1,3,26,29,30,54–58]. RNAi is a conserved natural pathway that is triggered by various types of dsRNAs (often single-stranded RNAs folded into hairpin structures) and that results in the selective downregulation of transcripts with sequence complementarity to one strand of the dsRNA[23]. Natural sources of dsRNAs include endogenous miRNAs[59] and exogenous linear dsRNAs that are typically introduced into cells by invading viruses[60–62]. Artificial targeted gene knockdown is achieved by the delivery of a wide range of designed RNAi reagents[55,63], including long dsRNAs[24], siRNAs[25], shRNAs[26] and miRNA-embedded shRNAs[27,28]. The delivery of RNAi reagents is achieved by transfection of pre-synthesized RNA (for siRNAs and dsRNAs), by transfection of DNA (which encodes a promoter-driven shRNA or shRNAmir) or by viral transduction methods using lentiviral, retroviral or transposon constructs

Figure 1 | **Molecular mechanisms underlying gene perturbation via lentiviral delivery of RNA interference reagents, Cas9 nuclease and dCas9 transcriptional effectors.** **a** | Lentiviral transduction begins with the fusion of virus particles with the cell membrane and the insertion of the single-stranded RNA (ssRNA) viral genome into the cell cytoplasm. A reverse transcriptase then converts the ssRNA genome into double-stranded DNA (dsDNA) that is imported into the nucleus and integrates into the host cell genome. Short hairpin RNA (shRNA) or single guide RNA (sgRNA) transgenes are then expressed from an RNA polymerase III (Pol III) or Pol II promoter. **b** | For shRNA transgenes, maturation involves a series of nucleolytic processing steps that result in cytoplasmic small interfering RNA (siRNA) with sequence complementarity to the target mRNA. Drosha processing is required for reagents consisting of shRNAs embedded in microRNA precursors (shRNAmirs) but is usually bypassed for simple stem–loop shRNA reagents. Gene silencing is achieved by siRNA recruitment to the RNA-induced silencing complex (RISC) for mRNA degradation and translational inhibition. **c**,**d** | By contrast, both the Cas9 nuclease and catalytically inactive Cas9 (dCas9)-mediated transcriptional modulation act in the nucleus. The transgene-encoded Cas9–sgRNA complex targets a genomic locus through sequence complementarity to the 20-bp sgRNA spacer sequence (part **c**). For Cas9 nuclease-mediated knockout, double-strand break (DSB) formation is followed by non-homologous end-joining (NHEJ) DNA repair that can introduce an indel mutation and a coding frameshift. For dCas9-mediated transcriptional modulation, the modification of expression (white arrows) depends on the exact type of fusion of either dCas9 or sgRNA (part **d**) (FIG. 2). These induced nuclear events, together with endogenous transcript degradation and dilution through cell division, will result in a new steady-state expression level in the cytoplasm.

Short hairpin RNA
(shRNA). Small RNAs forming hairpins that can induce sequence-specific silencing in mammalian cells through RNA interference, both when expressed endogenously and when produced exogenously and transfected into the cell.

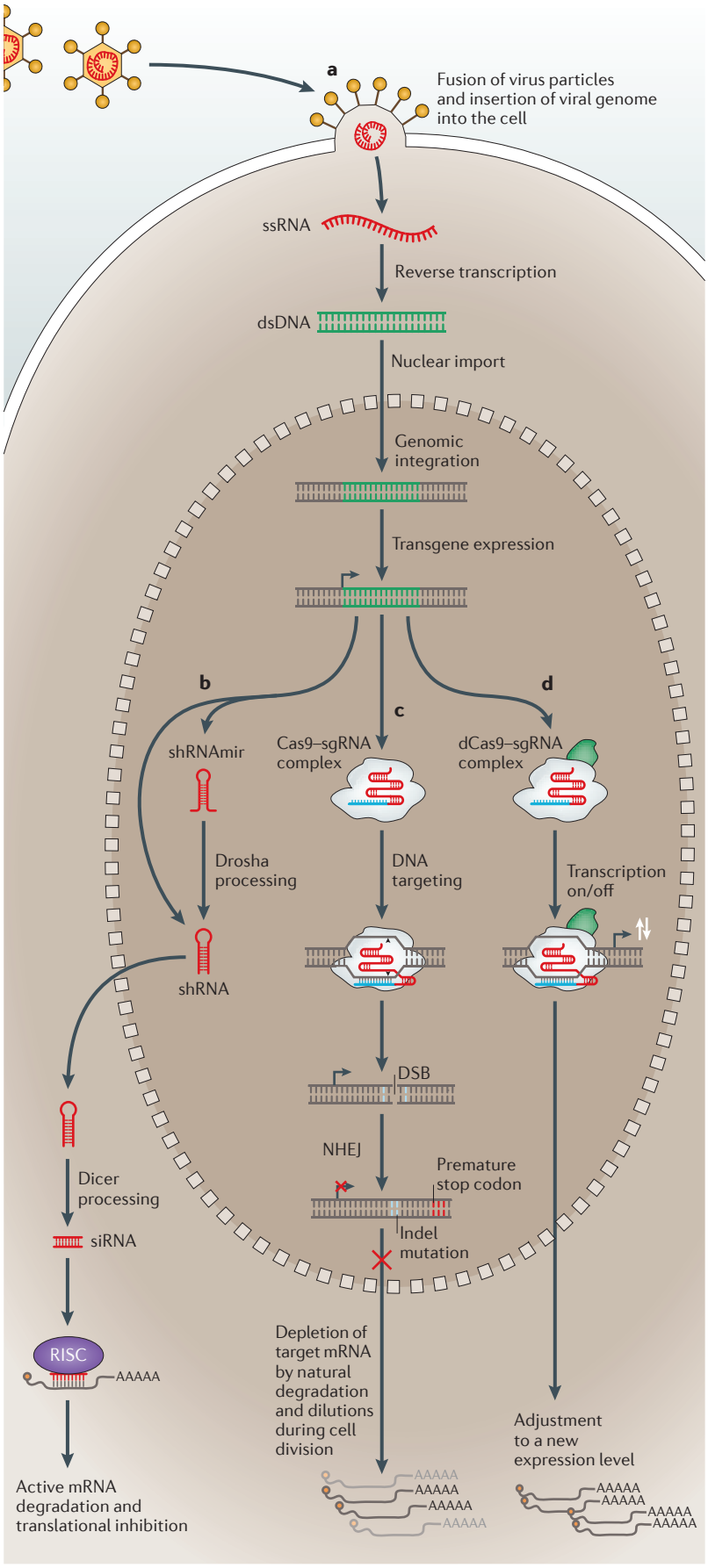
microRNA
(miRNA). Small RNA molecules processed from hairpin-containing RNA precursors that are produced from endogenous miRNA-encoding genes. miRNAs are 21–23 nucleotides in length and, through the RNA-induced silencing complex (RISC), they target and silence mRNAs containing imperfectly complementary sequences.

Indel
(Insertion and deletion). Mutations due to small insertions or deletions of DNA sequences.

Single guide RNA
(sgRNA). An artificial fusion of CRISPR (clustered regularly interspaced short palindromic repeat) RNA (crRNA) and transactivating crRNA (tracrRNA) with critical secondary structures for loading onto Cas9 for genome editing. It functionally substitutes the complex of crRNA and tracrRNA that occurs in natural CRISPR systems. It uses RNA–DNA hybridization to guide Cas9 to the genomic target.

Nonsense-mediated decay
(NMD). An mRNA surveillance mechanism that degrades mRNAs containing nonsense mutations to prevent the expression of truncated or erroneous proteins.

© 2015 Macmillan Publishers Limited. All rights reserved



with a cloned shRNA or shRNAmir cassette (FIG. 1). In contrast to RNA polymerase III (Pol III)-driven expression of shRNAs or sgRNAs, Pol II-driven expression of shRNAmirs can be temporally controlled and genetically restricted across tissues[63]. Most RNAi reagents are nucleolytically processed by the enzyme Dicer into functional siRNAs. Before processing by Dicer, shRNAmirs require nuclear processing by Drosha–DGCR8, but this step is usually bypassed with other reagents[63]. Regardless of the reagent type, the resultant siRNAs are then loaded into the RNA-induced silencing complex (RISC), which is guided to the target mRNA molecule by the siRNA to initiate mRNA degradation or translational inhibition[23].

Catalytically inactive Cas9 for transcriptional modulation. In addition to gene knockout that is mediated by the error-prone repair of targeted DSBs and RNAi-based gene knockdowns, catalytically inactive Cas9 (dCas9) and various fusions of either dCas9 or sgRNAs with transcriptional activator, repressor and recruitment domains have been used to modulate gene expression at targeted loci without introducing irreversible mutations to the genome. The dCas9-based transcriptional inhibition and activation systems are commonly referred to as CRISPRi and CRISPRa, respectively (FIG. 2). dCas9 by itself can have a repressive effect on gene expression, which is probably due to steric hindrance of the components of the transcription initiation and elongation machinery[64,65] (FIG. 2Aa). Although this approach has been successful in *Escherichia coli*, the degree of repression achieved in mammalian cells has been modest[64–68]. Chromatin-modifying repressor domains have been fused to dCas9 in an attempt to improve repression in mammalian cells[66] (FIG. 2Ab). However, the magnitude of repression displayed high variability across sgRNAs even with these fusion proteins[66]. To achieve a more robust effect, sgRNA libraries tiling the upstream regions of genes were constructed, and the variability in the measured effect on transcription was used to infer rules for the design of more-potent repressive sgRNAs[48]. These rules included the sgRNA target location relative to the transcription start site, the length of the protospacer and the spacer nucleotide composition features[48]. Although dCas9-mediated repression and RNAi-based tools seem to result in a similar molecular effect, dCas9 repression occurs by inhibiting transcription, whereas RNAi acts on the mRNAs in the cytoplasm. These differences might result in varying cellular responses.

Whereas loss-of-function screens can be conducted using a variety of both established and new Cas9-based tools, gain-of-function screens have been limited to cDNA overexpression libraries[69]. The coverage of such libraries is incomplete owing to the difficulty of cloning or expressing large cDNA constructs. Furthermore, these libraries often do not capture the full complexity of transcript isoforms, and they express genes independently of the endogenous regulatory context. To facilitate Cas9-based gain-of-function screens, synthetic activators were constructed by fusing dCas9 with transcriptional activation domains such as VP64 or p65 (REFS 68,70–73) (FIG. 2Ba). However, these fusions

© 2015 Macmillan Publishers Limited. All rights reserved

CRISPRi
An engineered transcriptional silencing complex based on catalytically inactive Cas9 (dCas9) fusions and/or single guide RNA (sgRNA) modification.

CRISPRa
An engineered transcriptional activation complex based on catalytically inactive Cas9 (dCas9) fusions and/or single guide RNA (sgRNA) modification.

only led to modest activation when delivered with a single sgRNA in mammalian cells. The delivery of multiple sgRNAs targeting the same promoter region improved target gene activation[70–72], but this was still not reliable enough to implement genome-wide activation screens. To amplify the signal of dCas9 fusion effector domains, a repeating peptide array of epitopes fused to dCas9 was developed together with activation effector domains fused to a single-chain variable fragment (ScFv) antibody[74] (FIG. 2Bb). Similar to the repression screen, a tiling approach was then used to infer rules for potent sgRNAs, followed by the design of a genome-wide library and the implementation of an activation screen[48].

We recently took advantage of a crystal structure of Cas9 in complex with a guide RNA and target single-stranded DNA (ssDNA)[75] to rationally design an efficient Cas9 activation complex composed of a dCas9 fusion protein and modified sgRNA[49] (FIG. 2Bc). This design was guided by the following principles: the use of alternative attachment positions to recruit endogenous transcription machineries more effectively; the mimicking of natural transcriptional activation mechanisms by recruiting multiple distinct activators that act in synergy to drive transcription; and the identification of design rules for efficient positioning of the Cas9 activation complex on the promoter. We used this design to implement a genome-wide gain-of-function screen[49] to identify genes that confer vemurafenib resistance in melanoma cells when upregulated.

Modified scaffolds with different RNA-binding motifs were recently developed for both activation and repression of gene expression[76] (FIG. 2C). A combination of these scaffolds enabled the execution of complex synthetic transcriptional programmes with the simultaneous activation and repression of different genes.

The most apparent advantage of dCas9-mediated transcriptional activation is that induction originates from the endogenous gene locus (unlike expression from an exogenous cDNA construct). Yet, the extent to which synthetic transcriptional modulators preserve the complexity of transcript isoforms and different types of feedback regulation remains to be tested[77,78]. In one tested case[49], two transcript isoforms were expressed at equal levels, suggesting that transcript complexity can be preserved. One important advantage of cDNA expression vectors is the ability to easily express mutated genes without modifying the endogenous genomic loci.

Libraries and screening strategies

Functional screens in cultured cells are conducted in two general formats: arrayed or pooled (FIG. 3). In an arrayed format, individual reagents are arranged in

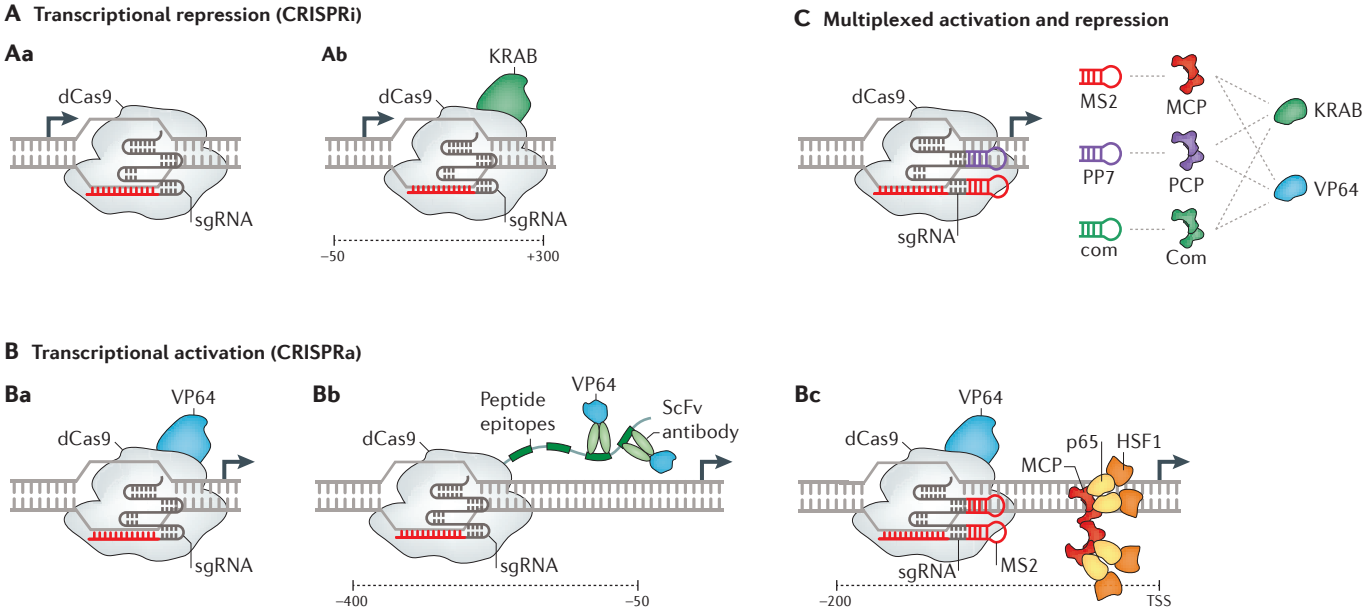


Figure 2 | **dCas9-mediated transcriptional modulation.** The different ways in which catalytically inactive Cas9 (dCas9) fusions have been used to synthetically repress (CRISPRi) or activate (CRISPRa) expression are shown. All approaches use a single guide RNA (sgRNA) to direct dCas9 to a chosen genomic location. **A** | To achieve transcriptional repression, dCas9 can be used by itself (whereby it represses transcription through steric hindrance)[64–68] (part **Aa**) or can be used as part of a dCas9–KRAB transcriptional repressor fusion protein[48,66] (part **Ab**). **B** | For transcriptional activation, various approaches have been implemented that involve the VP64 transcriptional activator. One approach is a dCas9–VP64 fusion protein[68,70–73] (part **Ba**). In an alternative method aimed at signal amplification, dCas9 is fused to a repeating array of peptide epitopes, which modularly recruit multiple copies of single-chain variable fragment (ScFv) antibodies fused to transcriptional activation domains[48,74] (part **Bb**). Another approach is a dCas9–VP64 fusion protein together with a modified sgRNA scaffold with an MS2 RNA motif loop. This MS2 RNA loop recruits MS2 coat protein (MCP) fused to additional activators such as p65 and heat shock factor 1 (HSF1)[49] (part **Bc**). **C** | Multiplexed activation and repression was implemented using an array of modified sgRNAs with different RNA recognition motifs (MS2, PP7 or com) and corresponding RNA-binding domains (MCP, PCP or Com) fused to different transcriptional effector domains (KRAB or VP64)[76]. TSS, transcriptional start site. Parts **Bb** and **C** adapted from REF. 48 and REF. 76, respectively, Cell Press; part **Bc** adapted from REF. 49, Nature Publishing Group.

© 2015 Macmillan Publishers Limited. All rights reserved

multiwell plates with a single reagent (or a small pool of reagents) per well. As each reagent is separately prepared, arrayed resources are more expensive and time consuming to produce than reagents for pooled screening, and conducting arrayed screens can require special facilities that use automation for the handling of many plates. However, in arrayed screens, where each well has a single known genetic perturbation, a much wider range of cellular phenotypes can be investigated using fluorescence, luminescence and high-content image analysis[54,79–81] (FIG. 3).

For arrayed screens, reagents can be delivered by either transfection or viral transduction. Using transfection, a large amount of plasmid DNA encoding the RNA reagent (or pre-synthesized RNA reagent) is delivered into cells, resulting in transiently high levels of functional RNA reagents (sgRNAs, shRNAs or siRNAs) until the transfected reagents are diluted out through cell division and degradation. Using viral transduction, the multiplicity of infection (MOI) can be kept low such that most cells receive a single virus that is stably integrated. These distinct kinetics of reagent expression from transfection versus viral transduction approaches can result in differences in target specificity (discussed below).

Screening reagents in pooled formats are easier to produce owing to the availability of oligonucleotide library synthesis technologies[82,83]. *In silico*-designed libraries are synthesized as a highly complex pool of oligonucleotides. These oligonucleotides are then cloned as a pool to create a plasmid library that is used for virus production and screening[26]. Unlike the transfection and viral transduction options of arrayed screens, pooled screens are limited to low-MOI viral delivery. Stable transgene integration in pooled formats facilitates screen readout using NGS. This is carried out by preparing genomic DNA from the cell population, sequencing across the sgRNA-encoding or shRNA-encoding regions of the viral integrants, and then mapping each sequencing read to a pre-compiled table of the designed sgRNA or shRNA library. This results in the quantification of the relative proportion of different integrated library constructs in the cell population.

Pooled screens are less expensive and labour intensive than arrayed screens. However, both approaches still require proficiency in molecular biology, tissue culture and data analysis. It is easier to carry out screens that require long culture times in pooled formats than in arrayed formats, as the latter often use small culture volumes (for example, 384-well plates) and require special robotic equipment for passaging many plates at once. In addition, pooled approaches enable screening in *in vivo* environments[56–58,84–86]. Conversely, pooled approaches are limited to growth phenotypes (that is, effects on cell proliferation or survival) or to cell-autonomous phenotypes that are selectable by cell sorting as fluorescence or cell surface markers.

Recent Cas9–sgRNA screens[44–49] in mammalian cell culture used a pooled screening approach with libraries that ranged from 10^3 to 10^5 sgRNAs. All of these libraries contained sgRNA redundancy (multiple distinct sgRNAs that target the same gene) and targeted either human or mouse genomes (TABLE 1). They all used cell growth as a phenotype and showed both positive and negative selection results.

In positive selection screens, a strong selective pressure is introduced such that there is only a low probability that cells without a relevant survival-enhancing perturbation will remain following selection. Commonly, positive selection experiments are designed to identify perturbations that confer resistance to a drug, toxin or pathogen. One example is a screen for host genes that are essential for the intoxication of cells by anthrax toxin[47]. In this case, most sgRNAs are depleted owing to the strong selective pressure of the toxin, and only a small number of cells, which are transduced with sgRNAs that introduce a protective mutation, survive and proliferate. As very few hits are usually expected and resistant cells continue to proliferate, the signal is strong and easy to detect in pooled approaches.

In negative selection, the goal is to identify perturbations that cause cells to be depleted during selection; such perturbations typically affect genes that are necessary for survival under the chosen selective pressure. The simplest negative selection screen is continued growth for an extended period of time: in this case, the depleted cells are those carrying reagents that target genes that are essential for cell proliferation. These genes can be found by comparing the relative frequency of each sgRNA between a late time point and an earlier one. Negative selection screens almost always require greater sensitivity to changes in the representation of library reagents, as the depletion level is more modest and the number of depleted genes is larger (for example, essential genes). Moreover, when using Cas9 nuclease, there is a chance that not all mutations will abolish gene function owing to small in-frame mutations, resulting in a mixed phenotype. One important application of negative selection screens is the identification of gene perturbations that selectively target cancer cells which harbour known oncogenic mutations; these 'oncogene addictions' might serve as possible drug targets[87,88].

Target specificity

Target specificity is an important point of consideration for all gene perturbation systems (TABLE 2). It consists of the ratio between on-target efficacy and unintended off-target effects, which is manifested by the consistency between unique reagents that target the same gene. On-target efficacy is a measure of how well a reagent can modify the expression of its intended gene target. Off-target effects include the perturbation of unintended genetic elements and global cellular responses. Target specificity will depend on the exact experimental settings. For example, as the concentrations of Cas9 and sgRNA affect target specificity[89], transient transfections will differ from low-MOI transductions in target specificity.

Gene targeting reagent consistency. One of the encouraging results observed in the initial Cas9-mediated knockout screens[44–47] was that, for the top-scoring genes, a high percentage of unique sgRNAs designed to target the same genes were enriched following

© 2015 Macmillan Publishers Limited. All rights reserved

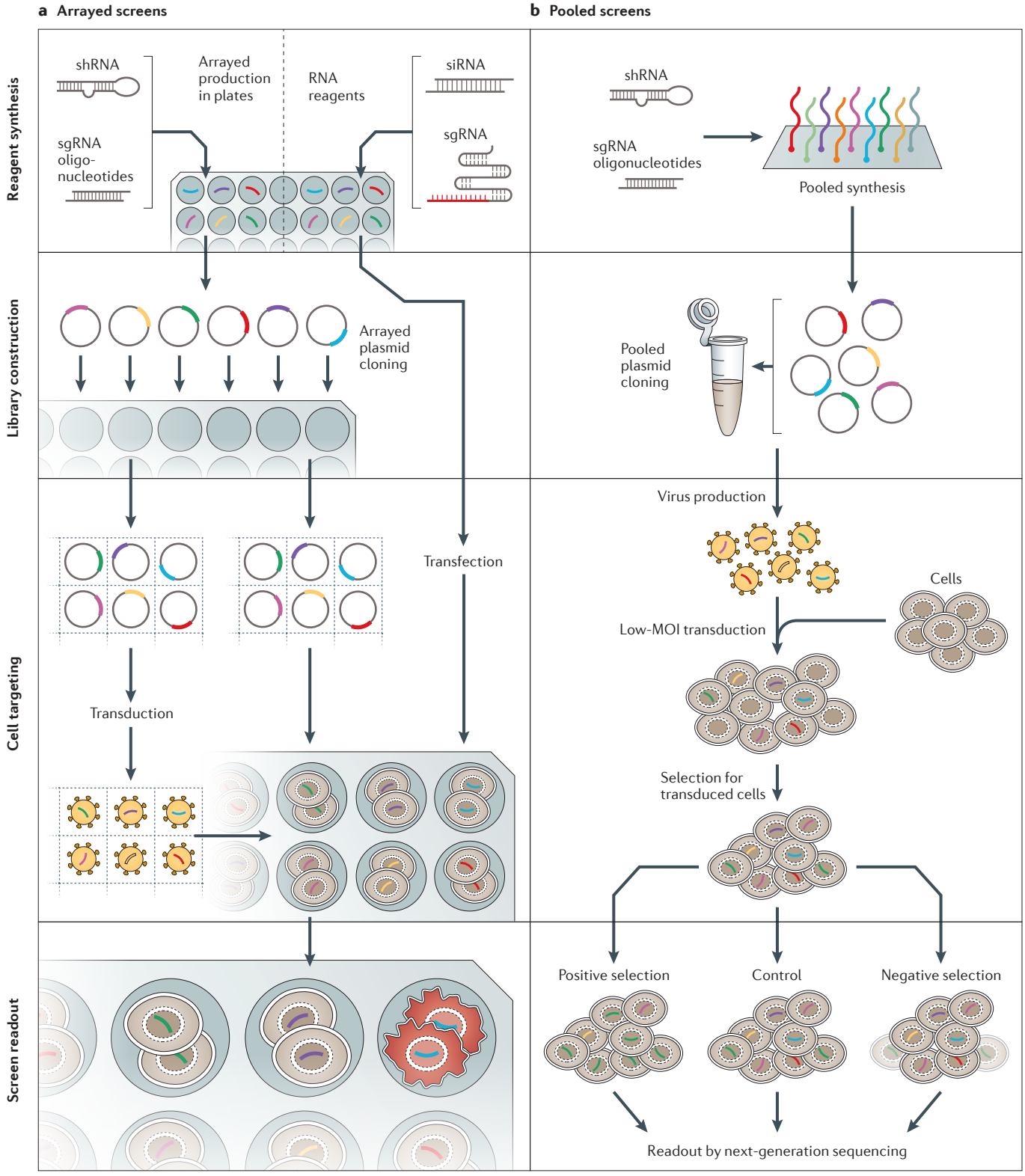


Figure 3 | **Screening strategies in either arrayed or pooled formats.** Genetic screens follow two general formats that differ in the way in which the targeting reagents are constructed and how cell targeting and readout is carried out. **a** | In arrayed screens, reagents are separately synthesized and targeting constructs are arranged in multiwell plates. Cell targeting is also conducted in multiwell plates using either transfection or viral transduction. Screen readout is based on cell population measurements in individual wells. **b** | In pooled screens, reagents are usually synthesized and constructed as a pool. Viral transduction limits transgene copy number (ideally, one perturbation per cell), and viral integration enables readout through PCR and next-generation sequencing. Readout is based on the comparison of the abundance of the different genomically integrated transgene reagents between samples. MOI, multiplicity of infection; sgRNA, single guide RNA; shRNA, short hairpin RNA; siRNA, small interfering RNA.

© 2015 Macmillan Publishers Limited. All rights reserved

Table 1 | **Experimental parameters of recent Cas9-mediated genetic screens**

Cas9 delivery	Cas9 protein	sgRNA library size	Number of targeted genes	Coverage (sgRNAs per gene)	Cell lines	Species	Positive or negative selection	Refs
Clonal isolation of stably integrated cells	Cas9 nuclease	73,151	7,114	10 and tiling sgRNAs for ribosomal genes	KBM7; HL60	Human	Both	45
Delivery with the sgRNA library	Cas9 nuclease	64,751	18,080	3 or 4 on average	A375; HUES62	Human	Both	44
Clonal isolation of stably integrated cells	Cas9 nuclease	87,897	19,150	4 on average	mESC	Mouse	Both	46
Clonal isolation of stably integrated cells	Cas9 nuclease	873	291	3	HeLa	Human	Positive	47
Polyclonal selected cell population	dCas9 repression complex	206,421	15,977	10 per TSS	K562	Human	Both	48
Polyclonal selected cell population	dCas9 activation complex	198,810	15,977	10 per TSS	K562	Human	Both	48
Polyclonal selected cell population	dCas9 activation complex	70,290	23,430	3 per TSS	A375	Human	Both	49

dCas9, catalytically inactive Cas9; mESC, mouse embryonic stem cell; sgRNA, single guide RNA; TSS, transcription start site.

positive selection. One example is a screen carried out to identify gene knockouts that confer resistance to the chemotherapy etoposide[45]. As DNA topoisomerase 2A (TOP2A) creates cytotoxic DSBs during treatment with etoposide, TOP2A depletion results in drug resistance. Impressively, all ten distinct sgRNAs for the *TOP2A* gene showed high levels of enrichment in drug-treated samples. This level of consistency is rarely observed in RNAi-based screens, resulting in the generation of very large, high-coverage RNAi reagent libraries[90]. We have observed similar results[44], in which a high percentage of sgRNAs for the top-scoring gene hits showed a strong phenotypic effect in a screen for resistance to the RAF inhibitor vemurafenib. We directly compared these results with a previous vemurafenib resistance screen using RNAi (shRNA)[91]. Interestingly, we found that the top ten hits of both screens (based on RIGER[92] analysis) shared only a single gene and that reagent consistency was much higher for the hits in the Cas9 screen (78% versus 20% of reagents enriched). In another study that aimed to identify genes involved in susceptibility to 6-thioguanine (6-TG) and susceptibility to *Clostridium septicum* α-toxin in mouse embryonic stem cells[46], both known and novel hits were found. Similarly, a higher percentage of sgRNAs were able to produce a phenotype than in shRNA knockdown when validated using individual sgRNAs for the top hits. In our positive selection vemurafenib screen, we also found a high validation rate with six of seven of the top hits reproducing the pooled screen results in arrayed-format drug titration curves[44]. Although these results are promising, more side-by-side comparisons with RNAi-based screens using different phenotypes and established RNAi screening platforms and libraries[55,93] are needed. In addition, the main results to be emphasized by the recent Cas9-knockout screens have been obtained using strong positive selection pressure. There is still a need for more-extensive validation and comparison to RNAi tools using negative selection experiments.

Despite the high consistency in strong positive selection screens, sgRNAs can still have large variations in efficiencies. This difference can be partially predicted by sgRNA sequence features[45,53] and chromatin accessibility at the target site[94], and can be used in the design of more-efficient libraries[53]. Although it is tempting to infer quantitative phenotypic information from growth-based Cas9-knockout genetic screens (for example, assigning fitness measures to gene knockouts), it is important to realize that quantitative differences in depletion or enrichment of the sgRNA-encoding constructs might result from differences in sgRNA efficiencies that cause earlier or later knockouts.

Achieving high levels of reagent consistency for dCas9-based transcriptional modulation is more challenging, as the effect of different sgRNAs will be affected by the relative distance to the transcription start site in a manner that might differ between genes. For both repression and activation, library design was guided by the unbiased testing of sets of sgRNAs[48,49]. Reassuringly, using a similar RAF inhibition positive selection experiment, we observed high levels of consistency between unique activating sgRNAs[49].

On-target loss of function and reagent efficacy. Continuous expression of the Cas9 nuclease using low-MOI lentiviral transduction can result in near-complete allelic modification owing to the irreversibility of the genomic modification[44,46,47], as long as no transgene silencing occurs. However, error-prone DSB repair will result in different mutations in different cells, and there is no guarantee that every mutation will abolish gene function. For example, small in-frame indels might not disrupt gene function. Given that every cell usually has more than one gene copy, this will result in a multimodal distribution that consists of defined null, heterozygote and wild-type expression states (FIG. 4a). This is in contrast to RNAi and dCas9 reagents that modulate transcription, which are expected to have similar effects

© 2015 Macmillan Publishers Limited. All rights reserved

Table 2 | **Features of the different perturbation tools used for targeted genetic screens**

	Loss of function			Gain of function	
	Cas9 nuclease	**CRISPRi**	**RNAi tools**	**CRISPRa**	**cDNA overexpression**
Type of perturbation	Indel mutation in the target DNA that generally results in a complete knockout owing to a coding frameshift	Repression of gene expression by dCas9-mediated transcriptional inhibition	Repression of gene expression by targeting the mRNA molecule for degradation and translational inhibition	Activation of gene expression by dCas9-mediated recruitment of transcriptional activation domains to TSSs	Exogenous overexpression of cloned cDNA constructs
Expected off-target effects	Additional unexpected indels in the genome	Repression of additional genes and effects on chromatin	Repression of additional mRNAs owing to partial 'seed' matching and imprecise Dicer processing; global effects owing to saturation of endogenous RNAi machinery (mostly relevant to siRNA transfections)	Expression of additional genes and effects on chromatin	Not many gene-specific off-target effects; global effects on translation owing to strong expression of a single gene
On-target efficacy	With continuous expression, near-complete allelic modification can be achieved in a short time frame	Inhibition level depends on the choice of sgRNA and the basal expression level of the target gene	Repression efficacy depends on the choice of RNAi tool and the specific targeting sequence	Activation level depends on the choice of sgRNA and the basal expression level of the target gene	High expression of most cDNA constructs owing to expression from the same promoter
Constitutive versus conditional expression	Cas9 expression can be made conditional	Cas9 expression can be made conditional	Only Pol II-driven RNAi reagents can be conditionally expressed	Cas9 expression can be made conditional	cDNA constructs can be conditionally expressed
Reversibility of perturbation	Irreversible	Reversible	Reversible	Reversible	Reversible
Refs	44–47	48	1	48,49	69

dCas9, catalytically inactive Cas9; Pol II, RNA polymerase II; RNAi, RNA interference; sgRNA, single guide RNA; siRNA, small interfering RNA; TSS, transcription start site.

across transduced cells, resulting in a general shift in the continuous expression distribution (FIG. 4b). This difference will not be apparent from mean expression measurements in bulk cell populations. It is worth noting that, in practice, we and others have observed an almost complete level of gene knockout at the protein level[44,46] for a limited set of tested proteins. This can be explained by additional repressive effects of Cas9 binding by steric hindrance, large in-frame deletions that still abolish gene function or a higher sensitivity to mutations at these loci. Interestingly, the distribution of indel sizes can vary between targeted loci[44–46,95,96] and can be partially predicted by the DSB flanking sequences[95], suggesting that modifications at different loci will result in different percentages of disruptive mutations and that such information can be incorporated in future libraries to achieve higher knockout efficacy.

Direct comparison of the phenotypes following Cas9 versus shRNA targeting demonstrated stronger effects of Cas9 in a few tested cases. This was shown both in pooled formats for dCas9-mediated transcriptional repression[48] and in arrayed validation[44,46] for the Cas9 nuclease. This suggests a greater efficacy of individual sgRNAs than shRNA in these cases. An advantage of using Cas9 nuclease over transcriptional modulation approaches is that mutations are irreversible and are not affected by subsequent transgene silencing. However, in RNAi, it is easier to monitor and isolate cells that harbour the intended expression perturbation. This can be achieved by co-delivery of the RNAi reagent and a reporter but, when using Cas9, sgRNA expression does not indicate the duration and magnitude of the actual genetic perturbation.

Off-target activity. Characterizing off-target effects and enhancing the specificity of both Cas9 and RNAi reagents continue to be major challenges for improving both research and clinical applications.

For Cas9-mediated genome editing, early reports demonstrated that Cas9 tolerates mismatches between the sgRNA and the target sequence across the whole recognition site in a manner that depends on the mismatch positions, number of mismatches and nucleotide identity[73,89,97–99]. In our design of genome-wide libraries, we used early empirical mismatch data[89] to choose sgRNAs with minimal predicted off-target activity[100]. Much work is still required in order to fully characterize Cas9 off-target effects. For example, recent work has suggested that small insertions or deletions ('bulges' in the sgRNA or DNA target) can also be tolerated[99].

Unbiased methods to detect Cas9-induced DSBs and Cas9-binding events are providing a more refined picture of where Cas9 binds and induces unintended modifications. Initial attempts to map off-target genome modifications using whole-genome sequencing revealed a low

© 2015 Macmillan Publishers Limited. All rights reserved

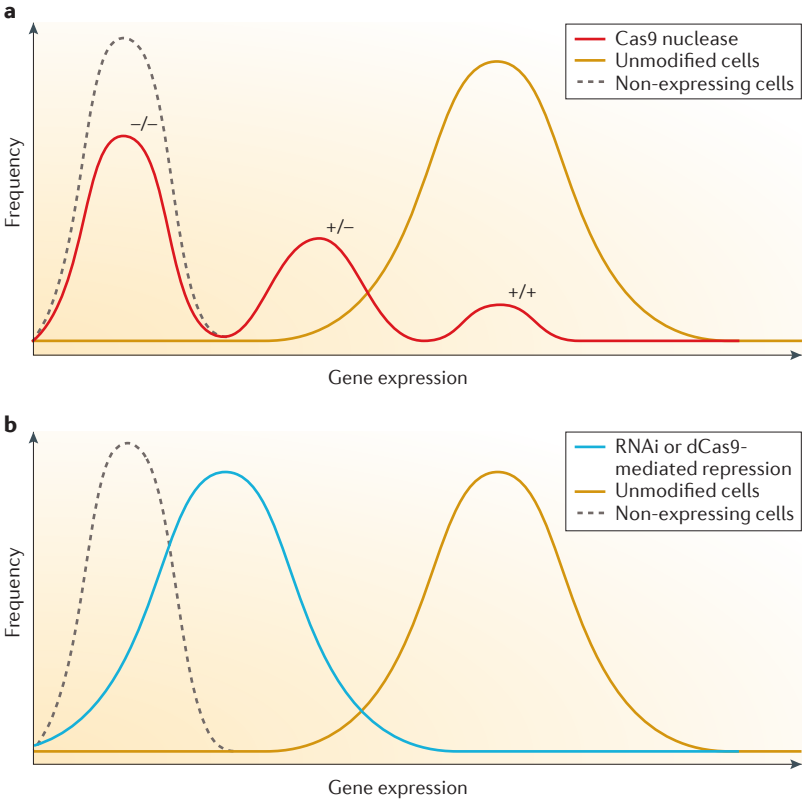


Figure 4 | **Distinct expression distributions for knockdown and knockout of a gene. a** | Theoretical target gene expression distribution following knockout mediated by lentiviral-delivered Cas9 nuclease is shown. This assumes an 80% level of allelic mutations that abolish gene function, combining out-of-frame and large deletions, close to complete allele modification rate and diploid cells. Although most cells will have a complete knockout in both alleles, some cells will retain at least one copy of a functional allele. **b** | Theoretical target gene expression distribution following catalytically inactive Cas9 (dCas9)-mediated transcriptional repression or RNA interference (RNAi)-mediated knockdown is shown. All transduced cells experience a similar perturbation that results in a shift in the target gene expression distribution.

incidence of off-target modifications[101,102]. However, this approach is limited by sequencing coverage to detect low-frequency events. Recently, unbiased detection of DSBs[103,104] revealed unexpected off-target activity that could not have been predicted using the current computational tools. Additional experiments using such unbiased methods will provide a better understanding of Cas9 target specificity. Another unbiased approach is mapping of dCas9 binding using chromatin immunoprecipitation followed by NGS (ChIP–seq)[94,105]. Such studies revealed a surprisingly large number of off-target binding events mediated by short PAM-proximal homology between the guide RNA and target sequence. Reassuringly, when this off-target binding occurs for catalytically active Cas9 it is not typically sufficient to induce DSBs, probably because the transient binding and imperfect matching of sgRNA to the target sequence is insufficient for DNA cleavage[106]. This raises concern that transcriptional modulation screens might be affected by this high incidence of transient off-target binding. However, dCas9-mediated transcriptional repression was shown to be sensitive to even a few mismatches[48], and genome-wide expression profiling exhibited specific effects for both activation and repression[48,66]. Moreover, large control sets of sgRNAs did not show any phenotypic off-target effects for both activation and repression of transcription[48]. For future library designs, specificity could be further improved using sgRNA modifications[107,108], double-nicking approaches[73,109], synthetic Cas9 protein design with improved specificity[75,110] and the use of different Cas9 orthologues[111,112].

For RNAi-based screening strategies, the characterization and avoidance of off-target effects have been subject to extensive investigation in recent years[33,35,113,114]. Early gene expression profiling studies revealed that unique siRNA reagents targeting the same genes displayed siRNA sequence-driven effects rather than signatures of target gene modulation, hinting at low target specificity[35]. This was later realized to occur as processed siRNAs enter the natural miRNA pathways that target transcripts with 3′ untranslated region (3′ UTR) sequences that have complementarity to the 5′ region of the siRNA[34]. Targeting can occur even when only eight nucleotides of the siRNA match, an effect that is similar to the 'seed region' in miRNA targeting. Recently, seed effects alone were used to identify host factors that are required for the infection of human cells by various different pathogens[115]. Although these reports were based on the transfection of large amounts of synthetic siRNA, gene silencing using different RNAi reagents, such as low-MOI transductions of shRNA or shRNAmir, would not necessarily be prone to the same level of off-target effects. Ongoing efforts to design algorithms for the more accurate prediction of targets of both endogenous miRNAs and exogenous RNAi triggers can improve both the design of RNAi reagent libraries and data analysis[116]. Finally, advances in the mechanistic understanding of miRNA biogenesis[117,118] can facilitate improved design of RNAi reagents and expression vectors that will avoid imprecise Dicer processing and produce higher levels of functional siRNAs.

To summarize, although off-target effects are a major concern for both Cas9 and RNAi approaches, they depend on the exact experimental settings and can be minimized by better mechanistic understanding and refinement of the currently used tools. Off-target effects are a major concern in clinical applications: when attempting to correct a disease-associated gene in a patient, a rare off-target mutation could potentially be toxic or oncogenic. By contrast, in genetic screens, false-positive hits owing to off-target perturbations can be easily avoided by requiring that multiple distinct reagents targeting the same genetic element display the same phenotype.

Practical considerations

Many of the technical details for conducting a genome-scale screen using Cas9 are similar to RNAi screens. These have been extensively discussed in other reviews[3,26,30,55]; thus, we focus here on topics that are specific to the use of Cas9.

False-positive
Pertaining to screening results: in a screen that results in a set of putative gene hits associated with a phenotype, a false positive is a gene that is predicted to be associated but that is actually not associated with the phenotype.

© 2015 Macmillan Publishers Limited. All rights reserved

Cas9 delivery. The most commonly used Cas9 protein, from the bacterium *S. pyogenes*, is a large protein that is encoded by a 4.1-kb coding sequence. This suggests two delivery approaches for Cas9-mediated genetic screens. The first involves the delivery of only Cas9 (viral integration or knock-ins) to generate a stable Cas9-expressing, clonal or polyclonal cell line, followed by cell expansion and the delivery of an sgRNA-only library. Clonal cell lines have the advantage that a line with high Cas9 expression levels can first be selected. However, generating a clonal cell line is not necessarily possible for all cell lines, and cells can accumulate mutations during expansion from a single cell. The second approach comprises simultaneous delivery of both Cas9 and sgRNAs using library vectors that encode both components. Although the first approach can be easily applied in immortalized cell lines, it is less feasible in primary cells that are not easily expanded in culture. For the second approach, delivery of both Cas9 and sgRNA in a single virus is challenging because viral titres can be low owing to the size of the *cas9* gene. We have recently improved the titre of the single virus system[100], thus enabling easier screening applications in primary cells or cells that are difficult to transduce. An additional option is to use a *cas9*-transgenic mouse[119], which circumvents the need for Cas9 delivery for *in vivo* or mouse-derived primary-cell screening applications.

Adeno-associated virus (AAV) vectors have advantages for *in vivo* and gene therapy applications, as they do not integrate into the genome and are thus less likely to induce oncogenesis. For pooled screening applications, the non-integrating nature of AAV vectors is less favourable because genomic integration is used to read out the abundance of the different perturbation reagents in a heterogeneous population of selected cells. However, AAV-based pooled screens might have advantages in certain *in vivo* applications: for non-dividing cells, the viral episome can be used for NGS readout. As the combined size of *S. pyogenes* Cas9 and the sgRNA cassette is already near the packaging limit of AAV, efficient *in vivo* editing by Cas9 AAV delivery requires either the delivery of two separate AAV vectors[120] or a single vector system using a smaller Cas9 orthologue from the bacterium *Staphylococcus aureus*, which we have recently adapted for *in vivo* genome editing[111].

Culture time before selection for efficient targeting. Success of Cas9 nuclease knockout screens requires a high genomic modification rate with a culture time that will suffice to deplete most of the proteins. Measurements of allelic modification rates in the first published screens demonstrated close to complete allelic modification after approximately 10 days across several gene targets[44–47]. There is no guarantee that all cell lines will display similar results, and it is important to measure allelic modification rates as a function of time across several genomic loci before using a cell line for screening.

Additional time needs to be added for the depletion of perturbed proteins. In contrast to RNAi that acts directly on the mRNA by actively degrading it, both Cas9 nuclease and dCas9-mediated protein depletion modulate transcription in the nucleus. This is combined with endogenous mRNA degradation and dilution owing to cell proliferation, and results in a slower change in mRNA levels (FIG. 1). This difference might be small in rapidly dividing cells, but depleting stable proteins in post-mitotic or even slowly dividing cells can require longer culture times post-transduction. The mode of delivery can also have an effect on the required time for gene perturbation. For example, arrayed format transfection of synthetic siRNA libraries[121] results in faster knockdown than lentiviral transduction, which requires subsequent transgene expression and nucleolytic processing to generate mature siRNAs.

Interaction with cellular machinery. The dependence on endogenous cellular pathways can introduce limitations when designing a perturbation screen. RNAi-based tools depend on an active endogenous RNAi pathway, whereas Cas9 tools act by exogenous delivery of all components (with the exception of endogenous NHEJ mechanisms, which are required for indel formation in knockout screens but which are not needed thereafter). The RNAi pathway has been associated with a wide variety of cellular processes ranging from host–pathogen interactions and cellular differentiation, to cancer[122]. Additionally, genes that are directly involved in RNAi activity cannot be continuously targeted efficiently using synthetic RNAi reagents; therefore, they may be missed if they are involved in the screened phenotypes. dCas9-mediated transcriptional repression screens can serve as a good alternative for knockdown screens in these cases, as this type of silencing is expected to use fewer endogenous pathways (FIG. 2), thus reducing the chance of having disruptive interactions between the targeted genetic element and the targeting tool.

An additional concern is the global effect of the targeting reagents on cellular physiology. The delivery of large amounts of exogenous siRNAs might saturate the endogenous RNAi system, resulting in additional toxic effects[33]. Although this is a major concern in arrayed siRNA transfection experiments, it might be less relevant to low-MOI viral-based shRNA or shRNAmir delivery. Cas9 expression in cells, and the external induction of DSBs, has not been studied in depth, and more work is still needed to establish that there are no disruptive or toxic effects.

Challenges and future outlook

Initial Cas9-mediated screens displayed remarkable results, with high levels of guide consistency, genomic modification, hit confirmation and strong phenotypic effects[44–49]. Despite these promising results, there are many aspects of using Cas9 for functional genomics that require further study. These include investigation into the cellular response to Cas9 delivery and activity in cells, and the demonstration of the same high levels of sgRNA consistency across a wider range of cell models and phenotypes. There is also a need for the

© 2015 Macmillan Publishers Limited. All rights reserved

False-negative
Pertaining to screening results: in a screen that results in a set of putative gene hits associated with a phenotype, a false negative is a true hit that was missed.

unbiased estimation of false-negative rates, as it is not clear how many of the sgRNA reagents in a particular computationally designed library actually perturb the intended targets. Although the high consistency of hit sgRNAs per gene suggests that this percentage is quite high, there is still a need for an unbiased test across multiple genomic locations. In addition, negative selection screens for growth phenotypes remain a challenge, which might be addressed by improving the efficiency of sgRNAs[53], developing more-sensitive screen readout methods and improving the statistical analysis tools.

Knockout, knockdown and activation screens are complementary methods (TABLE 2) that together will contribute to a more complete understanding of biological systems. For example, genes that retain function at low expression levels will be unlikely to display an obvious phenotype upon knockdown and might therefore be missed in knockdown screens. By contrast, genes that are essential for cell viability cannot be assessed for their contribution to additional cellular phenotypes using complete knockout; partial knockdown will be useful in these cases. In addition, as gene regulatory networks are highly interconnected and contain multiple feedback loops, the cellular phenotype in response to knockout and knockdown can be markedly different.

Screening opportunities using Cas9 extend beyond coding genes. Custom-designed sgRNA libraries can be used for the unbiased discovery of regulatory sequences by tiling sgRNAs throughout a non-coding genomic region. The delivery of multiple sgRNAs[42,72,123] can facilitate screening for epistatic effects between pairs of genes[124] or can be used to induce more-disruptive genetic modifications such as microdeletions. It is also possible to study the effects of perturbing non-coding RNAs. In this case, nuclease-induced DSBs might be suboptimal, as translational frameshift and NMD are less relevant. Instead, deletion approaches using two sgRNAs, or effective dCas9-mediated transcriptional repression, might be more suitable. In addition, fusing Cas9 to additional effector domains can facilitate high-throughput screens for phenotypic effects of additional epigenetic modifications[68]. Another type of high-throughput assay used Cas9 combined with HDR to conduct saturation mutagenesis experiments within an endogenous locus, thus expanding the possibilities of studying sequence-encoded regulatory information[125].

NGS has revolutionized our ability to read information from the genome, including the DNA sequence itself, the state of the transcriptome and the epigenome[126,127]. With these new insights into the genome, there is a need to understand the function of genetic elements through perturbation. Cas9-mediated screens will have an important role in drawing causal links between genetic architecture and phenotypes, and will enhance our ability to decipher cellular function in health and disease.

1. Boutros, M. & Ahringer, J. The art and design of genetic screens: RNA interference. *Nature Rev. Genet.* **9**, 554–566 (2008).
2. Kile, B. T. & Hilton, D. J. The art and design of genetic screens: mouse. *Nature Rev. Genet.* **6**, 557–567 (2005).
3. Grimm, S. The art and design of genetic screens: mammalian culture cells. *Nature Rev. Genet.* **5**, 179–189 (2004).
4. Jorgensen, E. M. & Mango, S. E. The art and design of genetic screens: *Caenorhabditis elegans*. *Nature Rev. Genet.* **3**, 356–369 (2002).
5. St Johnston, D. The art and design of genetic screens: *Drosophila melanogaster*. *Nature Rev. Genet.* **3**, 176–188 (2002).
6. Patton, E. E. & Zon, L. I. The art and design of genetic screens: zebrafish. *Nature Rev. Genet.* **2**, 956–966 (2001).
7. Forsburg, S. L. The art and design of genetic screens: yeast. *Nature Rev. Genet.* **2**, 659–668 (2001).
8. Page, D. R. & Grossniklaus, U. The art and design of genetic screens: *Arabidopsis thaliana*. *Nature Rev. Genet.* **3**, 124–136 (2002).
9. Shuman, H. A. & Silhavy, T. J. The art and design of genetic screens: *Escherichia coli*. *Nature Rev. Genet.* **4**, 419–431 (2003).
10. Sundaram, M. V. The love–hate relationship between Ras and Notch. *Genes Dev.* **19**, 1825–1839 (2005).
11. Nüsslein-Volhard, C., Frohnhöfer, H. G. & Lehmann, R. Determination of anteroposterior polarity in *Drosophila*. *Science* **238**, 1675–1681 (1987).
12. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
13. Driever, W. *et al.* A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37–46 (1996).
14. Haffter, P. *et al.* The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1–36 (1996).
15. Schneeberger, K. Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature Rev. Genet.* **15**, 662–676 (2014).
16. Copeland, N. G. & Jenkins, N. A. Harnessing transposons for cancer gene discovery. *Nature Rev. Cancer* **10**, 696–706 (2010).
17. Dupuy, A. J., Akagi, K., Largaespada, D. A., Copeland, N. G. & Jenkins, N. A. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* **436**, 221–226 (2005).
18. Rad, R. *et al.* PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* **330**, 1104–1107 (2010).
19. Kotecki, M., Reddy, P. S. & Cochran, B. H. Isolation and characterization of a near-haploid human cell line. *Exp. Cell Res.* **252**, 273–280 (1999).
20. Carette, J. E. *et al.* Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
21. Guo, G., Wang, W. & Bradley, A. Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* **429**, 891–895 (2004).
22. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
This paper reports the discovery of RNAi in *Caenorhabditis elegans*.
23. Ketting, R. F. The many faces of RNAi. *Dev. Cell* **20**, 148–161 (2011).
24. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
25. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genet.* **3**, 737–747 (2002).
26. Root, D. E., Hacohen, N., Hahn, W. C., Lander, E. S. & Sabatini, D. M. Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nature Methods* **3**, 715–719 (2006).
27. Silva, J. M. *et al.* Second-generation shRNA libraries covering the mouse and human genomes. *Nature Genet.* **37**, 1281–1288 (2005).
28. Chang, K., Elledge, S. J. & Hannon, G. J. Lessons from nature: microRNA-based shRNA libraries. *Nature Methods* **3**, 707–714 (2006).
29. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
30. Moffat, J. & Sabatini, D. M. Building mammalian signalling pathways with RNAi screens. *Nature Rev. Mol. Cell Biol.* **7**, 177–187 (2006).
31. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
32. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
33. Jackson, A. L. & Linsley, P. S. Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Rev. Drug Discov.* **9**, 57–67 (2010).
34. Birmingham, A. *et al.* 3′ UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nature Methods* **3**, 199–204 (2006).
35. Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nature Biotech.* **21**, 635–637 (2003).
36. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
37. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
This paper reports that Cas9 facilitates the cleavage of target DNA in bacterial cells.
38. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
This paper reports that processing of CRISPR RNA is facilitated by small non-coding transactivating crRNA (tracrRNA).
39. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
This paper reports that the Cas9 system is modular and can be transplanted into distant bacterial species to target plasmid DNA.
40. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).

© 2015 Macmillan Publishers Limited. All rights reserved

41. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
This paper, along with reference 40, characterizes Cas9-mediated DNA cleavage *in vitro*. This paper also shows that Cas9 can cleave DNA *in vitro* using chimeric sgRNAs containing a truncated tracrRNA.
42. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
43. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
References 42 and 43 describe the successful harnessing of Cas9 for genome editing.
44. Shalem, O. *et al.* Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
45. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
References 44 and 45 describe the development of lentiviral genome-scale sgRNA libraries and the application for positive and negative selection genetic screening in human cells.
46. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nature Biotech.* **32**, 267–273 (2014).
This paper describes the development of lentiviral genome-scale sgRNA libraries and the application for positive and negative selection genetic screening in mouse cells.
47. Zhou, Y. *et al.* High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
48. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
This paper describes the development and application of lentiviral genome-scale dCas9-mediated gene activation and repression for gain-of-function and loss-of-function screening.
49. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
This paper describes structure-guided engineering of a robust Cas9-based transcriptional activator and the development of a genome-scale sgRNA library for gain-of-function genetic screening.
50. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nature Rev. Genet.* **15**, 321–334 (2014).
51. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR–Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
52. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **14**, 8096–8106 (1994).
53. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR–Cas9-mediated gene inactivation. *Nature Biotech.* **32**, 1262–1267 (2014).
54. Echeverri, C. J. & Perrimon, N. High-throughput RNAi screening in cultured cells: a user's guide. *Nature Rev. Genet.* **7**, 373–384 (2006).
55. Mohr, S. E., Smith, J. A., Shamu, C. E., Neumüller, R. A. & Perrimon, N. RNAi screening comes of age: improved techniques and complementary approaches. *Nature Rev. Mol. Cell Biol.* **15**, 591–600 (2014).
56. Schramek, D. *et al.* Direct *in vivo* RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science* **343**, 309–313 (2014).
57. Beronja, S. *et al.* RNAi screens in mice identify physiological regulators of oncogenic growth. *Nature* **501**, 185–190 (2013).
58. Zhou, P. *et al. In vivo* discovery of immunotherapy targets in the tumour microenvironment. *Nature* **506**, 52–57 (2014).
59. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
60. Maillard, P. V. *et al.* Antiviral RNA interference in mammalian cells. *Science* **342**, 235–238 (2013).
61. Li, Y., Lu, J., Han, Y., Fan, X. & Ding, S.-W. RNA interference functions as an antiviral immunity mechanism in mammals. *Science* **342**, 231–234 (2013).
62. Burgess, D. J. Small RNAs: antiviral RNAi in mammals. *Nature Rev. Genet.* **14**, 821 (2013).
63. Pan, Q., van der Laan, L. J. W., Janssen, H. L. A. & Peppelenbosch, M. P. A dynamic perspective of RNAi library development. *Trends Biotechnol.* **30**, 206–215 (2012).
64. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
65. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protoc.* **8**, 2180–2196 (2013).
66. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
67. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR–Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
68. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
69. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nature Methods* **8**, 659–661 (2011).
70. Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nature Methods* **10**, 977–979 (2013).
71. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR–Cas9-based transcription factors. *Nature Methods* **10**, 973–976 (2013).
72. Cheng, A. W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163–1171 (2013).
73. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature Biotech.* **31**, 833–838 (2013).
74. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
75. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
76. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
77. Bregman, A. *et al.* Promoter elements regulate cytoplasmic mRNA decay. *Cell* **147**, 1473–1483 (2011).
78. Trcek, T., Larson, D. R., Moldón, A., Query, C. C. & Singer, R. H. Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* **147**, 1484–1497 (2011).
79. Hasson, S. A. *et al.* High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy. *Nature* **504**, 291–295 (2013).
80. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283–1298 (2006).
81. Neumann, B. *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nature Methods* **3**, 385–390 (2006).
82. LeProust, E. M. *et al.* Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res.* **38**, 2522–2540 (2010).
83. Cleary, M. A. *et al.* Production of complex nucleic acid libraries using highly parallel *in situ* oligonucleotide synthesis. *Nature Methods* **1**, 241–248 (2004).
84. Malina, A. *et al.* Repurposing CRISPR/Cas9 for *in situ* functional assays. *Genes Dev.* **27**, 2602–2614 (2013).
85. Zender, L. *et al.* An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer. *Cell* **135**, 852–864 (2008).
86. Rudalska, R. *et al. In vivo* RNAi screening identifies a mechanism of sorafenib resistance in liver cancer. *Nature Med.* **20**, 1138–1146 (2014).
87. Cheung, H. W. *et al.* Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl Acad. Sci. USA* **108**, 12372–12377 (2011).
88. Whitehurst, A. W. *et al.* Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* **446**, 815–819 (2007).
89. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotech.* **31**, 827–832 (2013).
90. Bassik, M. C. *et al.* Rapid creation and quantitative monitoring of high coverage shRNA libraries. *Nature Methods* **6**, 443–445 (2009).
91. Whittaker, S. R. *et al.* A genome-scale RNA interference screen implicates *NF1* loss in resistance to RAF inhibition. *Cancer Discov.* **3**, 350–362 (2013).
92. Birmingham, A. *et al.* Statistical methods for analysis of high-throughput RNA interference screens. *Nature Methods* **6**, 569–575 (2009).
93. Hoffman, G. R. *et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in *BRG1*-deficient cancers. *Proc. Natl Acad. Sci. USA* **111**, 3128–3133 (2014).
94. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotech.* **32**, 670–676 (2014).
95. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nature Methods* **11**, 705–706 (2014).
96. Hendel, A. *et al.* Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep.* **7**, 293–305 (2014).
97. Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature Biotech.* **31**, 839–843 (2013).
98. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nature Biotech.* **31**, 822–826 (2013).
99. Lin, Y. *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **42**, 7473–7485 (2014).
100. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nature Methods* **11**, 783–784 (2014).
101. Veres, A. *et al.* Low incidence of off-target mutations in individual CRISPR–Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**, 27–30 (2014).
102. Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12–13 (2014).
103. Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotech.* **33**, 179–186 (2015).
104. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nature Biotech.* **33**, 187–197 (2015).
105. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotech.* **32**, 677–683 (2014).
106. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
107. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
108. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR–Cas nuclease specificity using truncated guide RNAs. *Nature Biotech.* **32**, 279–284 (2014).
109. Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
110. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
111. Ran, F. A. *et al. In vivo* genome editing with *Staphylococcus aureus* Cas9. *Nature* (in the press).
112. Esvelt, K. M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature Methods* **10**, 1116–1121 (2013).
113. Qiu, S., Adema, C. M. & Lane, T. A computational study of off-target effects of RNA interference. *Nucleic Acids Res.* **33**, 1834–1847 (2005).
114. Buehler, E. *et al.* siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci. Rep.* **2**, 428 (2012).
115. Franceschini, A. *et al.* Specific inhibition of diverse pathogens in human cells by synthetic microRNA-like oligonucleotides inferred from RNAi screens. *Proc. Natl Acad. Sci. USA* **111**, 4548–4553 (2014).
116. Zhong, R. *et al.* Computational detection and suppression of sequence-specific off-target phenotypes from whole genome RNAi screens. *Nucleic Acids Res.* **42**, 8214–8222 (2014).
117. Gu, S. *et al.* The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing *in vivo*. *Cell* **151**, 900–911 (2012).

© 2015 Macmillan Publishers Limited. All rights reserved

118. Frank, F., Sonenberg, N. & Nagar, B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465**, 818–822 (2010).
119. Platt, R. J. *et al.* CRISPR–Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
120. Swiech, L. *et al. In vivo* interrogation of gene function in the mammalian brain using CRISPR–Cas9. *Nature Biotech.* **33**, 102–106 (2015).
121. Dorsett, Y. & Tuschl, T. siRNAs: applications in functional genomics and potential as therapeutics. *Nature Rev. Drug Discov.* **3**, 318–329 (2004).
122. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
123. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
124. Bassik, M. C. *et al.* A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. *Cell* **152**, 909–922 (2013).
125. Findlay, G. M., Boyle, E. A., Hause, R. J., Klein, J. C. & Shendure, J. Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* **513**, 120–123 (2014).
126. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc. Natl Acad. Sci. USA* **111**, 6131–6138 (2014).
127. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* **306**, 636–640 (2004).

Acknowledgements
The authors thank L. Solomon for help with illustrations, J. Wright for manuscript review and members of the Zhang Laboratory for discussions. O.S. is supported by a Klarman Family Foundation Fellowship. N.E.S. is supported by a Simons Center for the Social Brain Postdoctoral Fellowship and by the National Human Genome Research Institute (NHGRI) of the US National Institutes of Health under award number K99-HG008171. F.Z. is supported by the US National Institute of Mental Health (NIMH) (DP1-MH100706), the US National Institute of Neurological Disorders and Stroke (NINDS) (R01-NS07312401), a US National Science Foundation (NSF) Waterman Award, the Keck, Damon Runyon, Searle Scholars, Klingenstein, Vallee, Merkin, Simons, and New York Stem Cell Foundations, and Bob Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator.

Competing interests statement
The authors declare competing interests: see Web version for details.

© 2015 Macmillan Publishers Limited. All rights reserved