Genome-scale CRISPR pooled screens

Neville E. Sanjana [a, b]

[a] New York Genome Center, New York, NY 10013, USA
[b] Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10012, USA

ARTICLE INFO

ABSTRACT

Genome editing technologies such as clustered regularly interspaced short palindromic repeats (CRISPR) systems have ushered in a new era of targeted DNA manipulation. The easy programmability of CRISPR using short oligonucleotides enables rapid synthesis of large-scale libraries for functional genetic screens. Here we present fundamental concepts and methods for pooled CRISPR screens and review biological results from recent genome-scale loss-of-function and gain-of-function screens. We also discuss new frontiers in pooled screens, including novel effector domains for functional screens and applications in the noncoding genome.

Pooled genetic screens, where each cell receives a different genetic perturbation prior to a phenotype-based selection, are a powerful technique for rapidly identifying specific genome elements that influence a selected phenotype. A single successful screen may provide a treasure trove of new genotype–phenotype interactions that can be used as a launching point for follow-up studies on the most promising gene targets. The key idea behind pooled screens is that phenotypic selection results in an enrichment (or depletion) of genetic perturbations relevant to the phenotype. An example of a selected phenotype can be resistance to a drug or expression of a particular cell surface receptor. In this way, pooled screens can search over a space of thousands of perturbations (e.g., genome scale), providing a substantially simpler technique than arrayed format screens where each perturbation is delivered in a separate well.

Traditionally, to generate the pool of perturbations, these screens have relied on DNA mutagenesis (induced through chemical mutagens, mobile genetic element insertion, or radiation) or manipulations at the transcript level such as RNA interference (RNAi). However, pooled screens using DNA mutagenesis can require painstaking work to map the mutation sites. Pooled screening at the transcript level with RNAi also presents challenges such as incomplete knockdown and large off-target effects [1,2].

Recently, the easy programmability of microbial clustered regularly interspaced short palindromic repeats (CRISPR) systems has created a new opportunity for performing pooled screens at the DNA level in a targeted fashion [3,4]. In their native context, CRISPR nucleases function as a prokaryotic immune system, cleaving foreign DNA from phage or plasmids before they can damage the host cell [5–7]. The nuclease is guided to a specific DNA sequence using a short, single-stranded RNA that is complementary to the target DNA. Genome editing applications using Streptococcus pyogenes Cas9 [8,9], the most commonly used CRISPR system, replace the endogenous prokaryotic RNA components with a synthetic single-guide RNA (sgRNA) that has a 20-bp sequence complementary to the target site [10]. After the Cas9–sgRNA ribonucleoprotein complex recognizes the target sequence in the genome, it creates a double-strand break (DSB). DSB repair mechanisms, such as non-homologous end-joining (NHEJ), can delete or add a few bases during the repair process. When the NHEJ-mediated repair occurs in a coding region, this can introduce a frameshift mutation where the net result is gene loss of function. As reviewed elsewhere, the Cas9 system results in high-efficiency genome modification in diverse genetic model (and non-model) organisms [11].

Using this RNA-guided system to produce targeted loss-of-function mutations, it is possible to build large, genome-scale libraries of CRISPR reagents for high-throughput pooled screens. The combination of DNA-level manipulation with easy targeting/programmability presents a new avenue for understanding the function of genome (and epigenome) elements in a wide variety of functional assays.

Technological underpinnings of CRISPR screens

Within months of the initial use of Cas9 for mammalian genome editing [8,9], several groups (including our own) took advantage of existing pooled synthesis and cloning techniques to construct large Cas9–sgRNA lentiviral libraries using methods similar to those previously used for large RNAi lentiviral libraries [3,4,12]. Typically, libraries are designed with multiple sgRNAs targeting each gene. Consistent changes in multiple sgRNAs can be used to increase confidence in a particular candidate gene. Briefly, libraries are synthesized as DNA and cloned into plasmids to produce lentivirus. After designing sgRNA guide sequences to target different genes, the oligonucleotides are synthesized in a pool of typically 10^4 to 10^5 different guides. Several different commercial platforms exist for producing pooled oligonucleotides (e.g., CustomArray, Twist Bioscience, Agilent Technologies), and synthesis fidelity and length is constantly improving. The pooled oligos are cloned into the sgRNA cassette of the backbone, and the resulting lentivirus expresses the Cas9 protein and target-specific sgRNA (Fig. 1A); alternatively, for higher viral titer, the small sgRNA can be delivered on a separate virus from the large Cas9 protein [13].

Lentivirus is produced from the cloned plasmid pool and then used to transduce mammalian cells for screening. Because lentiviruses integrate into the genome, the viral integrant serves as a tag for readout of which sgRNA construct is delivered to a particular cell. Viruses are introduced at a low multiplicity of infection (MOI) in order to ensure that cells receive only a single sgRNA from the library pool. The frequency of observing any particular tag (sgRNA) before and after phenotypic selection is the key parameter measured during a pooled screen. This frequency can be computed by taking a representative sample of the population, polymerase chain reaction (PCR) amplifying the lentiviral cassettes and their sgRNAs, and then counting the frequency of each sgRNA after next-generation sequencing (Fig. 1B). Ideally, the initial distribution of the library is as uniform as possible so that, after selection, any depletion or enrichment of specific sgRNAs is readily identifiable. In reality, biases can be introduced at many different stages, such as library synthesis, cloning, viral transduction, cell sampling, and PCR amplification, making it important to carefully control each stage of library handling before sequencing.

After selection, changes in sgRNA representation (enrichment or depletion) indicate potential candidates for further study (Fig. 1B). A key concept in screen analysis is that the strongest *gene* candidates are those that have *multiple sgRNAs* simultaneously enriched/depleted. In enrichment screens, cells carrying specific sgRNAs are selected for. Representative examples of enrichment are drug resistance or toxin screens, where a gene knockout promotes cell survival after exposure to a drug/toxin, and fluorescence-activated cell sorting (FACS)-based selection of cells with a fluorescent gene reporter to find mutations that modulate gene expression. By contrast, depletion screens can identify essential genes, where loss of function is incompatible with continued cell survival, and synthetic lethal interactions, where typically nonessential genes become essential when cells carry a particular (usually oncogenic) mutation in a different gene. Depletion analysis in a drug or toxin screen can also be used to find gene targets that enhance sensitivity to the drug or toxin.

Applications of pooled CRISPR screens

Over the past 2 years, pooled CRISPR screens have been deployed in diverse phenotypic assays: cancer drug resistance [3,4,14], bacterial toxin resistance [12], West Nile virus-induced cell death [15], mitochondrial metabolism [16], identifying essential and synthetic lethal genes in cancer cell lines [17–19], identifying genetic drivers of metastasis in an in vivo screen [20], and understanding gene networks in immune cells [21,22]. In addition, two catalytically inactive versions of Cas9 with arrays of transcriptional activation domains (in conjunction with sgRNA libraries targeting promoter regions) have facilitated genome-scale *gain-of-function* screens (Fig. 2) [23,24]. To illustrate the utility of pooled CRISPR screens, here we focus on two examples of loss-of-function screens: an enrichment screen and a set of related depletion screens.

One of the first pooled screens was a positive selection screen for drug resistance in a human BRAF mutant (V600E) melanoma cell line using a genome-scale library targeting approximately 18,000 genes with approximately 65,000 sgRNAs [3]. This BRAF gain-of-function mutation is found in more than 50% of malignant melanomas, and vemurafenib, a U.S. Food and Drug Administration-approved BRAF inhibitor, was shown to induce apoptosis preferentially in cells with the mutant form of BRAF [25]. In the clinic, vemurafenib results in reduction of solid tumors, but within weeks resistance to the drug develops, which is followed by tumor regrowth and a poor prognosis. Identifying mechanisms of resistance can help to inform the design of new treatment strategies to combine targeted BRAF inhibitors with other drugs. In the BRAF mutant melanoma cells, the genome-scale CRISPR screen identified both of the previously established loss-of-function mechanisms of vemurafenib resistance and several novel gene targets [26,27]. Importantly, the consistency of different reagents targeting the same gene was higher in the CRISPR screen than in the same screen using a genome-scale lentiviral RNAi library [27]. For example, the 10 most highly enriched genes in the RNAi screen had, on average, only 20% of the reagents targeting each gene enriched, indicating little agreement between different RNAi reagents designed to target the same gene. The CRISPR screen had, on average, 80% of reagents targeting each gene enriched for the 10 most highly enriched genes. This large difference in consistency suggests that the false-positive rate is lower using CRISPR reagents (further confirmed in a recent detailed comparison between screening technologies [28]) and, thus, gives us greater confidence in the identified gene candidates. As expected, individual loss-of-function mutations in several candidates from the CRISPR screen were shown to confer vemurafenib resistance in follow-up experiments [3], whereas only one gene from the RNAi screen could be validated [27]. A practical advantage of the lower false-positive rate is that genome-wide CRISPR screens can be conducted with smaller libraries (and thus using fewer cells) than a comparable genome-wide RNAi screen.

Two recent articles used depletion screens in multiple cell lines to estimate the percentage of genes in the human genome that are essential [17,18]. By measuring sgRNAs that are consistently depleted across multiple cell lines, they estimated that 1700 genes (~10% of all protein coding genes) are essential, which is approximately 5-fold higher than previous estimates from RNAi-based screens [29,30]. One reason for this increase is that RNAi knockdown is incomplete, whereas CRISPR is able to modify all copies of a gene when delivered using lentivirus and expressed constitutively [3,17]. This advantage is also relevant when comparing CRISPR
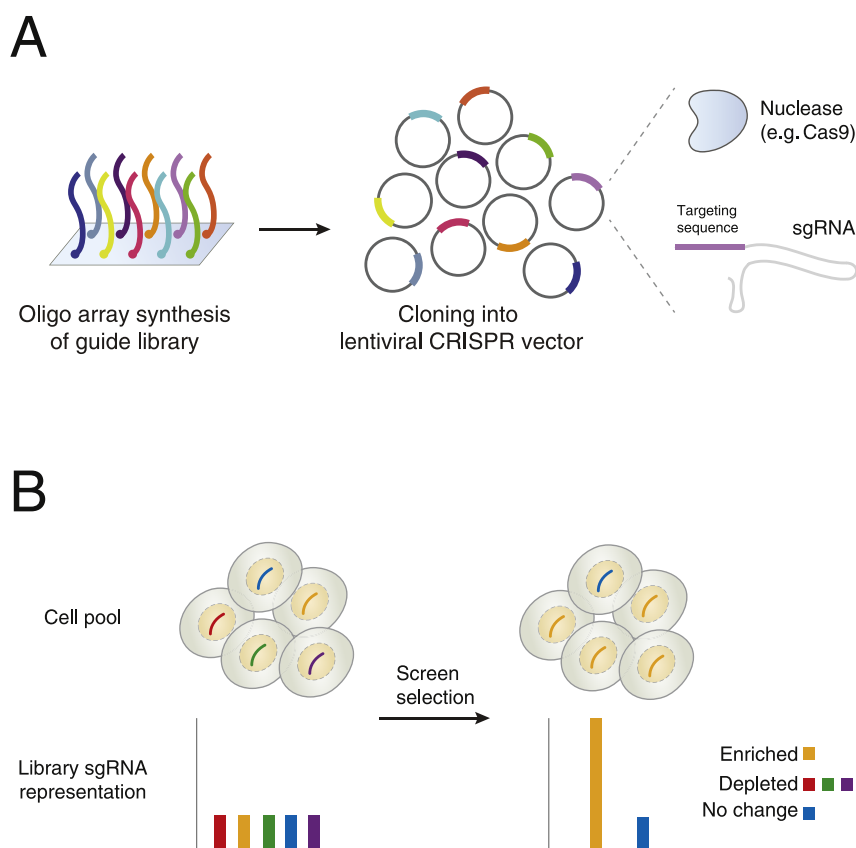
Fig.1. Synthesis of CRISPR libraries and changes in single-guide RNA (sgRNA) representation after enrichment or depletion. (A) Single-stranded DNA oligonucleotides are synthesized on an array. Each one contains a specific targeting sequence (e.g., ~20-bp spacer sequence) flanked by universal sequences for PCR amplification. After amplification, they are cloned into a lentiviral vector that contains a cassette for the approximately 20-bp spacer (expressed as part of the sgRNA) and that also encodes the nuclease (e.g., Cas9). After viral production and transduction, the nuclease and the sgRNA are expressed in the target cells. (B) Library representation in a pool of cells before and after phenotypic selection is shown. Initial library representation after viral transduction is uniform across different CRISPR reagents. After screen selection, the representation of the library has changed, whereby 1 sgRNA (yellow) is enriched, 3 sgRNAs (red, green, and purple) are depleted, and 1 sgRNA (blue) did not change in abundance. At each time-point, the representation of the library is assayed by taking a sample of the cell population, extracting genomic DNA, and PCR amplifying from the genomically integrated lentiviral cassettes. After next-generation sequencing of the pool, the frequency of each sgRNA is counted.

screens with other DNA-based screening technologies: Wang and coworkers used both CRISPR knockout and a gene trap technique using retroviral insertions in the near-haploid KBM7 cell line [17]. The gene trap screen was unable to detect any essential genes on chromosome 8, which is the only diploid chromosome in KBM7 cells. Because gene trap screens rely on random insertion in the genome, complete knockout is possible only in haploid cells, which have only one copy of each gene. This issue is especially relevant in cancer cell lines where multi-copy gene amplifications are more common. Finally, by comparing essential genes across different cell lines, both articles identified core essential genes and a much smaller set of context-dependent essential genes that were essential only in particular lines [17,18].

New frontiers in CRISPR screens

Since their initial development, pooled CRISPR screens have delivered new insights into basic biology and in applied, clinically relevant domains, yet there are still many opportunities for further development of this technology. Although in vivo screens have established a first step toward understanding cell–cell interactions [20], most CRISPR screens have been performed with a single cell type in isolation. New screens could replace drug selection with a cell-based selective pressure (e.g., T cell-mediated targeting of cancer cells) or combine library-transduced cells with other disease-relevant cell types (e.g., cancer cells cultured with endothelial or immune cells). It also is possible to deliver different pooled libraries to different populations of cells, capitalizing on recent cell type-specific gene expression data from resources such as the Genotype–Tissue Expression (GTEx) project and the Allen Brain Atlas [31,32]. For example, in a co-culture system of neurons and glia, each cell type could separately be transduced with a specific library (targeting either neuron- or glia-relevant genes) before being cultured together. Along similar lines, combinatorial delivery of two or more sgRNAs in a single construct would allow targeting of multiple genes within one cell (Fig. 2). However, it will be necessary to use smaller targeted libraries to avoid a combinatorial explosion: A genome-scale CRISPR library with 10^5 distinct sgRNAs would require (at 10^3 cells/construct) an impractical 10 trillion cells to screen all possible pairs of sgRNAs, but a library with 10^2 to 10^3 distinct sgRNAs is feasible given current screen handling and cell culture capacities.

The success in repurposing CRISPR using transcriptional activation domains for gain-of-function screens suggests that several other kinds of screens might be possible by attaching different epigenetic effector domains to a catalytically inactive CRISPR nuclease. These could act to modulate chromatin states or alter other epigenetic elements at specific genomic locations [33,34].
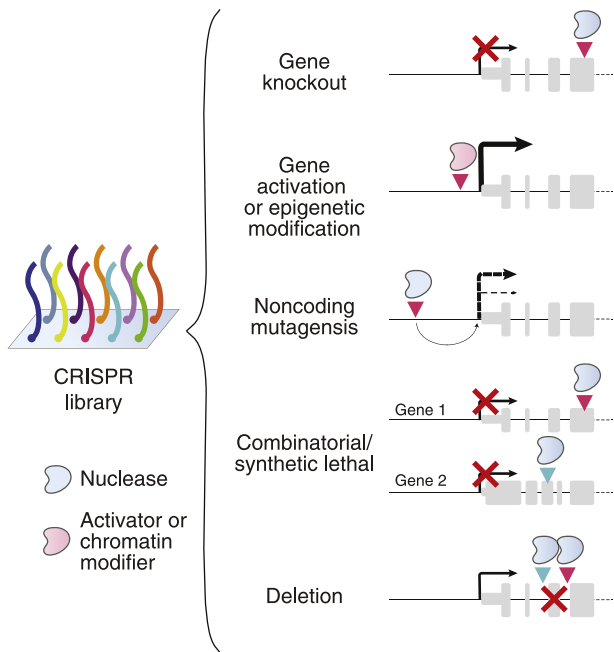
Fig. 2. Applications of CRISPR libraries for gene manipulation. Targeting of CRISPR nuclease/activator combined with sgRNAs from the libraries to specific gene features can be used for different kinds of screens. The specific screening application dictates the library sgRNA design and the type of CRISPR enzyme (e.g., nuclease or activator) used. *Gene knockout*: The sgRNA targets the CRISPR nuclease (e.g., Cas9) to an exon, where resulting frameshift mutations can trigger nonsense-mediated decay of messenger RNA and block production of functional protein products. *Gene activation or chromatin modification*: The sgRNA targets the CRISPR activator/epigenetic modifier complex (e.g., a catalytically inactive form of Cas9 with other functional domains added directly or as part of a larger complex) near the promoter of the gene to drive overexpression from the endogenous gene locus. *Noncoding mutagenesis*: The sgRNA targets the CRISPR nuclease to a putative regulatory element in the noncoding genome. *Combinatorial/synthetic lethal*: Sets of sgRNAs are delivered into the same cell target CRISPR nucleases to the exons of different genes to examine possible interactions between loss of function for multiple genes. *Deletion*: Pairs of sgRNAs are delivered into the same cell target CRISPR nucleases to nearby regions that, at some frequency, delete the intervening segment. This can be used to remove either coding elements such as exons (as depicted) or to delete larger noncoding elements (e.g., intergenic noncoding RNAs) that are more difficult to manipulate using single sgRNA mutagenesis.

Another consideration is *where* in the genome to target, either with nuclease or other effector domains. Currently, there has been little work on using pooled CRISPR screening approaches to manipulate noncoding genome elements such as noncoding RNAs (e.g., micro RNAs [miRNAs], large intergenic noncoding RNAs [lincRNAs]), enhancers, repressors, promoters, and structural elements (e.g., CCCTC binding factor [CTCF] binding sites) [13,20,35]. Delivery of multiple guides with a nuclease may enable large deletions, which would be useful for assessing the importance of larger noncoding elements. Both dual sgRNA-based deletions and single sgRNA scanning mutagenesis will be helpful for locating and understanding functional elements in the noncoding genome. From a technological standpoint, most CRISPR screens have focused on loss of function because of the high efficiency of NHEJ-mediated DSB repair after nuclease cutting, but it would be more desirable to precisely knock-in specific human genetic variants (e.g., single-nucleotide polymorphisms [SNPs], copy number variations [CNVs]) at predefined locations. Although doable at a single genomic locus (cf. CRISPR-based pooled saturating mutagenesis of three residues in the gene BRCA1 [36]), this type of precision editing is not yet possible at multiple genomic loci; however, given the rapid pace of genome engineering technology development, this may change soon.

Compared with RNAi and other pooled screening technologies, CRISPR pooled screens have demonstrated higher consistency between different reagents targeting the same genetic element, and because of this they have already found broad applicability and produced new biological insights. Given the many promising future directions, CRISPR pooled screens have enormous potential to improve our understanding of the genome and to identify the causal role of genome elements in both normal and disease states.

Acknowledgments

References

[1] C.J. Echeverri, P.A. Beachy, B. Baum, M. Boutros, F. Buchholz, S.K. Chanda, et al., Minimizing the risk of reporting false positives in large-scale RNAi screens, Nat. Methods 3 (2006) 777–779.

[2] S. Singh, X. Wu, V. Ljosa, M.-A. Bray, F. Piccioni, D.E. Root, et al., Morphological profiles of RNAi-induced gene knockdown are highly reproducible but dominated by seed effects, PLoS One 10 (7) (2015) e0131370.

[3] O. Shalem, N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T.S. Mikkelsen, et al., Genome-scale CRISPR–Cas9 knockout screening in human cells, Science 343 (2014) 84–87.

[4] T. Wang, J.J. Wei, D.M. Sabatini, E.S. Lander, Genetic screens in human cells using the CRISPR–Cas9 system, Science 343 (2014) 80–84.

[5] F.J.M. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, J. Mol. Evol. 60 (2005) 174–182.

[6] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiol. Read. UK 151 (2005) 653–663.

[7] R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, et al., CRISPR provides acquired resistance against viruses in prokaryotes, Science 315 (2007) 1709–1712.

[8] P. Mali, L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, et al., RNA-guided human genome engineering via Cas9, Science 339 (2013) 823–826.

[9] L. Cong, F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, et al., Multiplex genome engineering using CRISPR/Cas systems, Science 339 (2013) 819–823.

[10] M.M. Jinek, K.K. Chylinski, I. Fonfara, M.M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science 337 (2012) 816–821.

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

[12] H. Koike-Yusa, Y. Li, E.-P. Tan, M.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.

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

[14] J. Shi, E. Wang, J.P. Milazzo, Z. Wang, J.B. Kinney, C.R. Vakoc, Discovery of cancer drug targets by CRISPR–Cas9 screening of protein domains, Nat. Biotechnol. 33 (2015) 661–667.

[15] H. Ma, Y. Dang, Y. Wu, G. Jia, E. Anaya, J. Zhang, et al., A CRISPR-based screen identifies genes essential for West-Nile-Virus-induced cell death, Cell Rep. 12 (2015) 673–683.

[16] K. Birsoy, T. Wang, W.W. Chen, E. Freinkman, M. Abu-Remaileh, D.M. Sabatini, An essential role of the mitochondrial electron transport chain in cell proliferation is to enable aspartate synthesis, Cell 162 (2015) 540–551.

[17] T. Wang, K. Birsoy, N.W. Hughes, K.M. Krupczak, Y. Post, J.J. Wei, et al., Identification and characterization of essential genes in the human genome, Science 350 (2015) 1096–1101.

[18] T. Hart, M. Chandrashekhar, M. Aregger, Z. Steinhart, K.R. Brown, G. MacLeod, et al., High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities, Cell 163 (2015) 1515–1526.

[19] C.M. Toledo, Y. Ding, P. Hoellerbauer, R.J. Davis, R. Basom, E.J. Girard, et al., Genome-wide CRISPR–Cas9 screens reveal loss of redundancy between PKMYT1 and WEE1 in glioblastoma stem-like cells, Cell Rep. 13 (2015) 2425–2439.

[20] S. Chen, N.E. Sanjana, K. Zheng, O. Shalem, K. Lee, X. Shi, et al., Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis, Cell 160 (2015) 1246–1260.

[21] O. Parnas, M. Jovanovic, T.M. Eisenhaure, R.H. Herbst, A. Dixit, C.J. Ye, et al., A genome-wide CRISPR screen in primary immune cells to dissect regulatory networks, Cell 162 (2015) 675–686.

[22] J.L. Schmid-Burgk, D. Chauhan, T. Schmidt, T.S. Ebert, J. Reinhardt, E. Endl, et al., A genome-wide CRISPR screen identifies NEK7 as an essential component of NLRP3 inflammasome activation, J. Biol. Chem. 291 (2016) 103–109.

[23] S. Konermann, M.D. Brigham, A.E. Trevino, J. Joung, O.O. Abudayyeh, C. Barcena, et al., Genome-scale transcriptional activation by an engineered CRISPR—Cas9 complex, Nature 517 (2015) 583—588.

[24] L.A. Gilbert, M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, et al., Genome-scale CRISPR-mediated control of gene repression and activation, Cell 159 (2014) 647—661.

[25] M. Das Thakur, D.D. Stuart, The evolution of melanoma resistance reveals therapeutic opportunities, Cancer Res. 73 (2013) 6106—6110.

[26] S. Huang, M. Hölzel, T. Knijnenburg, A. Schlicker, P. Roepman, U. McDermott, et al., MED12 controls the response to multiple cancer drugs through regulation of TGF-β receptor signaling, Cell 151 (2012) 937—950.

[27] S.R. Whittaker, J.-P. Theurillat, E. Van Allen, N. Wagle, J. Hsiao, G.S. Cowley, et al., A genome-scale RNA interference screen implicates NF1 loss in resistance to RAF inhibition, Cancer Discov. 3 (2013) 350—362.

[28] B. Evers, K. Jastrzebski, J.P.M. Heijmans, W. Grernrum, R.L. Beijersbergen, R. Bernards, CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes, Nat. Biotechnol. (2016), http://dx.doi.org/10.1038/nbt.3536.

[29] B. Luo, H.W. Cheung, A. Subramanian, T. Sharifnia, M. Okamoto, X. Yang, et al., Highly parallel identification of essential genes in cancer cells, Proc. Natl. Acad. Sci. USA. 105 (2008) 20380—20385.

[30] R. Marcotte, K.R. Brown, F. Suarez, A. Sayad, K. Karamboulas, P.M. Krzyzanowski, et al., Essential gene profiles in breast, pancreatic, and ovarian cancer cells, Cancer Discov. 2 (2012) 172—189.

[31] GTEx Consortium, The Genotype—Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans, Science 348 (2015) 648—660.

[32] E.S. Lein, M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, et al., Genome-wide atlas of gene expression in the adult mouse brain, Nature 445 (2007) 168—176.

[33] P.I. Thakore, A.M. D'Ippolito, L. Song, A. Safi, N.K. Shivakumar, A.M. Kabadi, et al., Highly specific epigenome editing by CRISPR—Cas9 repressors for silencing of distal regulatory elements, Nat. Methods 12 (2015) 1143—1149.

[34] N.A. Kearns, H. Pham, B. Tabak, R.M. Genga, N.J. Silverstein, M. Garber, et al., Functional annotation of native enhancers with a Cas9—histone demethylase fusion, Nat. Methods 12 (2015) 401—403.

[35] M.C. Canver, E.C. Smith, F. Sher, L. Pinello, N.E. Sanjana, O. Shalem, et al., BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Nature 527 (2015) 192—197.

[36] 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.