Published online 23 March 2015

Nucleic Acids Research, 2015, Vol. 43, No. 7 ***3407–3419***
doi: 10.1093/nar/gkv226

SURVEY AND SUMMARY

Advances in CRISPR-Cas9 genome engineering: lessons learned from RNA interference

Rodolphe Barrangou[1,†], Amanda Birmingham[2,†], Stefan Wiemann[3], Roderick L. Beijersbergen[4], Veit Hornung[5] and Anja van Brabant Smith[2,*]

[1]Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC 27695, USA, [2]Dharmacon, part of GE Healthcare, Lafayette, CO 80026, USA, [3]Division of Molecular Genome Analysis, and Genomic & Proteomics Core Facility, German Cancer Research Center, 69120 Heidelberg, Germany, [4]The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands and [5]Institute of Molecular Medicine, University Hospital, University of Bonn, 53128 Bonn, Germany

Received January 15, 2015; Revised March 4, 2015; Accepted March 5, 2015

ABSTRACT

The discovery that the machinery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 bacterial immune system can be repurposed to easily create deletions, insertions and replacements in the mammalian genome has revolutionized the field of genome engineering and reinvigorated the field of gene therapy. Many parallels have been drawn between the newly discovered CRISPR-Cas9 system and the RNA interference (RNAi) pathway in terms of their utility for understanding and interrogating gene function in mammalian cells. Given this similarity, the CRISPR-Cas9 field stands to benefit immensely from lessons learned during the development of RNAi technology. We examine how the history of RNAi can inform today's challenges in CRISPR-Cas9 genome engineering such as efficiency, specificity, high-throughput screening and delivery for *in vivo* and therapeutic applications.

INTRODUCTION

From early classical genetic studies to present-day molecular ones, the ability to modulate gene content and expression has been essential to understanding the function of genes within biological pathways and their correlation with disease phenotypes. The discovery of RNAi and its reduction to practice in mammalian cells in the early to mid 2000's made reverse genetics approaches feasible on a genome scale in higher eukaryotes (1). In the last 24 months, another gene modulation technique, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 genome engineering (referred to as CRISPR-Cas9), has emerged; in that remarkably brief window, this approach has proven to be a powerful tool for studying individual gene function, performing genome-wide screens, creating disease models and perhaps developing therapeutic agents (2). These lightning advances have largely followed the path blazed by RNAi studies and we argue that further leverage is to be gained by examining relevant successes and failures in the last 14 years of RNAi.

RNAi and CRISPR-Cas9 have many clear similarities. Indeed, the mechanisms of both use small RNAs with an on-target specificity of ∼18–20 nt. Both methods have been extensively reviewed recently (3–5) so we only highlight their main features here. RNAi operates by piggybacking on the endogenous eukaryotic pathway for microRNA-based gene regulation (Figure 1A). microRNAs (miRNAs) are small, ∼22-nt-long molecules that cause cleavage, degradation and/or translational repression of RNAs with adequate complementarity to them (6). RNAi reagents for research aim to exploit the cleavage pathway using perfect complementarity to their targets to produce robust down-regulation of only the intended target gene. The CRISPR-Cas9 system, on the other hand, originates from the bacterial CRISPR-Cas system, which provides adaptive immunity against invading genetic elements (7). Generally, CRISPR-Cas systems provide DNA-encoded (7), RNA-mediated (8), DNA- (9) or RNA-targeting(10) sequence-specific targeting. Cas9 is the signature protein for Type II CRISPR-Cas systems (11), in which gene editing is me-

*To whom correspondence should be addressed. Tel: +1 720 890 5188; Fax: +1 303 604 3286; Email: anja.smith@ge.com
†These authors contributed equally to this work.

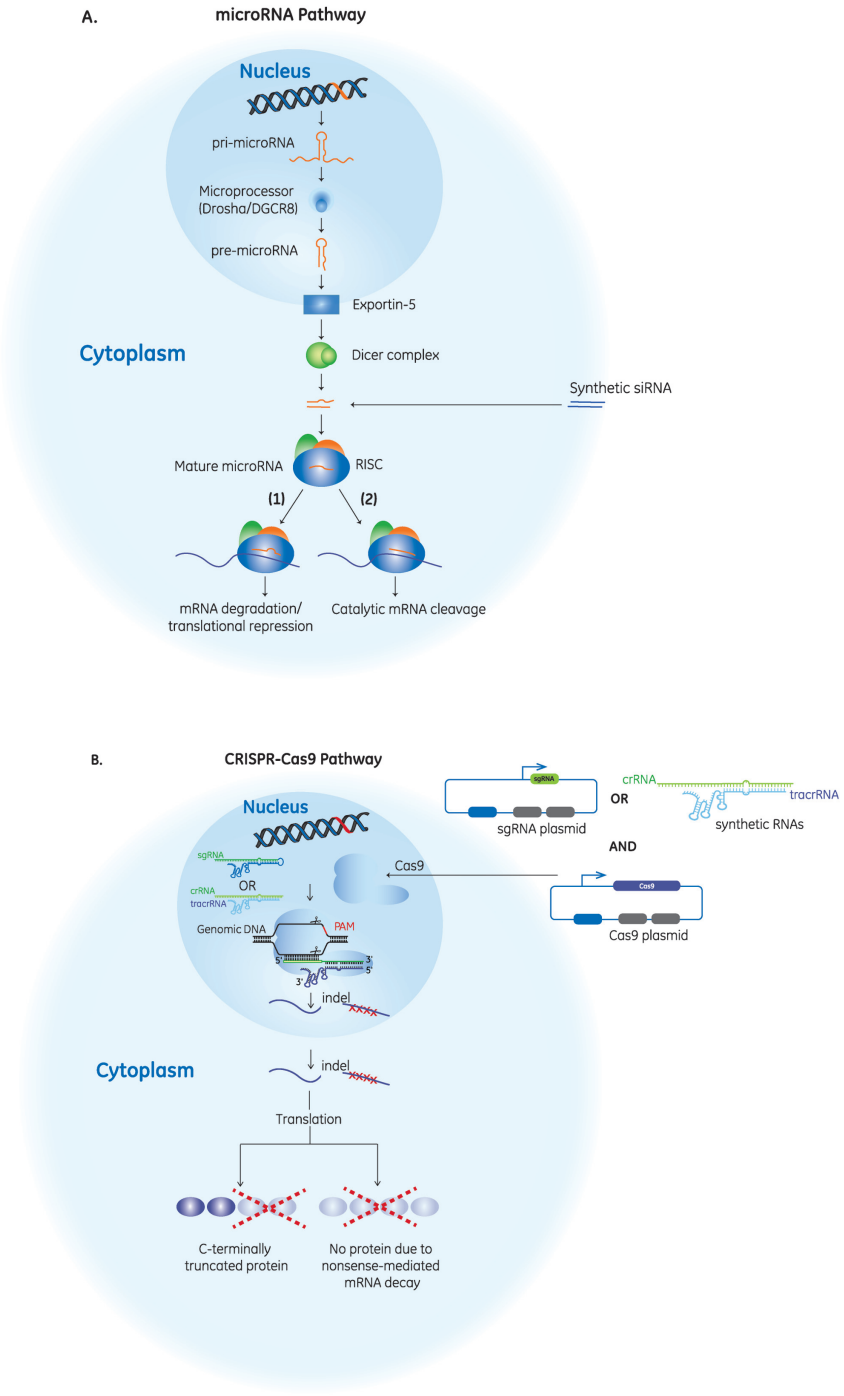

Figure 1. The RNAi and CRISPR-Cas9 pathways in mammalian cells. (**A**) miRNA genes code for primary miRNAs that are processed by the Drosha/DGCR8 complex to generate pre-miRNAs with a hairpin structure. These molecules are exported from the nucleus to the cytoplasm, where they are further processed by Dicer to generate ∼22-nt-long double-stranded mature miRNAs. The RNA duplex associates with an Argonaute (Ago) protein and is then unwound; the strand with a more unstable 5′ end (known as the guide strand) is loaded into Ago to create the RNA-induced silencing complex (RISC) while the unloaded strand is discarded. Depending on the degree of complementarity to their targets, miRNAs cause either transcript cleavage and/or translational repression and mRNA degradation. siRNAs directly mimic mature miRNA duplexes, while shRNAs enter the miRNA pathway at the pre-miRNA hairpin stage and are processed into such duplexes. (**B**) CRISPR-Cas9-mediated genome engineering in mammalian cells requires crRNA, tracrRNA and Cas9. crRNA and tracrRNA can be provided exogenously through a plasmid for expression of a sgRNA, or chemically synthesized crRNA and tracrRNA molecules can be transfected along with a Cas9 expression plasmid. The crRNA and tracrRNA are loaded into Cas9 to form an RNP complex which targets complementary DNA adjacent to the PAM. Using the RuvC and HNH nickases, Cas9 generates a double-stranded break (DSB) that can be either repaired precisely (resulting in no genetic change) or imperfectly repaired to create a mutation (indel) in the targeted gene. There are a myriad of mutations that can be generated; some mutations will have no effect on protein function while others will result in truncations or loss of protein function. Shown are mutations that will induce a frameshift in the coding region of the mRNA (indicated by red X's), resulting in either a truncated, non-functional protein or loss of protein expression due to nonsense-mediated decay of the mRNA.

diated by a ribonucleoprotein (RNP) complex consisting of a CRISPR RNA (crRNA) (8) in combination with a *trans*-activating CRISPR RNA (tracrRNA) (12) and a Cas9 nuclease (13–16) that targets complementary DNA flanked by a protospacer-adjacent motif (PAM) (17–19). The molecular machinery from the CRISPR-Cas9 bacterial immune system can be repurposed for genome editing in mammalian cells by introduction of exogenous crRNAs and tracrRNAs or a single guide RNA chimeric molecule (sgRNA) which combines crRNA and tracrRNA sequences, together with the Cas9 endonuclease to create a double-strand break (DSB) in the targeted DNA (16,20–22) (Figure 1B). The DSB is repaired either by non-homologous end joining (NHEJ) or homology-directed repair (HDR) (23). The error-prone NHEJ pathway typically generates small insertions or deletions (indels) that are unpredictable in nature, but frequently cause impactful and inactivating mutations in the targeted sequence; conversely, the HDR pathway is useful for precise insertion of donor DNA into the targeted site.

Both RNAi and CRISPR-Cas9 have experienced significant milestones in their technological development, as highlighted in Figure 2 (7–14,16–22,24–51) (highlighted topics have been detailed in recent reviews (2,4,52–58)). The CRISPR-Cas9 milestones to date have mimicked a compressed version of those for RNAi, underlining the practical benefit of leveraging similarities to this well-trodden research path. While RNAi has already influenced many advances in the CRISPR-Cas9 field, other applications of CRISPR-Cas9 have not yet been attained but will likely continue to be inspired by the corresponding advances in the RNAi field (Table 1). Of particular interest are the potential parallels in efficiency, specificity, screening and *in vivo*/therapeutic applications, which we discuss further below.

EFFICIENCY

Work performed during the first few years of intensive RNAi investigations demonstrated that, when taking 70–75% reduction in RNA levels as a heuristic threshold for efficiency (59), only a small majority of siRNAs and shRNAs function efficiently (24,60) when guide strand sequences are chosen randomly. This observation led to the development in 2004 of rational design algorithms for siRNA molecules (Figure 2), followed later by similar algorithms for shRNAs. These methods have been able to achieve ∼75% correlation and >80% positive predictive power in identifying functional siRNAs (61) but have been somewhat less effective for shRNAs (62) (perhaps because in most cases, shRNAs produce less knockdown than do siRNAs, likely due to a smaller number of active molecules in each cell). crRNAs also vary widely in efficiency: reports have demonstrated indel (insertion and deletion) creation rates between 5 and 65% (20,25), though the average appears to be between 10 and 40% in un-enriched cell populations. Indeed, a growing amount of evidence suggests a wide range of crRNA efficiency between genes and even between exons of the same gene, yielding some 'super' crRNAs that are more functional (26,27). However, such high-functioning crRNAs are likely to make up only a small percentage of those randomly selected for any given gene (28).

Following the RNAi playbook, efforts to develop rational design algorithms for crRNA have already begun: Doench *et al.* (28) assayed thousands of sgRNAs in a functional assay and identified sequence features that predict sgRNA activity. In addition, the CRISPR-Cas9 field has moved quickly to determine relevant structures and details on the separate binding and cleavage processes (29–30,63), thus avoiding a deficit of mechanistic information that impeded early efforts to predict RNAi efficiency. From these early CRISPR-Cas9 studies, it is already understood that the efficiency of the crRNA:tracrRNA:Cas9 RNP complex is determined by PAM-dependent, crRNA-driven Cas9 binding to target DNA as well as crRNA sequence complementarity to the target DNA (particularly in the first PAM-proximal 8–10 nt, the 'seed' region of the crRNA). Nonetheless, many details of CRISPR-Cas9 activity remain to be learned. Because an efficient sgRNA or crRNA must not only create a DSB in the sequence of interest, but also create a mutation that results in functional disruption of the resulting protein or non-coding structure, the type of indel as well as its position along the length of the coding sequence are likely to be important; thus, effective rational design efforts for functional knockouts may need to incorporate attributes governing these characteristics as well. Furthermore, more than a decade of efforts in the RNAi field demonstrates that even a well-characterized system is not necessarily easy to predict: crRNA design algorithms, like those for siRNAs and shRNAs, may plateau in their predictive power at a level that still necessitates testing of multiple reagents per target in order to guarantee selection of a functional one.

It is worth noting that efficient gene silencing by an RNAi reagent requires a process of ongoing, active repression; this mechanism may be impaired by gene-specific factors that increase mRNA turnover and/or decrease RNA-induced silencing complex (RISC) turnover. As such, there are some genes for which no RNAi reagents can be found that qualify as effective. In contrast, efficiency of a crRNA depends upon the probability of a one-time event—the editing of a DNA site. Because of this difference, the success of rational design efforts may be less critical to the CRISPR field on an individual gene basis, as it will be widely possible to find an effective crRNA if one is willing to evaluate enough treated cells; highly effective crRNA designs, however, will still be necessary for genome-scale studies, as discussed further below.

SPECIFICITY AND OFF-TARGET EFFECTS

Perhaps in no other area are the lessons of RNAi as obvious as in that of specificity. While RNAi was originally hailed as exquisitely specific (64), subsequent research has shown that in some circumstances it can trigger non-specific effects and/or sequence-specific off-target effects (65). Many non-specific effects seen with this approach are mediated by the inadvertent activation of pattern recognition receptors (PRRs) of the innate immune system that have evolved to sense the presence of nucleic acids in certain sub-cellular compartments. siRNA length, certain sequence motifs, the absence of 2-nt 3′ overhangs and cell type are important fac-

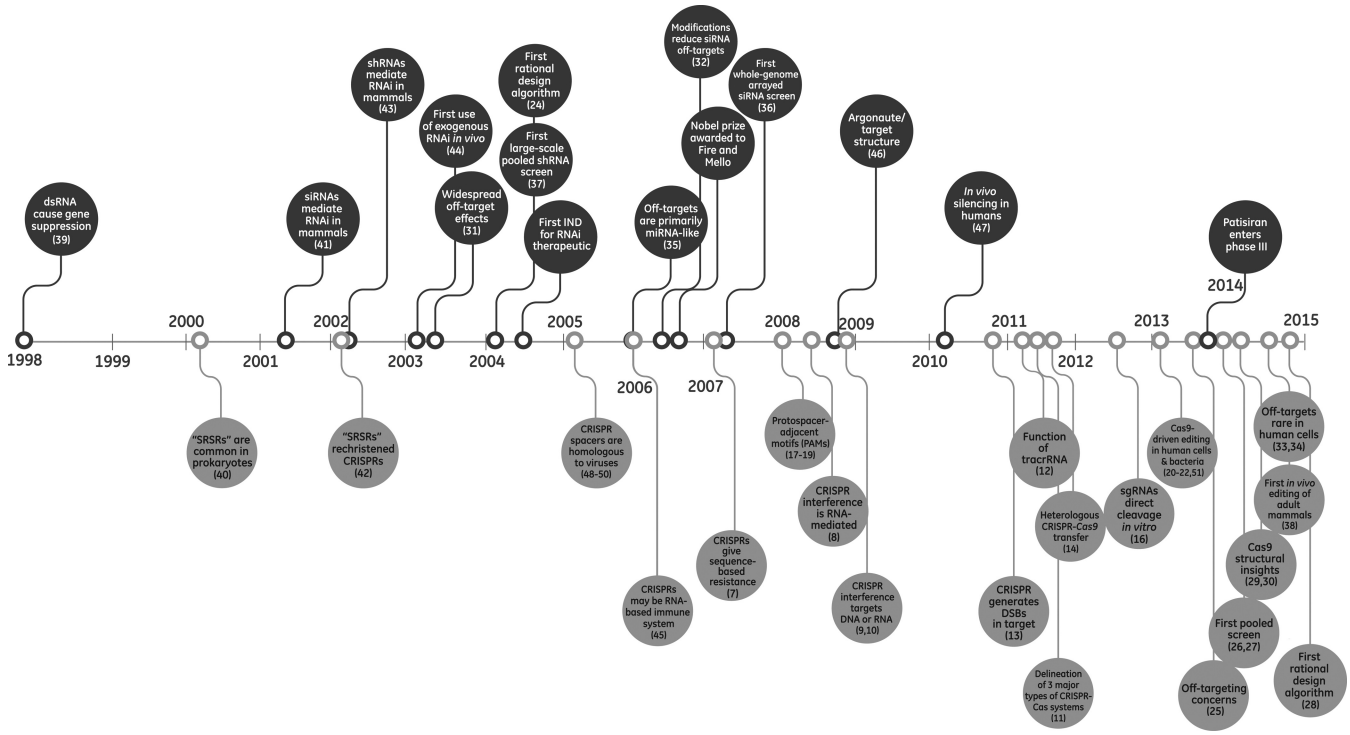

Figure 2. Timeline of milestones for RNAi and CRISPR-Cas9. Milestones in the RNAi field are noted above the line and milestones in the CRISPR-Cas9 field are noted below the line. These milestones have been covered in depth in recent reviews (2,4,52–29).

Table 1. Summary of improvements in the CRISPR-Cas9 field that can be anticipated by corresponding RNAi advances

Milestone	RNAi	CRISPR
IND application	2004	This step is undoubtedly imminent. The drug that was the subject of the first RNAi IND failed clinical trials when its effect was shown to be due to non-RNAi-related mechanisms; especially since CRISPR therapeutics require the delivery not only of a targeting RNA but of exogenous Cas9 (delivered as DNA, mRNA or protein), pharmaceutical developers must avoid allowing history to repeat itself.
Off-target driver identification	2006	Current work is characterizing the nature and extent of the PAM-proximal crRNA 'seed'. Until it is complete, novel outcomes must be demonstrated using multiple reagents to the same target, as is routinely done for RNAi. Once the crRNA seed is understood, researchers should determine whether it could be leveraged to develop sequence-specific off-target controls such as RNAi's C911 controls.
Off-target-reducing modifications	2006	While effective specificity-enhancing chemical modifications for CRISPR may have to wait until off-target drivers are more fully understood, synthetic crRNAs should be modifiable by precisely the same methods as synthetic siRNAs.
Large-scale arrayed screening	2007	Genome-wide arrayed screens using CRISPR are likely to be more challenging because the percentage of edited cells is typically lower than for RNAi. Nonetheless, CRISPR screening and analysis practices will build on and extend those designed for RNAi screening, just as the latter did with those for small-molecule screening.
in vivo use (human)	2010	As CRISPR-driven editing in adult human cells has already been achieved, *in vivo* human use seems inevitable. Efficacious delivery, including that of the exogenous Cas9 protein (or Cas9 mRNA) necessary to make integration-less DNA modifications, is likely to present a significant hurdle. Novel delivery formulations developed in pursuit of RNAi therapeutics will undoubtedly be among those tried first.
Phase III entry	2014	CRISPRa and other dCas9-based approaches raise the hope of addressing conditions untreatable purely via RNAi-like down-regulation while retaining the reversible nature of RNAi. The two modalities might profitably be used in parallel.

tors for induction of the mammalian interferon response (66–68). Additionally, the general perturbation of cellular or tissue homeostasis by the delivery process itself can also trigger unwanted responses (most likely secondary to innate immune damage-sensing pathways) such as the wide-spread alteration of gene expression caused by cationic lipids, especially when used at high concentrations (69). Such non-specific effects associated with delivery will still exist for CRISPR-Cas9 but can likely be overcome by minimizing lipid concentration as is now routinely done in RNAi studies. Similarly, the introduction of chemical modifications into the backbone of an siRNA duplex (e.g. 2′-O-methyl ribosyl) can block the recognition of RNA molecules by PRRs (66,70–71), so such modifications may also address

innate immune system recognition caused by synthetic crRNAs. Researchers would do well to investigate whether additional effects may result, potentially in a cell-line or cell-type dependent manner, as a response to creation of DSBs or the abundant expression of Cas9 or an sgRNA molecule (such as that seen when strong shRNA expression outcompetes that of endogenous miRNAs, leading to the breakdown of cellular regulation (72)). These types of non-specific off-target effects have already been reported with other genome engineering techniques (e.g. zinc finger nucleases (73)) but the ease-of-use and simplicity of the CRISPR-Cas9 system should allow researchers to address these types of questions fully in the near future.

RNAi can also produce sequence-specific off-target effects, which were initially described in early 2003 (31), but whose potential impact was not fully appreciated until well after the method had become a widely used research and screening technique (e.g. (74)). Cleavage-based off-targeting, which occurs when RISC encounters an unintended transcript target with perfect or near-perfect complementarity to its guide strand, can induce knockdown equivalent to that of intended target down-regulation and was originally hypothesized to be the main cause of sequence-specific off-target effects. It took several years to determine that these effects were in fact primarily caused by RNAi reagents acting in a 'miRNA-like' fashion, down-regulating unintended targets by small (usually <2-fold) amounts primarily through seed-based interactions with the 3′ UTR of those unintended targets. Because miRNA-like off-targeting is generally seed-based and all transcripts contain matches to a variety of 6–8-base motifs, such off-targeting can affect tens to hundreds of transcripts. Furthermore, if the RNAi reagent contains a seed mimicking that of an endogenous miRNA, the off-targeting may affect the pathway or family of targets evolutionarily selected for regulation by that miRNA. It is not possible to design RNAi reagents that do not contain seed regions found in the transcriptome's 3′ UTRs and the non-seed factors that conclusively determine whether or not a seed-matched transcript is in fact off-targeted have not yet been identified. Both rational design and chemical modifications such as 2′ O-methyl ribosyl substitutions can mitigate seed-based off-target effects (32), but without a full solution, specificity remains a well-known pain point for RNAi users.

Inspired by these concerns, an initial evaluation of the off-target potential of CRISPR-Cas9 was published within months of the technique's debut and work to refine these early findings has continued apace. Studies have revealed some sequence flexibility, and tolerance for mismatches and bulges, that have generated concerns about specificity and sequence-directed off-target cleavage (25,75–79). Several variations of CRISPR-Cas9 have been developed to address specificity including paired nickases (77), short sgRNAs (76) and Cas9 fused to FokI (80), and the rapid advances in understanding CRISPR-Cas9 mechanism and structure are likely to further fuel such developments. Recent papers, however, have uncovered very few to no off-target mutations that can be attributed to CRISPR-Cas9 (33,34) and conclude that clonal artifacts that derive from isolating CRISPR-Cas9-edited cells may be a larger concern. This apparent discrepancy between prediction and reality, while encouraging, highlights a treacherous pitfall in studying off-target gene editing: to date the primary approach has been to predict putative off-target sites and then search for editing at those sites, but this approach risks falling prey to the 'streetlight effect', in which one searches only where it is easy to look. The RNAi field learned this lesson painfully: early off-target prediction efforts focused on strong overall complementarity as a determinant and thus largely failed to identify genes that were actually off-targeted due to short, seed-based complementary (35). Unless the CRISPR-Cas9 field learns from RNAi's mistakes, it is in danger of repeating the same very one, especially as CRISPR-Cas9 specificity has recently also been shown to depend on an as-yet-not-fully-defined seed region (20,81). At this stage, computational predictions of putative off-target gene-editing sites are at best questionable guesses and thus cannot be depended upon in assessing the full effect of off-targeting. Unfortunately, while RNAi specificity studies are limited to the transcriptome, analysis of CRISPR-Cas9 specificity requires identifying effects that may occur anywhere in the genome, generally through next-generation sequencing. This process is costly and depends upon non-trivial data analysis and processing, so it remains unclear whether the field has the will to commit to this work. Until such time as a less biased understanding of CRISPR-Cas9 off-targeting emerges, researchers are advised to emulate the best-practice of RNAi by using multiple crRNAs or sgRNAs in order to show redundancy of phenotype by multiple reagents targeting the same gene, thus ensuring that the phenotype is due to on-target effects rather than off-target effects.

GENOME-SCALE SCREENING TOOLS

Interest in genome-scale CRISPR-Cas9-based screening has blossomed, with some pooled screening resources already available (26,27) and arrayed ones likely to emerge in the near future. Because genome editing screens will also be affected by a large number of the factors that make RNAi screens more challenging than small-molecule ones (82), practitioners would do well to study the hard-won victories in this field since the first published whole-genome synthetic lethal screen for sensitization to paclitaxel (36) (Figure 2) before diving into these costly experiments.

Of particular importance is evaluating whether the lower efficiencies seen using CRISPR-Cas9 are sufficient to generate a desired phenotype in the screening assay—that is, determining whether the phenotype is detectable in the targeted cell population. In this regard, two factors are of special concern: the ploidy of the gene locus of interest (as tumor cell lines are often aneuploid) and the likelihood of disrupting the reading frame by the induced mutation (since +3 or −3 indels would not serve this purpose). Taking these factors into account, the chance of obtaining a high percentage of cells that have a functional knockout in a bulk cell culture is relatively low under typical screening conditions. Consequently, it is unlikely that traditional arrayed loss-of-signal screens such as those common in RNAi will be widely feasible in bulk-transfected cells using CRISPR-Cas9. Nevertheless, the CRISPR-Cas9 technology may have an advantage in screens for which a complete

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.