Regulatory RNA-assisted genome engineering in microorganisms
Tong Si\textsuperscript{1}, Mohammad HamediRad\textsuperscript{1} and Huimin Zhao\textsuperscript{1,2,3,4}

Regulatory RNAs are increasingly recognized and utilized as key modulators of gene expression in diverse organisms. Thanks to their modular and programmable nature, \textit{trans-}acting regulatory RNAs are especially attractive in genome-scale applications. Here we discuss the recent examples in microbial genome engineering implementing various \textit{trans-}acting RNA platforms, including sRNA, RNAi, asRNA and CRISPR-Cas. In particular, we focus on how the scalable and multiplex nature of \textit{trans-}acting RNAs has been used to tackle the challenges in creating genome-wide and combinatorial diversity for functional genomics and metabolic engineering applications. Advances in computational design and context-dependent regulation are also discussed for their contribution in improving fine-tuning capabilities of \textit{trans-}acting RNAs.

Addresses
\textsuperscript{1} Department of Chemical and Biomolecular Engineering, Carl R Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, United States
\textsuperscript{2} Department of Chemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, United States
\textsuperscript{3} Department of Biochemistry, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, United States
\textsuperscript{4} Department of Bioengineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, United States

Corresponding author: Zhao, Huimin (zhao5@illinois.edu)

Introduction
Engineered microorganisms are emerging as solutions to many global challenges that we are facing today by providing environmental remediation, medical applications, and renewable chemical production [1–3]. To achieve these engineering objectives, intensive reprogramming of cellular networks is needed, especially upon the introduction of foreign functionalities such as heterologous pathways and synthetic gene circuits. Given the complexity of biological systems, cellular reprogramming requires both identification of genetic targets on a genome scale and optimization of their expression in a concerted manner (multiplex editing). Such practices underlie the concept of ‘genome engineering’, in which large-scale and multiplex genetic modifications are created, by contrast with the more traditional term ‘genetic engineering’, in which only a small number of genes are modulated [4–8].

Within this context, regulatory RNAs are increasingly implemented for comprehensive understanding and engineering of genomes [9,10,11]. Interacting with DNA, RNA, protein and metabolic molecules, RNAs can regulate every step of gene expression [11]. Specifically, \textit{trans-}acting RNA-mediated gene regulation is scalable and predictable, as the recognition of target DNA or RNA sequences is primarily governed by the Watson–Crick base-pairing principle [6,10**]. Nowadays, genome-wide libraries of RNA molecules or RNA expression cassettes can be readily designed and manufactured, facilitating screening and identification of relevant genes to a given phenotype. Moreover, it is possible to introduce several regulatory RNAs simultaneously in one cell and hence enable multiplex targeting. This capability is crucial in understanding and engineering non-linear interactions within complex genetic networks, as exemplified as combinatorial optimization of multiple gene targets [12,13].

In this review, we first survey examples of various \textit{trans-}acting RNA regulatory mechanisms in microbial genome engineering, including small RNA (sRNA), antisense RNA (asRNA), RNA interference (RNAi) and the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associated proteins (Cas) system. Both genome-scale screening and multiplex targeting applications will be emphasized. We then discuss some notable technological advances in implementing RNA-based genome engineering, including computational design and context-dependent regulation.

\textbf{Trans-acting regulatory RNA platforms sRNA, RNAi and asRNA}
Various \textit{trans-acting} regulatory RNAs employ different mechanisms (Table 1). In terms of mode-of-action and target sequences, sRNA, RNAi and asRNA modulate gene expression post-transcriptionally by targeting mRNA molecules, whereas CRISPR exerts transcriptional modulation via interactions with DNA.
In *Escherichia coli*, Hfq-associated synthetic sRNAs have been implemented for metabolic engineering ([Figure 1](#fig1)) [14**]. Hfq is a common RNA-binding protein to facilitate sRNA–mRNA complex formation and mRNA degradation. A modular synthetic sRNA template was designed based on MicC, which is a naturally occurring sRNA in *E. coli* [14**]. This sRNA template consists of a common scaffold sequence to recruit the Hfq protein, and a custom target-binding sequence that is complimentary to the translation initiation region (TIR) of a given gene to attenuate translational initiation ([Figure 1](#fig1)). The versatility of this approach was demonstrated in 14 different *E. coli* strains, through combinatorial repression of four gene targets, screening of a medium-sized sRNA library, and fine-tuning of knockdown levels [14**]. The ability to regulate gene expression without modification of genome sequences greatly accelerate the prototyping of production hosts.

RNAi is a common gene silencing mechanism in eukaryotes ([Figure 2a](#fig2)) [9]. However, the model yeast, *Saccharomyces cerevisiae*, lacks a native RNAi pathway [13]. The recent reconstitution of RNAi machinery in *S. cerevisiae* enables the use of this powerful tool for yeast genome engineering [15]. Enzymatically generated genomic DNA fragments were cloned into a pair of convergent constitutive promoters to construct double-stranded RNA libraries for genome-wide silencing ([Figure 2b](#fig2)) [16**]. Screening the resultant library led to identification of suppressors of a telomere-defect mutation and genetic determinants for improved tolerance towards lignocellulosic hydrolysate inhibitors [16**,17]. Furthermore, RNAi-assisted genome evolution (RAGE) enables continuous improvement of a target phenotype, by applying a rationale similar to directed evolution for protein engineering ([Figure 2b](#fig2)). During one round of RAGE, several single knockdown modifications were first identified by RNAi screening, and then integrated in the yeast genome to create new parent strains for the next round of engineering ([Figure 2b](#fig2)). In this way, multiplex knockdown mutations were accumulated to improve complex traits in *S. cerevisiae* [16**]. The ease of RNAi library construction allows iterative genome-wide screening which is otherwise prohibitively tedious and time-consuming using traditional gene knockout approaches, and therefore lead to discovery of synergistic genetic mutations that are not accessible previously.

Compared with sRNA and RNAi, the use of asRNAs has been primarily limited to genome-wide knockdown screening. For example, a paired-termini expression vector was used to express stable asRNAs derived from an *E. coli* genomic library for a loss-of-function screen [18]. Although also applied in a collection of gram-positive bacteria (as reviewed in [6]) and a yeast pathogen [19], the lack of a clear understanding of gene silencing mechanism by asRNAs hinders the widespread application of asRNA screening.

**CRISPR**

Whereas most *trans*-acting RNAs target mRNA molecules, the CRISPR–Cas system provides a RNA-mediated interference system that cleaves DNA ([Figure 3a](#fig3a)) [20,21**]. Specifically, the type II *Streptococcus pyogenes* CRISPR–Cas system offers a simple configuration that contains only a single endonuclease (Cas9) [21**]. *Trans*-activating CRISPR RNA (tracrRNA) facilitates the processing of CRISPR RNA (crRNA) by Cas9, and then the Cas9–crRNA–tracrRNA complex introduces a double-stranded break (DSB) at the targeted genome locus.
Mechanism and application of RNAi in S. cerevisiae. (a) RNA-induced silencing complex (RISC) seeks and degrades mRNA targets under the guide of small interfering RNAs (siRNAs), and therefore leads to gene silencing. (b) In RNAi-assisted genome engineering (RAGE), yeast genomic DNA fragments are inserted between a pair of convergent promoters to synthesize long double-stranded RNAs for RISC activation. Then, following genome-wide RNAi screening, the identified beneficial cassettes conferring improved phenotypes are integrated to create new parent strains for the next round of RAGE. Iterative rounds of RAGE result in accumulation of multiplex knockdown targets for directed genome evolution [14**].

(Figure 3a). The tracrRNA and crRNA molecules can also be engineered into a single guide RNA (sgRNA) [21*]. The specificity of target recognition is determined by the guide sequence in the crRNA followed by a short sequence called protoscaler adjacent motif (PAM) [21*].

The DSBs introduced by CRISPR–Cas can substantially improve genome editing efficiency by stimulating the process of non-homologous end joining (NHEJ) or homologous recombination (HR) [22]. This versatile framework has already been applied in a wide range of organisms including bacteria [23], yeasts [24], plants [25], animals [26,27], and human cells [22,28].

In microbial genome engineering, multiplex genome editing is desirable to accelerate strain prototyping and combinatorial optimization. Traditional approaches rely on sequential incorporation of multiple modifications, which can be very tedious and time-consuming. On the other hand, improved editing efficiency and multiplex nature of the CRISPR–Cas system may help to tackle this challenge. In S. cerevisiae, one-step multiple gene disruption was achieved using the homology-integrated CRISPR–Cas (HI-CRISPR) system (Figure 3b) [29*]. Homologous knockout donors for premature stop codon introduction were inserted before the 5' end of the guide sequences in crRNAs, and such configuration permitted integration of all the necessary CRISPR elements in a single ultrahigh copy plasmid (Figure 3b). Three genes were simultaneously disrupted with an efficiency ranging from 27% to 100% after a single transformation in 4–6 days [29*], whereas traditional methods require six rounds of transformation and may take approximately six weeks. In another study, a collection of 31 S. cerevisiae strains covering all possible knockout combinations of five genes in an isopenoiid biosynthesis pathway were rapidly constructed via CRISPR [30]. The CRISPR–Cas system was also employed to disrupt multiple alleles of the same gene in polyploid industrial S. cerevisiae strains [31]. Besides gene disruption, multiplex integration was attempted. In S. cerevisiae, six DNA fragments bearing 11 genes for muconic acid biosynthesis were inserted at three genomic loci where DSBs were generated by CRISPR [32].
Although integration efficiency requires further improvement (currently 4.2%) to reduce the effort in identifying correct constructs, substantial decrease in the time commitment (6-fold in this study) to prototype large biochemical pathways can greatly accelerate pathway engineering [32]. Likewise, multiplex mutagenesis by CRISPR has also been demonstrated in bacteria, including *E. coli*, *Tatumella citrea*, *Streptococcus pneumonia* and *Streptomyces* [23,33,34]. In particular, the dual cleavage system was recently implemented in chromosomal deletions in *Streptomyces* species, where genomic DNA portions of various sizes (from 20 bp to 30 kb) were removed following DSBs generation in two genomic loci [34]. This new capability of rapid, efficient and precise chromosomal editing may greatly facilitate natural product research in *actinobacteria*.

The CRISPR–Cas system has also been expanded for gene regulation without altering the target DNA sequences. To achieve this goal, a deactivated Cas9 (dCas9) protein with abolished endonuclease activity was created (Figure 3a) [35**]. It was found that efficient gene silencing can be achieved when the non-template DNA strand was targeted by dCas9/sgRNA complex, which blocked transcriptional initiation or elongation [35**]. This so-called CRISPR interference (CRISPRi) has been successfully demonstrated in *E. coli*, *actinomyces*, yeast and mammalian cells [35**,36,37]. Also, fine-tuning of gene repression can be realized via introduction of mismatches between crRNA and target sequences, as well as by variation of target sequences [38]. In addition, when fused with transcription factors, dCas9 can accomplish more modes of gene regulation (Figure 3a). For example, gene activation can be achieved in *E. coli* and yeast, when dCas9 is fused to the omega subunit of bacterial RNA polymerase and yeast transcriptional activator domains, respectively [36–39]. Likewise, gene silencing efficiency can be improved by fusing dCas9 to transcription repressors or chromatin silencers [36]. Together, the dCas9-based gene regulation platform has demonstrated great versatility in terms of host selection and fine-tuning capability.

### Technology advances in RNA-assisted genome engineering

#### Computational design

For the computational design of *trans*-acting regulatory RNAs, one aim is to predict gene modulation efficiency. For example, a quantitative relationship between the repression capability and the binding energy was established for sRNA regulators in *E. coli*, allowing rational design for tunable gene regulation [14**]. Another aim is to design synthetic RNA reagents with optimized targeting efficacy and minimized off-target effect to create genome-wide libraries. For RNAi, in addition to direct construction using genomic libraries [16**], it is also possible to computationally design synthetic pools of RNAi sequences, and the readers are directed to a more comprehensive review for available tools [40]. For CRISPR, there are a few web servers that can design target sequences for microbes including *E. coli*, *S. cerevisiae* and *Aspergillus* [41–43]. In particular, E-CRISPR ranks targeting sequences of the same gene according to specificity, annotation and efficacy scores, and is able to create genome-scale libraries in a few hours thanks to its fast algorithms (http://www.e-crisp.org/E-CRISPR/) [41]. CRISPRdirect offers application program interface (API) to allow generation of many targeting sequences in an automated manner, and therefore enables genome-scale design (http://crispr.dbcls.jp/) [43]. Although genome-wide screening using CRISPR has not been demonstrated in microbial hosts, in the future, these computational tools are indispensable to design target sequence libraries on a genome scale.

#### Context-dependent regulation

One limitation of *trans*-acting regulatory RNAs is that they lack the innate ability to sense cellular signals. Such capability may be incorporated in three ways. First, inducible promoters can be used to control the synthesis of regulatory RNAs and their accessory proteins. For example, two components of CRISPRi, dCas9 and guide RNA, were subjected to the control of an inducible promoter, and this configuration linked gene repression to the presence of the inducer [35**]. In another case, the synthesis of a hairpin RNAi construct was driven by quorum-sensing promoters in *S. cerevisiae*. Dynamic gene knockdown was achieved where the target genes were only silenced when cell culture reached a high population density [44]. Second, allosteric modulation can be implemented by creating a hybrid RNA molecule that consists of both a *cis*-ligand-sensing motif and a *trans*-target-recognition motif. Ligand binding to the *cis*-motif led to structural rearrangement that allowed the *trans*-recognition motif to form a functional conformation [45**]. Third, signal sensing capability can be programmed into the accessory proteins of regulatory RNAs. In a light-activated CRISPR–Cas9 effector (LACE) system, dCas9 and a transcriptional activator domain were fused with cryptochrome 2 (*CRY2*) and cryptochrome-interacting basic-helix–loop-helix (CIB1) proteins from *Arabidopsis thaliana*, respectively. Upon blue light irradiation, the heterodimerization of *CRY2* and CIB1 recruited the activator domain to dCas9 that binds to a target locus, and therefore achieved spatiotemporal gene activation [46,47]. Furthermore, the development of split Cas9 also permits regulated activation of CRISPR dependent on the inducible assembly of a functional Cas9 [48].

#### Conclusions

Regulatory RNA-assisted genome engineering has become an emerging platform for large-scale and multiplex modifications of microbial genomes. However, there are certain limitations requiring further development. For example, most RNA regulation mechanisms repress gene expression. Though RNA-mediated gene overexpression...
methods are being developed [36–38,49], general principles are still missing to design activating RNAs on a genome scale. Also, off-target effect is a legitimate concern to most RNA platforms [9,10**,19], and therefore precautions have to be taken when interpreting genome-wide screening results.

In the future, the discovery and adaptation of more RNA regulatory mechanisms will provide a rich source to develop new engineering platforms. In addition, although genome-wide screening using CRISPR in microorganisms has not been reported, the practice of CRISPR-mediated functional screening in human cells clearly shows the potential [50]. Moreover, the application of RNA-mediated genome engineering in less-studied microbial hosts will facilitate the development of new chassis. Together, we foresee an accelerating trend in implementing regulatory RNAs in the understanding and engineering of microbial genomes.

Acknowledgements

We gratefully acknowledge financial support from Defense Advanced Research Program Agency (HR0011-14-C-0069), Roy J. Carver Charitable Trust (13-4257), and Carl R. Woese Institute for Genomic Biology at the University of Illinois at Urbana-Champaign for financial support in our development and application of genome-scale engineering tools.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as: • of special interest ** of outstanding interest


A comprehensive review on the design and application of non-coding RNA devices in microbial engineering.


The first example of rationally designed sRNA platform in E. coli for metabolic engineering applications.


This study presents the first report of RNAI screening in S. cerevisiae. In addition, iterative rounds of RNAI screening enabled the use of directed evolution strategy for genome engineering in yeast.


An updated, comprehensive review on the CRISPR-Cas9 system.


This study implemented an intelligent design where homologous donors were inserted before the 5' end of guide sequences in the crRNA array, allowing one-step delivery of all necessary CRISPR elements on a single plasmid into yeast cells for multi-gene knockout.


The inception of CRISPRi. Reversible gene regulation is enabled using a catalytically inactive Cas9 to interfere with transcriptional processes.


This work created RNA fusions combining trans-acting and cis-acting modules, which permit the integration of sensing and targeting capabilities of RNA devices.


