A modular RNA interference system for multiplexed gene regulation

Ari Dwijayanti[1,2], Marko Storch[1,2,3], Guy-Bart Stan[2,4]*, Geoff S. Baldwin[1,2]*

[1] Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

[2] Imperial College Centre for Synthetic Biology, Imperial College London, London SW7 2AZ, UK

[3] London Biofoundry, Imperial College Translation & Innovation Hub, London W12 0BZ, UK

[4] Department of Bioengineering, Imperial College London, London SW7 2AZ, United Kingdom

* To whom correspondence should be addressed. Email: g.baldwin@imperial.ac.uk; g.stan@imperial.ac.uk

Keywords: sRNA; post-transcriptional regulation; modular design; multiplexed control; synthetic biology

ABSTRACT

The rational design and realization of simple-to-use genetic control elements that are modular, orthogonal and robust is essential to the construction of predictable and reliable biological systems of increasing complexity. To this effect, we introduce **m**odular **A**rtificial **R**NA **i**nterference (mARi), a rational, modular and extensible design framework that enables robust, portable and multiplexed post-transcriptional regulation of gene expression in *Escherichia coli.* The regulatory function of mARi was characterized in a range of relevant genetic and cellular contexts and was shown to be independent of other genetic control elements and the gene of interest, as well as growth-phase and strain type. Importantly, the extensibility and orthogonality of mARi enables the simultaneous post-transcriptional regulation of multi-gene systems as both single-gene cassettes and poly-cistronic operons. To facilitate adoption, mARi was designed to be directly integrated into the modular BASIC DNA assembly framework. We anticipate that mARi-based genetic control within an extensible DNA assembly framework will facilitate metabolic engineering, layered genetic control, and advanced genetic circuit applications.

INTRODUCTION

Synthetic biology aims to make the engineering of biological systems more predictable, efficient and reliable (1–3). With this goal in mind, our ability to predictably combine genetic parts into higher-level functional biological/biochemical systems of increasing complexity is essential to improve the efficiency of the Design-Build-Test-Learn cycle (4, 5). In this work, we focus on expanding the toolbox of modular biomolecular control elements that function at the post-transcriptional level, specifically small non-coding RNAs (sRNAs). The ubiquity of sRNA-based control in critical cellular processes such as homeostasis and adaptation demonstrates their importance for the design and control of biological systems (6–10). The sRNAs act to coordinate and synchronize multiple signals through sequence-specific and transient RNA-RNA interactions (11, 12), typically leading to down-regulation of target gene expression (13–16).

In bacterial systems, *trans*-acting sRNAs rather than *cis*-acting sRNAs have been reported to be especially effective in silencing gene expression due to their longer half-life (17–19). The programmability of *trans*-acting sRNAs makes them attractive candidates for the design of synthetic biological circuits for metabolic engineering and synthetic biology purposes (7, 20–23). Despite the fundamentally attractive features of sRNA-based regulatory systems, current approaches for implementing specific targeted artificial sRNAs are hampered by the need for bespoke design (21, 23, 24), leader sequences (25), or target insertions (20, 22).

Here, we set out to design a universal, modular framework for the facile implementation of *trans*-acting sRNA-based control using verified, re-usable target sequences, namely **m**odular **A**rtificial **R**NA **i**nterference (mARi). BASIC DNA assembly provides a modular, standardized and automatable (26) framework for genetic structure based on orthogonal, computationally designed linkers (27). We exploited this by targeting the DNA linkers used in the construction of gene expression cassettes and operons. In our mARi design, a modified seed sequence specific to the linker upstream of the target Gene of Interest (GOI) was fused to a native sRNA scaffold containing a host factor-1 (Hfq) binding site (23, 24) (**Fig. 1a, 2a**). The integration of mARi-based regulation into a modular DNA assembly method offers a simple yet powerful strategy for implementing RNAi regulatory systems. Since standardized linkers are used in the assembly process, gene expression can be controlled by expression of an mARi that is cognate to the linker sequence upstream of the target gene (**Fig. 1**). We characterize the post-transcriptional regulation of mARi in the context of various genetic design parameters including transcript ratio, molecular copy number, spatial organisation, host strain, and growth phase. Finally, mARi sequence variants were further expanded and their implementation was demonstrated for simultaneous regulation in a multi-gene system with different genetic architectures.

MATERIALS AND METHODS

Bacterial strains and outgrowth conditions

The following *E. coli* strains were used for plasmid construction and expression: DH5α (F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* ϕ80d*lacZ*ΔM15 Δ (*lacZYA-argF*)U169, hsdR17 ($r_K^- m_K^+$), λ^-), DH10b [F-*mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ (*ara leu*) 7697 *gal*U *gal*K *rps*L *nup*G λ-], BL21 (DE3) (F^-*omp*T *hsd*S_B (r_B^-, m_B^-) *gal dcm* (DE3)), and BL21star (DE3) (F^-*omp*T *hsd*S_B (r_B^-, m_B^-) *gal dcm rne*131 (DE3)). *E. coli* colonies were grown from single-colony isolates in LB (Luria Bertani) medium supplemented with the appropriate antibiotic, shaken at 220 rpm and grown at 37°C (unless specified otherwise). The antibiotics used for plasmid maintenance were Carbenicillin and Kanamycin at a concentration of 100 µg/ml and 50 µg/ml, respectively.

Design and analysis of mARi sequences

The sequences of UTR-RBS linkers used in BASIC DNA assembly (27) were designed and validated using R2o DNA designer (28). Secondary structures of the mARis were predicted using RNAFold WebServer (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and visualized using VARNA GUI (29) (http://varna.lri.fr/index.php?lang=en&page=tutorial&css=varna). The predicted interactions between mARis and their mRNA targets were simulated through IntaRNA (30, 31) (http://rna.informatik.unifreiburg.de/IntaRNA/Input.jsp). The Free binding energy of mARi and its mRNA target site was estimated using a web-based service of the RNA folding software DINAMelt (32) (http://mfold.rna.albany.edu/?q=DINAMelt/Two-state-melting) using default parameters. The percent identity of mARi-UTR pairs was calculated using the EMBOSS needle method (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). The predicted expression strength of the UTR-RBS linker was simulated using EMOPEC (33) and RBS calculator v2.0 (34, 35) with the input sequence starting from the Transcription Start Site (+1) to 100 bp of the downstream CDS. The off-target effect was computed using CopraRNA (31, 36, 37) (http://rna.informatik.uni-freiburg.de/CopraRNA/Input.jsp). NC_000913 (*E. coli* MG1655), NC_010473 (*E. coli* DH10b),

NC_012892 and NC_012971 (*E. coli* BL21 (DE3)) were used as inputs references for host-background off-target prediction.

Plasmid assembly

Plasmids were constructed using the BASIC DNA assembly method (27, 38). New DNA parts with prefix and suffix sequences were synthesized as gBlocks by Integrated DNA Technologies (IDT), or gene fragments by TWIST Bioscience, or derived by PCR mutagenesis. The gBlocks and gene fragments were directly cloned into pJET1.2 or pUC AmpR_plasmid as BASIC bioparts. Linkers for BASIC assembly were provided by Biolegio or IDT **(Supplementary Table 6)**. All constructs were transformed into chemically competent *E. coli* DH5α. Sequence verification of BASIC bioparts was carried out using Sanger sequencing (Source Bioscience). PCR mutagenesis for generating new BASIC parts were carried out using Phusion Polymerase (NEB) and phosphorylated oligonucleotides (IDT). Details of the BASIC bioparts used to construct the plasmids are provided in **Supplementary Table 5**. Maps of plasmids used in this study are provided in **Supplementary Fig. 7**.

Flow cytometry assay

Cellular expression of sfGFP and mCherry was performed in 96 well plates with biological triplicates for each construct. *E. coli* carrying an empty backbone (containing only an origin of replication and an antibiotic resistance gene) was used as a negative control. *E. coli* colonies were inoculated from glycerol stock into LB media supplemented with appropriate antibiotics. The colonies were grown for 16 h at 30°C and 600 rpm in a benchtop shaker (Mikura). The culture was then diluted 200x (10 µl into 190 µl, then 10 µl into 90 µl) in LB media using an automated liquid handling robot (CyBio Felix). The plate was grown at 37°C and 600 rpm in a benchtop shaker (Mikura). After 6 h of incubation, the Abs600 was measured. Abs600 typically reached values in the range 0.6–0.7. Two microliters of liquid culture were taken and diluted into 200 µl Phosphate Buffer Saline (PBS) supplemented with 2 mg/ml Kanamycin. The fluorescence data of single cells for the dual reporters system (Fig 4b & d) was collected using BD Fortessa flow cytometer with a 488 nm excitation laser and filter 530/30 for sfGFP; 561 nm excitation laser and filter 610/20 for mCherry. For the triple reporters

system (Fig 4c), fluorescence was measured using Attune NxT flow cytometer with a 488 nm excitation laser and channel BL1 530/30 for sfGFP; 561 nm excitation laser and channel YL2 620/15 for mCherry; 405 nm excitation laser and channel VL1 440/50 for mTagBFP. In total 10,000 events were collected for each sample. The data was stored as an FCS file 3.0 and analysis was done using FlowJo V10. Single cell population gating was performed after plotting FSC-H against SSC-H and histograms of each channel of fluorescence. The chosen gating covered 90-98% of the total population. The fluorescence intensity of the sample was calculated by subtracting the geometric mean fluorescence of the control (strains with empty backbone). The normalised fluorescence was calculated by dividing the fluorescence intensity of the sample with mARi by the fluorescence of the sample without mARi expression.

Plate reader assay

For continuous growth assay, samples were first diluted 200x in LB media (10 µl into 190 µl, then 10 µl into 90 µl) using an automated liquid handling robot (CyBio Felix); the plate was then incubated in a microplate reader (Clariostar, BMG Labtech) with continuous shaking at 37°C and 600 rpm for 12-24 h. Absorbance at 600 nm and GFP fluorescence (F:482-16/F:530-40) were measured every 15 min. The normalised Abs600 was calculated by subtracting the mean Abs600 of media only control from the Abs600 of each well. The fluorescence intensity was corrected by subtracting the mean fluorescence value of negative control at equivalent Abs600. The value Fl/Abs600 was calculated by dividing the value of corrected fluorescence by corrected Abs600. Statistical analysis was calculated in Prism v.8.0 (GraphPad). The values were compared using a Student's t-test for unpaired comparisons and one-way ANOVA.

RESULTS

Design of a modular Artificial RNA interference regulation system

The mARi sequences were designed with two core components: the seed sequence and an sRNA scaffold containing a Hfq binding site (**Fig. 1a, 2a**). The scaffold sequence is responsible for recognition by RNA chaperones, such as the Hfq protein, which are highly conserved as part of natural cellular regulation (39, 40) in a wide range of organisms (41–44). The inclusion of this RNA chaperone binding site in the sRNA structure has previously been shown to increase the silencing performance of sRNA-based gene repression (45). One of the intensively-studied Hfq-dependent sRNAs, MicC was chosen from sRNA scaffolds naturally found in *E. coli* (46) and *Salmonella enterica* (16). The nature of the multiple targets of MicC based sRNAs in controlling the expression of outer membrane proteins (OmpC and OmpD) can be exploited to target any mRNA sequence of interest. Based on this architecture, the seed sequences can be modified while preserving the native MicC scaffold (21, 23, 24) (**Fig. 2a**).

The seed sequence within mARi creates a specific base-pairing interaction to a target sequence in the mRNA of interest, resulting in its post-transcriptional silencing. To standardize and enhance the modularity of mARi, the seed sequence was designed to target expression cassettes constructed via BASIC DNA assembly (27). BASIC assembly relies on standardised linkers between interchangeable DNA parts (**Fig. 1a**) (27). Expression cassettes and operons can be tuned using functional DNA linkers that encode a ribosome binding site (RBS) within a defined 5' untranslated region (UTR; UTR-RBS linker; **Fig. 1b**). Since BASIC UTRs are computationally designed to be orthogonal to the host genome and exclude secondary structure (28), they provide an ideal target for mARi and reduce off-target effects.

We designed mARi sequences to target the translation initiation region of mRNA for three main reasons: (i) the 5’ UTR sequences are relatively AU rich (31-43% GC; **Supplementary Table 4**) and AU rich sequences are preferentially bound by Hfq(47); (ii) targets located in the translation initiation region have been shown to exhibit a low off-target effect in *trans*-encoded sRNA-based regulation

(45), which is important for regulator specificity/orthogonality; (iii) a 5' UTR target site location offers potential modularity since it is independent of different GOIs and is compatible with the modular UTR-RBS linkers in the BASIC design framework (27). To evaluate this approach, seed sequences were designed to target different positions in the translation initiation region (**Fig. 2b**). Positions 1-3 addressed gene-independent targets within the BASIC framework, while position 4 was specific to the gene of interest, in this case *superfolder-gfp* (*sfgfp*; **Fig. 2b**). The reverse complements of the cognate target sequences were combined with the MicC scaffold to create a series of mARi sequences (**Supplementary Table 1**).

The design of the full sequence of each mARi variant was computationally evaluated to ensure that it met the requirements for effective repression. Firstly, the designed sequence for each mARi was made compatible with BASIC DNA assembly (27), specifically avoiding the forbidden restriction site *Bsa*I. Secondly, to achieve effective repression and to have a high affinity binding between the seed and target sequences, the GC content and binding energy of the seed sequences were calculated and found to vary from 34.62% to 43.8% and -35.7 kcal/mol to -55.2 kcal/mol, respectively. The binding energy has a positive correlation to the binding affinity and observed repression capability (23, 45), and previous studies suggested a typical binding energy of -30 to -40 kcal/mol (23, 24) was required for effective repression activity. However, lower binding energy (less than -40 kcal/mol) sRNAs have been experimentally validated and suggested to achieve higher repression (45). Homology length determines mARi specificity and target discrimination (45): shorter binding sites have a lower binding affinity and may cause off-target effects (23, 24), while longer binding sequences may have a complex secondary structure that could cause binding to the mRNA target to be less thermodynamically favorable (45). All mARis used here were designed to have perfect complementarity in their seed sequences to their target mRNA. (**Supplementary Fig. 1b**).

To investigate the repression efficiency of mARi targets, sfGFP expression cassettes were assembled with a constitutive promoter P_{J23101_BASIC} and UTR-A-RBSc linker, while mARi expression was driven by the constitutive P_{J23101} (**Fig. 1b**). Measurement of reporter expression demonstrated that all

mARi target sequences led to repression of sfGFP, compared to control cells lacking mARi (**Fig. 2c**). The intra-gene target at position 4 showed significant repression as anticipated (21, 23, 24). However, targeting this position would require bespoke design and optimisation for each new expression cassette; a less desirable strategy from a modular engineering perspective. Position 1, which targets the UTR upstream of the RBS was as effective as position 4 (**Fig. 2c**). In fact, targeting this standardised UTR component of the BASIC framework is ideal for modular engineering since it is independent of both the GOI and RBS.

To evaluate the independence of position 1 UTR target, a set of constructs with five different RBS strengths (RBSa-RBSe) (**Supplementary Table 2**) controlling sfGFP translation were assembled both with and without the position 1 mARi cassette all within the same UTR-A context. The *sfgfp* gene-only constructs resulted in a series of fluorescence intensities reflecting RBS strength variation (**Fig. 2d**). The position 1 UTR-A targeted mARi cassette reduced the fluorescence signal in all cases and the relative repression was found to be constant across the series (**Fig. 2d**). This demonstrated that the repression activity of mARi-A at position 1 was independent of RBS strength, an important feature for RBS design flexibility and modularity of mARi.

mARi activity is tuneable via genetic design parameters

To evaluate the features of mARi, we characterised its responsiveness to genetic design parameters and its robustness to different contexts and conditions. Two key determinants modulating the repression activity of this post-transcriptional regulator are (i) the degradation rate of RNA complexes and (ii) the ratio of transcript abundance (48). We focused on investigating the relative abundance of mARi in relation to its target mRNA. Unlike previous work where inducible expression of transcripts was used for simulating the shifting balances of transcript abundance (13, 37, 38), here we constitutively expressed both transcripts to test the impact of different transcript levels at steady state. Relative transcript abundance of mARi and mRNA at steady state was estimated from the relative promoter strength driving both regulator and target (**Supplementary Fig. 2b**). To assess the

impact of the mARi:mRNA ratio, a matrix of designs with four different promoter strengths for both mARi and mRNA were constructed (**Fig. 3a,b**). The resulting sfGFP expression showed a consistent trend, with sfGFP fluorescence decreasing as the mARi:mRNA ratio increased (**Fig. 3c, Supplementary Fig. 2c, Supplementary Table 3**). The maximum repression was observed when an excess of mARi was present in relation to its target mRNA, presumably because an excess of mARi is required to saturate binding of target mRNA. On the other hand, a low expression ratio led to weak repression, presumably due to an insufficient amount of mARi being available to inactivate mRNA target (45, 48).

A further variable affecting intracellular transcript abundance is plasmid copy number used in the system. Designs with both high (blue box in **Fig. 3a**) and low (grey box in **Fig. 3a**) expression ratios were selected to exemplify different levels of transcript abundance. These systems were first assembled with both mARi and its target in a single plasmid, with varying plasmid copy numbers from low to high: pSC101, p15A and pMB1 (50) (**Fig. 3d left panel**). Expression of mARi in either high or low expression ratio repressed sfGFP expression for all tested plasmid copy numbers, demonstrating that the mARi-based regulatory system is functional at different plasmid copy numbers (**Fig. 3e**).

We next reasoned that by placing the mARi on a second, higher copy vector, repression should be more efficient. To test this hypothesis, the mRNA expression cassette was placed on a p15A backbone, while mARi was placed on the higher copy pMB1 and pUC backbones, with both high- and low-expression ratios (**Fig. 3d right panel**). In this experimental design, the transcript expression ratio was governed by the combination of plasmid copy number and transcription rate of mARi relative to mRNA. Despite mARi being driven by a stronger promoter and being on a higher copy vector in all cases, the double-plasmid systems all showed less repression than that observed with the corresponding single-plasmid systems (**Fig. 3f**). This suggests a role for spatial organization of transcripts in post-transcriptional repression, as anticipated from recent theoretical analysis(51). The fact that the sRNA-based regulator has a shorter half-life and stability than mRNA may further reduce the mARi effectiveness when it is used in a double plasmid system with a broader physical transcript distribution.

mARi-based regulation is robust to different genetic contexts

Unlike other biomolecular regulators (e.g. transcription factors, TALEs, zinc fingers, or CRISPRi) (5, 41, 42), mARi only requires transcriptional activity. Thus, the expression of a short sequence for mARi would be expected to have a low cellular burden and not significantly impact host cell growth. To test this, the high transcript ratio for mARi in a p15A backbone and control plasmid without mARi expression were used to compare the effect of mARi overexpression on host fitness. We sought to test this repression system in four commonly used *E. coli* strains: DH5α, DH10b, BL21 (DE3), and BL21star (DE3). The high expression of mARi did not affect bacterial growth across different host strains, indicating the mARi-based repression system has a low production cost, and does not impact significantly on host cells (**Supplementary Fig.3**).

We then sought to evaluate the silencing activity of mARi in these *E. coli* K12 and B strains. The mARi-based gene regulation utilizes the native *E. coli* Hfq chaperone and degradosome complex (e.g. RNase E) as helper proteins to regulate gene repression. The intracellular abundance of these components is influenced by several factors, including the genotype and growth phase. For instance, the abundance of Hfq is constant during the exponential growth, but reduces when *E. coli* cells reach stationary phase (40, 54–57).

Functional characterisation of mARi at early stationary phase, 8 h (**Supplementary Fig. 3**), showed about 60-80% repression activity in both *E. coli* K-12 and B strains (**Fig. 3g**), while the silencing activity in the B strains was higher than in the K-12 strains for comparable plasmid copy numbers. (**Supplementary Fig. 4a**). The mARi system was active across different phases of cell growth (**Fig. 3h**): the *E. coli* K-12 strain (DH10b cells) showed relatively stable repression activity in both mid-exponential and early stationary growth, whereas a slightly increased repression activity was observed in early stationary phase for the BL21 (DE3) and BL21star (DE3) strains.

Expanding a set of orthogonal mARis through seed sequence modification

We sought to create a set of mARi-based regulators based on its underlying modular design principles coupled with the modular UTR-RBS linkers used in BASIC to construct both expression cassettes and operons. Five different UTR-RBS linkers were identified using R2oDNA Designer (28), whilst preserving the same medium strength RBS (RBSc). Computational evaluation of the five selected UTR (UTR-A-E) sequences demonstrated minimal similarity/identity scores between the set (**Supplementary Fig. 5a).** Low sequence similarity of non-cognate pairs is desirable to achieve minimal cross-talk interaction and off-target effects (45). The target 35 bp upstream of the RBS sequence were evaluated for their GC content and free binding energy of the target/seed sequences (**Supplementary Table 4)**. The GC content of the seed sequence is in a range of 31% to 43% and resulted in a free binding energy of -48.9 kcal/mol to -57.2 kcal/mol, which were in the favourable range for RNA-RNA interaction (45, 58). The seed sequences were also predicted to have minimal off-target interactions with three different *E. coli* genomes: *E. coli* MG1655, DH10b, and BL21 (DE3) **(Supplementary Table 4).**

To experimentally validate target specificity, we constructed all possible combinations of mARi and target UTR upstream of *sfgfp*; in each case maintaining the same promoter (**Fig. 4a**). The expression of sfGFP was significantly repressed in the presence of the cognate mARi systems relative to the non-cognate pairings (**Fig. 4a, Supplementary Fig.5c**), confirming their specificity with reduced cross-reactivity between each pair.

Modularity and orthogonality of mARi enables simultaneous and multiplexed gene regulation

After demonstrating that mARis exhibit specific interaction towards cognate UTRs in the context of a single gene, we evaluated their ability to operate in a multiplexed reporter gene design. Initially, this was evaluated with both reporter genes (i.e. *sfgfp* and *mcherry*) organized as separate transcriptional units (**Fig. 4b**). Specific and independent repression by each mARi towards its cognate target was observed, with UTR-A-*sfgfp* being selectively targeted by mARi-A and UTR-B-*mcherry* being

targeted by mARi-B (**Fig. 4b** and **Supplementary Fig. 6a**). Further, specific and independent repression was also confirmed when three mARis (mARi-A, mARi-B, and mARi-C) were employed to target UTR-A, UTR-B, and UTR-C preceding the *sfgfp*, *mcherry*, and *mtagbfp* genes, respectively (**Fig. 4c** and **Supplementary Fig. 6b**).

Two sets of mARis were used to regulate gene expression in a two-gene operon with UTR-A-RBSc driving the expression of *sfgfp*, upstream of UTR-B-RBSc driving the expression of *mcherry*, constitutively expressed by the J23101_BASIC promoter (**Fig. 4d**). With both genes on a single transcript, the targeted repression for the first gene (*sfgfp*) is strongly affected by its cognate mARi-A, with only a minor affect from mARi-B, which targets the downstream *mcherry*; the downstream *mcherry* is most strongly attenuated by its cognate mARi-B, but it is also significantly attenuated by repression of the upstream *sfgfp* with full attenuation only reached with both mARi-A and B (**Fig. 4d** and **Supplementary Fig. 6c**).

The mARi-based repression of a polycistronic mRNA provides some insight into its mechanism of action. Facilitated degradation would be expected to induce a similar repression level on both genes, but when only one of the reporter genes was targeted there was a differential effect: a non-targeted gene downstream of the mARi target exhibited significant repression, while a non-targeted gene upstream of the mARi target exhibited only slight repression (**Fig. 4d**). This behaviour is consistent with steric hindrance of the RBS being the main mode of action rather than active degradation. The silencing effects on the downstream *mcherry* were also additive, suggesting that repression at each UTR-RBS position was independent; it further suggests that read-through was possible from the upstream RBS, even when the downstream RBS was repressed, but this was occluded when both UTRs were bound by their respective mARis.

DISCUSSION

Here we report the development, characterization, and implementation of a modular post-transcriptional regulation system based on *trans*-acting sRNAs. The modular design of mARi is

inherited from the highly composable structure of the natural MicC sRNAs scaffold (23, 24). Modification of the seed sequence, whilst retaining the sRNA scaffold, has enabled the reprogramming of natural sRNAs with different targets of interests that do not exist in nature (24, 25, 59). Here, we utilised this modularity for targeted gene repression by directing mARis to standardized UTRs in the translation initiation region of the mRNA target. We demonstrated that targeted repression by binding to standardized sequences in the 5'-UTR upstream of the RBS were independent from both RBS and GOI contexts.

The computationally designed orthogonal UTR sequences of the BASIC UTR-RBS linkers proved to be ideal target sites for mARis (**Fig. 2c,d** and **4**). The standardised linker sequences that define the UTR have been computationally generated and validated to ensure their orthogonality in the DNA assembly process, as well as to the *E. coli* host (28). Therefore, by targeting the BASIC DNA linkers, it was anticipated that the mARi would exert orthogonal, post-transcriptional gene regulation with minimal cross-interaction to non-cognate targets, bioparts, plasmid backbones, and host genotypes (28). The modular nature of the design framework meant facile variation of genetic design parameters and it was observed that repression was modulated by both relative transcript abundance and spatial organisation (**Fig. 3**). Furthermore, the orthogonal basis of the design framework removed the need to design or insert bespoke target sites upstream of target genes, a drawback of previous works focused on reusable *trans*-encoded small RNAs (7, 20–22, 25).

It is also important to note that the designed mARis and UTR-RBS linkers omit an RNase E binding site, which is essential for RNase E-dependent cleavage. The elimination of RNase E in the mARi design was anticipated to reduce background degradation of unpaired mARis and improve their half-lives. The exclusion of an RNase E cleavage site in the mARi design may be beneficial for the implementation of an orthogonal repression system in various *E. coli* strains, with a response that is independent of the host genotype. Indeed, the system was observed to function across different *E. coli* strains and growth phases, demonstrating both its portability and robustness, notably with

minimal difference between BL21 (DE3) and BL21star (DE3), the latter being deficient in RNase E (**Fig. 3**).

The modular nature of the design enabled control of different genes and facilitated multiplexing of target genes. In total, five pairs of orthogonal mARis/UTRs were created and tested in this work. These mARis were applied to multiplexed and simultaneous post-transcriptional regulation in multi-gene systems, including both multiple transcriptional units and an operon architecture (**Fig. 4**). Importantly, this implementation was carried out simply by assembling regulators and targets using BASIC DNA assembly without re-optimisation of the regulator for different target genes. This extensibility is essential for the simultaneous regulation of multiple genes in metabolic engineering, layered genetic control, and advanced genetic circuit applications, while scalability is possible through our recently demonstrated low cost automation platform for BASIC assembly (26).

Author Contributions

A.D.: Designed and performed all experiments and data analysis, devised the experimental strategy and wrote the manuscript.

M.S.: Devised experimental strategy and design, assisted with the experiments and data interpretation and wrote the manuscript.

G.-B.S.: Contributed to the experimental strategy, assisted with experimental design and data interpretation and wrote the manuscript.

G.S.B.: Contributed to the experimental strategy, assisted with the experimental design and data interpretation and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Acknowledgements

A.D. received a PhD scholarship from Indonesia Endowment Fund for Education (LPDP). G.-B.S. gratefully acknowledges the support of the UK EPSRC through the EPSRC Fellowship for Growth EP/M002187/1 and of the Royal Academy of Engineering through the Chair in Emerging Technology programme. G.S.B. acknowledges the support of UK Research and Innovation through the Engineering and Physical Sciences Research Council [EP/R034915/1]; as well as the EU through H2020 [820699].

References

1. Cameron,D.E., Bashor,C.J. and Collins,J.J. (2014) A brief history of synthetic biology. *Nat. Rev. Microbiol.*, 10.1038/nrmicro3239.
2. Purnick,P.E.M. and Weiss,R. (2009) The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.*, **10**, 410–422, 10.1038/nrm2698.
3. Slusarczyk,A.L., Lin,A. and Weiss,R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.*, 10.1038/nrg3227.
4. Kelwick,R., MacDonald,J.T., Webb,A.J. and Freemont,P. (2014) Developments in the tools and methodologies of synthetic biology. *Front. Bioeng. Biotechnol.*, **2**, 10.3389/fbioe.2014.00060.
5. Xiang,Y., Dalchau,N. and Wang,B. (2018) Scaling up genetic circuit design for cellular computing: advances and prospects. *Nat. Comput.*, **17**, 833–853, 10.1007/s11047-018-9715-9.
6. Storz,G., Vogel,J. and Wassarman,K.M. (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell*, 10.1016/j.molcel.2011.08.022.
7. Huang,H.H., Qian,Y. and Del Vecchio,D. (2018) A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nat. Commun.*, **9**, 1–12, 10.1038/s41467-018-07899-z.
8. Thompson,K.M., Rhodius,V.A. and Gottesman,S. (2007) σE regulates and is regulated by a small RNA in Escherichia coli. *J. Bacteriol.*, **189**, 4243–4256, 10.1128/JB.00020-07.
9. Tu,K.C., Long,T., Svenningsen,S.L., Wingreen,N.S. and Bassler,B.L. (2010) Negative Feedback Loops

Involving Small Regulatory RNAs Precisely Control the Vibrio harveyi Quorum-Sensing Response. *Mol. Cell*, **37**, 567–579, 10.1016/j.molcel.2010.01.022.

10. Brosse,A., Korobeinikova,A., Gottesman,S. and Guillier,M. (2016) Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. *Nucleic Acids Res.*, **44**, 9650–9666, 10.1093/nar/gkw642.
11. Nitzan,M., Rehani,R. and Margalit,H. (2017) Integration of Bacterial Small RNAs in Regulatory Networks. *Annu. Rev. Biophys.*, **46**, 131–148, 10.1146/annurev-biophys-070816-034058.
12. Beisel,C.L. and Storz,G. (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiol. Rev.*, **34**, 866–882, 10.1111/j.1574-6976.2010.00241.x.
13. Levine,E., Zhang,Z., Kuhlman,T. and Hwa,T. (2007) Quantitative Characteristics of Gene Regulation by Small RNA. *PLoS Biol.*, **5**, 10.1371/journal.pbio.0050229.
14. Rice,J.B., Balasubramanian,D. and Vanderpool,C.K. (2012) Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proc. Natl. Acad. Sci.*, **109**, E2691–E2698, 10.1073/pnas.1207927109.
15. Morita,T. and Aiba,H. (2011) RNase E action at a distance: Degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev.*, **25**, 294–298, 10.1101/gad.2030311.
16. Pfeiffer,V., Papenfort,K., Lucchini,S., Hinton,J.C.D. and Vogel,J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.*, **16**, 840–846, 10.1038/nsmb.1631.
17. Man,S., Cheng,R., Miao,C., Gong,Q., Gu,Y., Lu,X., Han,F. and Yu,W. (2011) Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Res.*, **39**, e50–e50, 10.1093/nar/gkr034.
18. Nakashima,N., Tamura,T. and Good,L. (2006) Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli. *Nucleic Acids Res.*, **34**, 10.1093/nar/gkl697.
19. Engdahl,H.M., Lindell,M. and Wagner,E.G.H. (2001) Introduction of an RNA stability element at the 5′-end of an antisense RNA cassette increases the inhibition of target RNA translation. *Antisense*

Nucleic Acid Drug Dev., **11**, 29–40, 10.1089/108729001750072100.

20. Ghodasara,A. and Voigt,C.A. (2017) Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Res.*, **45**, 8116–8127, 10.1093/nar/gkx530.

21. Kelly,C.L., Harris,A.W.K.K., Steel,H., Hancock,E.J., Heap,J.T. and Papachristodoulou,A. (2018) Synthetic negative feedback circuits using engineered small RNAs. *Nucleic Acids Res.*, **46**, 9875–9889, 10.1093/nar/gky828.

22. Ceroni,F., Furini,S., Stefan,A., Hochkoeppler,A. and Giordano,E. (2012) A synthetic post-transcriptional controller to explore the modular design of gene circuits. *ACS Synth. Biol.*, **1**, 163–71, 10.1021/sb200021s.

23. Na,D., Yoo,S.M., Chung,H., Park,H., Park,J.H. and Lee,S.Y. (2013) Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nat. Biotechnol.*, **31**, 170–174, 10.1038/nbt.2461.

24. Yoo,S.M., Na,D. and Lee,S.Y. (2013) Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli. *Nat. Protoc.*, **8**, 1694–707, 10.1038/nprot.2013.105.

25. Lahiry,A., Stimple,S.D., Wood,D.W. and Lease,R.A. (2017) Retargeting a Dual-Acting sRNA for Multiple mRNA Transcript Regulation. *ACS Synth. Biol.*, **6**, 648–658, 10.1021/acssynbio.6b00261.

26. Storch,M., Haines,M.C. and Baldwin,G.S. (2020) DNA-BOT: A low-cost, automated DNA assembly platform for synthetic biology. *Synth. Biol.*, 10.1093/synbio/ysaa010.

27. Storch,M., Casini,A., Mackrow,B., Trewhitt,H., Ellis,T. and Baldwin,G.S. (2015) BASIC: a new Biopart Assembly Standard provides accurate, single-tier DNA assembly for synthetic biology. *ACS Synth. Biol.*, 10.1021/sb500356d.

28. Casini,A., Christodoulou,G., Freemont,P.S., Baldwin,G.S., Ellis,T. and MacDonald,J.T. (2014) R2oDNA Designer: Computational design of biologically-neutral synthetic DNA sequences. *ACS Synth. Biol.*, 10.1021/sb4001323.

29. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA : Interactive drawing and editing of the RNA secondary structure. **25**, 1974–1975, 10.1109/TCBB.2008.28.

30. Busch,A., Richter,A.S., Backofen,R., Group,B., Freiburg,A. and D-,F. (2008) IntaRNA : efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. **24**, 2849–2856, 10.1093/bioinformatics/btn544.

31. Wright,P.R., Georg,J., Mann,M., Sorescu,D.A., Richter,A.S., Lott,S., Kleinkauf,R., Hess,W.R. and Backofen,R. (2014) CopraRNA and IntaRNA: Predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.*, **42**, 119–123, 10.1093/nar/gku359.

32. Markham,N.R. and Zuker,M. (2005) DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, **33**, 577–581, 10.1093/nar/gki591.

33. Bonde,M.T., Pedersen,M., Klausen,M.S., Jensen,S.I., Wulff,T., Harrison,S., Nielsen,A.T., Herrgård,M.J. and Sommer,M.O.A. (2016) Predictable tuning of protein expression in bacteria. *Nat. Methods*, **13**, 233–236, 10.1038/nmeth.3727.

34. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950, 10.1038/nbt.1568.

35. Espah Borujeni,A., Channarasappa,A.S. and Salis,H.M. (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.*, **42**, 2646–2659, 10.1093/nar/gkt1139.

36. Raden,M., Ali,S.M., Alkhnbashi,O.S., Busch,A., Costa,F., Davis,J.A., Eggenhofer,F., Gelhausen,R., Georg,J., Heyne,S., *et al.* (2018) Freiburg RNA tools: A central online resource for RNA-focused research and teaching. *Nucleic Acids Res.*, **46**, W25–W29, 10.1093/nar/gky329.

37. Wright,P.R., Richter,A.S., Papenfort,K., Mann,M., Vogel,J., Hess,W.R., Backofen,R. and Georg,J. (2013) Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci.*, **110**, E3487–E3496, 10.1073/pnas.1303248110.

38. Storch,M., Casini,A., Mackrow,B., Ellis,T. and Baldwin,G.S. (2017) BASIC: A simple and accurate modular DNA assembly method.

39. Vytvytska,O., Moll,I., Kaberdin,V.R., Gabain,A. Von and Bla,U. (2000) Hfq (HF1) stimulates ompA mRNA decay by interfering with ribosome binding.

40. Večerek,B., Moll,I. and Bläsi,U. (2005) Translational autocontrol of the Escherichia coli hfq RNA chaperone gene. *RNA*, **11**, 976–984, 10.1261/rna.2360205.

41. Franze De Fernandez,M.T., Eoyang,L. and August,J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qβ-RNA. *Nature*, **219**, 588–290.

42. Sun,T., Li,S., Song,X., Pei,G., Diao,J., Cui,J., Shi,M., Chen,L. and Zhang,W. (2018) Re-direction of carbon flux to key precursor malonyl-CoA via artificial small RNAs in photosynthetic Synechocystis sp. PCC 6803. *Biotechnol. Biofuels*, **11**, 1–16, 10.1186/s13068-018-1032-0.

43. Huis in 't Veld,R.A.G., Kramer,G., van der Ende,A., Speijer,D. and Pannekoek,Y. (2017) The Hfq regulon of Neisseria meningitidis. *FEBS Open Bio*, **7**, 777–788, 10.1002/2211-5463.12218.

44. Puerta-Fernández,E. and Vioque,A. (2011) Hfq is required for optimal nitrate assimilation in the cyanobacterium Anabaena sp. strain PCC 7120. *J. Bacteriol.*, **193**, 3546–3555, 10.1128/JB.00254-11.

45. Hoynes-O'Connor,A. and Moon,T.S. (2016) Development of Design Rules for Reliable Antisense RNA Behavior in E. coli. *ACS Synth. Biol.*, **5**, 1441–1454, 10.1021/acssynbio.6b00036.

46. Chen,S., Zhang,A., Blyn,L.B. and Storz,G. (2004) MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli. *J. Bacteriol.*, **186**, 6689–6697, 10.1128/JB.186.20.6689.

47. Vogel,J. and Luisi,B.F. (2011) Hfq and its constellation of RNA. *Nat. Rev. Microbiol.*, **9**, 578–589, 10.1038/nrmicro2615.

48. Mitarai,N., Benjamin,J.-A.M., Krishna,S., Semsey,S., Csiszovszki,Z., Massé,E. and Sneppen,K. (2009) Dynamic features of gene expression control by small regulatory RNAs. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 10655–10659, 10.1073/pnas.0901466106.

49. Massaiu,I., Pasotti,L., Casanova,M., Politi,N., Zucca,S., Angelis,M.G.C. De and Magni,P. (2015) Quantification of the gene silencing performances of rationally- designed synthetic small RNAs. *Syst. Synth. Biol.*, 10.1007/s11693-015-9177-7.

50. Jahn,M., Vorpahl,C., Hübschmann,T., Harms,H. and Müller,S. (2016) Copy number variability of

expression plasmids determined by cell sorting and Droplet Digital PCR. *Microb. Cell Fact.*, 10.1186/s12934-016-0610-8.

51. Teimouri,H., Korkmazhan,E., Stavans,J. and Levine,E. (2017) Sub-cellular mRNA localization modulates the regulation of gene expression by small RNAs in bacteria. *Phys. Biol.*, **14**, 10.1088/1478-3975/aa69ac.

52. Nielsen,A.A.K., Segall-Shapiro,T.H. and Voigt,C.A. (2013) Advances in genetic circuit design: Novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.*, **17**, 878–892, 10.1016/j.cbpa.2013.10.003.

53. Bradley,R.W., Buck,M. and Wang,B. (2015) Tools and principles for microbial gene circuit engineering. *J. Mol. Biol.*, 10.1016/j.jmb.2015.10.004.

54. Azam,T.A., Iwata,A., Nishimura,A., Ueda,S. and Ishihama,A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *J. Bacteriol.*, **181**, 6361–6370.

55. Kajitani,M., Kato,A., Wada,A., Inokuchi,Y. and Ishihama,A. (1994) Regulation of the Escherichia coli hfq gene encoding the host factor for phage Q(β). *J. Bacteriol.*, **176**, 531–534.

56. Yoon,S.H., Han,M., Jeong,H., Lee,C.H., Xia,X., Lee,D., Shim,J.H., Lee,S.Y., Oh,T.K. and Kim,J.F. (2012) Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. *Genome Biol.*, **13**, R37, 10.1186/gb-2012-13-5-r37.

57. Moon,K. and Gottesman,S. (2011) Competition among Hfq-binding small RNAs in Escherichia coli. *Mol. Microbiol.*, **82**, 1545–1562, 10.1111/j.1365-2958.2011.07907.x.

58. Lee,Y.J. and Moon,T.S. (2018) Design rules of synthetic non-coding RNAs in bacteria. *Methods*, **143**, 58–69, 10.1016/j.ymeth.2018.01.001.

59. Sharma,V., Yamamura,A. and Yokobayashi,Y. (2012) Engineering artificial small RNAs for Conditional gene silencing in escherichia coli. *ACS Synth. Biol.*, **1**, 6–13, 10.1021/sb200001q.

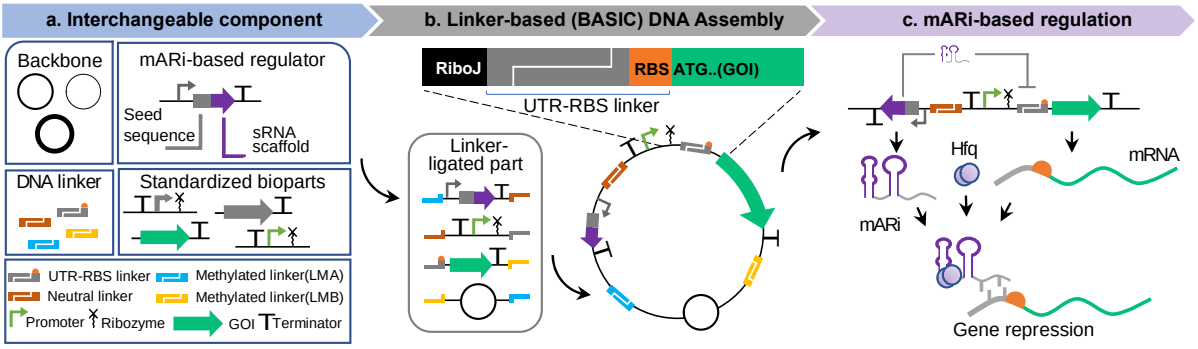

Fig. 1. mARi-based regulation integrated into the BASIC modular design and assembly framework. (**a**) mARi-based regulators composed of a promoter, seed sequence (grey), sRNA scaffold (purple) and terminator. These regulators were created as interchangeable components together with standardized DNA parts (i.e. plasmid backbone, promoter, and coding sequence) and computationally designed DNA linkers (i.e. neutral, UTR-RBS, and methylated linker) used in the BASIC DNA assembly framework (27). (**b**) Plasmids were constructed from the interchangeable DNA parts using linker-based DNA assembly: a representative assembled plasmid shows a functional UTR-RBS linker controlling expression of a downstream gene of interest (GOI). (**c**) Expression of mARi leads to repression of a target gene with a cognate UTR by base-pairing between the 5'untranslated region and mARi seed sequence.

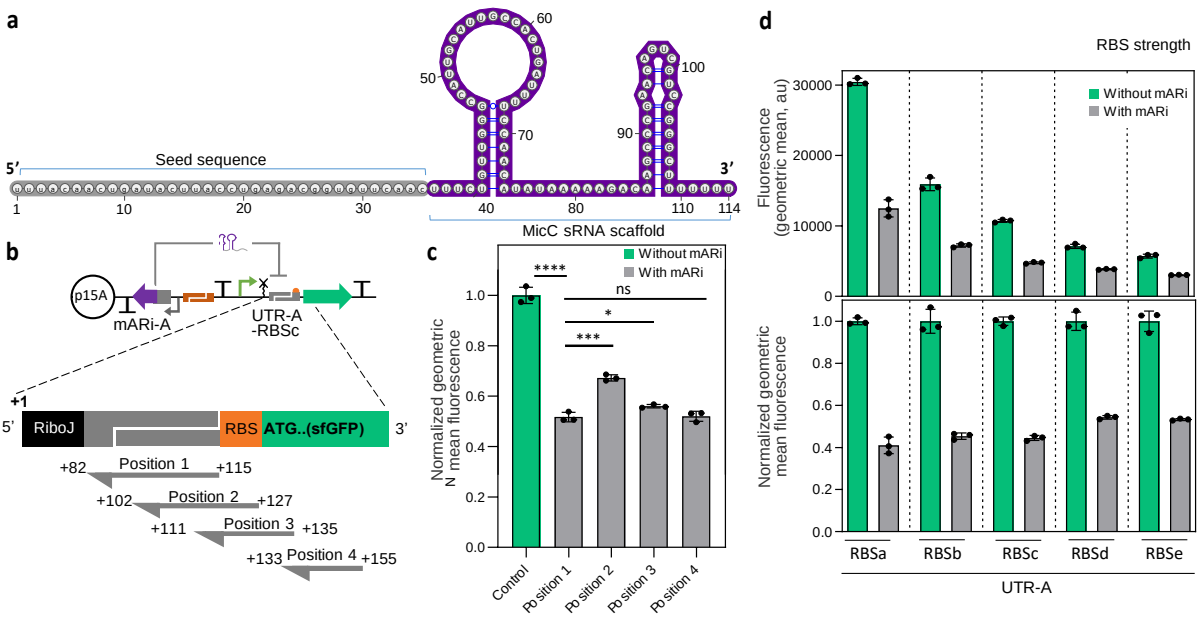

Fig. 2. Design and development of mARi-based gene regulation.

(**a**) Sequence and secondary structure of mARi containing a seed sequence (grey) and MicC RNA scaffold (purple). (**b**) Schematic design of the target site selection schema for mARi-mediated repression. Four positions within the translation initiation region were selected as target sites (Positions 1-4; **Supplementary Table 1**). The numbers indicate the relative position of mARi downstream of the transcription start site (+1). (**c**) Functional characterization of mARi-based gene regulation showing the silencing activity of mARi-A for the different target positions vs a sfGFP-only control without mARi. (**d**) The fluorescence intensity (top) and normalized fluorescence (bottom) when position 1 mARi-A was used to target UTR-A with varying RBS strengths. All sfGFP fluorescence measurements were performed by flow cytometry after 6 h growth. Data are shown with error bars for the mean ± SD of triplicate measurements (black dots). Statistically significant differences were determined using Student's t-test (**** represents $p<0.0001$, *** represents $p<0.001$, * represents $p<0.1$, ns represents not significant).

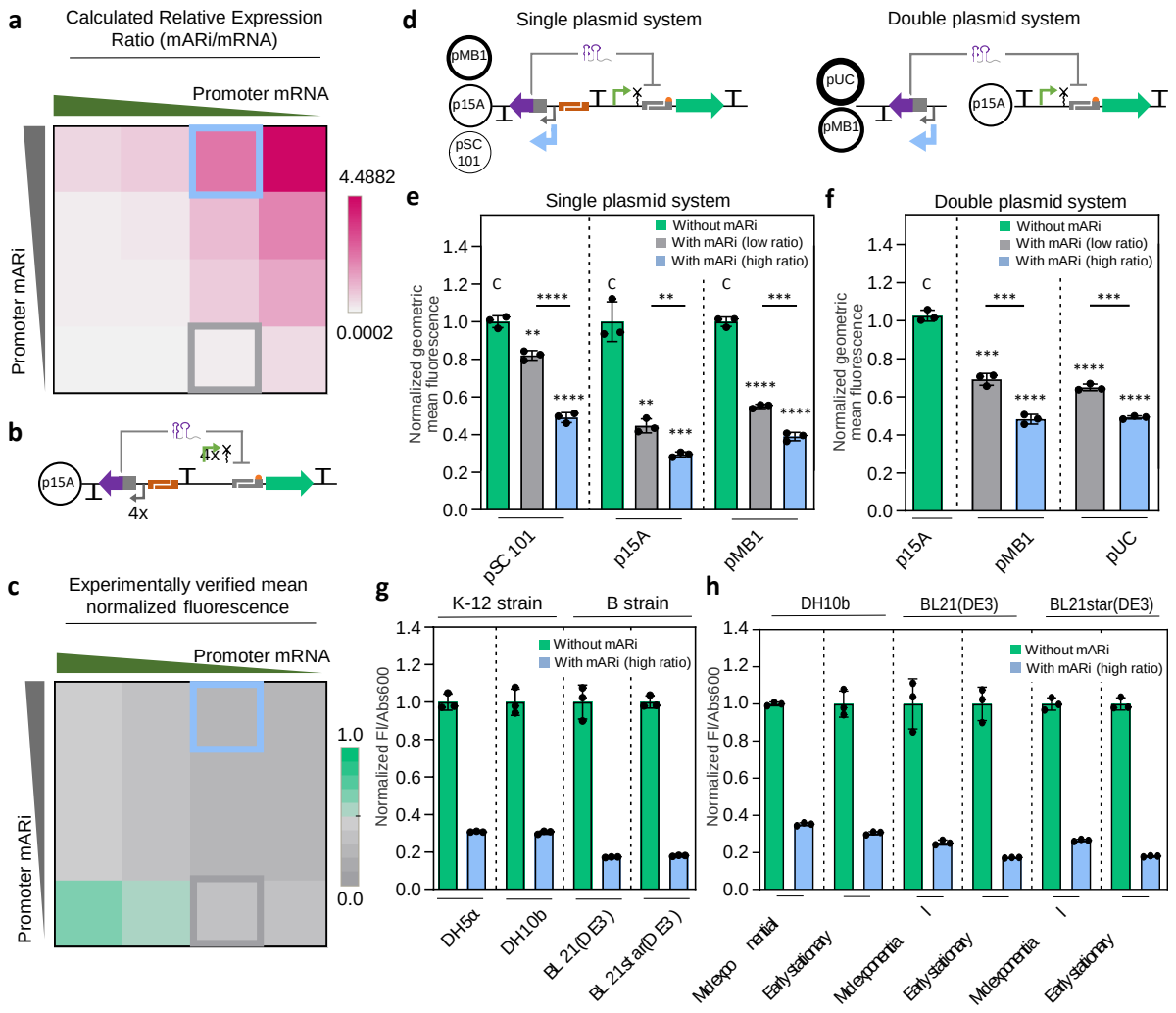

Fig. 3. Characterization of mARi-A across different genetic architectures.

(**a**) Heat-map showing relative expression ratio calculated from the relative strength of promoters driving mARi and mRNA expression across a matrix of constructs (**Supplementary Fig. 2b**). The representative conditions used for set-point high and low expression ratios are highlighted with a blue square (P_{J23119}-mARi:$P_{J23101BASIC}$-mRNA) and grey square (P_{J23101}-mARi:$P_{J23101BASIC}$-mRNA), respectively. (**b**) Schematic of mARi repression constructs used to investigate the effect of transcript ratios using four different strength constitutive promoters driving both mARi and target mRNA in a single plasmid system. (**c**) Heat-map showing the measured relative expression levels (mean normalized fluorescence) of the mARi expression ratio matrix from (**a**). The calculated relative expression ratio and repression activity of mARi are provided in **Supplementary Table 3**. (**d**) Schematic of the mARi-based repression system with high and low expression ratios constructed in single- and double-

plasmid systems with different copy numbers. Reporter expression was measured for the single (**e**) and double (**f**) plasmid systems for high and low ratio mARi against a control construct without mARi; sfGFP fluorescence measurements were performed by flow cytometry assay. (**g**) Reporter expression of the high-ratio mARi was measured in different *E. coli* strains against a control construct without mARi. (**h**) The repression activity of mARi in representative strains (DH10b, BL21 (DE3), and BL21star (DE3)) across different growth phases (**Supplementary Fig. 3**). Assays in **g** and **h** were performed by plate reader assay. All error bars show the mean ± SD of triplicate measurements (black dots). Statistically significant differences were determined against control without mARi expression (C) using Student's t-test (**** represents $p<0.0001$, *** represents $p<0.001$, ** represents $p<0.01$).

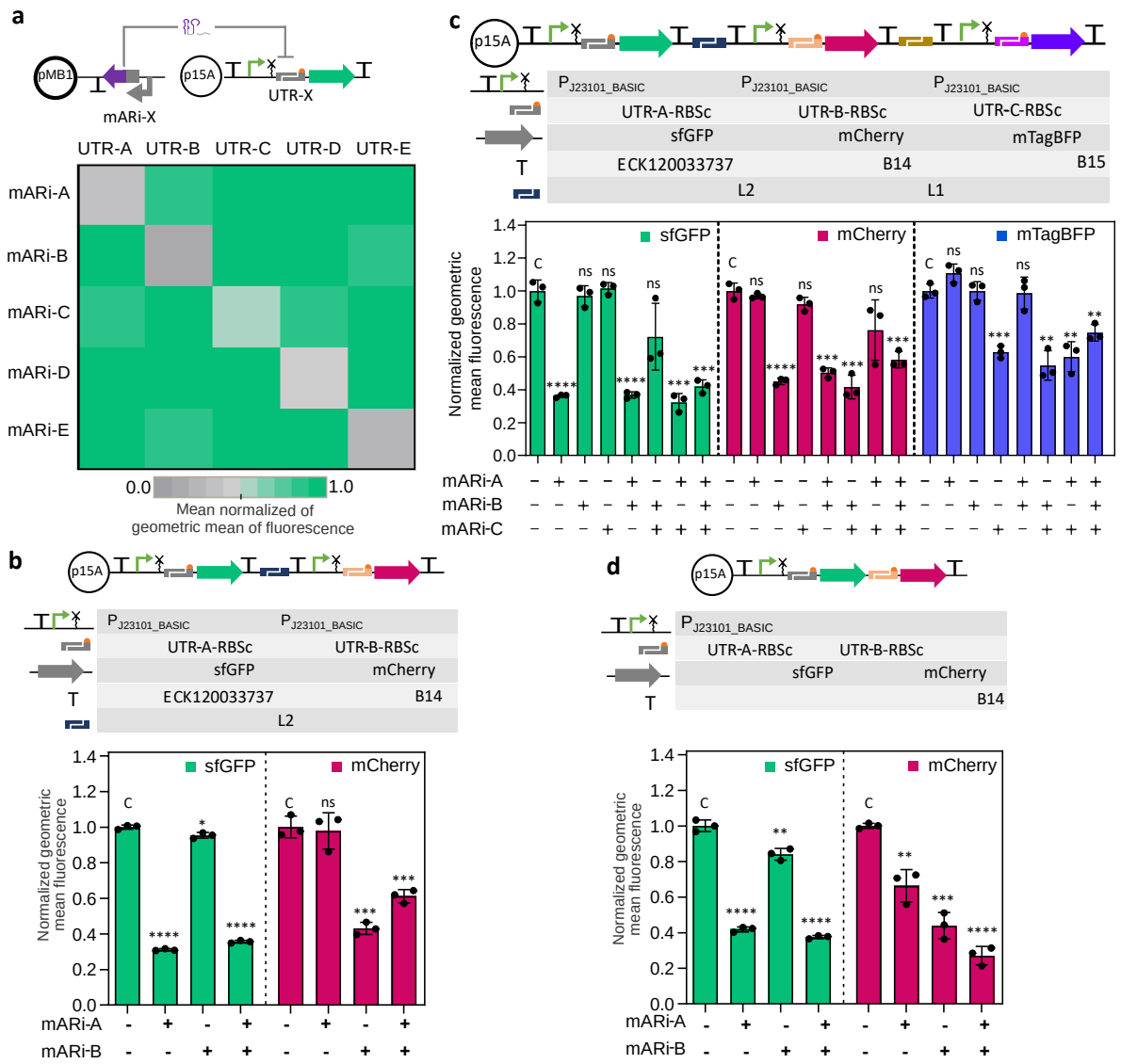

Fig. 4. mARi enables multiplex and simultaneous control of gene expression in a multi-gene system. (**a**) Fluorescence response of matrix of all possible mARi/UTR pairs measured by flow cytometry assay with triplicate repeats (**Supplementary Fig. 5c**). A schematic of experimental design to evaluate the target specificity of modular mARi-mRNA pairs is shown. (**b**) The fluorescence response of dual reporters in a double transcriptional unit system when combined with all possible mARi-A and mARi-B regulators was measured by flow cytometry assay. Schematic of reporters in dual transcription units is shown. (**c**) The fluorescence response of reporters in a triple transcriptional unit system when combined with all possible mARi-A, mARi-B, mARi-C regulators was measured by flow cytometry assay. Schematic of reporters in triple transcription units is shown. (**d**) The fluorescence response of dual reporters in an operon system when combined with all possible mARi-A and mARi-B regulators was

measured by flow cytometry assay. Schematic of dual reporters in an operon system is shown. All error bars show the mean ± SD of triplicate measurements (black dots). Statistically significant differences were determined against control without mARi expression (C) using Student's t-test (**** represents $p<0.0001$, *** represents $p<0.001$, ** represents $p<0.01$, * represents $p<0.1$, ns represents not significant).