Methods to regulate gene expression programs in bacterial cells are limited by the absence of effective gene activators. To address this challenge, we have developed synthetic bacterial transcriptional activators in *E. coli* by linking activation domains to programmable CRISPR-Cas DNA binding domains. Effective gene activation requires target sites situated in a narrow region just upstream of the transcription start site, in sharp contrast to the relatively flexible target site requirements for gene activation in eukaryotic cells. Together with existing tools for CRISPRi gene repression, these bacterial activators enable programmable control over multiple genes with simultaneous activation and repression. Further, the entire gene expression program can be switched on by inducing expression of the CRISPR-Cas system. This work will provide a foundation for engineering synthetic bacterial cellular devices with applications including diagnostics, therapeutics, and industrial biosynthesis.
Bacteria are attractive targets for a wide variety of engineering applications. Bacterial strains with the ability to utilize carbon sources like CO₂, CO, methane, or lignocellulose, and alternative energy sources such as light or H₂ could provide the foundation for cost-effective and environmentally-friendly industrial biosynthesis. Microbial communities, such as those that reside in the human gut, play an important role in human health and disease, and tools to engineer these bacteria have great potential as both diagnostics and therapeutics. To harness, regulate, and modify the behavior of these and other bacteria, there is a compelling need to develop genetic tools to control gene expression and implement complex, multi-gene regulatory programs. Ideally, we want to build circuits that can regulate many genes at once, dynamically respond to external inputs or the internal state of the cell, and be easily reprogrammed to explore different functional architectures. While capabilities to edit and modify genomes are rapidly expanding, our ability to encode a precisely-defined and dynamically-responsive gene expression program with cis-regulatory sequences at the DNA level remains difficult. Thus, we sought to develop synthetic transcription factors in bacteria, which could be coupled to programmable DNA binding domains and controlled by inducible promoters to engineer complex, dynamically-responsive multi-gene expression programs.

Synthetic control of gene expression has recently become much more straightforward with the emergence of programmable transcription factors using the CRISPR-Cas system (Fig. 1). A catalytically-inactive Cas9 (dCas9) protein can be used to target specific DNA sequences with guide RNAs (gRNAs) that recognize their targets based on predictable Watson-Crick base pairing. This approach can be used to repress genes by physically blocking RNA polymerase (CRISPR interference or CRISPRi)⁶,⁷. To activate genes (CRISPR activation or CRISPRa), the CRISPR complex can be linked to a transcriptional activator by direct fusion to dCas9 or via recruitment domains on the gRNA⁷–¹¹. In bacteria, however, there are very few transcriptional activation domains that have been reported to be effective when fused to modular DNA binding domains. Bacterial two-hybrid systems have been constructed with pairs of candidate interacting proteins separately fused to RNA polymerase subunits and DNA binding proteins. It is also possible to fuse RNA polymerase subunits directly to DNA binding domains to activate transcription²–⁴. One of the RNA polymerase subunits, RpoZ, has been coupled to the CRISPR system to activate gene expression⁷,¹³–¹⁸. For comparison, in eukaryotic systems there are many effective activators and CRISPRa has been extensively used in a variety of applications¹⁹. The paucity of reports of CRISPRa in bacteria suggests that RpoZ may not be effective as a general activator of transcription, or that we lack a complete understanding of the design rules to predictably activate gene expression in bacteria.

To develop an improved toolkit for gene activation in bacteria, we screened a broad set of candidate proteins for transcriptional activity in E. coli. We identified several proteins that can effectively activate gene expression when recruited via the CRISPR-Cas system. Using the most effective activator, SoxS, we can activate one target gene while simultaneously repressing a different target gene with CRISPRi, and we can control the entire multi-gene expression program with inducible promoters driving CRISPR-Cas system components. We find that gene activation in E. coli is highly sensitive to the location of the gRNA target site, consistent with prior results², and suggesting a possible explanation for why it has been difficult to develop synthetic activators in bacteria. Finally, we show that bacterial CRISPRa can be used to increase the output of a heterologous ethanol biosynthesis pathway. These results provide a framework for implementing CRISPRa in bacteria with a wide variety of potential applications. Further, because SoxS interacts with a highly conserved site on RNA polymerase, our bacterial CRISPRa toolkit may be portable to a broad range of bacterial species.

Results

Identifying transcriptional activation domains for bacteria. To recruit transcriptional activators to the CRISPR-Cas system, we used gRNAs that are extended with hairpin sequences to recruit RNA binding proteins (RBPs), which are in turn fused to candidate activators (Fig. 1a)¹⁰. These modified gRNAs, termed scaffold RNAs (scRNAs), encode both the target sequence and the regulatory action to execute at that target. By using scRNAs to recruit activators (CRISPRa) to some genes and gRNAs to physically block RNA polymerase (CRISPRi) at other genes, we can encode complex expression programs where some genes are activated and others are repressed (Fig. 1b)¹⁰.

To screen for potential activators, we first constructed an E. coli strain with a genomically-integrated, weakly-expressed GFP reporter gene. The upstream region of the reporter gene includes a number of potential gRNA target sites and is identical to that used previously to evaluate the dCas9-RpoZ fusion protein (see Supplementary Methods)². We targeted the CRISPR-Cas complex to the W108 gRNA site located 91 bases upstream of the transcriptional start site (TSS), as this site previously demonstrated the strongest activation with dCas9-RpoZ⁷. For candidate activators we chose endogenous transcriptional regulators SoxS, H2, and controlled by inducible promoters to engineer complex, dynamically-responsive multi-gene expression programs.

**Fig. 1** CRISPR activation in bacteria enables complex multi-gene expression programs. a To activate gene expression, we target a CRISPR-Cas complex upstream of a target gene. dCas9 binds a scaffold RNA (scRNA), which is a modified gRNA that encodes both the target sequence and an RNA hairpin to recruit effector proteins that interact with RNA polymerase. The schematic depicts a 1x MS2 scRNA containing an MS2 RNA hairpin, which binds the MS2 coat protein (MCP) that is fused to candidate activator proteins¹⁰. b Combining CRISPRi with CRISPRa enables multi-gene expression programs for simultaneous activation and repression. scRNAs that recruit activators can target genes for activation, while gRNAs targeted within a gene result in CRISPRi-based repression. If the CRISPR-Cas system components are controlled by inducible promoters, the entire gene expression program can be dynamically regulated.
MarA, Rob, and CAP\textsuperscript{20,21}; hijackers (i.e. bacteriophage or transposon effectors) TetD, λcII, GP33, N\textsubscript{4}SSB, and AsiA\textsuperscript{22–26}; and RNA polymerase subunits RpoZ (ω), RpoD (σ\textsubscript{70}), and the N-terminal domain of RpoA (α\textsubscript{NTD})\textsuperscript{12,13,27}. Of these candidates, α\textsubscript{NTD}\textsuperscript{12}, RpoZ\textsuperscript{7,13}, and AsiA\textsuperscript{26} have been previously reported to activate transcription when fused to heterologous DNA binding domains, and other candidates were selected based on literature reports suggesting that they could recruit RNA polymerase to activate transcription. Each of these candidate proteins was fused to the MS2 coat protein (MCP), which binds to an MS2 hairpin on the scRNA (Fig. 2a)\textsuperscript{10}.

Several candidate activators produced significant GFP reporter expression in late stationary phase expression, with the largest effects from SoxS and TetD, and smaller but still detectable GFP expression from λcII, α\textsubscript{NTD}, and RpoZ (Fig. 2a). Activation with MCP-RpoZ is significantly increased in a ΔpoZ host strain, consistent with that observed previously for other RpoZ fusion proteins, including dCas9-RpoZ (Supplementary Fig. 1a)\textsuperscript{7,13}. We also obtained significant GFP expression with the T4 bacteriophage activator AsiA (Supplementary Fig. 1b) by co-transforming with a σ\textsubscript{70} F563Y mutant, which prevents toxicity that occurs when AsiA is expressed alone and inhibits the endogenous σ\textsubscript{70} subunit\textsuperscript{26}. The most effective activator, SoxS, is a member of the AraC family of transcription factors. SoxS is expressed during oxidative stress and activates expression of a number of genes by binding to RNA polymerase in a pre-recruitment complex and scanning the genome to find DNA targets\textsuperscript{20,28}. This mechanism of action could explain the effectiveness of SoxS as a candidate activator for CRISPRa.

To confirm that the observed GFP expression arises from recruitment of the candidate activator upstream of the GFP reporter gene, we performed several negative controls with the SoxS activator. First, we demonstrated that activation requires the

**Fig. 2** Effector proteins can activate reporter gene expression in *E. coli*. a The CRISPRa complex targets a GFP reporter gene (sfGFP, superfolder GFP) driven by a weak BBa_J23117 promoter. The gRNA target is the W108 sequence, located 91 bases upstream of the transcription start site (TSS). Several candidate activator proteins fused to MCP result in significant increases in GFP expression, including SoxS and TetD. The dotted red line indicates the background fluorescence level observed in the parental MG1655 *E. coli* strain containing the GFP reporter gene. GFP levels were measured in late stationary phase. Similar trends with smaller overall effects were observed in exponential phase. b All components of the CRISPRa complex must be present for gene activation. Significant GFP expression is observed when dCas9, the 1x MS2 scRNA, and MCP-SoxS are all expressed. When MCP-SoxS is omitted (no activator) or the MS2 hairpin is removed (no MS2), there is no significant GFP expression. Modest GFP expression is observed when an off-target scRNA is used that has no target site in this strain (RR2, Supplementary Table 2), suggesting that overexpression of SoxS may have some off-target gene activation effects. Values reported are GFP fluorescence levels measured by flow cytometry. Values are median±s.d. for at least three biological replicates (specific values are indicated by black dots).
presence of the activator protein. When dCas9 and the 1x MS2 scRNA are expressed without MCP-SoxS, the CRISPR-Cas complex can bind upstream of the GFP reporter, but there is no significant GFP expression (Fig. 2b). Similarly, when the gRNA lacks the 1x MS2 recruitment hairpin, there is no significant GFP activation. Finally, we expressed dCas9, MCP-SoxS, and an off-target 1x MS2 scRNA and observed weak but detectable GFP expression (Fig. 2b). This result suggests that a small fraction of the GFP expression observed with MCP-SoxS arises from non-specific gene activation, which is plausible given that SoxS is an endogenous regulator of transcription in E. coli with many gene targets and a relatively degenerate binding site.

Optimizing activity of the SoxS activator. To optimize the activity of CRISPRa with SoxS and minimize off-target effects at endogenous SoxS gene targets, we systematically varied several design parameters of our system. First, we modified the scRNA structure to optimize it for bacterial expression. In our original design, the MS2 hairpin is appended at the 3′ end of the gRNA sequence, just downstream of an endogenous tracr terminator hairpin. Removing this terminator hairpin while retaining the 3′ MS2 hairpin leads to a 1.5–2-fold increase in GFP expression, likely due to increased steady-state levels of the full length scRNA (Fig. 3a). For comparison, in eukaryotic cells the same modified scRNA lacking the terminator reduces CRISPRa-mediated reporter gene expression ~2-fold (Supplementary Fig. 2a), possibly because the terminator hairpin interacts with Cas9. Presumably, the same detrimental effect is present in bacteria, but it is outweighed by the benefit from ensuring that the entire scRNA construct is expressed. Using this optimized scRNA design (scRNA.b1), we then varied the length of the amino acid linker connecting MCP and SoxS. 5 and 10 amino acid linkers increased GFP expression by ~2–3-fold compared to our original 2 amino acid linker (Fig. 3b). To increase activity further, we tested an scRNA design with two MS2 hairpins (2x MS2), but observed no significant increase in GFP expression.
compared to the 1x MS2 scRNA (Supplementary Fig. 2b). We also tested an alternative gRNA design in which MS2 hairpins are embedded within existing hairpins in the gRNA (sgRNA 2.0)11. We observed no significant GFP expression with sgRNA 2.0 (Supplementary Fig. 2c), which was surprising given that this design is highly effective in eukaryotic cells.

To minimize potential off-target effects at endogenous SoxS gene targets, we attempted to decouple its endogenous DNA binding activity from its transcriptional activation function. We identified several candidate residues at the SoxS DNA binding interface that are known to disrupt activity at endogenous sites when mutated33,34. We expected that if transcriptional activation could be decoupled from endogenous DNA binding, we would retain activity with our CRISPRa system even with SoxS mutants that are defective for activity at endogenous promoters. For two SoxS mutants, R93A and S101A, we observed no loss in activity in a CRISPRa assay; rather, we observed a 2-fold increase in GFP expression compared to wild type (wt) SoxS (Fig. 3c). The increased activity could arise if less SoxS is sequestered at endogenous DNA sites in these mutants. In contrast, the mutations R40A and F88A result in a substantial loss of GFP fluorescence in a CRISPRa assay, which could result from perturbations to protein structure or stability. To assess whether we effectively decoupled transcriptional activation from endogenous DNA binding, we plotted CRISPRa versus endogenous promoter activity for the series of SoxS mutants (Fig. 3d). To directly measure endogenous activity at SoxS gene targets when MCP-SoxS fusion proteins are expressed, we used LacZ fusion constructs for two different SoxS target genes, zwf and fumC (Supplementary Fig. 3), which are representative of class I and class II SoxS target genes, respectively34. The mutants that retain CRISPRa function, R93A and S101A, are ~2-fold weaker in endogenous activity than wt SoxS at zwf and 4-fold weaker at fumC. Thus, with SoxS R93A or S101A, we can partially decouple transcriptional activation from endogenous DNA binding. It may be possible to further decouple transcriptional activation from DNA binding with additional SoxS mutations.

We proceeded to test a fully optimized system with the 1xMS2 scRNA.b1 design, a longer 5 amino acid linker between MCP and SoxS, and the SoxS R93A mutant to reduce non-specific activity at endogenous SoxS targets. We observed a substantial increase in GFP fluorescence using an on-target guide RNA and a significant decrease in background activity with an off-target guide RNA (Supplementary Fig. 4) compared to the original, unoptimized system (Fig. 2b). Using RT-qPCR, we observed a 50-fold increase in GFP mRNA levels relative to negative control strains (Supplementary Fig. 5). Importantly, we observed no significant growth burdens associated with expression of the CRISPRa system (Supplementary Fig. 6a, b). Further, while we typically measured GFP levels in late stationary phase, we found similar trends in gene activation with smaller overall effects when GFP levels were measured in exponential phase (Supplementary Fig. 6c). We therefore proceeded with MCP-(5aa)-SoxS93A and 1x MS2 scRNA.b1 in future experiments. In some cases, we also used the functionally equivalent 1x MS2 scRNA.b2 design (Fig. 3a), which differs by one base (see Supplementary Methods). In direct comparisons, this fully optimized SoxS-based CRISPRa system performs better than the previously reported dCas9-RpoZ bacterial CRISPRa system1. Activation with SoxS is effective in an MG1655 strain where dCas9-RpoZ activity is undetectable, and in a ∆rpoZ MG1655 strain the SoxS-based system outperforms dCas9-RpoZ by >2-fold (Supplementary Fig. 7).

Inducible and simultaneous control of multiple genes. With CRISPR-Cas transcriptional programs, it is straightforward to target multiple genes for activation or repression6,8,10,35, although simultaneous activation and repression has not yet been demonstrated in bacteria. Using our optimized bacterial CRISPRa system, we tested whether we could simultaneously activate one gene while repressing another. We constructed an E. coli strain with two genomically-integrated fluorescent reporter genes, a weakly expressed GFP and a strongly expressed RFP. We expressed dCas9, MCP-(5aa)-SoxS93A, a 1x MS2 scRNA.b1 for GFP activation, and a gRNA targeting RFP for CRISPRi repression (Fig. 4a). When both the scRNA and the gRNA are expressed, we observe a simultaneous increase in GFP expression and decrease in RFP expression, and the magnitudes of the effects are indistinguishable from those observed when each gRNA is expressed alone (Fig. 4a). Thus, when all components are expressed constitutively, we can effectively control multiple genes simultaneously and have different effects at individual gene targets.

To develop the capability to dynamically regulate multi-gene programs, we first attempted to control the CRISPRa system with inducible promoters. In previous work, using Tet-inducible promoters to control CRISPRi in bacteria required pTet controlling both dCas9 and the gRNA6. We tested whether a similar strategy could effectively control CRISPRa. When dCas9 and the 1x MS2 scRNA are controlled by pTet and MCP-(5aa)-SoxS93A is expressed constitutively, however, we observed leaky GFP expression in the absence of the anhydrotetracycline (aTc) inducer (Fig. 4b), likely due to leaky expression from pTet36. In contrast, when either MCP-(5aa)-SoxS93A alone or all three components of the CRISPRa system are controlled by pTet, we observed no leaky GFP expression and pTet-inducible GFP levels comparable to that observed with constitutive expression (Fig. 4b). GFP levels for pTet-inductions were measured in late stationary phase cultures; in early stationary phase cultures we observed modestly weaker overall GFP induction compared to that observed with constitutive promoters. We also tested arabinose-inducible pBAD promoters as an alternative to pTet. With pBAD controlling all three components of the CRISPRa system, we observed inducible GFP expression, but at a level twofold weaker than that observed with constitutive expression. Alternatively, with pBAD controlling both dCas9 and the 1x MS2 scRNA, and constitutive expression of MCP-(5aa)-SoxS93A, we observed strong inducible GFP expression with no significant leaky expression in the absence of arabinose (Fig. 4b). These results were obtained with cultures in early stationary phase, unlike the results with pTet which were obtained in late stationary phase. In preliminary experiments with arabinose-inducible CRISPRi, we observed leaky repression in late stationary phase cultures, possibly because glucose depletion can relieve catabolite repression of the pBAD promoter37. We therefore performed the arabinose-inducible experiments in early stationary phase cultures, where no leaky CRISPRi repression was observed. Taken together, these results indicate that effective inducible control of CRISPRa can be achieved with pTet controlling MCP-(5aa)-SoxS93A, or with pBAD controlling dCas9 and the 1x MS2 scRNA.

With inducible control over CRISPRa and CRISPRi, we can implement a two-gene switch from ON/OFF to OFF/ON. We used separate pTet promoters to control all components of the system: dCas9, MCP-(5aa)-SoxS93A, the 1x MS2 scRNA for GFP, and the gRNA for RFP. Upon addition of aTc, GFP expression increases while RFP expression decreases (Fig. 4c). Alternatively, with pBAD promoters controlling dCas9, the 1x MS2 scRNA for GFP, and the gRNA for RFP, and constitutively expressed MCP-(5aa)-SoxS93A, we could also inducibly switch GFP on and RFP off (Fig. 4c).
Fig. 4 Multi-gene inducible control. a Expressing multiple gRNAs allows simultaneous regulation of multiple genes. Using an E. coli strain with integrated GFP and RFP reporters, a 1x MS2 scRNA targets GFP for activation by recruiting MCP-(5aa)-SoxSR93A, while an unmodified gRNA targets RFP for repression. When both gRNAs are expressed, GFP expression increases while RFP expression decreases, and the observed effects are similar to those observed when each gRNA is expressed alone. OT indicates an off-target control for the J106 target site (Supplementary Table 2), which is not present in this strain. Values reported are GFP fluorescence levels measured by flow cytometry. Values are median±s.d. for three biological replicates (specific values are indicated by black dots).

b CRISPRa can be inducibly controlled with pTet or pBAD (Ara) promoters. In each case, different components of the CRISPRa system (i.e. dCas9, the 1x MS2 scRNA, or MCP-(5aa)-SoxSR93A) are controlled by constitutive (c) or inducible (pTet, T or pBAD, A) promoters. pTet is induced with aTc and pBAD is induced with arabinose. For pTet inductions, cultures were harvested after overnight growth (late stationary phase), while for arabinose inductions cultures were harvested after 6 h (early stationary phase). Values reported are GFP or RFP fluorescence levels measured by flow cytometry. Values are median±s.d. for three biological replicates (specific values are indicated by black dots).

c CRISPRa and CRISPRi can be simultaneously induced with pTet or pBAD (Ara) promoters. For pTet-induction, all components of the CRISPR system are controlled by pTet. For arabinose induction the pBAD promoter controls dCas9 and the guide RNAs, while MCP-(5aa)-SoxSR93A is constitutively expressed. The RFP axis in the Ara panel is smaller than all other RFP axes in this figure because we observed consistently lower fluorescent protein expression in early stationary phase cultures versus late stationary phase cultures, even in the absence of CRISPRi. Values reported are GFP or RFP fluorescence levels measured by flow cytometry. Values are median±s.d. for three biological replicates, except for the +aTc tet inductions, for which two biological replicates were obtained (specific values are indicated by black dots).
Bacterial CRISPRa is highly sensitive to gRNA target site. In prior reports of bacterial CRISPRa with dCas9-RpoZ, reporter gene expression depended on the distance of the gRNA target from the TSS, with peak expression occurring at a site located ~90 bases upstream. To determine if a similar dependence applies to different activators recruited via 1x MS2 scRNA.b2 to the CRISPR-Cas complex, we designed a synthetic promoter (J1) that has gRNA target sites with the necessary PAM sequences every 10 bases on both strands upstream of a weak BBa_J23117 promoter driving an RFP reporter gene (Fig. 5a). With the MCP-(5aa)-SoxS<sub>93A</sub> activator, we observed a sharp dependence of RFP expression on target site position, with significant RFP expression occurring for sites at 80–90 bases upstream of the TSS on the non-template strand and at 60–80 bases upstream on the template strand (Fig. 5b). To determine if activators with weak activity in our initial assays (Fig. 2a) might be more effective at different target sites, we tested MCP fusion proteins of TetD, λcII, and α<sub>NTD</sub>-MCP with this promoter. We observed peak activities for these activators at target site positions similar to the most effective sites for SoxS (Fig. 5b), and in no case did we observe dramatic increases in activity relative to our initial assays (Fig. 2a).

Similarly, we tested whether an alternative gRNA design (sgRNA 2.0) with MS2 sites embedded within internal hairpins in the gRNA and the MCP-(5aa)-SoxS<sub>93A</sub> activator would be more effective at different target sites; we observed no detectable CRISPRa activity with this gRNA design at any target site. Finally, we tested an alternative RNA hairpin (1x PP7 scRNA.b1), that recruits the PP7 coat protein (PCP) fused to SoxS<sub>93A</sub>. MCP and PCP are structurally homologous proteins, but the MS2 and PP7 RNA hairpins differ significantly and interact in distinct orientations with their cognate binding proteins. Thus, we do not expect activators recruited by MS2 or PP7 to be positioned in...
similar orientations when recruited to the CRISPR-Cas complex. Perhaps because of this structural difference, PCP-(5aa)-SoxSR93A behaved very differently than MCP-(5aa)-SoxSR93A. The PCP construct was relatively ineffective at all sites on the non-template strand and moderately effective on the template strand (Fig. 5b). These results suggest that bacterial CRISPRa is highly sensitive to target position upstream of the TSS, and that it may be possible to find different combinations of activators, recruitment domains, and target sites that perform better than the most effective SoxSR constructs tested here. Because all of the activators tested here interact directly with RNA polymerase, albeit at different protein interfaces, we suggest that strong activation requires a CRISPR target site that allows RNA polymerase to bind to the minimal promoter (i.e. the −35 and −10 sites) and the activator simultaneously, which puts a significant constraint on the upstream target sites that can be used for CRISPRa.

To extend the range of effective target sites, we hypothesized that a longer, flexible linker between MCP and SoxSR93A might allow activators to recruit RNA polymerase to the promoter from a broader range of target sites. We therefore tested linkers with 2, 5, 10, or 20 amino acids. Extending the original 5 amino acid linker to 20 amino acids should lengthen the linker by ~57 Å (assuming a worm-like chain model for a flexible peptide linker with 3.8 Å/residue)39. Because target sites on the J1 promoter are spaced ~33 Å apart (10 bp spacing with 3.3 Å/bp for B-DNA), we expect that the extended linker should broaden the range of target sites by at least 10–20 bp in either direction from the optimal range observed with the 5 amino acid linker. Surprisingly, we observed no significant broadening of the effective target site range with longer linkers (Fig. 5c). We observed maximal activity with 5 and 10 amino acid linkers, while both 20 and 20 amino acid linkers showed reduced activity but no significant difference in the position of effective target sites. While it remains to be seen if it is possible to design a bacterial CRISPRa complex that functions over a broader target range, simply extending linker lengths between MCP and the SoxSR93A activator was not effective.

**Activating a metabolic gene cluster for ethanol biosynthesis.** An important practical application of bacterial CRISPRa will be to enable complex and dynamic multi-gene circuits to control biosynthetic pathways. The capability to express pathway enzymes while repressing competing enzymes may lead to significant improvements in product yields10. As a proof of concept, we tested whether we could upregulate a heterologous pdc adhB gene cassette from Zymomonas mobilis, which converts pyruvate to ethanol and has previously been used for ethanol production in E. coli (Fig. 6).11,42 When we targeted the heterologous gene cassette with our CRISPRa system, we observed a threefold increase in ethanol production relative to cells without CRISPRa (Fig. 6). This result suggests that CRISPRa-based transcriptional control can be used to activate biosynthetic pathways. We expect that combining CRISPRa-based control of heterologous biosynthetic genes with CRISPRi on competing endogenous genes will enable rapid explorations of a large space of genetic circuit architectures to improve biosynthesis yields, and further improvements may be realized by targeting additional metabolic genes for activation or repression, and by coupling the system to dynamically-regulated or inducible promoters.

**Discussion**

In this work, we have identified multiple synthetic transcriptional activators compatible with CRISPRa in E. coli, including SoxS, TetD, and AsiA. The most effective activator, SoxS, is substantially stronger than the previously reported RpoZ-based synthetic activator and does not require additional host-strain modifications for function. SoxS may be an effective activator in part because it interacts with the C-terminal domain of RpoA, which is connected to RNA polymerase by a relatively flexible tether12,43. Using SoxS and other activators, we find a surprisingly sharp dependence on the CRISPR target site distance from the TSS for effective gene activation, with the optimal gRNA sites positioned within a narrow window roughly 60–90 bases upstream of the transcription start site of a heterologous reporter gene. This range is strikingly smaller than the broad target site range observed for effective CRISPRa in eukaryotic cells, where many target sites in the 1–500 base range upstream of the TSS are effective44. One notable difference between bacteria and eukaryotes is that the effective bacterial activators identified in this and other work all appear to directly interact with RNA polymerase, while in eukaryotic cells transcription factors may act over longer distances via indirect chromatin modifications. Nevertheless, even given these constraints the ability to perform effective CRISPRa in bacteria will open significant avenues for engineering bacterial systems. Most importantly, we expect that
controlling branch points in metabolic networks with simultaneous gene activation and repression will improve biosynthetic yields beyond that obtained simply by constitutively over-expressing heterologous pathways. Further, we can use inducible promoters to control the timing of gene expression, and we can build more sophisticated dynamic gene expression programs using protein or RNA biosensors to sense and respond to cellular metabolic states.\textsuperscript{15,30}

Two future challenges remain for broad applicability of CRISPRa in bacteria. First, it will be necessary to identify predictive rules for targeting and activating gene expression at endogenous sites. In preliminary experiments, we have observed only modest increases in gene expression at endogenous sites, and the position of effective target sites does not necessarily correspond to the optimal target sites that we observed for heterologous promoter activation. It is possible that our observed distance dependence is not generalizable to different promoters, that endogenous regulatory factors interfere with CRISPRa, or that different types of promoters require distinct transcriptional activation domains. In practice, if predictive rules for targeting arbitrary endogenous gene targets become available, we envision using gene editing to introduce heterologous promoters at target sites of interest, which will enable us to regulate endogenous genes as part of a multi-genome CRISPR/a control program.

A second outstanding challenge for bacterial CRISPRa is to develop a system that is portable across bacterial species of high commercial and industrial value, such as strains that have the ability to utilize carbon sources like CO\(_2\), CO, methane, or lignocellulose, and alternative energy sources such as H\(_2\) or light.\textsuperscript{1,2} For gene repression, CRISPRi has been successfully used in a broad range of bacterial species,\textsuperscript{16,47} suggesting that the programmable DNA targeting component of the system is portable. For CRISPRi, many of our candidate gene activators interact with motifs on bacterial RNA polymerase that are highly conserved across a broad range of bacteria,\textsuperscript{51} including several that are relevant for industrial biosynthesis or microbiome engineering. In particular, SoxS interacts with a surface on the RNA polymerase subunit that is well-conserved in gammaproteobacteria, alphaproteobacteria, bacteroides, gram-positive bacteria, and even to a lesser extent in cyanobacteria (Supplementary Fig. 8). This conserved interface may allow CRISPRi systems developed in \textit{E. coli} to be ported to non-model bacteria with a wide range of useful biological functions.

**Methods**

**Bacterial strain construction and manipulation.** Plasmid constructs were cloned and \textit{E. coli} cells were cultured using standard molecular biology methods. Gene knockouts and integrations of fluorescent reporters were performed by recombining in a strain with a genomically-integrated lambda red system under the control of a temperature-sensitive promoter (NM700).\textsuperscript{52-54} Modified genome fragments were then transferred to MG1655 by P1 transduction\textsuperscript{26}. The stGFP reporter was integrated at the \textit{ nfsA} locus. The mRFP reporter (for CRISPRi experiments) was integrated at \textit{rhaA}.\textsuperscript{26} The \textit{ΔrpoZ} strain was constructed by recombining a pKD13-derived Kan\(_R\) linear PCR cassette with 36 base homology overlaps to flanking sites at the \textit{rpoZ} locus. The \textit{ΔrpoZ::KanR} knockout genomic fragment was transferred to MG1655 by \textit{P1} transduction, and the FRT-flanked Kan\(_R\) cassette was eliminated by transformation with the \textit{pCP20} helper plasmid to express FLP recombinase, followed by incubation at 42 °C to cure the plasmid.

**CRISPR-Cas systems.** CRISPR-Cas system components were delivered on plasmids as described in Supplementary Methods. Standalone CRISPR-Cas system components are included in the Supplementary Methods along with a list of the engineered \textit{E. coli} strains (Supplementary Table 1). Guide RNA target sites are listed in Supplementary Table 2.

**Flow cytometry.** Cells were inoculated in EZ-RDM (Teknova) supplemented with appropriate antibiotics and grown in 96-deep-well plates at 37 °C, 220 RPM overnight, to late stationary phase cultures were then diluted 1:40 in PBS and analyzed on a MACSQuant VYB flow cytometer (Milenyi Biotec). To enrich for single cells, a side scatter threshold trigger (SSC-H) was applied. To gate for single bacterial cells, we first selected events along the diagonal of the SSC-H vs. SSC-A plot.\textsuperscript{23} We then excluded events that appeared on the edges of the SSC-A vs. FSC-A plot and events that appeared on the edge of the fluorescence histogram (Supplementary Fig. 9).

For inducible CRISPR system construction with \textit{pTet} or \textit{pBAD} promoters, strains were inoculated in 3 mL LB medium supplemented with antibiotics and grown overnight at 37 °C, 220 RPM. On the next day, they were diluted 1:100 in 500 mL EZ-RDM supplemented with antibiotics and induced with 1 μM anhydrotetracycline (aTc) for \textit{pTet} or 100 mM IPTG for \textit{pBAD}. Non-induced controls were prepared for each strain. For \textit{pTet}, cells were grown at 37 °C, 220 RPM overnight. For \textit{pBAD}, cells were grown at 37 °C, 220 RPM for 6 h. In both cases, fluorescence was assayed via flow cytometry as described above.

**LacZ reporter assays.** LacZ reporter assays for \textit{zwf} and \textit{fumC} promoters were performed following a previously reported protocol with minor modifications.\textsuperscript{60,61} Specifically, cultures were grown for 18 h in EZ-RDM supplemented with antibiotics. Absorbance at 600 nm (OD\(_{600}\)) was measured from 150 μL samples in a Biotek Synergy HTX plate reader. 20 μL aliquot from each overnight culture was diluted to 80 μL pre-warmed solution of 0.2 mM KCI, 2 mM MgSO\(_4\), 0.8 mg/mL hexacycl/triethylammonium bromide, 0.4 mg/mL sodium deoxycholate, 5.4 μL/mL β-mercaptoethanol. Samples were incubated at 30 °C for 20 min. To initiate the reaction, 600 μL of substrate solution [60 mM Na\(_2\)HPO\(_4\), 40 mM NaH\(_2\)PO\(_4\), 1 mg/mL n-octylphenol-β-D-galactopyranoside, 0.02% Triton X-100] was added. β-mercaptoethanol was added at 30 °C until visible color developed, at which point 700 μL of stop solution [1 M Na\(_2\)CO\(_3\)] was added and the reaction time was recorded. Samples were centrifuged for 10 min to pellet cell debris, and the supernatants were removed. Absorbance at 420 nm (A\(_{420}\)) was measured from 150 μL samples in a plate reader. LacZ activity was calculated according to the formula: Miller Units = (1000 × A\(_{420}\)) / (OD\(_{600}\) × 0.02 mL × time).

**J1 reporter sequence design.** Custom Python scripts were used to generate a tiling array containing an NNG PAM site every 10 nucleotides on each strand according to the following base unit: 5′-NNNNCNNNNGG-3′. 10000 sequences of length 500 nt were generated by randomly sampling \textit{N} nucleotides, adjusted for endogenous GC content in \textit{E. coli} (BNID:100532) (http://bionumbers.hms.harvard.edu/bionumber.aspx?id=100558)\textsuperscript{24}. We discarded any sequences with four consecutive identical nucleotides. Sequences containing known transcription factor binding sites (using the experimental TF binding site database from RegulonDB\textsuperscript{65}) were also discarded. Of the remaining sequences, we arbitrarily chose one, labeled as J1, and confirmed that it had no detectable homology to the \textit{E. coli} genome. We placed a 170 bp fragment of this sequence upstream of the weak \textit{BbaJ23117} constitutive promoter (http://partscigen.org).

**Plate reader experiments.** For fluorescent reporter experiments with the J1 reporter, cells were inoculated in 2 mL EZ-RDM supplemented with appropriate antibiotics and grown at 37 °C, 220 RPM overnight. OD\(_{600}\) and observed fluorescence values were measured in a Biotek Synergy HTX plate reader (Excitation: 420 nm (A\(_{420}\)) was measured from 150 μL samples in a plate reader. LacZ activity was calculated according to the formula: Miller Units = (1000 × A\(_{420}\)) / (OD\(_{600}\) × 0.02 mL × time).

**Ethanol fermentations.** \textit{E. coli} MG1655 cells were transformed with the pCD355 plasmid containing the \textit{pdc adhB} gene cassette from \textit{Z. mobilis} under the control of a weak promoter (Supplementary Table 3) and with or without components of the CRISPRa system. Transformed cells were inoculated in 5 mL LB (2% glucose) supplemented with appropriate antibiotics and grown overnight at 37 °C, 220 RPM. OD\(_{600}\) measurements were taken from the overnight cultures, and 2.5 OD·mL of the culture was diluted into 50 mL LB supplemented with antibiotics. Cultures were grown aerobically for 18 h at 37 °C, 220 RPM. OD\(_{600}\) measurements were taken, and 15 OD·mL was pelleted and resuspended into 25 mL M9 media with 2% dextrose and antibiotics in a 125 mL flask. Cultures were incubated at 37 °C without shaking for 4 days. Ethanol concentrations in supernatants were measured with an Ethanol Oxidation Kit (R-Biopharm). We counted viable cells by plating on selective media before and after fermentation and observed no decrease in viable cells after the fermentation. To evaluate plasmid stability over the course of the fermentation, we performed mini-preps on cells after the fermentation. Analytical restriction digests indicated no large-scale recombination events, and no point mutations were detected by sequencing.

**Quantitative RT-PCR.** Strains were inoculated in 5 mL LB supplemented with appropriate antibiotics and grown overnight at 37 °C, 220 RPM. Cultures were then

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-04901-6 | www.nature.com/naturecommunications
diluted 1:100 in 5 mL EZ-RDM supplemented with antibiotics and grown to OD600 0.5 (using 150 µL samples in flat clear bottomed 96-well plates in a Biotek Synergy HFTX reader). Cultures were pelleted, flash-frozen in liquid nitrogen, and stored at –80 °C. Total RNA was extracted using an Aurum Total RNA Mini Kit (Bio-Rad). 1 µg of RNA was converted to cDNA using Script reverse transcriptase (Bio-Rad) in 20 µL reactions. qPCR was performed using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad) with a 58 °C annealing temperature and 10 µL reaction volumes. qPCR reactions were performed in triplicate on a CFX Connect (Bio-Rad) using 0.5 ng of cDNA, 400 nM primer concentration, and 15 s extension time. 16 S rRNA was used for normalization. Control samples without a template and without reverse transcriptase were analyzed to confirm gene-specific amplification and absence of gDNA contamination. Primer sequences are listed in Supplementary Table 4. Expression levels for each gene were calculated relative to 16S using the ΔΔCT method.

Code availability. Custom Python code to generate wiring arrays with PAM sequences (detailed in ”1 reporter sequence design” section of Methods) is available upon request.

Data availability. All data from this study are available upon request.

Received: 5 February 2018 Accepted: 1 June 2018

Published online: 27 June 2018

References

1. Haynes, C. A. & Gonzalez, R. Rethinking biological activation of methane and conversion to liquid fuels. Nat. Chem. Biol. 10, 331–339 (2014).
2. Lan, E. I. & Liao, J. C. Microbial synthesis of n-butanol, isobutanol, and other higher alcohols from diverse resources. Bioresour. Technol. 135, 339–349 (2013).
3. Mimeo, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a human commensal bacterium, Bacteroides thetaiotaomicron, to sense and respond to stimuli in the murine gut microbiota. Cell Syst. 1, 62–71 (2015).
4. Whitaker, W. R., Shepherd, E. S. & Sonnenburg, J. L. Tunable expression tools enable single-cell strain distinction in the gut microbiome. Cell 169, 538–546.e12 (2017).
5. Riglar, D. T. et al. Engineered bacteria can function in the mammalian gut long-term as live diagnostics of inflammation. Nat. Biotechnol. 35, 653–658 (2017).
6. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183 (2013).
7. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. 41, 7429–7437 (2013).
8. Gilbert, L. A. et al. CRISPR-mediared modular RNA-guided regulation of transcription in eukaryotes. Cell 154, 442–451 (2013).
9. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963 (2013).
10. Zalatan, J. G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell 160, 339–350 (2015).
11. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588 (2015).
12. Dove, S. L., Jong, J. K. & Hochschuld, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. Nature 386, 627–630 (1997).
13. Dove, S. L. & Hochschuld, A. Conversion of the omega subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. Genes. Dev. 12, 745–754 (1998).
14. Dove, S. L. & Hochschuld, A. A bacterial two-hybrid system based on transcription activation. Methods Mol. Biol. 261, 231–246 (2004).
15. Otoupal, P. B., Erickson, K. E., Escalas-Bordoy, A. & Chatterjee, A. CRISPR perturbation of gene expression alters bacterial fitness under stress and reveals underlying epistatic constraints. ACS Synth. Biol. 6, 94–107 (2017).
16. Peters, J. M. et al. Bacterial CRISPR: accomplishments and prospects. Curr. Opin. Microbiol. 27, 121–126 (2015).
17. Choi, K. R. & Lee, S. Y. CRISPR technologies for bacterial systems: Current achievements and future directions. Biotechnol. Adv. 34, 1180–1209 (2016).
18. Peng, R. et al. CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus. Microb. Cell. Fact. 15, 18 (2015).
19. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. 17, 5–15 (2016).
20. Martin, R. G., Gillette, W. K., Martin, N. I. & Rosner, J. L. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in Escherichia coli. Mol. Microbiol. 43, 355–370 (2002).
21. Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). J. Mol. Biol. 293, 199–213 (1999).
22. Griffith, K. L., Becker, S. M. & Wolf, R. E. Characterization of TetD as a transcriptional activator of a subset of genes of the Escherichia coli SoxS/MarA/Rob regulon. Mol. Microbiol. 56, 1103–1117 (2005).
23. Jain, D. et al. Crystal structure of bacteriophage lambda cII and its DNA binding protein (CAP). Science 275, 1655–1657 (1997).
24. Gregory, B. D., Deighan, P. & Hochschuld, A. An artificial activator that contacts a normally occluded surface of the RNA polymerase holoenzyme. J. Mol. Biol. 353, 497–506 (2005).
25. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning mechanism of transcription activation by SoxS of Escherichia coli: the dominance of DNA binding mutations of SoxS. J. Mol. Biol. 344, 1–10 (2004).
26. Miller, A., Wood, D., Ebrigh, R. H. & Rothman-Denes, L. B. RNA polymerase β subunit: a target of DNA binding-independent activation. Science 275, 7429–7436 (1998).
27. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. 177, 4121–4130 (1995).
28. Chao, J. A., Patksorsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA-protein complex. Nat. Struct. Mol. Biol. 15, 103–105 (2008).
29. Zhou, H.-X. Polymer models of protein stability, folding, and interactions. Biochemistry 43, 2141–2154 (2004).
30. Tan, S. Z. & Prasher, K. L. Dynamic pathway regulation: recent advances and methods of construction. Curr. Opin. Chem. Biol. 41, 28–35 (2017).
31. Ingram, L. O., Conway, T., Clark, D. P., Sewell, G. W. & Preston, J. F. Genetic engineering of ethanol production in Escherichia coli. Appl. Environ. Microbiol. 53, 2420–2425 (1987).
32. Alberthum, F. & Ingram, L. O. Efficient ethanol production from glucose, lactose, and xylose by recombinant Escherichia coli. Appl. Environ. Microbiol. 55, 1943–1948 (1989).
33. Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. Domain perturbation of gene expression alters bacterial fitness under stress and reveals underlying epistatic constraints. ACS Synth. Biol. 6, 94–107 (2017).
34. Peters, J. M. et al. Bacterial CRISPR: accomplishments and prospects. Curr. Opin. Microbiol. 27, 121–126 (2015).
35. Choi, K. R. & Lee, S. Y. CRISPR technologies for bacterial systems: Current achievements and future directions. Biotechnol. Adv. 34, 1180–1209 (2016).
36. Peng, R. et al. CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus. Microb. Cell. Fact. 17, 15 (2018).
Acknowledgements

We thank Maureen Thomason, Mary Lidstrom, Frances Chu, Willy Voje, Jason Stevens, Chuhern Hwang, and members of the Zalatan and Carothers groups for technical assistance, advice, and helpful discussions. The NM700 strain was a gift from Nadim Majdalani and Susan Gottesman. This work was supported by a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (J.G.Z.), an NSF Award MCB 1517052 (J.M.C.), and a University of Washington Presidential Innovation Award (J.M.C.)

Author contributions

J.G.Z. conceived the project. C.D., J.F., J.M.C., and J.G.Z. designed experiments, analyzed data, and wrote the manuscript. C.D., J.F., and A.P. performed experiments.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04901-6.

Competing interests: The authors declare no competing interests.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2018

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.