Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression
Alec AK Nielsen¹, Thomas H Segall-Shapiro¹ and Christopher A Voigt

Cells use regulatory networks to perform computational operations to respond to their environment. Reliably manipulating such networks would be valuable for many applications in biotechnology; for example, in having genes turn on only under a defined set of conditions or implementing dynamic or temporal control of expression. Still, building such synthetic regulatory circuits remains one of the most difficult challenges in genetic engineering and as a result they have not found widespread application. Here, we review recent advances that address the key challenges in the forward design of genetic circuits. First, we look at new design concepts, including the construction of layered digital and analog circuits, and new approaches to control circuit response functions. Second, we review recent work to apply part mining and computational design to expand the number of regulators that can be used together within one cell. Finally, we describe new approaches to obtain precise gene expression and to reduce context dependence that will accelerate circuit design by more reliably balancing regulators while reducing toxicity.

Addresses
Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding authors: Voigt, Christopher A (cavoigt@gmail.com)
¹ These authors contributed equally to this work.

Introduction
Cells naturally control gene expression using a variety of RNA, protein, and DNA-modifying regulators [1–3]. It was recognized early on that interactions between these regulators could lead to computational operations that are analogous to electronic circuits [4–6]. Genetic engineers have since attempted to build synthetic circuits that would implement artificial programs of gene expression. This could have a revolutionary impact on biotechnology, such as programming bacteria to individually respond to transient conditions in a bioreactor [7], designing therapeutic cells to sense and respond to diseased states within the human body [8–12], or engineering smart plants that can respond to conditions in the environment [13]. However, building synthetic circuits remains one of the greatest challenges in the field, where even simple circuitry is labor intensive to build and lacks the performance of its natural counterparts. As a result, synthetic genetic circuits have been slow to appear in biotechnology applications [14].

There are several reasons why genetic circuit design has been challenging compared to other areas in genetic engineering. First, circuits require precise tuning in the expression levels of their component regulators [15]. This is less essential when engineering cells to make small molecules or individual proteins, where genes tend to be maximally expressed. Second, regulators are prone to being toxic and, even when slight, this can inhibit growth and lead to evolutionary instability and a reduction in strain performance. Third, the regulatory interactions within a circuit all occur within the cell and crosstalk between them or with the host can impact the desired circuit behavior [16]. Fourth, there are few design rules for the systematic improvement of circuit performance (speed, dynamic range, robustness, and cell-to-cell variability). Finally, the physical construction of circuits requires the assembly of many parts, which until recently, has been technically challenging [17–19]. Often, these parts appear in genetic contexts that are different than that in which they were characterized and this can lead to interference [7].

In this review, we focus on recent advances in synthetic circuit design for bacteria. There have been other reviews looking at circuit design for eukaryotes and higher organisms [14,20,22–24]. We have divided the review into three sections. In the first section, we describe recent advances in different types of circuit design. Next, we describe how the toolbox of regulators has expanded, both in increasing the number of characterized regulators from different families, as well as the discovery of new biochemistries that can be harnessed for synthetic regulatory networks. Finally, we review new approaches to obtain precision expression control and its potential impact on building sophisticated circuitry.

I. Advanced circuit designs

To date, most of the genetic circuits that have been constructed are so small that there has been little need to utilize advanced concepts or algorithms in their design. As they get more sophisticated, however, it will become more difficult to identify a pattern of regulatory interactions that can produce a desired function. To this end,
approaches from electrical digital and analog circuit design have begun to be applied. Realizing these designs requires that regulators be functionally connected. This will require better control over their response functions (the input-output relationship), as well as handling other circuit characteristics such as retroactivity and evolutionary instability.

Adult digital circuits

Digital circuits produce signals at discrete levels (most commonly, high and low or 1 and 0), as opposed to operating in a continuous range. Their advantage is in their designability; there are many design tools that can abstract a desired circuit function into a large assembly of logic gates [25]. This designability comes at the cost of size and power requirements. Many digital gates may be needed to produce a computational function compared to what would be required if continuous variables were allowed. In terms of genetic circuits, this manifests as more DNA, regulators, and energetic resource expenditure [26–28].

Many genetic circuits have been built that produce Boolean logic functions, or ‘logic gates’ [29–31]. Note that while these are often described as being digital, all of these circuits exhibit analog features (so-called fuzzy logic), where there is a constant change in output. This can be used as the basis for the construction of analog circuitry (next section).

If genetic logic gates are designed to have inputs and outputs that have the same signal, they can be layered to produce more complex computational operations. In practice, this signal is transcriptional activity, where the inputs and outputs are promoters. This approach is modular but slow, with each layer requiring a step of transcription and translation with a timescale of 20 min [32]. Further, if one of the signals skips a layer, this can produce a fault where the output is transiently incorrect. Such faults have been exploited in the construction of pulse-generating genetic circuits, in the form of incoherent feed-forward loops [33–35].

There have been several studies to layer logic gates to produce more complex functions. This is closely related to work to build cascades through the connection of gates in a linear series [32,36,37]. As a proof-of-principle, a 4-input AND gate was built by layering three 2-input AND gates along with additional layers that contain the four sensors and an output [28] (Figure 1(a)). It has also been shown that a set of orthogonal NOR gates can be layered to form different logic operations by permuting the input and output promoters to produce different wiring diagrams [38]. Both of these examples perform relatively simple computations that could be designed by hand and both required more gates than would be needed using other types of regulators. It would require significantly larger circuits to realize the benefits of layered digital gates and computational design automation [25,39,40].

Analog circuits

Analog circuits operate with continuous signals. In electronic circuits, they are used when there are limitations in the number of components or power that can be used (e.g., in medical devices) [41]. This comes at a cost of designability, where each circuit has to be individually designed and simulated, which limits the size and flexibility of the circuits. In practice, every genetic circuit—natural or synthetic—is analogous to some degree and this needs to be accounted for in their design. The question is to what extent the design of genetic circuits can benefit from the principles used for building analog electronic circuits.

The value in considering analog circuit design was recently demonstrated [42]. The authors of the study were able to implement circuits that computed mathematical functions that would otherwise require many digital gates, including logarithm and power-law functions, and continuous addition and division (Figure 1(b)). Their circuit generates a wide dynamic range response function using a positive feedback loop and a second promoter to titrate away the activator. This response computes the logarithm of inducer concentration and the introduction of a second positive feedback loop produces a circuit that computes log-domain addition of two inducers. A log-domain division circuit was further engineered by having the two feedback loops compute the ratio between the two inducers. Remarkably, all of these arithmetic functions were computed using only two transcription factors.

Recombinase-based memory and logic

Logic gates based on transcription factors often exhibit analog features, with graded switch transitions. In contrast, highly digital switches can be built using recombinases that catalyze a sequence-specific change in the orientation of a unit of DNA, where each orientation corresponds to a different signal level. Recombinases have been used as the basis for a number of synthetic circuits [43–45] and have been layered to form a cascade [46]. Previously, the recombinases used were either irreversible (where the inversion is unidirectional) or reversible (where the same recombinase catalyzes both directions). Recently, a rewriteable switch has been built based on a system where an integrase catalyzes the switch in one direction and an integrase/excisionase pair catalyzes the reverse reaction [47]. This is a significant improvement in that it allows the signal to both hold permanently and also switch back to the initial state.

Multiple recombinases have been built into circuits that function as ‘memory logic’ devices, where the rearrangement of DNA is triggered by two input inducers [48,49].
Advanced genetic circuit designs. Recent progress in building more complex genetic circuits has been enabled through development of new part families and more sophisticated circuit architectures. (a) An intracellular 4-input AND gate built by layering three 2-input logic gates. Each 2-input logic gate works by expressing an activator from one input and a chaperone from the other input that complex and activate the output. The entire 4-input AND gate’s output is high only when all four inputs (arabinose, IPTG, AHL, and aTc) are present. Adapted from [28]. (b) An analog circuit that computes log-domain addition of arabinose and AHL concentrations was built using two wide dynamic input range feedback loops (AraC and LuxR) that express a common output. The dynamic input ranges for each feedback loop were extended by providing a ‘shunt’ plasmid with a promoter that titrates away the transcription factor from the feedback loop. Adapted from [42**]. (c) Recombinase memory circuits effect stable inversions of genetic regulatory elements. An irreversible 2-input AND memory circuit was built by expressing recombinases Bxb1 and phiC31 with AHL and aTc, such that two unidirectional terminators are flipped into a non-terminating orientation when both recombinases are expressed. This circuit stably maintains its output state for several days. Adapted from [48*]. (d) CRISPR-based gene regulation uses a catalytically inactive Cas9 protein (dCas9) and a targeting short guide RNA (sgRNA) to guide the sgRNA–dCas9 complex to specific DNA loci. By targeting various regions of a bacterial promoter, dCas9 was shown to repress transcription initiation using a highly programmable targeting mechanism. Adapted from [95**].

In one paper, two recombinases (Bxb1 and phiC31) irreversibly invert promoters, unidirectional terminators, and GFP. Only when transcription initiation, termination, and GFP are in the correct orientation is fluorescent output seen, and memory circuits for all irreversible 2-input logic functions were successfully constructed [48*] (Figure 1(c)). In a second paper, two recombinases (Bxb1 and TP901-1) invert or excise terminators and promoters to implement six irreversible 2-input logic functions [49].

There are several advantages to this approach in building logic switches. The gate response has a larger dynamic range that is easier to connect to downstream gates and the signal levels are more easily distinguishable than transcriptional gates (two different DNA orientations versus the presence or absence of a regulator). The hold state is also permanent, surviving over many generations and even after cell death. These gates could be layered as with those based on transcription factors and the size of
the DNA per gate is comparable. However, there are two disadvantages with using recombinases. First, they are not true logic in that a transient but temporally separate induction of the two signals leads to a permanent change in the output. Also, they tend to be slow, with each layer requiring up to eight hours [43] to complete.

Control of the response function
Building genetic circuits requires that the parts’ response functions match—that is, the output range of upstream parts maps onto the correct input range of downstream parts. In the ‘Precision gene expression’ section, we look at methods based on changing the expression levels of the regulators. Here, we discuss new approaches to change the shape of part response functions, including the basal level of the off state, dynamic range, and cooperativity [15]. These approaches could be applied to digital or analog circuits and to parts from different regulator families.

One of the biggest practical problems in constructing complex circuits is that the basal level of activity from a part’s off-state (‘leak’) is often sufficient to trigger the next layer of a circuit. This has been difficult to control, but there are several promising new approaches. First, leakage can be minimized using riboregulation to suppress translation. This has been used effectively to minimize the uninduced expression of toxic proteins [50]. A similar approach is to use small RNAs (sRNAs) to bind an RNA chaperone Hfq and the target mRNA to both inhibit translation from the mRNA and also target it for destruction by RNase E [51]. Second, in natural prokaryotic genetics, leakiness is controlled using 5’-terminators that come directly after a promoter to attenuate transcription [52], and a similar strategy could be used in synthetic systems to reduce leaky transcription of mRNA.

Increasing the nonlinearity of a response curve can also be important in circuit design. Nonlinearity occurs naturally via cooperativity, but this can be challenging to engineer de novo. An easier approach is to incorporate interactions that sequester the regulator at the DNA [53], RNA, or protein level [54]. The level of effector must exceed the binding capacity of the competitively sequestering partner to pass the threshold, bind the cognate partner, and take action. Examples of competitive binding partners include anti-sigma sequestering sigma-factors [55, 68], sRNA sequestering mRNA [56], and decoy operators sequestering DNA-binding proteins [53].

Buffering retroactivity
Genetic gates within a cell inevitably share resources; for example, they utilize the host RNA polymerase and ribosomes. Shared resources can cause coupling between gates that are otherwise unconnected and can cause a downstream gate to affect the behavior of an upstream one [57, 58]. It has been experimentally shown that increasing the number of downstream operators can impact the response time and threshold of a gate—an effect termed ‘retroactivity’ [58]. Since the size of genetic circuits has been small, this has not yet emerged as a significant problem, but retroactivity is expected to worsen as the circuit size increases and when the output of a gate is connected to many downstream circuits or actuators (‘fan-out’).

An interesting approach for insulating modules from retroactivity has been mathematically investigated [57]. While the general solution is well known from control theory (high input amplification and high negative feedback), the authors of this study looked at specific biological mechanisms that could fill the role. Surprisingly, simple circuit modifications like a non-leaky input promoter (for high input amplification) and rapid degradation of the transcription factor (for negative feedback) effectively insulate retroactivity in mathematical models.

Engineering evolutionary stability
The selection of a particular circuit topology and genetic implementation impacts its evolutionary stability. Recent work has begun to illuminate the design choices that lead to instability, which can in turn be used to guide future designs. It has been observed that if the resting state of a gate requires the expression of an active regulator, the gate is more evolutionarily unstable [59]. Thus, selecting an architecture that minimizes the number of regulators that have to be expressed at a given time could increase stability [60]. The modes by which cells reject circuits are also becoming more well understood. Unsurprisingly, the use of plasmids and the repetition of DNA sequences in a design leads to instability [7, 60, 61]. Re-using strong double terminators within a genetic construct has been found to be particularly unstable and terminator diversification dramatically increases the number of generations before a circuit is lost due to homologous recombination [61–62].

Additionally, a recent large-scale effort to ascertain which heterologous genes are toxic in Escherichia coli has begun to shed light on how synthetic constructs affect host fitness [63]. By examining cloning gaps in over 9.3 million sequencing clones from 393 microbial genomes, more than 15 000 genes were found to have toxic expression products. Through subsequent validation and analysis, new restriction enzymes, toxin–antitoxin systems, and toxic small RNAs were discovered. The authors even observed the existence of toxic DNA-binding motifs that likely titrate away DnaA and inhibit normal replication. This wealth of information about toxic DNA elements is a fantastic resource for predicting how genes may affect host systems, and could guide which regulators can be effectively mined for use in circuits.
II. Classes of regulators

An explosion in the number of well-characterized regulators available for building genetic circuits has occurred over the past few years. Before this, there were relatively few regulators available (e.g., LacI, TetR, AraC, and CI) and these were re-used in many designs. A major goal has been to expand the number of regulatory parts that are orthogonal, that is, do not cross react with each other such that they can be used together in a circuit [64,65]. This has been achieved via two approaches. First, bioinformatics and whole gene DNA synthesis have been used to access regulators from sequence databases (‘part mining’) [66]. Second, families of regulators have been characterized that are conducive to the rational design of orthogonal sets (e.g. RNA antisense regulation, zinc finger proteins, transcription activator like effectors, and CRISPR-Cas9). Computational methods have played a role both to predict the orthogonality of regulators identified in databases as well as in structure-guided design. Collectively, this has resulted in well over one hundred regulators that could theoretically be used together in a single large circuit in a bacterium.

Despite efforts to standardize and collect data surrounding biological parts [59,67], the majority of the information on different regulators is buried in individual papers, making it difficult to choose parts for desired circuit designs. In Table 1, we show data comparing 16 characterized regulator families currently available for building genetic circuits. For each family, the number of characterized orthogonal regulators is shown (with a metric of crosstalk), along with the dynamic range that has been achieved with that regulator. Note that the table focuses on bacteria and there is more data for some families in eukaryotic cells.

Protein regulators of transcription

Proteins that directly bind DNA to regulate transcription make up the majority of the regulatory parts available for use in bacteria (Figure 2(a)). One way in which proteins can regulate transcription is by initiating transcription at promoters. The native E. coli RNA polymerase (RNAP) can be directed to new promoters by expressing sigma factors from other organisms [55,68]. A large set of orthogonal sigma factors has been generated through part mining, in which sigma factors from many organisms were synthesized and their activities characterized [68]. Alternatively, the phage RNAP from T7 is often used and the promoter specificity of this polymerase has been changed to generate new parts through rational design and part mining [69–71] as well as random mutagenesis [72,73**].

Activators upregulate transcription by binding to a promoter to recruit RNAP. Classically, there are a number of natural activator proteins that have been used in genetic engineering, such as λ CI and LuxR. A small library of CRP activators was built by using bioinformatics to direct mutations at residues responsible for operator specificity [74]. Part mining has been applied to identify activators that require a second chaperone protein for activity and this has been used as the basis for building AND gates [28].

Repressors block transcription by blocking the binding or progression of RNAP. Recently, there have been efforts to increase the number of orthogonal repressors available for circuit design. To expand the LacI family, mutations were made to specific DNA residues in the binding site and DNA binding residues in the protein and a set of 5 orthogonal repressors was selected [75]. Part mining has been applied to expand the number of available TetR homologues and this led to the identification of an orthogonal set of 16 repressors [38].

There are several classes of transcription factors that have modular protein structures that facilitate their engineering to target particular DNA sequences. Zinc finger proteins (ZFPs) and transcription activator like effectors (TALEs) have such a structure and have been particularly successful in being used in eukaryotic cells [76–80,81**]. It has been surprisingly difficult to get these regulators to work in bacteria, but there are published examples of ZFPs being used as repressors and activators [82] and a TALE as a repressor in E. coli [83*].

RNA regulators of translation

RNA parts that regulate translation take advantage of the fact that RNA base pairing follows a simple code that is computationally predictable [84] (Figure 2(b)). Two parts families of this type utilize antisense sRNA binding to alter the accessibility of the ribosome binding site (RBS) controlling translation initiation. RNA-IN/OUT parts consist of a modified natural system [85] in which an RNA molecule base pairs to the 5 ’-end of an mRNA (including the RBS) such that the ribosome cannot initiate translation [86**]. The orthogonal set of these regulators was increased through computational design and experimentally confirmed. A second part family uses trans-activating RNAs that work by disrupting a secondary structure that blocks the RBS by default, leading to translational activation [87,88*]. Finally, it has been shown that the expression of modified 16S RNA that has been engineered to bind a non-canonical Shine Delgarno sequence can recruit ribosomes and this has been used as the basis to build gates [89–91].

When designing gates, the challenge with using RNA that acts on the level of translation is that it is difficult to convert an RNA input to an RNA output of the same form. Therefore, the resulting gates cannot be layered. An approach to this problem is to use a cis element that converts a translational signal into a transcriptional one [92*]. This component utilizes a modified sequence from
Table 1

<table>
<thead>
<tr>
<th>Characteristics of regulatory part families currently available for constructing genetic circuits in bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Characterized</td>
</tr>
<tr>
<td>family size</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DNA modification</td>
</tr>
<tr>
<td>Recombinases</td>
</tr>
<tr>
<td>Nonlinear,</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>T7 polymerases</td>
</tr>
<tr>
<td>TnA family</td>
</tr>
<tr>
<td>repressors</td>
</tr>
<tr>
<td>LacI repressors</td>
</tr>
<tr>
<td>Chaperone</td>
</tr>
<tr>
<td>activators</td>
</tr>
<tr>
<td>ECF sigma factors</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A set of anti-sigma factors enables sequestration</td>
</tr>
<tr>
<td>TALE repressors</td>
</tr>
<tr>
<td>Zinc finger</td>
</tr>
<tr>
<td>activators</td>
</tr>
<tr>
<td>Zinc finger</td>
</tr>
<tr>
<td>repressors</td>
</tr>
<tr>
<td>Converted RNA-IN/OUT</td>
</tr>
<tr>
<td>PT181 attenuators</td>
</tr>
<tr>
<td>Attenuators can be repeated to increase signal</td>
</tr>
<tr>
<td>dCas9 repression</td>
</tr>
<tr>
<td>Requires dCas9 expression</td>
</tr>
<tr>
<td>dCas9-Cas activation</td>
</tr>
<tr>
<td>Requires dCas9-Cas expression</td>
</tr>
<tr>
<td>Translation regulation</td>
</tr>
<tr>
<td>Additional large sets of RNAs computationally designed</td>
</tr>
<tr>
<td>Activating riboregulators</td>
</tr>
<tr>
<td>O-ribosymes</td>
</tr>
</tbody>
</table>

This table presents a brief comparison of a number of regulatory part families that have been characterized and are available for use in building genetic circuits in bacteria. Each part is defined as a trans-acting element and the target of this element: for example a transcriptional repressor and its binding site. Part families were chosen to either have at least three characterized members or be based on a technology proven to be extendable (TALE, zinc finger, and CRISPR-based parts). The characteristics were derived from the cited sources as follows:

Characterized family size indicates the number of parts of this type that have been characterized for crosstalk. A ‘+’ indicates that this parts family is based off of a technology proven to enable orthogonal, programmable DNA binding and so the parts set may be predictably extendable.

Maximum dynamic range is the largest reported fold change between the on and off states (i.e. with the trans-acting element present and absent) of a single member of the part family.

The largest tested parts sets show the largest number of parts in each family that have been shown to function above specific thresholds of orthogonal range. Orthogonal range is a conservative measurement of the orthogonality of a parts set; it represents the fold change between the on-target effect of a part and the worst off-target effect on that part. For example, the ‘T7 polymerases’ family has a 3 part set that functions above 10x orthogonal range, meaning that there is a group of three polymerases where each activates its target promoter to a level more than 10x the level that either of the other two polymerases activate it.

Notes:

a Numerical data was used from this reference for the part family in this row.

b Data computed from bar or line plots in this reference was used in this row.

d Data was read from colored orthogonality grids in this reference. Note that the numbers in this row may be less accurate because of uncertainties in this method.

d While only two recombinases have been tested together at a time, at least seven have been used in genetic circuits in bacterial systems [44–49].

The regulation of transcription is a critical process in biological systems. The regulation of transcription can be achieved through various mechanisms, including the use of RNA-based regulatory elements. The use of RNA-based regulation allows for a more dynamic and fine-tuned control over gene expression. RNA regulons of transcription are a type of regulatory mechanism that utilizes RNA molecules to control gene expression. These systems can be used to design novel genetic circuits that can be programmed to respond to specific stimuli. One example of such a system is the use of the CRISPR-Cas9 system, which can be engineered to target specific DNA sequences and induce transcriptional silencing. Another example is the use of RNA-based activators, which can be designed to bind to specific RNA sequences and activate transcription. These systems can be used to design novel genetic circuits that can be programmed to respond to specific stimuli.
The diversity of genetic regulatory parts available for building bacterial genetic circuits. Schematics and representative data for a selection of different circuit-building mechanisms are shown. The colored regions in the schematic indicate the variable regions that make up each part. A solid line surrounding a part type indicates that there is a well-characterized, orthogonal set of parts of this type available for circuit building. A dashed line indicates that there is a proof-of-concept part. The data shown either demonstrates the orthogonality and size of a parts family, or if that is not available, it shows proof-of-concept activity. (a) Protein parts that act on transcription include natural repressors and activators, phage polymerases, sigma factors, and repressors and activators based on programmable DNA binding proteins. Figures adapted from: Natural repressors [38,75], Phage polymerases [69], Natural activators [28], Sigma factors [68], Programmable repressors [83], Programmable activators [82]. (b) RNA parts that act on translation include orthogonal ribosomes, antisense sRNA repressors, and activating riboregulators. Figures adapted from: Orthogonal ribosomes [89], antisense sRNA repressors [86**], activating riboregulators [88**]. (c) RNA parts that act on transcription include sRNA regulators converted to affect transcription, antisense RNA transcriptional attenuators, dCas9 repression, and dCas9 activation. Figures adapted from: Converted sRNA regulators [92**], RNA transcription attenuators [99**], dCas9 repression [95**], dCas9 activation [96**]. (d) Natural recombinases have been used to modify DNA and implement logic. Figure adapted from [48**].

A protein that uses a small guide RNA to target a DNA sequence as part of the CRISPR/Cas bacterial immunity system [94]. Normally, Cas9 functions as a nuclease and cleaves DNA, but it was shown that if the nuclease activity is mutated, then the complex blocks RNAP (Figure 1(d)). This can be used either as a repressor or as an activator by fusing an activation domain to Cas9 [95**,96**]. An advantage of this system is the potential to design vast numbers of orthogonal regulators by building guide RNAs that target different ‘operator’ sequences. The cross reactions that may result from this approach are just beginning to be characterized and understood [97,98]. Once it is shown that the Cas9-based systems can be layered, this will become a powerful toolbox for
circuit engineering. However, a practical challenge with using this system is the acute toxicity of Cas9 when expressed in many organisms.

A more developed approach for RNA to control transcription is based on the PT181 attenuation system [99**,100,101]. When an antisense RNA is present and binds a target sequence on a transcript RNA, the nascent transcript folds into a transcriptional terminator and attenuates the message. This part family was expanded both through part mining and random mutagenesis, including utilizing some of the orthogonal RNA pairs from the RNA-IN/OUT system. These attenuators have been shown to be fully composable into cascades and logic gates [100].

Proteins that modify DNA

Thus far, DNA modification has been implemented in bacteria by using recombinases to invert or excise segments of DNA (Figure 2(d)). This mechanism is commonly found in bacteria, phages, and mobile genetic elements, providing a diversity of natural parts to exploit in synthetic systems [102].

The current parts for engineering DNA flipping and excision are all natural recombinases, which vary in a number of ways. The most commonly used recombinases in genetic engineering are the simple tyrosine recombinases Cre and Flp [103]. Tyrosine recombinases can flip the region between their recognition sites in both directions, leading to an even distribution of orientations. Additionally, if their recognition sites are oriented in the same direction, they catalyze DNA excision. These two recombinases have been shown to function properly and orthogonally in bacteria [46]. The invertases FimB and Hin have also been used together in a bacterial system [44]. These invertases are similarly bidirectional, but they lack the capability for excision. Finally, phage integrases are a class of recombinases that catalyze unidirectional flipping, which leads to the accumulation of a specific DNA orientation [104]. Three integrases (Bxb1, C31, and TP091-1) have been characterized and used to build bacterial circuits [48*,49]. Additionally, a number of these integrases have matching excisionases, which allow for the reversal of DNA flipping [47**,104].

In addition to part mining natural recombinases, there has been progress in generating new families of recombinase parts through modifications. One such approach is to iteratively make mutations to the DNA binding region in the recombinase and select for proteins with new specificities [105,106]. Another promising method is to generate new parts by creating recombinase fusions with zinc finger and TALE DNA binding domains [107,108].

Finally, while current efforts have focused on recombinases, other mechanisms of DNA modification may also hold promise for building circuitry. Recently, there have been successful efforts to selectively methylate DNA in bacteria [109*] using a modular zinc finger design. An in vivo means of reading methylation state would open up these parts for use in genetic circuits. This seems feasible, as bacteria are known to contain many regulatory systems that respond to DNA methylation [110,111].

### III. Precision gene expression

Building, tuning, and connecting genetic circuits requires the ability to engineer precise changes in gene expression. Further, when gates are combined to build a circuit their genetic context changes, potentially impacting function [112]. There have been many recent advances in the development of ‘tuning knobs’ that allow for the fine-tuned control of transcription and translation (Figure 3). These can take the form of part libraries or computational tools. Further, insulators have been developed that reduce interference between parts. This is leading to a redefinition of the classical expression cassette.

**Tuning knobs for expression**

**Promoters.** Libraries of constitutive promoters for different species have been built by mutating the −10 and −35 RNAP binding regions of the promoter or the region affecting DNA melting (−10 to +2) [113**,114**,115–118]. Advances in oligonucleotide synthesis have enabled these libraries to become very large. For example, >10 000 combinations of promoters and 5′-UTR’s were built from a pooled oligonucleotide library and screened by combining cell sorting and deep sequencing (flow-seq) [113**]. Computational models of promoters have also been developed that are based on the free energy of RNAP binding to the −10/−35 sites and promoter melting [119*,120*]. These show promise in predicting promoter strength, however, a complete model that balances all contributions has yet to be built.

**Ribosome binding sites.** The RBS is a part that is relatively simple to tune to achieve different expression levels. As a result, it has been broadly applied to tuning response functions for building genetic circuits [28,35,38,121,122,123*]. The ribosome makes contacts with the Shine–Dalgarno sequence and start codon and binding is influenced by RNA base-pairing, the spacing between these regions, and the mRNA secondary structure. The RBS Calculator is a computational tool based on a biophysical model that balances these contributions [124,125]. There are additional terms that influence the strength of the RBS and several of these, including the role of the standby site, have been characterized and incorporated into new versions of the software [125,126]. There is much to be learned from non-canonical RBSs [127] including leaderless RNAs [128] and a better understanding of these processes could improve biophysical models. Further, libraries of 5′-UTRs that
A modern expression cassette comprising genetic tuning knobs and insulators. (a) A cassette is shown for precise expression of a gene of interest (GOI). Insulating parts are highlighted in blue. The cassette is shown using symbols from the synthetic biology open language visual (SBOLv). (b) Biophysical models of transcription based on the RNAS binding energy have been constructed to complement empirically characterized promoter libraries. Figure adapted from [120]. (c) The RBS Calculator provides a computational framework for designing RBS sequences of a given strength based on a thermodynamic model of translation initiation. Figure adapted from [124]. (d) Terminator strength is partially informed from a biophysical model relating various aspects of the terminator sequence. Figure adapted from [62]. (e) Tunable copy-number plasmids allow for a wide range of gene expression based on the repA/CoIE2 and pir/RPl6 plasmid systems. Figure adapted from [130]. (f) Promoter-insulating sequences have been shown to reduce the change in promoter activity when upstream and downstream sequences are introduced to the flanking promoter context. Figure adapted from [132]. (g) Ribozymes have been used to improve the predictability of gene expression by reducing promoter/5'-UTR coupling effects. Figure adapted from [112]. (h) Bicistronic RBS designs cause the rank order of expression constructs to be more predictable compared to single RBSs. Figure adapted from [133]. (i) RNase III sites in the 3'-UTR reduce the variability in gene expression for reporters coupled with libraries of terminators. Figure adapted from [129].
include the RBS have been measured [113**,114***], and a recent technique to tune RBS strength using hypermutable sequence repeats between the Shine-Dalgarno region and start codon was used to explore expression parameters for a bistable switch [123*].

**Terminators.** During transcription, mRNA is released when RNAP reaches a terminator. Terminators are important in circuit design for three reasons. First, they offer a means to tune expression by modulating read-through and could potentially decrease leaky expression. Second, many terminators are required when gates are combined to build circuits. Circuits can have many transcription units, each of which needs strong termination to avoid interference with other circuit elements. These terminators must be sequence diverse to avoid recombination. Third, the recombinase-based memory circuits utilize unidirectional terminators as a core part of their design. To address these needs, there have been major efforts to use part mining to glean large libraries of terminators from bacterial genomes [62,129].

**Origins.** To finely control plasmid replication beyond the standard plasmid systems, tunable-copy-number plasmids can be used. Increasing the expression of transacting replication factors (repA and pir) increases the plasmid copy number (for ColE2 and R6K origins, respectively). Using this system, a library of ‘DIAL’ strains that constitutively express replication factors from the genome has been constructed that can yield between 1 and 250 plasmid copies by transforming the corresponding plasmids into the different strains [130**].

**Insulators to buffer the impact of genetic context**
Part function is impacted by genetic context, or the sequences of the neighboring parts [112,113**,131, 132*,133**]. In turn, this can impact the response function of the entire genetic circuit. Neighboring genetic context effects can have two forms. First, there is a direct interference of one part type on another. For example, the strength of an RBS is influenced by the promoter and the first codons of the expressed gene. Second, when two parts are combined a new function can appear at their interface. For example, promoters have been inadvertently constructed through the assembly of two parts containing an LVA-degradation tag, a DNA barcode, and a BioBrick scar [134*]. To overcome this issue, insulator parts have been developed to diminish the effect of genetic context. Figure 3 shows a conceptual re-visit of the expression cassette, where insulators are strategically placed between key parts.

The first insulators in the diagram are based on bidirectional terminators, which flank the expression cassette to reduce transcriptional read-through into or out of the expression cassette [62]. The next type of insulation addresses the issue that promoters used in synthetic biology are typically too small and only capture the −35 and −10 regions. This can cause the promoter to have different strengths depending on the up and downstream sequence. Longer promoters should be used that at least encompass the UP element (−35 to −64) that binds the α-subunit of RNAP [119*,135]. Taking this further, it has been shown that the addition of upstream insulator sequences (up to −105) and downstream insulator sequences (down to +55) further standardizes promoter activity [132*].

The transcription start site of promoters is not always known and a ‘promoter part’ is rarely annotated to end at the +1 position. This can be further complicated by the observation that there is sometimes a distribution of mRNA produced from the same promoter and single mutations to the promoter can change the start site. Additionally, the RBS is particularly sensitive to changes in the 5′-UTR, even by a single nucleotide. Both self-cleaving ribozymes [112] and CRISPR RNA-processing [131] have been used to physically cut and detach the variable 5′-UTR from the mRNA, thereby shortening and making constant the 5′-mRNA context. These tools are particularly important when combining gates to build a transcriptional circuit, where the promoter inputs and output of each gate occur in new contexts.

A bicistronic RBS sequence has been shown to reduce the impact of the protein coding sequence on the 5′-UTR secondary structure including the RBS [133**]. A small ‘leader’ peptide-coding sequence with its own RBS is positioned upstream from the gene of interest such that the peptide sequence overlaps the RBS for the downstream gene and the peptide’s stop codon ends at the gene of interest’s start codon. The natural helicase activity of ribosomes loaded at the first RBS unfolds the mRNA secondary structure near the second RBS that controls gene expression, decoupling the translation initiation rate of the second RBS from the downstream coding sequence.

Lastly, the variability in gene expression when coupled with various terminators can be reduced by encoding an RNase III site in the 3′-UTR of the mRNA [136,129]. Post-transcriptional processing of mRNAs by RNase III standardizes the 3′-end of the mRNA, such that the sequence and secondary structure of the cleaved RNA can no longer contribute to mRNA stability and degradation.

**Conclusion**
Genetic circuit design is at an inflection point regarding the size and sophistication of computational operations that can be implemented within living cells. The first phase of the field involved the *ad hoc* construction of individual circuit functions (e.g., an oscillator or a gate) by piecing together the necessary biochemistries. For an
extended period, the complexity of these circuits remained relatively flat as the rules for composing a circuit were explored [14]. After this phase, there have been painstaking efforts to expand the number of available regulators. During this time, there has been highly technical work to better understand how to control and insulate expression by revisiting old paradigms, like the expression cassette. Collectively, these efforts have yielded large sets of regulators that have been thoroughly characterized and for which the rules of assembly are better understood. This is expected to lead to a period, over the next few years, where there is sudden scale-up in the complexity of circuits. Ultimately, this will lead to circuit design as a regular component of genetic engineering projects, along with protein, pathway, and strain engineering.

Acknowledgements
CAV, AAKN, and THSS are supported by Life Technologies and the US National Science Foundation Synthetic Biology Engineering Research Center (SynBERC). AAKN and THSS are supported by a National Defense Science and Engineering Graduate Fellowship, and THSS is supported by a Hertz Foundation Fellowship.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


This work introduces Eugene, a programming language specifically designed for synthetic biology. Eugene has the capability to precisely specify different parts and their properties, aid in their assembly into devices using rule-based constraints, and interface with other tools for simulation or DNA assembly.


Through clever circuit design, this research exploits the analog nature of biochemical reactions to compute arithmetic and logarithmic functions in an elegant fashion.


A single digital memory unit is created by using an integrase and an excision from bacteriophage to selectively flip a region of DNA. The memory unit is shown to be able to be switched back and forth without degrading, and can hold its state for over 100 cell generations.

In this paper, recombines were used to conditionally invert genetic regulatory elements to implement every 2-input digital logic function. While the states cannot be reversed, the output signal shows no sign of decay after several days.

Trans-activating RNA that bind cis-repressing RNA was used for engineering many aspects of microbial protein production. The authors achieved tight control of gene expression to tune expression, effected fast response times, and reduced gene expression leakage.


Tandem repeats of a DNA operator were used in this study as ‘decays’ to sequester transcription factor activator and increase the steepness of the activator’s response. This raises an interesting design strategy for increasing the non-linearity of parts for use in digital logic.

Mathematical models of protein-sequestration were used in this study to demonstrate that molecular titration of effectors into an inactive state effects an ultrasensitive response. This type of sequestration can be used to increase the ultrasensitivity of a transcriptional factor’s response curve.

Sequestration of sigma factors using anti-sigma was used to increase the ultrasensitivity of a sigma factor positive feedback loop. The result is a bistable latch with tunable switching boundaries.


In this landmark paper, biochemical solutions to ameliorate the issue of ‘retroactivity’ were mathematically investigated. The two strategies described were: first, transcriptional amplification with protein degradation negative feedback, and second, phosphorylation amplification with dephosphorylation negative feedback.

The impact of retroactivity on gene circuits is further examined, with a focus on how it alters the dynamics of systems. Single cell and population measurements are used to confirm the modeled effects.


This study measured the evolutionary stability of simple genetic circuits and determined the genetic basis for circuit failure. It was found that repeating certain sequences (e.g., terminators used in multiple locations) exhibited high rates of homologous recombination, accelerating the rate of evolutionary escape. By diversifying the sequences and reducing expression, longer evolutionary half-lives were observed.


By analyzing a large library of genome fragments that failed to clone properly in E. coli, the authors identify over 15,000 toxic gene products. Further experiments are carried out to identify the functions of a number of these genes.


Phage assisted continuous evolution (PACE) is a clever technique to evolve genes using the fitness of a bacteriophage life cycle in a lagoon of bacteria. Using PACE, T7 RNA polymerases were evolved that recognize specific promoters over the course of only a few days.


In this study, the authors use an algorithm to design a large set (180) of TALE binding sites that are predicted to be orthogonal to each other and the human genome. In vivo testing of eight of the designed sequences shows that they are active and orthotoghol.


In this paper, a TALE scaffold was shown to outperform the widely used LacI repressor in blocking transcription. Because TALEs can be modularly engineered to bind specific sequences, they could provide an enormous library of orthogonal bacterial transcription factors.


By analyzing the activity over 500 RNA-IN-RNA-OUT antisense-regulator pair mutants, the authors predict the potential performance of thousands of additional antisense-regulator pairs for use in genetic circuits.


This paper describes the rational design, characterization, and use of an expanded set of trans-activating riborepressors to control metabolic flux.


In this work, the architecture from the E. coli tna operon is used to create control elements where the translation rate of a downstream gene is related to the translation rate of an upstream ‘leader peptide’. This element is used with a number of translation regulators and in some cases leads to a dramatic increase in the dynamic range and orthogonal-ity of those parts.


This paper uses the nuclease-deficient CRISPR-Cas system to reversibly repress and terminate transcription, followed by thorough analysis of off-target effects.


Building on the original dCas9-repression research, this group further characterizes the CRISPR system’s ability to repress. They also extend the system to achieve activation of bacterial promoters.


To generate RNA effectors that can regulate transcription, a number of chimeric fusions of the P181 transcriptional attenuator and antisense RNA transcriptional regulators are constructed and tested. Eleven new attenuators are produced, with a set of seven that are mutually orthogonal.
Advances in genetic circuit design


This work presents the first programmable, site-specific methylation system that has been built for use in *E. coli*. It is shown to be capable of achieving well over 50% methylation at specific sites on plasmid DNA in vivo.


Even very simple parts can yield complicated behavior when connected. In a tour de force effort, this was demonstrated by measuring all combinations of 114 *E. coli* constitutive promoters and 111 RBSs of known strength. While these are the simplest parts available, only 64% of the combined parts function as expected based on their individual strengths.


An important aspect of genetic engineering is quantification of part activity and reliability. This paper from the BIOFAB uses a simple ANOVA framework to analyze combinations of promoters, 5’-UTRs, and reporters.


This paper presents the first comprehensive computational model for promoters in bacteria. A model is built for the promoters in *E. coli*, but the methodology could be adapted to other types of promoters and other species.


A thermodynamic model is presented for the binding of *E. coli* n70 RNAP to the –10 and –35 regions of a promoter. The model is able to accurately predict how mutations impact promoter strength.


This study showed that simple sequence repeats (SSRs) between the Shine-Dalgarno region and the start codon of an mRNA are hypermutable and modulate the strength of the adjacent RBS. These SSRs were used to explore the expression parameter space of a bistable switch.


The strains described in this work constitute express genomically integrated trans-acting replication factors that regulate plasmid copy number. By developing strains that express different levels of the replication factors, the authors produced a set of strains to rapidly optimize gene expression on a corresponding plasmid.


This research demonstrates the genetic context-dependence of bacterial promoters on both upstream and downstream sequence. The authors provide a series of insulated promoters to be used in synthetic biology designs.


This work introduces bicistronic RBS design as an engineering concept. The authors demonstrate that the inclusion of a second upstream RBS and leader peptide in the 5'-UTR reduces the context dependence of the downstream RBS, thus allowing for more precise control of expression levels.


Composing two BioBrick sequences that include an LVA-degradation tag and a DNA-barcode resulted in the emergence of a constitutive promoter between the parts. This work highlights the potential issues with parts assembly in the absence of sequences.


Zinc finger activators and repressors are characterized and used to construct OR, NOR, AND, and NAND gates in mammalian cells. For the AND and NAND gates, a split intein is used to reconstitute zinc finger function from two separately expressed protein fragments.