Biomolecule, Pathway, and Circuit Engineering

(Biomolecular Engineering)

5885 Hollis Street, 4th Floor, Emeryville, CA 94608
Phone: +1.510.871.3272 Fax: +1.510.245.2223

This material is based upon work supported by the National Science Foundation under Grant No. 1818248.

Biomolecule, Pathway, and Circuit Engineering

Summary

Biomolecule, Pathway, and Circuit Engineering focuses on the importance, challenges, and goals of engineering individual biomolecules themselves to have expanded or new functions. Successful progress would be demonstrated by production of functional macromolecules on-demand from both natural and non-natural building blocks, targeted design of complex circuits and pathways, and control over the dynamics of regulatory systems.

Introduction and Impact

At the molecular level, the functional richness, complexity, and diversity of biology can be localized predominantly to large "macro"-molecules (nucleic acids and proteins) and secondary metabolites. Indeed, evolution has produced and leveraged biomolecules and their assemblies to achieve extraordinarily sophisticated natural functions far surpassing our current engineering capabilities. If researchers are able to efficiently design, generate, synthesize, assemble, and regulate biomolecules in ways that rival the functional complexity of natural counterparts, but with user-defined functions, then all areas of bioengineering and synthetic biology should benefit.

The challenge of crafting biomolecules, pathways, and circuits that carry out user-defined functions has historically been an exercise in building out from what exists in nature to what doesn't. Certainly, this mode of bioengineering will be important going forward and will see transformations as our knowledge of and ability to harvest what exists in nature increases. Likewise, this mode of bioengineering will advance as our ability to take natural components and bring them to new functions improves, both in the ambitiousness of the functions we can reach (that is, how different they are from natural functions) and the scale with which we can reach them. Under this framework, we outline a number of transformative tools, technologies, and goals centered on parts prospecting, high-throughput measurement, and computational and evolutionary design approaches, to both better understand how natural parts work and rapidly improve upon them to reach user-defined functions. We should also keep in mind that as synthetic biology advances, what exists in nature may no longer be the only framework from which we can extract starting points for building out. Indeed, fundamentally new biological components of our own creation, for example ones containing fully unnatural chemical building blocks, might introduce entirely new categories of what exists to biology and so we must develop tools to use and design from those categories. Therefore, we also define a number of transformative tools, technologies, and goals that will allow us to exploit these truly new categories of biological matter.

The roadmap for *Biomolecule, Pathway, and Circuit Engineering* addresses the engineering of individual biomolecules to have expanded or new functions and the combination of biomolecular parts into macromolecular assemblies, pathways, and circuits that carry out a larger function, both *in vivo*, in cell culture systems, and *in vitro*, in cell-free and/or purified settings. The roadmap operates from the definition that 1) biomolecules are ***made by*** natural or engineered biological systems; 2) biomolecules are ***made from*** natural simple building blocks or engineered variants of those building blocks; and 3) the production of biomolecules can predominantly be genetically encoded. The roadmap uses the broad definition that macromolecular assemblies operate as complexes of physically-interacting individual biomolecules, that pathways are

combinations of biomolecules that achieve a coordinated function, and that circuits are combinations of biomolecules that achieve regulatory control or dynamic information processing. Under these definitions, typical biomolecules include natural and engineered variants of existing macromolecules (e.g., DNA, RNA, proteins, lipids, and carbohydrates), as well as new biopolymers containing unnatural nucleotides and amino acids; typical macromolecular assemblies include self-assembling protein nanostructures or nucleoprotein complexes; typical pathways include collections of natural or engineered enzymes that produce desired secondary metabolites; and typical circuits include natural or engineered regulatory modules that control gene expression in a dynamical fashion. We note that the boundaries between this section and host engineering (see **Host and Consortia Engineering**) can easily blur, but offer the practical and subjective classification guideline that this section treats bioengineering problems where the key innovations can be localized to manipulating and understanding individual molecules and their assemblies in contrast to manipulating and understanding the dynamics of large networks of molecules.

Transformative Tools and Technologies

Computational macromolecular design

Computational design of biomolecules with specific functions is a major area of research in synthetic biology. Advances in this area should eventually result in the on-demand generation of any specific molecular function, including catalysis and intermolecular interactions at the heart of biomolecular, pathway, and circuit engineering. Within computational design, protein, DNA, and RNA engineering have advanced the furthest, so we discuss computational design challenges through the lens of these particular macromolecules with the understanding that similar advances can be made for all biomolecules.

Protein design

Computational protein design is a discipline aimed at identifying specific sequences that have desired three-dimensional shapes or function, ideally exploiting the speed and cheapness of *in silico* computation to do so. Contrary to experimental methods such as directed evolution (laboratory evolution), computational biomolecular design aspires to “virtually” identify likely functional, while aiming to eliminate non-functional, molecules without producing and directly testing them. Computational biomolecular design has advanced to the point where defined structures and binding interactions can be constructed, but improvements are needed in expanding 1) the range and effectiveness of protein functions that can be designed, and 2) the success rate.

A critical aspect of designing functional proteins is the ability to accurately predict structure from sequences, which remains especially challenging for large proteins (>125 amino acids), beta-sheet topology, long-range contacts, and membrane proteins. Closely homologous proteins in nature have a backbone rmsd (root-mean-square deviation) <3 angstroms (Reva, Finkelstein, & Skolnick, 1998), so an rmsd of <3 angstroms between a computationally predicted structure (whether folding an existing sequence or designing a new sequence) and its actual structure solved through X-ray crystallography is a biologically justified metric for

success. Currently, there are already several cases where computationally-predicted structures give an atomic level accuracy better than 2.5 angstroms, but regularly achieving such accuracy, especially for large proteins and with a diversity of structural features, remains a critical challenge. Furthermore, the most successful computational platforms still rely on homology to existing proteins at various levels of resolution. And even at the single-residue resolution, there is still reliance on existing protein structures – for example, conformational rotamers in a leading protein design platform, Rosetta, are partly scored based on their frequency in the PDB (Alford et al., 2017; Davey, Damry, Goto, & Chica, 2017). Therefore, the types of proteins that can currently be designed are still ones close to natural proteins. Moving farther and farther away from natural structures should result in both a better understanding of protein biophysics and new scaffolds specialized for new applications.

In terms of the types of protein functions that can be effectively designed, enzyme activity presents a major current challenge. One reason for this is that enzymes may rely on intricate molecular dynamics for catalysis that are difficult to capture in current design platforms; currently only single-residue conformational dynamics have been engineered by computational design (Davey et al., 2017). Better addressing the challenge of enzyme design would enable broad advances in synthetic biology. For example, enzymes are at the heart of metabolic pathway engineering goals. Quantifiable metrics in enzyme design can be based on diffusion-limited rate constant (k_{cat}/K_M) improvements. This limit for enzymes is ~10^9 $M^{-1}s^{-1}$, natural enzymes average a k_{cat}/K_M of approximately 10^5 $M^{-1}s^{-1}$, but computationally designed enzymes have k_{cat}/K_M values that are usually around three orders of magnitude lower than natural enzymes (Bar-Even et al., 2011; Kuo-chen & Shou-ping, 1974). A major goal of computational enzyme design should be to routinely achieve the k_{cat}/K_M values of natural enzymes for artificial (user-defined) reactions.

The success rate of protein structure prediction and computational protein design is still low. This significantly limits broad adoption of protein design by the biomolecular engineering community. Because of the inaccuracy and imperfection of the molecular mechanics force-field underlying protein design, highly trained experts are often needed to curate computational design to select the ones that will be tested experimentally. The scant availability of such experts limits the broad deployment of protein design within the industrial and academic communities. Leveraging high-throughput experimental screening of large numbers of computational designs is a way to alleviate such limitations, but result in very high cost for design projects, which in turn restrict application of design to the most well-funded academic and industrial institutions. For example, the typical rate of success for enzyme design is in the low percent range. A rate of success greater than 50% would therefore have a tremendous impact.

Achieving these goals will require progress on multiple aspects of design, which can be categorized as physics-based or knowledge-based. Physics-based design approaches will advance through the improvement of the molecular mechanics force-fields and knowledge-based design approaches will advance through the curation of very large datasets of positive and negative design outcomes to enable the further development of machine learning techniques that extract design models from data. A combination of physics- and knowledge-based advances may be required to maximize the success rate for computational protein

design. Physics-based molecular dynamics simulations can incorporate protein dynamics that may be at the heart of certain protein functions, but the amount of computational power required make it impossible to perform dynamics simulation on all design candidates at all scales. The ability to incorporate coarse-grained or full-atom dynamics in the design stage would critically enable the design of enzymes with high catalytic activities without the need to use laboratory evolution post-design.

Computational nucleic acid design

There has been considerable interest in designing nucleic acids (DNA, RNAs) and nucleic acid machines to carry out custom function (e.g., binding, sensing, catalysis, regulation) because nucleic acids are arguably uniquely programmable due to their reliance on base pairing for secondary structure, while allowing a wide range of sophisticated structural elements through tertiary and non-canonical structures. Several breakthrough technologies have emerged recently based on RNA-protein complexes that rely on base-pair guided interactions, including RNA silencing, CRISPR genome editing and gene activation and repression, and therapeutics that target pre-mRNA splicing. Nevertheless, compared to computational protein design, nucleic acid design, especially design of RNAs at the non-canonical and tertiary structural level is underdeveloped. In addition, the design of complexes that mix RNA and proteins, or 'nucleoproteins', remains particularly underdeveloped. A major goal of the field should be to resolve this gap.

A number of computational approaches have been developed to be able to predict secondary and tertiary structures from primary sequences, and more recently, increased interest in nucleic acid (DNA/RNA) nanotechnology has emerged given the potential of developing computational rules that can lead to nucleic acids that assemble into complex shapes. Most of the recent successes in computational nucleic acid design leverage our knowledge of secondary structure thermodynamics to design at the level of RNA secondary and canonical Watson-Crick base pairing interactions. In order to expand computational nucleic acid design, several major areas of improvement are needed including: 1) incorporating RNA folding kinetics into design algorithms, both from the standpoint of designing efficient folding pathways, as well as designing structures that can dynamically change in response to local or global environmental changes and specific interactions with other molecules (Espah Borujeni, Mishler, Wang, Huso, & Salis, 2016; Espah Borujeni & Salis, 2016); 2) designing at the level of three dimensional structure; 3) incorporation of non-canonical interactions (i.e. Hoogstein base pairing, nucleotide-backbone pairing, etc.) into design approaches (Das, Karanicolas, & Baker, 2010); 4) incorporating the growing number of synthetic nucleotide chemistries (i.e., Hachimoji codes (Hoshika et al., 2019)) within nucleic acid design; and 5) integrating frameworks for RNA and protein design (Leistra, Amador, Buvanendiran, Moon-Walker, & Contreras, 2017) to predict and design structure-function relations for nucleic acids alone and in the context of riboprotein complexes and hybrid structures. Together, this will unlock the ability to design new and powerful target functionalities including: RNA-ligand binding pockets for new ligands relevant to biosensor design; catalytic sites for improved RNA catalysis in ribozymes and within the ribosome for new functions such as bespoke gene editing tools and templated unnatural polymer biosynthesis, respectively (Carothers, Goler, Juminaga, & Keasling, 2011); higher-order

dynamic RNAs that can change global folding patterns in response to stimuli relevant to single-molecule molecular logics; improved RNA-protein complexes for RNA-guided gene editors (e.g., CRISPR systems); new classes of RNA-protein nanomachines that can perform cell-like functions such as cargo sorting and transport; and RNAs that post-transcriptionally control gene expression in a targeted way by directly regulating stability of entire clusters of mRNAs via designed RNA-RNA interactions (we note that this can be particularly relevant to the engineering and optimization of metabolic pathways and complex phenotypes in a variety of hosts (Leistra, Curtis, & Contreras, 2019)).

Evolutionary macromolecular engineering

Evolution is a powerful bioengineer, but the natural evolutionary process is slow. New directed evolution platforms for rapid optimization of nucleic acids, proteins, pathways, and circuits towards desired functions are needed. Metrics for effective directed evolution include: 1) fold-improvement over starting point function, as well as absolute-level of function that can be evolved; 2) the types of functions that can be evolved; and 3) scale (how many experiments can be run simultaneously). Although there are many applications of directed evolution for all types of biomolecules, there are two particularly demanding testbeds for directed evolution technologies discussed below: protein enzyme evolution and the specific binding of nucleic acids to small molecules or proteins.

Enzyme evolution

Only a small fraction of directed enzyme evolution experiments give increases in k_{cat}/K_M that bring activities within range of natural enzymes (Goldsmith & Tawfik, 2017). Empirically, the most extensive directed evolution efforts yielding orders of magnitude improvement in k_{cat}/K_M have typically required about ten mutations (Goldsmith & Tawfik, 2017), but classical directed evolution methods rarely traverse adaptive mutational pathways with >5 mutations. Similar metrics aiming for ten mutations guide the evolution of binding proteins, where this scale of mutation is necessary for achieving extremely high-affinity binders (picomolar) from weak binders (micromolar). Therefore, platforms for directed protein evolution that can routinely yield variants with ten or more adaptive mutations are desirable. We note that the related problem of evolving a new protein-based biosensor such as a transcription factor that binds to a new ligand is related to the K_M problem for enzyme evolution. Thus, improved technologies for evolving enzymes will have great impact in other areas of biomolecular engineering.

DNA/RNA aptamer evolution

DNA or RNA sequences that can bind to specific proteins or small molecule ligands are commonly referred to as aptamers. While there was an initial push to evolve new aptamers via techniques such as SELEX (Ellington & Szostak, 1990; Tuerk & Gold, 1990), the field has existing challenges in terms of the chemical diversity of small molecules that can be targeted with aptamers, as well as incorporating aptamers into functional RNA molecules such as biosensors or gene regulators. With regards to the challenge of diversity, aptamers have been largely limited to those targets that are already known to bind to RNA well (such as nucleotide analogs and other co-factors for which natural aptamers exist), or to compounds that can be

easily immobilized on solid supports. In terms of aptamers in functional RNA molecules, there is some recent progress in incorporating new aptamers into fluorescent RNA biosensors as well as new classes of RNA regulators called riboswitches, but progress is still hampered by a lack of understanding of ligand-mediated allosteric effects that alter RNA structure (Carlson & Lucks, 2019; Villa, Su, Contreras, & Hammond, 2018). Both of the challenges could be fruitfully addressed by new evolution methods that selected for RNA binding interactions in the context of functional molecules - for example by new selection methods that used the binding of free ligand (i.e., not bound to a solid support) to trigger a regulatory event (e.g.., activation of transcription) that could be selected for. In addition, evolutionary methods that can support the use of non-natural nucleic acids could further enhance the diversity of chemical interactions and structural motifs available to offer new ligand binding properties. We note that progress towards this goal would also impact the recent and growing interest in developing small molecule drugs for RNA targets (Palacino et al., 2015).

Development of platform technologies for evolutionary macromolecular engineering

Platforms for directed evolution that can traverse long mutational pathways are critical for crossing fitness valleys. Given the length of functional biopolymers (i.e., proteins), evolution (and any design strategy) is (and always will be) a highly limited search through sequence/fitness space. Given the ruggedness of fitness landscapes, this results in fixation of suboptimal sequences that represent local fitness maxima. Crossing fitness valleys to reach more-fit maxima is always a critical challenge in directed evolution. To cross these valleys, multi-mutation pathways, which can be elicited by fluctuating or changing selection conditions or spatial structure, are critical. In addition, when directed evolution is used to improve multi-gene metabolic pathways, the number of beneficial mutations available/needed to achieve an optimal function increases compared to the evolution of single enzymes. As a result, directed evolution systems capable of traversing long mutational pathways are here needed, too. Developments in continuous evolution systems and automation should allow directed evolution to address such demanding biomolecular engineering goals by accessing long mutational pathways at scale (Badran & Liu, 2015; Halperin et al., 2018; Ravikumar, Arzumanyan, Obadi, Javanpour, & Liu, 2018; Zhong & Liu, 2019).

Evolution platforms require selection. Therefore, the types of functions that one can evolve biomolecules to achieve depend on the availability of high-throughput screens and genetic selections for those properties. Potential selection systems for genetic growth-based selection are abundant - as the propagation of biomolecular variants with desired properties can be linked to cell survival through synthetic genetic circuits - but the reliability of these selections and the difficulty in designing new ones vary widely (Higgins & Savage, 2018). *In vitro*, screening throughput is lower than for *in vivo* selection, but there is more control over what is screened for through the precision of FACS and assays involving droplet sorting systems or microtiter plates. Therefore, on the one hand, there is a need for more and more general *in vivo* selection systems, for example, custom transcriptional biosensors for arbitrary small molecules to select biomolecules (enzymes, RNAs) that produce desired products, two-hybrid systems, nucleic acid-protein interaction selection systems, display systems (e.g., ribosome), and other binding-based selection systems to find custom affinity reagents or interaction inhibitors; as well

as diversification methods (e.g., PACE (Esvelt, Carlson, & Liu, 2011; Halperin et al., 2018; Ravikumar, Arrieta, & Liu, 2014; Ravikumar et al., 2018), evolvR (Esvelt et al., 2011; Halperin et al., 2018; Ravikumar et al., 2014, 2018)), and OrthoRep (Esvelt et al., 2011; Halperin et al., 2018; Ravikumar et al., 2014, 2018)) that operate *in vivo* to match the possible throughput of selection. On the other hand, there is a need for highly-streamlined and high-throughput *in vitro* screening technologies that will likely utilize new technologies in microfluidics and DNA barcoding and sequencing. We note that for certain types of nucleic acid based selections such as SELEX, one can carry out *in vitro* selections with extremely large libraries (>10^{14} variants), which greatly increase the likelihood of recovering functional molecules (Carothers, Oestreich, Davis, & Szostak, 2004). Improving the throughput of other *in vitro* or *in vivo* screening technologies to enable more than the current 10^6-10^8 variants to be screened could significantly improve the ability to evolve molecules with properties that are difficult to access using SELEX.

In terms of scale, classical methods usually limit one researcher to carry out no more than ten independent biomolecular evolution experiments in parallel, especially if multiple rounds of mutation and selection are needed. Evolution platforms that can exceed this scale are therefore desired. Continuous *in vivo* evolution platforms can achieve scale by requiring only serial passaging for selection, as serial passaging of cells is highly scalable. Continuous *in vitro* nucleic acid evolution platforms similarly scale (Blind & Blank, 2015). Although automated methods for selecting macromolecules, including RNA aptamers (Cox et al., 2002), were first developed more than 15 years ago, the relatively limited capabilities and high costs of those robotic platforms hampered widespread adoption. Looking ahead, laboratory automation may begin to more easily increase the scale of evolutionary macromolecular engineering undertakings as many steps such as PCR diversification, transformations, and selection have become much more readily automated (Linshiz et al., 2013; Shih et al., 2015).

Finally, we note that evolutionary macromolecular engineering and computational design are highly complementary. Not only can computational design provide starting points for evolutionary engineering of both proteins and nucleic acids (especially powerful since gene synthesis is scalable so many starting points can be exactly made), evolutionary methods - in particular highly scalable ones - can generate large sets of successful (and unsuccessful) outcomes to train computational algorithms. Indeed, machine-learning in protein engineering is a rapidly expanding area of research that we believe holds great promise and is highly synergistic with the large datasets that can be provided through continuous evolution experiments and high-throughput sequence-function mapping experiments (Yang, Wu, & Arnold, 2018).

Collection and curation of more biomolecular parts

While we have emphasized how one may design and evolve new biomolecular parts, there is already a rich existing collection of natural biomolecules that nature offers. Proper prospecting and curation of parts from the rapidly growing number of genomes sequences is a valuable strategy to complement design and evolutionary approaches. Even if design and evolutionary approaches rapidly advance, there is still the need for good starting points for design and evolution to modify and these starting points come from parts collections.

As parts collections expand, including through the addition of more and more synthetic variants, characterization and curation become crucial. Standardized methods for measuring the performance of particular parts is therefore essential. This is especially important for parts controlling gene expression, which form the basis of biological circuit design. Host specificity, environmental effects, modularity, and tunability of parts are all critical aspects in biological circuit design.

Unnatural nucleotide and amino acid polymerization systems

The construction of macromolecules that contain unnatural building blocks would be broadly useful for new therapeutics, materials, and biocontainment strategies. Systems for PCR and transcription of fully unnatural nucleotide-containing genes of up to 400 base pairs is an aspirational but reasonable, metric the field should aim for. At this length, unnatural aptamer and aptazyme polymers could be regularly evolved and engineered. Systems capable of handling even longer sequences (1000 base pairs) would be useful as new information polymers capable of encoding unnatural proteins and sustaining genetic codes based on new genetic alphabets (Martin et al., 2018). Expanded genetic code systems for translation of fully unnatural amino acid containing proteins with more than 200 amino acids and/or proteins with at least four distinct unnatural amino acid building blocks would also be an aspirational quantifiable goal for the field. This goal would open up new categories of research in biomaterial production and evolution and further motivate the expansion of genetic codes, a key area of synthetic biology with a wide range of applications from biomolecular engineering to biocontainment.

BIOMOLECULE, PATHWAY, AND CIRCUIT ENGINEERING

Goal	Breakthough Capability	Milestone

On-demand design, generation, and evolution of macromolecules for desired functions.

2 Years	5 Years	10 Years	20 Years
***De novo* prediction of RNA structure, protein structure, and complexes of DNAs/RNAs and proteins from primary sequence and the ability to make accurate predictions of mutability and effect of mutations from structure.**			
Reliably predict the structure of 300-amino acid proteins and 200-nucleotide RNA domains within 5 Ångstroms from primary sequence.	Reliable *de novo* prediction of RNAs and proteins containing non-canonical structures.	Routine prediction of structures for 500-amino acid proteins and 200-nucleotide RNA domains within 3 Ångstrom.	Routine prediction of structures for 3,000-amino acid proteins, protein-protein and RNA-protein interactions, and protein and RNA-protein complexes.
		Design proteins that fold correctly 50% of the time and RNA-protein complexes that form correctly 20% of the time.	
Improve force-field and backbone-sampling algorithms and capabilities to capture force-fields of post-transcriptionally- and post-translationally-modified nucleosides and amino acids.	Routine redesign of ligand binding sites and/or aptamers for custom ligands with a greater than 50% success rate.	Modeling and design of chromatin states that can be manipulated to change function.	Routine prediction of protein function from structure.
***De novo* design and/or prediction of macromolecular dynamics and dynamic macromolecular structures.**			
Improving computational models of RNA dynamics that can incorporate experimental data.	Incorporating co-transcriptional (for RNA) and co-translational (for protein) processes into design algorithms.		Routine design of enzymes with high activities.
	Design of intrinsic regulatory control into biomolecules.		Modeling and design of dynamic RNA nanomachines that can engage with and manipulate the chormatin states of living systems.
	Design of dynamic and responsive protein-RNA nanomachines.		
	Routine design of large proteins, beta topologies, membrane proteins, and loops.		Modeling and design of dynamic DNA-RNA-protein condensates that can expand beyond the functionality of natural condesates.
	Routine design of protein complexes.		
High-throughput integrated computational, experimental, and evolutionary schemes for refinement of desired biomolecule functions.			
Durable and high-mutation-rate *in vivo* continuous DNA mutagenesis and evolution systems in model organisms.	Durable and high-mutation-rate *in vivo* continuous DNA mutagenesis and evolution systems in non-model organisms.	Full control over all statistical properties of DNA diversification *in vivo*.	*De novo* DNA synthesis *in vivo* with single-cell sequence control.
		Direct sequencing of proteins and carbohydrates.	Ability to select for any function, including those conferred by: A) small molecules, lipids, or carbohydrates, and; B) proteins or nucleic acids.

Special considerations for on-demand design, generation, and evolution of macromolecules that rely on non-canonical/unnatural building blocks.

2 Years	5 Years	10 Years	20 Years
PCR, reverse transcription, cellular replication, and transcription of fully unnatural nucleotide-containing genes of up to 400 base pairs.			
Identification of "missing" functionality or functionalities in A-T-G-C base pairs.	Improved *in vitro* manipulation of unnatural nucleic acids.	Biosynthesis of unnatural nucleotides.	Establishment of organisms capable of full replication, maintenance, and transcription of a plasmid or artificial chromosome made up entirely of unnatural bases.
	Expansion of unnatural nucleotide toolkit.		

Expanded genetic code systems for translation of >100-amino acid proteins containing fully-unnatural amino acids, and proteins with at least four, distinct unnatural amino acid building blocks.			
Create proteins that are capable of gaining new, therapeutically-useful activities through unnatural amino acids.	Efficient biosynthesis of proteins containing three or more distinct unnatural amino acid building blocks.	Biosynthesis of unnatural amino acids.	Templated biosynthesis and evolution of new polymers with large user-selected sets of unnatural building blocks *in vivo*.

Holistic, integrated design of multi-part genetic systems (i.e., circuits and pathways).

Design of highly-stable, large genetic systems (genomes) with targeted expression levels in a host organism or cell type, incorporating system-wide effects.			
Incorporate gene expression interactions into predictable design of prokaryotic genetic systems.	Incorporate gene expression interactions into predictable design of eukaryotic genetic systems.	Discovery and characterization of mechanistic interactions at the systems-level affecting protein activities inside cells.	Whole-tissue or whole-cell, nucleotide-resolution simulations encompassing several layers of models predicting gene regulatory, metabolic, and system-level behaviors.

Ability to rationally engineer sensor suites, genetic circuits, metabolic pathways, signaling cascades, and cell differentiation pathways.			
Reliable engineering of genetic circuits with more than 10 regulators for sophisticated computations.	Reliable engineering of novel, many-enzyme pathways utilizing combinations of bioprospected enzymes with well-characterized kinetics.	Reliable expression of redesigned synthases to produce secondary metabolites.	Simultaneous, tunable, timed expression of many transcription factors controlling mammalian cell state.
	Five-time improvement and expansion of inducers/promoters for model organisms that respond to environmental inputs and any intracellular metabolite.	Computational design of protein-ligand and RNA-ligand interfaces suitable for engineering protein-based or RNA-based sensors.	
	Utilize machine-learning approaches to use the vast amount of uncurated literature results within pathway design.		

Integrated design of RNA-based regulatory systems for cellular control and information processing.

Porting nucleic acid strand displacement technology into cellular systems with RNA instantiations.			
RNA implementation of strand displacement cascades in bacteria.	RNA implementation of strand displacement cascades in eukaryotic systems.	Engineer computational RNA strand displacement networks in mammalian systems.	Computational design of RNA strand displacement neural networks that process the transcriptome.
	Engineer 'universal' computational strand displacement architectures using strand displacement in bacteria.		Engineer RNA neural networks that dynamically reprogram cell state.

Porting successes in computationally designed bacterial RNA-based genetic regulators into eukaryotic and mammalian systems.			
First generation eukaryotic RNA-based gene regulators that utilize RNA:RNA interactions and/or strand-displacement and achieve 10-fold change in gene expression.	Second generation eukaryotic RNA-based gene regulators that are suitable for computational design to create libraries that are highly-orthogonal and high-performing, achieving 100's-fold change in gene expression.	Expand RNA modification apparatus to modify non-natural RNA alphabets to enhance their functional properties.	Engineering enzymes that can perform non-natural RNA modifications to further expand the chemical repertoire of what is possible and extend RNA ligand recognition, catalysis and genetic control.
Creation of RNA modification machinery that allows programmable site-specific modifications of RNA, focusing on naturally abundant modifications.	Use RNA modifications for programming or fine-tuning RNA functions.		

2 Years **5 Years** **10 Years** **20 Years**

Roadmap Elements

Goal 1: On-demand design, generation, and evolution of macromolecules for desired functions.

[Current State-of-the-Art]: Currently, the mapping of structure and function of a macromolecule from the primary sequence is the critical challenge towards achieving on-demand design, generation, and evolution. Computational DNA, RNA, and protein design has advanced to the point where defined structures, binding interactions, and enzymatic activity can be constructed, especially for proteins. Still, substantial improvements are needed in expanding: 1) the range and effectiveness of macromolecular functions that can be designed, and 2) success rate. *De novo* computational protein design, PDB-informed protein design strategies, origami-based nucleic acid structure design, physics-based design of RNA switches, machine-learning strategies that deduce molecular contacts for protein and nucleic acid folding from multiple sequence alignments, and hybrid approaches have enjoyed considerable success and hold significant promise.

Evolutionary or semi-rational approaches have advanced to the point where substantial improvements can be gained via a wealth of directed evolution, continuous evolution, and library based approaches, often coupled with computation and modelling, but only when suitable macromolecules (i.e., those that possess some function along the axis of the desired function) have been previously identified. However, creating effective and scalable diversification systems, effective selection and screening systems, reaching *de novo* evolution of function, and expanding the scope and throughput of selection/screening systems with the ability to directly select/screen for the exact function remain critical challenges.

[Breakthrough Capability 1]: De novo prediction of RNA structure, protein structure, and complexes of DNAs/RNAs and proteins (from primary sequence) and the ability to make accurate predictions of mutability and effect of mutations from structure.

- **2 years: Reliably predict (greater than a 50% success rate) the structure of 300-amino acid proteins and 200-nucleotide RNA domains within 5 Ångstroms from primary sequence.**
    - [Bottleneck]: Existing methods for both RNA and protein structure design rely heavily on macromolecules of known structure.
        - [Potential Solution]: Machine learning with coevolutionary models on large multiple sequence alignments of homologous RNAs and proteins to extract structure from sequence alone.
        - [Potential Solution]: Better understand structures in the sequence space of non-biological RNA and proteins such as de novo designed structures.
    - [Bottleneck]: There are no methods capable of predicting RNA-protein complexes at even modest resolution from primary sequence.
        - [Potential Solution]: Use of high-throughput technologies for mapping RNA-protein crosslinks, nucleotide-resolution chemical mapping of RNA components, and rapid cryo-EM of RNA-protein complexes to guide and test computational modeling.

- **2 years: Improve force-field and backbone-sampling algorithms and include capabilities to capture force-fields of post-transcriptionally- and post-translationally-modified nucleosides and amino acids.**
    - [Bottleneck]: Conformational dynamics for protein design need to be improved (especially for hydrogen-bonding and electrostatic interactions between protein residues) to more accurately capture interactions responsible for protein structure, stability and function; similarly, algorithms that sample potential protein conformations required for function need to be sped up and improved.
        - [Potential Solution]: Gather large datasets of mutants or designed protein sequences and their experimentally characterized activity and use machine learning/data science techniques to develop improved molecular mechanics force-fields.
        - [Potential Solution]: Investigate alternative sampling algorithms for protein backbone by developing protein design software able to take advantage of commodities parallel computing architectures such as general purpose GPUs and cloud-based FPGAs.
    - [Bottleneck]: Force fields for molecular dynamics simulations of RNAs and RNA-protein complexes need to be improved.
        - [Potential Solution]: Detailed biophysical characterization of synthetic model systems designed to push the force fields to their limits of predictability.
        - [Potential Solution]: Development of design methodologies that leverage angstrom-level RNA and RNA-protein complex simulations.
    - [Bottleneck]: Even when a global conformational minimum is sampled, there is no guarantee that a force-field will correctly identify it as such because computational protein design algorithms sacrifice scoring accuracy for speed often by only considering pairwise interactions.
        - [Potential Solution]: Systematic errors caused by the assumption that pairwise interactions are sufficient to define protein folds must be identified, and score terms that can quickly approximate those errors should be implemented to improve accuracy.
- **5 years: Reliable *de novo* prediction (greater than a 50% success rate within 5 Ångstrom r.m.s.d.) of RNAs and proteins containing non-canonical structures (including irregular protein loops and RNA aptamers).**
    - [Bottleneck]: There exists a large variety in possible loop conformations, making them difficult to effectively sample. The problem is particularly important for RNA and RNA-protein complexes where functional tertiary folds are dictated by idiosyncratic structures.
        - [Potential Solution]: Utilize knowledge of RNA and protein sequence-structure relationships from the PDB to limit conformational search space; explore refinement strategies that couple this with physics-based score functions, molecular dynamics simulations, and employ new generation Monte Carlo sampling methods (Watkins et al., 2018).

        - [Potential Solution]: Incorporation of failed designs, based on structural data, into models.
- **5 years: Routine redesign of ligand binding sites and/or aptamers for custom ligands with a greater than 50% success rate.**
    - [Bottleneck]: Motif-based approaches to ligand binding site design is limited by how much structural information of the target ligand is available.
        - [Potential Solution]: Fragmenting ligands allows for the formation of larger motif libraries, enabling the design of completely novel ligand binding sites.
        - [Potential Solution]: Extensions of new Monte Carlo methods (Watkins et al., 2018) for atomically-accurate RNA and protein structure prediction to sample sequence and structure simultaneously.
        - [Potential Solution]: Computational medium-resolution pre-design of large 3D RNA and protein shapes that encapsulate all surfaces of target ligands, integrated with high-throughput combinatorial screening focused at macromolecule-ligand interfaces for which current computational approaches are not yet accurate.
- **10 years: Routine prediction of structures for 500-amino acid proteins and 200-nucleotide RNA domains within 3 Ångstrom.**
    - [Bottleneck]: Diversity of conformations.
- **10 years: Design proteins and RNAs that fold correctly 50% of the time and RNA-protein complexes that form correctly 20% of the time.**
    - [Bottleneck]: High-throughput methods to assess protein and RNA folding.
        - [Potential Solution]: Improved dynamic light scattering, high-throughput circular dichroism, high-throughput crystallography, or cryo-electron-microscopy assays.
- **10 years: Modeling and design of chromatin states that can be manipulated to change function.**
    - [Bottleneck]: Incomplete functional characterization of natural chromatin.
        - [Potential Solution]: Engineered platforms to rapidly interrogate hundreds of structural, enzymatic, and synthetic chromatin proteins.
- **20 years: Routine prediction of structures for 3,000-amino acid proteins (such as PKSs, etc.), protein-protein and RNA-protein interactions, and protein and RNA-protein complexes (re-engineered ribosomes, spliceosomes, etc.).**
    - [Bottleneck]: Diversity of conformations.
- **20 years: Routine prediction of RNA and protein function from structure.**
    - [Bottleneck]: While it is possible to predict the function of unknown RNAs and proteins by homology to molecules with known function, *ab initio* functional prediction will be a major challenge.
        - [Potential Solution]: No clear path to a solution can be envisioned, but it will likely involve the ability to computationally model interaction networks between arbitrary sets of biomolecules simply based on their sequences.

        - [Potential Solution]: Achieving a better understanding of the full repertoire of possible protein and RNA functions is needed.

[Breakthrough Capability 2]: De novo design and/or prediction of macromolecular dynamics and dynamic macromolecular structures.

- **2 years: Improving computational models of RNA dynamics that can incorporate experimental data.**
    - [Bottleneck]: There is a lack of rigorous physical models that can incorporate experimental characterization of RNA structure and physicochemical data into models of RNA folding dynamics.
        - [Potential Solution]: Expansion of RNA folding dynamics physicochemical modeling toolsets to incorporate experimental data.
        - [Potential Solution]: Machine-learning based models of RNA structure based off large-scale experimental characterization datasets.
- **10 years: Incorporating co-transcriptional (for RNA) and co-translational (for protein) processes (and including cellular factors that participate in these processes) into design algorithms.**
    - [Bottleneck]: A lack of principles of co-transcriptional RNA folding and co-translational protein folding that can be incorporated into design algorithms.
        - [Potential Solution]: Approaches to use model systems along with a variety of techniques (high-throughput chemical biology, biophysical) to uncover the required principles.
    - [Bottleneck]: No RNA or protein design algorithms incorporate co-transcriptional or co-translational folding dynamics into the design process.
        - [Potential Solution]: Incorporate the design principles learned from the study of these processes into these algorithms.
        - [Potential Solution]: Develop appropriately coarse-grained models that can efficiently simulate co-transcriptional and co-translational folding.
- **10 years: Design of intrinsic regulatory control into biomolecules (e.g., allostery).**
    - [Bottleneck]: Long-range interactions in proteins are difficult to capture in computational protocols because of the enormous amount of conformational sampling that would be required, propagation of error, and limitations in scorefunction accuracy over long ranges.
        - [Potential Solution]: A focus on short-range allosteric interactions may be necessary; statistical approaches to understanding long-range allosteric interactions will be useful for future regulatory design.
    - [Bottleneck]: We lack a sufficient understanding of how ligand-, protein-, and RNA-RNA binding can dynamically alter RNA structure in either equilibrium or out-of-equilibrium RNA folding regimes.
        - [Potential Solution]: Development and validation of approaches that can map RNA-ligand and protein- and RNA-RNA interactions at atomic resolution and in high-throughput.

        - [Potential Solution]: Development and validation of approaches that can extract RNA folding sub-population information to uncover principles of ligand-, protein-, and RNA-RNA mediated conformational changes.
    - [Bottleneck]: There are few RNA aptamers that can sense ligands with Kd's that are sub-micromolar, required for many applications.
        - [Potential Solution]: Expansion into non-natural nucleic acid chemistries to expand the structure and chemical diversity of aptamers.
- **10 years: Design of dynamic and responsive protein-RNA nanomachines.**
    - [Bottleneck]: It is challenging to image the three-dimensional structure of self-assembled protein-RNA nanostructures within the cell.
        - [Potential Solution]: Application of high-resolution techniques (super-resolution imaging, cryo-EM) to model synthetic protein-RNA nanostructures.
        - [Potential Solution]: Use of high-throughput techniques (in-cell chemical probing) to validate proper cellular assembly.
    - [Bottleneck]: Primitives for converting molecular binding (e.g. ligands, RNAs) into changes in a three-dimensional protein-RNA nanostructure are underdeveloped.
        - [Potential Solution]: Incorporate known natural motifs that allow ligand and RNA-mediated switching into protein-RNA nanostructures.
    - [Bottleneck]: RNA-protein interactions required for nanomachines with most sophisticated functions (e.g., dynamic control over protein complexes, cell signaling pathways) are challenging to engineer.
        - [Potential Solution]: Harness improved tools for predicting RNA-protein interactions and integrate them into design of dynamic RNA nanomachines.
- **10 years: Routine design of large proteins, beta topologies, membrane proteins, and loops.**
    - [Bottleneck]: Although these challenges have recently been addressed, the computational methods are too nascent to ensure that successful design is achieved routinely.
        - [Potential Solution]: Continued exploration of these computational methods will begin to elucidate the potential for success and existing limitations.
    - [Bottleneck]: Designing functional proteins requires successful prediction of not just the topology, but also the precise positioning of the elements within that topology.
        - [Potential Solution]: Explore new approaches to designing specific conformations within an existing topology that satisfy user-specified parameters, such as angles between secondary structure elements.
- **10 years: Routine design of protein complexes.**
    - [Bottleneck]: Predicting and modelling protein-protein interactions is difficult.
        - [Potential Solution]: Continued development of co-evolutionary models, physics models, and design platforms.

    - [Bottleneck]: Stronger influence of environment: in contrast to the design of individual proteins, complexes require molecules to assemble in a sea of other molecules.
        - [Potential Solution]: Improved molecular dynamics simulations.
- **20 years: Routine design of enzymes with high activities (i.e., k_{cat}/K_M > 10^5 1/M*s).**
    - [Bottleneck]: Most powerful protein design platforms don't address molecular dynamics well, and protein dynamics are fundamentally challenging to capture.
        - [Potential Solution]: Ability to at-will engineer enzyme specificity, including to understand what enzymes exist, understand principles behind what exists, and map domain and sequence/functions.
        - [Potential Solution]: Improve multi-state design algorithms which are aimed at designing proteins with multiple interchanging conformations.
    - [Bottleneck]: Successful catalysis often requires considerations other than conformation and residue positioning, such as active site electrostatics.
        - [Potential Solution]: Explore the use of more accurate but computationally expensive simulations, such as quantum mechanical calculations, to determine the optimal electrostatic environment for a desired reaction; couple this knowledge with constraints on active site electrostatics during the design process. Alternatively, use knowledge from existing enzymes that catalyze similar reactions to guide these constraints.
- **20 years: Modeling and design of dynamic RNA nanomachines that can engage with and manipulate the chromatin states of living systems.**
    - [Bottleneck]: *De novo* understanding of RNA structure-function relationship is currently beyond reach and measuring manipulations of chromatin states in complex living systems adds an additional layer of uncertainty.
- **20 years: Modeling and design of dynamic DNA-RNA-protein condensates that can expand beyond the functionality of natural condensates.** For example, heterochromatin, mediator, and Pol II nuclear condensates that govern transcription initiation.
    - [Bottleneck]: Robust, physics-based models of condensate interactions that can be extrapolated to predict unknown functionality.

[Breakthrough Capability 3]: High-throughput integrated computational, experimental, and evolutionary schemes for refinement of desired biomolecule functions including enzymatic activity and binding. For related reading, please see *Gene Editing, Synthesis, and Assembly*, which contains information regarding DNA diversification and library synthesis techniques that can be combined with *in vivo* diversification and assay/selection schemes described here.

- **2 years: Durable and high-mutation-rate *in vivo* continuous DNA mutagenesis and evolution systems in model organisms.**
    - [Bottleneck]: Increase mutation rates for *in vivo* continuous random mutation systems.
        - [Potential Solution]: Lower biases for *in vivo* continuous mutation systems.

- **5 years: Durable and high-mutation-rate *in vivo* continuous DNA mutagenesis and evolution systems in non-model organisms.**
    - [Bottleneck]: Portability of *in vivo* continuous mutation systems.
        - [Potential Solution]: Parts mining for autonomous genetic systems.
- **10 years: Full control over all statistical properties of DNA diversification *in vivo*.**
    - [Bottleneck]: Libraries of DNA can be created through *in vivo* continuous mutagenesis systems such as evolvR; but mutation rates and preferences are not precisely controlled.
        - [Potential Solution]: Future generations of programmable mutagenesis tools will likely include enhanced activities such that mutation rate and outcome can be introduced in a more controlled fashion.
- **10 years: Direct sequencing of proteins and carbohydrates.**
    - [Bottleneck]: Current instrumentation tools and technologies.
        - [Potential Solution]: High-throughput mass spectrometry that unambiguously identifies protein variants or carbohydrate linkages in a complex mixture.
    - [Bottleneck]: Limited techniques appropriate for direct sequencing.
        - [Potential Solution]: Massively parallel detection and sequencing of proteins and carbohydrates using principles from high-throughput DNA sequencing adapted to other molecules through, for example, labeled primary-sequence specific affinity reagents.
- **20 years: *De novo* DNA synthesis *in vivo* with single-cell sequence control.**
    - [Bottleneck]: Individual cells can be programmed to synthesize specific DNA sequences *de novo*, for example through light-triggered template-independent DNAPs.
        - [Potential Solution]: Specialized collections of non-template-dependent DNA polymerases and DNA editing enzyme whose exact activities are controllable by increasingly penetrant forms of energy, starting with light.
- **20 years: Ability to select for any function, including those conferred by: 1) small molecules, lipids, or carbohydrates; and 2) proteins or nucleic acids, including biophysical properties or properties not easily tied to growth.**
    - [Bottleneck]: Technology to tie production of small molecule, lipids, or carbohydrates to a selection.
        - [Potential Solution]: Cell adhesion on a surface; glycoarrays, lectin arrays.
        - [Potential Solution]: Creation of biosensors in a cell to link product to cell death or sortable phenotype (e.g., fluorescence).
        - [Potential Solution]: Improve small molecule detection and collection via capillary electrophoresis.
    - [Bottleneck]: Technology to select for any macromolecular function or property (e.g., fold, shape).
        - [Potential Solution]: Synthetic use of natural channels, transporters, quality control systems that naturally discriminate these properties in living systems.

Goal 2: Special considerations for on-demand design, generation, and evolution of macromolecules that rely on non-canonical/unnatural building blocks.

[Current State-of-the-Art]: While DNA, RNA, and proteins containing natural building blocks are readily synthesized using natural biological machinery and the rules of templated biosynthesis, DNA, RNA, and proteins containing modified or unnatural building blocks, including ones that recapitulate post-translational modifications, are difficult to access. Only certain unnatural building blocks are available, only a few distinct unnatural building blocks can be used simultaneously, and the length of fully unnatural polymers that can be produced is extremely low compared to natural counterparts. Overcoming these bottlenecks will lead to new biomolecules with expanded functions stemming from the diversity of non-canonical and unnatural building blocks that could become available to synthetic biology.

The design, generation, and evolution of macromolecules containing unnatural building blocks relies on the achievement of the same capabilities as the production of wholly-natural macromolecules. The following reflects the special considerations necessary for the utilization of unnatural building blocks.

[Breakthrough Capability 1]: PCR, reverse transcription, cellular replication and transcription of fully unnatural nucleotide-containing genes of up to 400 base pairs. At this length, unnatural aptamer and aptazyme polymers could be regularly evolved and engineered.

- **2 years: Identification of "missing" functionality or functionalities in A-T-G-C base pairs.**
    - [Bottleneck]: Previous work in this field has focused on achieving unnatural base pair incorporation rather than on the incorporation of "useful" bases with specialized chemical functionalities in mind (e.g., metal chelators, novel functional groups etc.).
        - [Potential Solution]: Potentially useful chemical functionalities should be enumerated.
- **5 years: Improved *in vitro* manipulation of unnatural nucleic acids.**
    - [Bottleneck]: Evolution of unnatural aptamers, allosteric regulators and aptazymes requires reverse transcription in order to complete cycles of synthesis and selection.
        - [Potential Solution]: Evolve/engineer reverse transcriptases that can incorporate the array of unnatural nucleotides and be able to be easily adjusted to incorporate additional chemistries as they are developed.
- **5 years: Expansion of unnatural nucleotide toolkit.**
    - [Bottleneck]: At present, transcription and translation of DNA containing unnatural base pairs has been achieved only in the context of a single specialized unnatural base pair and only in *E. coli.*
    - [Bottleneck]: Success in this endeavor required extensive optimization.
        - [Potential Solution]: Begin to explore alternative (previously explored) unnatural base pairs in the context of the optimized conditions, especially

ones that do not perturb the double helical structure of DNA and can be incorporated in any sequence context.

- **10 years: Biosynthesis of unnatural nucleotides.**
  - [Bottleneck]: Current unnatural base pairs must be chemically synthesized, which could limit the ability to use them in large scale applications.
    - [Potential Solution]: Engineered biosynthetic pathways capable of generating non-natural bases *in vivo.*
  - [Bottleneck]: Considerations must be made regarding the risks and impacts of release of non-natural nucleic acids into nature.
- **20 years: Organisms capable of full replication, maintenance, and transcription of a plasmid or artificial chromosome made up entirely of unnatural bases.**
  - [Bottleneck]: Transcripts should confer useful function for the organism and also be made entirely of unnatural bases.

[Breakthrough Capability 2]: Expanded genetic code systems for translation of >100-amino acid proteins containing fully-unnatural amino acids, and proteins with at least four, distinct unnatural amino acid building blocks.

- **2 years: Create proteins that are capable of gaining new, therapeutically-useful activities through unnatural amino acids.**
  - [Bottleneck]: Efficiency and scale of protein expression with expanded genetic code systems.
    - [Potential Solution]: Specialized strains, specialized tRNA/aminoacyl-tRNA synthetase systems, specialized ribosomes, and genomically-recoded organisms.
- **5 years: Efficient biosynthesis of proteins containing three or more distinct unnatural amino acid building blocks.**
  - [Bottleneck]: Not enough free codons to hijack for unnatural amino acid building blocks.
    - [Potential Solution]: Genomically-recoded strains that free up redundant codons, orthogonal ribosomes that can utilize special tRNAs for special messages, or orthogonal organellar genetic codes.
  - [Bottleneck]: Not enough mutually orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pairs for unnatural building block incorporation.
    - [Potential Solution]: Design or scalable evolution of new mutually orthogonal sets of aaRS/tRNA pairs for genomically-recoded organisms or for orthogonal ribosomes; hijacking of organellar genetic codes and associated aaRS/tRNA pairs.
- **10 years: Biosynthesis of unnatural amino acids.**
  - [Bottleneck]: Current unnatural amino acids are chemically synthesized in most cases, which could limit the ability to use them in large scale applications.
    - [Potential Solution]: Engineered biosynthetic pathways capable of generating unnatural amino acids *in vivo*.

- **20 years: Templated biosynthesis and evolution of new polymers with large user-selected sets of unnatural building blocks *in vivo*.**
    - [Bottleneck]: Ribosome, elongation factor, and aaRS/tRNA engineering for new building blocks and polymer linkage chemistries beyond peptide bonds.

Goal 3: Holistic, integrated design of multi-part genetic systems (i.e., circuits and pathways).

[Current State-of-the-Art]: A long-standing goal of molecular engineering is to create components that can control genetic processes. There are commonly used predictive models that can design short genetic parts to control gene expression processes both within cells and in cell-free *in vitro* systems. These parts are then combined into larger genetic systems (operons, regulons) to create desired cellular functions, including sensors, genetic circuits, transporters, multi-enzyme metabolic pathways, organelle compartments, and orthogonal expression systems. These multi-part genetic systems are particularly useful, as they are used by natural organisms as a central information processing component to sense and respond to changing internal and external conditions. Thus, being able to *de novo* design or engineer these systems offers many points of control of fundamental cellular processes. However, there are several challenges that have been encountered in trying to develop multi-part genetic systems. For example, there are many coupled interactions, between adjacent parts or between distant genetic modules, that alter system function in unpredictable and undesired ways. Therefore, new approaches are needed to correctly design large genetic systems, taking into account these poorly understood mechanisms, within an even larger genomic background.

[Breakthrough Capability 1]: Design of highly-stable, large genetic systems (genomes) with targeted expression levels in a host organism or cell type, incorporating system-wide effects.

- **2 years: Incorporate gene expression interactions into predictable design of prokaryotic genetic systems.**
    - [Bottleneck]: Incomplete understanding of coupled expression dynamics within single or multi-protein genetic systems, accounting for changes in transcription, translation, and mRNA decay.
        - [Potential Solution]: Develop sequence-to-function models of transcription initiation, elongation, and transcription factor binding dynamics.
        - [Potential Solution]: Develop sequence-to-function models of mRNA decay rates, incorporating coupled interactions between transcription, translation, and the several mRNA decay pathways.
        - [Potential Solution]: Develop sequence-to-function models of mRNA translation, including improved ribosome-mRNA free energy models as well as incorporating the kinetics of RNA folding and Ribosome Drafting mechanisms.
        - [Potential Solution]: Incorporating design rules quantifying criteria for optimal functioning of genetic systems, including increased genetic stability and minimal resource load.

        - [Potential Solution]: Develop models of host resource limitations that can indirectly couple gene expression processes.
- **5 years: Incorporate gene expression interactions into predictable design of eukaryotic genetic systems.**
    - [Bottleneck]: Incomplete understanding of coupled expression dynamics within single or multi-protein genetic systems, accounting for changes in epigenetics, transcription, translation, mRNA decay, and splicing.
        - [Potential Solution]: Develop sequence-to-function models of transcription initiation, including the dynamics of epigenetic modifications.
        - [Potential Solution]: Develop sequence-to-function models of eukaryotic mRNA processing including capping, alternative splicing, poly-adenylation, packaging and nuclear export.
        - [Potential Solution]: Develop sequence-to-function models of mRNA decay rates, incorporating coupled interactions between transcription, translation, and the several mRNA decay pathways.
        - [Potential Solution]: Develop sequence-to-function models of mRNA translation, including pausing, upstream open reading frames as well as incorporating the kinetics of RNA folding.
        - [Potential Solution]: Incorporating design rules quantifying criteria for optimal functioning of genetic systems, including increased genetic stability and minimal resource load.
- **10 years: Discovery and characterization of mechanistic interactions at the systems-level affecting protein activities inside cells.**
    - [Bottleneck]: A variety of poorly characterized systems-level interactions affect protein activities inside cells, including expression resource allocation, metabolic resource allocation, local changes to the cellular environment (pH, crowding, etc.), and local changes to protein folding and activity (e.g., co-translational folding, allosteric regulation).
        - [Potential Solution]: Development of layered mechanistic models accounting for changes in expression resource & metabolite levels (e.g., RNAP, TFs, EFs, NTPs, ribosomes, tRNAs, amino acids, and chaperones).
        - [Potential Solution]: Modeling and experimental methods to elucidate allosteric regulation of individual enzymes, and RNA/protein regulators.
- **20 years: Whole-tissue or whole-cell, nucleotide-resolution simulations encompassing several layers of models predicting gene regulatory, metabolic, and system-level behaviors.**
    - [Bottleneck]: Incomplete enumeration and quantification of biophysical and chemical interactions controlling system-wide behaviors.
        - [Potential Solution]: Systematic development and validation of models quantifying each layer of interactions, followed by critical testing of simulation predictions incorporating multiple modeling layers; a “Genome Calculator”.

[Breakthrough Capability 2]: Ability to rationally engineer sensor suites, genetic circuits, metabolic pathways, signaling cascades, and cell differentiation pathways.

- **2 years: Reliable engineering of genetic circuits with more than ten regulators for sophisticated computations.**
    - [Bottleneck]: The availability of orthogonal, programmable, non-repetitive regulators with desired gene regulatory effects.
        - [Potential Solution]: Toolboxes of highly non-repetitive, regulated promoters, including large sets of mutually orthogonal promoter/regulator pairs.
        - [Potential Solution]: Toolboxes of highly non-repetitive transcription factors or CRISPR genetic parts, including mutually orthogonal transcription factors and CRISPR systems.
        - [Potential Solution]: Toolboxes of highly non-repetitive RNA regulators of gene expression that can control transcription, translation and mRNA degradation.
        - [Potential Solution]: Toolboxes of highly non-repetitive and mutually orthogonal recombinases.
        - [Potential Solution]: Predictive models at nucleotide to systems resolution, capable of predicting how combinations of genetic parts lead to desired computations (e.g., analog, digital, signal processing, pattern recognition).
- **5 years: Reliable engineering of novel, many-enzyme pathways utilizing combinations of bioprospected enzymes with well-characterized kinetics.**
    - [Bottleneck]: Predicting enzymatic reactions and kinetics from amino acid/nucleic acid sequence remains difficult. Many protein or RNA-based enzymes are promiscuous with varying substrate selectivities.
        - [Potential Solution]: The kinetics and substrate selectivities of large families of bio-prospected enzymes could be characterized to identify rules for quantifying enzyme-substrate promiscuity.
        - [Potential Solution]: Enzyme databases could be greatly expanded to incorporate predicted enzyme-substrate promiscuities, enabling the design of more novel and exotic multi-enzyme pathways.
- **5 years: Five-time improvement and expansion of inducers/promoters for model organisms that respond to environmental inputs and any intracellular metabolite.**
    - [Bottleneck]: We are still relatively limited to chemical inducers of expression.
        - [Potential Solution]: Catalog known signals that microbes respond to including potential genetic parts necessary.
        - [Potential Solution]: Expand technology to include other non-conventional inputs (e.g., light, electricity); two-year goals would include focus on improving known inputs (e.g., optogenetic tools) and conducting initial tests with new technologies.
        - [Potential Solution]: Develop RNA and protein biosensors that respond to a variety of non-natural chemical inputs.

- **5 years: Utilize machine-learning approaches to use the vast amount of uncurated literature results within pathway design.**
    - [Bottleneck]: The performance of pathway design software critically dependent on the quality and amount of enzymatic and biochemical data available; although curated databases such as MetaCyc (Caspi et al., 2018) (MetaCyc is available at https://metacyc.org) or BRENDA (Jeske, Placzek, Schomburg, Chang, & Schomburg, 2019) (BRENDA is available at https://www.brenda-enzymes.org/) exist for enzymes involved in metabolism, a large amount of previous experimental results are "buried" in the literature, including records of enzyme specificity and other characterizations, and therefore impossible to utilize for pathway design.
        - [Potential Solution]: Develop natural language processing (NLP) and machine learning approaches to extract and characterize the relevant information from the literature of the last four decades.
- **10 years: Creation of optogenetic tools for *in vivo* RNA post-transcriptional control to allow for easy control of any gene expression process through mRNA (You & Jaffrey, 2015).**
    - [Bottleneck]: There are currently no optogenetic RNA aptamers that can be used *in vivo*.
        - [Potential Solution]: The development of light activated chromophore ligands that are compatible with cellular biochemistry and can be easily synthesized by the cell.
        - [Potential Solution]: The design or evolution of aptamers that bind to only one conformation of the chromophore, thus giving the basis for optogenetic switching of RNA structure.
        - [Potential Solution]: Enzymatic processes to covalently link RNA sequences to chromophores in a selective manner so that optogenetic chromophore transitions can be cycled as is the case with optogenetic protein mechanisms.
        - [Potential Solution]: Take advantage of emerging optogenetic variants of CRISPR/Cas systems, now ported (via Cas13-like effectors) to bind to RNA with light-controlled affinity.
    - [Bottleneck]: There are currently no optogenetic RNA regulatory mechanisms that can be used *in vivo*.
        - [Potential Solution]: Couple the above developed optogenetic aptamer system to RNA regulatory mechanisms that can control a range of gene expression processes.
- **10 years: Reliable expression of redesigned synthases to produce secondary metabolites, including polyketides and non-ribosomal peptides.**
    - [Bottleneck]: Many important natural products are synthesized by modular multi-enzyme synthases but these complexes are difficult to express, particularly so in engineered forms where specific modules are recombined/designed in order to alter productive specificity.

        - [Potential Solution]: Structural analysis, driven by advances in cryo-EM, to inform structural modeling of hierarchical assembly rules.
        - [Potential Solution]: High-throughput assembly/assay of natural polyketide synthases (PKSs) and non-ribosomal peptides (NRPs) to build functional PKSs/NRPSs that produce unnatural molecules and incorporate successes/failures into model.
- **10 years: Computational design of protein-ligand and RNA-ligand interfaces suitable for engineering protein-based or RNA-based sensors.**
    - [Bottleneck]: Potential energy models for nucleic acid interactions require improvements in charge screening and ionic effects.
    - [Bottleneck]: The number of RNA-ligand interactions that have been characterized at atomic resolution is sparse, prohibiting informatics-based approaches to design RNA-ligand interactions.
        - [Potential Solution]: Develop and apply high throughput methods to characterize RNA-ligand interactions.
    - [Bottleneck]: Many candidate RNA/protein sequences must be experimentally characterized to find one that binds well to a targeted ligand.
        - [Potential Solution]: Maximally informative measurements of designed proteins and RNAs could be utilized to further improve potential energy functions and model predictions.
- **20 years: Simultaneous, tunable, timed expression of many transcription factors controlling mammalian cell state.**
    - [Bottleneck]: Insufficient mapping between transcription factor expression levels, regulated promoter activities, epigenetic modifications, and overall regulation of protein and metabolite levels.
    - [Bottleneck]: Insufficient temporal expression control over multi-protein genetic systems in eukaryotic systems, particularly mammalian systems.
        - [Potential Solution]: Development of improved genetic circuits controlling temporal expression of desired proteins in mammalian cells, stimulated by small molecule, cell contact, or cell cycle conditions.
        - [Potential Solution]: Systematic characterization of human, genome-wide transcription factor binding sites, affinities, and gene regulatory effects.

Goal 4: Integrated design of RNA-based regulatory systems for cellular control and information processing.

[Current State-of-the-Art]: There is a unique opportunity to construct information processing genetic circuitry out of RNA molecules versus proteins due to three factors: 1) a deep history of molecular computation with nucleic acids *in vitro* (Cherry & Qian, 2018; Qian & Winfree, 2011); 2) the concept of nucleic acid strand displacement as a simple, yet highly modular and programmable mechanism to maintain and propagate changes of molecular state; and 3) the emergence of RNA secondary structure design tools that can implement new designs using these paradigms. Within the realm of *in vitro* nucleic acid 'circuits', there has been great progress in developing *in vitro* systems that can process information in much the same way as

genetic regulatory circuits. In these systems, chemical state is defined as the concentration of specific nucleic acid species, and these states can be changed through designed interactions that can process information, such as logic evaluation (Seelig, Soloveichik, Zhang, & Winfree, 2006), or even perform complex computational tasks (Cherry & Qian, 2018; Qian & Winfree, 2011). One specific paradigm for programming these interactions is via nucleic acid strand displacement - a method by which specific nucleic acid (DNA or RNA) hybrids can exchange strands with each other to change the abundance of specific hybrid species. They offer a powerful paradigm for molecular computation in that they leverage the existing DNA/RNA structure design tools and can be abstracted into high-level programming languages. In addition, RNA has additional advantages in being able to enhance functional properties through expanded non-natural nucleic acid chemistries, and by leveraging the wide range of RNA modifications present especially in eukaryotic systems. Overall, we know a great deal more about the fundamental principles of programming reaction cascades with RNAs than we do for proteins, due to the latter being governed by specific protein interactions that are harder to generalize. We therefore highlight RNA circuit design as its own goal while covering protein-based circuit design in ***Goal 3: Holistic, integrated design of multi-part genetic systems (i.e., circuits and pathways)***.

While there is great promise in RNA circuit design, and some progress made in porting the most basic elements of strand displacement into designed RNA regulators of gene expression (toehold switches (Green, Silver, Collins, & Yin, 2014) and small transcription activating RNAs (STARs) (Chappell, Westbrook, Verosloff, & Lucks, 2017)), the full repertoire of strand displacement capability has so far not been fully ported to living cellular systems. This represents the major challenge of this goal, with the accordingly great opportunity to expand the programmable molecular control over cellular systems through porting strand displacement into cells.

[Breakthrough Capability 1]: Porting nucleic acid strand displacement technology into cellular systems with RNA instantiations.

- **2 years: RNA implementation of strand displacement cascades in bacteria.**
    - [Bottleneck]: Difficulties in designing RNAs that are stable and robustly fold into desired structures needed for strand displacement within the cell.
        - [Potential Solution]: Identification of new RNA-protecting motifs compatible with strand displacement.
        - [Potential Solution]: Development and use of RNA-targeting CRISPR nucleases (e.g., dCas13a) for site-specific modification of RNA bases to enhance stability.
        - [Potential Solution]: Development of alternative, strand-displacement motifs that are more compatible with the cellular context.
        - [Potential Solution]: Further advances of in-cell structure probing techniques to inform design principles for RNA strand displacement regulators.

        - [Potential Solution]: Evolutionary optimization using RNA-targeting CRISPR systems to activate mRNAs or regulatory outputs to activate gene expression for selection.
    - [Bottleneck]: Cellular RNA cascades require large concentration ratios to be effective.
        - [Potential Solution]: Design and implement RNA cascades that leverage catalytic interactions.
        - [Potential Solution]: Develop design rules to integrate the components of current multi-strand cascades into a single strand.
    - [Bottleneck]: Cellular RNAs can have high secondary structures that make them challenging to use as inputs for strand displacement cascades.
        - [Potential Solution]: Implement RNA signal transducers that modify cellular RNA structures to be compatible.
    - [Bottleneck]: Filling gaps in RNA parts and their performance. There remains a high level of discrepancy in RNA part performance, particularly across different cell types and in the context of intracellular conditions. For example, dynamic range (On/Off state function) is a critical parameter for gene regulators that determines the stringency of control and their utility for constructing gene networks but varies in different contexts for many existing parts.
        - [Potential Solution]: Creation of high-performing RNA parts explicitly characterized for robustness and engineered and tested at magnesium concentrations that are physiologically relevant for intracellular use.
- **5 years: RNA implementation of strand displacement cascades in eukaryotic systems.**
    - [Bottleneck]: Eukaryotic RNAs are subject to more complex regulation and are subject to additional processing (e.g., nuclear transport, capping, polyadenylation, splicing, microRNA pathways).
        - [Potential Solution]: Develop the equivalent of compact and simplified bacterial RNA regulators that distill the complexity of these interactions to function robustly in mammalian systems.
    - [Bottleneck]: RNA concentrations in eukaryotic cells are lower than those in prokaryotic cells and reduce RNA-RNA interactions.
        - [Potential Solution]: Implement RNA strand displacement cascades using RNAs that are confined to the nucleus or that are tagged to remain in the nucleus.
        - [Potential Solution]: Make use of catalytic RNA systems as described above.
- **5 years: Engineer ‘universal’ computational strand displacement architectures using strand displacement in bacteria.**
    - [Bottleneck]: The most sophisticated strand-displacement architectures require reversible interactions for error correction.
        - [Potential Solution]: Leverage existing architectures by implementing strand displacement that allow reversible interactions, or develop a new

strand-displacement architecture that is amenable to cellular requirements.
    - [Bottleneck]: Strand-displacement neural networks require precise control over the relative stoichiometry of strands in the network.
        - [Potential Solution]: Develop methods to encode the complete strand-displacement RNA network in a single strand of RNA.
    - [Bottleneck]: Signal conditioning methods are required in the propagating cascade for proper network function.
        - [Potential Solution]: Engineer methods of RNA signal thresholding and amplification required for signal conditioning.
- **10 years: Engineer computational RNA strand displacement networks in mammalian systems.**
    - [Bottleneck]: Methods implemented for encoding RNA strand-displacement networks in bacteria may no longer function in mammalian cells.
        - [Potential Solution]: Adopt RNA network encoding that is specific to mammalian cells that accommodates differences in RNA processing, polyadenylation, and potential changes in ribozyme activity.
    - [Bottleneck]: Lower concentrations of RNAs in mammalian cells prevents proper network function.
        - [Potential Solution]: Encode the RNA strand-displacement network within a single strand of RNA.
- **20 years: Computational design of RNA strand displacement neural networks that process the transcriptome.**
    - [Bottleneck]: Strand displacement networks need to take as input transcriptomic RNAs that are improperly folded and processed, partially degraded, and interacting with other cellular machinery.
        - [Potential Solution]: Design of strand displacement interaction architectures that are robust to errors in inputs.
        - [Potential Solution]: Design of error correction computational layers that can process ‘messy’ input signatures.
- **20 years: Engineer RNA neural networks that dynamically reprogram cell state.**
    - [Bottleneck]: Strand displacement networks cannot interface with systems that can regulate or alter the cellular state.
        - [Potential Solution]: Design of strand displacement output systems that can interact with (i.e., activate and/or repress) a range of native or engineered cellular systems.

[Breakthrough Capability 2]: Porting successes in computationally designed bacterial RNA-based genetic regulators into eukaryotic and mammalian systems.

- **2 years: First generation eukaryotic RNA-based gene regulators that utilize RNA:RNA interactions and/or strand-displacement and achieve 10-fold change in gene expression.**
    - [Bottleneck]: Current state-of-the-art RNA-based gene regulators in bacteria utilize RNA structures to control transcription (STARs) and translation (Toeholds,

and small RNAs) that are fundamentally different in eukaryotes and do not directly port.
    - [Potential solution]: Understand how initiation, elongation and termination of both transcription and translation can be manipulated through formation of RNA structures and interactions.
    - [Potential solution]: Uncover how natural RNA processing (splicing, polyadenylation, 5' capping) specific to eukaryotes can be regulated through formation of RNA structures and interactions.
    - [Potential solution]: Investigate orthogonal mechanisms of gene expression and RNA processing in eukaryotic cells, for example, that use machinery from naturally occurring viruses.
- **2 years: Creation of RNA modification machinery that allows programmable site-specific modifications of RNA, focusing on naturally abundant modifications (N6-methyl adenosine, 2'-O-methylation, pseudouridine).**
    - [Bottleneck]: We currently have no ability to program RNA modifications inside living cells and natural modification machines lack sequence flexibility.
        - [Potential solution]: Create CRISPR-based systems able to perform guide RNA directed post-transcriptional modification of RNA.
        - [Potential solution]: Investigate natural RNA modification machinery to identify programmable and/or sequence flexible mechanisms.
- **5 years: Second generation eukaryotic RNA-based gene regulators that are suitable for computational design to create libraries that are highly-orthogonal and high-performing, achieving 100's-fold change in gene expression.**
    - [Bottleneck]: Design rules of how specific RNA structures and sequences control gene expression and RNA processing in eukaryotes is poorly understood.
        - [Potential solution]: Couple DNA library synthesis with high-throughput screening to rapidly explore sequence-function landscapes of different cellular process.
- **5 years: Use RNA modifications for programming or fine-tuning RNA functions.** Examples include using modifications to dynamically control how RNAs interact, control gene expression, and form nanostructures.
    - [Bottleneck]: While RNA modifications are known to be abundant and diverse, we have almost no idea about their functional consequences. This fundamentally limits our ability to harness modified nucleotides as building blocks or control points.
        - [Potential solution]: Combine *in vitro* synthesized and modified RNAs with cell-free protein synthesis systems or purified reactions to systematically investigate effects of modifications.
        - [Potential solution]: Apply programmable RNA modification tools *in vivo* to determine functional consequences of RNA modifications.
        - [Potential solution]: Use next-generation sequencing technologies and statistical analysis to understand how modifications perturb endogenous RNA processing, structure formation, interactions and translation.

- **10 years: Expand RNA modification apparatus to modify non-natural RNA alphabets to enhance their functional properties.**
    - [Bottleneck]: It is unknown how RNA modification machinery would tolerate non-natural RNA alphabets.
        - [Potential solution]: Engineer and evolve RNA modification machinery to diversify substrate specificity.
- **20 years: Engineering enzymes that can perform non-natural RNA modifications to further expand the chemical repertoire of what is possible and extend RNA ligand recognition, catalysis and genetic control.**
    - [Bottleneck]: Current non-natural RNA modifications have to be synthesized *in vitro*, and cannot be produced inside cells.
        - [Potential Solution]: Engineer protein enzymes or ribozymes capable of generating non-natural RNA modifications in cells.
        - [Potential Solution]: Engineer biosynthetic pathways that can produce substrates of non-natural RNA modification reactions *in vivo*.

References

Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RL, Das R, Baker D, Kuhlman B, Kortemme T, Gray, J. J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, *13*(6), 3031–3048. https://doi.org/10.1021/acs.jctc.7b00125

Ali, H., & Khan, E. (2018). Trophic transfer, bioaccumulation, and biomagnification of non-essential hazardous heavy metals and metalloids in food chains/webs—Concepts and implications for wildlife and human health. *Human and Ecological Risk Assessment: An International Journal*, 1–24. https://doi.org/10.1080/10807039.2018.1469398

AlQuraishi, M. (2019). AlphaFold at CASP13. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btz422

Badran, A. H., & Liu, D. R. (2015). In vivo continuous directed evolution. *Current Opinion in Chemical Biology*, *24*, 1–10. https://doi.org/10.1016/j.cbpa.2014.09.040

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., & Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, *50*(21), 4402–4410. https://doi.org/10.1021/bi2002289

Bier, E., Harrison, M. M., O'Connor-Giles, K. M., & Wildonger, J. (2018). Advances in Engineering the Fly Genome with the CRISPR-Cas System. *Genetics*, *208*(1), 1–18. https://doi.org/10.1534/genetics.117.1113

Blind, M., & Blank, M. (2015). Aptamer selection technology and recent advances. *Molecular Therapy. Nucleic Acids*, *4*, e223. https://doi.org/10.1038/mtna.2014.74

Boeing, P., Leon, M., Nesbeth, D. N., Finkelstein, A., & Barnes, C. P. (2018). Towards an Aspect-Oriented Design and Modelling Framework for Synthetic Biology. *Processes (Basel, Switzerland)*, *6*(9), 167. https://doi.org/10.3390/pr6090167

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature Biotechnology*, *36*(10), 1005–1015. https://doi.org/10.1038/nbt.4238

Carlson, P. D., & Lucks, J. B. (2019). Elements of RNA design. *Biochemistry*, *58*(11), 1457–1459. https://doi.org/10.1021/acs.biochem.8b01129

Carothers, J. M., Goler, J. A., Juminaga, D., & Keasling, J. D. (2011). Model-driven engineering of RNA devices to quantitatively program gene expression. *Science*, *334*(6063), 1716–1719. https://doi.org/10.1126/science.1212209

Carothers, J. M., Oestreich, S. C., Davis, J. H., & Szostak, J. W. (2004). Informational complexity and functional activity of RNA structures. *Journal of the American Chemical Society*, *126*(16), 5130–5137. https://doi.org/10.1021/ja031504a

Chappell, J., Westbrook, A., Verosloff, M., & Lucks, J. B. (2017). Computational design of small transcription activating RNAs for versatile and dynamic gene regulation. *Nature Communications*, *8*(1), 1051. https://doi.org/10.1038/s41467-017-01082-6

Cherry, K. M., & Qian, L. (2018). Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, *559*(7714), 370–376. https://doi.org/10.1038/s41586-018-0289-6

Clark, D. S., & Blanch, H. W. (1997). *Biochemical Engineering (Chemical Industries)* (2nd ed., p. 716). New York, New York: Crc Press.

Costello, Z., & Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *Npj Systems Biology and Applications*, *4*, 19. https://doi.org/10.1038/s41540-018-0054-3

Cox, J. C., Hayhurst, A., Hesselberth, J., Bayer, T. S., Georgiou, G., & Ellington, A. D. (2002). Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. *Nucleic Acids Research*, *30*(20), e108. https://doi.org/10.1093/nar/gnf107

Cuperus, J. T., Groves, B., Kuchina, A., Rosenberg, A. B., Jojic, N., Fields, S., & Seelig, G. (2017). Deep learning of the regulatory grammar of yeast 5' untranslated regions from 500,000 random sequences. *Genome Research*, *27*(12), 2015–2024. https://doi.org/10.1101/gr.224964.117

Das, R., Karanicolas, J., & Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. *Nature Methods*, *7*(4), 291–294. https://doi.org/10.1038/nmeth.1433

Davey, J. A., Damry, A. M., Goto, N. K., & Chica, R. A. (2017). Rational design of proteins that exchange on functional timescales. *Nature Chemical Biology*, *13*(12), 1280–1285. https://doi.org/10.1038/nchembio.2503

de Kok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt D, Shapland EB, Serber Z, Dean J, Newman JD, Chandran, S. S. (2014). Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synthetic Biology*, *3*(2), 97–106. https://doi.org/10.1021/sb4001992

Dehingia, M., Adak, A., & Khan, M. R. (2019). Ethnicity-Influenced Microbiota: A Future Healthcare Perspective. *Trends in Microbiology*, *27*(3), 191–193. https://doi.org/10.1016/j.tim.2019.01.002

DeLoache, W. C., Russ, Z. N., & Dueber, J. E. (2016). Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nature Communications*, *7*, 11152. https://doi.org/10.1038/ncomms11152

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, *346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, *346*(6287), 818–822. https://doi.org/10.1038/346818a0

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *Plos One*, *3*(11), e3647. https://doi.org/10.1371/journal.pone.0003647

Espah Borujeni, A., Mishler, D. M., Wang, J., Huso, W., & Salis, H. M. (2016). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Research*, *44*(1), 1–13. https://doi.org/10.1093/nar/gkv1289

Espah Borujeni, A., & Salis, H. M. (2016). Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society*, *138*(22), 7016–7023. https://doi.org/10.1021/jacs.6b01453

Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, *472*(7344), 499–503. https://doi.org/10.1038/nature09929

Fan, W., Guo, Q., Liu, C., Liu, X., Zhang, M., Long, D., Xiang, Z., Zhao, A. (2018). Two mulberry phytochelatin synthase genes confer zinc/cadmium tolerance and accumulation in transgenic Arabidopsis and tobacco. *Gene*, *645*, 95–104. https://doi.org/10.1016/j.gene.2017.12.042

Farré, G., Blancquaert, D., Capell, T., Van Der Straeten, D., Christou, P., & Zhu, C. (2014). Engineering complex metabolic pathways in plants. *Annual Review of Plant Biology*, *65*, 187–223. https://doi.org/10.1146/annurev-arplant-050213-035825

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner N, Wilson ML, Adam L, Anderson JC, Bartley BA, Beal J, Chandran D, Chen J, Densmore D, Endy D, Grünberg R, Hallinan J, Hillson NJ, Johnson JD, Kuchinsky A, Lux M, Misirli G, Peccoud J, Plahar HA, Sirin E, Stan GB, Villalobos A, Wipat A, Gennari JH, Myers CJ, Sauro, H. M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology*, *32*(6), 545–550. https://doi.org/10.1038/nbt.2891

Gantz, V. M., & Bier, E. (2015). Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, *348*(6233), 442–444. https://doi.org/10.1126/science.aaa5945

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), E6736–43. https://doi.org/10.1073/pnas.1521077112

Gibson, D. G. (2011). Enzymatic assembly of overlapping DNA fragments. *Methods in Enzymology*, *498*, 349–361. https://doi.org/10.1016/B978-0-12-385120-8.00015-2

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gilbert, J. A., & Melton, L. (2018). Verily project releases millions of factory-reared mosquitoes. *Nature Biotechnology*, *36*(9), 781–782. https://doi.org/10.1038/nbt0918-781a

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, *154*(2), 442–451. https://doi.org/10.1016/j.cell.2013.06.044

Goldsmith, M., & Tawfik, D. S. (2017). Enzyme engineering: reaching the maximal catalytic efficiency peak. *Current Opinion in Structural Biology*, *47*, 140–150. https://doi.org/10.1016/j.sbi.2017.09.002

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews. Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: de-novo-designed regulators of gene expression. *Cell*, *159*(4), 925–939. https://doi.org/10.1016/j.cell.2014.10.002

Grunwald, H. A., Gantz, V. M., Poplawski, G., Xu, X.-R. S., Bier, E., & Cooper, K. L. (2019). Super-Mendelian inheritance mediated by CRISPR-Cas9 in the female mouse germline. *Nature*, *566*(7742), 105–109. https://doi.org/10.1038/s41586-019-0875-2

Haitjema, C. H., Solomon, K. V., Henske, J. K., Theodorou, M. K., & O'Malley, M. A. (2014). Anaerobic gut fungi: Advances in isolation, culture, and cellulolytic enzyme discovery for biofuel production. *Biotechnology and Bioengineering*, *111*(8), 1471–1482. https://doi.org/10.1002/bit.25264

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., & Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, *560*(7717), 248–252. https://doi.org/10.1038/s41586-018-0384-8

Ham, T. S., Dmytriv, Z., Plahar, H., Chen, J., Hillson, N. J., & Keasling, J. D. (2012). Design, implementation and practice of JBEI-ICE: an open source biological part registry platform and tools. *Nucleic Acids Research*, *40*(18), e141. https://doi.org/10.1093/nar/gks531

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, El Assal DC, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Ben Guebila M, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, El Assal LP, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Aragón Artacho FJ, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson BØ, Thiele I, Vlasov, V. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, *14*(3), 639–702. https://doi.org/10.1038/s41596-018-0098-2

Higgins, S. A., & Savage, D. F. (2018). Protein science by DNA sequencing: how advances in molecular biology are accelerating biochemistry. *Biochemistry*, *57*(1), 38–46. https://doi.org/10.1021/acs.biochem.7b00886

Hillson, N., Caddick, M., Cai, Y., Carrasco, J. A., Chang, M. W., Curach, N. C., Bell DJ, Le Feuvre R, Friedman DC, Fu X, Gold ND, Herrgård MJ, Holowko MB, Johnson JR, Johnson RA, Keasling JD, Kitney RI, Kondo A, Liu C, Martin VJJ, Menolascina F, Ogino C, Patron NJ, Pavan M, Poh CL, Pretorius IS, Rosser SJ, Scrutton NS, Storch M, Tekotte H, Travnik E, Vickers CE, Yew WS, Yuan Y, Zhao H, Freemont, P. S. (2019). Building a global alliance of biofoundries. *Nature Communications*, *10*(1), 2040. https://doi.org/10.1038/s41467-019-10079-2

Hooker, S. E., Woods-Burnham, L., Bathina, M., Lloyd, S. M., Gorjala, P., Mitra, R., Nonn L, Kimbro KS, Kittles, R. (2019). Genetic ancestry analysis reveals misclassification of commonly used cancer cell lines. *Cancer Epidemiology, Biomarkers & Prevention*. https://doi.org/10.1158/1055-9965.EPI-18-1132

Hoshika, S., Leal, N. A., Kim, M.-J., Kim, M.-S., Karalkar, N. B., Kim, H.-J., Bates AM, Watkins NE, SantaLucia HA, Meyer AJ, DasGupta S, Piccirilli JA, Ellington AD, SantaLucia J, Georgiadis MM, Benner, S. A. (2019). Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science*, *363*(6429), 884–887. https://doi.org/10.1126/science.aat0971

Hsiao, V., Cheng, A., & Murray, R. M. (2016). *Design and application of stationary phase combinatorial promoters*. MurrayWiki. Retrieved from http://www.cds.caltech.edu/~murray/preprints/hcm16-seed_s.pdf

Huang, A., Nguyen, P. Q., Stark, J. C., Takahashi, M. K., Donghia, N., Ferrante, T., Dy AJ, Hsu KJ, Dubner RS, Pardee K, Jewett MC, Collins, J. J. (2018). BioBits™ Explorer: A modular synthetic biology education kit. *Science Advances*, *4*(8), eaat5105. https://doi.org/10.1126/sciadv.aat5105

Hughes, R. A., & Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harbor Perspectives in Biology*, *9*(1). https://doi.org/10.1101/cshperspect.a023812

Jakobson, C. M., Chen, Y., Slininger, M. F., Valdivia, E., Kim, E. Y., & Tullman-Ercek, D. (2016). Tuning the catalytic activity of subcellular nanoreactors. *Journal of Molecular Biology*, *428*(15), 2989–2996. https://doi.org/10.1016/j.jmb.2016.07.006

Jeske, L., Placzek, S., Schomburg, I., Chang, A., & Schomburg, D. (2019). BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research*, *47*(D1), D542–D549. https://doi.org/10.1093/nar/gky1048

Johns, N. I., Gomes, A. L. C., Yim, S. S., Yang, A., Blazejewski, T., Smillie, C. S., Smith MB, Alm EJ, Kosuri S, Wang, H. H. (2018). Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nature Methods*, *15*(5), 323–329. https://doi.org/10.1038/nmeth.4633

Jones, H. P., Holmes, N. D., Butchart, S. H. M., Tershy, B. R., Kappes, P. J., Corkery, I., Aguirre-Muñoz A, Armstrong DP, Bonnaud E, Burbidge AA, Campbell K, Courchamp F, Cowan PE, Cuthbert RJ, Ebbert S, Genovesi P, Howald GR, Keitt BS, Kress SW, Miskelly CM, Oppel S, Poncet S, Rauzon MJ, Rocamora G, Russell JC, Samaniego-Herrera A, Seddon PJ, Spatz DR, Towns DR, Croll, D. A. (2016). Invasive mammal eradication on islands results in substantial conservation gains. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(15), 4033–4038. https://doi.org/10.1073/pnas.1521179113

Karim, A. S., & Jewett, M. C. (2016). A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering*, *36*, 116–126. https://doi.org/10.1016/j.ymben.2016.03.002

Kedzierska, K., Valkenburg, S. A., Doherty, P. C., Davenport, M. P., & Venturi, V. (2012). Use it or lose it: establishment and persistence of T cell memory. *Frontiers in Immunology*, *3*, 357. https://doi.org/10.3389/fimmu.2012.00357

Kim, E. Y., & Tullman-Ercek, D. (2014). A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. *Biotechnology Journal*, *9*(3), 348–354. https://doi.org/10.1002/biot.201300391

Kong, W., Meldgin, D. R., Collins, J. J., & Lu, T. (2018). Designing microbial consortia with defined social interactions. *Nature Chemical Biology*, *14*(8), 821–829. https://doi.org/10.1038/s41589-018-0091-7

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods*, *11*(5), 499–507. https://doi.org/10.1038/nmeth.2918

Kuo-chen, C., & Shou-ping, J. (1974). Studies on the rate of diffusion-controlled reactions of enzymes. Spatial factor and force field factor. *Scientia Sinica*, *27*(5), 664–680.

Kyrou, K., Hammond, A. M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A. K., Nolan, T., Crisanti, A. (2018). A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*, *36*(11), 1062–1066. https://doi.org/10.1038/nbt.4245

Lane, R. S., & Quistad, G. B. (1998). Borreliacidal Factor in the Blood of the Western Fence Lizard (Sceloporus occidentalis). *The Journal of Parasitology*, *84*(1), 29. https://doi.org/10.2307/3284524

Leistra, A. N., Amador, P., Buvanendiran, A., Moon-Walker, A., & Contreras, L. M. (2017). Rational modular RNA engineering based on in vivo profiling of structural accessibility. *ACS Synthetic Biology*, *6*(12), 2228–2240. https://doi.org/10.1021/acssynbio.7b00185

Leistra, A. N., Curtis, N. C., & Contreras, L. M. (2019). Regulatory non-coding sRNAs in bacterial metabolic pathway engineering. *Metabolic Engineering*, *52*, 190–214. https://doi.org/10.1016/j.ymben.2018.11.013

Li, M. Z., & Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature Methods*, *4*(3), 251–256. https://doi.org/10.1038/nmeth1010

Linshiz, G., Stawski, N., Poust, S., Bi, C., Keasling, J. D., & Hillson, N. J. (2013). PaR-PaR laboratory automation platform. *ACS Synthetic Biology*, *2*(5), 216–222. https://doi.org/10.1021/sb300075t

Long, M. R., Ong, W. K., & Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology*, *34*, 135–141. https://doi.org/10.1016/j.copbio.2014.12.019

Looi, F. Y., Baker, M. L., Townson, T., Richard, M., Novak, B., Doran, T. J., & Short, K. R. (2018). Creating disease resistant chickens: A viable solution to avian influenza? *Viruses*, *10*(10). https://doi.org/10.3390/v10100561

Lubertozzi, D., & Keasling, J. D. (2009). Developing Aspergillus as a host for heterologous expression. *Biotechnology Advances*, *27*(1), 53–75. https://doi.org/10.1016/j.biotechadv.2008.09.001

Ma, S., Saaem, I., & Tian, J. (2012). Error correction in gene synthesis technology. *Trends in Biotechnology*, 30(3), 147–154. https://doi.org/10.1016/j.tibtech.2011.10.002

Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C., & Pfleger, B. F. (2015). Synthetic biology toolbox for controlling gene expression in the cyanobacterium Synechococcus sp. strain PCC 7002. *ACS Synthetic Biology*, *4*(5), 595–603. https://doi.org/10.1021/sb500260k

Martin, R. W., Des Soye, B. J., Kwon, Y.-C., Kay, J., Davis, R. G., Thomas, P. M., Majewska NI, Chen CX, Marcum RD, Weiss MG, Stoddart AE, Amiram M, Ranji Charna AK, Patel JR, Isaacs FJ, Kelleher NL, Hong SH, Jewett, M. C. (2018). Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nature Communications*, *9*(1), 1203. https://doi.org/10.1038/s41467-018-03469-5

McCarty, N. S., & Ledesma-Amaro, R. (2019). Synthetic biology tools to engineer microbial communities for biotechnology. *Trends in Biotechnology*, *37*(2), 181–197. https://doi.org/10.1016/j.tibtech.2018.11.002

McDermott, J., & Hardeman, M. (2018). Increasing Your Research's Exposure on Figshare Using the FAIR Data Principles. *Figshare*. https://doi.org/10.6084/m9.figshare.7429559.v2

Medema, M. H., van Raaphorst, R., Takano, E., & Breitling, R. (2012). Computational tools for the synthetic design of biochemical pathways. *Nature Reviews. Microbiology*, *10*(3), 191–202. https://doi.org/10.1038/nrmicro2717

Molteni, M. (2019, March 10). 23andMe's New Diabetes Test Has Experts Asking Who It's For | WIRED. Retrieved May 21, 2019, from https://www.wired.com/story/23andmes-new-diabetes-test-has-experts-asking-who-its-for/

Moore, S. J., MacDonald, J. T., Wienecke, S., Ishwarbhai, A., Tsipa, A., Aw, R., Kylilis N, Bell DJ, McClymont DW, Jensen K, Polizzi KM, Biedendieck R, Freemont, P. S. (2018). Rapid acquisition and model-based analysis of cell-free transcription-translation reactions from nonmodel bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), E4340–E4349. https://doi.org/10.1073/pnas.1715806115

Morrell, W. C., Birkel, G. W., Forrer, M., Lopez, T., Backman, T. W. H., Dussault, M., Petzold CJ, Baidoo EEK, Costello Z, Ando D, Alonso-Gutierrez J, George KW, Mukhopadhyay A, Vaino I, Keasling JD, Adams PD, Hillson NJ, Garcia Martin, H. (2017). The Experiment Data Depot: A Web-Based Software Tool for Biological Experimental Data Storage, Sharing, and Visualization. *ACS Synthetic Biology*, *6*(12), 2248–2259. https://doi.org/10.1021/acssynbio.7b00204

Muthusaravanan, S., Sivarajasekar, N., Vivek, J. S., Paramasivan, T., Naushad, M., Prakashmaran, J., Gayathri V, Al-Duaij, O. K. (2018). Phytoremediation of heavy metals: mechanisms, methods and enhancements. *Environmental Chemistry Letters*, *16*(4), 1–21. https://doi.org/10.1007/s10311-018-0762-3

Nahar, N., Rahman, A., Nawani, N. N., Ghosh, S., & Mandal, A. (2017). Phytoremediation of arsenic from the contaminated soil using transgenic tobacco plants expressing ACR2 gene of Arabidopsis thaliana. *Journal of Plant Physiology*, *218*, 121–126. https://doi.org/10.1016/j.jplph.2017.08.001

Naran, K., Nundalall, T., Chetty, S., & Barth, S. (2018). Principles of immunotherapy: implications for treatment strategies in cancer and infectious diseases. *Frontiers in Microbiology*, *9*, 3158. https://doi.org/10.3389/fmicb.2018.03158

National Academies of Sciences, Engineering, and Medicine, Division on Earth and Life Studies, Board on Life Sciences, Board on Chemical Sciences and Technology, & Committee on Strategies for Identifying and Addressing Potential Biodefense Vulnerabilities Posed by Synthetic Biology. (2018). *Biodefense in the age of synthetic biology*. Washington (DC): National Academies Press (US). https://doi.org/10.17226/24890

National Research Council (US) Committee on Assessing the Importance and Impact of Glycomics and Glycosciences. (2012). *Transforming glycoscience: A roadmap for the future*. Washington (DC): National Academies Press (US). https://doi.org/10.17226/13446

National Research Council (US) Committee on Industrialization of Biology: A Roadmap to Accelerate the Advanced Manufacturing of Chemicals, Board on Chemical Sciences and Technology, Board on Life Sciences, Division on Earth and Life Studies. (2015). *Industrialization of biology: A roadmap to accelerate the advanced manufacturing of chemicals.* Washington (DC): National Academies Press (US). https://doi.org/10.17226/19001

National Research Council (US) Committee on Scientific Evaluation of the Introduction of Genetically Modified Microorganisms and Plants into the Environment. (1989). *Field testing genetically modified organisms: framework for decisions*. Washington (DC): National Academies Press (US). https://doi.org/10.17226/1431

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross D, Densmore D, Voigt, C. A. (2016). Genetic circuit design automation. *Science*, *352*(6281), aac7341. https://doi.org/10.1126/science.aac7341

Niu, D., Wei, H.-J., Lin, L., George, H., Wang, T., Lee, I.-H., Zhao HY, Wang Y, Kan Y, Shrock E, Lesha E, Wang G, Luo Y, Qing Y, Jiao D, Zhao H, Zhou X, Wang S, Wei H, Güell M, Church GM, Yang, L. (2017). Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science*, *357*(6357), 1303–1307. https://doi.org/10.1126/science.aan4187

Pacheco, A. R., Moel, M., & Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communications*, *10*(1), 103. https://doi.org/10.1038/s41467-018-07946-9

Palacino, J., Swalley, S. E., Song, C., Cheung, A. K., Shu, L., Zhang, X., Van Hoosear M, Shin Y, Chin DN, Keller CG, Beibel M, Renaud NA, Smith TM, Salcius M, Shi X, Hild M, Servais R, Jain M, Deng L, Bullock C, McLellan M, Schuierer S, Murphy L, Blommers MJ, Blaustein C, Berenshteyn F, Lacoste A, Thomas JR, Roma G, Michaud GA, Tseng BS, Porter JA, Myer VE, Tallarico JA, Hamann LG, Curtis D, Fishman MC, Dietrich WF, Dales NA, Sivasankaran, R. (2015). SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nature Chemical Biology*, *11*(7), 511–517. https://doi.org/10.1038/nchembio.1837

Palluk, S., Arlow, D. H., de Rond, T., Barthel, S., Kang, J. S., Bector, R., Baghdassarian HM, Truong AN, Kim PW, Singh AK, Hillson NJ, Keasling, J. D. (2018). De novo DNA synthesis using polymerase-nucleotide conjugates. *Nature Biotechnology*, *36*(7), 645–650. https://doi.org/10.1038/nbt.4173

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante T, Ma D, Donghia N, Fan M, Daringer NM, Bosch I, Dudley DM, O'Connor DH, Gehrke L, Collins, J. J. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, *165*(5), 1255–1266. https://doi.org/10.1016/j.cell.2016.04.059

Pearl, J. (2018). Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining' ' - WSDM '18* (pp. 3–3). New York, New York, USA: ACM Press. https://doi.org/10.1145/3159652.3176182

Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D., & Kosuri, S. (2018). Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science*, *359*(6373), 343–347. https://doi.org/10.1126/science.aao5167

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Qian, L., & Winfree, E. (2011). Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, *332*(6034), 1196–1201. https://doi.org/10.1126/science.1200520

Rai, P. K., Lee, S. S., Zhang, M., Tsang, Y. F., & Kim, K.-H. (2019). Heavy metals in food crops: Health risks, fate, mechanisms, and management. *Environment International*, *125*, 365–385. https://doi.org/10.1016/j.envint.2019.01.067

Ravikumar, A., Arrieta, A., & Liu, C. C. (2014). An orthogonal DNA replication system in yeast. *Nature Chemical Biology*, *10*(3), 175–177. https://doi.org/10.1038/nchembio.1439

Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., & Liu, C. C. (2018). Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell*, *175*(7), 1946–1957.e13. https://doi.org/10.1016/j.cell.2018.10.021

Reva, B. A., Finkelstein, A. V., & Skolnick, J. (1998). What is the probability of a chance prediction of a protein structure with an rmsd of 6 A? *Folding & Design*, *3*(2), 141–147. https://doi.org/10.1016/S1359-0278(98)00019-4

Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., Lee D, Huang CL, Chandrasegaran S, Cai Y, Boeke JD, Bader, J. S. (2017). Design of a synthetic yeast genome. *Science*, *355*(6329), 1040–1044. https://doi.org/10.1126/science.aaf4557

Ross, M. J., & Coates, P. T. (2018). Using CRISPR to inactivate endogenous retroviruses in pigs: an important step toward safe xenotransplantation? *Kidney International*, *93*(1), 4–6. https://doi.org/10.1016/j.kint.2017.11.004

Schellenberger, J., Lewis, N. E., & Palsson, B. Ø. (2011). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, *100*(3), 544–553. https://doi.org/10.1016/j.bpj.2010.12.3707

Seelig, G., Soloveichik, D., Zhang, D. Y., & Winfree, E. (2006). Enzyme-free nucleic acid logic circuits. *Science*, *314*(5805), 1585–1588. https://doi.org/10.1126/science.1132493

Sethuraman, N., & Stadheim, T. A. (2006). Challenges in therapeutic glycoprotein production. *Current Opinion in Biotechnology*, *17*(4), 341–346. https://doi.org/10.1016/j.copbio.2006.06.010

Shih, S. C. C., Goyal, G., Kim, P. W., Koutsoubelis, N., Keasling, J. D., Adams, P. D., Hillson, N. J., Singh, A. K. (2015). A versatile microfluidic device for automating synthetic biology. *ACS Synthetic Biology*, *4*(10), 1151–1164. https://doi.org/10.1021/acssynbio.5b00062

Sid, H., & Schusser, B. (2018). Applications of gene editing in chickens: A new era is on the horizon. *Frontiers in Genetics*, *9*, 456. https://doi.org/10.3389/fgene.2018.00456

Smith, H. O., Hutchison, C. A., Pfannkoch, C., & Venter, J. C. (2003). Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(26), 15440–15445. https://doi.org/10.1073/pnas.2237126100

Stark, J. C., Huang, A., Hsu, K. J., Dubner, R. S., Forbrook, J., Marshalla, S., Rodriguez F, Washington M, Rybnicky GA, Nguyen PQ, Hasselbacher B, Jabri R, Kamran R, Koralewski V, Wightkin W, Martinez T, Jewett, M. C. (2019). BioBits Health: Classroom Activities Exploring Engineering, Biology, and Human Health with Fluorescent Readouts. *ACS Synthetic Biology*, *8*(5), 1001–1009. https://doi.org/10.1021/acssynbio.8b00381

Stark, J. C., Huang, A., Nguyen, P. Q., Dubner, R. S., Hsu, K. J., Ferrante, T. C., Anderson M, Kanapskyte A, Mucha Q, Packett JS, Patel P, Patel R, Qaq D, Zondor T, Burke J, Martinez T, Miller-Berry A, Puppala A, Reichert K, Schmid M, Brand L, Hill LR, Chellaswamy JF, Faheem N, Fetherling S, Gong E, Gonzalzles EM, Granito T, Koritsaris J, Nguyen B, Ottman S, Palffy C, Patel A, Skweres S, Slaton A, Woods T, Donghia N, Pardee K, Collins JJ, Jewett, M. C. (2018). BioBits™ Bright: A fluorescent synthetic biology education kit. *Science Advances*, *4*(8), eaat5107. https://doi.org/10.1126/sciadv.aat5107

Stephens, N., Di Silvio, L., Dunsford, I., Ellis, M., Glencross, A., & Sexton, A. (2018). Bringing cultured meat to market: Technical, socio-political, and regulatory challenges in cellular agriculture. *Trends in Food Science & Technology*, *78*, 155–166. https://doi.org/10.1016/j.tifs.2018.04.010

Sundstrom, E. R., & Criddle, C. S. (2015). Optimization of Methanotrophic Growth and Production of Poly(3-Hydroxybutyrate) in a High-Throughput Microbioreactor System. *Applied and Environmental Microbiology*, *81*(14), 4767–4773. https://doi.org/10.1128/AEM.00025-15

Takahashi, M. K., Chappell, J., Hayes, C. A., Sun, Z. Z., Kim, J., Singhal, V., Spring KJ, Al-Khabouri S, Fall CP, Noireaux V, Murray RM, Lucks, J. B. (2015). Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems. *ACS Synthetic Biology*, *4*(5), 503–515. https://doi.org/10.1021/sb400206c

Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, *249*(4968), 505–510. https://doi.org/10.1126/science.2200121

Van Enennaam, A. (2018, June 12). Use of Gene Editing to Introduce the Polled Trait into Elite Germplasm. Retrieved February 26, 2019, from https://www.dairyherd.com/article/use-gene-editing-introduce-polled-trait-elite-germplasm/

Venayak, N., von Kamp, A., Klamt, S., & Mahadevan, R. (2018). MoVE identifies metabolic valves to switch between phenotypic states. *Nature Communications*, *9*(1), 5332. https://doi.org/10.1038/s41467-018-07719-4

Villa, J. K., Su, Y., Contreras, L. M., & Hammond, M. C. (2018). Synthetic biology of small rnas and riboswitches. *Microbiology Spectrum*, *6*(3). https://doi.org/10.1128/microbiolspec.RWR-0007-2017

Watkins, A. M., Geniesse, C., Kladwang, W., Zakrevsky, P., Jaeger, L., & Das, R. (2018). Blind prediction of noncanonical RNA structure at atomic accuracy. *Science Advances*, *4*(5), eaar5316. https://doi.org/10.1126/sciadv.aar5316

Weinhandl, K., Winkler, M., Glieder, A., & Camattari, A. (2014). Carbon source dependent promoters in yeasts. *Microbial Cell Factories*, *13*, 5. https://doi.org/10.1186/1475-2859-13-5

Wen, K. Y., Cameron, L., Chappell, J., Jensen, K., Bell, D. J., Kelwick, R., Kopniczky M, Davies JC, Filloux A, Freemont, P. S. (2017). A Cell-Free Biosensor for Detecting Quorum Sensing Molecules in P. aeruginosa-Infected Respiratory Samples. *ACS Synthetic Biology*, *6*(12), 2293–2301. https://doi.org/10.1021/acssynbio.7b00219

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A.,

Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Yang, K. K., Wu, Z., & Arnold, F. H. (2018). Machine learning-guided directed evolution for protein engineering. Retrieved from https://arxiv.org/abs/1811.10775v2

Yin, G., Garces, E. D., Yang, J., Zhang, J., Tran, C., Steiner, A. R., Roos C, Bajad S, Hudak S, Penta K, Zawada J, Pollitt S, Murray, C. J. (2012). Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs*, *4*(2), 217–225. https://doi.org/10.4161/mabs.4.2.19202

You, M., & Jaffrey, S. R. (2015). Designing optogenetically controlled RNA for regulating biological systems. *Annals of the New York Academy of Sciences*, *1352*, 13–19. https://doi.org/10.1111/nyas.12660

Zhong, Z., & Liu, C. C. (2019). Probing pathways of adaptation with continuous evolution. *Current Opinion in Systems Biology*. https://doi.org/10.1016/j.coisb.2019.02.002