EBRC Engineering Biology Research Consortium

Engineering Biology

A Research Roadmap for the Next-Generation Bioeconomy

June 2019

roadmap.ebrc.org

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.

EBRC

June 2019

Roadmap Contributors

Leadership Team

Emily Aurand
EBRC
Director of Roadmapping

Jay Keasling
University of California, Berkeley
Working Group Chair
Energy Sector Lead

Douglas Friedman
EBRC
Executive Director

Technical Theme Leads

Howard Salis
Pennsylvania State University
Gene editing, synthesis, and assembly

Chang Liu
University of California, Irvine
Biomolecule, pathway, and circuit engineering

Pamela Peralta-Yahya
Georgia Institute of Technology
Host and consortia engineering

James Carothers
University of Washington
Host and consortia engineering

Adam Arkin
University of California, Berkeley
Data Integration, Modeling, and Automation

Cynthia Collins
Rensselaer Polytechnic Institute
Data Integration, Modeling, and Automation

Application and Impact Sector Leads

Ute Galm
Zymergen, Inc.
Industrial Biotechnology

Carrie Cizauskas
Zymergen, Inc.
Industrial Biotechnology

Karmella Haynes
Emory University
Health & Medicine

Albert Lu
Corteva Agriscience
Food & Agriculture

David Savage
University of California, Berkeley
Environmental Biotechnology

Roadmapping Working Group

Visu Annaluru
Corteva Agriscience

Roel Bovenberg
DSM

Paul Carlson
Cornell University

Lydia Contreras
University of Texas at Austin

Paul Freemont
Imperial College of London

Fumiaki Hamazato
Hitachi High-Technologies Corporation

Michael Jewett
Northwestern University

Ahmad Khalil
Boston University

Jens Klaus Plassmeier
BASF Corporation

Hans Roubos
DSM

Jeffrey Sampson
Agilent Technologies

Matthew Wook Chang
National University of Singapore

Roadmap Contributors

Jason Aulds, Scientist, US Department of Defense
Ania Baetica, Postdoctoral Researcher (El-Samad Lab), University of California, San Francisco
Cassandra Barrett, Graduate Student (Haynes Lab), Arizona State University
Allison Berke, Scientist, The Good Food Institute
Daniel Brauer, Graduate Student (Francis Lab), University of California, Berkeley
Kathryn Brink, Graduate Student (Tabor Lab), Rice University
James Chappell, Assistant Professor, Rice University
Yvonne Chen, Assistant Professor, University of California, Los Angeles *
George Church, Professor, Harvard University
Austin Cole, Graduate Student (Ellington Lab), University of Texas
Rhiju Das, Associate Professor, Stanford University
Jason Delborne, Associate Professor, North Carolina State University *
Doug Densmore, Associate Professor, Boston University *
James Diggans, Director, Bioinformatics and Biosecurity, Twist Bioscience Corporation *
John Dueber, Associate Professor, University of California, Berkeley *
Mary Dunlop, Assistant Professor, Boston University *
Christopher Earnhart, Joint Product Lead, US Department of Defense
Robert Egbert, Staff Scientist, Pacific Northwest National Laboratory
Andrew Ellington, Professor, University of Texas at Austin *
Steven Evans, Retired, Dow AgroSciences *
Mike Fero, Chief Executive Officer, TeselaGen Biotechnology, Inc. *
Clem Fortman, Director of Security Engagements, Engineering Biology Research Consortium
Emma Frow, Assistant Professor, Arizona State University *
Larry Gilbertson, Scientist & Genomics Strategy Lead, Monsanto
Jimmy D. Gollihar, Synthetic Biologist, US Army Research Laboratory
Alexander Green, Assistant Professor, Arizona State University
Tahlia Haeuser, Scientist, Zymergen *
Emily Hartman, Graduate Student (Francis Lab), University of California, Berkeley
Nathan Hillson, Staff Scientist, Lawrence Berkeley National Laboratory *
Cyan James, Medical and Life Sciences Analyst, Peraton, support to US Dept. of Defense
Richard Johnson, Chief Executive Officer, Global Helix LLC *
Terry Johnson, Associate Teaching Professor, University of California, Berkeley *
Kristala Jones Prather, Professor, Massachusetts Institute of Technology *
Richard Kitney, Professor, Imperial College of London *
Cody Krivacic, Graduate Student (Kortemme Lab), University of California, San Francisco
Natalie Kuldell, Executive Director, BioBuilder Educational Foundation *
Joshua Leonard, Associate Professor, Northwestern University *
Julius Lucks, Associate Professor, Northwestern University *
June Medford, Professor, Colorado State University *
Jeremy Mills, Assistant Professor, Arizona State University
Richard Murray, Professor, California Institute of Technology *
Vincent Noireaux, Professor, University of Minnesota *
Michael P. O'Keefe, Principal, Science & Technology, LMI, support to US Dept. of Defense
Kenneth Oye, Associate Professor, Massachusetts Institute of Technology *

Megan Palmer, Senior Research Scholar, Stanford University *
C. Nicole Rosenzweig, Division Chief, US Department of Defense
Heike Sederoff, Professor, North Carolina State University
Patrick Shih, Assistant Professor, University of California, Davis
Pamela Silver, Professor, Harvard University *
Michael Smanski, Assistant Professor, University of Minnesota *
Liz Specht, Associate Director of Science & Technology, The Good Food Institute
Ranjan Srivastava, Professor, University of Connecticut
Mark Styczynski, Associate Professor, Georgia Institute of Technology *
Kersh Thevasundaram, Graduate Student (Chang Lab), University of California, Berkeley
Kurt Thorn, Director of Data Science, Zymergen *
Danielle Tullman-Ercek, Associate Professor, Northwestern University *
Alon Wellner, Postdoctoral Researcher (Liu Lab), University of California, Irvine
Wilson Wong, Associate Professor, Boston University
Peng Yin, Professor, Harvard University *
Jesse Zalatan, Assistant Professor, University of Washington
Alexandre Zanghellini, Co-Founder and CEO, Arzeda
Jeremy Zucker, Computational Scientist, Pacific Northwest National Laboratory

* Denotes EBRC Member

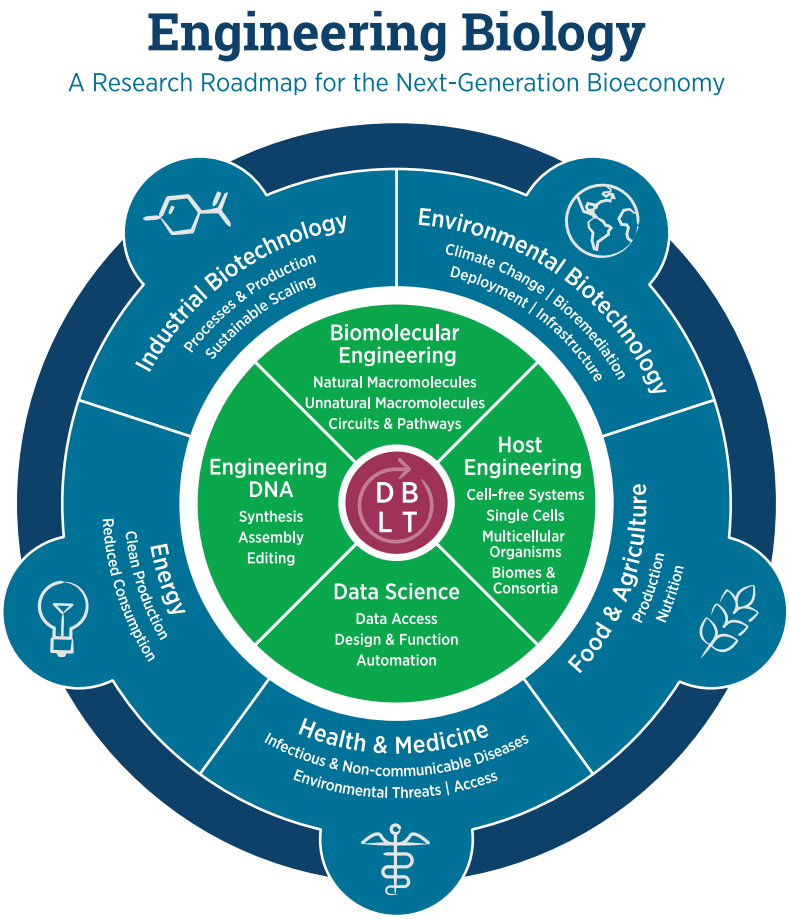
Table of Contents

Overview

Introduction

The EBRC Roadmap is a critical assessment of the current status and potential of engineering biology. It is intended to provide researchers and other stakeholders (including government funders) with a compelling set of technical challenges and opportunities in the near and long term. Our ongoing roadmapping process was initiated in response to the recommendations put forth in the 2015 National Academies report, *Industrialization of Biology*, and at the request of the National Science Foundation and other US government stakeholders, including the National Science & Technology Council. With this inaugural release of the Roadmap, EBRC will provide a "go-to" resource for engineering/synthetic biology research and development.

Working with the broader EBRC community, the Technical Roadmapping Working Group led the development of the roadmap scope and content. Collective insight and substantive contributions were provided by more than 80 leading scientists and engineers, including academic, industry, and student members of EBRC and from the broader research community. Since mid-2018, the working group has held six workshops and countless teleconferences to develop the content and engage discussion around the roadmap. The result is a collaborative effort that represents the engineering biology research community's vision for the future of the field.



The Roadmap

The matrixed framework of the roadmap considers challenges, bottlenecks, and other limitations observed or predicted in the research, development, and application of advancements in engineering biology tools and technologies toward addressing broad societal challenges. The roadmap's four technical themes form the foundation of engineering biology research and technology and illustrate where our current abilities lie and what we might achieve in the next 20 years. Complementarily, the five roadmap application and impact sectors demonstrate the breadth and impact of technical advancements in real-world application areas and exemplify how engineering biology tools and products could be oriented towards some of the most complex problems we face as a society. The technical themes represent a "bottom-up" approach focusing on tool and technology innovations to move the field forward, while the five application and impact sectors are a "top-down" look at how engineering biology could contribute toward addressing and overcoming national and global challenges.

Technical Themes

The technical themes of the roadmap focus on four key areas of engineering biology research and development: 1) Gene editing, synthesis, and assembly; 2) Biomolecule, pathway, and circuit engineering; 3) Host and consortia engineering; and 4) Data Integration, Modeling, and Automation. Each technical theme has a series of transformative tools and technologies - platform technical areas in which advancement is both necessary and will have a significant impact on revolutionizing the field - and aspirational goals that drive the progression of research and development. Construction of the technical roadmap is accomplished through delineation of milestones at 2, 5, 10, and 20 years and each milestone is elaborated by anticipated or imagined bottlenecks and creative potential solutions. The 2-year and 5-year milestones are intended to signify objectives that can be reached with current or recently implemented funding programs, as well as existing infrastructure and facilities resources. The 10-year and 20-year milestones are expected to be more ambitious achievements that may require (and thus, result in) significant technical advancements and/or increased funding and resources and new and improved infrastructure. The goals and breakthrough capabilities therefore represent the visionary 20+ year aspirations for the technical themes.

In the case of *Gene editing, synthesis, and assembly*, the future holds the potential for rapid, *de novo* synthesis of entire genomes, which is reflected in goals focused on the manufacture of high-fidelity oligonucleotides at 10,000-oligomer length, design and assembly of megabase clonal DNA fragments, and highly advanced precision gene editing without off-target effects. *Biomolecule, pathway, and circuit engineering* focuses on the activity and assembly of individual biomolecules into networks to carry out larger functions of the cell, and how we might design, create, and evolve these macromolecules from both natural and non-natural building blocks, with the end goal of achieving integrated, controlled circuits and pathways. Envisioned is the ability to routinely design and predict the structure and function of macromolecules, biosynthesis of unnatural amino acids and other building blocks, and controlled expression of transcription factors determining cell state. This blends and transitions into the larger and more complex functions of

individual cells, whole organisms, and biomes described in *Host and consortia engineering*, where we detail necessary advancements to achieve customized cell-free systems and synthetic cells, on-demand production and control of single cells and multicellular organisms with highly-defined functions, and finally the engineering of multi-genome systems and biomes. While, classically, engineering biology has focused on microorganisms as tools for production, we aim to expand this vision to include cells as the products themselves, and provide an outlook for enabling complex engineering of plants, animals, and multi-organism systems. Cornerstone to enabling engineering and production of designed genomes, non-natural biomolecular circuits, and customized cells and organisms, is the integration of advanced data analysis, design, and data modeling. *Data Integration, Modeling, and Automation* highlights the transformative potential of integrated biological data models, design frameworks for biomolecules, hosts, and organismal communities, and the promise of automating the design-build-test-learn process. Together these four technical areas establish the foundation from which we can enable the rapid progression and expansion of engineering biology into industry and application.

Application and Impact Sectors

The roadmap also illustrates many potential applications of engineering biology, and demonstrates the possible use and impact of these tools and technologies to address and overcome societal challenges, through a focus on five sectors: 1) Industrial Biotechnology; 2) Health & Medicine; 3) Food & Agriculture; 4) Environmental Biotechnology; and 5) Energy. We frame these possibilities through the lens of solving pervasive societal challenges, including enabling and establishing a cleaner environment, supporting the health and well-being of growing populations, and accelerating innovation and economic viability of industry. From each framing societal challenge, we consider the science and engineering aims and objectives for engineering biology that may be necessary or instrumental to overcoming the challenge and we identify potential discrete technical achievements towards the objective. These technical achievements in each sector reflect our four technical themes, allowing for future mapping of the technical milestones to the engineering biology sector objectives.

The roadmap details many potential applications of engineering biology toward establishing the U.S. as a global leader in the bio-based economy is the basis from which *Industrial Biotechnology* looks at sustainable manufacturing, novel product discovery and development, and integrated workflows and pipelines for biology-based products and materials. *Health & Medicine* focuses not only on developing and advancing tools for fighting disease, but also improving wellbeing through bio-integrated lifestyle technologies, offering new options to people with disabilities through engineered cell systems, and addressing the damage of environmental threats to health. Producing more food that is healthier and increasingly nutrient-dense is the focus of *Food & Agriculture*, including advancing the production of food and nutrients from atypical and under-utilized sources, such as microorganisms, insects, alternative plant species, and “clean meats”. Engineering biology achievements in *Environmental Biotechnology* will enable advancements in bioremediation, resource-recovery, the deployment of engineered organisms, and bio-enabled and bio-built infrastructure, contributing to cleaner lands, waters, and air. And finally, *Energy* focuses on the production of energy-dense and carbon-neutral biofuels, and

tools and products to reduce energy use and consumption. Combined, the Roadmap sectors suggest the potential breadth and utility of engineering biology tools and technologies in creating a better world.

The Way Ahead

As our roadmapping efforts progress, there is opportunity to expand each technical theme, to better capture the considerable breadth of engineering biology research, especially in related fields, such as bioinformatics, biomedical engineering, physics, and microscopy, to spur greater integration. A further focus on pipeline technologies, such as advanced robotics and machine learning, bio-integrated electronics, precision microfluidics and high-throughput screening, DNA barcoding, and novel biosensors, may elaborate near-term technical milestones and provide incentive for additional investment into emerging technologies. The greater roadmap framework and organization has been designed such that there is nearly-unlimited space to investigate the overlap between advancements in engineering biology tools and technologies and the impact and utility of these tools and technologies toward solving economic, industrial, and environmental problems; to this end, future EBRC roadmapping efforts are likely to include additional, intricate integration of the technical themes and application sectors. Similarly, our roadmapping efforts aim to be an evergreen process through consideration and assessment of the identified milestones to track attainment, and updates to acknowledge and explore new avenues and technical landscapes.

With this technical roadmap, EBRC hopes to establish a resource for the research and research-support community, including educators, policymakers, and funding bodies, that portrays the importance and impact of engineering biology tools and technologies and highlights community-identified areas for innovation and advancement.

Linking to Societal Considerations and Integrating Public Values

The current roadmap highlights a suite of technical ambitions and potential pathways for advancing the engineering of biology. It represents the initial outcome of a broad consultation of the scientific and engineering community to identify goals for the field and offers a vision of the technical possibilities and challenges of engineering biology in the service of broad social, ecological, and economic priorities. However, the current material does not explicitly address the policy and social environments that new technologies and tools must navigate if they are to make a difference in the real world.

The roadmap exists on a yet-unspecified landscape of social priorities, cultural preferences, ethical minefields, political traditions, and economic realities. Interdisciplinary scholarship in fields complementary to science and engineering can be engaged in critically evaluating and framing the goals and aims of this technical roadmap. As the roadmap evolves, it will be important to question how engineering biology is utilized to address these global problems, and ways that novel technologies might create new problems or exacerbate existing social and political inequalities. To maximize the possibility of positive outcomes, technological pursuits will need to be coupled with ongoing study and negotiation of the social, cultural, political, and economic landscapes for which they are being designed. Doing this will require leveraging expertise in disciplines beyond science and engineering, including the arts, humanities, and social and behavioral sciences.

With this technical roadmap as a first step, EBRC hopes to establish a resource for the research and research-support community, including educators, policymakers, and funding bodies, that portrays the importance and impact of engineering biology tools and technologies and highlights community-identified areas for innovation and advancement. This will include outlining areas of research and forms of stakeholder engagement needed to steer the field towards desirable societal outcomes. We can expect, over time, for the goals and priorities of this roadmap to change. We emphasize that broad interdisciplinary and ongoing engagement with a wide range of partners will be essential to guiding research trajectories in the most meaningful directions. EBRC will engage current and future supporters of this roadmap to prioritize such integrative and interdisciplinary considerations as part of its ongoing development.

Security: Managing Risk through Engagement

The development of synthetic biology technologies and their applications alters the overall biological security landscape. In order to best assess the opportunities for science and technology advancement, it is important to frame the advances in terms of their impacts on national and global security. This is especially relevant given that we encourage the further advance of technologies that can improve our economic security, while not challenging our physical security. While no technology is risk-free, by being able to carry out rational assessments of integrated risk it should be possible for policy makers and others to better assess whether investments are likely to enhance prosperity and security.

The roadmap has been written in a manner that implicitly considers security and has been, and continues to be, reviewed to identify any potential for misuse of the engineering biology tools, technologies, and applications outlined in the roadmap. The proposed roadmap elements are written at a level of detail that are not expected to directly enable malicious misuse; however, potential security concerns that could arise as the field achieves the technical goals put forth in the roadmap have been identified. Ongoing efforts will provide necessary description and context for these concerns, to be integrated with the current roadmap and/or published separately, as well as consideration of security issues independent of the present material. The goal of such current and ongoing security assessments are first and foremost to ensure that we do not enable malicious actors or activities through our work. Furthermore, it is incumbent on us to engage directly with the security community, policymakers, and other stakeholders about concerns we anticipate may emerge and how the EBRC community, through the Security Working Group, is engaging and collaborating with others to mitigate these concerns. Ongoing stakeholder engagement in the social and political context in which these technologies are developed is essential to ensuring that this roadmap does not introduce new vulnerabilities, and EBRC and its partners are committed to strengthening that engagement.

In concert with strategies to mitigate potential misuse, an overarching goal of this roadmap is to develop pathways for technology developments that has potential to enhance both economic and national security. These include strategies to incorporate security design into technologies from the outset, enable improved regulatory science, and protect human health, the environment and the economy. We remain strong advocates for diligent consideration of potential security considerations, by all stakeholders, in their pursuit to advance engineering biology.

Technical Themes

Gene Editing, Synthesis, and Assembly

(Engineering DNA)

Gene Editing, Synthesis, and Assembly

Summary

Gene Editing, Synthesis, and Assembly focuses on the development and advancement of tools to enable the production of chromosomal DNA and the engineering of entire genomes. Advancements are needed in the design and construction of functional genetic systems though the synthesis of long oligonucleotides, assembly of multiple fragments, and precision editing with high specificity.

Introduction and Impact

Fundamentally, an organism's sensing, metabolic, and decision-making capabilities are all encoded within their genome, a very long double-stranded DNA molecule. By changing an organism's genome sequence, we have the ability to rationally alter these cellular functions, and thereby engineer them to address a myriad of societal challenges. The ability to rationally alter DNA sequences, combining gene editing, DNA synthesis, and DNA assembly, are therefore considered a cornerstone capability of engineering biology, enabling us to construct engineered genetic systems to reprogram organisms with targeted functions. Advances in gene editing, synthesis, and assembly have significant transformative impacts on all sectors impacted by engineering biology by broadening the complexity and breadth of functionality that can be introduced into an engineered organism.

The market for synthesized DNA is both mature and ripe for disruption. Using existing technologies, several service providers currently synthesize single-stranded DNA molecules (oligonucleotides) and double-stranded DNA molecules (DNA fragments). They actively compete across several criteria, including cost-per-DNA base pair, sequence fidelity, turnaround time, confidentiality of intellectual property, and customer service. However, several early-stage technologies have the potential to dramatically alter the commercial landscape by enabling the manufacture of much longer DNA fragments at significantly reduced costs.

Gene Editing, Synthesis, and Assembly highlights several technological routes to achieving the overall goal of manufacturing mega-base length DNA molecules, and designing genes and genomes with desired functionalities. We also illustrate how new technological developments in one process (e.g., oligonucleotide synthesis, or coupled synthesis and sequencing) can directly lead to improvements in downstream processes (e.g., DNA fragment synthesis).

Transformative Tools and Technologies

Oligonucleotide synthesis technologies

Currently, phosphoramidite-based chemistry is the predominant approach for synthesizing oligonucleotides. Even after significant optimization, per-cycle synthesis yields are about 99.5%, meanwhile synthesis of a 200-nucleotide oligonucleotide has a yield of only 35% (Hughes & Ellington, 2017). New technologies seek to improve this process by: 1) synthesizing thousands of oligonucleotides in parallel, using either on-chip supports or within tiny microtiter wells (Kosuri & Church, 2014); or 2) improving synthesis processivity by replacing the

phosphoramidite-based chemistry, for example, using enzyme catalysis (e.g., terminal deoxynucleotidyl transferases) to extend primers with defined nucleotides (Palluk et al., 2018). Clearly, achieving picomole production of 1000-mer oligonucleotides with error-free sequences would significantly improve the overall DNA assembly protocol.

Technologies for oligonucleotide assembly into non-clonal DNA fragments

Currently, multiple 60- to 200-mer oligonucleotides are assembled into non-clonal DNA fragments using a combination of annealing, ligation, and/or polymerase chain reaction. The cost of synthesizing non-clonal DNA fragments is $0.10 to $0.30 per base pair, depending on size and complexity. DNA fragments between 300 and 1800 base pairs can be synthesized by multiple providers and DNA fragments up to 5800 base pairs can be synthesized by select providers at increased cost. Errors are introduced whenever two oligonucleotides form undesired base pairings, when two oligonucleotides are incorrectly ligated together, or when DNA polymerases extend a synthesized DNA fragment with an incorrect nucleotide. Certain sequence determinants will increase the error rate, resulting in a mixture of undesired fragments. Computational sequence design can reduce the frequency of these errors. Mismatch repair enzymes may be added (with added cost) to eliminate DNA fragments with mis-paired nucleotides, for example, as a result of mis-annealing or DNA polymerase errors. This process has been scaled up to assemble thousands of non-clonal DNA fragments per day. The purification of full-length, error-free DNA fragments remains a challenge. Utilizing longer oligonucleotides (see Oligonucleotide synthesis technologies) would enable the synthesis of longer non-clonal DNA fragments with the same error rate. New technologies utilizing nanopore sequencing have the potential to couple sequencing and purification at single-molecule resolution.

Multi-fragment DNA assembly techniques for clonal genetic systems and genomes

Currently, multiple DNA fragments (300 to 3000 base pairs long) are assembled into large genetic systems (10,000 to 1,000,000 base pairs long) using single-pot DNA assembly techniques that combine cocktails of bio-prospected and/or engineered enzymes, including exonucleases, endonucleases, DNA polymerases, ligases, and/or recombinases (Gibson et al., 2009; Hughes & Ellington, 2017). Enzyme costs are currently about $25 per assembly. Assembled DNA is then introduced into cells for clonal separation and replication. Most assembly techniques have essential sequence determinants, for example, regions of overlapping homology or flanking Type IIs restriction sites (Engler, Kandzia, & Marillonnet, 2008). Errors are introduced when two fragments anneal together at incorrect overlap regions, when two fragments are mis-ligated at incorrect ligation junctions, or when DNA polymerases incorporate incorrect nucleotides during DNA synthesis. Computational sequence design can limit the frequency of errors. A major challenge for DNA assembly is the trial-and-error identification of a full-length, error-free genetic system. For example, an optimized assembly technique with a per-junction efficiency of 90% will assemble a 10-part (3000 base pairs per part) system with 35% yield. At the same per-junction efficiency, assembling a 1,000,000 base pair genome from 3000 base pair DNA fragments will have a miniscule yield of 5.2×10^{-14} %. This limitation to DNA assembly has motivated the synthesis of longer non-clonal DNA fragments

(see Technologies for oligonucleotide assembly into non-clonal DNA fragments). For example, 1,000,000 base pair genomes could be assembled from 10,000 base pair, 30,000 base pair, or 50,000 base pair DNA fragments with a 0.002%, 2.7%, or 11% efficiency, respectively. If longer non-clonal DNA fragments are unavailable, then hierarchical approaches to DNA assembly are required, which increases the number of DNA assembly reactions and verification costs.

Sequencing costs become significant once assembled genetic systems are large and/or assembly yields are exceedingly small. For example, after assembling a 30,000 base pair genetic system with a 35% yield, it is necessary to sequence at least seven clonal isolates to achieve at least a 95% chance of identifying a fully-correct one. At low throughput, this cost is about $1000 (using Sanger sequencing). Using next generation sequencing, this cost can be greatly reduced to about $0.70, but only when a large amount of DNA (2 billion base pairs) is sequenced at the same time (Goodwin, McPherson, & McCombie, 2016). Similarly, if a 1,000,000 base pair genome is assembled from 30,000 base pair fragments with a 2.7% yield, then it would be necessary to sequence 100 clonal isolates to achieve a 93% chance of identifying a fully correct one (about $275 in sequencing costs). Finally, hierarchical DNA assembly can be performed by first assembling and purifying smaller genetic systems (e.g., up to 30,000 base pairs) and then using them to perform a multi-fragment assembly to build larger genetic systems (e.g., up to 35 five times larger than the smaller systems) (Richardson et al., 2017). Hierarchical DNA assembly increases sequencing costs by a multiplier roughly equal to the number of hierarchical cycles. Overall, DNA assembly costs are greatly reduced by utilizing longer non-clonal DNA fragments and by parallelizing operations such that at least 2 billion base pairs of DNA are verified across multiple DNA assembly reactions.

GENE EDITING, SYNTHESIS, AND ASSEMBLY

Goal	Breakthough Capability	Milestone

Manufacture thousands of very long oligonucleotides with high fidelity.

Highly efficient oligonucleotide synthesis to increase the number, length, and fidelity of oligonucleotides.			
Robustly synthesize one million 200-mer oligonucleotides with a per-nucleotide error rate of fewer than one in 500 nucleotides.	Robustly synthesize 1,000-mer oligonucleotides with a per-nucleotide error rate of fewer than one in 1,000 nucleotides.	Reduce per-nucleotide error rates for 1,000-mer oligonucleotide synthesis to fewer than one in 5,000 nucleotides.	Synthesize 10,000-mer oligonucleotides at 99.99% cycle efficiency within one minute with a per-nucleotide error rate of fewer than one in 30,000 nucleotides.

Many-fragment DNA assembly with simultaneous, high-fidelity sequence validation.

Predictive design of DNA sequences for improved assembly of longer, more information-rich DNA fragments.			
Coupled design of DNA sequences to optimize nucleotide composition to support synthesis, while maintaining genetic system function.	Incorporate machine learning to identify poorly-understood problematic sequences and process conditions.	Design algorithms that identify optimal synthesis strategies for assembling megabase-length genetic systems.	Design algorithms for optimal one-pot-assembly of billions of unique genomic/chromosomal variants with defined sequences.
Methods for one-step, simultaneous assembly and sequence-verification of long DNA fragments.			
Reliable assembly of 10,000 base pair non-clonal DNA fragments.	Reliable assembly and verification of 10,000 base pair clonal DNA fragments.	Reliable assembly and verification of 100,000 base pair clonal DNA fragments.	Reliable assembly and verification of 1,000,000 to 10,000,000 base pair clonal DNA fragments.
Pipelined synthesis, assembly, and functional testing of engineered genetic systems.			
Achieve desired functionalities in lower-fidelity, error-prone genetic systems.	Achieve reliable Design-for-Testing in engineered genetic systems.	Achieve readily-swappable modules within large genetic systems.	Achieve one-month Design-to-Test cycles for megabase-length genetic systems.

Precision genome editing at multiple sites simultaneously with no off-target effects.

Ability to reliably create any precise, defined edit(s) (single nucleotide polymorphisms or gene replacement) with no unintended editing in any organism, with edits ranging from a single base change to the insertion of entire pathwa.			
Ability to generate any defined single base pair change in model organisms.	High efficiency editing (> 90%) across the genome with no off-target activity.	High-efficiency gene insertion or deletion of moderately large changes (< 10 kb) via homologous recombination.	Precise, parallel editing or regulatory modifications (10 to 1000 modifications) across model and non-model organisms, including plants and animals.
Precise, predictable, and tunable control of gene expression for many genes inside diverse cells and organisms across different timescales.			
Achieve long-lasting gene repression and activation.	Ability to regulate expression in non-model organisms.	Technologies to monitor and manipulate genetic and epigenetic mechanisms controlling tissue-wide and organism-wide expression levels over time.	Ability to precisely regulate gene expression in whole-body organisms, with single-cell resolution using dynamic or static control.
Ability to reproducibly deliver editing cargo efficiently and specifically to a given target cells or tissues, and control dosage and timing of the editing machinery.			
Improve editors to function without sequence requirements with activity comparable to 2019 state-of-the-art capabilities.	Routine use of editors without detectable off-target effects.	Enhance specificity of delivery modalities for high efficiency (>90% efficient) editing of cells in a defined tissue.	Quantitative, specific, and multiplexed editing of any site, in any cell, in any organism.
2 Years	**5 Years**	**10 Years**	**20 Years**

Roadmap Elements

Goal 1: Manufacture thousands of very long oligonucleotides with high fidelity.

[Current State-of-the-Art]: Existing synthesis chemistries manufacture oligonucleotides up to 200 nucleotides long with cycle efficiencies of 99.5% and yields of 35%. Parallel synthesis of oligonucleotides is carried out on solid supports, producing up to 300,000 oligonucleotides with defined sequences (Hughes & Ellington, 2017; Kosuri & Church, 2014).

[Breakthrough Capability]: Highly efficient oligonucleotide synthesis to increase the number, length, and fidelity of oligonucleotides.

- **2 years: Robustly synthesize one million 200-mer oligonucleotides with a per-nucleotide error rate of fewer than one in 500 nucleotides.**
    - [Bottleneck]: Scaling-up the production of chip-based or semiconductor-based oligonucleotide synthesis chemistries.
        - [Potential Solution]: Microfabrication of nanotiter plates and patterned nanometer-scale chips.
        - [Potential Solution]: Improved process dynamics taking into account inherent stochasticity and improved electronic control of reaction chemistries.
- **5 years: Robustly synthesize 1,000-mer oligonucleotides with a per-nucleotide error rate of fewer than one in 1,000 nucleotides.**
    - [Bottleneck]: Current phosphoramidite-based chemistries have peaked at 99.5% per-nucleotide efficiencies, resulting in only 0.66% yields when producing 1,000-mers; efficiencies must be 99.9% to achieve more than 35% yields and cycle times must also be reduced for commercial scalability.
        - [Potential Solution]: Enzyme-based, non-templated synthesis (e.g., via terminal deoxynucleotidyl transferases) has the potential to achieve greater than 99.9% per-nucleotide efficiencies and synthesis rates exceeding one nucleotide-per-second.
- **10 years: Reduce per-nucleotide error rates for 1,000-mer oligonucleotide synthesis to fewer than one in 5,000 nucleotides.**
    - [Bottleneck]: Non-templated DNA synthesis is currently slow with lower fidelity than templated synthesis and improvements in enzyme substrate selectivity or substrate availability are needed to control sequence-specific synthesis.
        - [Potential Solution]: Significant bioprospecting, rational design, and directed evolution of enzymes responsible for non-templated DNA synthesis can improve selectivity and increase catalytic efficiencies.
- **20 years: Synthesize 10,000-mer oligonucleotides at 99.99% cycle efficiency within one minute with a per-nucleotide error rate of fewer than one in 30,000 nucleotides.**
    - [Bottleneck]: Multiple synergistic improvements are needed, including improved non-templated DNA polymerases, fast substrate switching at the nanoliter scale, multi-nucleotide same-cycle addition, and electronic control of substrate selection.

- [Potential Solution]: Inspiration from natural DNA polymerases, ligases, recombinases, and helicases, working together in a dynamic molecular machine, potentially using non-natural nucleotides for greater specificity.

Goal 2: Many-fragment DNA assembly with simultaneous, high-fidelity sequence validation.

[Current state-of-the-art]: Oligonucleotides are assembled into double-stranded DNA fragments up to 6000 base pairs long using *in vitro* techniques (e.g., polymerase cycling assembly, ligation cycling) as well as *in vivo* techniques (yeast-mediated homologous recombination), producing non-clonal DNA fragments (Gibson, 2011; Li & Elledge, 2007; Richardson et al., 2017; Smith, Hutchison, Pfannkoch, & Venter, 2003). Clonal (isogenic) fragments are then identified using a combination of enzyme-based removal of mismatched base pairs (e.g., MutS) and DNA sequencing (Sanger or NGS). Multiple verified DNA fragments are then assembled together into longer fragments (10,000 to 100,000 base pairs long) using hierarchical approaches employing DNA assembly techniques (e.g., Gibson assembly, ligation cycling reaction, Golden Gate). Megabase length DNA is then assembled from 100,000 base pair fragments using yeast-mediated homologous recombination. More detailed descriptions of commonly used techniques follow:

- **Polymerase Cycling Assembly (PCA)** is a method to assemble larger DNA constructs from shorter oligonucleotides (Smith et al., 2003). PCA is an efficient method for assembling constructs between 200 to 1,000 base pairs in length. The process is similar to PCR, but utilizes a set of overlapping “seed” oligonucleotides that are designed to hybridize to one another leaving gaps that are then filled in using a thermostable DNA polymerase. The oligonucleotides are generally 50 to 100 nucleotides in length to ensure uniqueness in the hybridization with their complement. The reactions are cycled from ~60 and ~95 C^{o} for 15 to 30 cycles. The full-length assembled product is then usually amplified by PCR using two terminal-specific primers. PCA is an efficient method for assembling constructs between 200 and 1,000 base pairs in length and can be performed in individual tubes or multiplexed using microtiter well plates.
- **Emulsion PCA** is a method developed by Sriram Kosuri for highly multiplexing the assembly of larger constructs from small amounts of shorter DNA fragments (Plesa, Sidore, Lubock, Zhang, & Kosuri, 2018). In this method, the oligos required for a given construct are designed with a unique barcode on the terminus which specifically hybridizes with a complementary barcoded attached to a bead from a complex pool of oligonucleotides. The bead mixture is then emulsified into picoliter-sized droplets containing a Type IIs restriction endonuclease (RE), dNTPs, and a thermostable DNA polymerase. The oligonucleotides are released from the bead by the Type II RE and then assembled by PCA through thermal cycling of the emulsion. Using this method, thousands of specific constructs can be assembled in a single emulsion tube depending upon the number uniquely barcoded beads.
- **Ligase Cycling Assembly (LCA)** is a method to assemble larger DNA constructs from shorter oligonucleotides or double-stranded DNA fragments (de Kok et al., 2014). LCA is an efficient method for assembling constructs between 500 and 10,000 base pairs in length. LCA assembly uses shorter, single-stranded bridging oligonucleotides that are complementary to the termini of adjacent DNA fragments that are to be joined using a thermostable ligase. Like PCA, LCA

utilizes multiple temperature cycling to denature, re-anneal, and then ligate the fragments to assemble the larger DNA construct, and can be performed in individual tubes or multiplexed using microtiter well plates.

- **Gibson Assembly** is a method to assemble larger DNA constructs from shorter oligonucleotides or double-stranded DNA fragments (Gibson et al., 2009). Gibson assembly is an efficient method for assembling constructs up to many tens of kilobase-pairs in length. This method, which is isothermal, utilizes up to 15 double-stranded DNA fragments having around 20 to 40 base pair overlaps with the adjacent DNA fragments. The DNA fragments are first incubated with 5' to 3' exonuclease, resulting in single-stranded regions on the adjacent DNA fragments that can anneal in a base pair-specific manner. The gaps are then filled in with a DNA polymerase and the final nicks closed with a DNA ligase. This method can be performed in individual tubes or multiplexed using microtiter well plates.

Importantly, the final fidelity (error rate) of the assembled constructs using the above methods are at the mercy of the quality of the input oligonucleotides. These methods usually incorporate some type of error reduction or correction methods which include removing errored duplexes (mismatches and insertions) with the MutS protein after denaturation and reannealing of the construct, or degradation of the error containing DNA using T7 or CEL endonuclease. (For review please see: 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)

[Breakthrough Capability 1]: Predictive design of DNA sequences for improved assembly of longer, more information-rich DNA fragments.

- **2 years: Coupled design of DNA sequences to optimize nucleotide composition to support synthesis, while maintaining genetic system function.**
    - [Bottleneck]: Many genetic systems contain polymeric sequences, long repeats, and non-canonical DNA structures that inhibit the assembly process.
        - [Potential Solution]: Genetic systems can be rationally designed to eliminate problematic sequence elements, while maintaining their function, thus reducing their "synthesis complexity".
        - [Potential Solution]: Toolboxes of highly non-repetitive genetic parts can be designed and characterized to enable design of non-repetitive genetic systems.
- **5 years: Incorporate machine learning to identify poorly-understood problematic sequences and process conditions.**
    - [Bottleneck]: The complete list of sequence elements that inhibit DNA assembly is not fully known and the process conditions leading to undesired byproducts are not well understood.
        - [Potential Solution]: Machine learning algorithms have the ability to identify problematic DNA sequences and undesired process conditions that lead to inefficient DNA assembly.

- **10 years: Design algorithms that identify optimal synthesis strategies for assembling megabase-length genetic systems.**
    - [Bottleneck]: The functions of some genetic system components are more strictly reliant on problematic sequences and trade-offs between design-for-function versus design-for-synthesis are likely.
        - [Potential Solution]: Design algorithms can identify regions with problematic sequences and identify optimal strategies for mixing and matching megabase-length assembly strategies accounting for these regions.
- **20 years: Design algorithms for optimal one-pot-assembly of billions of unique genomic/chromosomal variants with defined sequences.**
    - [Bottleneck]: Mixtures of oligonucleotides can be used to construct combinatorial libraries of DNA fragments, though assembling those fragment libraries into diversified mega-base genetic systems has not been achieved.
        - [Potential Solution]: Parallel evaluation of sequence design criteria across billions (or trillions) of potential sequence variants can be carried out; as the diversification of libraries increase, the number of sequence variants increases combinatorially.

[Breakthrough Capability 2]: Methods for one-step, simultaneous assembly and sequence-verification of long DNA fragments.

- **2 years: Reliable assembly of 10,000 base pair non-clonal DNA fragments.**
    - [Bottleneck]: The availability of high-fidelity long oligonucleotides, the optimization of process conditions, and the presence of problematic sequences.
        - [Potential Solution]: Higher fidelity 100-mer and 200-mer oligonucleotides.
        - [Potential Solution]: Identification of optimal process conditions and removal (by design) of problematic sequence elements.
- **5 years: Reliable assembly and verification of 10,000 base pair clonal DNA fragments.**
    - [Bottleneck]: Low assembly yields and decoupled sequencing leads to more costly hierarchical processes with higher failure rates.
        - [Potential Solution]: Enzyme-based selection (e.g., via MutS) can eliminate DNA fragments containing errors**.**
        - [Potential Solution]: Approaches using simultaneous DNA synthesis and sequencing can rapidly sort DNA fragments, excluding fragments with errors (for example, using nanopore-based sequencing and dynamic pore flicking).
- **10 years: Reliable assembly and verification of 100,000 base pair clonal DNA fragments.**
    - [Bottleneck]: Reliable, low-cost assembly of clonal 10,000 base pair fragments.
        - [Potential Solution]: Higher efficiency, ten-part assemblies using lower-cost, clonal 10,000 base pair DNA fragments.
        - [Potential Solution]: Extra long read sequencing for verification of 100,000 base pair fragments (e.g., nanopore sequencing).

- **20 years: Reliable assembly and verification of 1,000,000 to 10,000,000 base pair clonal DNA fragments.**
    - [Bottleneck]: Reliable, low-cost assembly of clonal 100,000 base pair fragments.
        - [Potential Solution]: *In vivo*, yeast-mediated assembly of clonal 100,000 base pair fragments into megabase-length genetic systems.
        - [Potential Solution]: Extra long read sequencing for verification of 1,000,000 base pair fragments (e.g., nanopore sequencing).

[Breakthrough Capability 3]: Pipelined synthesis, assembly, and functional testing of engineered genetic systems.

- **2 years: Achieve desired functionalities in lower-fidelity, error-prone genetic systems.**
    - [Bottleneck]: Unpredictable relationship between synthesis and assembly errors versus undesired functional outcomes.
        - [Potential Solution]: Elimination of problematic sequences via rational design and incorporation of robust, mutation-invariant design into genetic systems.
        - [Potential Solution]: Routine application of low-cost -omics technologies to verify the functions of genetic systems (e.g., DNA-Seq, RNA-Seq, Ribo-Seq, and metabolomics).
- **5 years: Achieve reliable Design-for-Testing in engineered genetic systems.**
    - [Bottleneck]: Costly to assay diverse genetic functions to verify desired behaviors.
        - [Potential Solution]: Synthesized and assembled genetic systems can directly incorporate a suite of sensors and genetic circuits for self-testing of genetic system function; sensor-circuit outputs could be tailored for desired high-throughput assays, including surface display, Flow-Seq, and RNA-Seq.
- **10 years: Achieve readily-swappable modules within large genetic systems.**
    - [Bottleneck]: Synthesis of megabase-length genetic systems may contain commonly used and re-used genetic modules.
        - [Potential Solution]: Previously synthesized and assembled genetic modules (of more than 100,000 base pair fragments) can be re-used in downstream processes and models can be developed to predict inter-module interactions and overall system function.
- **20 years: Achieve one-month Design-to-Test cycles for megabase-length genetic systems.**
    - [Bottleneck]: Design algorithms, synthesis chemistries, assembly techniques, simultaneous sequencing, and functional testing must be seamlessly integrated within a commercially viable suite of services with fast turnaround times.
        - [Potential Solution]: A combination of cooperative horizontal service providers and well-integrated vertical service providers operating within a healthy commercial ecosystem.

Goal 3: Precision genome editing at multiple sites simultaneously with no off-target effects.

[Current State-of-the-Art]: A variety of current tools can be used for DNA sequence edits and for non-editing-based genome engineering including gene regulation and chromatin engineering. Transcription activator-like effector nucleases (TALEN)-based or clustered regularly interspaced short palindromic repeats (CRISPR)-based genome engineering techniques introduce site-specific nicks or double-stranded breaks, which are then repaired using natural repair pathway (Doudna & Charpentier, 2014). Additional state-of-the-art editing technologies include adeno-associated virus (AAV)-mediated homologous recombination and meganuclease activity. With CRISPR and TALEN technologies, up to six distinct sites, and up to 15,000 identical sites, have been targeted simultaneously, with efficiencies ranging from 2% to 90%. Gene regulation is achieved through site-specific DNA-binding proteins (zinc-finger proteins, transcription activator-like effectors, and Cas proteins), which fuse to gene regulatory domains to carry out activation or repression of desired genes. In these cases, up to six distinct genes have been targeted for regulation, with repression magnitudes up to 300-fold (knock-down) and activation magnitudes up to 20-fold (knock-up) (L. A. Gilbert et al., 2013; Qi et al., 2013).

[Breakthrough Capability 1]: Ability to reliably create any precise, defined edit(s) (single nucleotide polymorphisms or gene replacement) with no unintended editing in any organism, with edits ranging from a single base change to the insertion of entire pathways.

- **2 years: Ability to generate any defined single base pair change in model organisms.**
    - [Bottleneck]: Performance of current editing technology and known-unknowns (e.g., chromatin, double-stranded break repair) and unknown-unknowns regarding basic biology.
        - [Potential Solution]: Improved base editing enzymes capable of catalyzing all possible nucleotide transitions.
        - [Potential Solution]: Engineered nucleases and recombinases to control repair of double-stranded breaks using either non-homologous end-joining or homologous recombination.
- **5 years: High-efficiency editing (beyond 90%) across the genome with no off-target activity.**
    - [Bottleneck]: A better understanding of canonical protospacer adjacent motif (PAM) specificities, non-canonical R-loop formation, types of double-stranded breaks, the effects of DNA supercoiling, and the double-strand DNA repair pathway mechanisms.
        - [Potential Solution]: A quantitative, predictive understanding of the coupled chromatin, editing, and repair interactions.
        - [Potential Solution]: A suite of genome editors covering all possible nucleotide (PAM) specificities.
        - [Potential Solution]: Genome editors with improved on-target and reduced off-target effects.

        - [Potential Solution]: Design algorithms for predicting single guide RNA (sgRNA) guide RNA sequences, sgRNA concentrations, and genome editor concentrations to achieve desired on-target activities with minimal off-target activities.
    - [Bottleneck]: Need for a better understanding of chromatin effects on editing in higher order systems.
        - [Potential Solution]: Fusion of epigenetic effectors to Cas9.
- **10 years: High-efficiency gene insertion or deletion of moderately large changes (but less than 10 kilobases) via homologous recombination.**
    - [Bottleneck]: The ability to manipulate double-stranded DNA break repair at high efficiency in non-model cells.
        - [Potential Solution]: Single-effector base editors that catalyze insertions and deletions with high efficiency.
        - [Potential Solution]: Improving the targeted delivery of DNA repair templates to nuclei in model and non-model organisms.
        - [Potential Solution]: Inducible control of DNA break repair pathways with significantly higher efficiency.
        - [Potential Solution]: Retron-mediated synthesis of DNA repair templates from co-delivered crRNAs, sgRNAs or RNPs.
    - [Bottleneck]: Ability to efficiently deliver large DNA constructs into cells.
        - [Potential Solution]: Expand viral packaging size.
        - [Potential Solution]: Develop and/or enhance non-viral DNA delivery tools and technologies.
- **20 years: Precise parallel editing or regulatory modifications (10 to 1000 modifications) across model and non-model organisms, including plants and animals.**
    - [Bottleneck]: Co-expressing many crRNAs or sgRNAs within arrays can trigger genetic instability, due the presence of many repetitive sequences.
        - [Potential Solution]: Toolboxes of highly non-repetitive CRISPR genetic parts can be designed and characterized, enabling the design of many-sgRNA arrays that do not trigger genetic instability.

[Breakthrough Capability 2]: Precise, predictable, and tunable control of gene expression for many genes inside diverse cells and organisms across different timescales.

- **2 years: Achieve long-lasting gene repression and activation.**
    - [Bottleneck]: Insufficient quantitative understanding of transcriptional regulation, epigenetic mechanisms, and cross-regulatory interactions.
        - [Potential Solution]: Quantitative characterization of additional genome editors with diverse protein fusions for more potent CRISPRa and CRISPRi.
        - [Potential Solution]: Systematic characterization of gene regulatory effects when changing RNP binding site locations, affinities, and RNP-PIC interactions.

        - [Potential Solution]: Engineered pioneer transcription factors that can reliably generate desired epigenetic states.
- **5 years: Ability to regulate expression in non-model organisms.**
    - [Bottleneck]: A quantitative understanding of promoter-specific, epigenetic-specific, and tissue-specific gene regulatory interactions.
        - [Potential Solution]: Organism- and tissue-specific promoters to express CRISPR components in desired non-model organisms.
        - [Potential Solution]: Organism- and tissue-specific protein fusion domains for potent activation or repression of genes in non-model organisms.
        - [Potential Solution]: Improved delivery and/or expression of CRISPR components into non-model organisms.
- **10 years: Technologies to monitor and manipulate genetic and epigenetic mechanisms controlling tissue-wide and organism-wide expression levels over time.**
    - [Bottleneck]: Insufficient methodologies for measuring and quantifying epigenetic, mRNA level, protein level, and metabolite level changes at genome-wide, single-cell resolution with sufficient precision to be relevant to clinical phenotypes.
        - [Potential Solution]: Single-cell epigenetic modification (histone modification, lncRNA, nucleosome position), protein level, and metabolite level determination.
        - [Potential Solution]: Single-cell, primary cell line measurements quantifying epigenetic modifications, protein levels, and metabolite levels, for example, using clinical tissues.
- **20 years: Ability to precisely regulate gene expression in whole-body organisms, including humans, with single-cell resolution using dynamic or static control.** This capability excludes germline genome editing.
    - [Bottleneck]: Improved delivery and expression of large genetic constructs in primary cell lines.
        - [Potential Solution]: Coupled transfection, genome editing, and genome repair to insert large genetic constructs into site-specific regions within chromosomes.
    - [Bottleneck]: Poorly understood tissue-specific cell states.
        - [Potential Solution]: Predictive models quantifying relationships between genetic, epigenetic states, and cellular phenotypes.
        - [Potential Solution]: Organ-, tissue-, and site-specific chromatin effectors with predictable/controllable effects on chromatin structure, gene expression, and gene accessibility.

[Breakthrough Capability 3]: Ability to reproducibly deliver editing cargo efficiently and specifically to a given target cells or tissues, and control dosage and timing of the editing machinery.

- **2 years: Improve editors to function without sequence requirements (such as protospacer adjacent motif (PAM) sequences) with activity comparable to 2019 state-of-the-art capabilities.**
    - [Bottleneck]: Current editors are large and prone to off-target activity.
        - [Potential Solution]: "Version 2.0" of high-fidelity editors.
        - [Potential Solution]: Smaller editors suitable for enhanced delivery.
        - [Potential Solution]: Regulated editors that are active in a context-dependent fashion (cell type, small molecule regulation, etc.).
- **5 years: Routine use of editors without detectable off-target effects (less than 0.001% off-target editing).**
    - [Bottleneck]: Limit of detection for current sequencing technologies.
        - [Potential Solution]: Technology improvement to lower detection limits and achieve more targeted detection.
        - [Potential Solution]: Reliable assays to discover potential off-target sites.
- **10 years: Enhance specificity of delivery modalities for high efficiency (>90% efficient) editing of cells in a defined tissue.**
    - [Bottleneck]: Current delivery modalities have low cell-type specificity.
        - [Potential Solution]: Large scale development of virus engineering for tropism/specificity.
        - [Potential Solution]: RNPs engineered with cell-type specificity via receptor interactions, and other modalities.
    - [Bottleneck]: Editing *in vivo* often leads to low efficiency of edits.
        - [Potential Solution]: Enhanced viral delivery with long half-life and low immunogenicity (likely needed for every organism of interest, including plant viruses and animal viruses).
        - [Potential Solution]: Viruses with large capacity (up-to or greater than 10 kilobase capacity) needed to deliver editor, guide RNAs, and any donor DNA molecule.
        - [Potential Solution]: Engineered effector complexes (such as RNPs) that can be delivered directly *in vivo* and maintain activity.
- **20 years: Quantitative, specific, and multiplexed editing of any site, in any cell, in any organism.**
    - [Bottleneck]: Specificity, efficiency, genetic stability, and off-target effects all pose challenges.
        - [Potential Solution]: Continued improvement of delivery vectors for plants, non-model animals, and other organisms.
        - [Potential Solution]: Development of toolboxes of non-repetitive CRISPR components to enable highly-multiplexed editing without triggering genetic instability.

- [Potential Solution]: Tools for editing in humans and animals for therapeutics, specifically those that overcome or mask the immune response.

Biomolecule, Pathway, and Circuit Engineering

(Biomolecular Engineering)

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*.

Host and Consortia Engineering

(Host Engineering)

Host and Consortia Engineering

Summary

Host and Consortia Engineering spans the development of cell-free systems, synthetic cells, single-cell organisms, multicellular tissues and whole organisms, and microbial consortia and biomes. Development of robust cell-free systems capable of diverse reactions, domestication and use of many single-cell hosts, targeted modification of multicellular organisms, and manipulation of microbial consortia.

Introduction and Impact

Engineering biology has delivered new tools to engineer microorganisms, plants, and animal cell lines. There are now entirely new ways to construct hosts to perform tasks that nature cannot accomplish. While many of these efforts have focused on 'traditional' hosts represented by model microbes like *E. coli* and *S. cerevisiae*, there is a wealth of potential if the unique capabilities of a broader range of microbes can be harnessed for useful purposes. These might include microbes that are photosynthetic, such as cyanobacteria (Markley, Begemann, Clarke, Gordon, & Pfleger, 2015), that can utilize non-sugar feedstocks such as methane or lignocellulose (Haitjema, Solomon, Henske, Theodorou, & O'Malley, 2014; Sundstrom & Criddle, 2015), or that can be engineered to produce and secrete complex macromolecules more efficiently than model hosts. The possibility that stable multi-organism consortia and biomes of defined compositions could be constructed is particularly tantalizing.

Cell-free biology has been a staple of life science research for more than 50 years. More recent technological achievements have created cell-free gene expression systems that can produce protein at titers reaching grams/liter, that can be constructed from non-model organisms, and that are greatly minimizing the time needed to prototype systems and circuits via the design-build-test cycle. The model-driven construction of complex cell-free systems as hosts could result in programmable hosts for advanced biosensing, for on-demand biomanufacturing, and even for the bottom-up construction of synthetic cells.

At present, engineering complex functions in non-model hosts remains difficult because many of the tools and approaches developed for model organisms cannot be applied with the same efficacy in non-model organisms. The profound impact that the development of similar tools and approaches for engineering non-model organisms would bring to two of the world's most important industrial sectors, energy and chemical production, more than justify attention to these challenges. It will be crucial to ensure that sufficient attention is given to biosecurity and biodefense risks that will rise along with improved capabilities for engineering diverse microbes (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). The pharmaceutical industry has long biomanufactured fermentation-based natural products in microbes and therapeutic proteins from mammalian cell culture, both of which could be dramatically improved through advancements in host engineering.

Compared to engineering in single cell hosts, the state-of-the-art for engineering multicellular systems and organisms is less well-developed. To date, these approaches have

been primarily aligned with natural reproduction, where genetically identical cells and tissues are created by editing the gametes or embryos of plants or animals. Gene editing methodologies introducing biochemical and molecular changes have already resulted in plants and animals with desirable characteristics that may be difficult to obtain through traditional breeding techniques. Further developing the capacity to reliably, and selectively, edit and modify multicellular eukaryotes could be transformative for a broad range of environmental and agricultural applications.

Host and Consortia Engineering focuses on the advancement of tools and technologies required for the characterization and engineering of host cells and organisms, and the integration and interaction of these systems and the environment. This includes developing methods, tools, and models to: 1) generate synthetic cells and cell-free systems to accomplish tasks and processes that cannot be accommodated by existing natural hosts; 2) enable organismal transformation, modification, and reprogramming of cellular chemistry, biology, and transport; 3) predict and integrate inputs and outcomes from environmental signals; and 4) enable the control, definition, and determination of differentiation, three-dimensional architecture, and other aspects of complex multicellular systems and biomes.

Transformative Tools and Technologies

Cell-free systems

Cell-free biology is the activation of complex biological processes without using intact living cells. While used for more than 50 years across the life sciences as a foundational research tool, a recent technical renaissance has made possible high-yielding cell-free gene expression systems (that produce protein in excess of grams/liter), the development of cell-free platforms from non-model organisms, and multiplexed strategies for rapidly assessing biological design-build-test cycles. These advances provide exciting opportunities to profoundly transform engineering biology through new approaches to model-driven design of genetic circuits, fast and portable sensing of compounds, on-demand biomanufacturing, building cells from the bottom up (i.e., synthetic cells), and next-generation educational kits. Key opportunities lie in understanding, harnessing, and expanding the capabilities of biological systems. For example, through the use of cell-free systems to inform cellular design, the efficiency of DNA synthesis can be amplified so that many different genes (encoding many different biomolecules) can be synthesized. The ability of cell-free systems to transcribe and translate a piece of DNA without the need to clone it into a specific vector and transforming into an organism (with all the limits associated with DNA transformation efficiencies) enables cell-free systems to shorten the time to testing, thus speeding up the overall design-build-test cycle, and enabling the scalable prototyping of gene function. However, to date, there are limited numbers of large datasets available that allow comparison of part performance between the cell-free environment and in cells (*in vivo*). One need is to make data and models available to the community so that others can build and test improved models leveraging already developed systems and data. In another direction of research, there is a need to investigate the use of cell-free systems in manufacturing. Imagine how rapid access to vaccines and therapeutics in remote settings could change lives, and how new biomanufacturing paradigms suitable for use in low resource

settings might promote better access to costly drugs through decentralized production. “Just-add-water” freeze-dried, cell-free systems could offer a disruptive approach to emerging and re-emerging diseases threats. It is a paradigm shifting concept.

Tools for engineering and characterization of host organisms

Today, we have a number of host organisms for which we have a satisfactory understanding of their metabolism and sufficient genetic tools that we can use for reliable engineering. However, there are a number of applications that beg for more suitable chassis. There is a need to develop new tools for existing organisms, as well as entirely new platform organisms, and capabilities compatible with high-throughput, data-driven workflows that are becoming increasingly favored in industrial biotechnology. Key capabilities needed across organisms include reliable transformation methods for plasmid delivery and genome integration, and well-characterized genetic parts (including promoters and terminators) to regulate gene expression (Johns et al., 2018). Predictive models for gene and protein expression-timing and -levels are also needed. Through the successful engineering of a broader library of host cells and multicellular organisms, we can increase the number of reporters and tools to better understand biology, establish new living sensors and sentinels, and expand the production of polymers, metabolites, and numerous other products.

Host onboarding and transformations

One area of host engineering where progress still needs to be achieved is the on-boarding of engineered genetic sequences, circuits, and pathways into host cells. Crucial to this is a detailed understanding of the central dogma machinery and a systems-level understanding of host physiology such that genes can be reliably expressed and (synthetic) pathways incorporated without negatively impacting fitness. This will include an understanding and prediction of the endogenous gene regulator elements, including transcription factors, important DNA cis elements, regulatory RNAs, and the role of chromatin and epigenetic markings. Furthermore, there is the need for fully sequenced and annotated genomes for the majority of organisms; this can be extended to fully annotated metabolic pathways and enzyme activities. Advancements in genetic transformations (or viral transductions) and the ability to manipulate the genome, and ultimately, the ability to transplant chromosomes, would enable more robust design, control, and/or domestication of host organisms and their functional cellular machinery.

HOST AND CONSORTIA ENGINEERING

Goal	Breakthough Capability	Milestone

Cell-free systems capable of natural and/or non-natural reactions.

2 Years	5 Years	10 Years	20 Years
Ability to build reproducible and comparable cell-free systems for practical applications in bioengineering and biomanufacturing from multiple organisms, including non-model hosts.			
Complete characterization of the general effects of cell-growth harvest conditions and extract preparation parameters on bacterial cell-free extract behavior.	Complete standardization of common-use bacterial cell-free system.	Complete library of user-defined reaction components that could be used in a customizable cell-free system.	Consistent ability to generate cell-free systems from any organism or a subset of organisms that make all types of desired products.
Ability to build a cell, including the molecular subsystems that enable the processes of DNA replication, transcription, translation, energy regeneration, and membrane construction.			
Demonstrated ability to synthesize all components encoded by a (minimal/synthetic) cell using cell-free systems.	Demonstrate and design a minimal genome that could support the construction of a cell, including regulation.	Ability to have ribosomes make ribosomes in a cell-free system.	Engineer compartmentalization and communication strategies for the design of (synthetic) cells.
	Ability to build metabolic modules capable of supporting long-lasting energy regeneration.	Expand the chemistry of living systems to make chemical reactions not possible with biological chemistry alone.	Replace test tubes with chemically-defined, standardized micro-vesicles to compartmentalize processes.
Long-lasting, robust, and low-cost cell-free system for protein synthesis and biomanufacturing.			
Identify reagent instabilities in cell-free systems across multiple organisms and all biological kingdoms.	Alleviate reagent instabilities and prolong the half-life of cell-free reagents from a few hours to several days using inexpensive substrates.	Stabilize catalysts to facilitate cell-free reactions on the order of weeks.	Robust and scalable production of cell-free systems that last for weeks.
	Avoid inhibition (poisoning) of cell-free reactions by byproducts or the desired products.		
Ability to use cell-free systems to inform cellular design of genetic parts and circuits.			
Ability to use next-generation sequencing read-outs to quantitatively map performance of genetic designs in cell-free systems.	Ability to identify new genetic parts in cell-free systems for any bacterial host to facilitate forward engineering in cells.	Ability to identify new genetic circuits in cell-free systems for any bacterial host to facilitate forward engineering in cells.	Ability to identify new genetic circuits in cell-free systems for any eukaryotic host to facilitate forward engineering in cells.
			Accelerate the development of any non-model host into useful chassis organisms for engineering biology with cell-free systems.
Decentralized, portable, on-demand sensing and manufacturing using cell-free systems.			
Ability to use safe lysates low in endotoxin for sensing and manufacturing objectives.	Demonstrate portability (two-year storage duration of freeze-dried reactions without loss of functionality) of cell-free systems.	Point-of-care cell-free protein production system ready for validation by the Food and Drug Administration.	Point-of-care cell-free protein therapeutic and vaccine production system ready for validation by the Food and Drug Administration.
	Increase productivity and rate of cell-free reactions.		

Ability to manufacture any targeted glycosylated protein or metabolite using cell-free biosynthesis.			
Ability to build modular, versatile cell-free platforms for glycosylation pathway assembly.	Expanded set of glycosylation enzyme-variants that efficiently install eukaryotic glycans.	Expanded set of enzymes capable of glycosylating metabolites *in vitro*.	Ability to produce any glycosylated protein therapeutics and vaccines at the point-of-care in less than one week.
	Production of bacterial glycoconjugate vaccines in cell-free systems.	Cell-free pipelines to produce and assess the functionality of diverse, human glycosylated protein therapeutics	

On-demand production of single-cell hosts capable of natural and non-natural biochemistry.

Ability to grow any host, anytime, in a controlled and regulated setting.			
Establish protocols for the development of media that support cellular viability for non-model organisms.	Develop robust, high-throughput screens for rapidly assaying useful properties in libraries of organisms.		
Robust screening of useful chassis beyond model organisms.	Use output of high-throughput screens/sensors and computer control to amplify a signal or expand a cell line that produces a product of interest.		

Routine domestication of non-model organisms through DNA delivery and genetic modification.			
Catalog and assay current methodologies and tools for carrying out DNA delivery in microbial/mammalian systems and plant systems.	Development of well-characterized and robust insertion sites in plant genomes.	Develop high-throughput, targeted editing and rapid-genome-evolution tools that couple genetic changes to phenotypic changes.	Routine genetic manipulation of any non-model host in less than one week from first isolation.
Develop high-throughput methods that can be done in parallel for DNA delivery (using standard methods) into non-model hosts.	Develop high-throughput, genome-wide editing tools for non-model organisms.		
Establish a suite of gene-editing tools for the rapid insertion and/or deletion of genetic elements in diverse primary mammalian cells.	Establish robust temporal and/or spatial control of gene expression in mammalian cells.	Develop universal approaches to transforming any plant.	
Characterize basic DNA parts for expression strength in non-model organisms.	Develop broad-host-range vectors for a variety of model and non-model organisms.		

Ability to build and control small molecule biosynthesis inside cells by design or through evolution.			
Identify model organisms for performing specific types of chemistries or organisms that have native precursor biosynthesis pathways for specific classes of molecules.	Construct a limited number of model host organisms for synthesizing all-natural products.	Software and hardware for optimizing titer, rate, and yield of any product produced by any host.	On-demand construction of single cell organisms for production of nearly any molecule of interest, including organic chemicals and polymers.
Precise temporal control of gene expression for well-studied systems.	Construction of single-cell organisms for production of unnatural derivatives of natural products.		
	Temporal control over multiplexed regulation of many genes in parallel.		

2 Years	5 Years	10 Years	20 Years

Spatial control over, or organization of, metabolic pathways in cells and construction of unnatural organelles.			
Tools to target heterologous proteins to various subcellular compartments.	Inducible synthesis of organelles.	Methods and tools to reprogram transport of metabolites and compartmentalization of biochemical reactions.	
		Alter chemical conditions within the organelle/microcompartment.	
	Gain-control for selective permeability in and out of the organelle.	Multiple orthogonal organelles/ microcompartments in the same cell for compartmentalizing different parts of a pathway.	

Production and secretion of any protein with the desired glycosylation or other post-translational modifications.		
One or more microbial hosts capable of producing laboratory-scale quantities of a single glycoform of a desired protein.	A few microbial hosts capable of secreting functional versions of proteins with no post-translational modifications.	Ubiquitous control of post-translational modification in a diverse array of hosts.

On-demand fabrication and modification of multicellular organisms.

Ability to control differentiation and de-differentiation of cells within a population.	
On-demand, reproducible functionalization of simple micro-tissues or micro-consortia made up of two or more engineered cell types.	Programmable and regulatable pathways that can be induced to differentiate or de-differentiate somatic cells.

Ability to characterize and control the three-dimensional architecture of multicellular systems.			
Characterize existing tissue components and standardize measurements to evaluate function.	Identification of novel 3D scaffold designs that can lead to desirable cellular properties.	Create modular, synthetic communication circuits that can be implemented in tissues to allow for control of new or existing cellular communication systems.	Bottom-up design and construction of whole organs at the centimeter-length scale.

Ability to achieve stable non-heritable changes in somatic cells.			
Routine delivery of biomolecule "effectors" (i.e., DNA, RNA, proteins) into slowly-dividing or non-dividing cells.	Generation of effective artificial epigenetic chromosomal states and maturation of the emerging field of chromatin engineering.	Ability to generate cell states that are stable and effective after the inducer/effector is removed in certain model tissues.	Nimble adaptation of somatic cell engineering technologies to any natural tissue at any developmental stage.

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

Ability to make predictable and precise, targeted, heritable changes through germline editing.			
Complete sequence of select host genomes to allow design of targets for gene editing.	Efficient germline transformation systems developed in targeted hosts.	Ability to coordinate engineered multicellular functions in intact organisms via orthogonal communication systems.	Routine, on-demand, efficient germline editing for any targeted host of interest at high-throughput scale.
	Ability to deliver transgene constructs to most (>90%) somatic cells in a higher organism to rapidly prototype transgenic phenotypes.		
	Temporally controlled transgene expression that works on the scale of generations.		
Define and validate tissue-specific DNA parts in plants.	Efficient gene editing in differentiated cells.	On-demand gene editing of organisms with desired traits.	
	Ability to domesticate engineered biological parts to confer immune tolerance in immunocompetent organisms.		

Generation of biomes and consortia with desired functions and ecologies.

Ability to control cell-to-cell communication between different species.			
Tightly-controlled promoter-response regulator systems that enable intra- and inter-species cellular communication.	Synthetic cell-to-cell communication elements and networks that function in a broad range of host organisms.	Signal-response pathways that function in synthetic communities of 5-10 organisms, employing a variety of pathway types and host species.	Ability to produce engineered microbes that can reliably invade and coexist within a complex community and manipulate the consortium/ biome function and behavior.
Ability to characterize, manipulate, and program three-deminsional architecture of the biome.			
Use of existing technologies to better understand the species composition and collective components of microbial communities and consortia.	Non-destructive, 3D visualization of microbial communities from a broad range of environments.	Ability to manipulate the 3D architecture of natural or engineered communities using external inputs.	Programmed communities that self-assemble into a desired 3D architecture.
Ability to control and/or define the function of an engineered microbial community/biome.			
Ability to combine species with specialized functions to enable the production of desired products.	Assembly of consortia to produce desired molecules/products, considering community-level metabolic flux.	Plug-and-play assembly of consortia to produce desired molecules/products from specific starting materials, considering community level metabolic flux and organism-to-organism communication.	On demand assembly of consortia that are programmed to respond dynamically.
Targeted modification of an existing microbiome to enable new functions or address dysbiosis through the addition, removal, or reorganization of the community members.			
Use of existing technologies to characterize functions of microbial communities from a broad range of environments.	Characterize how select microbiomes respond to changes in the environment.	Predictive models of microbiome function and response to a broad range of environmental and ecological changes.	Ability to modify an existing biome or consortia as desired.

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

Roadmap Elements

Goal 1: Cell-free systems capable of natural and/or non-natural reactions.

[Current State-of-the-Art]: Cell-free synthetic biology is emerging as a transformative approach to understand, harness, and expand the capabilities of natural biological systems. The foundational principle is that complex biomolecular transformations are conducted without using intact cells. Instead, crude cell lysates, or extracts, are used, which provides a unique freedom-of-design to control biological systems for a wide array of applications. For example, cell-free protein synthesis (CFPS) systems have been used to decipher the genetic code, prototype genetic circuits (Moore et al., 2018; Takahashi et al., 2015) and metabolic pathways (Karim & Jewett, 2016), enable portable diagnostics (Pardee et al., 2016; Wen et al., 2017), facilitate on-demand biomolecular manufacturing (Pardee et al., 2016), produce antibody therapeutics at the commercial scale (Yin et al., 2012), and enable advances in education (Huang et al., 2018; Stark et al., 2019, 2018). The recent surge of applications has revitalized interest in cell-free systems, especially in areas where limits imposed by the organism may impede progress. Despite these advances, several barriers limit advancement of the field. Specifically, there are opportunities to: (i) standardize lysate generation approaches, (ii) enable decentralized manufacturing of complex therapeutics and vaccines, (iii) establish design principles for genetically-encoded biosensors to rationalize their engineering for addressing global sustainable development goals (e.g., food and water security) including portable and on-demand strategies, (iv) reduce costs of cell-free reactions by enabling long-lived protein expression, (v) generate large datasets and quantitative models to allow comparison of part performance between the cell-free environment and in cells (*in vivo*), (vi) synthesize more complex classes of proteins such as glycoproteins, (vii) construct synthetic cellular machines (e.g., ribosomes) and biosynthetic modules to both understand life and lead to new manufacturing paradigms, and (viii) build cells from the bottom up.

[Breakthrough Capability 1]: Ability to build reproducible and comparable cell-free systems for practical applications in bioengineering and biomanufacturing from multiple organisms, including non-model hosts.

- **2 years: Complete characterization of the general effects of cell-growth harvest conditions and extract preparation parameters on bacterial cell-free extract behavior (e.g., protein synthesis and native genetic regulators).**
    - [Bottleneck]: Inability to produce cell-free systems in a scalable and standardized fashion, protocols for cell-free systems generation vary from lab-to-lab, and reactions are carried out at very small volumes (1-10µl) that limit the extent to which we can characterize the systems; for example, we lack an understanding of how the metabolic state of an extract impacts energy usage, protein synthesis, and DNA/RNA regulation.
        - [Potential Solution]: Arrive at a standardized bacterial cell-free system extract generation and reaction protocol that is robust, inexpensive, and would allow their routine scalable production; this includes identification of energy mixtures robust to various extract preparation procedures.

        - [Potential Solution]: Use -omics tools (e.g., mass spectrometry, next generation sequencing) to understand batch-to-batch and lab-to-lab lysate quality and composition variability.
        - [Potential Solution]: Define an equipment set for bacterial lysate preparation (i.e., lysis technique).
- **5 years: Complete standardization of common-use bacterial cell-free system.**
    - [Bottleneck]: *Ad hoc* use of different extract preparation procedures, cell-free reaction conditions, and reporter construct architectures lead to challenges in making reproducible and comparable cell-free systems.
        - [Potential Solution]: Design cell-free system physiochemical composition that is robust to different process parameters, or identify compositions for each defined extract preparation method.
        - [Potential Solution]: Identify and implement measurement techniques needed to facilitate reproducibility (e.g., standard assay for quantifying activity across labs).
        - [Potential Solution]: Rigorously characterize and understand how reaction geometries and surface area-to-volume impact cell-free system performance.
        - [Potential Solution]: Create a plasmid repository specific to bacterial cell-free systems (including T7 constructs and *E. coli* promoters), with a description of performance (such as protein expression kinetics).
        - [Potential Solution]: Develop a quantitative ‘culture’ to cell-free system practices.
- **10 years: Complete library of user-defined reaction components for use in a customizable cell-free system.**
    - [Bottleneck]: Gaps in understanding critical components and how they vary across species and kingdoms (e.g., prokaryotes vs. eukaryotes).
        - [Potential Solution]: Assess critical parameters with design-of-experiments and machine learning across strains and procedures.
        - [Potential Solution]: Develop an online tool to customize the components of a cell-free reaction for a given organism and approximate RNA/protein output, reveal components that should have been included, and application-specific assembly (metabolic engineering vs. protein synthesis vs. circuits).
    - [Bottleneck]: Ability to regenerate energy and cofactors with native metabolism is constrained in non-*E. coli* platforms.
        - [Potential Solution]: Assess metabolic pathways that could be activated to facilitate energy regeneration.
        - [Potential Solution]: Develop exogenous cofactor regeneration modules to drop in and out when native ones are missing.
        - [Potential Solution]: Incorporate genome-scale models to predict energy and cofactor regeneration systems most suitable for a new host.

- **20 years: Consistent ability to generate cell-free systems from <u>any organism or a subset of organisms that make all types of desired products</u>, including all biological kingdoms and DNA programmed cell-free systems at-scale.**
    - [Bottleneck]: Gaps in understanding critical components of host lysates that make each unique and which to use for a specific, desired product.
        - [Potential Solution]: Identify the critical species-specific components (using -omics approaches) and experimentally validate these components (including transcription factors, accessory proteins for translation, and metabolic modules).
    - [Bottleneck]: Production of lysate relies on scaled-up culturing of sometimes recalcitrant cells.
        - [Potential Solution]: Understand which extract preparation protocol parameters should/can be tuned to match a new organism (i.e., trace the physiology of an organism to the extract preparation parameters).
    - [Bottleneck]: Identify units of operation required for the scale-up and scale-out production of cell-free systems.
        - [Potential Solution]: Identify quality assurance and quality control metrics at each unit of operation that would allow the scale-up of the process.

[Breakthrough Capability 2]: Ability to build a cell, including the molecular subsystems that enable the processes of DNA replication, transcription, translation, energy regeneration, and membrane construction.

- **2 years: Demonstrated ability to synthesize all components encoded by a minimal or synthetic cell using cell-free systems.**
    - [Bottleneck]: Multiple minimal genomes necessary for building a cell have been proposed but none have yet been demonstrated to work in this context.
        - [Potential Solution]: Define a set of possible minimal genomes that could enable self-replication of a minimal cell and make these component parts.
        - [Potential Solution]: Test each component of a minimal cell individually (including metabolic modules, regulation by most types of regulators, etc.).
- **5 years: Demonstrate and design a minimal genome that could support the construction of a cell, including regulation.**
    - [Bottleneck]: We do not yet know how to build genomes *de novo*, and native genomes have built-in regulation that may not be suitable for engineered biological systems.
        - [Potential Solution]: Create and refactor modularized pathways for building a minimal cell.
        - [Potential Solution]: Show examples of a synthetic self-regulating gene cluster.
        - [Potential Solution]: Exploit CRISPR as a potential universal regulation mechanism to regulate synthetic minimal genomes.

- **5 years: Ability to build metabolic modules capable of supporting long-lasting energy regeneration.**
    - [Bottleneck]: Typical cell-free reactions that serve as the basis for minimal cells use an energy-rich environment of nucleotide triphosphates and high energy phosphate bond donors for chemical energy; these systems are short-lived, expensive, and can lead to inhibitory byproducts.
        - [Potential Solution]: Create generalized approaches for cost-effective, long-lived energy regeneration modules that can integrate with synthetic cells, which may require compartmentalized systems.
        - [Potential Solution]: Integrate efficient physical ATP regeneration systems in synthetic cells, such as photosynthesis.
- **10 years: Ability to have ribosomes make ribosomes in a cell-free system.**
    - [Bottleneck]: Constructing ribosomes built completely of *in vitro* synthesized parts has remained elusive.
        - [Potential Solution]: Develop conditions that facilitate co-synthesis of all ribosome proteins and ribosomal RNA, first in extracts, then in purified systems.
- **10 years: Expand the chemistry of living systems to make chemical reactions not possible with biological chemistry alone.**
    - [Bottleneck]: The palette of biological chemistry is smaller than chemistry.
        - [Potential Solution]: Build hybrid biological-chemical systems.
        - [Potential Solution]: Expand the chemistry of genetically-encoded systems.
- **20 years: Engineer compartmentalization and communication strategies for the design of synthetics cells.** (For related reading, please see Goal 2, Breakthrough Capability 4: Spatial control over (or organization of) metabolic pathways in cells and construction of unnatural organelles.)
    - [Bottleneck]: Lack a precise physical understanding of how the composition of the compartment influences encapsulation, transport, and retention of various types of cargo.
        - [Potential Solution]: Quantify the relationship between compartment composition and encapsulation efficiency, permeability, and stability as a function of compartment cargo.
    - [Bottleneck]: Communication strategies for synthetic cells are not yet well-developed.
        - [Potential Solution]: Identify and implement mechanisms for communication between engineered compartments that allows efficient, reliable, multi-channel signaling between synthetic cells.
        - [Potential Solution]: Identify complementary signaling modules with reliable performance in a cell-free environment.

- **20 years: Replace test tubes with chemically-defined, standardized micro-vesicles to compartmentalize processes.**
    - [Bottleneck]: Inability to generate stable liposomes with the correct size and membrane-permeability to generate micro-vesicles (synthetic cells) with different chemical environments.
        - [Potential Solution]: Use microfluidics for precise production of chemically-defined vesicles.
        - [Potential Solution]: Exploit non-natural compartments, such as block copolymers, as a means to make mechanically robust compartments with negligible non-specific permeability.

[Breakthrough Capability 3]: Long-lasting, robust, and low-cost cell-free system for protein synthesis and biomanufacturing.

- **2 years: Identify reagent instabilities in cell-free systems across multiple organisms and all biological kingdoms.**
    - [Bottleneck]: Stability of reagents is not well understood in non-*E. coli*-based cell-free systems.
        - [Potential Solution]: The full spectrum of metabolic activity during cell-free reactions over time can be characterized in multiple systems, including those emerging cell-free systems, to identify reagent instabilities.
- **5 years: Alleviate reagent instabilities and prolong the half-life of cell-free reagents from a few hours to several days using inexpensive substrates.**
    - [Bottleneck]: High-energy phosphate compounds are still commonly used to fuel cell-free systems (e.g., PEP, 3PGA); these compounds are expensive, lead to inhibitory phosphate concentrations, and only enable bursts of ATP regeneration instead of long-lived energy regeneration.
        - [Potential Solution]: Central metabolism, non-phosphorylated energy substrates, or light driven approaches (e.g., from bacteriorhodopsin), among others, could be used to enable long-lived energy regeneration without inhibitory byproducts; should be tested in extracts made from organism across all kingdoms and identify key factors (such as, concentrations, additional reagents, etc.) for robust alternative energy activation.
    - [Bottleneck]: Cofactor regeneration and balancing strategies preclude long-term activation of energy metabolism to fuel metabolic activity.
        - [Potential Solution]: Molecular purge valves implemented in crude extracts could maintain redox balance.
        - [Potential Solution]: Develop schemes for balancing ATP, and derivatives including NAD(P) and FAD.
        - [Potential Solution]: Create non-natural cofactors, and requisite engineered enzymes, that have better stability properties.
    - [Bottleneck]: Reagent instabilities limit reaction time.
        - [Potential Solution]: Genomic modifications to extract source strains can be carried out to stabilize substrates.

- **5 years: Avoid inhibition (poisoning) of cell-free reactions by byproducts or the desired products.**
    - [Bottleneck]: Small molecule byproducts of metabolism, such as phosphate, inhibit cell-free reactions.
        - [Potential Solution]: Engineer systems to avoid the accumulation of chemical inhibitors.
        - [Potential Solution]: Develop product siphoning strategies and dialysis-like strategies to remove inhibitors from the system.
        - [Potential Solution]: Engineer molecular complexes that are resistant to byproducts and chemical products.
- **10 years: Stabilize catalysts to facilitate cell-free reactions on the order of weeks.**
    - [Bottleneck]: Cell-free reactions terminate because substrates and cofactors are depleted, byproducts accumulate to inhibit the reaction, or the catalysts become inactivated; of these, enzyme stability represents a significant technical and economic hurdle to technology development in this space.
        - [Potential Solution]: Characterize catalyst instabilities such as, tolerance to byproducts and inhibitors.
        - [Potential Solution]: Increase stability of enzymes involved in cell-free systems.
        - [Potential Solution]: Generate design criteria for choosing catalysts (i.e., identifying select organisms from which to derive lysates and enzymes) for desired applications.
        - [Potential Solution]: Develop reactor designs and bioprocesses that continuously replenish the source of catalysts.
- **20 years: Robust and scalable production of cell-free systems that last for weeks.**
    - [Bottleneck]: Without enzymes or lysates that are stable on the order of weeks, significant fractions of carbon will otherwise be used in generating the biocatalysts required of these systems.
        - [Potential Solution]: Integrate knowledge of the system and innovations to facilitate long-lived reactions.

[Breakthrough Capability 4]: Ability to use cell-free systems to inform cellular design of genetic parts and circuits.

- **2 years: Ability to use next-generation sequencing read-outs to quantitatively map performance of genetic designs in cell-free systems.**
    - [Bottleneck]: Sequence-function data are limited by colorimetric and fluorescent read-outs.
        - [Potential Solution]: Develop next-generation, deep-sequencing-based approaches for monitoring transcription and translation in cell-free reactions, including transcription factor metabolite interactions.

- **5 years: Ability to identify new genetic parts in cell-free systems (including promoters, ribosome binding sites, and terminators) for any bacterial host to facilitate forward engineering in cells.**
    - [Bottleneck]: There are limited numbers of sufficiently-large datasets available that allow comparison of genetic part performance and the development of modeling frameworks between the cell-free environment and *in vivo*.
        - [Potential Solution]: Develop cell-free systems for 20 industrially-relevant organisms that could form a testbed to establish libraries of new genetic parts and how to accelerate design.
        - [Potential Solution]: Develop a repository of genetic parts for cell-free systems, including performances.
        - [Potential Solution]: Develop a quantitative modeling platform specific to cell-free systems that takes into account the advantages and limitations of cell-free expression.
- **10 years: Ability to identify new genetic circuits in cell-free systems for any bacterial host to facilitate forward engineering in cells.**
    - [Bottleneck]: Resource limitations in cell-free systems (e.g., energy and cofactor regeneration) constrain the construction of multi-gene systems encoded by complex genetics, as well as time-dynamics needed to assess their function.
        - [Potential Solution]: Create robust, long-lived cell-free systems that can be routinely used for activating and characterizing multi-gene circuits.
    - [Bottleneck]: There are limited numbers of sufficiently-large datasets available that allow comparison of genetic circuit performance and the development of modeling frameworks between the cell-free environment and *in vivo*.
        - [Potential Solution]: Develop cell-free systems for 20 industrially-relevant organisms which could form a testbed to establish libraries of new genetic circuits.
- **20 years: Ability to identify new genetic circuits in cell-free systems for any eukaryotic host to facilitate forward engineering in cells.**
    - [Bottleneck]: Lack of knowledge of how to activate essential components in eukaryotic cell-free systems.
        - [Potential Solution]: Make more lysates from eukaryotes and eukaryotic-like systems to be able to assess essential components.
    - [Bottleneck]: Transcription and translation are typically combined in cell-free systems requiring viral sequences (e.g., internal ribosome entry site) for translation initiation rather than a 5'cap and polyA tail.
        - [Potential Solution]: Develop strategies to compartmentalize transcription to better mimic the natural process.
- **20 years: Accelerate the development of any non-model host into useful chassis organisms for engineering biology with cell-free systems.**
    - [Bottleneck]: Sufficient transcription and translation activity is necessary to assess genetic designs and metabolic pathways.

        - [Potential Solution]: Streamline and provide generalized approaches to enable sufficient cell-free activity for gene expression and biosynthesis from diverse species.
        - [Potential Solution]: Use mass spectrometry to determine the proteome composition and metabolite composition of non-model organisms to accelerate the optimization of cell-free systems.
    - [Bottleneck]: Expand cell-free systems to poorly explored ares, such as extremophiles, for engineering biology far from standard conditions.
        - [Potential Solution]: Select a set of extremophiles that can be grown in laboratories and easily lysed.

[Breakthrough Capability 5]: *Decentralized, portable, on-demand sensing and manufacturing using cell-free systems.*

- **2 years: Ability to use safe lysates low in endotoxin for sensing and manufacturing objectives.**
    - [Bottleneck]: Toxicity of cell-free components is not well understood.
        - [Potential Solution]: Evaluate toxicity of cell-free components.
    - [Bottleneck]: Lipopolysaccharides, also known as lipoglycans and endotoxins, are large molecules consisting of a lipid and a polysaccharide composed of O-antigen which can cause toxicity of products manufactured using bacteria.
        - [Potential Solution]: Cost-effective strategies to remove endotoxin is required to ensure safety.
        - [Potential Solution]: Create bacterial species that are detoxified by design.
- **5 years: Demonstrate portability (such as two-year storage of freeze-dried reactions without loss of functionality) of cell-free systems.**
    - [Bottleneck]: A major limitation of traditional sensors and centralized medicine manufacturing is that the products must be refrigerated; cell-free systems can be freeze-dried for potential room-temperature storage and distribution, but suffer from activity loss in certain conditions.
        - [Potential Solution]: Create stable, freeze-dried systems for storage without a cold-chain by using cryoprotectants and process modifications.
        - [Potential Solution]: Demonstrate multiple reaction formats (i.e., pellets, gels, paper, etc.) stable for 2-5 years.
- **5 years: Increase productivity and rate of cell-free reactions.**
    - [Bottleneck]: Manufacturing medicines in rapid response to emerging and re-emerging threats requires fast protein synthesis rates and reactions that can be completed in minutes to hours.
        - [Potential Solution]: For translation outputs, increase the overall catalyst concentration in the reaction to enhance reaction rates.
        - [Potential Solution]: When possible, develop transcriptional outputs (such as for sensors) which offer a significant speed improvement (minutes versus hours).

- **10 years: Point-of-care cell-free protein production system ready for validation by the Food and Drug Administration (FDA).**
    - [Bottleneck]: Completely automated operation platform to manufacture and purify proteins suitable for the FDA has not yet been fully validated.
        - [Potential Solution]: Build integrated units and measure the repeatability (of the same unit) and reproducibility (between units) of the system.
- **20 years: Point-of-care cell-free protein therapeutic and vaccine production system ready for validation by the Food and Drug Administration (FDA).**
    - [Bottleneck]: Point-of-care synthesis and administration of glycoprotein therapeutics and vaccines requires robust production and purification methods.
        - [Potential Solution]: Develop simple, portable purification systems which can reliably produce FDA-compliant vaccines and therapeutics from cell-free production systems and demonstrate efficiency in animal models.

[Breakthrough Capability 6]: Ability to manufacture any targeted glycosylated protein or metabolite using cell-free biosynthesis.

- **2 years: Ability to build modular, versatile cell-free platforms for glycosylation pathway assembly.**
    - [Bottleneck]: Many of the most important components of glycosylation pathways are associated with cellular membranes and cannot be recapitulated easily in cell-free systems.
        - [Potential Solution]: Develop and optimize efficient strategies use of oligosaccharyltransferases (including eukaryotic versions) to transfer pre-built sugars from lipid-linked oligosaccharides *in vitro.*
        - [Potential Solution]: Develop methods to select and assemble a set of soluble enzymatic tools capable of producing therapeutically-relevant glycoproteins with desired properties *in vitro* from simple, commercially available activated-sugar building blocks.
- **5 years: Expanded set of glycosylation enzyme-variants that efficiently install eukaryotic glycans.**
    - [Bottleneck]: Synthesis of complex human glycans in cell-free systems is constrained by the available set of well characterized enzymes; existing characterizations do not provide sufficient tools to go from a set of sugar monomers and a design to a glycoprotein of interest.
        - [Potential Solution]: Expand the glycoengineering toolkit by characterizing glycoslytransferases to assemble sets of well-characterized enzymes that can reliably produce desired glycoproteins.
        - [Potential Solution]: Engineer existing glycosylation enzymes for desired activities when naturally occurring enzymes with desired functionalities are not available (e.g., engineering of the bacterial oligosaccharyltransferase to accept the eukaryotic core glycan would enable the efficient production of the eukaryotic core glycan in bacterial systems).

- **5 years: Production of bacterial glycoconjugate vaccines in cell-free systems.**
    - [Bottleneck]: The production of glycoconjugate vaccines against bacteria require the culturing of pathogenic strains recalcitrant to bioengineering and the use of non-specific conjugation chemistry.
        - [Potential Solution]: Express diverse bacterial O-antigen pathways for characterization and production of lipid-linked oligosaccharides required for vaccines and use bacterial oligosaccharyltransferases to site-specifically attach these oligosaccharides *in vitro*.
        - [Potential Solution]: Use the greater control afforded by cell-free protein synthesis systems to produce FDA-approved vaccine carrier and antigen proteins, many of which cannot be produced outside of pathogenic organisms.
- **10 years: Expanded set of enzymes capable of glycosylating metabolites *in vitro*.**
    - [Bottleneck]: Glycosylation of therapeutically-relevant metabolites (including many antibiotics) is often required for desirable pharmacokinetic/dynamic properties, but many of the enzymes which perform these glycosylation activities are unknown or can only be expressed in their native host strain.
        - [Potential Solution]: Use cell-free protein synthesis to screen the activity of metabolite-targeting glycosyltransferases on existing bioactive compound libraries to understand their specificities and improve the therapeutic utility of small molecule products.
    - [Bottleneck]: Substrate limitations lead to inherent inefficiencies.
        - [Potential Solution]: Develop metabolic models of lysate hosts to identify genetic knockouts that enhance glycan production and carry out such modifications.
- **10 years: Cell-free pipelines to produce and assess the functionality of diverse, human glycosylated protein therapeutics.**
    - [Bottleneck]: Development timelines of glycoprotein therapeutics are slowed by the need to produce these products in mammalian cell lines.
        - [Potential Solution]: Develop, optimize, and implement strategies to create more than ten unique and homogeneous glycan structures on proteins by cell-free methods (e.g., trimannose core eukaryotic glycan, afucosylated galactose-terminated core glycan, fucosylated sialic-acid terminated core glycan, biantennary glycan, etc.); once synthesized, these products can studied and optimized using therapeutic functionality assays.
- **20 years: Ability to produce any glycosylated protein therapeutics and vaccines at the point-of-care in less than one week.**
    - [Bottleneck]: Intentional engineering of glycan structures and the synthesis of novel structures is constrained by identifying sets of enzymes for the manufacture of the sugar structures.
        - [Potential Solution]: Develop cell-free systems capable of rapidly and robustly producing any defined glycoprotein on-demand.

Goal 2: On-demand production of single-cell hosts capable of natural and non-natural biochemistry.

[Current State-of-the-Art]: Current tools and technologies for on-demand production of organisms are limited by the number and scope of transformation capabilities, continuous and rapid production capability, and the lack of secure public repositories for academics and industry containing the necessary organismal design and characterization information. Today, we have a number of host microbes for which we have a satisfactory, though not extensive, understanding of their metabolism and sufficient genetic tools that we can use for reliable engineering. Engineering of plant and animal cells is expanding, especially given particular applications (e.g., CAR-T cell engineering), but still faces significant bottlenecks.

[Breakthrough Capability 1]: Ability to grow any host, anytime, in a controlled and regulated setting.

- **2 years: Establish protocols for the development of media that support cellular viability for non-model organisms.**
    - [Bottleneck]: Inability to grow most organisms on earth in non-natural environments (i.e., inside the lab) and a lack of knowledge of conditions for growing novel or non-model organisms.
        - [Potential Solution]: Database collection and mapping: a better understanding of how to find what is important for organismal growth and survival, potentially including the ability to apply artificial intelligence to recognize patterns.
        - [Potential Solution]: Develop a general protocol for identifying viable media and create databases of media that work with known organisms; leverage machine learning using various databases (e.g., 16S rRNA profiling to identify evolutionary relationships between known and unknown species) to help generate suggestions for potential media.
- **2 years: Robust screening of useful hosts beyond model organisms.**
    - [Bottleneck]: There are a few hosts that are the primary chassis for engineering applications, but they are not always the best choice for biosynthesis of all potential products.
        - [Potential Solution]: Expand beyond well-studied model organisms to those that are closely related including the development of tools for genetic manipulation and transformation, as necessary.
    - [Bottleneck]: Inability to identify specific organisms that could be beneficial for the production of specific products, metabolites, intermediates, and specific catalytic reactions.
        - [Potential Solution]: Development of rapid assays to map desired activity to potential capability.
        - [Potential Solution]: Inexpensive metabolomics analysis that lends itself to profiling of all metabolites in a cell. Current solutions that are ongoing

include developing of specific sensors (i.e., RNAs) to detect specific metabolites (though this is low-throughput).

- **5 years: Develop robust, high-throughput screens for rapidly assaying useful properties in libraries of organisms.**
    - [Bottleneck]: Availability of high-throughput screens for functions/products that cannot be screened by color or that cannot be selected using growth/viability assays.
        - [Potential Solution]: Incorporate protein or RNA sensors into cells for a particular small molecule that the cell may be engineered to produce or consume.
- **5 years: Use output of high-throughput screens/sensors and computer control to amplify a signal or expand a cell line that produces a product of interest.**
    - [Bottleneck]: There are few (or no) systems for simultaneously measuring the output of a sensor and then using a computer to expand production of the desired product.
        - [Potential Solution]: Develop biosensors to detect one or more particular desired outputs and gene expression systems that allow for computer control (e.g., light, inducer-repressor).

[Breakthrough Capability 2]: Routine domestication of non-model organisms through DNA delivery and genetic modification. (For related reading, please see *Gene editing, Synthesis, and Assembly.)*

- **2 years: Catalog and assay current methodologies and tools for carrying out DNA delivery in microbial/mammalian systems (e.g., viral vectors, conjugations, biochemical methods) and plant systems (e.g., *Agrobacterium*-, biolistic-, nanomaterial-based methods).**
    - [Bottleneck]: These methods have not been systematically compared among organisms resulting in a lack of clarity as to when one approach may be superior, or even viable, compared to another.
        - [Potential Solution]: Better aggregation of the information available for these different organisms, especially if a standard set of similar broad host vectors can be used (as a control) to standardize these.
        - [Potential Solution]: Develop bacteriophages that can be useful for engineering large number of organisms.
        - [Potential solution]: Understand the biological basis for what enables some species (e.g., plants) to be more amenable to transformation and genetic modification.
- **2 years: Develop high-throughput methods that can be done in parallel for DNA delivery (using standard methods) into non-model hosts.**
    - [Bottleneck]: Currently a limitation in the cloning process.
        - [Potential Solution]: Methods exist, but they could be improved and more widespread.

- **2 years: Establish a suite of gene-editing tools for the rapid insertion and/or deletion of genetic elements in diverse primary mammalian cells.**
    - [Bottleneck]: Genome-editing tools, particularly CRISPR/Cas technology, have enabled efficient genetic modification of a variety of immortalized cell lines, but primary mammalian cells are often more difficult to engineer with high efficiency and at scale.
        - [Potential Solution]: Develop non-toxic gene-delivery methods (viral or non-viral) utilizing reagents and equipment that are compatible with clinical manufacturing and/or high-volume cell modification.
- **2 years: Characterize basic DNA parts for expression strength in non-model organisms, specifically a larger library of plants.**
    - [Bottleneck]: Basic characterized promoters with characterized expression strengths has not yet been carried out in a systematic manner. Previous efforts have been piecemeal and the transferability of parts (e.g., promoters) between different plant species has not been well explored.
        - [Potential Solution]: A large-scale, community driven project to standardize and characterize parts will dramatically advance the state of plant engineering.
- **5 years: Development of well-characterized and robust insertion sites in plant genomes.**
    - [Bottleneck]: The majority of plant engineering efforts rely on random insertion of transgenes into the genome, resulting in the necessity to screen and characterize transformants -- a very laborious process when working with plants.
        - [Potential Solution]: Develop CRISPR-based genome-editing tools that reliably get targeted insertions with high efficiency.
- **5 years: Develop high-throughput, genome-wide editing tools for non-model organisms.**
    - [Bottleneck]: Gene editing tools are not always specific or work at all in certain organisms.
        - [Potential Solution]: Screen/develop new CRISPR-based genome editing proteins.
- **5 years: Establish robust temporal and/or spatial control of gene expression in mammalian cells.**
    - [Bottleneck]: Relatively few transcriptional, post-transcriptional, or translational regulatory devices exist for robust gene-expression control in mammalian cells; the vast majority of systems make use of inducible promoters that are often leaky or require precise tuning, which cannot be done in many practical applications such as patient-derived cells.
        - [Potential Solution]: Develop and catalogue a suite of core promoters and response elements, RNA-based regulatory devices, protein-degradation tags, among others, and record standardized, quantitative information on their performance (e.g., basal expression level with enzymatic vs. fluorescent reporters, fold induction, degradation rate, etc.).

- **5 years: Develop broad-host-range vectors for a variety of model and non-model organisms.**
    - [Bottleneck]: Lack of broad-host-range vectors that function for many different organisms requires the development of vectors specific for each organism.
        - [Potential Solution]: Engineer new host vectors targeting broader range of organisms; information from NCBI (e.g., plasmids sequences) could potentially be used to get guidance as to functional capabilities/parts that can be effective in different vector design for different organisms.
- **10 years: Develop high-throughput, targeted editing and rapid genome-evolution tools that couple genetic changes to phenotypic changes.**
    - [Bottleneck]: It is difficult to evolve non-model (and frankly, model) organisms to a desired phenotype.
        - [Potential Solution]: Develop general biosensors for particular desired phenotypes for use in non-model organisms.
        - [Potential Solution]: Develop mutator proteins coupled to sensors to evolve non-model organisms until they achieve some desired phenotype.
- **10 years: Develop universal approaches to transforming any plant.**
    - [Bottleneck]: Current plant transformation techniques work on a limited subset of plants.
        - [Potential Solution]: Better understand the biological basis for barriers in transformation in plants.
        - [Potential Solution]: Develop novel approaches that are species or host agnostic in incorporating and delivering DNA.
        - [Potential Solution]: Develop methods that are tissue-culture-independent.
- **20 years: Routine genetic manipulation of any non-model host in less than one week from first isolation.**
    - [Bottleneck]: It can take years to make a non-model into a host for heterologous gene expression.
        - [Potential Solution]: Use the tools developed to achieve previous milestones to address this bottleneck.

[Breakthrough Capability 3]: Ability to build and control small molecule biosynthesis inside cells by design or through evolution.

- **2 years: Identify model organisms for performing specific types of chemistries or organisms that have native precursor biosynthesis pathways for specific classes of molecules.**
    - [Bottleneck]: Academics and companies have constructed heterologous hosts for a limited number of chemical classes, but there are many other chemical classes that need hosts with precursor pathways constructed; additionally, some of these precursor production hosts may not be ideal for particular applications or environments.
        - [Potential Solution]: Use bioinformatics and screening to identify organisms that might be particularly useful for producing a chemical under a particular environmental condition or that already has precursor

pathways for a particular class of chemicals; build a database of those organisms.
        - [Potential Solution]: Development of robust plant hosts in order to bridge the gap before trying to engineer plant metabolic pathways into microbes.
- **2 years: Precise temporal control of gene expression for well-studied systems.**
    - [Bottleneck]: Many desired chemicals are toxic to growth of the producing host. Avoidance of this results in separation of growth and production phases which has been achieved by changing the media conditions (Clark & Blanch, 1997); however, such approaches are costly, may require the introduction of extra chemicals that are difficult to remove from the desired products, and cannot easily accommodate cellular heterogeneity, or be used to fine tune the shift from growth to production phases.
        - [Potential Solution]: Develop several types of expression systems for controlling the timing of gene expression and thus the timing of chemical production; ultimately, decouple growth and reproduction from energy and carbon metabolism and product generation (Venayak, von Kamp, Klamt, & Mahadevan, 2018).
    - [Bottleneck]: The necessary genetic regulatory elements should be sufficiently orthogonal to the host to permit tuning, but nonetheless responsive to changes in cell state.
        - [Potential Solution]: Identifying, and then re-engineering, diverse sets of protein and RNA regulators capable of activating or repressing gene expression in response to targeted changes in growth and nutrient state would expand these capabilities (Hsiao, Cheng, & Murray, 2016; Weinhandl, Winkler, Glieder, & Camattari, 2014); integrating the responses of these regulatory elements, perhaps through the use of engineered information processing networks, would permit the construction of circuitry for fine-tuning the timing of gene expression in microbial hosts.
- **5 years: Construct a limited number of model host organisms for synthesizing all-natural products.**
    - [Bottleneck]: There is a need for hosts capable of producing all natural products found in nature; there is no need for a single host to produce all of the natural products, but rather at least one host for every class of natural products.
        - [Potential Solution]: Use the database of possible hosts, genetic control systems and modular pathways to develop a range of hosts for each class of natural products; these hosts should produce high levels of the precursors to natural products.
- **5 years: Construction of single-cell organisms for production of unnatural derivatives of natural products.**
    - [Bottleneck]: There are few engineered cells that produce unnatural derivatives of natural products and the methods for engineering organisms to produce unnatural natural products are nascent.

        - [Potential Solution]: Construct hosts with the modular natural precursor pathways and engineered or evolved enzymes to create hosts that can produce unnatural natural products.
        - [Potential Solution]: Construct hosts with promiscuous enzymes (either natural, engineered, or evolved enzymers) and feed unnatural precursors that are incorporated into the final product.
- **5 years: Temporal control over multiplexed regulation of many genes in parallel.**
    - [Bottleneck]: It remains difficult to target gene expression to a particular growth phase or culture time in order to maximize production of a desired natural or unnatural product.
        - [Potential Solution]: Develop a catalog of promoters (including timing and gene expression strength) that are activated in various phases of cultivation.
        - [Potential Solution]: Identify promoters that are activated at the same time or with the same environmental queue and that can be used with multiple genes (i.e., multi-step biosynthetic pathway).
- **10 years: Software and hardware for optimizing titer, rate, and yield of any product produced by any host.**
    - [Bottleneck]: It is challenging and time-consuming to optimize the titer, rate and yield of a desired product in any host.
        - [Potential Solution]: Develop a software suite that considers the metabolic pathway(s) and host and incorporate a wide variety of measurements and can predict that changes that need to be made to the metabolic pathway add the host to optimize titer, rate, and yield.
- **20 years: On-demand construction of single cell organisms for production of nearly any molecule of interest, including organic chemicals and polymers.**
    - [Bottleneck]: Reliance on nature's single-cell organisms that are not ideally suited for producing a particular chemical.
        - [Potential Solution]: Use retrobiosynthesis software as well as genetic control system design software to *de novo* design microbial cells that can produce nearly any organic chemical; the cells would be designed for the processing environment to enable inexpensive purification of the final product.

[Breakthrough Capability 4]: Spatial control over, or organization of, metabolic pathways in cells and construction of unnatural organelles.

- **2 years: Tools to target heterologous proteins to various subcellular compartments.**
    - [Bottleneck]: There is a need for modular targeting tags that can be appended to any peptide sequence as N- or C-terminal fusions with efficient targeting to a desired compartment. The challenges are finding sequences that provide modularity; in other words, the ability to perform targeting on any protein sequence of interest, and compartmentalize a high percentage of total expressed protein, preferably at high expression levels.

        - [Potential Solution]: Most organelles have identified targeting sequences, but few satisfy the dual requirements of high efficiency and modularity; high-throughput screens for peptide sequences can be conducted using enzyme sequestration assays (DeLoache, Russ, & Dueber, 2016).
        - [Potential solution]: Bacterial organelles, known as microcompartments, have several known targeting tags for cargo enzymes and encapsulation is tunable with expression levels (Jakobson et al., 2016); however, distinct tags may use the same mechanism for loading, making it difficult to reliably control targeting of multiple enzymes into the same compartment. Additional biophysical characterization of existing and putative tags can be carried out using existing technologies to identify those with unique loading mechanisms or interaction partners.
- **5 years: Inducible synthesis of organelles.** Synthetic organelles (or compartments) that can be made on-demand and/or provide the means to compartmentalize different heterologous protein cargos would allow the engineer to mimic tissue compartmentalization in a single-cell microbe.
    - [Bottleneck]: Achieving sufficient understanding of the formation of organelles and microcompartments.
        - [Potential Solution]: Placing necessary organelle biogenesis factors under the expression control of a heterologous, inducible promoter should provide on-demand organelle production.
        - [Potential Solution]: Inducible control of bacterial microcompartment-forming proteins leads to some well-formed particles, but other factors must be identified to make sufficient quantities and do so robustly. These factors may be environmental factors or additional chaperone proteins, and can be identified by using a high-throughput screen, such as a flow cytometry-based assay that links cellular fluorescence to successful compartment formation (Kim & Tullman-Ercek, 2014).
- **5 years: Gain-control for selective permeability in and out of the organelle.**
    - [Bottleneck]: Many organelles and microcompartments will naturally have permeability to molecules that are desired to be partitioned either outside or inside the organelle (e.g., pathway intermediates); these organelles and the microcompartments will need to be modified to have lower permeability to these molecules.
        - [Potential Solution]: Selective transport of desired metabolites (e.g., pathway substrate, cofactors) must be engineered into the organelle/microcompartment by manipulating the proteins that govern the movement of small molecules across the membrane/shell.
- **10 years: Methods and tools to reprogram transport of metabolites and compartmentalization of biochemical reactions.**
    - [Bottleneck]: Prokaryotes and eukaryotes contain multiple compartments that can be utilized for sequestering sensitive chemistries or metabolites; unfortunately, it is not always clear which compartments to use.

        - [Potential Solution]: Catalogue organelles/microcompartments with favorable properties for desired applications/products (a suite of organelles will likely be of value to meet the needs of varied applications).
    - [Bottleneck]: Tools for engineering organelles and microcompartments are not readily available.
        - [Potential Solution]: Develop gene expression tools for a variety of different organelles in eukaryotes or for microcompartments in prokaryotes.
        - [Potential Solution]: Engineer multiple pathways to function in concert within the same organelle/microcompartment for optimal performance (e.g., cofactor balancing, co-substrate production, condensation reactions).
    - [Bottleneck]: Identifying parameters to optimize and means of optimization of microcompartment morphology.
        - [Potential Solution]: Genetically control the morphology of organelle/microcompartment (e.g., size, number) to achieve increased capacity for product cargo. Low capacity for enzymes is a limitation for many organelles; a solution would be the genetic manipulation of, for example, the shell protein expression levels or an organelle biogenesis pathway, to increase this capacity.
- **10 years: Alter chemical conditions within the organelle/microcompartment.**
    - [Bottleneck]: Altering the chemical environment of a compartment's lumen would require successful completion of efficient import of heterologous protein, reduction of small molecule permeability, and, for organelles, the functional trafficking of membrane proteins to the organelle's membrane.
        - [Potential Solution]: Combination of heterologous small molecule transport, enzyme, and compartmentalization of both enzymes and small molecule substrate, intermediates, and products will allow alteration of the chemical environment of the lumen.
- **10 years: Multiple orthogonal organelles/microcompartments in the same cell for compartmentalizing different parts of a pathway.**
    - [Bottleneck]: Achieving compartmentalization of different cargo in distinct organelles or microcompartments within the same cell demands the ability to generate such compartments with distinct "addresses" or protein import machinery recognizing targeting sequences orthogonal to the other native and synthetic organelles in the cell.
        - [Potential Solution]: Multiple natural organelles using different protein importomers can be utilized; however, the lumenal conditions are likely not optimal for the needs of each compartmentalized activities. Alternatively, engineered organelles with the ability to import orthogonal, distinct pools of protein in separate organelles should be achievable by either performing directed evolution on the native protein importomers or targeting heterologous importomers native to other organelles.

        - [Potential Solution]: Bacteria employ multiple distinct microcompartments within the same cell, and these are encoded by distinct operons which includes genes to make the protein shell boundary; it is feasible, but yet to be demonstrated, that these distinct microcompartments could be engineered simultaneously using existing techniques.

[Breakthrough Capability 5]: Production and secretion of any protein with the desired glycosylation or other post-translational modifications. (National Research Council (US) Committee on Assessing the Importance and Impact of Glycomics and Glycosciences, 2012)

- **2 years: One or more microbial hosts capable of producing laboratory-scale quantities of a single glycoform of a desired protein.**
    - [Bottleneck]: Compared to protein synthesis, post-translational modification and machinery is not well understood and any given protein may exist in the cell in multiple glycoforms, making advanced study challenging; this problem is particularly prevalent in the production of biologic pharmaceuticals (e.g., biosimilars) where the complex array of glycosylation patterns can cause differences in efficacy and safety (Sethuraman & Stadheim, 2006).
        - [Potential Solution]: Understand the pathways involved in glycosylation of the protein of interest and develop these pathways to enable tightly controlled synthesis of a single glycoform.
- **5 years: A few microbial hosts capable of secreting functional versions of proteins with no post-translational modifications.**
    - [Bottleneck]: There are relatively few bacterial hosts that have well developed protein secretion systems. Those that exist are primarily in Gram positive hosts, and are not necessarily compatible with all desired protein products; gram negative hosts, on the other hand, have an outer membrane that must also be crossed before exiting the cell, and secreting proteins across both the inner and outer membranes remains a challenge at the efficiencies required for commercial application.
        - [Potential Solution]: Develop new bacterial hosts with machinery specifically designed for protein secretion.
    - [Bottleneck]: The systems that do exist are either required for cellular function and thus can only be repurposed to a limited extent, or are highly regulated and difficult to activate or keep activated.
        - [Potential Solution]: Utilize systems biology to understand and remove the regulatory controls on secretion systems.
- **20 years: Ubiquitous control of post-translational modification (including glycosylation of multiple sites with multiple sugars) in a diverse array of hosts.**
    - [Bottleneck]: Post-translational modifications are often considered the “analog” control in the cell; as such, they can carefully tune how an individual protein interacts with its environment. The lack of understanding of this control makes affecting post-translational modification for desired use particularly challenging. The ability to tune the *analog* portion of the cell, however, can allow for more

efficient systems, better synthetic yields, and a broader set of uses than would otherwise be possible.

Goal 3: On-demand fabrication and modification of multicellular organisms.

[Current State-of-the-Art]: Currently for multicellular systems and organisms, our technology is closely aligned with natural reproduction: we edit gametes or embryos, and rely on natural processes to differentiate genetically-identical cells into tissues. Gene editing methodology allows substantial improvements and inclusion of novel biochemical and molecular changes. Today our engineering abilities in plants are limited to stable integration of small genetic circuits (fewer than 200 kb). Examples of engineered modification of animals include genome editing of chicken embryos to produce virus resistance (Looi et al., 2018; Sid & Schusser, 2018), and the inactivation of porcine endogenous retroviruses in pigs for human organs transplants (Niu et al., 2017; Ross & Coates, 2018), but significant work is necessary before we are able to selectively edit and modify multicellular eukaryotes with confidence and consistency.

Generally, we need a better understanding of cell-to-cell interactions and to establish stable modifications within multicellular systems. Current state-of-the-art in multicellular engineering includes tools and technologies for some plants (Farré et al., 2014) and fungi (such as *Aspergillus* (Lubertozzi & Keasling, 2009)), but most advances toward this goal have occurred in the engineering of a single cell type within a multicellular organism (for example, the introduction of the Polled trait into dairy cattle breeds (Van Enennaam, 2018)) and germline engineering. An emerging technology in multicellular system engineering is cell-scaffolding (loading specialized cells onto engineered matrices) and enabled control over the three-dimensional shape and structure of a system. Advancements in engineering for tissue- and organ-on-a-chip technologies are also helping to bring about advancements in this area.

For related reading, please see *Biomolecule, Pathway, and Circuit Engineering*, ***Goal 3: Holistic, integrated design of multi-part genetic systems (i.e., circuits and pathways)***.

[Breakthrough Capability 1]: Ability to control differentiation and de-differentiation of cells within a population.

- **2 years: On-demand, reproducible functionalization of simple micro-tissues or micro-consortia made up of two or more engineered cell types.**
    - [Bottleneck]: Variations in cell culture quality between trials and institutions.
        - [Potential Solution]: Investigate incubation conditions to identify and mitigate environmental sources of variability in cell behavior and growth.
        - [Potential Solution]: Identify informative biomarkers.
- **5 years: Programmable and regulatable pathways that can be induced to differentiate or de-differentiate somatic cells.**
    - [Bottleneck]: Gaps in understanding which genes and networks can be altered to control cell behavior.
        - [Potential Solution]: Experimental tools to co-regulate any desired set of multiple gene targets, such as via engineered transcription factors.

[Breakthrough Capability 2]: Ability to characterize and control the three-dimensional (3D) architecture of multicellular systems.

- **2 years: Characterize existing tissue components and standardize measurements to evaluate function.**
    - [Bottleneck]: Inconsistent reporting of matrix and cell performance resulting from differences in composition and tissue geometry.
        - [Potential Solution]: Generate a library of known cell types, matrices, and exogenous signalling molecules and characterize all combinations under identical conditions and geometries; measurements should include characteristics such as stress-strain response, degradation rates in serum, and immunogenicity.
    - [Bottleneck]: Discovery of novel exogenous signalling molecules to regulate cell and tissue behavior.
        - [Potential Solution]: Evaluate specific factors (via high-throughput chemical screening) that can be supplemented to existing 3D matrices/scaffolds to induce drastic changes in cellular behavior (such as morphology, differentiation, tissue composition, matrix alignment).
- **5 years: Identification of novel 3D scaffold designs that can lead to desirable cellular properties.**
    - [Bottleneck]: Limited capacity for nutrient delivery (~100μm by diffusion alone).
        - [Potential Solution]: Develop synthetic microvascular networks, either by self-assembly of endothelial cells and pericytes or 3D patterning of tissues; these networks must be able to support cell growth within the scaffold and be robust to changes in tissue composition (including the introduction of additional cell types, mechanical forces, or chemical factors).
    - [Bottleneck]: Heterogeneous cell seeding within a large scaffold.
        - [Potential Solution]: Enable and advance formation and production of extracellular matrix and methods to improve seeding (for example, forced flow of cells into tissue).
- **10 years: Create modular, synthetic communication circuits that can be implemented in tissues to allow for control of new or existing cellular communication systems.**
    - [Bottleneck]: Cell heterogeneity within tissue and ability to target only the cell type(s) of interest.
        - [Potential Solution]: Link integration or expression of a circuit to expression of a synthetic molecule in the target cell type.
    - [Bottleneck]: Robustness against perturbations, including the various signalling molecules expressed in tissue.
        - [Potential Solution]: Build in redundant control to genetic circuits, and leverage advances in biomolecular design to use components with large induction ratios and minimal cross-talk with other circuit components.

- **20 years: Bottom-up design and construction of whole organs at the centimeter-length scale.**
    - [Bottleneck]: Nutrient delivery in organs.
        - [Potential Solution]: Assemble synthetic arterioles and venuoles to interface with capillaries, then couple these synthetic vascular networks with cell culture media perfusion strategies.
    - [Bottleneck]: Large-scale assembly of substructures into complete tissues.
        - [Potential Solution]:
    - [Bottleneck]: Understanding the principles of organ *design*; most efforts to date focus on recapitulation of existing organs or a reduced set of functions performed by a given organ.
        - [Potential Solution]: Prototyping synthetic organs to substitute for, complement, or enhance native organ function in a manner beyond recapitulation of evolved biology.

[Breakthrough Capability 3]: Ability to achieve stable non-heritable changes in somatic cells.

- **2 years: Routine delivery of biomolecule "effectors" (i.e., DNA, RNA, proteins) into slowly-dividing or non-dividing cells.**
    - [Bottleneck]: Lack of technologies for homogenous delivery of macromolecules into tissues.
        - [Potential Solution]: Further development of cell-penetrating nanoparticles and exosomes.
- **5 years: Generation of effective artificial epigenetic chromosomal states and maturation of the emerging field of chromatin engineering.**
    - [Bottleneck]: Incomplete functional characterization of natural chromatin.
        - [Potential Solution]: Engineered platforms to rapidly interrogate hundreds of structural, enzymatic, and synthetic chromatin proteins.
    - [Bottleneck]: Uncertain causal relationship between genome/epigenome and cell behavior.
        - [Potential Solution]: Coordination of statistical genome/ epigenome association studies (GWAS/ EWAS) with experimental reconstruction of states to test and validate associations.
- **10 years: Ability to generate cell states that are stable and effective after the inducer/effector is removed in certain model tissues.**
    - [Bottleneck]: Gaps in understanding cell "homeostasis" and how biochemical processes inside cells are interconnected and reinforce each other.
        - [Potential Solution]: Development of experimentally-supported predictive (systems biology) models to predict the long-term impact of an artificial perturbation.
    - [Bottleneck]: Gene-editing dependency; too much focus on transcriptional regulation in the nucleus.
        - [Potential Solution]: New tools to control self-perpetuating "post-translational" states (such as RNA and protein modification).

        - [Potential Solution]: Advances in organelle engineering (especially for mitochondria).
- **20 years: Nimble adaptation of somatic cell engineering technologies to any natural tissue at any developmental stage.**
    - [Bottleneck]: Gaps in understanding cell-type and tissue-type-specific barriers that enable cells to resist conversions.
        - [Potential Solution]: Further advancement of systems biology methods to quickly identify appropriate target genes, proteins, and molecular networks.

[Breakthrough Capability 4]: Ability to make predictable and precise, targeted, heritable changes through germline editing.

- **2 years: Complete sequence of select host genomes to allow design of targets for gene editing.**
    - [Bottleneck]: Genetic variation.
        - [Potential Solution]: Sequencing of specific transformation target lines.
- **2 years: Define and validate tissue-specific DNA parts in plants.**
    - [Bottleneck]: There has been a dearth of plant DNA parts (e.g., promoters) that have been systematically characterized. Although many genes have been described from transcriptome datasets as tissue-specific, the validation and characterization of their tissue specificity will be required for future plant synthetic biology efforts.
        - [Potential Solution]: Systematic characterization of various tissue-specific promoters in various plant species.
- **5 years: Efficient germline transformation systems developed in targeted hosts.**
    - [Bottleneck]: Transformation systems are limited and optimization is slow; efficiency is such that the molecular analysis burden is high.
        - [Potential Solution]: Increase transformation efficiency through new vector design components that will stimulate cell division during the time the DNA is introduced into the cell OR enable improvement in high-throughput molecular analysis platforms to screen for those with the correct edits.
- **5 years: Ability to deliver transgene constructs to most (>90%) somatic cells in a higher eukaryote organism to rapidly prototype transgenic phenotypes.**
    - [Bottleneck]: Higher eukaryotes have relatively long timescales of organism development, making phenotype development too slow for effective research.
        - [Potential Solution]: Improve somatic cell nuclear transfer to embryos to speed multi-locus genome engineering for non-model organisms with long generation times.
    - [Bottleneck]: Existing gene delivery technologies reach only a subset of cells in an intact organism.
        - [Potential Solution]: Enhance gene delivery technologies to approach organism-scale delivery.

- **5 years: Temporally controlled transgene expression that works on the scale of generations.** For example, kill switches that are activated only after a defined number of generations.
    - [Bottleneck]: Robust molecular time-keeping methods.
        - [Potential Solution]: Design and implementation of robust synthetic cell cycle oscillators and other molecular timers.
    - [Bottleneck]: Gene expression platforms that confer stable expression across multiple cell divisions (such as in primary cells).
        - [Potential Solution]: Development of stable, controllable, heritable extra-genomic expression platforms, including artificial chromosomes.
    - [Bottleneck]: Spontaneous silencing of transgene constructs being expressed over long periods of time.
        - [Potential Solution]: Synthetic epigenetic mechanisms that interfere with or block natural silencing mechanisms.
- **5 years: Efficient gene editing in differentiated cells.**
    - [Bottleneck]: DNA folding is a physical blockade against gene-editing enzymes.
        - [Potential Solution]: Engineering the editing enzymes and/or helper proteins to unfold DNA.
    - [Bottleneck]: Bacterial CRISPR particles induce an immunogenic response.
        - [Potential Solution]: Discovery of non-immunogenic variants.
        - [Potential Solution]: Development of “coating” particles or chemical tags.
    - [Bottleneck]: Some cell types carry heterogeneous, naturally-modified genomes (such as immune cells).
        - [Potential Solution]: Delivery of gene expression cassettes that are not integrated into the chromosome.
- **5 years: Ability to domesticate engineered biological parts to confer immune tolerance in immunocompetent organisms.**
    - [Bottleneck]: Introduction of foreign proteins can induce immune rejection in immunocompetent organisms; the rules governing how such rejection is elicited by synthetic biology parts and how it may be circumvented are not yet clear.
        - [Potential Solution]: Generation toolboxes of “stealthed” parts that are unlikely to elicit immune rejection.
        - [Potential Solution]: Development of strategies for inducting active immune tolerance of synthetic biology parts.
- **10 years: Ability to coordinate engineered multicellular functions in intact organisms via orthogonal communication systems.**
    - [Bottleneck]: Generating synthetic analogs of coordinated processes, such as wound healing or immune protection, will likely require communication between engineered cells; co-opting native cell-cell signaling mechanisms is likely to exhibit cross-talk and cross-regulation with native systems.
        - [Potential Solution]: Generation of libraries of mutually orthogonal synthetic signaling molecules and receptors that can confer coordination across various length scales within an organism.

- **10 years: On-demand gene editing of organisms with desired traits.**
    - [Bottleneck]: Gene editing efficiency is low for multiple edits; limitations in what sequences can be edited due to CRISPR target-recognition constraints.
        - [Potential Solution]: Develop new CRISPR or other engineered enzymes that have expanded recognition sequences and efficiencies.
- **20 years: Routine, on-demand, efficient germline editing for any targeted hosts of interest at high-throughput scale.**
    - [Bottleneck]: Different and diverse transformation systems needed across species.
        - [Potential Solution]: Develop a process that is automated from preparation of embryos to transformation to selection/identification of successfully-edited embryos.

Goal 4: Generation of biomes and consortia with desired functions and ecologies.
Note: We make the distinction between biomes and multicellular organisms (see ***Goal 3: On-demand fabrication and modification of multicellular organisms***) through the definition of a biome as containing organisms (including multicellular organisms) with different genomes.

[Current State-of-the-Art]: While a few microbiome systems are well characterized, such as rhizobium for nitrogen fixation, we are still struggling to understand and how and why consortia of microbes cooperate in nature. Systems with mutual metabolic dependencies (synthetic heterotrophs) have enabled the construction of engineered consortia that are stable in laboratory settings (Pacheco, Moel, & Segrè, 2019). Our ability to produce synthetic interactions is possible with some ongoing efforts; for example, a small number of synthetic microbial consortia have been created as model systems, consisting of 2-3 different organisms (Kong, Meldgin, Collins, & Lu, 2018; McCarty & Ledesma-Amaro, 2019). Bioremediation and waste-water treatment demonstrate the principles that consortia can be used industrially, while probiotics and fecal microbe transplants demonstrate the principle that the composition of gut flora can be manipulated. Industrial startups in this space are emerging at a rapid pace, but our ability to make targeted changes, such as adding or removing a single organism, in an existing microbiome are very limited. Overall, our ability to understand and manipulate systems with specific functions or to remediate biomes and consortia that cease to function as desired is very limited.

[Breakthrough Capability 1]: Ability to control cell-to-cell communication between different species.

- **2 years: Tightly-controlled promoter-response regulator systems that enable intra- and inter-species cellular communication.**
    - [Bottleneck]: Limited technologies for the exchange of biochemical information within a population of cells.
        - [Potential Solution]: Use available cell-cell communication regulators to enable cell-cell communication in broader range of organisms.
        - [Potential Solution]: Expand communication systems by using a broader range of natural quorum-sensing/communication modules (such as acyl-

homoserine lactones, autoinducer-2, peptide-based signaling, and metabolic signaling).

- **5 years: Synthetic cell-to-cell communication elements and networks that function in a broad range of host organisms.**
    - [Bottleneck]: Current synthetic communication systems have been engineered to function in a limited set of organisms.
        - [Potential Solution]: Identify cell-cell communication elements in non-traditional hosts, characterize, and modify for use in synthetic circuits.
        - [Potential Solution]: Engineered membranes (specifically, receptors) to transmit information via cell-cell contact.
        - [Potential Solution]: Engineered transmission of secreted biomolecules and exosomes.
- **10 years: Signal-response pathways that function in synthetic communities of 5-10 organisms, employing a variety of pathway types and host species.**
    - [Bottleneck]: Cross-talk between communication elements.
        - [Potential Solution]: Combine communication modules in a manner that minimizes cross-talk; employ metabolic signaling to coordinate behavior across population as needed.
    - [Bottleneck]: Feasibility of enabling signaling between all community members.
        - [Potential Solution]: Design networks with essential community-level coordination using a limited set of communication modules.
- **20 years: Ability to produce engineered microorganisms that can reliably invade and coexist within a complex community and manipulate the consortium/biome function and behavior.**
    - [Bottleneck]: Ecological understanding of complex microbiomes, rules of coexistence, cooperation, and competition.
        - [Potential Solution]: Characterization of broad ranging natural, as well as synthetic, communities during environmental and ecological changes.
        - [Potential Solution]: Employ cell-cell communication or metabolic interactions to enable engineered cells to be accepted into/required by the target community.

[Breakthrough Capability 2]: Ability to characterize, manipulate, and program the three-dimensional (3D) architecture of a biome (i.e., the "ecosystem" of a natural or manipulated biome containing multiple species).

- **2 years: Use of existing technologies (including metagenomics, transcriptomics, proteomics, and mass spectrometry) to better understand the species composition and collective components of microbial communities and consortia.**
    - [Bottleneck]: Need data from diverse ecosystems, environments to build predictive models.
        - [Potential Solution]: Generate and include data from many environments so that models can integrate information with respect to the community, what they are doing, and the environment they are inhabiting.

- **5 years: Non-destructive, 3D visualization of microbial communities from a broad range of environments.**
    - [Bottleneck]: New technologies needed to report and visualize 3D structures and functions of consortia.
        - [Potential Solution]: Adapt imaging, sequencing, -omics technologies to characterize natural or engineered systems and their dynamics.
        - [Potential Solution]: Develop new reporter systems, such as cell-based sensors, to assess and quantify function and/or 3D organization.
- **10 years: Ability to manipulate the 3D architecture of natural or engineered communities using external inputs (such as molecules, temperature, or pH).**
    - [Bottleneck]: Limited understanding of how communities respond dynamically to environmental changes, especially in non-homogeneous systems.
        - [Potential Solution]: Characterization of how natural and engineered communities respond to environmental changes; build spatio-temporal models that incorporate genomic, functional, and environmental outcomes.
    - [Bottleneck]: Ability to add sensing and actuation capabilities to any cell type in a community setting.
        - [Potential Solution]: Targeted gene editing approaches that deliver only to a specific organismal-member of a community.
- **20 years: Programmed communities that self-assemble into a desired 3D architecture.**
    - [Bottleneck]: Tools that enable desired stratification and self-organization (or reorganization) of microbial communities.
        - [Potential Solution]: Use strategies from multi-organismal cell development, such as the generation and sensing of gradients and motility.
        - [Potential Solution]: Specific binding between cells using extracellularly-displayed proteins to build or lock in specific levels of organization.

[Breakthrough Capability 3]: Ability to control and/or define the function of an engineered microbial community/biome.

- **2 years: Ability to combine species with specialized functions to enable the production of desired products.**
    - [Bottleneck]: Growth rates and conditions ideal for production may vary between species.
        - [Potential Solution]: Apply synthetic ecological approaches, including identifying optimal growth conditions for the community, such as engineering mutualistic interactions to control community composition.
- **5 years: Assembly of consortia to produce desired molecules/products, considering community-level metabolic flux.**
    - [Bottleneck]: Ideal division of labor within the consortium is difficult to predict.

        - [Potential Solution]: Develop metabolic engineering approaches to separate processes (such as reducing metabolic load, balancing redox and cofactor use), ideally amongst community members.
- **10 years: Plug-and-play assembly of consortia to produce desired molecules/products from specific starting materials, considering community level metabolic flux and organism-to-organism communication.** For example, developing consortia of different microbial species that are grown/fermented together to create a desired product.
    - [Bottleneck]: Optimal growth/production conditions of the individual community members are likely to be different.
        - [Potential Solution]: Screen for, or predict, conditions that are optimal for the community; tune relative population densities through inoculation ratios and via feedback (cell-to-cell communication).
        - [Potential Solution]: Engineer community members to function optimally under target bioreactor/process conditions.
- **20 years: On-demand assembly of consortia that are programmed to respond dynamically, such that they can use different feedstocks, metabolize toxins or toxic byproducts, or produce different products in response to endogenous (system) or exogenous (user) cues.**
    - [Bottleneck]: New strategies needed for holistic engineering of consortia that can work under a broad range of conditions off the shelf.
        - [Potential Solution]: Develop reliable building blocks (organisms, communication modules, sensors, enzymes, metabolic processes) that can be recombined reliably and adapted for specific applications.

[Breakthrough Capability 4]: Targeted modification of an existing microbiome to enable new functions or address dysbiosis - at the host, community, or environment level - through the addition, removal, or reorganization of the community members.

- **2 years: Use of existing technologies (including metagenomics, transcriptomics, proteomics, and mass spectrometry) to characterize functions of microbial communities from a broad range of environments.**
    - [Bottleneck]: Need data from diverse ecosystems, environments to build predictive models.
        - [Potential Solution]: Generate and include data from many environments so that models can integrate information with respect to the community, including major functions and surrounding ecosystem.
- **5 years: Characterize how *select* microbiomes respond to changes in the environment, including the addition of toxins, the introduction new organisms (pathogens or commensals), and the selective removal of species from the community.**
    - [Bottleneck]: Ability to selectively remove species from a biome.
        - [Potential Solution]: Targeted anti-microbials.

- **10 years: Predictive models of microbiome function and response to a broad range of environmental and ecological changes.**
    - [Bottleneck]: Need to be able to undertake modeling and comparative pathway analysis to determine most robust, prioritized, and resilient systems.
        - [Potential Solution]: Controlled laboratory experiments or observational studies where microbiome function is determined as a function of community composition and environment.
        - [Potential Solution]: Machine learning approaches to determine whether there are patterns between microbiomes of interest (with respect to both form and function); follow-up with strategies designed to improve mechanistic understanding, as needed.
- **20 years: Ability to modify an existing biome or consortia as desired.**
  Biome/consortia modifications include: 1) adding functions such as the ability to sense the environment and coordinate responses for defined outcomes (pathogen defense, substrate transformation, and biosensing), and 2) manipulating composition of host-associated communities to address dysbiosis or add new functions.
    - [Bottleneck]: Need controlled invasion of non-native organism with desired properties or the ability to reintroduce a host-associated organism after engineering.
        - [Potential Solution]: *In situ* gene editing to add function to a community with minimal disruption.
    - [Bottleneck]: Ability to selectively add and remove community members, including engineered cells, and have them persist in a community for a desired length of time and at a desired population density.
        - [Potential Solution]: Kill switches (to remove organisms), auxotrophy/complementation and shared metabolisms (to retain organisms).
    - [Bottleneck]: Need an integrated understanding of microbial community ecology and function; ability to predict how adding a new member with desired functions will affect community health and stability.
        - [Potential Solution]: Focus on developing organisms that can integrate into host consortia and deliver the required functions; recent evidence suggests that function is more important that what organism is carrying it out.

Data Integration, Modeling, and Automation

(Data Science)

Data Integration, Modeling, and Automation

Summary

Data Integration, Modeling, and Automation focuses on robust, systematic use of the design, build, test, learn methodology to create complex systems. Progress requires a purpose-built computational infrastructure that supports DBTL biological processes, the ability to predict design outcomes, and optimize manufacturing processes at scale.

Introduction and Impact

Applications of engineering biology have grown beyond chemical production to include the generation of biosensor organisms for the lab, animal, and field, modification of agricultural organisms for nutrition and pest/environmental resilience, production of organisms for bioremediation, and live cell and gene/viral therapies. The rapid expansion of the field has resulted in new tools and new approaches; however, we are still challenged by the need for novel and more robust computational tools and models for engineering biology. For example, improved models of synthetic systems and of their interaction with their host organisms will facilitate more successful engineering and broader application.

The foundation of a viable design and manufacturing process for, or using, engineering biology is automation, which requires a complete description of a biological system's components, data to describe the system's function and interconnections, and computational models to predict the impact of environmental parameters on the system's behavior. For each stage and interface of the design-build-test-learn framework, we need to specify the new data and algorithms that drive experimental design, clarify the assay frameworks that allow computational diagnosis of outcomes, assure that metrology is high quality and comparable across sites, integrate frameworks that allow algorithmic prediction of process and performance improvements, and build interfaces to drive both automated and human-in-the-loop design improvements.

This information infrastructure for biological design is in a nascent state compared to engineering disciplines such as mechanical and electrical engineering, due to the recent emergence of the biomanufacturing field. A critical bottleneck is a lack of established "design rules," core aspects of biological and biomolecular function that apply to diverse systems and applications. Furthermore, technologies for the utilization, manufacture, and deployment of biological systems are still under development. These roadblocks have hampered the development of standard computational frameworks to represent and store information about biological components, predict system behavior, and diagnose failures. Therefore, widespread automation remains out of reach.

Data Integration, Modeling, and Automation proposes a roadmap towards efficiently scaling engineering biology applications from the design, build, test, and learn cycle to the efficient and reproducible creation of individual biological components, to intracellular systems, multicellular systems, and their operation in diverse environments. This includes access to a standard information and modeling ecology to support biological design, manufacture, and quality control/diagnosis parallel to those that exist in chemical and other engineering disciplines, but which respect the core differences inherent in the biological substrate; standard and accessible frameworks that support the effective development and use of information on biological system

and component function that are a necessary foundation for widespread biological design; models and tools for simulating the behavior of biological components and their interconnected systems in their diverse deployment environments that are necessary to support predictive design of these systems and diagnosis of their failures; and manufacturing process design and optimization tools with similarly attached information systems that are needed to ensure cost- and time- effective and scalable production of designed systems with minimal errors. All these systems should ideally be connected through findable, accessible, interoperable, and reusable (FAIR) data and process modeling efforts so that the community can benefit from their combined experience and work-products. Together, the protocols, metrology, and computational elements of the design-build-test-and-learn process can be continually improved.

Transformative Tools and Technologies

Integrated biological designs and data models

The foundation for design is knowledge of the components with which a design can be built and the environmental constraints under which the designed system will operate. While data can often be sparse for biological systems, there has been significant work in representing data about biomolecular function for both basic biology and engineering, including genome organization, gene regulatory-network function, metabolic pathways, and other aspects of biological function and phenotype. However, the specializations necessary to enable effective design across scales, from submolecular to mixed communities of cells in complex environments, is lacking.

The design of proteins and nucleic acids for desired functions has been a long standing biotechnological goal. There has been great progress in computational design for gene expression control, molecularly responsive nucleic acid structures, and protein structures; however, the reliability of these tools is still relatively low and the functional classes accessible for design are limited compared to those required. The current status calls for renewed scaling efforts in biomolecular characterization so that data driven methods of design can properly expand, new data-driven design algorithms and designs-of-experiments to predict such molecules, and better physics-based biomolecular design algorithms.

While design tools for metabolic engineering and gene regulatory network engineering have improved greatly over the last decade, they are still relatively limited to a small number of model organisms, a limited set of regulator families, and relatively well-characterized metabolic pathways. Current tools also have relatively primitive methods for incorporating multi-omic and other biological data to constrain their predictions, and tools for informative designs-of-experiments are lacking. Further, only recently have models of coupling to host resources and toxicity, issues of relative fitness and evolutionary robustness, and cross-organism pathway design been considered. The operation and design of mixed communities is in a primitive state.

There are almost no standardized computational approaches to ensure that the biological systems produced are measured sufficiently to prove effective and reliable function, to diagnose failures, and to predict what parameters or components must change to make the design models better match the observations and meet design goals. Integrated biological data

models will be required to understand, predict and control the effect of engineering these systems at all levels and time scales.

Integration of -omics and machine learning for the design-build-test-learn (DBTL) cycle

Rapid advances in fields that leverage supervised machine learning have owed their success to the existence of massive amounts of annotated data. Data that will inform integrated biological data models will include measurements of circuit behavior in a cellular context, continuous measurements of transcriptome, proteome, and metabolome at the single-cell level, measurements that inform bioprocessing at scale, and measurements of the effect of engineered organisms on ecological scales.

Beyond the accumulation of data, theoretical impediments also prevent machine learning from accelerating the DBTL cycle. Suppose X is a set of multi-omics measurements, and Y is the yield of the desired bioproduct. By training on many multi-omics datasets and yields, a machine learning algorithm should be able to take a new multi-omics dataset X' and predict the corresponding yield Y'. However, the critical question in the DBTL cycle is how to use measurements made in the current design to improve the design of the next iteration. That is, measurements X' of the design are not being asked to predict the yield Y' associated with that design. They are instead being asked to predict the yield Y* of a proposed design for which no data X* yet exists. Because the current generation of machine learning methods are powerless to address counterfactuals, new machine learning algorithms are needed that incorporate causal inference to identify interventions that would yield answers to the fundamental questions that drive the DBTL cycle (Pearl, 2018).

While existing multi-omics measurements can provide many features, and collect observations on those features in a sufficiently high-throughput manner to fully exploit the DBTL, several major inter-linked challenges are data visualization, integration, mining, and modeling. Creation of design libraries to exercise design space is needed. Multi-omics aspects are useful, but they are generally operated on one design at a time. There is a challenge in library creation and scaling -omics measurements for these libraries for machine learning techniques to work. Further, ideally molecular and cellular functions have been characterized allowing the design-of-experiments to be chosen to minimize the number of manufactured variants that cover the most informative parametric space. The challenge therefore is: 1) having sufficiently characterized components for effective design of experiments; 2) having sufficient information about the cellular function and environmental factors to constrain the machine learning models; 3) having sufficient high quality measurement bandwidth for the design-of-experiment to work; and 4) using machine-learning models to select the next parameter sets to try.

While the variety of available software is enabling more standardized circuit design, there are fewer tools available for multi-omics data analyses, data interrogation, data mining and machine learning. However, such approaches have recently been validated, where combining proteomics and metabolomics data and machine learning allowed the prediction of pathway dynamics that outperformed well-established and existing methods (Costello & Martin, 2018). Furthermore, two recent groundbreaking studies identified design principles for optimizing translation in *Escherichia coli* and the principle regulatory sequences of 5'

untranslated regions in yeast using machine learning approaches and large-scale measurements (Cambray, Guimaraes, & Arkin, 2018; Cuperus et al., 2017).

BioCAD tools and design-of-experiment (DoE) approaches

In many other industries, the maturation of computer-aided design (CAD) systems have dramatically increased the productivity of the designer, improved the quality of the design, improved communications through documentation, and created shareable databases for manufacturing. To achieve the level of sophistication of design automation employed in industries such as automotive, shipbuilding, or aerospace, significant progress must be made in laying the foundation for computer-aided design for biology (BioCAD) software tools and data standards to support the DBTL cycle. For example, the Synthetic Biology Open Language or SBOL allows *in silico* DNA models for synthetic biology to be represented (Galdzicki et al., 2014). Other examples of integrated BioCAD tools are Diva BioCad and the TeselaGen BioCAD/CAM platform (Boeing, Leon, Nesbeth, Finkelstein, & Barnes, 2018), an 'aspect-oriented' BioCAD design and modelling framework, and Cello (Nielsen et al., 2016) for gene circuit design automation. Many of these software tools are also currently being integrated into biological foundry automation suites, such as the Agile Biofoundry, in order to accelerate these processes. In addition, there is an increasing use of Design-of-Experiment (DoE) approaches for determining the most efficient experimental testing and measurement strategies (such as JMP statistical software from SAS). Such tools are distinct but complement BioCAD tools.

However, with the rapid growth and uptake of liquid handling automation and medium-throughput analytics in biofoundries, there is an increasing need to establish standardized protocols and reference materials to enable reproducibility and standardized measurements. There is also a need to develop numbers and range of software tools to allow interoperability of hardware building on platforms like ANTHA, as well as common data formats for measurements that can be used for machine learning, and standardized metadata and annotations to compare designs between laboratories and companies. The increasing use of large-scale libraries and high-throughput automation (such as microfluidic platforms) will inevitably lead to a data-deluge which will pose challenges in terms of data storage, data standards, data sharing and data visualisation.

A number of frameworks have recently been developed to aid engineers in turning designs of their biomolecules, pathways, and hosts into a set of formal automatable manufacturing operations. Further, these tools optimize for reliability and correctness of synthesis and efficiency in cost and time-of-production. Some of these link directly into the biomolecular and pathway/host design tools to choose optimal "DNA" parts to meet those design goals. However, there are not yet sophisticated tools supporting manufacture of high-complexity structured libraries for design-of-experiments.

Design tools are at their most powerful when the requirements, limitations, and desired outcome of a given design problem can be flexibly and completely specified in domain-specific languages (DSLs). These languages can and should support defining metrics against which designs can be optimized. Metrics could include, but are not limited to: yield, titer, efficiency, costs, environment, and longevity, among many others. Given the multiple scales at which design software will be asked to operate (such as for individual genetic networks, whole-cell

models, cell-to-cell interactions, and up to entire ecosystems), scale-specific DSLs may be appropriate. These languages must be highly expressive but remain digitally interpretable, including support for simulation of designs against encoded requirements as a means for selection among competing design candidates. These languages may also allow for the storage of experimental results that could be formally compared to the specification to determine whether a given design satisfies the encoded requirements.

Automation of 'Build' and 'Test'

To increase throughput, capacity, and reproducibility, physical and informatic automation efforts have been applied to the Build and Test portions of the biological engineering DBTL cycle. The use of (traditional, acoustic, and microfluidic) liquid handling robotics to prepare molecular biology reactions (e.g., PCR, DNA assembly) is representative of Build physical automation. Test physical automation includes parallel arrays of bioreactors integrated with liquid-handlers for automated real-time control (e.g., pH, feeding) and periodic culture sampling (for offline analysis). Sample tracking (through laboratory information management systems - LIMS), automated protocol design/selection, and data analysis pipelines are characteristic of Build and Test informatic automation. The extent of process automation can range from semi-manual (i.e., stand-alone automated unit operations that interface through a human operator), to full automation (autonomous integrated unit operations). Semi-manual and full-automation each have advantages: with semi-manual automation, there is process flexibility and decreased operational complexity; fully-automated platforms allow high-throughput and "24/7" operations; neither process is always preferable to the other.

Sample-independent performance, unit operation de-coupling, and operational "good-enough" thresholds enable process automation. Sample-independent methods are more amenable to automation due to sample-to-sample performance robustness and the direct enablement of method scale-out/parallelization. Representative methods include sequence-independent DNA assembly methods (vs. traditional sequence-dependent cloning strategies), microbial landing-pad strategies that enable the same DNA construct-encoded gene cluster to be productively deployed across phylogeny (rather than a bespoke construct for each organism), next-generation DNA sequencing methods (vs. primer-directed Sanger sequencing), and methods for preparing a single sample for multiple -omics analyses (global or targeted metabolomics, proteomics, and/or lipidomics). Very few methods are completely sample-independent, however, and it is important to have alternative method(s) for samples that prove to be problematic for the preferred method. Since technologies (including methods, software, and instrumentation) change very quickly, and significant effort is needed to adapt an existing, or create a new, automation method, unit operation de-coupling is crucial. The automation of any step in a process should ideally be unaffected by a technological change in an upstream or downstream step, otherwise all coupled steps need to be re-developed if any one step changes. In practice, this is difficult to achieve. For example in Build, it is not yet generally possible to Design any DNA sequence for fabrication without being sensitive to the limits of technology and method of fabricating the DNA (e.g., how sequence-independent or not the DNA synthesis/assembly technology actually is). An important automation-enabling approach is to set "good-enough" thresholds. Automated unit operations often process samples in batches, and a

key operational decision or stage-gate is to determine what to do with the (anticipated minority) of samples that fail to be successfully processed. One approach is to set a threshold, and as long as that threshold of samples are successful, to proceed with the successful samples and drop the failed ones. It is, of course, possible, and in some cases desirable or necessary, to re-queue the failed samples (potentially with an alternative method), but at some point repetitively failed samples must be abandoned or they will cumulatively drive the automated workflow to a halt.

Towards the desired impact of Build and Test automation increasing efficiencies, rates, scope, reliability, and reproducibility, there remain considerable challenges and associated opportunities and needs for improvement. These challenges, for example, include that technologies change rapidly leading to process instability and the need to chronically re-develop automation - like the Red Queen telling Alice she must run to stand still. Additional challenges include: that instrumentation differences across facilities limit automation method transferability; that the use and reliance upon automation can pose an operational robustness risk if an instrument fails (and if there is no instrument redundancy); and that *a priori* it can be difficult to predict which type of method might work effectively for a specific sample. Improvements are needed to better understand how transferable automated methods are across facilities and instruments, how to develop methods that are more suitable and robust to automation (i.e., less sample dependent), to further de-couple unit operations, and to further application of automation approaches, for example, to the Build of transcription/translation systems, biomes, and tissues.

Future requirements of engineering biology databases

A mature computational infrastructure for biodesign requires powerful access to information about biological parts and systems, their environments, their manufacturing processes, and their operations in and beyond the laboratory in which they are created. This in turn requires findable, accessible, interoperable, and reusable data that enable effective aggregation information on biological systems, their environments, and their processes of manufacture, and the establishment of standard models of data processing and analysis that allow open-development and scalable execution.

One of the key enablers of any data-intensive field is the production of computational frameworks capable of supporting findable, accessible, interoperable, and re-usable (FAIR) data and programmatic execution. Adherence to such principles means that informational products developed at one location can be found and used at another. Results can be checked, combined, and leveraged. While all data cannot be public and open, frameworks that *support* this option enable and strengthen work both within and among organizations and individuals.

In order to (re)use the vast amount of measurements we expect to capture in future engineering biology experiments, new databases will need to adhere to these FAIR Principles:

- Findable:
    - Data and metadata are assigned globally unique and persistent identifiers.
    - Data are described with rich metadata.
    - Metadata clearly and explicitly include the identifier of the data they describe
    - (Meta)data are registered or indexed in a searchable resource

- Accessible
    - (Meta)data are retrievable by their identifier using a standardized communication protocol
    - Metadata should be accessible even when the data is no longer available
- Interoperable
    - (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
    - (Meta)data use vocabularies that follow FAIR principles
    - (Meta)data include qualified references to other (meta)data
- Reusable
    - (Meta)data are richly described with a plurality of accurate and relevant attributes

For Engineering biology, these principles apply across the DBTL cycle: Designs should be FAIR to enable characterization across many different organisms, conditions, and implementations for many different teams. Build protocols should be FAIR to ensure reproducibility, and multi-omics measurements across many different studies of the same organism must be FAIR in order to accumulate enough Test data for Learn activities. (For related reading, please see: Wilkinson, M. D., et. al., (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18 and, for a related graphic, please see 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)

DATA INTEGRATION, MODELING, AND AUTOMATION

Goal	Breakthough Capability	Milestone

Establish a computational infrastructure where easy access to data supports the DBTL process for biology.

Established standard and accessible repositories for biomanufacturing data and analysis methods.

Have developed a system of robust communication between academia and industry surrounding engineering biology data access and needs.	Biomanufacturing-specific data standards and repositories.
Develop findable, accessible, interoperable, and reusable (FAIR) data standards and open repositories for engineering biology.	

Common computational infrastructure for finding biological data and common APIs for search and analysis.

Demonstrate common data search and interchange among current biological and chemical repositories and existing microbial biofabrications.	Produce a common library of open design tools for more open medical and agricultural environments.
Produce a common library of open design tools, built upon standard APIs, and supported by portable/virtualized execution environments.	

End-to-end, industry-normed design software platforms for engineered biological systems.

Develop industry-accepted, sharable assessments of current data tools and uses in reducing cost and increasing reliability of executing the DBTL cycle.	Create an industry-accepted, open-source or publically-accessible version of industrially-relevant DBTL software and data.

Establish functional prediction through biological engineering design at the biomolecular, cellular, and consortium scale.

Fully-automated molecular design from integrated, large-scale design data frameworks.

Structure- and comparative analysis-based libraries for automated directed evolution, with feedback of large-scale results to algorithms.	Automated designs for integrated manufacturing to enable more successful, iterated workflows.	Use of large-scale design data in integrated frameworks.	Design and integration of thousands of critical catalytic activities into proteins for a set of model hosts and creation of standard tools for allosteric control of these activities.
	Large-scale design data generation to inform next-generation algorithms for molecular design.		

Use of enzyme promiscuity prediction algorithms to design biosynthetic pathways for any molecule (natural or non-natural).

Retro-biosynthesis software that can identify any biological or biochemical route to any organic molecule.	Data integration for certain classes of enzymes and pathways and predictable host-specific expression in model organisms.	Integrated data that allows on-demand characterization, standardization, insertion, and deployment of natural and non-natural pathways.

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

<table>
<tr><th colspan="4">Scalable, data-driven host design for complex environments that enable high-level production of natural biomolecules.</th></tr>
<tr><td>Ability to make and screen multiple host mutations for epistasis mapping and synthetic interactions, making large-scale host optimization possible.</td><td>Thematic design rules for host system engineering inferred from data.</td><td colspan="2" rowspan="3">Data-driven domestication of any new host for new activities in any environment and scale.</td></tr>
<tr><td rowspan="2">Better data on physiology and fitness in deployment environments suitable for informing design.</td><td>Tools to acquire and transfer data to a novel host to inform both genetic-domestication and prediction and determination of function.</td></tr>
<tr><td>Novel design tools to support host design for more complex, natural (non-laboratory) environments.</td></tr>
<tr><th colspan="4">Enabled design of functional, self-supporting ecosystems.</th></tr>
<tr><td>Data-driven tools for selecting organisms for synthetic assemblies to achieve resistant, resilient activity.</td><td colspan="2" rowspan="3">Integration of molecular, pathway, and host design to create and build models of genetically-engineered communities that function predictably, in the context of deployment ecology.</td><td>Ability to design and build functional, enclosed, self-supporting ecosystems of multiple engineered microbial species for efficient industrial production.</td></tr>
<tr><td>Direct data collection for the most important communities in human, agriculture, and complex bioreactor work sufficient for informing design.</td><td rowspan="2">Ability to design, model, and engineer microbial consortia to simultaneously and efficiently produce multiple products of interest with minimal by-products and waste.</td></tr>
<tr><td>Modeling tools to identify cross-organismal networks and ecological interactions.</td></tr>
</table>

Establish optimal manufacturing processes from the unit-operation to the integrated-screening scale.

<table>
<tr><th colspan="4">Standardized informatics tools, data, and automation platforms for efficient and collaborative use and integration of data in order to develop novel products more quickly.</th></tr>
<tr><td rowspan="2">Establish communications and networks to develop democratized platforms for data exchange and automation across industry and academia.</td><td>Democratized platform for data exchange related to standard/ model microorganisms.</td><td>Democratized platform for non-model organisms and microbial communities.</td><td rowspan="2">Full machine learning capabilities and ability of algorithms to run greater than 90% of the DBTL+automation cycle.</td></tr>
<tr><td>Initial development of a non-model organism database to integrate predicted pathways and -omics data for production.</td><td>Democratized suite of platforms that can be utilized across different model systems.</td></tr>
</table>

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

Roadmap Elements

Goal 1: Establish a computational infrastructure where easy access to data supports the DBTL process for biology.

[Current State-of-the-Art]: The establishment of a computational infrastructure where easy access to data supports the DBTL process for biology is sometimes called data ecology. This means easy access to data and validated models of biological systems, the processes by which they are modified and manufactured, and their reciprocal impact on the environment(s) in which they are deployed. At the core, such access requires both databases of this information and standards that ensure the right information is captured for design. These standards then allow common infrastructures, including applications, programming, and interfaces, for finding, transporting, and analyzing this data. Standards support interoperability of information, portability and reuse of data, tools, and materials, collaboration among teams because of the common communication of data, tools and results, and help to ensure quality, since data and tools in standard formats can be checked for errors in more automated ways. Biological design presents special challenges in that the systems are far more diverse with much less controlled information about them, their operations and interactions with their environment are exceptionally complex in the whole compared to electronic systems (though the engineered aspects tend to be only a small part of the system), and the principles for design and manufacture are evolving rapidly and are highly application specific. The differences among engineering a microbe for production of a high-value chemical, engineering a T-cell for treating a specific cancer, and engineering a plant for growth and productivity in diverse field environments, are large and have different requirements for information and analysis.

Despite the complexities of the data ecology landscape, engineering biologists are increasingly familiar with a large number of key biological information resources. These national repositories and workbenches include those available from NCBI and EBI (REFSEQ, PUBMED, and SWISSPROT), to established repositories of key biological measurement types (PDB, SRA, GEO, ARRAYExpress, and IMG) and more volatile stores like MG-RAST or MicrobesOnline, to knowledge representation sites like METACYC, KEGG and BRENDA that together have been exceptionally important to interpretation of biological data. These are backed by strong data standards groups and ontological development that ensure that data is “represented” using a common language, with the appropriate organized characteristics to support automated statistical and semantic analysis. Further, there are attempts to unify the object ID space so that genes, genomes, taxa, chemicals, etc., can be uniformly labeled and cross referenced and searched across data sets and systems.

Various individual analytical tools and more integrated data and analysis workbenches have begun to arise. General purpose open systems, like KBase and Galaxy, serve different needs, but allow users to extend and share analytical capabilities and data that cross basic and applied biology and biotechnology. The Experimental Data Depot (Morrell et al., 2017) and the Joint BioEnergy Institute Inventory of Composable Elements (Ham et al., 2012) serve as a repositories and representation of data about bioengineered systems, numerous individual genetic device designers like RBSCalculator, and more integrated design systems like Cello, are also available. Further, there has been some effort in the synthetic biology community to

develop standards for interchangeable data, including the Synthetic Biology Open Language (SBOL), the Systems Biology Markup Language (SMBL), and others.

Currently, there are very few widely used integrated computational DBTL-support systems, and of these, they rarely advantage themselves of the large number of diverse biological data and analysis resources. Despite some standards efforts, they remain rather siloed and use idiosyncratic technologies for data representation and analysis execution that hinders community use and development. Further, current focus has been, understandably, on the basic design and construction of pathways and less on scalable production/formulation and on understanding post-deployment behaviors such as differences in operation outside the laboratory, failure modes in real environments, and tracking of designed biological objects in the environment and determining their sources and ecological impact (though there are examples of each of these). There is opportunity for the engineering/synthetic biology community to better advantage itself of the investments being made in other fields of quantitative and systems biology, medicine, chemical process engineering, and environmental science, and to establish its own best practices and standards for its unique aims.

There are three main activities associated such an effort that also deeply involve the experimental practice of synthetic biology and biological manufacture: 1) establishing strong standards for representation of synthetic biological objects, experimental design and process control structures, and measurements of these objects and their outcomes in a series of increasingly complex environments from initial laboratory creation to the sites of their application - these standards should adhere to FAIR (findable, accessible, interoperable, reusable) conventions and computation representations parseable and analyzable within the frameworks built for general computational data science (i.e., utilizing standards for ontology, ID space, data formats (e.g., RDF, JSON), and metadata for provenance); 2) demonstrating scalable computational libraries and infrastructure for repositing, searching, transporting, and aggregating/organizing these data types for analysis; and 3) the establishment of open, scalable software platforms that accelerate efficient, predictable design by enabling integrated access to the appropriate biological data, presented in design-oriented ways, and supported by a community-extensible set of tools whose results can be compared and contrasted to determine best practice over time. In each of these cases, the roadmap calls for starting with designs that operate in single organisms in laboratory conditions and scale out to multicellular systems deployed in more open conditions.

[Breakthrough Capability 1]: Established standard and accessible repositories for biomanufacturing data and analysis methods.

- **2 years: Have developed a system of robust communication between academia and industry surrounding engineering biology data access and needs.**
    - [Bottleneck]: Lack of connection and communication between information systems beyond engineering biology.
        - [Potential Solution]: Identify core needs for common data/model access spanning molecular and organismal biology, biomanufacturing processes, and tracking operation in deployment.

        - [Potential Solution]: In collaboration with existing biological data groups including those from NCBI/EBI/USDA, etc., develop biological design-oriented access and standards for data spanning, for example, protein structure, genomics, genotype-phenotype data, and treatment/disease data.
        - [Potential Solution]: In collaboration with existing chemical and materials data groups, develop biological design-oriented access and standards for data.
- **2 years: Develop findable, accessible, interoperable, and reusable (FAIR) data standards and open repositories for engineering biology.**
    - [Bottleneck]: Lack of standards for data exchange and communication.
        - [Potential Solution]: Prioritize linkage to existing biological databases; identify the need for establishment of new repositories.
- **5 years: Biomanufacturing-specific data standards and repositories.**
    - [Bottleneck]: Lack of universal agreement of standard parameters and repositories to prioritize.
        - [Potential Solution]: Coordinated effort to obtain input, decision, and agreement on a set of standards and repositories to use.

[Breakthrough Capability 2]: Common computational infrastructure for finding biological data and common APIs for search and analysis.

- **5 years: Demonstrate common data search and interchange among current biological and chemical repositories and existing microbial biofabrications.**
    - [Bottleneck]: Lack of agreed upon approach to common data search and diversity among current repositories.
- **5 years: Produce a common library of open design tools, built upon standard APIs, and supported by portable/virtualized execution environments to demonstrate best-practice interoperable biomanufacturing software.**
    - [Bottleneck]: Missing incentives for development and use of open design tools.
- **10 years: Produce a common library of open design tools for more open medical and agricultural environments.**
    - [Bottleneck]: Availability of design tools geared specifically to the unique requirements of biomedical and agricultural data collection and use.

[Breakthrough Capability 3]: End-to-end, industry-normed design software platforms for engineered biological systems.

- **5 years: Develop industry-accepted, sharable assessments of current data tools and uses in reducing cost and increasing reliability of executing the DBTL cycle.**
    - [Bottleneck]: Missing incentives for industry coordination and collaboration on tools assessment.
- **10 years: Create an industry-accepted, open-source or publically-accessible version of industrially-relevant DBTL software and data.**
    - [Bottleneck]: Agreement on standards and interoperability as well as complete understanding of industry needs required to enable full utilization.

Goal 2: Establish functional prediction through biological engineering design at the biomolecular, cellular, and consortium scale.

[Current State-of-the-Art]: ROSETTA, MOE, and NAMD are representative software platforms for biomolecular structure-based design and for the simulation of small molecules and peptides to proteins and larger systems. Google DeepMind's recent success at CASP13 (AlQuraishi, 2019) demonstrated that machine-learning approaches are also increasingly effective for biomolecular structure prediction, and it is anticipated that design and simulation will increasingly integrate physics- and structure-based modeling with statistical comparative- and screening-based data. Existing software tools are largely sufficient to design protein libraries to experimentally explore molecular space, predict protein domains and other structural boundaries, and leverage comparative (meta)genomics to build deep sets of sequence orthologs for important protein classes and suggest tolerable/efficacious mutation locations. Current limitations of these software include dependencies upon imperfect force-fields, a lack of full quantitative and allosteric modeling and parallel computation, and insufficient design-of-experiments support and structural coverage for statistical analyses. While it seems likely that high-throughput screening combined with machine learning may provide a data-driven approach to identifying function from sequence without resorting to first principles or ground-up approaches, measuring molecular activity at scale remains a key bottleneck.

The design of organisms with a targeted metabolic function (e.g., overexpression of a single biomolecular species) requires computational tools that: 1) identify sets of proteins that can convert readily available molecules to high value products, each protein performing one of a series of chemical modifications; and 2) identify best sets of enzymes and their stoichiometry that can work together as parts of pathways in the context of cellular metabolism. On the pathway level, genome-scale metabolic models link genotype to phenotype through the reconstruction of the complete metabolic reaction network of an organism. This technique can be used to define theoretical production limits and design and test new microbial strains *in silico*. This approach has been especially effective for predicting and improving metabolite production rates in heterologous biosynthetic pathways. Flux Balance Analysis (FBA), Flux Variability Analysis (FVA), and minimization of metabolic adjustment (MOMA) have been successfully used, in combination with genome-scale metabolic models, to predict cell growth, flux distribution, product synthesis, and to guide host design. A MATLAB toolbox called COBRA ("COnstraint-Based Reconstruction and Analysis; (Heirendt et al., 2019)) provides a convenient framework to simulate and analyze the phenotypic behavior of a genome-scale stoichiometric model (Schellenberger, Lewis, & Palsson, 2011), and retrobiosynthesis tools such as BNICE ("Biochemical Network Integrated Computational Explorer") and RetroPath are used to design new or improved biochemical pathways (Medema, van Raaphorst, Takano, & Breitling, 2012). In these design tools, software identifies novel metabolites, reactions, and whole pathways by predicting promiscuity based on classification of enzymes according to their chemical action. On the cellular level, a wide variety of host design tools have been developed for identification of gene targets for knockout, overexpression, or downregulation, introduction of non-native enzymatic reactions, and elimination of competing pathways in order to improve the cellular phenotypes (Long, Ong, & Reed, 2015). Pathway and host improvements achieved from these design tools are often non-intuitive and non-obvious. And, while genome-scale metabolic

models have been important for metabolic engineering efforts with organic compounds, advances are still required to transform the bioeconomy.

When it comes to community and consortia design, we are primarily in a state of data gathering and developing a baseline understanding of microbial communities across diverse locations/ecosystems, thus tools for multi-scale modeling at multicellular, organismal, and population levels have yet to be developed.

[Breakthrough Capability 1]: Fully-automated molecular design from integrated, large-scale design data frameworks.

- **2 years: Structure- and comparative analysis-based libraries for automated directed evolution, with feedback of large-scale results to algorithms.**
    - [Bottleneck]: Lack of shared libraries and robust assessment of computational approaches to directed evolution.
- **5 years: Automated designs for integrated manufacturing to enable more successful, iterated workflows.**
    - [Bottleneck]: Lack of integration of automation design tools.
- **5 years: Large-scale design data generation to inform next-generation algorithms for molecular design.**
    - [Bottleneck]: Insufficient standards and coordination among data generators to create robust datasets that can be successfully used for design.
- **10 years: Use of large-scale design data in integrated frameworks.**
    - [Bottleneck]: Lack of standardized datasets that can be integrated into diverse frameworks.
- **20 years: Design and integration of thousands of critical catalytic activities into proteins for a set of model hosts and creation of standard tools for allosteric control of these activities.**
    - [Bottleneck]: The current lack of standardized data, integration across platforms, and models of unknown catalytic activity put this currently make part, pathway, model integration far out of reach.

[Breakthrough Capability 2]: Use of enzyme promiscuity prediction algorithms to design biosynthetic pathways for any molecule (natural or non-natural).

- **2 years: Retro-biosynthesis software that can identify any biological or biochemical route to any organic molecule.**
    - [Bottleneck]: There are a nearly infinite number of chemicals that we want to produce using engineered hosts; however, the routes (biological-only or a combination of biological and chemical) to these chemicals are not always known or easy to imagine.
        - [Potential Solution]: Develop retrobiosynthesis software for all *known* metabolic pathways in all life forms and integrate that software with retrosynthesis software of all chemical catalysis to develop pathways to nearly any organic chemical.

- **5 years: Data integration for certain classes of enzymes and pathways and predictable host-specific expression in model organisms.**
    - [Bottleneck]: Limited integrated data to link pathway activity and expression.
        - [Potential Solution]: Specialized pathway optimization tools for these pathways and molecules.
- **20 years: Integrated data that allows on-demand characterization, standardization, insertion, and deployment of natural and non-natural pathways.**
    - [Bottleneck]: Lack of data integration and standards for data sharing.

[Breakthrough Capability 3]: Scalable, data-driven host design for complex environments that enable high-level production of natural biomolecules.

- **2 years: Ability to make and screen multiple host mutations for epistasis mapping and synthetic interactions, making large-scale host optimization possible.**
    - [Bottleneck]: Limits of data integration and databases of characterized genetic and pathway/circuit interactions.
        - [Potential Solution]: Backed-by-design tools based on merging pathway knowledge and experiential databases.
- **2 years: Better data on physiology and fitness in deployment environments suitable for informing designs in validated lab-scale simulations that meet activity, persistence, and ecological impact goals.**
    - [Bottleneck]: Limited data availability and lack of a coordinated collection effort.
- **5 years: Thematic design rules for host system engineering inferred from data.**
    - [Bottleneck]: Limits of data integration and feedback into design for desired production of molecules from discrete pathways/circuits in select organisms.
        - [Potential Solution]: Tools for specific host system optimization given production/activity class of target molecules (including sensors, regulators, and pathways).
- **5 years: Tools to acquire and transfer data to a novel host to inform both genetic-domestication and prediction and determination of function.**
    - [Bottleneck]: Limited data and predictive models for cross host domestication and function determination.
- **5 years: Novel design tools to support host design for more complex, natural (non-laboratory) environments.**
    - [Bottleneck]: The currently available tools and datasets to integrate host design and ecological data are limited and not standardized for cross-domain analysis.
- **20 years: Data-driven domestication of any new host for new activities in any environment and scale.**
    - [Bottleneck]: A diverse dataset and robust algorithms to fully model domestication of any potential host.

[Breakthrough Capability 4]: Enabled design of functional, self-supporting ecosystems.

- **2 years: Data-driven tools for selecting organisms for synthetic assemblies to achieve resistant, resilient activity.**
    - [Bottleneck]: Lack of open-source tool development focused directly on organism selection.

- **2 years: Direct data collection for the most important communities in human, agriculture, and complex bioreactor work sufficient for informing design.**
    - [Bottleneck]: Lack of standardized framework for data collection.
- **2 years: Modeling tools to identify cross-organismal networks and ecological interactions.**
    - [Bottleneck]: Insufficient data to support models in complex environments.
- **10 years: Integration of molecular, pathway, and host design to create and build models of genetically-engineered communities that function predictably, in the context of deployment ecology.**
    - [Bottleneck]: Inability to infer or determine cellular- and subcellular-level mechanistic-modes due to computational complexity.
        - [Potential Solution]: Develop more comprehensive algorithms for modeling purposes that specifically take advantage of domain specific knowledge, algorithmic advances leveraging parallelization, and hardware advances, such as the use of specialized electronic circuits.
- **20 years: Ability to design and build functional, enclosed, self-supporting ecosystems of multiple engineered microbial species for efficient industrial production.**
    - [Bottleneck]: Lack of data, tools, and standards for the production and dissemination of data-driven design-build integration.
- **20 years: Ability to design, model, and engineer microbial consortia to simultaneously and efficiently produce multiple products of interest with minimal by-products and waste.**
    - [Bottleneck]: Lack of understanding, data, and models on how complex consortia interact and the implications of such interactions that can affect engineering goals.

Goal 3: Establish optimal manufacturing processes from the unit-operation to the integrated-screening scale.

[Current State-of-the-Art]: Current state-of-the-art capabilities for generalized biofabrication reside primarily in large, well-established organizations (such as the Broad Institute) and biotechnology companies (such as Zymergen). The state-of-the-art, however, is still a somewhat ad-hoc assemblage of product-oriented tools, customized software local to that institution, proprietary data sets, custom automation solutions, and customized data-logging and analysis systems. Often, industry views its proprietary approach to data flow and informatics as unique and as a large part of their value-proposition and tends to sequester informatics gains to particular institutions. However, to address this, there is a rapidly growing number of public-funded, non-commercial bio-foundries, which has recently resulted in the establishment of the Global Biofoundries Alliance (Hillson et al., 2019). The aims of the Alliance are to establish open technology platforms that will allow the sharing of automation workflows and protocols, software, reference materials and best practices which may lead to new standards for measurement and data, as well as global capacity for establishing optimal manufacturing processes for synthetic biology.

[Breakthrough Capability]: Standardized informatics tools, data, and automation platforms for efficient and collaborative use and integration of data in order to develop novel products more quickly.

- **2 years: Establish communications and networks to develop democratized platforms for data exchange and automation across industry and academia.**
    - [Bottleneck]: Lack of standards for data exchange and communication, lack of standards of automation platforms, and extreme cost of automation for implementation in non-industrial settings.
        - [Potential Solution]: Dialogue between industry and academic scientists to develop standardized, cheaper, automated platforms for high-throughput experimentation of commonly used microbes.
        - [Potential Solution]: Development of a greater number and better-connected industry-academic consortia to share ideas, equipment, and platforms.
- **5 years: Democratized platform for data exchange related to standard/model microorganisms.**
    - [Bottleneck]: Lack of standards for exchange of design information and communication with automated systems for both build-execution and test-data-acquisition for commonly used microbes.
        - [Potential Solution]: Launch new industrial-academic consortia or partnerships that leverage shared automation technology and platforms.
        - [Potential Solution]: Start leveraging consortia to develop industry- and academia-wide data-logging, -analysis, and -sharing standards.
        - [Potential Solution]: Create standards-based approach to data exchange with access to integrated, generalized databases.
        - [Potential Solution]: Incorporation of design-of-experiments strategies that integrate the economics of obtaining data and required standards for data precision and accuracy.
        - [Potential Solution]: Collaborations to develop new data analysis tools specific to biology.
        - [Potential Solution]: Miniaturization of automated host engineering and analytical systems and integration into desktop machines able to re-engineer microbes. These machines may act genome-wide, in an automated fashion without human interference, i.e., from design to organism in one iteration.
- **5 years: Initial development of a database of organisms beyond *E. coli* and *S. cerevisiae* (i.e., a database of non-model organisms), that leverages existing databases, to integrate predicted pathways, and -omics data that confirm specific production of a compound of interest.**
    - [Bottleneck]: There are many organisms that might be useful in a particular environment or for producing a particular chemical; however, identifying the most useful hosts beyond current model organisms is challenging.

        - [Potential Solution]: Leverage existing databases (Biocyc, KEGG, etc.) to construct a database of non-model organisms that focuses on the known functionality of the non-model organism.
- **10 years: Democratized platform for non-model organisms and microbial communities.**
    - [Bottleneck]: Lack of standards for exchange of design information and communication with automated systems for both build-execution and test-data-acquisition for non-model microbes.
        - [Potential Solution]: Extend industry-academic consortia, and platforms/data sharing solutions for model organisms (5 year milestones) to problems associated with non-model organisms.
        - [Potential Solution]: Extend microfluidics approaches to experiments with non-model organisms.
- **10 years: Democratized suite of platforms that can be utilized across different model systems.**
    - [Bottleneck]: Full biological characterization of greater libraries of organisms, including transferability of design, modeling, and engineering strategies between organisms.
        - [Potential Solution]: Extend industry-academic consortia, and platforms/data sharing solutions for model organisms (5 year milestones) to problems associated with non-model organisms.
        - [Potential Solution]: Extend microfluidics approaches to experiments with non-model organisms.
- **20 years: Full machine learning capabilities and ability of algorithms to run greater than 90% of the DBTL+automation cycle.**
    - [Bottleneck]: Current understanding of the principles behind optimal design for complex systems is still limited.
        - [Potential Solution]: Improve predictability of complex systems through many years of iterations of current algorithms; much of this solution should happen organically honing of machine learning and artificial intelligence algorithms continues over time, particularly with increased access to standardized data and democratized automation platforms.
        - [Potential Solution]: Work to create new data, coding, and analytical languages that better capture the rules of biology.

Application and Impact Sectors

Industrial Biotechnology

Industrial Biotechnology

Industrial Biotechnology focuses on technical challenges relevant to industrial use of synthetic biology and the establishment of the United States as a global leader in the bio-based economy. Applications of engineering biology in this sector focus on increasing market share of bio-based products by US companies through the use of synthetic biology approaches, making sustainable manufacturing processes cost-competitive, accelerating innovation and discovery with respect to making new products and technologies, and generating products at scales necessary for economic viability. The goal is to stimulate investment in infrastructure, streamline production, and advance our use of engineering biology to improve lives. Many of the Challenges, Aims, and Objectives of the roadmap for *Industrial Biotechnology* are influenced by the findings of *Industrialization of biology: A roadmap to accelerate the advanced manufacturing of chemicals.* 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, & National Research Council. (2015). Washington (DC): National Academies Press (US).

Societal Challenge 1: Enable next-generation production through sustainable, cost-competitive, flexible, and efficient manufacturing processes.

- Science/Engineering Aim 1: Better use of abundant, renewable substrates to make specialty chemicals via economically viable processes.
    - Engineering Biology Objective 1: Modular systems (such as enzymes, consortia, and cell-free systems) that can adapt to different feedstocks and be easily modified to produce different target chemicals.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to edit genomes of microbial and fungal species that can rapidly degrade cellulosic biomass and other renewable feedstocks.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Rapid design and production of custom enzymes and enzyme pathways.
            - Assembled sets of proteins that can completely degrade sustainable feedstocks.
            - Regulatory components (including sensors and networks) that program the system to adapt to the feedstock, intermediates, and side products.
        - Host and Consortia Engineering Achievement:
            - Engineered microbial consortia with predictable composition, dynamics and function, to feed off of sequential byproducts in an (almost) closed-loop system.
        - Data Integration, Modeling, and Automation Achievement:
            - Novel analytics tools to enable prediction and manipulation of holistic microbial ecosystem function by incorporating both biological and environmental data.

        - Analytical tools to determine and predict matching of organism, strain, or pathway with feedstock/substrate source for best productivity, yield, and lowest cost.
    - Engineering Biology Objective 2: New enzymes and cells that work synergistically to degrade biomass and process by-products, possibly in combination with new chemical innovations.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Enzymes engineered to degrade renewable materials and process byproducts faster and more completely.
            - Metabolic or protein engineering approaches to enable complete use of all substrate components and byproducts.
        - Host and Consortia Engineering Achievement:
            - Increased protein secretion rates to enable on-demand enzyme synthesis and release.
            - Engineered microbial consortia with predictable composition, dynamics and function, to feed off of sequential byproducts in an (almost) closed-loop system.
        - Data Integration, Modeling, and Automation Achievement:
            - Prediction of protein and cell assemblies that will exhibit desired target production, considering both composition and how components are physically assembled.
            - Prediction and modeling of microbial consortia functioning, specifically synergistic pathways and byproduct recycling.
    - Engineering Biology Objective 3: Use of novel and lesser-used substrates/feedstocks for manufacturing processes that are more efficient and more environmentally-sustainable than currently available substrates. This engineering biology objective is aimed at developing new substrate and feedstock systems based solely on biotechnology industry needs, rather than utilizing offshoots of systems created for other purposes (such as farming or animal feeds).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improved methods for transformation of organisms with heterologous pathways.
        - Host and Consortia Engineering Achievement:
            - Ability to engineer currently used organisms and hosts to work well with feedstocks deemed promising from environmental and cost standpoints.
        - Data Integration, Modeling, and Automation Achievement:
            - Data analysis approaches combining sustainability analyses with strain/pathway/methodology product, yield, and efficiency to

determine promise of lesser-known feedstocks from both industrial-productivity and environmental-sustainability standpoints.
        - Modeling methods combining environmental and economic factors to determine the best ways to implement production of new feedstocks.
- Science/Engineering Aim 2: More efficient production of (bio)chemicals, bio-based products, and other specialty materials. The goal of this Aim is to lower energy usage, lower waste processing, and reduce water use in manufacturing and industrial settings.
    - Engineering Biology Objective 1: Implementation of computational approaches to assemble new systems that consider multiple facets of production. (Including feedstocks, target product, available facilities, biological component characteristics and limitations, among others.)
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
            - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Rapid design and production of custom enzymes and enzyme pathways.
            - Metabolic or protein engineering approaches to enable complete use of all substrate components.
        - Host and Consortia Engineering Achievement:
            - Engineered microbial consortia with predictable composition, dynamics and function, to feed off of sequential byproducts in an (almost) closed-loop system.
            - Development of fast-growing variants of non-model production hosts.
        - Data Integration, Modeling, and Automation Achievement:
            - Data integration methodology and approaches to describe and compare system performance.
            - Design-of-experiments approaches to obtain required data to enable prediction.
            - Artificial intelligence and/or machine learning approaches to predict how systems should be assembled considering production goals and constraints.
    - Engineering Biology Objective 2: Enable a broader range of microorganisms to be used in traditional biomanufacturing industries to expand the scope of natural product discovery and production.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.

        - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Rapid adaptation of enzymes to work in the context of different hosts.
    - Host and Consortia Engineering Achievement:
        - Development of fast-growing variants of non-model production hosts.
    - Data Integration, Modeling, and Automation Achievement:
        - Prediction of media components, additives, environmental conditions that promote growth of non-model production hosts from genomic data.
- Engineering Biology Objective 3: Experimental and computational approaches to increase growth rates, yield, and efficiency of production hosts.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
        - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Development of regulatory components that enable dynamic regulation to optimally balance growth and production, especially of toxic or high-energy-requiring products.
    - Host and Consortia Engineering Achievement:
        - Ability to engineer non-model production hosts with increased growth rates and improved yield and efficiency, especially under bioproduction conditions.
    - Data Integration, Modeling, and Automation Achievement:
        - Computational approaches to classify mutations found in slow-growing production hosts, to define if they are necessary or detrimental for cell growth and processing.
        - Automation approaches to screen new candidate hosts for fast growth and desired production rates.

- Science/Engineering Aim 3: Modular manufacturing to enable flexible, on-demand production of a range of target chemicals. The goal of this Aim is to reduce transportation costs of both feedstocks and products to improve economic feasibility, especially of lower-value chemicals like biofuels.
    - Engineering Biology Objective 1: Development of commercial systems for on-demand manufacturing of commodity and high-value chemicals. The ideal system will include on-demand capabilities for both upstream (production) and downstream (purification) elements of the process.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
            - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Assembled sets of proteins that can completely degrade sustainable feedstocks.
            - Regulatory components (including sensors and networks) that program the system to adapt to the feedstock, intermediates, and side products.
        - Host and Consortia Engineering Achievement:
            - Engineered microbial consortia with predictable composition, dynamics and function, to feed off of sequential byproducts in an (almost) closed-loop system.
        - Data Integration, Modeling, and Automation Achievement:
            - Novel analytics tools to enable prediction and manipulation of holistic microbial ecosystem function by incorporating both biological and environmental data.
            - Analytics tools and approaches to develop flexible manufacturing processes.
            - Widely adopted methods for defining reproducible workflows that can be used by cloud laboratories to embed protocols for implementation, characterization, and verification and validation of components, pathways/circuits, subsystems, cells, consortia, and multicellular organisms.
            - Modular field production facilities that can accommodate many manufacturing protocols.
            - Establish life cycle assessments to determine efficiency, sustainability, and feasibility of protocols and processes.

- Engineering Biology Objective 2: Develop hosts or consortia that can generate multiple products from a single process. The goal of this Objective is to generate product streams which can easily be separated, or modified, to produce different target products, such as by changing an environmental condition or one biological component.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
    - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
    - Community-level, metagenome editing.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Assembled sets of proteins that can completely degrade sustainable feedstocks.
    - Regulatory components (including sensors and networks) that program the system to adapt to the feedstock, intermediates, and side products.
    - Regulatory components that allow the user to easily switch between different target products.
  - Host and Consortia Engineering Achievement:
    - Ability to exert tight control over pathways (such as through dynamic metabolic engineering) that are not being used in production.
    - Engineered microbial consortia with predictable composition, dynamics, and function, to feed off of sequential byproducts in an (almost) closed-loop system.
    - Adaptable hosts optimized for production of multiple products.
  - Data Integration, Modeling, and Automation Achievement:
    - Novel analytics tools to enable prediction and manipulation of holistic microbial ecosystem function by incorporating both biological and environmental data.
    - Modeling and analytics tools for building systems with multiple objectives and constraints.
- Engineering Biology Objective 3: Engineer off-the-shelf hosts and microbial communities that can rapidly adapt and produce a target product(s) at high yield and high concentration. The goal of this objective is to enable production hosts that can rapidly adapt to different feedstocks, culture conditions, or toxic products and do so in increasingly closed-loop systems.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
    - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.

        - Community-level, metagenome editing.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Regulatory components (including sensors and networks) that program the community to adapt to the feedstock, intermediates, and side products.
    - Host and Consortia Engineering Achievement:
        - Engineer fast growing organisms that can rely on a variety of feedstocks.
        - Engineered microbial consortia with predictable composition, dynamics, and function, to feed off of sequential byproducts in an (almost) closed-loop system.
        - Ability to exert tight control over pathways (such as through dynamic metabolic engineering) that are not being used in production.
        - Reliable strategies for microbial community assembly that promote desired community composition and high-levels of productivity.
        - Engineered host organisms that can be stored without freezing and easily shipped.
        - Engineered hosts that produce fewer (or no) toxic by-products.
    - Data Integration, Modeling, and Automation Achievement:
        - Novel analytics tools to enable prediction and manipulation of holistic microbial ecosystem function by incorporating both biological and environmental data.
        - Automation strategies for assessing community composition and function dynamically.
        - Modular field production facilities that can accommodate many manufacturing protocols.

Societal Challenge 2: Scalable production of novel and existing products that are more sustainable and economically- and environmentally-friendly.

- Science/Engineering Aim 1: Improved ability to identify and make commodity, specialty, and high value chemicals and materials.
    - Engineering Biology Objective 1: Modeling, design, and test of pathways to make molecules and products that do not exist in nature.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Precise and automated generation of large collections of gene variants for pathway designs.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Design of enzymes for converting unnatural substrates.
            - At-will design of non-natural pathways for the production of novel products and materials.

        - Host and Consortia Engineering Achievement:
            - Creation of hosts suitable for the production of certain natural and synthetic molecule families.
            - Pathway integration processes for hosts that are easy and high-throughput.
            - Understanding of, and solutions for, product toxicity to the host.
        - Data Integration, Modeling, and Automation Achievement:
            - At-will design of non-natural pathways for the production of novel products and materials.
            - Automated robotic screening of desired phenotypes coupled with precise analytics of desired molecules and side products.
    - Engineering Biology Objective 2: Novel methodologies for discovery and optimization of existing metabolic pathways in host organisms.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Precise and automated generation of large collections of gene variants.
            - Gene editing capabilities for diverse production organisms.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Better enzyme and pathway design in industrial production hosts.
        - Host and Consortia Engineering Achievement:
            - Molecular tools for novel or unnatural production organisms.
        - Data Integration, Modeling, and Automation Achievement:
            - Automated robotic screening of desired phenotypes coupled with precise analytics of desired molecules and side products.
            - Accurate prediction of gene and enzyme function for poorly annotated genomes to allow more effective bioprospecting.
    - Engineering Biology Objective 3: Production of biologics in organisms at economically-viable scales.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Gene editing capabilities for diverse production organisms to ensure access to “the right tool for the job”.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Novel biosensors that respond to cheaper and/or more efficient inducer molecules (chemical pathway “on/off switches”).
            - Automated directed evolution of multi-gene pathways to make a single host species or strain as productive as possible (such as able to synthesize multiple products).
            - Engineer cell consortia for orchestrated production of multiple products simultaneously.
        - Host and Consortia Engineering Achievement:
            - Engineer host organisms that grow quickly during production and more efficiently use diverse feedstocks.

        - Conversion of side products to valuable commodities for enhancing economic value.
    - Data Integration, Modeling, and Automation Achievement:
        - Improved prediction and analysis of flux via metabolic pathways to determine maximum theoretical yields under different fermentation conditions.
- Engineering Biology Objective 4: Commercialization of new types of products using bio-inspired fabrication, including engineered living materials and devices.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Precision gene editing *in vivo*.
        - Reliable and efficient delivery vectors for gene editing agents.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Optimize pathways for the production of novel bio-polymer materials.
    - Host and Consortia Engineering Achievement:
        - Create engineered environments with three-dimensional structures and multiple cellular types ordered and arranged in a controlled manner throughout the structure.
        - Engineer hosts with controllable “kill switches”, decreased toxic byproducts, and decreased immunogenicity (especially for biosensors for multicellular hosts and/or environments).
    - Data Integration, Modeling, and Automation Achievement:
        - Enhanced prediction of compatibility between bio-fabricated materials and hosts/environments.

- Science/Engineering Aim 2: Manufacturing of consumable and infrastructure products via synthetic biology, including food, textiles, building materials, and packaging.
    - Engineering Biology Objective 1: Innovations in bio-based manufacturing of commodity products.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Highly efficient organismal gene and genome editing.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Enable more efficient and rapid engineering of orthogonal chemical communications pathways between cells.
            - Heterologous pathway transformation into organisms from different kingdoms.
        - Host and Consortia Engineering Achievement:
            - General molecular toolbox that can be applied to a wide range of hosts.
        - Data Integration, Modeling, and Automation Achievement:
            - Screening capabilities for bio-based production.
            - Databases and foundries for microbial isolates.

    - Engineering Biology Objective 2: Synthetic and natural organisms that can convert agricultural wastes into commodity products.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
            - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists to find hosts that are naturally more efficient or that have more efficient pathways that can be placed into heterologous hosts.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineered proteins and large complexes with efficient catalytic capabilities for conversion of bulk agricultural waste into products with downstream applicability.
            - Regulatory components (including sensors and networks) that program the organism to adapt to the feedstock, intermediates, and side products.
        - Host and Consortia Engineering Achievement:
            - General molecular toolbox that can be applied to a wide range of hosts.
            - Engineer hosts that can sustain production yield and efficiency under a wide range of stress conditions.
        - Data Integration, Modeling, and Automation Achievement:
            - Screening capabilities for bio-based production.
            - Libraries and foundries for microbial isolates.
            - Analysis and prediction of metagenomic data.
- Science/Engineering Aim 3: Safer production processes through reduced toxin use and toxic byproduct synthesis. (Including a reduction in downstream toxins from degradation of materials after use.)
    - Engineering Biology Objective 1: Risk, safety, and life-cycle assessments for bio-based manufacturing processes.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Safe, reliable, and efficient delivery vectors for gene editing agents.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineered biosensors for toxic compounds.
        - Host and Consortia Engineering Achievement:
            - Metabolic pathway engineering to prevent synthesis of toxic byproducts and/or to metabolize toxins.
        - Data Integration, Modeling, and Automation Achievement:
            - High-throughput screening approaches to assess toxin production.
            - Development of holistic models of risk/safety with respect to manufacturing process (including employees), environment, and consumers.

        - Develop new and better life cycle assessment models for biomolecule production.
    - Engineering Biology Objective 2: Incentives and regulations, when needed, to ensure the safety of those involved in production and of consumers, as well as overall process sustainability.
        - Data Integration, Modeling, and Automation Achievement:
            - Development of holistic models of risk/safety with respect to manufacturing process (including employees), environment, and consumers.
            - Undertake life cycle assessment to determine efficiency, sustainability, and feasibility of protocols and processes.
    - Engineering Biology Objective 3: Engineer processes, pathways, and enzymes to minimize waste production.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Ability to synthesize, edit, assemble, and deliver many genes and regulatory components in a single cell.
            - Ability to edit genomes of diverse hosts, including microbes, fungi, and protists.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Assembled sets of proteins that can completely degrade sustainable feedstocks.
            - Regulatory components (including sensors and networks) that program the system to adapt to the feedstock, intermediates, and side products.
        - Host and Consortia Engineering Achievement:
            - Engineered microbial consortia with predictable composition, dynamics, and function.
            - Metabolic pathway engineering to prevent synthesis of toxic byproducts and/or to metabolize toxins.
        - Data Integration, Modeling, and Automation Achievement:
            - Novel analytics tools to enable prediction and manipulation of holistic microbial ecosystem function by incorporating both biological and environmental data.
- Science/Engineering Aim 4: Better tools for rapidly translating desired products or process features into industrial workflows and accelerate time-to-market.
    - Engineering Biology Objective 1: Predictive and generalized models that allow laboratory-scale results to be accurately projected to industrial scale processes and vice-versa.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Methods for creating variant libraries that can be used for validating models of genetic circuits and pathways.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Ability to predict function from sequence.

        - Ability to design sequences for desired functions.
    - Host and Consortia Engineering Achievement:
        - Better methods for predicting interactions between components and for designing components and subsystems that behave as expected.
    - Data Integration, Modeling, and Automation Achievement:
        - Higher-throughput data collection and analysis.
        - Better tools for predictive modeling across scales and environments.
        - Ability to estimate robustness of circuits and pathways to genetic, host, and environmental context.
- Engineering Biology Objective 2: Common (and reproducible) standards for biological components and subsystems that enable re-use and efficient component suppliers.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Methods for modular assembly of subsystems and replacement of components required to reconfigure subsystem interfaces.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Protein libraries that allow modular replacement of domains to mix and match functions required for circuit and pathway engineering.
    - Host and Consortia Engineering Achievement:
        - Methods of defining “modules”, and flexible interconnection of modules, that maintain the desired function of the module independent of the operation of other parts of the circuit or genetic, host, environmental context.
    - Data Integration, Modeling, and Automation Achievement:
        - Methods for modeling components/subsystems that allows better characterization and prediction of effects of interaction with other modules and cellular resources, as well as the effects of uncertainty (such as context or system noise).
        - Automated techniques for assembly and characterization of complex circuits consisting of thousands of individual elements (organized as interacting subsystems).
- Engineering Biology Objective 3: Increased rate of design-build-test-learn cycles that combine design, modeling, prototyping, implementing, and characterization of components, pathways/circuits, subsystems, cells, consortia, and multicellular organisms.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Faster, lower-cost methods for creating genome-length sequences that are generated by design tools.

- Biomolecule, Pathway, and Circuit Engineering Achievement:
  - Better methods for design and characterization of proteins and multi-protein complexes for performing a wide variety of functions required at component to subsystem scales.
- Host and Consortia Engineering Achievement:
  - Ability to design, implement, and characterize circuits/pathways, subsystems, cells, consortia and multicellular organisms consisting of hundreds to millions of individual components through a modular, hierarchical framework that enables reuse of interacting components.
- Data Integration, Modeling, and Automation Achievement:
  - Layered and modular design abstractions and the modeling, characterization, and testing tools required to support the creation and use of components, pathways/circuits, and subsystems to create engineered cells, consortia, and multicellular organisms.

- Engineering Biology Objective 4: Deployment and improved use of automation for both research and translational activities to increase throughput.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Well-defined libraries with associated assembly and editing methods for genomic sequences that support the output of compilers and other design tools.
    - Increased cross-talk between geneticists and automation experts to hone efficient, high-throughput, automated laboratory protocols and workflows for gene editing and assembly and genetic library creation and exploration.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Well-defined libraries with associated assembly and editing methods for protein domains and molecular machines that support the output of compilers and other design tools.
  - Host and Consortia Engineering Achievement:
    - Engineered hosts with predictable composition, dynamics, and function.
    - Continually refined physical automation infrastructure and processes (including factories, robots, assembly lines, and workflows) to enable more efficient and modular high-throughput engineering of multiple different kinds of organisms, pathways, and product outputs.
  - Data Integration, Modeling, and Automation Achievement:
    - Widely adopted methods for defining reproducible workflows that can be used by cloud laboratories to implement protocols for implementation, characterization, and verification and validation of components, pathways/circuits, sub-systems, cells, multicellular organisms, consortia, and automation platforms.

Health & Medicine

Health & Medicine

Health & Medicine focuses on technical challenges relevant to the well-being of humans, non-human animals, and populations. Applications of engineering biology in this sector focus on preventing and eradicating disease and supporting longevity and quality of life. For related reading about tools and technologies that impact human and animal health, please see *Environmental Biotechnology* and *Food & Agriculture.*

Societal Challenge 1: Eradicate existing and emerging infectious diseases.

- Science/Engineering Aim 1: Mitigate the threat of microbial (non-viral) pathogens.
  - Engineering Biology Objective 1: Develop tools for rapidly and inexpensively diagnosing antimicrobial-resistant (AMR) susceptibilities and infections.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Error-free DNA synthesis for rapid, high-yield production of antibody proteins and sensors built from nucleic acids.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Rapid antibody development for detecting AMR pathogens.
    - Host and Consortia Engineering Achievement:
      - Develop cell-free systems to detect RNA signatures of AMR pathogen susceptibility.
      - Improve properties such as shelf-life and levels of protein expression of cell-free systems.
      - Develop cell/tissue models to screen and test anti-AMR interventions *in situ*.
    - Data Integration, Modeling, and Automation Achievement:
      - Improve prediction of AMR-conferring operons and markers, and their risk of transmission between organisms, to inform diagnostic tools.
      - Models for transforming -omics data to levels of susceptibility and resistance.
      - Improve identification, prediction, and modeling of characteristic pathways leading to resistance (for example, sequences of genetic changes).
      - Automate electronic reader systems for cheap and fast sequencing of AMR markers and patient-susceptibility biomarkers.
  - Engineering Biology Objective 2: Develop tools to treat microbial infections, overcome antimicrobial-resistance, and reduce the dependence upon antibiotics in humans, pets, livestock, and other animal populations.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Gene delivery systems targeted to specific pathogens.
      - Scaled-up synthesis of high-quality DNA encoding anti-microbial gene circuits.

- Biomolecule, Pathway, and Circuit Engineering Achievement:
  - Develop evolvable therapies (for example, phage therapy that evolves with the microbes).
  - Rapid design and synthesis of customized, targeted therapeutics (including endonucleases, lysins, endopeptidases, and proteases) for inhibiting pathogenic cell growth.
- Host and Consortia Engineering Achievement:
  - Engineer a more diverse gut microbiome to prevent potential pathogenicity and increase resistance to gastrointestinal tract infections.
  - Engineer organisms that can be used to seed the gut microbiome for creating *in situ* antibiotic products.
  - Design of cellular features to support successful, non-toxic delivery and stabilization of living therapies in the patient.
- Data Integration, Modeling, and Automation Achievement:
  - Improve prediction of evolution of novel antimicrobial resistance-conferring mutations.
  - Improve design and prediction of targeted therapeutics.
  - Develop methods for optimization of treatment strategies that stop or prevent the evolution, emergence, and/or dominance of resistant subpopulations of bacteria.

- Engineering Biology Objective 3: Reduce transmission of disease to humans from non-human animals.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - New tools for editing genes and pathways in insects and livestock that act as disease carriers and reservoirs.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Development of additional, ultra-low-cost animal vaccines, and new vaccines for diseases not currently covered.
  - Host and Consortia Engineering Achievement:
    - Engineer cells of insects and animals that act as disease carriers and reservoirs to attenuate pathogenicity and/or neutralize the pathogen (Lane & Quistad, 1998).
  - Data Integration, Modeling, and Automation Achievement:
    - Develop better models to predict emergence and evolution of antibiotic resistance under complex scenarios.
- Engineering Biology Objective 4: Genetically encode disease resistance (such as in livestock).
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Improve tools for genetic manipulation of animals.
  - Host and Consortia Engineering Achievement:
    - Engineer microbiomes to resist disease, such as through secretion of antimicrobial substances *in situ*.

- Engineer somatic cells for disease resistance; for example, by altering membrane components known to be points of attachment for certain pathogens, by enhancing immune memory to specific pathogens post-vaccine, or engineering chimeric antigen receptor (CAR) T cells for activity against fungal and other pathogens (Naran, Nundalall, Chetty, & Barth, 2018).

- Data Integration, Modeling, and Automation Achievement:
  - Computational identification of genes that confer disease resistance.

- Science/Engineering Aim 2: Diagnose and treat viral infections.
  - Engineering Biology Objective 1: Develop rapid, reliable diagnostics to detect viral infections.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Rapid, high-fidelity DNA synthesis for development and production of sensing technologies.
      - Gene editing technologies for building cell-based sensors.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Sensitive and specific molecular sensing technologies, such as molecular probes or DNA amplification.
    - Host and Consortia Engineering Achievement:
      - Cell-expressed reporters for infection.
      - Synthetic epigenetic silencing of viral DNA.
    - Data Integration, Modeling, and Automation Achievement:
      - Point-of-care (POC) and/or microfluidic systems for automating patient sample preparation.
      - Cheap and fast sequencing of viral infection markers and patient susceptibility biomarkers.
  - Engineering Biology Objective 2: Develop tools to treat viral infections.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - New gene editing tools to precisely neutralize or excise viral sequences from host genomes.
      - Scaled-up synthesis of high quality DNA encoding anti-viral gene circuits.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Rapid design and synthesis of customized, targeted endonucleases for inhibiting viral replication.
    - Host and Consortia Engineering Achievement:
      - Synthetic epigenetic silencing of viral DNA.
      - Development of cell and tissue models to screen and test anti-viral interventions *in situ*.
    - Data Integration, Modeling, and Automation Achievement:
      - Systems biology approaches to identifying critical molecular weaknesses in viral function as drug targets.

        - Simulations to predict pace and breadth of epidemics and impact of molecular interventions.
- Science/Engineering Aim 3: Develop new and better vaccines, other prophylactic tools, and production pipelines.
    - Engineering Biology Objective 1: Design antigens and adjuvants that improve immune memory.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Rapidly produce antigen variants for validation.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - *De novo* development of synthetic immunogenic antigens.
        - Host and Consortia Engineering Achievement:
            - Better understanding of the heterogeneity of immune memory longevity for different pathogens and different individuals.
            - Ability to increase longevity of specific memory T and B cells.
        - Data Integration, Modeling, and Automation Achievement:
            - Modeling and prediction of the correlation between adjuvants and immune memory.
            - Modeling and prediction of how immune memory formation varies between individuals (with the inclusion of characteristics such as race, ethnicity, geography, and socioeconomic status) for different pathogens.
    - Engineering Biology Objective 2: Develop nucleic acid- and other biomolecule-based vaccines (including hybrid biologic/polymer-based vaccines).
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Automated, large-scale, combinatorial DNA assembly and screening to identify nucleic acid construct designs enabling robust antigen expression.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Low-cost nucleic acid synthesis.
            - Increase antigen expression through gene expression engineering.
            - Enable nucleic acid systems for robust expression of multiple antigens.
            - Improve vector design and delivery methods.
        - Host and Consortia Engineering Achievement:
            - Engineer cells to produce low-levels of antigen to promote longevity of memory immune responses *in vivo*, while minimizing host immune response against the engineered cells (Kedzierska, Valkenburg, Doherty, Davenport, & Venturi, 2012).
        - Data Integration, Modeling, and Automation Achievement:
            - Use analytics and modeling to identify transcriptional and translational regulatory elements enabling enhanced protein expression.

        - Deep-learning techniques for understanding “sequence grammar” of regulatory elements.
    - Engineering Biology Objective 3: Enable and advance the use of plants, cell cultures, and cell-free systems to produce vaccines.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Increase DNA synthesis and fidelity to build and characterize promoters, circuits, and pathways for antigen production.
            - Scaled up synthesis of DNA-based antigens.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop gene expression tools to increase vaccine (antigen) yield in non-animal expression systems.
        - Host and Consortia Engineering Achievement:
            - Engineer host cells and cell-free systems for high fidelity production of vaccines.
        - Data Integration, Modeling, and Automation Achievement:
            - Employ modeling and analytical approaches to identify critical factors affecting vaccine production and quality.
- Science/Engineering Aim 4: Develop better population-scale surveillance methods for emerging infectious diseases and create technologies to rapidly address outbreaks and epidemics in real time.
    - Engineering Biology Objective 1: Advance engineering of biological tools to detect and track pathogen reservoirs and flow over time and space.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - DNA-based event recording (DNA barcoding).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Development of simple and cheap, biomolecule-based kits for surveillance and analysis.
        - Host and Consortia Engineering Achievement:
            - Next-generation live cell reporting systems.
        - Data Integration, Modeling, and Automation Achievement:
            - Advanced models for pathogen flow through the environment and populations in real time.
    - Engineering Biology Objective 2: Develop tools to rapidly characterize and respond to known and unknown pathogens in real time at population scales.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Diagnostics for nucleic acids indicative of the presence of specific pathogens, utilizing targeted DNA- and RNA-binding Cas editors.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Development of strain-specific vaccines in real time (i.e., during an outbreak).
        - Host and Consortia Engineering Achievement:
            - Engineer microbes that detect pathogenic antigens and react by secreting anti-pathogen factors.

        - Data Integration, Modeling, and Automation Achievement:
            - Advanced bioinformatics to quickly characterize emerging pathogens from genetic sequences and epigenetic markers.

Societal Challenge 2: Address non-communicable diseases and disorders. Regarding non-communicable diseases and disorders, we consider the advancement engineering biology tools and technologies to address cancer, addiction, obesity, neurodegenerative diseases, aging-related disorders, psychiatric disorders, heart disease, diabetes, and other genetic disorders and lifestyle diseases.

- Science/Engineering Aim 1: Measure molecular markers of disease.
    - Engineering Biology Objective 1: Develop biosensors for measuring metabolites, proteins, and other biomolecules *in vivo*.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - High-fidelity production of complex large, functional DNAs and RNAs (such as aptamers and riboswitches).
            - Efficient production of >1 kilobase biosensor genes and circuits that may contain repeat elements for sensing multiple input signals.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Identify biosensors for molecules for which there are currently no biosensors.
            - Identify fast and reliable biosensor readouts for *in vivo* applications.
            - Engineer memory circuits to record the presence of metabolites and proteins and the intensity and duration of those signals.
        - Host and Consortia Engineering:
            - Enable selective transfection/transduction and delivery of large biosensor sequences and circuits into host cells.
        - Data Integration, Modeling, and Automation Achievement:
            - Genome mining for biosensors.
            - Identify design principles to incorporate these biosensors into different hosts.
            - Leverage machine learning technologies to facilitate deconvolution and identification of biosensor signals.
- Science/Engineering Aim 2: Generate new drug therapies.
    - Engineering Biology Objective 1: Develop platforms for rapidly and effectively identifying drugs to treat non-infectious diseases.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Rapid and cost efficient synthesis of genetic circuits.
            - Parallel and error-free genome engineering of mammalian cell lines.

        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Development of novel, high-affinity agents, such as antibody proteins, nucleic acids, and other macromolecules, that bind drug targets.
            - Macromolecular adducts (chemical “tags”) to control the distribution and delivery of drugs within cells, tissues, and organs.
            - Novel modulators of cell pathways that show little or no off-target toxicity.
        - Host and Consortia Engineering Achievement:
            - Genetically-encoded reporters for real-time tracking of drug activity in cells, tissues, and microbiomes.
            - Microbial reporters to detect gastrointestinal tract cell stress signals.
        - Data Integration, Modeling, and Automation Achievement:
            - Develop automated, large-scale screening platforms for drug discovery.
            - Powerful associative analyses to link gene and protein networks to disease states.
            - Models to predict biased accumulation of drug in certain tissues based on the chemical and/or physical properties of the drug.
            - Drug-to-disease database and associated software/informatics tools to rapidly evaluate potential for drug repurposing.
    - Engineering Biology Objective 2: Identify patient-specific drugs.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Identify patient-specific genetic biomarkers.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Artificial co-evolution of macromolecular therapeutics which change in response to a patient’s unique biochemistry.
            - Anticipatory library of therapeutic variants that contain best-matches for patient-specific drug target variants.
        - Host and Consortia Engineering Achievement:
            - Development of patient-matched disease models (such as organoids).
        - Data Integration, Modeling, and Automation Achievement:
            - Develop libraries of drug efficacy correlated to de-identified patient biomarkers, used to identify promising patient-specific drugs.
            - Use modeling and bioinformatics to predict novel interventions on an individualized basis.
- Science/Engineering Aim 3: Develop and hone genetic engineering/gene therapies.
    - Engineering Biology Objective 1: Develop targeted delivery of gene therapies to specific tissues and cells.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Rapid and cost-efficient synthesis of genetic circuits.

        - Efficient DNA editing in mitochondria.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Non-immunogenic macromolecules and vesicles to deliver therapeutic DNA, RNA, and proteins to cells and tissues.
    - Host and Consortia Engineering Achievement:
        - Increase the payload size for DNA delivery vectors by at least ten-fold.
        - Produce optimal epigenetic imprinting patterns in induced pluripotent stem cells (iPSC) and artificially-differentiated cells.
    - Data Integration, Modeling, and Automation Achievement:
        - Generate models to predict efficiency of DNA/RNA delivery based on the structure of the payload and features of the target cell or tissue.
- Engineering Biology Objective 2: Regulate, control, and maintain gene therapies.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Rapid and cost efficient synthesis of genetic circuits.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Circuits that enable temporal control of gene therapy localization and activation.
    - Host and Consortia Engineering Achievement:
        - Prevent immune system from reacting to or eliminating gene therapy.
    - Data Integration, Modeling, and Automation Achievement:
        - Develop predictive models to determine optimal maintenance/scheduling of gene therapies.
        - Automation to rapidly design, build, and test circuit designs in mammalian cells.

- Science/Engineering Aim 4: Advance engineered cell systems (including the human microbiome and immune system), organs, and tissues to manage and treat disease and disease outcomes.
    - Engineering Biology Objective 1: Characterize, engineer, and manipulate different microbiota throughout the body for health purposes.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Re-code microbial genomes/chromosomes.
            - Targeted gene editing systems for specific microbes or cell types.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer enzymes to enhance or alter metabolism.
            - Engineer secretion systems for *in vivo* delivery of therapeutics from microbes.
        - Host and Consortia Engineering Achievement:
            - Rationally design and engineer microbial cells and communities.

        - Achieve short- and long-term, predictable tuning of the microbiome to deliver therapeutics, add functions and enzymes, and remove organisms.
    - Data Integration, Modeling, and Automation Achievement:
        - Advanced modeling of interactions between microbes within the microbiota and their host.
        - Ecological models that incorporate changes in host, microbes, and the local environment (more specifically, the location in gastrointestinal tract, in the skin, etc.), and enable prediction of therapeutic approach.
        - Develop models that focus on function (enzymes, pathways) to diagnose and predict dysbiosis.
        - Employ statistically rigorous models to differentiate correlation and causation with respect to changes in the microbiome, as correlations are still valuable for diagnostics but therapies and interventions should be focused where there is a causative or clearly functional link.
- Engineering Biology Objective 2: Create cell-autonomous genetic circuits to drive tissue formation and repair.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Achieve stable expression from synthetic transgenes.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Engineer macromolecules with predictable, robust, orthogonal dynamic behavior that demonstrate no unintended cross-interaction with other factors.
        - Engineer libraries of synthetic, orthogonal cell-communication mechanisms, including short-range communication (receptors) and long-range communication (morphogens).
    - Host and Consortia Engineering Achievement:
        - Engineer mechanisms to coordinate behavior of single cells in a population and interaction with the host (i.e., patient).
        - Customize the function and number of major cellular features, including cell surface proteins, the cytoskeleton, organelles, and chromosomes.
    - Data Integration, Modeling, and Automation Achievement:
        - Rapid single-cell -omics pipelines to understand the molecular and cellular recipes in development and tissue formation.
- Engineering Biology Objective 3: Engineer immune cell-based therapies.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Improve parallel and precise genome editing in primary immune cells.

- Biomolecule, Pathway, and Circuit Engineering Achievement:
  - Improve biosensor and genetic circuit designs to improve specificity, efficacy, and safety.
- Host and Consortia Engineering Achievement:
  - Engineer mechanisms to coordinate behavior of single cells and their interaction with the human host.
- Data Integration, Modeling, and Automation Achievement:
  - Increase the reliability of predicting protein, pathway, and circuit function from sequences to enable better biosensor, receptor, and genetic circuit designs.

- Engineering Biology Objective 4: Enable biocompatible allo- and xeno-transplant and implantation of synthetic or engineered (including "printed") tissues/organs.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Improve parallel and precise genome editing in recipient's immune system to establish or increase tolerance to the donor tissue/organs and immunize against cross-species disease transmission.
    - Improve parallel and precise genome editing in donor animals to reduce or remove immunogenicity and cross-species disease transmission.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Develop biosensors for identifying xenoreactive immune cells.
    - Engineer libraries of synthetic, orthogonal cell-communication mechanisms, including short-range communication (receptors) and long-range communication (morphogens).
    - Enable production of synthetic and engineered bioscaffolds for tissue regeneration.
  - Host and Consortia Engineering Achievement:
    - Engineer the recipient's immune system to be specifically tolerant of the implant without excessive immune suppression.
  - Data Integration, Modeling, and Automation Achievement:
    - Advanced modeling of interactions between implant/transplant and the host.
    - Rapid single-cell -omics pipelines to understand the molecular and cellular characteristics of development and tissue formation.

Societal Challenge 3: Address environmental threats to health, including toxins, pollution, accidents, radiation, exposure, and injury.

- Science/Engineering Aim 1: Integrate (bio)materials and living tissues to address injuries and navigate dangerous environments.
    - Engineering Biology Objective 1: Enable greater and more beneficial interaction of living cells and tissues with prosthetics.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Parallel, scalable, and cost-effective genome engineering to enable the use of allogeneic cell sources, as opposed to patient-specific sources.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop and rapidly produce biomolecule-based materials (biomaterials) that have improved physiological properties.
            - Develop biopolymers with physical durability to resist long-term wear and tear.
            - Achieve minimally-invasive control of synthetic gene and protein networks with light-programmable macromolecules (advanced optogenetics).
        - Host and Consortia Engineering Achievement:
            - Engineer cellular pathways, extracellular matrices, and connective tissues that enhance prosthetic compatibility without compromising health.
        - Data Integration, Modeling, and Automation Achievement:
            - Identify predictive, detectable, micro-scale biosignatures (biological outputs) that correlate with health, damage, or disease.
    - Engineering Biology Objective 2: Integrate wearable tech with living cells to sense and act upon threats to health.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Develop systems for reliable genomic integration of reporters that will sense specific cell states in high-risk populations, where the molecules/states can be sensed, analyzed, and acted upon externally (electronic or optic signaling).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop sensing and reporting systems that enable *in situ* detection of toxins and disease indicators, where the info can be sensed, analyzed, and acted upon externally (electronic or optic signaling) or in a more integrated fashion.
        - Host Engineering Achievement:
            - Develop probiotics and similar cell systems that can report to external devices.
            - Tune select cells or tissues to interact with stimuli from external (electronic) devices in a highly controlled manner.

        - Data Integration, Modeling, and Automation Achievement:
            - Develop and advance modeling and analytics to integrate information from wearable tech, medical sensors (like those for continuous glucose monitoring), and eventually *in vivo* sensors, to predict health, physical performance, toxin exposure, disease, other states of interest.
            - Use novel machine learning approaches to integrate different types of sensor data and address variation between people and populations.
            - Design and model systems that both sense and act upon threats, with reliable communication and data integration.
            - Expand and improve algorithms for estimating health states based on a limited set of measurable data.
    - Engineering Biology Objective 3: Engineer the immune system to improve allotransplant of tissues, organs, and limbs.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Achieve highly efficient, rapid genetic or epigenetic editing of the allograft genome with synthetic gene cassettes or whole chromosomes.
            - Achieve efficient co-editing of human leukocyte antigen (HLA) gene clusters to prevent allograft rejection.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Generate potent gene delivery vehicles for immune gene clusters (such as HLAs) and whole synthetic chromosomes.
            - Develop macromolecules to neutralize or mask non-self protein markers.
        - Host and Consortia Engineering Achievement:
            - Remove potent non-self antigens from allograft tissues/organs.
            - Replace non-self with “self” markers in allograft cells.
        - Data Integration, Modeling, and Automation Achievement:
            - Achieve data-driven molecular profiling of key antigens to identify engineerable donor tissue and support patient-to-allograft matching.
- Science/Engineering Aim 2: Develop systems to detect, identify, reverse, neutralize, and clear biochemical damage.
    - Engineering Biology Objective 1: Prevent, reverse, or neutralize microlesions induced by toxins, radiation, and other factors.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - High-fidelity production and delivery of DNA and non-coding RNAs to aid DNA damage repair.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Design robust, mutation-specific, base-editors.

        - Deliver synthetic repair machinery into the nucleus and mitochondria.
        - Design lipids and cell surface features to reverse cell membrane damage.
    - Host and Consortia Engineering Achievement:
        - Induce prophylactic genetic and epigenetic states in somatic cells prior to exposure (an interesting example of this might be conditioning astronauts for space exploration).
    - Data Integration, Modeling, and Automation Achievement:
        - Use data analytics and modeling to predict microlesion weak-spots (e.g., DNA, RNA, protein hotspots) to support anticipatory medical care.
- Engineering Biology Objective 2: Neutralization and clearance of toxic substances from the body.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Achieve highly-efficient, rapid gene editing to enable cells to detect and neutralize threats as needed.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Develop biomolecular reporters to track migration and accumulation of toxins through the body.
        - Generate macromolecules that neutralize and clear prions and protein plaques from the body.
        - Design high-affinity molecules to bind and clear toxins from the body (an interesting example of this might be a synthetic antivenom).
    - Host and Consortia Engineering Achievement:
        - Generate and engineer hosts and cell-free systems that can act as bio-factories to produce anti-toxins at practical scales.
    - Data Integration, Modeling, and Automation Achievement:
        - Models to predict symptoms, onset, and timing of poisoning, to inform the rational design of antidotes and treatment regimes.

Societal Challenge 4: Promote equitable access to healthcare, patient representation in research, democratization of medicine, and the development of personalized medicines.

- Science/Engineering Aim 1: Develop patient-specific testbeds for drug treatments to support patient representation and personalized medicine.
    - Engineering Biology Objective 1: Develop induced pluripotent stem cell (iPSC)-derived organoids as personalized models.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Deliver and stabilize the expression of synthetic DNA in iPSCs.
            - Use nucleases to efficiently edit very small numbers of cells with minimal error.

        - Identify and characterize differences in gene expression profiles between human primary tissues and iPSC-derived tissues.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Develop nanocarriers to efficiently deliver macromolecules into small numbers of cells.
    - Host and Consortia Engineering Achievement:
        - Develop patient-specific organ-on-a-chip devices to model individual patient response to drug treatments across organ systems.
        - Develop minimally-invasive methods to collect and culture useful cells.
    - Data Integration, Modeling, and Automation Achievement:
        - Establish databases of genetic and metabolic expression and activity profiles of iPSCs and iPSC-derived tissues.
- Engineering Biology Objective 2: Personalize medical treatments to human subpopulations and/or individuals. (Dehingia, Adak, & Khan, 2019; Hooker et al., 2019; Molteni, 2019)
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Develop fast, high-fidelity, on-demand synthesis of large fragments of customized DNA and RNA for clinical use.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Develop macromolecules and gene circuits that sense and report the local cellular or tissue environment.
    - Host and Consortia Engineering Achievement:
        - Develop living therapeutic cells that switch phenotypes in response to the local tissue environment.
        - Develop allergen-free platforms and cells for drug production.
    - Data Integration, Modeling, and Automation Achievement:
        - Model variants and alleles present in the human population to better understand and mitigate health challenges.
        - Identify useful semi-generalizable (familial or population-wide) features to help accelerate diagnoses and the design of treatment regimes.

- Science/Engineering Aim 2: Make cutting-edge therapy more available and affordable.
    - Engineering Biology Objective 1: Scale-up hard-to-produce therapeutic molecules, proteins, and cell therapies.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Parallel, scalable, and cost-effective genome engineering to enable the use of allogeneic cell sources, as opposed to patient-specific sources.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer temperature-stable, active macromolecule- and cell-based therapies.

- Design molecular features to support inexpensive, robust purification and processing of therapeutics.
- Develop circuits that behave robustly across different growth media.

- Host and Consortia Engineering Achievement:
  - Develop and characterize select microbes for advanced-therapeutics production in stationary phase.
- Data Integration, Modeling, and Automation Achievement:
  - Create publicly-accessible and encrypted databases of health-related data.
  - Identify critical metabolic/molecular bottlenecks and work-arounds for hard-to-produce therapeutics.

Food & Agriculture

Food & Agriculture

Food & Agriculture focuses on the tools and technologies impacting how we feed the Earth's people and animals. Engineering biology provides unique means and opportunities to support growing populations with more and different types of food, address changes to food security, diet, and demand, and reduce the impact of climate change and urban growth in a sustainable manner. Increasing challenges from biotic and abiotic stresses also significantly impact agricultural productivity and health. Concepts include increasing yield and sustainable productivity, while reducing the consumption of resources, including land, water, fertilizers, and pesticides. Achieving these Aims and Objectives will help to meet increased demands for nutrient-rich foods and healthy food animals.

Societal Challenge 1: Produce more food for a growing global population.

- Science/Engineering Aim 1: Improve agricultural yields by increasing crop efficiency and production.
    - Engineering Biology Objective 1: Improve photosynthesis efficiency in crops and other food plants.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Edit genes in the photosynthetic pathway for improved properties, including stability, catalytic activity, and substrate specificity.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improved efficiency of key enzyme(s) in the photosynthesis pathway.
        - Host and Consortia Engineering Achievement:
            - Introduce synthetic (heterologous or modified) enzymes/complexes/pathways to improve photosynthetic efficiency.
        - Data Integration, Modeling, and Automation Achievement:
            - Models for engineered photosynthesis pathway.
    - Engineering Biology Objective 2: Improve soil nutrients, water, and CO_2 use efficiency in crops and other food plants.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Gene and genome editing to modulate expression and properties of key proteins involved in transport, storage, mobilization, and usage.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve key protein properties in important pathways or associated regulatory/transcription factors that are directly or indirectly involved in transport, storage, mobilization, and usage.
        - Host and Consortia Engineering Achievement:
            - Increase or re-engineer microbiome and symbiotic interactions to introduce or enhance nitrogen fixation and nutrient mobilization.

- Science/Engineering Aim 2: Increase the availability and consistency of agricultural crop production by combating stressors and expanding consumable species.
    - Engineering Biology Objective 1: Maintain crop yield under abiotic stress.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Minimize change in crop development times expected under higher temperature by regulating plant hormones and developmental gene expression.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Introduce C4 pathways introduced into C3 organisms to maintain photosynthetic capacity with stress from temperature and temperature variation.
            - Improve efficiency of photosynthetic pathway components.
            - Reduce heat- and water-stress-response during critical reproductive periods, such as fruit and seed formation.
        - Host and Consortia Engineering Achievement:
            - Chloroplast engineering to improve/stabilize photosynthetic pathways.
            - Reduce transpiration rates due to increased temperatures.
        - Data Integration, Modeling, and Automation Achievement:
            - Modeling of crop response to complex environmental changes, including global climate change.
    - Engineering Biology Objective 2: Maintain crop yield under biotic stress.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Identify native insect and disease resistance traits (e.g., R genes) from non-crop species, or from non-domesticated “crop” species, and introduce to crop species.
            - Engineer durable resistance in crops from computationally-designed proteins modeled after natural resistances and improving therein.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer novel insecticidal proteins with different mode-of-action or spectrum-of-control beyond those in use today.
        - Host and Consortia Engineering Achievement:
            - Introduce metabolic or signalling pathways that improve or reinforce plant defense response to insect or disease pressure.
    - Engineering Biology Objective 3: Accelerate domestication of wild plant species.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Insert desirable traits (encoded by multiple genes or processes) from wild species into modern crops, or from modern crops into edible wild species, to improve plant products such as fruit size and yield, while maintaining genetic diversity and better protection from biotic/abiotic stresses.

        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve expression and activity of “domestication” genes to (re)introduce into wild species to generate new crops, increasing the number and variety of regional sources for calories and protein.
        - Host and Consortia Engineering Achievement:
            - Add new pathways to improve oils, proteins, and vitamin sources in “alternative” or newly domesticated crop species.
- Science/Engineering Aim 3: Improve the production and yield of meat from livestock and fish.
    - Engineering Biology Objective 1: Increase food animal yield, such as increased litter size, faster reproduction, and faster development.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Modify animal/fish genomes to enhance production of growth hormones.
            - Delete genes/proteins that may restrict muscle mass (e.g., myostatin) which can also reduce fat content.
            - Modify fecundity genes to increase litter size.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer synthetic proteins that promote growth hormone production.
        - Host and Consortia Engineering Achievement:
            - Introduce heterologous gene cassettes (inducible promoter-gene-terminator) into genome that encodes synthetic protein promoting growth hormone production.
        - Data Integration, Modeling, and Automation Achievement:
            - Model growth rate vs. hormone production/induction.
    - Engineering Biology Objective 2: Reduce infectious and non-infectious disease in food animals, including alternatives to antibiotics.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Modify animal genomes to auto-induce vaccination to common pathogens.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop synthetic peptides or proteins that protect animals from common viral and bacterial diseases.
        - Host and Consortia Engineering Achievement:
            - Introduce genes encoding synthetic peptides for disease resistance into genome of food animals to protect against common viral and bacterial disease.
        - Data Integration, Modeling, and Automation Achievement:
            - Improve modeling of disease transmission to predict optimal husbandry conditions for preventing disease.

    - Engineering Biology Objective 3: Engineer the rumen/livestock microbiome to improve digestion of feed for improved nutrient absorption and metabolism.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Introduce genes to increase or develop tunable populations of beneficial gut microbiota.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop enzymes that are efficient at breaking down complex carbohydrates at pH of cow rumen.
        - Host and Consortia Engineering Achievement:
            - Introduce novel or synthetic digestive enzymes/pathways to increase digestible fiber into most abundant microbiome species that aid in digestion or eliminate/inactivate anti-nutritive substances.
        - Data Integration, Modeling, and Automation Achievement:
            - Survey microbiome populations in cow rumen and identify most abundant species for manipulation.
            - Model microbiome establishment and interactions.
- Science/Engineering Aim 4: Enable and advance the production and availability of non-vertebrate animal food sources.
    - Engineering Biology Objective 1: Increase non-vertebrate yield, including faster reproduction and development.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Remove genes for non-essential functions from the perspective of engineering organisms as food (such as shell production, wing development).
            - Silence genes involved in dormancy or molting to increase development and reproductive cycle-time.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop synthetic anti-molting peptide that normally limits growth and reproductive cycle-time, to include in feed.
        - Host and Consortia Engineering Achievement:
            - Develop inducible system that increases the number of molts/year (especially at adult stage) to speed up development time and reach reproductive age more quickly.
            - Develop a system that allows crustacea to continuously molt by overexpressing steroid hormones.
    - Engineering Biology Objective 2: Increase non-vertebrate biomass.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Change or modulate hormonal pathways that limit growth by changing expression of pathway genes.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Design enzymes that degrade hormones to prevent development from larvae to adults (for insects food sources).

        - Host and Consortia Engineering Achievement:
            - Inducible system that allows tunable expression of proteins or full pathway to maximize growth rate or biomass accumulation.
            - Develop a system that allows crustacea to continuously molt by overexpressing steroid hormones.
- Science/Engineering Aim 5: Improve production of "clean meat". (Stephens et al., 2018)
    - Engineering Biology Objective 1: Increase the diversity, availability, and optimization of characterized and standardized cell lines used for clean meat.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Create cell line libraries (to include stem cells, myoblasts, etc.) with desired phenotypes, including taste, texture, and aroma.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Develop biosensors/reporters for cell lines that provide readouts for suboptimal performance or culture conditions to accelerate research and development.
            - Engineer synthetic pathways to render differentiation responsive to inexpensive triggers (e.g., a unique sugar, rather than a complex growth factor cocktail).
        - Host and Consortia Engineering Achievement:
            - Increase proliferation/division rate of cells to improve biomass accumulation rate.
            - Increase genetic stability of cells to maintain genetic integrity over generations.
            - Select for or engineer cells to exhibit higher propensity to differentiate down desired pathways (muscle, fat, etc.) and low propensity to pursue undesirable pathways (bone, tendon, etc.).
        - Data Integration, Modeling, and Automation Achievement:
            - Create database of cell line characteristics, including genomic, proteomic, and metabolomic data.
            - Establish shared dataset of metabolic parameters from a wide variety of cell lines under various growth conditions to facilitate systems biology approach to metabolic pathway engineering and modeling.
    - Engineering Biology Objective 2: Engineer renewable and alternative growth media to support cell growth and health.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Improve expression or specific activity of proteins in growth pathways in native organisms.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Maximize production of growth and maturation factors while minimizing impact on growth of hosts, i.e., on/off expression system development.

- Create multiple genes encoding various biosynthetic enzymes or constructing novel pathways to produce growth factors.
- Use rational design or directed evolution to engineer growth factor variants/mimics that are, for example, more stable, more potent, have higher binding affinity.
- Utilize small molecule screens or natural product screens to identify entirely new growth factor mimics.

■ Host and Consortia Engineering Achievement:
- Engineer yeast or other cost-effective hosts to produce growth factors and other small molecules useful in cell culture media.

■ Data Integration, Modeling, and Automation Achievement:
- Model maximum theoretical yield obtainable for growth factors produced in a variety of cost-effective hosts to enable educated host choices.
- Develop machine learning-informed algorithms for more sophisticated Design-of-Experiments to expand the explorable space for medium formulations with many variables.
- Incorporate empirical information from highly-parallelized microfluidics platforms to assess cell performance in novel formulations.
- Merge insights from spent media analysis with systems biology modeling of metabolic pathways to understand how to bias metabolism toward biomass and protein accumulation.

○ Engineering Biology Objective 3: Improve bioscaffolding materials and perfusion systems to grow larger, more complex, tastier, and thicker pieces of “meat”.

■ Biomolecule, Pathway, and Circuit Engineering Achievement:
- Create biodegradable and/or edible scaffold biomaterials from biomolecules.
- Evaluate a wide variety of native and modified biopolymers from the plant, fungal, and bacterial kingdoms for suitability to create tunable hydrogel scaffolds, specifically for properties like photoresponsiveness (photopolymerization, light-induced degradation, etc.) for fabricating more spatially defined scaffolds.
- Enzyme screening and engineering to make specific modifications to plant- or fungal-derived biopolymer scaffolds (for example, modified cellulose).

■ Host and Consortia Engineering Achievement:
- Engineer plants or other low-cost biomass platforms (such as fungal platforms) to express peptides that make scaffolds derived from them more amenable to animal cell attachment.
- Examine effects of growth conditions and strain selection on producing scaffolds with desirable properties from fungal hosts (such as mycelium or secreted proteins).

        - Engineer animal cells to produce enzymes and/or attachment molecules that enable a wider variety of scaffold materials or scaffold remodeling *in situ* to more closely mimic native extracellular matrix-cell interaction.
    - Data Integration, Modeling, and Automation Achievement:
        - Produce fluid dynamic models for *in silico* prediction of appropriate scaffold architecture, culture medium viscosity and flow rates, required nutrient and dissolved oxygen concentrations, among other properties for supporting thick tissue perfusion.
        - Develop empirically validated scaling factors for facilitating upscale from bench to production in tissue perfusion bioreactors.
- Science/Engineering Aim 6: Advance the quality of plant-based meat products and improve large-scale manufacturing capabilities.
    - Engineering Biology Objective 1: Diversify raw material supply chains and improve sensory performance ingredients for incorporation into plant-based meat.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Alter high-protein crops for even higher protein accumulation and/or bias toward accumulation of high-performing storage proteins for plant-based meat applications.
            - Alter lipid synthesis pathways of common oilseed crops to produce higher levels of high-value lipids that are traditionally scarce in plant sources (including saturated fats, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA)).
            - Edit crop plants for lower levels of secondary metabolites that negatively impact taste (such as saponins and other components of bitter or beany off-flavors).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Identify new methods for storage and/or scavenging to improve shelf life of less-highly-refined fractions (for example, to slow oxidation and rancidity in plant flours with residual oil).
            - Identify plant-sourced homologs or synthetic mimics of proteins or small molecules determined to play critical roles in the taste of various types of meat.
        - Host and Consortia Engineering Achievement:
            - Perform thorough genetic, phenotypic, and compositional characterization of underutilized crops to identify promising candidates for novel sources of proteins, flavorings, or other functional ingredients.
            - Engineer promising specialty crop candidates for plant protein sources to achieve the same agronomic yield gains, robustness to pests, and abiotic stress, that are already standard in commodity crops.

- Explore capability of microbial fermentation to improve sensory and functional properties of plant proteins and other raw materials through, for example, enzyme secretion and selective metabolism of undesirable components.
- Engineer high-efficiency microbial production hosts to produce functional proteins or flavoring ingredients through synthetic biology and novel metabolic pathway introduction.

■ Data Integration, Modeling, and Automation Achievement:

- Capture and store agronomic data along with raw material characterization data to enhance predictive capacity for how growing conditions, soil, weather, etc., affect the end product.
- Develop better analytical tools for assessing/predicting functionality or performance of complex fractions of plant ingredients.
- Identify compositional signatures (for example, through mass spectrometry or capillary electrophoresis) that can be used as analytical tools to characterize nuanced performance characteristics of lots of raw materials.

○ Engineering Biology Objective 2: Improve large-scale manufacturing and texturization of plant-based meat.

■ Gene Editing, Synthesis, and Assembly Achievement:

- Edit crop plants for higher efficiency of fractionation into protein, starch, fiber, and oil components.

■ Biomolecule, Pathway, and Circuit Engineering Achievement:

- Develop new food processing enzymes (rationally designed or through directed evolution) that are optimized for plant-based substrates to improve raw ingredient functionalization (for example, hydrolysis or cross-linking of plant proteins to improve solubility, water-binding capacity, and gelling).
- Develop methods for microencapsulating ingredients like flavors and fats in edible ingredients such that they remain protected and stable during high-pressure processing.

■ Host and Consortia Engineering Achievement:

- Engineer crop plants for modified proteins that have natively improved functional properties for plant-based meat processing (such as solubility, cross-linking capability and fat-binding capacity).

■ Data Integration, Modeling, and Automation Achievement:

- Develop mechanistic models of plant protein denaturation, alignment, and crystallization within the context of high-shear processing methods like extrusion, to inform the process variables for a given composition of input materials.

- Science/Engineering Aim 7: Engineer microorganisms for nutrient production.
    - Engineering Biology Objective 1: Engineer hosts to produce safer food components (e.g., processing aids) and ingredients (e.g., vitamins).
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Modify vitamin biosynthetic pathways in native production hosts to increase yield or stability of provitamins.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve specificity and properties of enzymes involved in (pro)vitamin biosynthesis to increase yield.
        - Host and Consortia Engineering Achievement:
            - Introduce and optimize (pro)vitamin biosynthesis pathways from different sources into yeast, microalgae, and bacteria.
            - Introduce transporters to help increase production rates.
            - Reduce metabolic regulation and controls to increase production of (pro)vitamins.
        - Data Integration, Modeling, and Automation Achievement:
            - Computational tools for enzyme design and engineering.

Societal Challenge 2: Increase and improve the nutritional content and value of food.

- Science/Engineering Aim 1: Increase the nutrient content in agricultural crops.
    - Engineering Biology Objective 1: Improve the quantity and quality of seed proteins and proteins from vegetative plant tissues.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Increase expression of proteins with desired amino acid content by introducing engineered genetic parts (including promoters and peptide tags).
            - Modulate seed-specific pathway genes to enhance protein accumulation in seed without impacting germination.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Design novel proteins that incorporate important amino acids that can be expressed in seed or improve protein solubility.
            - *De novo*, model-based creation of proteins with high nutritional value.
        - Host and Consortia Engineering Achievement:
            - Introduce proteins of high nutritional value into plant tissue.
        - Data Integration, Modeling, and Automation Achievement:
            - Model protein and oil flux in the seed to ensure optimal accumulation.
            - Produce better models of important proteins to understand where protein quality can be improved.

- Engineering Biology Objective 2: Increase bioavailability of proteins through engineering structure, amino acid composition, and removing anti-nutritive factors.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Edit plant protein sequences to increase digestibility through changes to structure and amino acid content or reduce expression of anti-nutritive factors.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Design new proteins to increase content of rare or needed amino acids and/or enhanced solubility.
    - Eliminate aspects of pathways that produce anti-nutritive factors that do not impact plant growth.
  - Host and Consortia Engineering Achievement:
    - Introduce or modulate pathways and feedback loops that play a role in protein degradation and amino acid scavenging.
- Engineering Biology Objective 3: Enable (increased) production of beneficial fatty acids and digestible fiber in staple crops.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Knockout genes/pathways that divert carbon away from fatty acid (FA) production in plants.
    - Reduce lignin biosynthesis in crops.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Improve enzymatic properties of proteins involved in conversion of FAs to long-chain polyunsaturated fatty acids (LC-PUFAs) to increase production and accumulation in plants.
    - Develop more effective enzymes that can reduce lignin content in crops.
  - Host and Consortia Engineering Achievement:
    - Introduce pathways or novel enzymes from heterologous systems that are missing in plants to convert FAs into omega-3 or omega-6 LC-PUFAs.
    - Integrate and/or replace poorly perfoming or less desireable pathway enzymes.
    - Overexpress lignin-degradation enzyme(s) under an inducible promoter.

- Science/Engineering Aim 2: Improve the healthiness of agricultural crops by enabling the reduction and elimination of toxins.
  - Engineering Biology Objective 1: Engineer and improve crops and other agricultural plants to prevent accumulation of heavy metals. (Ali & Khan, 2018; Fan et al., 2018; Muthusaravanan et al., 2018; Nahar, Rahman, Nawani, Ghosh, & Mandal, 2017; Rai, Lee, Zhang, Tsang, & Kim, 2019)
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Identify or generate regulators responsive to heavy metals.

        - Identify and remove, or modify specificity of, transporters involved in movement of heavy metals.
    - Host and Consortia Engineering Achievement:
        - Generate compartments or mechanisms for compartmentalization or sequestration and excretion of heavy metals.
    - Data Integration, Modeling, and Automation Achievement:
        - Protein modeling to identify putative sites for modification of transporter specificity, selectivity, and activity.
        - Automation in phenotyping.
- Engineering Biology Objective 2: Engineer and improve crops to enable reduction and/or elimination of toxins and allergens, such as gluten, peanut-allergen proteins, or other health-affecting factors. (Engagement Example: By engineering a reduction in the amount of asparagine in foods, such as potatoes, it is possible to decrease the formation of harmful acrylamides upon processing of those foods. See http://www.innatepotatoes.com for this example in action.)
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Generate mutations that delete or modify allergenic proteins or protein domains (such as peanut protein or gluten).
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Generate toxin specific sensors, reporters, and transcriptional regulators.
        - Modify metabolic pathways that generate anti-nutrients (such as phytate).
        - Modify endogenous biosynthetic pathways to metabolize or prevent production of toxins.
    - Host and Consortia Engineering Achievement:
        - Engineer mechanisms of sequestration and secretion.
        - Generate compartments, or mechanisms for compartmentalization or sequestration, and subsequent excretion of toxins.
    - Data Integration, Modeling, and Automation Achievement:
        - Protein modeling to identify putative sites for modification.
        - Flux analysis for modification of metabolic pathways.
        - Design and model signal perception and response systems for redirecting metabolism.
- Engineering Biology Objective 3: Improve food bio-processing to prevent and eliminate organic toxins (such as mycotoxins) in post-harvest environments.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Identify and engineer enzymes for bioconversion of toxins into non-toxic, consumable degradation products.
    - Host and Consortia Engineering Achievement:
        - Generate microbial or cell-free systems to degrade or sequester toxins.

        - Data Integration, Modeling, and Automation Achievement:
            - Protein modeling to identify enzymes with specificity, selectivity and activity for targeting toxins.
- Science/Engineering Aim 3: Increase the nutrient content and value from animal food sources.
    - Engineering Biology Objective 1: Increase bioavailability of non-vertebrate proteins.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Identify and edit genes to reduce chitin content in insects, crustaceans, and molluscs to reduce allergenic potential.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Edit relevant proteins for structure and expression.
        - Host and Consortia Engineering Achievement:
            - Reduce processing pathways that prevent protein availability.
            - Enable inducible expression of chitinase in transgenic insects at time of death.
    - Engineering Biology Objective 2: Increase non-vertebrate nutrient content.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Introduce genes for relevant nutrients (such as B12) in non-vertebrate animals.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer custom probiotics which produce relevant growth factors and micronutrients/vitamins.
        - Host and Consortia Engineering Achievement:
            - Host resilience optimized through genetically-determined pathogen resistance.

EBRC

June 2019

Environmental Biotechnology

Environmental Biotechnology

Environmental Biotechnology focuses on the technologies and tools to enable deployment of bioengineered systems in the land, air, water, and human landscapes for purposes related to remediation, natural resource management, environmental monitoring, and species management. In contrast to other sectors - including industry and agriculture - this technology is deployed in the environment in a process design that is inherently poorly controlled. Technical and societal challenges therefore reflect its broad and uncontrolled nature. The Challenges, Aims, and Objectives in this sector are highly parallel to those that also appear in *Energy* and *Health & Medicine*.

Societal Challenge 1: Address and mitigate climate change.

- Science/Engineering Aim 1: Enable adaptation of ecosystems to climate change.
    - Engineering Biology Objective 1: Enable and advance the production of drought-tolerant vegetation for growth on marginal land.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Genetic tools to edit non-model plants, including trees.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Plants engineered with waxy leaves to prevent evaporation.
            - Physiological root mechanisms for concentrating water and functioning in high-ionic-strength soils.
            - Engineer C4 pathways into C3 plants for increased drought- and heat-resistance.
        - Host and Consortia Engineering Achievement:
            - Transcription factor engineering to turn on/off water stress-related pathways.
            - Introduction of heterologous genes from other plants that enable moisture sensing (e.g., transporters, etc.).
    - Engineering Biology Objective 2: Enable and advance the production of self-fertilizing plants and/or cover crops.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Complex genotype construction in both model (e.g., maize) and non-model (e.g., clover, vetch) plants.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Optimize nitrogen fixation pathway enzymes to fit new host organisms.
        - Host and Consortia Engineering Achievement:
            - Transfer the nitrogen fixation pathway into plant.
            - Increase the nitrogen fixing activity of rhizobia-containing plants.
        - Data Integration, Modeling, and Automation Achievement:
            - Accurate prediction of nitrogen fixation pathways to insert into plant.
            - Multi-organism/multi-scale modeling of element cycling.

- Engineering Biology Objective 3: Engineering biomes for robust soils.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Genome engineering capabilities for implementing large numbers of targeted modifications in specific hosts that may have limited tools for transformation, modification, and programmable gene expression.
    - Genetic tools to engineer a variety of soil microbes.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Soil microbes engineered to produce extracellular polymers that stabilize the soil.
    - Biosensing systems to sense phosphate/nitrogen levels.
    - Engineered extracellular/intracellular mechanisms to concentrate critical elements and promote their long-term storage and/or release in soil, rather than dissolution and field run-off.
    - Engineer stable production of antibiotic and antifungal compounds to provide disease suppression for newly established crop plants.
  - Host and Consortia Engineering Achievement:
    - A variety of soil microbes that can be readily engineered (including transformation).
    - Produce stable, engineered microbial cultures in an agriculture setting.
    - Engineer microbial communities to sense water, carbon, and other nutrient contents of soils and secrete/consume appropriate nutrients to maintain the soil.
  - Data Integration, Modeling, and Automation Achievement:
    - Analysis of stability of engineered microbes in the soil.
    - Better understanding (analysis, modeling, and prediction) of microbial consortia in natural systems and how they interact and evolve over time and under different conditions.

- Science/Engineering Aim 2: Enable and advance carbon sequestration from the environment.
  - Engineering Biology Objective 1: Engineer soil biomes that more efficiently sequester carbon.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Genetic tools to engineer a variety of soil microbes.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Development of biopolymers that are not readily biodegraded and that can be secreted by engineered organisms into soil.
    - Host and Consortia Engineering Achievement:
      - Plants engineered to secrete carbonaceous materials into the soil for long-term carbon sequestration.
      - Microbes engineered to secrete recalcitrant biopolymers to extend sequestration periods.

        - Data Integration, Modeling, and Automation Achievement:
            - Environmental-scale modeling of biome population dynamics.
    - Engineering Biology Objective 2: Engineer highly productive plants for improved CO_2 removal and recycling from the environment.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Edit genes in the photosynthetic pathway for improved properties, including stability, catalytic activity, and substrate specificity.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve efficiency of key enzyme(s) in the photosynthetic pathway.
        - Host and Consortia Engineering Achievement:
            - Introduce synthetic (heterologous or modified) enzymes/complexes/pathways for photosynthetic efficiency, such as to increase wavelength absorption via engineered chromophores and to enable CO_2 concentrating mechanisms in C3 crops (i.e., C4-like or cyanobacterial-like mechanisms).
        - Data Integration, Modeling, and Automation Achievement:
            - Engineered-photosynthetic pathway modeling.
    - Engineering Biology Objective 3: Engineer organisms for improved methane removal and/or recycling from the environment.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Improve genetics in methanotrophs.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve heterologous expression of methane assimilation pathways.
            - Creation of non-natural pathways for methane utilization.
        - Host and Consortia Engineering Achievement:
            - Rapid genetic construction of methanotrophs.
            - Stable deployment of engineered methanotrophs in appropriate settings (such as well-heads).
        - Data Integration, Modeling, and Automation Achievement:
            - Predictive models of microbial ecology across a range of conditions and environments.
    - Engineering Biology Objective 4: Engineer recalcitrant, sinking marine phytoplankton for long-term carbon storage.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Multi-gene modification in non-model algae and cyanobacteria.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Redirection of photosynthate into recalcitrant biopolymers.
            - Enhance biological mechanisms for carbonate formation (i.e., fixing carbon in a non-reduced form).

        - Alter structural proteins and/or physiology in order to control phytoplankton density and subsurface deposition rate (i.e., sinking).
    - Host and Consortia Engineering Achievement:
        - Engineer hosts that, once they reach a certain density, aggregate/filament to accelerate sinking.
        - Reduce the need for iron, which is often limiting in the marine environment.
    - Data Integration, Modeling, and Automation Achievement:
        - Life cycle analysis of carbon turnover in the environment.

Societal Challenge 2: Expand tool sets for bioremediation and resource recycling.

- Science/Engineering Aim 1: Enable better, more advanced bioremediation of petrochemical pollutants including plastics.
    - Engineering Biology Objective 1: Engineer microorganisms to rapidly degrade hydrocarbons after an oil spill.
        - Gene Editing Achievement:
            - Engineer new genomic programs, such as combinations of synthetic auxotrophies, that increase the safety and reduce the risk of deploying engineered microbes in the field.
        - Biomolecular Engineering Achievement:
            - Synthesis of enzyme libraries to enable the identification of synthetic pathways with improved degradation performance.
            - Engineer improved transporters for hydrocarbon uptake.
            - Engineer catabolic enzymes with enhanced catalytic turnover.
            - Engineer orthogonal versions of critical enzymes that require xenobiotic molecules to function.
        - Host Engineering Achievement:
            - Hosts with improved tolerance to chemical insults.
            - Hosts that secrete engineered enzymes.
            - Hosts that selectively secrete natural, biodegradable detergents (e.g., bile salt-like detergents) to facilitate hydrocarbon turnover.
        - Data Integration, Modeling, and Automation Achievement:
            - Modeling of ecosystems (including ocean currents) to evaluate rapid deployment when necessary, such as post-oil spill.
    - Engineering Biology Objective 2: Engineer microorganisms to degrade recalcitrant plastics such as polyethylene terephthalate (PET).
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Engineer new genomic programs, such as combinations of synthetic auxotrophies, that increase the safety and reduce the risk of deploying engineered microbes in the field.

        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Synthesis of enzyme libraries to enable the identification of synthetic pathways with improved degradation performance.
            - Engineer enzymes that convert xenobiotic plastics into functional metabolic intermediates.
        - Host and Consortia Engineering Achievement:
            - Hosts that secrete engineered enzymes.
        - Data Integration, Modeling, and Automation Achievement:
            - Predictive modeling of ecosystem-wide ramifications of engineered-organism deployment into polluted ecosystems (including secondary/downstream and long-term effects) to evaluate regions of highest need and viability for organism deployment.
- Science/Engineering Aim 2: Improve bioremediation and revitalization of water and soil.
    - Engineering Biology Objective 1: Introduction of engineered microorganisms, plants, and animals (specifically fish and mussels) for watershed and wastewater remediation and revitalization.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Synthesis of gene clusters to produce relevant degradation machinery.
            - Strategies for efficient genome editing of naturally-occurring species to be re-introduced into the environment.
            - Engineer new genomic programs, such as combinations of synthetic auxotrophies, that increase the safety and reduce the risk of deploying engineered microbes in the field.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Design of novel enzymes to degrade xenobiotic small molecules such as toxins and antibiotics.
            - Design and construction of novel physiological mechanisms for the sequestration and processing of nitrogen and phosphate in run-off.
        - Host and Consortia Engineering Achievement:
            - Reliable integration of relevant pathways to be stable against genetic mutation.
            - Strategies to ensure the non-pathogenicity of the host (to the environment, over the long term), including the introduction of kill switches to remove engineered organisms once their work is finished.
            - Develop defined consortia of microbes and algae to most productively revitalize water sources.

        - Data Integration, Modeling, and Automation Achievement:
            - Predictive ecological models of species interactions to minimize the risk of uncontrolled cell growth in the environment/ecosystems.
    - Engineering Biology Objective 2: Plant platforms that can be engineered for remediation of a range of contaminated environments.
        - Gene Editing Achievement:
            - Plants transformed with pathways and metabolisms that enable the uptake of targeted contaminants and that have clearly visible 'markers' for public surveillance (for example, colors that clearly mark the plant as being genetically modified, so as to prevent people from eating these plants).
        - Biomolecular Engineering Achievement:
            - Enzymes engineered for efficient bioconversion and/or biosequestration of environmental contaminants.
        - Host Engineering Achievement:
            - Engineered variants of plants used for phytoremediation, such as Indian mustard (*Brassica juncea L.*), Willow (*Salix* species), Poplar (*Populus deltoides*), Indian grass (*Sorghastrum nutans*), and Sunflower (*Helianthus Annuus L.*), that do not permit the flow of DNA into wild-type plants.
        - Data Integration, Modeling, and Automation Achievement:
            - Design and modeling of genetic programs for pathways and engineered metabolisms in select organisms for a wide range of targeted environmental contaminants.
- Science/Engineering Aim 3: Enable more efficient and advanced resource recovery.
    - Engineering Biology Objective 1: Enable biorecovery of rare earth metals.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Knockout and gain-of-function systems to validate pathways/components and operation of sequestration involved in metal solubilization.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Detailed kinetics and stoichiometry of sequestration mechanisms, including interactions with relevant metals and competing ions.
            - Engineer orthogonal versions of critical enzymes that require xenobiotic molecules to function.
        - Host and Consortia Engineering Achievement:
            - Demonstration of solubilization and sequestration in relevant mesocosm studies.
        - Data Integration, Modeling, and Automation Achievement:
            - Complete modeling on most effective ecosystem for deployment of the organism, such as in an open-release scenario or in a

contained-use scenario which might be local (such as localized bioreactors at the point of recycling).

Societal Challenge 3: Controlled deployment of engineered organisms to improve ecosystem biodiversity, robustness, and the well-being of inhabitants.

- Science/Engineering Aim 1: Improve engineering of select insects for safe, effective environmental deployment.
    - Engineering Biology Objective 1: Design and produce insects with safe and effective gene drives to combat the spread of vector-borne infectious diseases. (Bier, Harrison, O'Connor-Giles, & Wildonger, 2018; Gantz & Bier, 2015; Gantz et al., 2015; Kyrou et al., 2018)
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Gene engineering capabilities for producing targeted sterility in insects/arachnids regardless of species.
            - Gene editing capabilities to enhance vector resistance to parasites.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improve ability to target vector reproductive capabilities without off-target effects and ride-along mutations.
        - Host and Consortia Engineering Achievement:
            - Ability to introduce genetically-encoded "kill switches" such as auxotrophies dependent on localized, environmentally-available compounds.
            - Ability to enhance host (vector) antibody production against specific pathogen antigens.
        - Data Integration, Modeling, and Automation Achievement:
            - Better predictive long-term environmental and disease models incorporating climate change data with sterile vector release programs.
            - Increase automation capabilities for gene editing and rearing large numbers of sterile vectors of different species.
    - Engineering Biology Objective 2: Characterize and engineer natural microorganisms to control insect populations. (J. A. Gilbert & Melton, 2018; National Research Council (US) Committee on Scientific Evaluation of the Introduction of Genetically Modified Microorganisms and Plants into the Environment, 1989)
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Engineer new genomic programs, such as combinations of synthetic auxotrophies, that increase the safety and reduce the risk of deploying engineered microbes in the field.
            - Knockouts to validate candidate gene(s) in *Wolbachia* (a common parasitic microbe) to confirm genotype-to-phenotype models, to

use *Wolbachia* as a biological tool for controlling insect populations.

- Biomolecule, Pathway, and Circuit Engineering Achievement:
  - Identify key biomolecular interactions involved in host-microbe interactions to prepare for future engineering.
- Host and Consortia Engineering Achievement:
  - Validate function of candidate native microbial organisms.
  - Genome engineering capabilities for implementing large numbers of targeted modifications in specific hosts that may have limited tools for transformation, modification and programmable gene expression.
- Data Integration, Modeling, and Automation Achievement:
  - Genotype-to-phenotype tools for what makes *Wolbachia* function for controlling vector reproduction/fitness and other physiological functions.
  - Genotype-to-phenotype tools for identifying candidate organisms from natural populations.
  - Automation in rearing/sexing of infected organisms to fit use requirements.

- Engineering Biology Objective 3: Increase diversification and resilience of capable insect pollinators.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Knockout or gain-of-function systems to validate semiochemical attractant/repellent pathways in candidate animals.
    - Identification of genes involved in susceptibility to various pollinator pathogens and/or toxins (e.g., honey bee susceptibility to *Varroa* mites).
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Ability to affect multiple genes, with minimal off-target physiological effects, to achieve increased resistance to pollinator pathogens and toxins.
  - Host and Consortia Engineering Achievement:
    - Demonstrate functional circuits conferring host-insect attraction in engineered insects.
    - Validate function, viability, and behavior of candidate native organisms.
  - Data Integration, Modeling, and Automation Achievement:
    - Accurate prediction of plant-insect pairs and potential ecological off-target pairs.
    - Increased automation capabilities for gene editing and rearing large numbers of disease-resistant pollinators.

- Science/Engineering Aim 2: Improved engineering of select plants and animals for safe, effective, environmental deployment.
    - Engineering Biology Objective 1: Design and produce animals with effective gene drives to produce sterile animals to control invasive species populations. (Grunwald et al., 2019; Jones et al., 2016)
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Gene engineering capabilities for producing targeted sterility regardless of species.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improved ability to target reproductive capabilities without off-target effects and ride-along mutations.
        - Host and Consortia Engineering Achievement:
            - Ability to introduce genetically-encoded “kill switches” such as auxotrophies dependent on localized, environmentally-available compounds.
        - Data Integration, Modeling, and Automation Achievement:
            - Better predictive long-term environmental and population models incorporating climate change data and engineered organism behavior/spread.
    - Engineering Biology Objective 2: Design and produce bio-containable engineered plant species. (National Research Council (US) Committee on Scientific Evaluation of the Introduction of Genetically Modified Microorganisms and Plants into the Environment, 1989)
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Develop and expand transformation tools and sequence databases to support editing in plants of interest.
            - Further develop and enhance cytoplasmic male sterility in select species.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Enable biocontainment through mitochondrial and chloroplast engineering.
        - Host and Consortia Engineering Achievement:
            - Develop *Agrobacterium* strain(s) capable of transforming new plants of interest at workable efficiency.
        - Data Integration, Modeling, and Automation Achievement:
            - Better predictive long-term environmental and population models incorporating climate change data and engineered organism behavior/spread.

    - Engineering Biology Objective 3: Use extant species homologues and genetic information for archiving and to achieve de-extinction for *select*, recently extinct (i.e., extinct for fewer than 150 years) species that have viable, available habitats.
        - Gene Editing Achievement:
            - Ability to effectively use CRISPR systems to combine varying amounts of genetic material from two species at the germline level.
        - Biomolecular Engineering Achievement:
            - Ability to grow germ cells for multiple species *in vitro* for easy gene editing.
            - Isolation and/or sequencing of genomic DNA from old, poorly preserved samples.
            - Develop CRISPR cassette delivery system to be able to effectively and efficiently edit embryos.
        - Host Engineering Achievement:
            - Develop viable hosts capable of successful breeding/pollinating, with life spans similar to original species.
        - Data Integration, Modeling, and Automation Achievement:
            - Better predictive long-term environmental and population models incorporating climate change data and engineered organism behavior/spread.
- Science/Engineering Aim 3: Develop high-resolution biosensors for environmental monitoring and detection.
    - Engineering Biology Objective 1: Develop better environmental biosensors for the detection of chemicals, radiation, temperature, pH, and water quality.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Large scale synthesis of biosensor variants, across kingdoms (from bacteria/archaea to plants).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Improved biosensor/genetic circuit designs for a multitude of sensor inputs.
            - Develop RNA or protein-based biosensors to detect and measure metabolites of interest.
        - Host and Consortia Engineering Achievement:
            - Create organisms that can act as multiplexing sensors capable of analyzing multiple environmental cues and providing measurable responses or combination of responses that may be deconvoluted to determine stimuli.
        - Data Integration, Modeling, and Automation Achievement:
            - Build more extensive and fully-sequenced metagenomics databases/libraries to enable searches for diverse functionalities across multiple gene clusters.

- Better enable real-time data feeds.

Societal Challenge 4: Enable sustainable, more environmentally-friendly materials and infrastructure development.

- Science/Engineering Aim 1: Enable sustainable production of environmentally-friendly consumable materials.
    - Engineering Biology Objective 1: Engineer organisms to produce durable, biodegradable materials (including plastics).
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Synthesis of gene clusters to produce and export materials.
            - Implement necessary genetic mutations to improve cell fitness during production.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer biological polymers with the desired properties (such as strength, flexibility, and permeability).
            - Engineer channels that secrete biopolymers to make purification easier and reduce potential toxic side-effects on host organism.
            - Develop recyclable and degradable plastics from monomers that can be readily produced from renewable resources using engineered organisms.
        - Host and Consortia Engineering Achievement:
            - Microbes that produce relevant monomers.
            - Microbes that polymerize monomers into desired polymer morphology (such as branching and chain length).
        - Data Integration, Modeling, and Automation Achievement:
            - Design and model genetic programs for pathways and engineered metabolisms in select organisms.
- Science/Engineering Aim 2: Enable the production of sustainable construction materials.
    - Engineering Biology Objective 1: Development and production of fast-growing trees that produce stronger, fire-resistant wood.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Vectors and transformation systems for a desirable trees and vegetation.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Enzymes and pathways that catalyze increased tannin deposition in plant tissue.
        - Host and Consortia Engineering Achievement:
            - Enable potential controlled propagation of engineered trees and vegetation.
        - Data Integration, Modeling, and Automation Achievement:
            - Identify genetic programs to improve resistance to fire through increased water uptake, tannin deposition, and taproot depth (similar to strategies employed by coastal redwoods).

        - Ability to predict and address factors limiting the scale-up of fast-growing, fire-resistant wood.
- Science/Engineering Aim 3: Improved management of the built environment using bio-designed and -enabled tools and technologies, including replacement of non-natural infrastructure with engineered organisms. (Engagement Example: By leveraging engineered enzymes during biofuel fermentation, byproducts of this process can be re-used to create more environmentally-friendly construction materials. See https://poet.com/asphalt for this example in action.)
    - Engineering Biology Objective 1: Engineer and produce organisms to replace or augment energy-consuming infrastructure or reduce energy consumption.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Enable tissue-specific gene expression in higher-order eukaryotes (such as in the twigs of deciduous trees or in needles of conifers).
            - Introduce luciferase and luciferin into a variety of trees (such as to create glowing plants and trees for natural, carbon-negative lighting in urban areas).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Produce and advance engineered enzymes and other biomolecules for energy-consuming processes.
            - Enable incorporation of photosynthetic pathways into infrastructure biomaterials (such as for the production of carbon-negative bio-concrete).
        - Host and Consortia Engineering Achievement:
            - Engineer tree branches and leaves to maximize canopy area to increase summer shading.
        - Data Integration, Modeling, and Automation Achievement:
            - Use modeling to design a tree with the optimum canopy area to increase summer shading.
            - Ability to design low-load (cyclic) bioluminescence pathways that generate luciferin/luciferase at pre-specified times and/or in response to external stimuli.

Energy

Energy

Energy focuses on the application of engineering biology tools and technologies to advance clean and affordable energy sources and to reduce overall energy consumption. To reduce the amount of carbon dioxide added to the atmosphere, energy will need to come from renewable sources, including waste gases (such as carbon dioxide and methane), microorganisms, and plants. Biology can be a source for renewable energy by providing biomass for electricity generation and for the production of highly energy-dense transportation fuels, and used to optimize processes to use less energy.

Societal Challenge 1: Produce affordable and clean energy.

- Science/Engineering Aim 1: Enable production of energy-dense and carbon-neutral transportation fuels from lignocellulosic feedstocks, oil crops, and agriculture and municipal wastes.
    - Engineering Biology Objective 1: Develop enzymes that can readily deconstruct lignin and cellulose/hemicellulose to monomers.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - High-throughput synthesis of large gene clusters (> 10 kilobases).
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer enzymes/pathways for production of hydrocarbons, including decarboxylating/decarbonylating enzymes, cofactor changes on enzymes, and new pathways/enzymes that conserve carbon/energy.
            - Engineer transporters to facilitate fuel export from the cell.
            - Improving enzymes for deconstruction of cellulosic biomass, including engineering cellulases and ligninases to be functional and stable in complex environments.
        - Host and Consortia Engineering Achievement:
            - Engineer microbes and/or consortia to efficiently express and secrete deconstruction enzymes.
        - Data Integration, Modeling, and Automation Achievement:
            - Technoeconomic and life cycle analysis models to determine sustainability of energy production.
            - Enzyme engineering models.
            - BioCAD models for designing gene expression.
    - Engineering Biology Objective 2: Further develop and advance oil crops that produce biofuels.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Edit chromosomes of oil crops to accumulate more and different types of oils.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer fatty acid synthases to produce fuels (fats) of a particular molecular weight.

- Host and Consortia Engineering Achievement:
  - Engineer oil crops to be drought tolerant and not require significant inputs of fertilizer.
- Data Integration, Modeling, and Automation Achievement:
  - Accurate prediction of factors leading to increased yields for oil crops.

- Engineering Biology Objective 3: Develop crops suited to specific climates (particularly marginal lands that would not be used to grow food) that require little water or fertilizer and can be readily deconstructed to aromatic and sugar monomers.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Plant chromosome synthesis.
    - Synthesis of complex (e.g., repeat) DNA.
    - Efficient CRISPR systems for plants.
    - Methods for efficiently editing organelle genomes.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Light energy conversion: engineer enzymes to more efficiently convert solar light to carbon/ATP.
    - Light capture: expand the range of solar spectrum wavelengths that can be captured by photosynthesis.
    - Improve CO_2 fixation by reducing 2-phosphoglycolate produced by RuBISCO O_2 fixation.
  - Host and Consortia Engineering Achievement:
    - Nitrogen fixation in plants.
    - Phosphate solubilization in plants.
    - Drought-tolerance traits in biomass crops.
    - Stable gene delivery to all plant tissues via viral vectors for prototyping of genetic designs.
    - Additional tools for controlling gene expression in plants, including large-scale knockout of unnecessary (for a given application or a particular environment) plant genes and pathways.
    - Metabolic pathways for producing bioproducts in plants (e.g., non-fuel products that could improve economics).
    - Methods to target and engineer specific microorganisms in plant microbiomes.
    - Co-regulation of plant and microbiome genes.
    - Develop microbiomes to aid nutrient uptake and water retention in soil.
  - Data Integration, Modeling, and Automation Achievement:
    - Models for identifying best geographic locations for energy crops.
    - Data collected at the plant- and field-level to understand growth and productivity.
    - Satellite imagery of plant productivity, land, and water use.

- Technoeconomic and life cycle analysis models to determine sustainability of energy production.
- BioCAD models for designing gene expression.
- Metabolic flux analysis of engineered organisms.

- Engineering Biology Objective 4: Develop microorganisms that can transform sugar and aromatic monomers into hydrocarbon-based liquid transportation fuels.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Synthesize large clusters of genes that encode metabolic pathways for various products (fuels, commodity chemicals, specialty chemicals, etc) ready to be transformed into any microbial host.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Engineer transporters to facilitate fuel export from the cell.
  - Host and Consortia Engineering Achievement:
    - Develop efficient pathways for production of liquid transportation fuels from metabolic intermediates.
    - Develop metabolic pathways in microbes that will allow them to simultaneously consume aromatic monomers (from lignin) and sugars.
  - Data Integration, Modeling, and Automation Achievement:
    - Metabolic flux analysis of engineered organisms.
    - Models of microbes in bioreactors to predict performance.
    - Technoeconomic and life cycle analysis models to determine sustainability of energy production.
    - BioCAD models for designing gene expression.

- Science/Engineering Aim 2: Enable production of energy-dense and carbon-neutral transportation fuels (and other products) from C1 feedstocks (particularly carbon dioxide, carbon monoxide, and methane).
  - Engineering Biology Objective 1: Engineer microorganisms that can transform carbon dioxide and electron sources to liquid fuels.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Efficient CRISPR systems.
      - Methods for efficiently editing organelle genomes.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Engineer enzymes/pathways for production of hydrocarbons, including decarboxylating/decarbonylating enzymes, cofactor changes on enzymes, and new pathways/enzymes that conserve carbon/energy.
      - Engineer transporters to facilitate fuel export from the cell.
    - Host and Consortia Engineering Achievement:
      - Develop efficient pathways for production of liquid transportation fuels from metabolic intermediates.

        - Develop tools to enable engineering of carbon concentration/fixation pathways in CO_2 fixers.
        - Large-scale knockout of unnecessary pathways.
        - Development of carbon transport and concentration mechanisms.
        - Organelle synthesis.
    - Data Integration, Modeling, and Automation Achievement:
        - Metabolic flux analysis of engineered organisms.
        - Models of microbes in bioreactors to predict performance.
        - Technoeconomic and life cycle analysis models to determine sustainability of energy production.
        - BioCAD models for designing gene expression.
- Engineering Biology Objective 2: Engineer photosynthetic microorganisms that can transform sunlight and carbon dioxide to transportation fuels.
    - Gene Editing, Synthesis, and Assembly Achievement:
        - Efficient CRISPR systems.
        - Methods for efficiently editing organelle genomes.
        - Synthesize and transform genomes of organelles.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
        - Light energy conversion: engineer microbes and algae to more efficiently convert solar light to carbon/ATP.
        - Light capture: expand the range of solar spectrum wavelengths that can be captured by photosynthesis.
        - Improve CO_2 fixation by reducing 2-phosphoglycolate produced by RuBISCO O_2 fixation.
        - Engineer enzymes/pathways for production of hydrocarbons, including decarboxylating/decarbonylating enzymes, cofactor changes on enzymes, and new pathways/enzymes that conserve carbon/energy.
        - Engineer transporters to facilitate fuel export from the cell.
    - Host and Consortia Engineering Achievement:
        - Develop tools for controlling gene expression in photosynthetic microbes.
        - Develop efficient pathways for production of liquid transportation fuels from metabolic intermediates.
        - Development of carbon transport and concentration mechanisms.
    - Data Integration, Modeling, and Automation Achievement:
        - Technoeconomic and life cycle analysis models to determine sustainability of energy production.
        - BioCAD models for designing gene expression.
        - Metabolic flux analysis of engineered organisms.
        - Models of microbes in photobioreactors to predict performance.

- Engineering Biology Objective 3: Develop microorganisms that can transform methane to liquid transportation fuels.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Efficient CRISPR systems.
    - Synthesize large clusters of genes that encode metabolic pathways for various products (including fuels, commodity chemicals, and specialty chemicals) ready to be transformed into any microbial host.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Recombinant expression of particulate and soluble methane monoxygenase (MMO) and demonstrate *in vivo* activity in a heterologous host.
    - Improve catalytic activity of MMO in native and heterologous hosts.
    - Engineer enzymes/pathways for production of hydrocarbons, including decarboxylating/decarbonylating enzymes, cofactor changes on enzymes, and new pathways/enzymes that conserve carbon/energy.
    - Engineer transporters to facilitate fuel export from the cell.
  - Host and Consortia Engineering Achievement:
    - Growth on alternative carbon sources to allow systematic genetic screening of critical enzymes for methane utilization.
    - More efficient and rapid transformation methods.
    - Large-scale knockout of unnecessary pathways.
    - Develop efficient pathways for production of liquid transportation fuels from metabolic intermediates.
  - Data Integration, Modeling, and Automation Achievement:
    - Metabolic flux analysis of engineered organisms.
    - Models of microbes in bioreactors to predict performance.
    - Technoeconomic and life cycle analysis models to determine sustainability of energy production.
    - BioCAD models for designing gene expression.
- Engineering Biology Objective 4: Engineer microorganisms that can transform carbon dioxide and electron sources to methane.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Efficient CRISPR systems.
    - Methods for efficiently editing organelle genomes.
    - Synthesize and transform genomes of organelles.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Light energy conversion: engineer microbes and algae to more efficiently convert solar light to carbon/ATP.
    - Light capture: expand the range of solar spectrum wavelengths that can be captured by photosynthesis.

- Improve CO_2 fixation by reducing 2-phosphoglycolate produced by RuBISCO O_2 fixation.
- Engineer enzymes/pathways for production of hydrocarbons, including decarboxylating/decarbonylating enzymes, cofactor changes on enzymes, and new pathways/enzymes that conserve carbon/energy.
- Engineer transporters to facilitate fuel export from the cell.
- Engineer enzymes/pathways for production of hydrogen (via water splitting).

■ Host and Consortia Engineering Achievement:
- Develop tools for controlling gene expression in photosynthetic microbes.
- Develop efficient pathways for production of liquid transportation fuels from metabolic intermediates.
- Development of carbon transport and concentration mechanisms.

■ Data Integration, Modeling, and Automation Achievement:
- Technoeconomic and life cycle analysis models to determine sustainability of energy production.
- BioCAD models for designing gene expression.
- Metabolic flux analysis of engineered organisms.
- Models of microbes in photobioreactors to predict performance.

- Science/Engineering Aim 3: Enable the efficient production of biomass for conversion to electricity.
  - Engineering Biology Objective 1: Develop crops that require little water or fertilizer and can be used to generate biomass designed for eventual electricity production.
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Plant chromosome synthesis.
      - Synthesis of complex (e.g., repeat) DNA.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Engineer proteins that improve the water acquisition in plant (aquaporins).
      - Engineer proteins that reduce water loss (uniport aquaporins).
    - Host and Consortia Engineering Achievement:
      - Improve plant transformation protocols.
      - Reduce generation time of model plant species.
    - Data Integration, Modeling, and Automation Achievement:
      - Risk assessment models for the impact of engineered plants on the surrounding ecology.
      - Methods to capture field data (such as growth, microbe composition) under non-standardized conditions over a long period of time.

- Engineering Biology Objective 2: Engineer organisms (including plants, microorganisms, and algae) to more efficiently convert solar light to fix carbon and produce ATP.
  - Gene Editing, Synthesis, and Assembly Achievement:
    - Plant chromosome synthesis.
    - Synthesis of complex (e.g., repeat) DNA.
  - Biomolecule, Pathway, and Circuit Engineering Achievement:
    - Improve efficiency of extracellular electron transfer from easy-to-grow microbes (e.g., *Shewanella*).
    - Light energy conversion: engineer microbes and algae to more efficiently convert solar light to carbon/ATP.
    - Light capture: expand the range of solar spectrum wavelengths that can be captured by photosynthesis.
    - Improve CO_2 fixation by reducing 2-phosphoglycolate produced by RuBISCO O_2 fixation.
    - Engineer plants with a higher content of lignin and lower cellulose/hemicellulose content to enable greater biomass production.
  - Host and Consortia Engineering Achievement:
    - Convert extracellular electron transfer of diverse redox potential metals from *Geobacter* species to *Shewanella* species.
    - Develop fast-growing, drought-tolerant grasses and trees for specific environments.
  - Data Integration, Modeling, and Automation Achievement:
    - Risk assessment model for the impacts of engineered plants on the surrounding ecology.
    - Methods to capture field data (such as growth, microbe composition) under non-standardized conditions over a long period of time.

Societal Challenge 2: Reduce global energy consumption.

- Science/Engineering Aim 1: Further develop energy-saving processes with biology.
  - Engineering Biology Objective 1: Develop and advance more efficient enzymes for everyday use (such as for laundry and dishwasher detergent).
    - Gene Editing, Synthesis, and Assembly Achievement:
      - Efficient CRISPR systems for more organisms.
    - Biomolecule, Pathway, and Circuit Engineering Achievement:
      - Improve proteases, cellulases, lipases, and amylases to enable/improve activity at low temperature, stability during storage, stability in aqueous detergent and activity in the presence of other enzymes.
    - Host and Consortia Engineering Achievement:
      - Efficient protein production and purification.

        - Data Integration, Modeling, and Automation Achievement:
            - Computational enzyme design.
            - Enzyme engineering models.
            - BioCAD models for designing gene expression.
- Science/Engineering Aim 2: Produce more energy-efficient crops that require less energy input for cultivation.
    - Engineering Biology Objective 1: Develop crops that require little water or fertilizer to reduce energy used for cultivation.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Efficient CRISPR systems for plants.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Drought-toleration synthetic circuits that close pores in response to environmental or externally-delivered signals.
        - Host and Consortia Engineering Achievement:
            - Produce model plants that have a higher yield.
            - Efficient plant transformation technologies (including for chloroplast and genome) for energy relevant plants.
        - Data Integration, Modeling, and Automation Achievement:
            - Risk assessment models for the impact of engineered plants on the surrounding ecology.
            - Methods to capture field data (such as growth, microbial composition) under non-standardized conditions over a long period of time.
    - Engineering Biology Objective 2: Expand the range of solar spectrum wavelengths that can be captured by photosynthesis.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Efficient CRISPR systems for plants.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Rational engineering and directed evolution of the photosynthetic light-harvesting complex.
        - Host and Consortia Engineering Achievement:
            - Efficient expression of functional designed enzymes.
        - Data Integration, Modeling, and Automation Achievement:
            - Modeling of photosynthesis and use of model to design a photosynthetic apparatus that captures more light.
- Science/Engineering Aim 3: Develop organismal bio-processes that enable the production of energy from (currently) atypical sources.
    - Engineering Biology Objective 1: Enable seawater fermentation.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Develop tools/methods for editing the chromosomes of microbes resident in seawater.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Biosynthetic pathways active in halophilic/hypersaline conditions.

        - Host and Consortia Engineering Achievement:
            - Develop robust genome engineering for halophiles (e.g., *Vibrio natriegens*).
            - Engineer resistance to phage/toxins.
            - Strict biocontainment strategies for potential engineered organisms deployment.
        - Data Integration, Modeling, and Automation Achievement:
            - Modeling of organism fitness in seawater.
    - Engineering Biology Objective 2: Enable wastewater fermentation.
        - Gene Editing, Synthesis, and Assembly Achievement:
            - Develop tools/methods for editing the chromosomes of microbes resident in wastewater treatment facilities.
        - Biomolecule, Pathway, and Circuit Engineering Achievement:
            - Engineer enzymes to convert fats in wastewater to hydrocarbons.
        - Host and Consortia Engineering Achievement:
            - Engineer resistance to phage/toxins.
            - Strict biocontainment strategies for potential engineered organisms deployment.
            - Engineer microbes for ammonium oxidation, denitrification, and polyphosphate accumulation.
        - Data Integration, Modeling, and Automation Achievement:
            - Model competition of microbes in wastewater treatment facilities to predict survival of engineered organisms.

References

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

Appendix

Appendix

A primer on DBTL for Engineering Biology

Engineering biology is a rapidly advancing discipline in which biological circuits and biochemical pathways with predicted functionality are implemented in living systems using systematic engineering workflows. A major difference between engineering/synthetic biology and classical engineering disciplines lies in the fact that engineered systems have been constructed from man-made and well-characterized building blocks in a “bottom-up” design strategy. In contrast, engineering biology often relies on partly characterized biological components that are implemented in extremely complex and dynamic living environments (cells and organisms) that are poorly understood. Because of this complexity, classical engineering approaches are only partly applicable to engineering biology. An iterative Design-Build-Test-Learn (DBTL) cycle has been developed that relies on data analytics and mathematical models with the goal of characterizing and controlling for the host response. Currently, the DBTL cycle is closely connected to the synthetic biology ecosystem, with many different companies working in different parts of the cycle.

The **DESIGN** process encompasses both biological design and operational design. For example, biological designs can specify desired cellular target functions, such as a cell that produces a complex natural product or that generates a detectable signal in response to an extracellular analyte. For operational design, the experimental procedures and protocols requires design. For example, the optimal amount of sample required to execute a specific experimental protocol to achieve required data capture. Required performance specifications must also be captured so that the process has a set of quantitative objectives to meet. To implement these functions in an organism then requires identifying appropriate biological parts (e.g., enzymes, reporters, regulatory sequences, etc.) that can be assembled to implement the desired function. Because the universe of biological parts is large and growing, standard registries that characterize these parts under a variety of different biological contexts and environmental physiological conditions and host organisms will be necessary. New approaches will be needed to specify effective design functions that can be used to drive the assembly of these components into functional assemblies. New mathematical and computational tools will be needed to solve these optimization problems and to specify appropriate constraints. Lastly, these optimal mathematical solutions will need to be implemented using optimal genetic parts to effectively map the space of potential solutions to the space of solutions that can actually be engineered. Design-of-experiment, or DoE, approaches could play an important role in efficiently searching for and assembling genetic parts and circuitry to enable the specified design with DNA sequences derived from either databases or the literature. As the search space is vast, DoE approaches still require choices to be made on what to search. Also, DoE approaches must be supplemented by computational methods to speed up the search for optimal genetic parts. The end of the design process is one or more DNA sequence(s) comprised of multiple genetic parts that generate the desired functions in a targeted biochemical, cellular, organismal, or biome context.

The **BUILD** process primarily consists of DNA assembly, incorporation of the DNA assembly in the host, and verification of the assembled sequence in the expected genetic context. The DNA build process iteratively assembles the DNA sequence specified in the Design process. The DNA assembly process uses molecular biology techniques, often aided by robotic automation, to combine multiple DNA fragments together and generally requires transformation into a host organisms for screening and verification of proper assembly. Build constructs are verified by DNA sequencing, restriction enzyme digests, and other techniques directed by software tools. Many design constructs require multiple hierarchical rounds of DNA assembly. For instance, round one may be used to assemble individual transcriptional units or large genes, round two may be used to assemble multiple individual transcriptional units to generate a biosynthetic pathway. The result of the DNA build process is a physical DNA molecule or, increasingly, a pooled library of DNA molecules that comprises the specified DNA sequence(s).

Delivery and verification of the DNA build within the desired host, or host build, is the second build process. This involves delivering the build genetic construct into the host organism, either as an independent genetic entity (e.g., a circular DNA plasmid or artificial chromosome), or by integration into a host chromosome. This is accomplished using standard molecular biology tools and is termed transformation. The efficiency of the transformation and selection of cells that contain the desired genetic sequence is often optimized through automation and a high-throughput screening process. When working with unstudied hosts, identifying amenable conditions for transformation and integration can require significant research. For example, host-onboarding and host optimization can require significant genetic manipulations of the host before testing, to remove adverse phenotypes and improve a host's utility for a specific design process. This could include the removal of the host's restriction endonuclease system or endogenous toxins, alterations to the membrane to improve phage susceptibility or alter immune modulation, or even inserting 'kill switches' or other biosafety features depending on the specific application.

The **TEST** process involves assessing whether the desired specified biochemical/cellular functions encoded in the designed DNA sequence have been achieved by the host organism or biome. This could also include testing genetic designs in multicellular transgenic organisms, although the scale and complexity of measurements required is challenging. For unicellular organisms, this requires growing the organism and assaying for the desired function (e.g. quantifying production of the desired product). Full validation of proper function and debugging non-functional designs may require substantially more intensive analysis, including tools such as proteomics, liquid chromatography-mass spectrometry, cas chromatography-mass spectrometry, and next-generation DNA/RNA sequencing. Measurements of, for example, product titer and yield, enzyme activities, cell phenotype, sensing thresholds and dynamic ranges, allows an assessment of the efficacy of the current design against the user-defined optimal target function. For bioprocessing, a major challenge is in scaling, which in a Test context requires measurements at small volume to inform large volume fermentation, an area of active research.

The **LEARN** process utilizes measured data and mathematical (statistical or mechanistic) models of the engineered biochemical, cellular, organismal, or biome context to

obtain actionable insights that can be used to generate better designs in the next iterations. For example, the integration of multi-omics data with metabolic models has been used to identify genetic interventions that improve titer, rate, and yield of engineered pathways. The cycle is then repeated until the user-defined target function is achieved.

The DBTL cycle thus provides an overall and iterative design framework to enable systematic design of biological systems at the genetic level as well as the elucidation of potential genetic design rules.

About EBRC and the Process of Creating the Roadmap

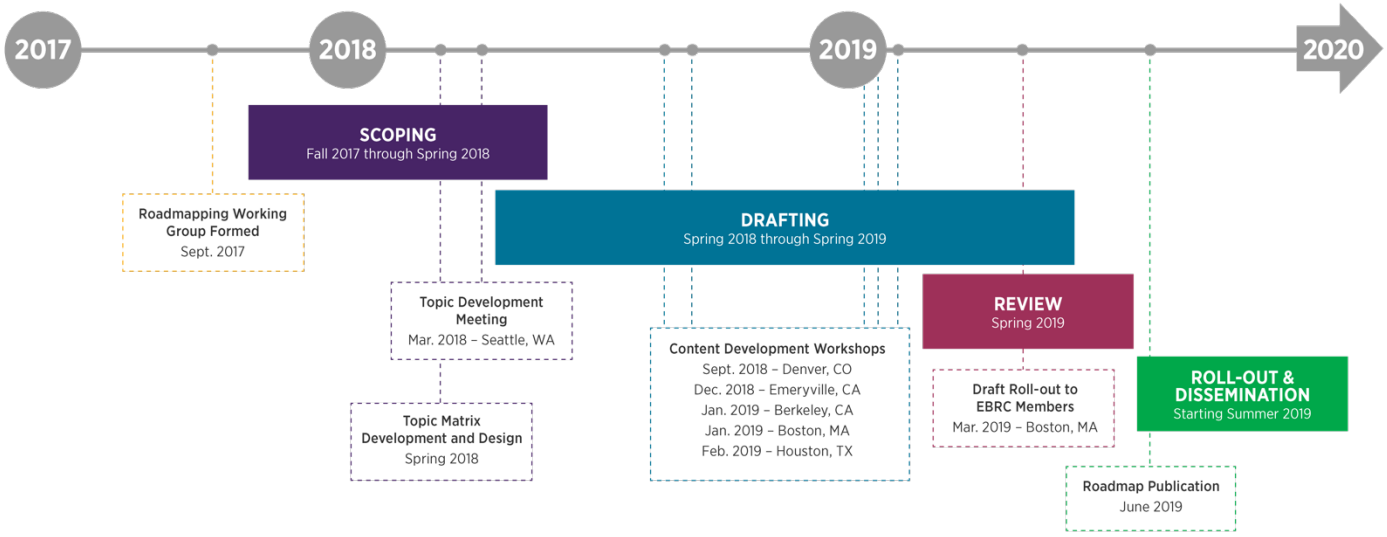
The roadmap has been developed using a community-driven process that emphasizes diversity and inclusivity. More than 80 members of the research community came together to make substantial contributions over the course of six workshops, three EBRC retreats, countless teleconferences, and immense individual effort.

About EBRC

EBRC comprises a membership of more than 80 academic faculty from nearly 40 institutions and scientists from more than a dozen companies and government agencies. The faculty, together with industry leaders and scientists, represent the leading talent in engineering biology research. They come from diverse scientific and engineering backgrounds including synthetic biology, bioengineering, chemical engineering, chemistry, nearly all life science disciplines, computer science, electrical engineering, and many others. The group also includes individuals from related fields including the social sciences, policy, and teaching and education.

The EBRC membership represents some of the nation's top scientists and engineers, and members' service roles and networks span the science and engineering enterprise, from interaction with policy makers, regulators, and funding agencies, to public and non-profit educational organizations (such as iGEM and BioBuilder), to scientific support organizations, including publishers and professional associations. These interactions give EBRC broad insight into engineering biology in the global enterprise.

Timeline



Our Process

Leading the large group of contributors was a core of 14 dedicated scientists and engineers. These leaders supported the group to generate more than 50,000 words of technical content in each of our sectors and themes. The additional working group members and other contributors worked together, and on an equal playing field, to develop this detailed roadmap for

the engineering biology community. The working group met regularly to discuss the content and scope of the Roadmap, with each member contributing valuable perspective and expertise.

The first several months were dedicated to scoping the project. This included wide-ranging discussions and input to determine how to organize a roadmap for such a diverse field. Ultimately, the four technical themes and five sectors were selected as our organizational tool.

The themes were then further developed using a bottom-up approach to understand the underlying technical topics, identify the transformative tools and technologies being developed as the current state-of-the-art, and ultimately to define goals, breakthrough capabilities, milestones, bottlenecks and potential solutions under each primary theme.

Concurrently, sectors were developed using a top-down methodology. Before thinking about the technology, we considered what major societal challenges were present in each sector, and then what science and technology (broadly) might help contribute to solutions. Only then did we consider how our field, engineering biology, could make a positive impact and what underlying technologies would need to be developed to get us there.

The roadmap itself was developed through an iterative process at a series of roadmapping workshops. Sections initially drafted at one workshop, were picked up by a wholly different group at another, allowing diverse community input to drive the content. Using an open, web-based approach, contributors were able to write, edit, comment, and question throughout the development process. Each section of the roadmap was reviewed by contributors that had not worked on the section originally. The content was then finalized and edited for consistency.