Data Integration, Modeling, and Automation

(Data Science)

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.

© 2019 Engineering Biology Research Consortium

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.

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