BIOAUTOMATION CHALLENGE 2022

May 11 2022

We are pleased to announce the 2022 Bioautomation Challenge Cohort. The Challenge offers life science researchers access to cloud labs in order to improve the reproducibility of life science research and gather large datasets, especially for groups focused on protein engineering. Nine groups have been selected, including groups spanning seven universities and three continents. This first cohort of awardees will pioneer research science in the cloud by identifying the best use cases of cloud labs in research, writing foundational open-source methods, and gathering the first cloud datasets.

Five groups are focused on the application of robotics and machine learning to protein engineering:

David Baker, University of Washington, USA.
Engineering de novo *protein binders in the cloud*

Kevin Esvelt, Massachusetts Institute of Technology, USA.
Mammalian protein binder engineering for pandemic preparedness

Yong Xin Zhao, Carnegie Mellon University, USA.
Virus like particle engineering

Huaiying Zhang, Carnegie Mellon University, USA.
Robotic quantification of protein phase diagrams

Heinrich Volschenk, Stellenbosch University, South Africa.
Antimicrobial engineering

Four groups will explore other applications of cloud labs in research science:

Michael Todhunter, Todhunter Scientifics, USA.
Closed-loop culture media engineering

George Church, Harvard, USA.
Robotic bacterial genome engineering using MAGE

William Greenleaf, Stanford, USA.
Standardized ATAC-seq sample preparation

Eske Willerslev, Cambridge University, United Kingdom.
Environmental DNA extraction for metagenomics

Groups will begin their cloud lab time between mid May and December 2022.

BIOAUTOMATION CHALLENGE 2022

David Baker, University of Washington

An automated, high-fidelity pipeline for measuring de novo *designed protein binding interactions*

NATE BENNETT, JOSEPH HARMAN, CHASE ARMER, BRIAN COVENTRY, XINRU WANG

Designing proteins that bind to arbitrary biological targets with high affinity and specificity remains an outstanding challenge in de novo protein design. We have previously designed high-affinity, hyperstable binders to multiple natural targets, including the SARS-CoV2 spike protein, influenza hemagglutinin, and numerous cancer targets, using a Rosetta-based computational binder design pipeline followed by iterative yeast display screening.

Despite these successes and recent advances in computational protein design, our per-binder success rate remains approximately 1%. The design protocol also often requires additional rounds of mutagenesis and sorting to mature initial hits into high-affinity binders. To enable the design of high-affinity protein binders using a single library sort, we aim to use bioautomation to generate a large, high-quality binding affinity dataset and use this dataset to both train a machine learning model to accurately predict binding affinities and to guide the improvement of our design pipeline.

BIOAUTOMATION CHALLENGE 2022

Kevin Esvelt, MIT
The Esvelt lab submitted two proposals

Evolving Receptor Decoy Mimics for Competitive Inhibition of Viruses

EMMA CHORY

This proposal seeks to use bioautomation and directed evolution to develop high affinity, yet endogenously-inactive mimics of sACE2 in order to create rapidly implementable therapeutics to combat SARS-CoV-2 and potential corona-like viruses. This approach would allow for the generation of scalable and translatable biologics, and provide a platform to rapidly course-correct for potential mutations that may arise in the future. Utilizing deep-learning with UniRep, we are able to rapidly improve de novo protein by screening a low-N (hundreds) of engineered protein mutants generated by error prone PCR. This proposal seeks to automate the transfection of libraries of subcloned protein variants, quantify their protein concentration, and subsequently characterize their binding affinities to generate protein receptor decoys. Using the SARS-COV2-Spike protein and ACE2 protein as a model, we can generate a protocol for the rapid engineering of protein receptor decoys for a host of human receptors that interact on the surface viral proteins.

Cloud-driven high-throughput biomolecular activity profiling at single-variant level

BO TU

The capability of precisely measuring biomolecular activity at single-variant level in a mixed pool of DNA sequences has important implications in the field of directed evolution, allowing direct mapping of biomolecular fitness landscape. Furthermore, such capability will enable rapid characterization of biological parts needed for increasingly complex genetic circuits and regulatory networks, and possibly acceleration in biological drug discovery. Our laboratory has recently developed Direct High-throughput Activity Recording and Measurement Assay (DHARMA), a method that enables highly-multiplexed single-variant activity measurement by directly recording biomolecular activity onto a segment of DNA contiguous to the corresponding activity-encoding sequence. Compared to conventional methods such as fluorescence activated cell sorting (FACS), which groups activities in a library into bins, our method can provide highly sensitive measurements of activities of individual library members without laborious instrument calibration against standard materials and reference strains. We aim to implement DHARMA on the cloud lab platform and create an end-to-end experimental/computational pipeline that determines individual biomolecular activities of a given list of DNA sequences.

BIOAUTOMATION CHALLENGE 2022

Yong Xin Zhao, Carnegie Mellon University

Development of Novel Vectors for Gene Therapy

YU HONG WANG, EMMA DIBERNADO

Through repurposing and engineering natural materials, gene delivery methods have achieved a broad range of applications over the past two decades. From clinical treatments such as vaccination and cancer therapy, to research-oriented uses such as optogenetics and stem cell engineering. Since the advent of CRISPR/Cas9 technology, efficient and accurate gene editing has exploded, yet their great promise has been bottlenecked by the vehicles to deliver them into mammalian systems. Viral capsids such as AAV and lentiviral vectors are established tools, but they have drawbacks such as limited packaging capacity, immunogenicity, scalability, and a laborious production process. While these challenges are driving the further development of these vectors, it is important to seek other potential vehicles. Here, I propose various candidate proteins that form Virus-Like Particles (VLPs) to overcome some of these shortcomings and a directed evolution scheme to optimize them that can be automated via the resources available at the Emerald Cloud Lab. Such automation methods will streamline production and screening, resulting in increased rigor and reproducibility of data. In parallel, we can capture a greater diversity of proteins, accelerate the identification of successful variants, and reduce the labor of production. As a result, we hope to create a production pipeline for VLPs and their subsequent usage as gene delivery vehicles that are scalable, efficient, and safe.

BIOAUTOMATION CHALLENGE 2022

Huaiying Zhang, Carnegie Mellon University

Mapping Protein Phase Diagrams

HUAIYING ZHANG

Similar to how oils separate from water, some proteins can spontaneously demix from their surroundings into condensed phases termed biomolecular condensates. Protein phase separation plays important roles in various biological processes such as stress response, transcription, and signaling. On the other hand, aberrant protein phase separation is linked to various maladies including HIV , cancer, and neurodegenerative diseases. As a result, a wave of startups has emerged to develop drugs targeting aberrant protein phase separation.

Several aspects of protein phase separation are functionally important and thus are attractive drug targets: the phase separation process, partition of molecules to the phase-separated condensates, and condensate material properties. Targeting any of these aspects requires understanding the factors that control normal and disease-related phase separation and screening for molecules that can modulate phase separation.
However, these endeavors are hindered by low-throughput in laboratory settings and the lack of standardization of methods to quickly find target windows for the many phase-separating proteins that are implicated in diseases. This is because protein phase separation is a system behavior and needs to be evaluated with a phase diagram that depicts the conditions under which the protein will phase separate

As such, the generation of protein phase diagrams is highly amenable to the automation capabilities provided by the Emerald Cloud Lab. Our goals are to use the Emerald Cloud Lab to:

Optimize purification conditions to increase the yield of phase separating proteins.
Produce multi-dimensional phase diagrams elucidating how parameters including concentration, protein domains, temperature, pH, and ionic strength influence protein phase behavior.

In summary, we will take advantage of the ECL to optimize purification for ‘sticky’ proteins and map their phase diagrams. This protocol can be adapted to produce phase diagrams for in vivo conditions which will enable us to quantify how phase separation is altered in disease-relevant parameters such as crowding in the cellular environment, presence of non-specific interactions of other biomolecules, post-translational modifications, and regulatory signaling of live cells. In addition, the protocol will allow us to screen for molecules that can disrupt protein phase behavior for cancer therapy. Beyond our lab, the standardized and high-throughput protocols can be readily used by our colleagues both in academia and industry to speed up discoveries in protein phase separation and therapy.

BIOAUTOMATION CHALLENGE 2022

Heinrich Volschenk, Stellenbosch University, South Africa

Developing microbial technologies for novel antimicrobial peptides

FRIEDE VAN DER BERG, DR KIM TROLLOPE

We aim to develop an automated method to prepare K. pastoris Cell-Free Protein Synthesis (CFPS) extracts and test their applicability in the engineering of an antimicrobial peptide that may find application in health and biomanufacturing.

Protein engineering is a molecular biology technique that makes use of recombinant DNA technologies and heterologous gene expression. However, design, construction and screening of microbially-expressed combinatorial mutation libraries is typically hampered by scale and complexity, necessitating development of advanced automation and optimization tools with improved efficiency and accuracy. At present, none of the available tools offer an automated and customizable design of mutagenic oligos for the construction of combinatorial gene libraries by gene synthesis methodologies. In addition, some of the published tools are no longer being supported, not freely available for widespread commercial use.

Combining automation with cell-free expression, peptide engineering and machine-learning provides a unique opportunity to significantly advance the discovery of novel AMPs and developing microbial technologies for their large scale production. Firstly, automation will enable the construction of large synthetic AMP variant libraries using synthetic oligo pools. Secondly, automation will enable high-throughput expression prototyping of AMP variant libraries using cell-free expression. Lastly, by combining machine learning approaches the data generated would greatly benefit developing predictive models for AMP functionality and production in microbial hosts. Unfortunately, despite recent exciting progress in synthetic biology and automation, the technology remains inaccessible for many in low- and middle-income countries due to the expensive reagents required for its manufacturing, as well as specialized equipment required for distribution and storage.

Our cell-free gene expression (CFE) of choice is Komagataella pastoris/phaffi. In K. pastoris expression screening workflows, the plating and selection of transformants is one of the costliest activities. A project design such as the proposed one stands to benefit greatly from the cost reduction that can be realized by successfully scaling down and screening genetic constructs by CFPS before one proceeds to production strain engineering. Clonal variability that is inherent to working with K. pastoris is removed as a variable during CFPS, which significantly reduces screening burdens and timelines. The number of constructs to be taken through to strain engineering would be significantly reduced after CFPS. K. pastoris provides numerous advantages as an expression host as it can perform post-translational modifications, grows in minimal, defined media that are animal origin free and secretes proteins to the culture medium which simplifies downstream processing. The development and automation of CFPS in eukaryotic K. pastoris thus has a lot to offer the scientific community as most CFPS development has been done using E. coli.

BIOAUTOMATION CHALLENGE 2022

Michael Todhunter, Todhunter Scientifics

Autonomous formulation and testing of cell culture media recipes with microscopy

MICHAEL TODHUNTER, ERIC CARLSON, BEA PIPPIN

Essentially all laboratory-based human biological research relies on maintaining cells in culture media that provides an appropriate chemical environment: nutrients, carbon sources, growth factors, and so on. The behaviors of cells - how long they replicate, how quickly they senesce, whether they differentiate, and what functional behaviors they exhibit - are contingent on their culture media environment. Most culture media recipes were designed at least 50 years ago when the goals of cell culture were much narrower. Today, we grow a wider variety of cells for a broader scope of purposes, but media formulation has not kept pace.

Culture media formulation is essential for several pressing topics in aging biology. First, senescence studies require low-stress (e.g. oxidative, glycative, or nitrogenous stress) media because stressful culture environments cause premature, non-physiological senescence. Second, the production of therapeutic biologics, such as heterochronic exosomes, requires culture media that supports sustained and metabolically intensive secretion. Third, specifying the differentiation of stem cells - a precondition for many cell therapies - requires culture media that can reconstitute specific stem cell niches. These avenues of research are stymied for lack of effective strategies to rapidly design bespoke culture media.

Traditional approaches are poorly suited to formulating culture media because media design space is very large and very dense. A culture medium may have 70 or more ingredients; some ingredients, like B27 or FBS, are themselves complex mixtures of even more ingredients. These ingredients are interdependent - e.g., the effect of insulin varies with the concentrations of glucose and IGF1, and the effect of retinoic acid varies with the concentrations of cholecalciferol and albumin. For a medium with 70 ingredients, assuming five relevant concentrations for each ingredient, the design space comprises recipes to explore. This design space greatly exceeds the capacity of factorial Taguchi methods or high-throughput screening, but it is well within the capacity of modern mathematical strategies such as gradient descent or Bayesian black box optimization.

I propose to fully automate this process. This protocol tests the effects of cell culture media on the growth and morphology of cultured cells. The protocol entails preparing media, thawing cells from frozen ampoules, culturing cells for a week, and high-throughput microscopy at four time points.

In the long run, I think this project helps address a fundamental problem with the pace of biological research. Biology today is a mashup of fast, data-rich omics research and slow, data-poor wet-lab research. I envision a future where biologists phrase research questions in a way that machines can answer. A single biologist could supervise numerous research automatons, running 24/7, performing experiments in parallel, iterating towards the biologist's goals. This is how we accelerate biological research, and I see no better place to start than changing how we formulate cell culture media.

George Church, Harvard

Systematic measurement of protein evolvability under an expanded genetic code

ANUSH CHIAPPINO-PEPE, RUSSEL MIRANDA VINCENT, HUSEYIN TAS, KAMESH NARASIMHAN

Genomic recoding refers to the process of replacement of codons with synonymous alternatives, resulting in synthetic genomes that can be used for a variety of biotechnological applications. Genomically recoded organisms (GROs) enable efficient incorporation of non-standard amino acids (nsAAs) in target proteins resulting in attractive properties such as novel catalysis, protein therapeutics, and biomaterials. Current approaches to engineer proteins with nsAAs involve rational engineering using computational protein design tools that are biased towards preliminary assumptions on the effect of nsAAs on a protein rather than generalizable knowledge on nsAA-based protein engineering. Such rational approaches can result in sub-optimal protein fitness with poor suitability towards industrial applications. In contrast, systematic measurement of protein fitness or activity across millions of variants in the presence of nsAAs could result in generalizable rules for nsAA incorporation in proteins. Herein, we propose to automate a multiplexed genome engineering (MAGE) to systematically measure the fitness of nsAA-containing essential proteins that may enable tight biocontainment in GROs, among other promising applications.

BIOAUTOMATION CHALLENGE 2022

William Greenleaf, Stanford

Evaluating the impact of genetic variations on the molecular phenotypes of human iPSCs using large-scale chromatin accessibility and transcriptomic assays

BETTY LIU

Over 85% of the phenotype-associated single nucleotide polymorphisms (SNPs) from GWAS studies lie within noncoding regions, the majority of which have no known functions. To bridge this gap in understanding between genotype and phenotype, expression quantitative trait loci (eQTL) studies have associated genetic variants to gene expression levels. eQTL data provides an additional layer of mechanistic information on how genetic variations lead to the expression changes in key genes relevant to the phenotype of interest. However, gene expression itself is a phenotype regulated by the interactions between the chromatin and a wide range of regulatory molecules (e.g. transcription factors), and gene expression data alone is not sufficient to reveal the epigenetic mechanism that drives downstream phenotypic variability.

The interactions of chromatin regulatory molecules occur at sites where the chromatin is accessible to these chromatin-binding factors. Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) is an efficient and robust technique used to evaluate genome-wide chromatin accessibility. Linking genetic variations to both chromatin accessibility and gene expression can define both the epigenetic “how” and the gene expression “what” of the downstream phenotypic differences. We aim to develop an automated high-throughput assay for measuring chromatin accessibility. We will use this assay to identify epigenetic effectors of disease-associated genetic variations at scale.

BIOAUTOMATION CHALLENGE 2022

Eske Willerslev, Cambridge, UK

Environmental DNA extraction for metagenomics

YUCHENG WANG

Our goal is the development of a robust yet flexible enough pipeline for the high-throughput extraction and sequencing library preparation of ancient environmental DNA, which refers to the highly fragmented ancient DNA molecules preserved in sedimentological substrates (e.g., marine and lake sediments, ice cores, permafrost, deposits in caves, rockshelters, and archaeological sites, coprolites, ancient gut contents and dental calculus) that can provide profound insights into the evolutionary history of whole ecosystems, but requires extensive "by-hand" laboratory treatment to extract and sequence. Our proposal centres on the automation of a modified phenol-chloroform extraction protocol. This type of extraction has long been a 'workhorse' for DNA isolation, though in comparison to more recent approaches (such as silica-binding spin columns), this method is time-consuming, involves hazardous reagents, and requires risky sample transfer between multiple tubes. Nonetheless, we have obtained results that demonstrate the strong utility of this method for ancient environmental DNA in our modified protocol provided here, and the large-scale application of this method through bioautomation would be highly desirable for our research, and within the rapidly growing field of palaeogenomics.