Event Agenda

Winter Q-Bio Conference 2025

Sun, Feb 16, 2025

5:00 PM - 8:00 PM

Opening Registration and Welcome Reception
Location: Ko Olina Ballroom Lanai

6:00 PM - 8:00 PM

Kids' Science! Bubble Lab
Location: Ko Olina Ballroom Lawn

Mon, Feb 17, 2025

6:40 AM - 8:40 AM

Registration and Breakfast
Location: Ko Olina Ballroom Lanai

8:40 AM - 10:40 AM

Opening Remarks and Morning Session I
Location: Ko Olina Ballroom 4/5

> **Hawaiian Morning Protocal**
> 8:40 AM - 9:00 AM
>
> **Opening Remarks: Galit Lahav & Hana El-Samad**
> 9:00 AM - 9:20 AM
>
> **Keynote: Frances Arnold | Innovation by Evolution: Bringing New Chemistry to Life**
> 9:20 AM - 9:55 AM
> Speaker: Frances Arnold
>
> **Keynote: Jennifer Doudna | Genome Editing the Future: Improving Human and Planet Health with CRISPR**
> 10:00 AM - 10:35 AM
> Speaker: Jennifer Doudna

10:40 AM - 11:00 AM

Coffee Break
Location: Ko Olina Ballroom Lanai

11:00 AM - 12:20 PM

Kids' Science! Playdough Planet and Crater Case
Location: Ko Olina Ballroom Lanai

11:00 AM - 12:20 PM

Morning Session II
Location: Ko Olina Ballroom 4/5

> **Keynote: Mo Khalil | Synthetic reconstitution of complex cellular function**
> 11:00 AM - 11:35 AM
> Speaker: Ahmad Khalil
>
> **William Shih: Single-molecule detection and identification via DNA nanotechnology**
> 11:40 AM - 11:55 AM
> Speaker: William Shih
>
> **Shirin Shivaei: Non-invasive imaging of cell-based therapies using acoustic reporter genes**
> 12:00 PM - 12:15 PM
> Speaker: Shirin Shivaei

12:20 PM - 1:40 PM

Lunch (on own)

1:40 PM - 3:00 PM **Afternoon Session I**

Location: Ko Olina Ballroom 4/5

> **Keynote: Alice Ting | Engineering proteins to map and manipulate cells**
> 1:40 PM - 2:15 PM
>
> **Zitong Jerry Wang: A cellular solution to a robotics problem**
> 2:20 PM - 2:35 PM
> Speaker: Jerry Wang
>
> **Alon Oyler-Yaniv: Proximal Negative Feedback as a Mechanism for Post Thymic T cell Selection**
> 2:40 PM - 2:55 PM
> Speaker: Alon Oyler-Yaniv

3:00 PM - 3:20 PM **Coffee Break**

Location: Ko Olina Ballroom Lanai

3:20 PM - 4:40 PM **Afternoon Session II**

Location: Ko Olina Ballroom 4/5

> **Keynote: Jamie Cate | Engineering the ribosome to make new sequence defined polymers**
> 3:20 PM - 3:55 PM
> Speaker: Jamie Cate
>
> **Rita Oliveira: Statistical design of a synthetic microbiome that suppresses a gut pathogen**
> 4:00 PM - 4:15 PM
> Speaker: Rita Oliveira
>
> **Natalie Sauerwald: Decomposition of phenotypic heterogeneity in autism reveals distinct and coherent genetic programs**
> 4:20 PM - 4:35 PM
> Speaker: Natalie Sauerwald

6:30 PM - 8:30 PM **Pau Hana**

Location: Ko Olina Ballroom Lanai

6:30 PM - 8:30 PM **Kid's Movie Night**

Location: Ko Olina Ballroom 2

8:30 PM - 10:00 PM **Karaoke!**

Location: Ko Olina Ballroom 2

Tue, Feb 18, 2025

6:30 AM - 8:30 AM **Breakfast**

Location: Ko Olina Ballroom Lanai

8:30 AM - 10:10 AM

Contributed Session I

Location: Ko Olina Ballroom 1

Paige Steppe: Engineered Bacterial Patterning Across Tumor Microenvironments

8:30 AM - 8:40 AM
Speaker: Paige Steppe

Elizabeth Libby: Tunable Intrinsically Disordered Regions for Engineering Precise and Robust Bacterial Ser/Thr Signaling Pathways

8:40 AM - 8:50 AM
Speaker: Elizabeth Libby

Konstantine Tchourine: Macroecological Laws Naturally Arise from Complex and Chaotic Species Dynamics

8:50 AM - 9:00 AM
Speaker: Konstantine Tchourine

Maggie Carroll: Engineering functional microbiomes by decoding environmental effects

9:00 AM - 9:10 AM
Speaker: Maggie Carroll

Erel Levine: Interactions among host and bacteria shape the composition, stability, and persistence of the worm microbiome

9:10 AM - 9:20 AM

Katie O'Connor: Targeting tumor fibrosis through host-microbial interfaces

9:20 AM - 9:30 AM
Speaker: Katie O'Connor

Alfonso Jaramillo: Autonomous learning in adaptive genetic circuits

9:30 AM - 9:40 AM
Speaker: Alfonso Jaramillo

Arun Chakravorty: Spatial Transcriptomics Reveals the Temporal Architecture of the Seminiferous Epithelial Cycle and Precise Sertoli-Germ Synchronization

9:40 AM - 9:50 AM
Speaker: Arun Chakravorty

Silas Boye Nissen: Cluster Assembly Dynamics Drive Fidelity of Planar Cell Polarity Polarization

9:50 AM - 10:00 AM
Speaker: Silas Boye Nissen

Alyssa Chiang: Engineered microbial consortium paradigm enables deployable whole-cell biosensors

10:00 AM - 10:10 AM
Speaker: Alyssa Chiang

8:30 AM - 10:10 AM **Contributed Session II**

Location: Ko Olina Ballroom 2

Kaleda Denton: Conformity to Continuous and Discrete Ordered Traits
8:30 AM - 8:40 AM
Speaker: Kaleda Denton

Yun Hao: A sequence-to-expression transfer learning framework for context-specific prediction of variant effects on expression
8:40 AM - 8:50 AM
Speaker: Yun Hao

Sandeep Kambhampati: TissueMosaic enables cross-sample differential analysis of spatial transcriptomics datasets through self-supervised representation learning
8:50 AM - 9:00 AM

Rohan Maddamsetti: Scaling laws revealed by computation of plasmid copy numbers across the tree of life
9:00 AM - 9:10 AM
Speaker: Rohan Maddamsetti

Gerald Pao: Finding biological signal integration on the surface of low dimensional manifolds
9:10 AM - 9:20 AM
Speaker: Gerald Pao

Shayna Holness: High intensity p-bodies formed in aged S. cerevisiae are a harbinger of cell death
9:20 AM - 9:30 AM
Speaker: Shayna Holness

Wojciech Szpankowski: Finding Biologically Significant Structures in PPI Networks
9:30 AM - 9:40 AM
Speaker: Wojciech Szpankowski

Jeff Drocco: Knowledge-Graph Guided Bayesian Active Learning for Discovering Top-K Genetic Interactions
9:40 AM - 9:50 AM
Speaker: Jeff Drocco

Betty Liu: A systematic investigation of transcription factor dose effects on the open chromatin reveals sequence determinants of dose response
9:50 AM - 10:00 AM
Speaker: Betty Liu

Andrew Savinov: High-throughput discovery of inhibitory protein fragments with AlphaFold
10:00 AM - 10:10 AM
Speaker: Andrew Savinov

10:10 AM - 10:30 AM **Coffee Break**

Location: Ko Olina Ballroom Lanai

10:30 AM - 12:20 PM **Kids' Science! Energy and Chain Reaction Machines**

Location: Ko Olina Ballroom Lanai

10:30 AM - 12:20 PM **Main Morning Session**

Location: Ko Olina Ballroom 4/5

Opening Remarks/Announcements
10:30 AM - 10:40 AM

Keynote: Neda Bagheri | Computational modeling of emergent spatiotemporal cell population dynamics
10:40 AM - 11:15 AM
Speaker: Neda Bagheri

Max Wilson: Optogenetic engineering of cellular signaling for discovery and therapy
11:20 AM - 11:35 AM
Speaker: Max Wilson

Keynote: Lingyan Shi | Metabolic Nanoscopy for Studying Aging and Diseases
11:40 AM - 12:15 PM
Speaker: Lingyan Shi

12:20 PM - 1:40 PM **Lunch (on own)**

1:40 PM - 3:20 PM **Main Afternoon Session**

Location: Ko Olina Ballroom 4/5

Keynote: Ilka Bischofs | Variability in the entry and exit from bacterial dormancy

1:40 PM - 2:15 PM
Speaker: Ilka Bischofs

Sabrina Spencer: Quiescence and senescence as graded states of cell-cycle withdrawal

2:20 PM - 2:35 PM
Speaker: Sabrina Spencer

Keynote: Arjun Raj | Can a cell learn?

2:40 PM - 3:15 PM

3:20 PM - 3:40 PM **Coffee Break**

Location: Ko Olina Ballroom Lanai

3:40 PM - 5:30 PM **Poster Session**

Location: Ko Olina Ballroom 3

7:30 PM - 10:30 PM **Pa'ina Haumana**

Location: Mekiko Cantina

Wed, Feb 19, 2025

6:30 AM - 8:30 AM **Breakfast**

Location: Ko Olina Ballroom Lanai

8:30 AM - 10:10 AM **Contributed Session I**

Location: Ko Olina Ballroom 1

Matthew Deyell: SatSeq: a scalable platform to map protein sequence to cellular and in vivo functions

8:30 AM - 8:40 AM
Speaker: Matthew Deyell

P. C. Dave Dingal: Co-evolution models predict specificity of protease-substrate interactions

8:40 AM - 8:50 AM

Negin Farzad: Human Lymph Node Cellular Senescence Atlas Reveals Age-Dependent Alteration in Germinal Center B Cell Function and Niches

8:50 AM - 9:00 AM
Speaker: Negin Farzad

Keisuke Ishihara: Hydrogel swelling drives cavity formation in multicellular spheroids

9:00 AM - 9:10 AM
Speaker: Keisuke Ishihara

Andrew Lu: Therapeutic protein circuits for cancer therapy

9:10 AM - 9:20 AM
Speaker: Andrew Lu

Shirin Shivaei: Deep-brain imaging of neuronal activity-dependent gene expression

9:20 AM - 9:30 AM
Speaker: Shirin Shivaei

Aika Toyama: Quantitative Analysis of Cyclin-CDK Interactions in Living Cells Using FCCS with Green and Near-infrared Fluorescent Protein

9:30 AM - 9:40 AM
Speaker: Aika Toyama

Toshimichi Yamada: Synthetic Organizer Cells Guide Development via Spatial and Biochemical Instructions

9:40 AM - 9:50 AM
Speaker: Toshimichi Yamada

Martin Tran: Lineage motifs as developmental modules for control of cell type proportions

9:50 AM - 10:00 AM
Speaker: Martin Tran

8:30 AM - 10:10 AM **Contributed Session II**

Location: Ko Olina Ballroom 2

Ksenia Sokolova: AI-Driven Framework for Integrating Genomic Variants into Individual Representations
8:30 AM - 8:40 AM
Speaker: Ksenia Sokolova

Tina Subic: Differentiable Loop Extrusion Model (dLEM) Provides a Physical, Genome-wide Description of Chromatin Contacts
8:40 AM - 8:50 AM
Speaker: Tina Subic

Arjuna Subramanian: Designing novel-to-nature protein building blocks for synthetic cells with foldtuned language models
8:50 AM - 9:00 AM
Speaker: Arjuna Subramanian

Razeen Shaikh: Optimal performance objectives in the highly conserved bone morphogenetic protein signaling pathway
9:00 AM - 9:10 AM
Speaker: Razeen Shaikh

Robyn Shuttleworth: Exploring the role of inflammatory cells and the extracellular matrix during tissue reprogramming
9:10 AM - 9:20 AM
Speaker: Robyn Shuttleworth

Vivek Behera: Conserved Principles of Spatial Biology Define Tumor Heterogeneity and Response to Immunotherapy
9:20 AM - 9:30 AM
Speaker: Vivek Behera

Anat Bren: Tradeoffs in bacterial physiology determine the efficiency of antibiotic killing
9:30 AM - 9:40 AM
Speaker: Anat Bren

Purushottam Dixit: The ability to sense the environment is heterogeneously distributed in cell populations
9:40 AM - 9:50 AM
Speaker: Purushottam Dixit

10:10 AM - 10:30 AM **Coffee Break**

Location: Ko Olina Ballroom Lanai

10:30 AM - 12:20 PM **Kids' Science! Chemical Reactions and Glowsticks**

Location: Ko Olina Ballroom Lanai

10:30 AM - 12:20 PM **Main Morning Session**

Location: Ko Olina Ballroom 4/5

Opening Remarks/Announcements
10:30 AM - 10:40 AM

Julia Salzman: Dual phenotype and functional prediction with genome-free statistical systems genomics
10:40 AM - 10:55 AM
Speaker: Julia Salzman

Keynote: Jacob Hanna | Synthetic Ex Utero Embryogenesis: from Naive Pluripotent Cells to Complete Developmental Models
11:00 AM - 11:35 AM
Speaker: Yaqub (Jacob) Hanna

Kate Galloway: Proliferation history and transcription factor levels drive direct conversion
11:40 AM - 11:55 AM
Speaker: Katie Galloway

Gavin Schlissel: Single-molecule analysis of morphogen diffusion revealed a novel mechanism of evolutionary diversification
12:00 PM - 12:15 PM
Speaker: Gavin Schlissel

12:20 PM - 1:40 PM **Lunch (on own)**

1:40 PM - 3:20 PM **Afternoon Session I**

Location: Ko Olina Ballroom 4/5

> **Keynote: Jennifer Oyler-Yaniv | Between TNFα and Proteostatic Stress Drives Cell Death and Guard Immunity**
> 1:40 PM - 2:15 PM
> Speaker: Jennifer Oyler-Yaniv
>
> **Zev Gartner: MAGIC matrices: freeform bioprinting materials to support complex and reproducible organoid morphogenesis**
> 2:20 PM - 2:35 PM
> Speaker: Zev Gartner
>
> **Keynote: Sydney Shaffer | The unseen history of cells in cancer evolution and treatment**
> 2:40 PM - 3:15 PM
> Speaker: Sydney Shaffer

3:20 PM - 3:40 PM **Coffee Break**

Location: Ko Olina Ballroom Lanai

3:40 PM - 5:00 PM **Afternoon Session II**

Location: Ko Olina Ballroom 4/5

> **Keynote: John Albeck | Visualizing metabolic heterogeneity and its integration with growth factor signaling in living cells**
> 3:40 PM - 4:15 PM
> Speaker: John Albeck
>
> **Ashley Laughney: Mapping the Emergent Functions of Proteins in Multicellular Systems**
> 4:20 PM - 4:35 PM
>
> **Maria Chikina: Towards interpretable sequence to function models**
> 4:40 PM - 4:55 PM
> Speaker: Maria Chikina

6:30 PM - 9:30 PM **Banquet!**

Location: Ocean Lawn

9:30 PM - 11:00 PM **Silent Disco!**

Location: Ko Olina Ballroom

Thu, Feb 20, 2025

6:30 AM - 8:30 AM **Breakfast**

Location: Ko Olina Ballroom Lanai

8:30 AM - 10:10 AM **Contributed Session I**

Location: Ko Olina Ballroom 1

Mohamad Abedi: Expanding the cell signaling space with more than a thousand de novo designed agonists
8:30 AM - 8:40 AM
Speaker: Mohamad Abedi Mohamad Abedi

Alexander Davies: Live-cell whole tissue models reveal sources of dynamic signaling heterogeneity and single cell drug response variation in the metastatic niche
8:40 AM - 8:50 AM
Speaker: Alexander Davies

Tiffany Zhou: Modeling Synchronized Lysis of Bacterial Populations in Spatially Extended Environments
8:50 AM - 9:00 AM
Speaker: Tiffany Zhou

Bryan Duoto: Quantitative RNA-Templated Size Control of Plant Virus-Like Particles
9:00 AM - 9:10 AM
Speaker: Bryan Duoto

Rikki Garner: Tissue fluidity: a double-edged sword for multicellular patterning
9:10 AM - 9:20 AM
Speaker: Rikki Garner

Stephanie Hartel: Quantifying Nutrition: A Metabolomic Perspective on Food Diversity and Classification
9:20 AM - 9:30 AM
Speaker: Stephanie Hartel

Oliver Inge: Combinatorial BMP4 and Activin direct choice between alternate routes to endoderm during human gastrulation
9:30 AM - 9:40 AM

Tavis Reed: Enhancing Context-Specific Interactome Mapping: Spatially Resolved Global Protein Interactome Networks and Interactome Homology in Viral Infections
9:40 AM - 9:50 AM
Speaker: Tavis Reed

David Van Valen: Understanding Kinase Substrate Interactions with Phospho-PCA and Deep Mutational Scanning
9:50 AM - 10:00 AM
Speaker: David Van Valen

8:30 AM - 10:10 AM **Contributed Session II**

Location: Ko Olina Ballroom 2

Kristen Naegle: Computational and synthetic experimental tools to understand domain-based tyrosine phosphorylation
8:30 AM - 8:40 AM
Speaker: Kristen Naegle

Benjamin Doran: Hierarchical organization across 60 million protein sequences reflects evolutionary constraints and differences of emergent properties
8:40 AM - 8:50 AM
Speaker: Benjamin Doran

Ivy Liu: Scalable, compressed phenotypic screening using pooled perturbations
8:50 AM - 9:00 AM
Speaker: Nuo (Ivy) Liu

Cordelia McGehee: An Application of Filippov Systems to Mathematical Modeling of Adaptive Chemotherapy Administration
9:00 AM - 9:10 AM
Speaker: Cordelia McGehee

Sadia Siddika Dima: Aggregation and DNA binding of Dorsal/NF-kappaB in early Drosophila embryos
9:10 AM - 9:20 AM
Speaker: Sadia Siddika Dima

Chenlei Hu: Scalable imaging-free spatial genomics through computational reconstruction
9:20 AM - 9:30 AM
Speaker: Chenlei Hu

Nik Kovinich: MYB SG2 and WRKY33 KEEP Motifs Mediate Essential Interactions of a Conserved Transcription Factor Network that Regulates Divergent Pathogen- Induced Biochemical Defenses
9:30 AM - 9:40 AM
Speaker: Nik Kovinich

Tess Marvin: Developing a Transfer Learning Model for Variant Pathogenicity Prediction
9:40 AM - 9:50 AM
Speaker: Tess Marvin

10:10 AM - 10:30 AM **Coffee Break**

Location: Ko Olina Ballroom Lanai

10:30 AM - 12:30 PM **Main Session and Closing Remarks**

Location: Ko Olina Ballroom 4/5

Zakary Singer: Engineered bacteria launch and control an oncolytic virus

10:30 AM - 10:45 AM
Speaker: Zakary Singer

Keynote: Marcella Gomez | A data-driven approach to modeling and control of wound state progression and healing outcomes

10:50 AM - 11:25 AM
Speaker: Marcella Gomez

Keynote: Brian Munsky | How crappy can our experiments be, without sacrificing uncertainty

11:30 AM - 12:05 PM
Speaker: Brian Munsky

Closing Remarks: Olga Troyanskaya

12:10 PM - 12:25 PM

12:30 PM - 12:30 PM **Meeting Adjourns**

Pau Hana: Mentorship Roundtable event
In the spirit of the Winter qBio Conference, we want to set the example for the science community by embracing diversity and inclusion. To this end, we are hosting our second annual “Pau Hana” – a roundtable event with lively conversations that serve to create a positive culture within the Winter qBio community. Attendees will rotate between tables to discuss a variety of topics, ranging from promoting equity in academia to developing fun lab culture.

Pa’ina Haumana: Young Scientist Networking event
In order to build community among young scientists at the Winter qBio Conference, we are hosting our third annual student-postdoc mixer. This event will foster connections across academic institutions and support open dialogue among peers.

Frances Arnold: Innovation by Evolution: Bringing New Chemistry to Life
Along with sharing some life experiences, I will explain why I love evolution for design. Chemistry encoded in DNA and optimized by evolution enables efficient, clean, sustainable routes to fuels, chemicals, materials, pharmaceuticals and more. Evolution not only optimizes, however--it can also innovate. We are learning to use evolution to create entirely new biocatalysts, ones that catalyze reactions unknown in biology and sometimes unprecedented in human-invented chemistry.

Jennifer Doudna: Genome Editing the Future: Improving Human and Planet Health with CRISPR
Fundamental research to understand how bacteria fight viral infections uncovered the function of CRISPR-Cas programmable proteins that detect and cut specific DNA or RNA sequences. CRISPR technology is now an indispensable tool in human, animal and agricultural research. Furthermore, the FDA’s approval of a CRISPR therapy for sickle cell disease marked the beginning of a new era in healthcare. I will discuss the scientific and societal advances that will expand both the applications and impact of genome editing across the globe.

Mo Khalil: Synthetic reconstitution of complex cellular function
There is a long and rich tradition of using bottom-up reconstitution approaches to study and understand biology. Recently, tools of synthetic biology are allowing us to scale the ambition of this approach, by providing the opportunity to reconstitute and control increasingly complex biochemical systems directly inside the cell. In this talk, I will describe our efforts to develop and apply synthetic reconstitution approaches to dissect cellular function across different levels of biological organization. Specifically, I will describe synthetic biology platforms we have developed to investigate and control (1) how transcription factors function within gene circuits to enable regulatory specificity in the eukaryotic genome, (2) how chromatin modifications collaborate to make robust and long-term epigenetic memories, and (3) an intriguing form of cooperation among individual cells. Throughout, I will show how the design principles we uncover and the technologies we develop are driving the development of methods to program precision and novel cellular functions for immune cell therapy and beyond.

William Shih: Single-molecule detection and identification via DNA nanotechnology
Signal amplification via molecular polymerization has proven to be a powerful strategy for the detection of nucleic-acid biomarkers, however most approaches are constrained either by the need for external instrumentation or limited sensitivity due to properties intrinsic to the molecular system design. Here we expand crisscross polymerization, a strategy previously shown to achieve robustly seed-dependent self-assembly of single- stranded DNA and DNA-origami monomers, to achieve autonomous, isothermal exponential amplification of crisscross ribbons through their concurrent growth and scission via toehold-mediated strand displacement. We demonstrate how this CrissCross Chain Reaction, or 3CR, can be used as a detection strategy through coupling to single- and double-stranded nucleic-acid targets. In addition to discussing crisscross, I'll also describe our efforts towards single-protein identification via mechanical fingerprinting. We combine DNA nanotechnology with single-molecule force spectroscopy to create a mechanically reconfigurable

DNA Nanoswitch Caliper capable of measuring multiple coordinates on single biomolecules with atomic resolution.

Shirin Shivaei: Non-invasive imaging of cell-based therapies using acoustic reporter genes
Cell-based therapies are rapidly emerging as a powerful therapeutic approach in medicine. The ability of engineered cells to traffic to and function at specific anatomical locations is a major aspect that differentiates them from traditional therapies. However, there is a lack of non-invasive, non-ionizing, cost-accessible methods to track therapeutic cells once deployed inside the body and to ensure their proper function. Here, we establish a platform for in vivo imaging of primary cell therapies using ultrasound – a ubiquitously accessible technology for high-resolution non-invasive imaging. We created and optimized a lentiviral delivery system to express acoustic reporter genes based on gas vesicles in mammalian cells such as T cells, showing that this results in robust ultrasound contrast. Additionally, we developed genetic circuits that link the acoustic signal to T cell activation state via activity-dependent promoters. We applied this technology in primary human T cells, using it to non-invasively track their accumulation and proliferation as a targeted therapy in a mouse tumor xenograft model and benchmarked it against commonly used standard terminal methods, such as immunohistology. By making it possible to visualize cell-based therapies and their function inside opaque living organs with unprecedented resolution and accessibility, this technology has the potential to significantly accelerate their development and effective use.

Zitong Jerry Wang: A cellular solution to a robotics problem
Beacon homing is a task where a robot navigates to a signal source in a complex environment. Cells face a similar challenge for navigating to a ligand source in tissue, where fluid flow and extracellular matrix binding break up ligand concentration gradients into discontinuous, patchy landscapes. We show that cells can solve this problem using a variant of a robotics algorithm called Bayes filtering. Through modeling and simulation, we demonstrate that adaptive redistribution of cell surface receptors observed in migratory cells implements the key steps of this algorithm. Compared to a uniform receptor distribution, this strategy improves navigation efficiency by 30-fold in simulated tissue environments. A unique aspect of the cellular implementation that departs from standard Bayes filtering leads to significant improvement in the performance of traditional robot localization algorithm.

Alon Oyler-Yaniv: Proximal Negative Feedback as a Mechanism for Post Thymic T cell Selection
Cytokine mediated interactions in the immune system are proximal in nature, extending only a few cell diameters from a site of production. Rather than being controlled by a single centralized hub, an immune response proceeds as the outcome of a collection of self-organized transient neighborhoods with different cellular and biochemical compositions.
Regulatory niches are transient structures composed of antigen presenting cells, conventional T cells (Tconvs), and regulatory T cells (Tregs) interacting in spatially isolated circuits. Tconvs can proliferate when their T cell receptor (TCR) binds to specific peptides presented by antigen presenting cells. In the presence of the activated T cell-derived cytokine IL-2, Tregs within the niche suppress T cell activation, forming a negative feedback loop—a common control mechanism that prevents runaway reactions and sharpens decision-making. The balance between TCR signaling strength and Treg-mediated suppression determines whether a T cell clone expands or dies out.
A successful immune response requires the expansion of T cell clones that respond strongly to an invading pathogen while preserving sufficient clonal diversity to guard against pathogen evolutionary immune evasion. The canonical model of inter-clonal competition posits that higher-affinity clones outgrow lower-affinity clones, in a process akin to natural selection. However, natural selection is typically slow, dependent on uncontrolled growth, and transiently permits the proliferation of weak, errant clones—dynamics that contrast with the observed clonal expansion and egress of T cells in secondary lymphoid organs.
Since multiple T cell clones can share a single regulatory niche, suppression operates both in cis—affecting the activated T cell—and in trans, influencing other nearby T cells. We hypothesize that regulatory niches function as distributed selection centers, promoting the expansion of high-affinity T cell clones while actively suppressing

the growth of lower-affinity clones. These dynamics foster a systemic response that is both potent—enriching strong clones within the repertoire—and robust, preserving clonal diversity and maintaining resistance to mutating pathogens.
To investigate this hypothesis, we combined theoretical models of regulatory niche formation with experimental studies of T cell selection and expansion. Our models revealed how niche-mediated feedback loops influence inter-clonal competition, selectively enriching high-affinity clones while suppressing weaker ones. These predictions were tested using in vitro co-culture systems and in vivo experiments that tracked T cell expansion dynamics in the presence or absence of trans-suppression. Together, this work demonstrates how regulatory niches shape the immune repertoire by balancing the enrichment of potent clones with the preservation of diversity, providing a deeper understanding of the mechanisms driving clonal selection and expansion.

Jamie Cate: Engineering the ribosome to make new sequence defined polymers
The ribosome is the universal translator of the genetic code and is shared across all life. Its core RNA sequences are among the most ancient in biology. We are pursuing a grand challenge in synthetic biology, repurposing the ribosome to make new sequence defined polymers with non-protein backbone chemistry. However, this requires tampering with some of the most conserved RNA elements in the ribosome catalytic center, as natural ribosomes only poorly incorporate new backbones into proteins. I will describe how we use ribosomal RNA phylogeny, deep learning, and structural biology to identify ways to improve engineering of the ribosome, and to achieve our goal of creating a new translator in biology.

Rita Oliveira: Statistical design of a synthetic microbiome that suppresses a gut pathogen
Microbiomes perform critical functions across many environments on Earth, including in the gut. The net output of the gut microbiome, which includes a wide range of metabolites, impacts host fitness and provides resistance to dense colonization by pathogens such as Klebsiella pneumoniae. However, when microbiota-mediated colonization resistance is compromised by antibiotics, K. pneumoniae expands markedly in the gut lumen, thereby enhancing the risk of healthcare-associated infections. In the case of K. pneumoniae, increasing resistance to carbapenem and broad-spectrum beta-lactam antibiotics leaves patients with very limited and sometimes no available treatment options. Therefore, alternative approaches targeting the reconstitution of the gut microbiota are currently being explored. Increasing experimental evidence suggests that administering consortia of live commensal bacterial species can optimize microbiome composition and function, leading to reduced disease severity and enhanced health. However, elucidating principles of consortia design is immensely challenging due to the complexity of microbe-microbe interactions and the need to elucidate their mechanisms of action against in suppressing pathogen gut colonization.
To address this, we used a diverse bank of gut commensal strains derived from healthy humans and targeted the clearance of the multi-drug-resistant K. pneumoniae. We engineered a functional synthetic microbiome using a novel approach that combines statistical inference with a ‘Design-Build-Test-Learn’ approach (‘DBTL+’), a framework used in synthetic biology to systematically and iteratively develop and optimize biological systems. This process was agnostic to mechanisms of action against the pathogen, bacterial interactions, or compositions of natural microbiomes and involved statistical inference focused solely on the presence or absence of strains. In just a single round of DBTL+, we successfully converged on a generative model of K. pneumoniae suppression. Statistical inference revealed 15 key strains crucial for community function. Combining these strains into a community (‘SynCom15’) suppressed K. pneumoniae across unrelated in vitro environments and matched the clearance ability of a whole stool transplant in a pre-clinically relevant mouse model of infection.
Our work introduces ‘statistical design’ as a potent concept for engineering synthetic microbiomes, offering a promising pathway for innovations in synthetic ecology. Through this research, we emphasize the potential of strain presence-absence in deriving principles of community design, paving the way for future applications in microbiome engineering.

Natalie Sauerwald: Decomposition of phenotypic heterogeneity in autism reveals distinct and coherent genetic programs

Unraveling the phenotypic and genetic complexity of autism is extremely challenging yet critical for understanding the biology, inheritance, trajectory, and clinical manifestations of the many forms of the condition. Here, we leveraged broad phenotypic data from a large cohort with matched genetics to characterize classes of autism and their patterns of core, associated, and co-occurring traits, ultimately demonstrating that phenotypic patterns are associated with distinct genetic and molecular programs. We used a generative mixture modeling approach to identify robust, clinically-relevant classes of autism which we validate and replicate in a large independent cohort. We link the phenotypic findings to distinct patterns of de novo and inherited variation which emerge from the deconvolution of these genetic signals, and demonstrate that class-specific common variant scores strongly align with clinical outcomes. We further provide insights into the distinct biological pathways and processes disrupted by the sets of mutations in each class. Remarkably, we discover class-specific differences in the developmental timing of genes that are dysregulated, and these temporal patterns correspond to clinical milestone and outcome differences between the classes. These analyses embrace the phenotypic complexity of children with autism, unraveling genetic and molecular programs underlying their heterogeneity and suggesting specific biological dysregulation patterns and mechanistic hypotheses.

Paige Steppe: Engineered Bacterial Patterning Across Tumor Microenvironments

We recently described a paradigm for engineering bacterial adaptation using plasmids coupled to the same origin of replication. In this study, we use plasmid coupling to generate spatially separated and phenotypically distinct populations in response to heterogeneous environments. Using custom microfluidic devices, we continuously tracked engineered populations along induced gradients, enabling an in-depth analysis of the spatiotemporal dynamics of plasmid coupling at both the micro- and mili- scales. Our observations reveal a pronounced and robust phenotypic separation within 4 h exposure to an opposing gradient of AHL and arabinose. Additionally, we optimized plasmid coupling for prospective applications in therapeutic delivery to tumors, focusing on environmental copy number adaptation across spatially distinct domains.

Elizabeth Libby: Tunable Intrinsically Disordered Regions for Engineering Precise and Robust Bacterial Ser/Thr Signaling Pathways

Creating robust biological parts for engineering signal transduction systems with defined activity is an important unsolved problem in designing sense-and-response systems for diverse applications, including biosensors and designer therapeutics. Bacterial Ser/Thr kinases offer some potential advantages over the well-studied two-component systems, including stability of Ser/Thr phosphorylation, separability of kinase and phosphatase activity, and modularity of substrates and receptors. However, these systems have not yet been adopted for synthetic biology due to a lack of downstream dedicated transcription factors, as well as a gap in our mechanistic knowledge of how kinase sequence translates to system output. To overcome these challenges, we used a combination of synthetic and quantitative biology approaches to create synthetic transcription factors and predictably engineer the activity of a Ser/Thr kinase. We developed modular synthetic transcription factors that respond to kinase activity based on phosphorylation of a short substrate drawn from an intrinsically disordered region. Using this system as a platform, we created a ~8k member transcription-factor library and used sort-seq to infer the activity of each variant. We found this substrate motif is highly plastic, with many variants showing increases in activity up to ~1.8-fold. A simple quantitative sequence-function map accurately predicts the activity of the transcription factors, providing a design framework. We demonstrated that the predictive power of the sequence-function map extends to the intrinsically disordered domain of the kinase itself. This enabled, for the first time, quantitative tuning of the activity of a Ser/Thr kinase in vivo, resulting in defined downstream gene expression. We then combined the engineered kinases with the synthetic transcription factors to develop synthetic bacterial Ser/Thr signaling pathways with a wide range of defined activity, including gains of up to >12-fold in response to signal. We applied this enhanced system to demonstrate sensing of a β-lactam antibiotic at ~1/200th the MIC, ~10 nM, bringing it within the range of many biologically active ligands for future sensing applications.

Konstantine Tchourine: Macroecological Laws Naturally Arise from Complex and Chaotic Species Dynamics

Recent studies revealed that microbiota dynamics follow multiple macroecological laws. Very similar macroecological laws, connecting various statistical descriptors of species' fluctuations and dynamics, have been previously described in ecologies of plants and animals. Despite their generality in nature, it is currently unclear why so many different ecosystems show strikingly similar macroecological relationships, and whether internal or external factors are primarily responsible for generating them. Here, we asked whether chaotic internal dynamics can simultaneously lead to multiple relationships observed in long-term and short-term microbiota dynamics. Our analyses show that very general constraints on species interactions and spatial migration parameters can simultaneously lead to multiple microbiological laws. Our investigation also provides several quantitative insights into the origins of variability of species abundances in ecology on various time scales. Overall, we demonstrate how these macroecological laws may mechanistically arise from biologically informed interaction-driven chaotic dynamics with a few common ecological constraints, providing an explanation for their broad prevalence in nature.

More specifically, our spatially-resolved generalized Lotka-Volterra ecosystem model included a number of spatially separated locations. Species interactions at each spatial location, which were all assumed to be well-mixed, were approximated by generalized Lotka-Volterra dynamics, and species were also able to migrate between the spatial locations. We used simulated annealing to optimize the five free hyperparameters describing aspects of species interactions, migration, and fluctuations in total carrying capacity. Notably, our model included no environmental stochasticity. The optimized hyperparameters resulted in the simulated species' trajectories that accurately approximated several macroecological laws, including Laplace scaling of daily abundance changes, residence and return times that scale as a power law with an exponential cutoff, power-law distributed rank-abundance distributions, anomalous diffusion Hurst scaling of long-term abundance drift, Taylor's law of long-term abundance fluctuations, and others, with accurate scaling exponents.

We used simulations with optimized hyperparameters together with experimental data to get crucial insights into the nature of macroecological relationships observed in multiple ecosystems. For example, we find that the abundance of each species behaves as a Gaussian process on short time scales, with changes possibly governed by random multiplicative processes, with a variance that changes in a scale-free manner every several weeks, and combines across longer time periods to form the overall Laplace distribution. We found that subdiffusive Hurst scaling is influenced by species spatial migration, and Taylor's law scaling is associated with the community matrix of interspecies interactions. We investigated this observation by analyzing publicly available long-term time series of bacterial abundances in baboon and human gut. This data revealed that Taylor's law scaling is strongly associated with functional capabilities of bacterial metabolic networks related to their ecological roles in the gut microbial community, consistent with the findings from our simulations about the relationship between Taylor's law scaling and species ecological interactions.

Maggie Carroll: Engineering functional microbiomes by decoding environmental effects

Bacterial microbiomes are complex biological systems whose functionality emerges from intricate interactions between microbial consortia and their environment. Designing effective microbial consortia could enable breakthroughs in microbiome- based therapeutics, bioremediation, and other applications. However, previous efforts in microbial biotherapeutic engineering have focused primarily on inter-bacterial interactions and the role of the environment remains poorly understood. This knowledge gap limits the predictability and applicability of microbiome functions, particularly when transitioning from in-vitro conditions to therapeutic and industrial settings.

To address this challenge, we adopted a data-driven approach that integrates insights from diverse in-vitro environments to uncover actionable design principles for microbiome engineering. Using statistical inference, we disentangled the contributions of microbial strains versus environmental factors in shaping community behavior. By understanding the environmental constraints that influence microbial consortia function, we gained insights into the robustness of these functions.

Key microbe-environment interactions were identified, enabling the development of strategies to rationally design microbial consortia with desired functional properties. Specifically, we identified environments that altered or abolished the ability of a designed consortium to suppress Klebsiella pneumoniae in mice. A predictive model trained across these environments allowed us to predict consortia suppression of Klebsiella

pneumoniae in novel in-vitro environments. This approach highlights the role of environmental epistasis in defining functional constraints and provides a path forward for designing resilient microbiomes. principles derived here establish a framework for engineering robust, adaptable symbionts capable of functioning across diverse and dynamic contexts.

Erel Levine: Interactions among host and bacteria shape the composition, stability, and persistence of the worm microbiome

Understanding the dynamics of microbiome assembly in vivo is crucial for engineering microbiomes de novo, accelerating microbiome recovery after perturbation, and introducing new species into existing communities. These dynamics are governed by interactions among microbes, host physiology, and the complex chemical and physical environment provided by the host. developing models that effectively integrate these factors and predict and facilitate control of assembly dynamics is complicated by the lack of real-time quantitative data from often-obscured habitats.

To address this challenge, we developed a microfluidics-based approach for real-time imaging of the gut microbiome in the worm C. elegans. Our system combines rapid multiplexed imaging, automated environmental control, and state-of-the-art image processing to monitor microbiome composition at single-bacterium resolution over extended periods in prescribed changing environments. We find that the gut microbiota organizes into two distinct states: a “fluid phase” of individual bacteria and a “condensate” composed of hundreds of bacteria, sometimes from multiple species. The formation, composition, and persistence of these condensates are influenced by bacterial interactions, host physiology and genetics, and the history of the microbiome. Our data show that microbiome assembly involves processes that span multiple temporal and spatial scales. To interpret these dynamics, we developed a modeling framework that contrasts the effect of environment, host, and microbial interactions. This framework highlights key distinctions between microbiome persistence and kinetic stability.

Katie O'Connor: Targeting tumor fibrosis through host-microbial interfaces

Bacterial therapeutics present a way to address drug delivery challenges in solid tumor environments (TMEs) by exploiting native adaptations of microbes to overcome inhibitory tumor phenotypes, such as fibrosis from a dysregulated extracellular matrix (ECM). Recent synthetic biology efforts have engineered bacteria to enzymatically degrade tumor ECM, alleviating drug perfusion barriers and immune trapping phenomena that are characteristic of fibrotic TMEs. However, off-target effects of enzymatic cytotoxicity remains a challenge yet to be addressed. Here we develop a novel control system utilizing bacterial quorum sensing to spatiotemporally regulate the enzymatic activity of ECM-degrading enzymes. We then evaluated this system in attenuated *Salmonella Typhimurium* in co-culture against *in vitro* human spheroid and patient-derived organoid platforms. By implementing spatiotemporal control of enzymatic production and subsequent enzymatic activity, bacteria were able to colonize model tumors to a much higher degree. Enzymatic activity was also found to selectively affect tumor microenvironments when co-culture with colorectal carcinoma organoids was compared to their matched-patient normal colon counterparts. Finally, resultant tumor phenotypes (i.e. breakdown of morphological integrity) were found to be proportional to bacterial seeding density, indicating potential for low-dosage in downstream translation. These results not only support bacteria as viable alternatives to alleviate canonical drug delivery challenges, but also encourage further coupling of payloads to bacterial population growth via quorum sensing systems.

Alfonso Jaramillo: Autonomous learning in adaptive genetic circuits

Adaptive gene circuits are engineered genetic networks designed to permanently and heritably modify their behavior in response to environmental stimuli, enabling living cells to perform complex functions without predefined programming. Traditional gene circuit designs rely on predefined DNA sequences, limiting their adaptability. We have engineered a type of adaptive gene circuit implementing a physical learning system that can self-adjust and evolve functions in response to environmental feedback. To achieve this, we introduced “memregulons”—minimal gene circuits that enable self-learning by incorporating analog memory directly into genetic networks. These circuits dynamically adjust their behavior through interactions with light and chemical inducers, allowing real-time modification of memory and fine-tuning of gene expression. By embedding learning

mechanisms within the genetic material, memregulons empower circuits to autonomously process inputs, evaluate outputs, and adapt when outcomes deviate from desired goals. This approach eliminates the need for encoding predefined functions in DNA, instead leveraging environmental interactions to guide functional evolution. We demonstrate that adaptive gene circuits equipped with memregulons can solve pattern recognition tasks and learn simple strategic behaviors, highlighting their potential for real-time adaptation. This capability offers transformative applications in bioremediation, biosensing, and therapeutic development, where living systems must effectively respond to changing and unpredictable conditions.

Arun Chakravorty: Spatial Transcriptomics Reveals the Temporal Architecture of the Seminiferous Epithelial Cycle and Precise Sertoli-Germ Synchronization
Spermatogenesis is characterized by the seminiferous epithelial cycle, a periodic pattern of germ cell differentiation with a wave-like progression along the length of seminiferous tubules. While key signaling and metabolic components of the cycle are known, the transcriptional changes across the cycle and the correlations between germ cell and somatic lineages remain undefined. Here, we use spatial transcriptomics via RNA SeqFISH+ to profile 2,638 genes in 216,090 cells in mouse testis and identify a periodic transcriptional pattern across tubules that precisely recapitulates the seminiferous epithelial cycle, enabling us to map cells to specific timepoints along the developmental cycle. Analyzing gene expression in somatic cells reveals that Sertoli cells exhibit a cyclic transcriptional profile closely synchronized with germ cell development while other somatic cells do not demonstrate such synchronization. Remarkably, in mouse testis with drug-induced ablation of germ cells, Sertoli cells independently maintain their cyclic transcriptional dynamics. By analyzing expression data, we identify an innate retinoic acid cycle, a network of transcription factors with cyclic activation, and signaling from germ cells that could interact with this network. Together, this work leverages spatial geometries for mapping the temporal dynamics and reveals a regulatory mechanism in spermatogenesis where Sertoli cells oscillate and coordinate with the cyclical progression of germ cell development.

Silas Boye Nissen: Cluster Assembly Dynamics Drive Fidelity of Planar Cell Polarity Polarization
The planar cell polarity (PCP) signaling pathway polarizes epithelial cells in the tissue plane by segregating distinct molecular subcomplexes to opposite sides of each cell, where they interact across intercellular junctions to form asymmetric clusters. The role of clustering in this process is unknown. We hypothesized that protein cluster size distributions could be used to infer the underlying molecular dynamics and function of cluster assembly and polarization. We developed a method to count the number of monomers of core PCP proteins within individual clusters in live animals, and made measurements over time and space in wild type and in strategically chosen mutants. The data demonstrate that clustering is required for polarization, and together with mathematical modeling provide evidence that cluster assembly dynamics dictate that larger clusters are more likely to be strongly asymmetric and correctly oriented. We propose that cluster assembly dynamics thereby drive fidelity of cell- and tissue-level polarization.

Alyssa Chiang: Engineered microbial consortium paradigm enables deployable whole-cell biosensors
Whole-cell biosensors (WCBs) offer an innovative, rapid, and cost-effective solution for biomonitoring in human health and environmental diagnostics. However, current WCB efforts predominantly focus on engineering genetic sensors optimized for detection of a single target analyte under controlled laboratory conditions. These approaches fall short under stressful environmental conditions and suffer from signal crosstalk, presenting considerable challenges for WCB deployment. To address these challenges, we propose a WCB framework of consortium biosensors, which leverage the cross-specificity of single-strain WCBs to enable identification of multiple targets through supervised learning classification. Specifically, we employ adaptive laboratory evolution of *E. coli* MG1655 to generate a genotypically and phenotypically diverse library of tractable strains exhibiting improved growth in seawater conditions. Equipped with sensors for five heavy metals of interest, the variant sensors were dynamically characterized in multiplexed, modular microfluidic devices, revealing significant signal diversity. Leveraging this diversity, we implement a simple machine learning classifier to decode the consortium's collective output to identify multiple analytes within cross-contaminated samples, outperforming the

predictive accuracy of traditional single-sensor systems. These results present a generalizable framework that could potentially facilitate the use of WCBs for practical, in-field settings outside the laboratory.

Kaleda Denton: Conformity to Continuous and Discrete Ordered Traits
Conformity and anti-conformity are cultural transmission biases observed across a range of species, including humans, birds, fish, and fruit flies. Traditional models of these biases have focused on cultural traits with unordered variants, such as behavioral strategies (e.g., cooperate / defect), the presence or absence of an innovation, or mate choice preferences in Drosophila. Fewer studies have explored (anti-)conformity to cultural traits with ordered variants, such as level of cooperation (low, medium, high) or proportion of time allocated to a task (0% to 100%). In these studies of ordered cultural traits, conformity is defined as a preference for the mean trait value in a population even if no members of the population have variants near this mean; e.g., 50% of the population has variant 0 and 50% has variant 1, producing a mean of 0.5. Here, we introduce models of conformity and anti-conformity to ordered traits, which can be either discrete or continuous. In these models, conformists prefer to adopt more popular cultural variants even if these variants are far from the population mean. To measure a variant's "popularity" in cases where no two individuals share precisely the same variant on a continuum, we introduce a metric called k-dispersal; this takes into account a variant's proximity to its k closest neighbors, with more "popular" variants being closer to others in trait space. We demonstrate through simulations that conformity to ordered traits need not produce a homogeneous population, as has previously been claimed. Under some combinations of parameter values, conformity sustains substantial trait variation over many generations. Furthermore, anti-conformity can drive pronounced polarization within populations. These findings highlight the complex effects of cultural transmission biases in shaping the distribution of ordered traits.

Yun Hao: A sequence-to-expression transfer learning framework for context-specific prediction of variant effects on expression
Interpreting context-specific transcriptional effects of genome variation is challenging yet critical for developing tailored disease risk prediction tools and treatments. One specific context receiving substantial interest in genomics is brain development given its implication in neuropsychiatric disorders. Existing sequence-to-expression models cannot make predictions for the diverse cell states involved in neurogenesis due to a lack of expression profiling samples of developing human brains. Here, we present Seal, an interpretable sequence-to-expression framework that leverages transfer learning-based neural network architecture to address the issue of limited training samples. Using TSS-proximal sequence as input, Seal can provide a comprehensive profiling of brain development in 26 tissues and 30 cell types over 7 developmental stages, a higher resolution than existing models. We demonstrate that Seal can achieve high accuracy and specificity in gene expression prediction, outperforming state-of-the-art models such as ExPecto and Enformer. We further demonstrate that Seal can prioritize causal eQTL variants. Among the eQTLs identified in fetal prefrontal cortex, Seal correctly predicts the direction of expression change for 71% of the top 100 variants with the strongest effect, again outperforming state-of-the-art models. Applying Seal to genetic variants from 1,790 autism spectrum disorder families, we provide insights into the transcriptional impact of de novo mutations in distinctive cell types of developing brain. Notably, proband-specific mutations exhibit stronger functional burden in principal neurons during the fetal stage. These analyses suggest Seal can generate mechanistic hypotheses for understanding the genetic basis of neuropsychiatric disorders.

Sandeep Kambhampati: TissueMosaic enables cross-sample differential analysis of spatial transcriptomics datasets through self-supervised representation learning
Spatial transcriptomics allows for the measurement of gene expression within native tissue context, thereby improving our understanding of how cell states are modulated by their microenvironment. Despite technological advancements, computational methods to perform comparative analysis across different samples and conditions are still underdeveloped. To address this, we introduce TissueMosaic (Tissue MOtif-based SpAtial Inference across Conditions), a self-supervised convolutional neural network designed to discover and represent tissue architectural motifs from multi-sample spatial transcriptomic datasets. TissueMosaic provides a data-driven basis to quantify the relatedness and dissimilarity of tissue motifs across different samples.

TissueMosaic further provides a framework to link these motifs to gene expression, enabling the study of how changes in tissue structure impact function. We formalize this analysis by developing a hypothesis testing framework to perform spatial differential expression analysis across conditions with multiple replicates. We can additionally identify and prioritize condition-specific tissue motifs, increasing our power to perform spatial differential expression despite the spatial heterogeneity inherent to tissues. We applied TissueMosaic to high resolution spatial transcriptomics datasets across multiple spatial transcriptomics technologies and tissues, including a mouse testis dataset with multiple wild-type and diabetic samples and a mouse thymus time course spanning Day 0 to Week 90 of mouse lifespan. In the testis, TissueMosaic discovered key genes in elongated spermatid cells exhibiting spatial dysregulation and associated with tissue disorganization due to diabetes. Within the complex changes in tissue architecture that occur throughout aging in the thymus, TissueMosaic highlights an immune activation signature associated with aggregation of antigen presenting cells during thymic involution. These findings underscore the potential of self-supervised learning to significantly advance spatial transcriptomics research. As the corpus of spatial transcriptomic data grows, TissueMosaic will enable large-scale cross-sample analyses to identify tissue motifs enriched in disease and genes that significantly covary with these motifs.

Rohan Maddamsetti: Scaling laws revealed by computation of plasmid copy numbers across the tree of life

Plasmids are extrachromosomal DNA elements that are foundational to biotechnology and drive key aspects of microbial dynamics, including the evolution and spread of clinical antibiotic resistance. Despite their importance, no comprehensive datasets on plasmid copy numbers exist. To address this gap, we invented Probabilistic Iterative Read Assignment (PIRA) to infer plasmid copy numbers at scale. We ran PIRA on all plasmid-bearing complete microbial genomes in the NCBI RefSeq database with linked short-read sequencing data in the Sequencing Read Archive (SRA). Our analysis of 4,540 bacterial genomes covers 10,261 plasmids and spans the microbial tree of life, revealing a universal anticorrelation between plasmid length and copy number. Based on this finding, we hypothesized that the functional properties of plasmids converge to the functional properties of chromosomes as they increase in length. Indeed, the fraction of DNA dedicated to protein-coding sequences converges to the fraction of DNA dedicated to protein-coding sequences on chromosomes. A similar relation holds for the number of metabolic genes encoded on plasmids. Theory and experiments are needed to better understand these empirical plasmid scaling laws.

Gerald Pao: Finding biological signal integration on the surface of low dimensional manifolds

Biological decision making frequently involves the integration of multiple signals where each of the inputs is insufficient to satisfy the conditions to trigger an event. An example of this is the G1 checkpoint in the cell cycle that requires centrosome duplication, nutrient availability, genomic integrity etc. simultaneously. The identification of each of the individual factors is difficult as they are often nonlinear combinations. Here we present a novel algorithm where many if not all variables determining a biological transition can be identified as a combination using time series analysis methods based on the Takens' theorem. With these we can identify combination of genes that allow the prediction of entry into cell cycle and predict when cell will enter into the cell cycle. The method that we call manifold dimensional expansion finds complex nonlinear relationships even in the near absence of correlation that allow prediction of future behavior. The model contains no latent variables as opposed to deep learning and allows for experimental testing of the generated hypotheses as it contains no latent variables. Furthermore the method is not sensitive to be trapped by local minima during optimization. The algorithm scales linearly with combinatorial complexity which increases exponentially which allows for the exploration of high dimensional processes. In the present we show an example of discovery of a cell cycle transcriptional response integration in which the individual genes involved show less than 10% correlation between the variables. However, we demonstrate empirically causal relationships between these variables, thus experimentally demonstrating the existence of causality without correlation in the cell cycle transcriptional response.

Shayna Holness: High intensity p-bodies formed in aged S. cerevisiae are a harbinger of cell death

Protein and RNA homeostasis are key components of cellular health, and the loss of homeostasis in both are well-established hallmarks aging of both individual cells and more complex multi-cellular organisms. RNA binding proteins (RBPs) in particular are known to be prone to aggregation, with specific aggregates associated with numerous neurodegenerative diseases (Parkinson's, Alzheimer's), and have been observed to aggregate in a variety of single cell aging studies. However, some RBPs coalesce to form cytoplasmic messenger ribonucleoprotein (mRNP) granules, such as p-bodies and stress granules, which have been shown to be beneficial to cell growth, especially in the face of environmental stress. Previous work revealed that ribosomal RBP (rRBP) aggregates and Hsp104p foci appear with greater frequency in older *S. cerevisiae*cells than in young cells, though they tend to appear in only a subset of cells with one of two distinct aging pathways, referred to as Mode 1 aging cells. By fluorescently tagging the p-body component Pab1 and observing its localization in single-cell aging experiments, we've determined that aging cells form p-bodies separate from and in addition to HSP104p aggregates (consistent with normal p-body formation). Though small, low fluorescence intensity, transient p-bodies form in young mother cells of both aging modes, Mode 1 cells also tend to experience larger, higher intensity, and more permanent p-bodies. These high intensity p-bodies are correlated with the future lifespan of the cells, which may indicate a transition in these mRNP granules from beneficial to pathological.

Wojciech Szpankowski: Finding Biologically Significant Structures in PPI Networks

Computational and comparative analysis of protein-protein interaction (PPI) networks enable understanding of the modular organization of the cell through identification of functional modules and protein complexes. These analysis techniques generally rely on topological features such as connectedness, based on the premise that functionally related proteins are likely to interact densely and that these interactions follow similar evolutionary trajectories. Significant recent work has focused on efficient algorithms for identification of modules and their conservation. One critical component of this infrastructure is a measure of the statistical significance of a network match (a conserved subgraph) or a *dense subcomponent*. Corresponding sequence-based measures, as presented last year during Q-Bio 2024, are key components of sequence matching tools. In this presentation we analytically quantify statistical significance of *largest dense components* and network matches in reference model graphs. We argue that such largest dense subgraphs occur a finite number of times in several positions of a graph, allowing us to assess statistical *correlation* in such biological networks. We also propose a method for evaluating statistical significance based on the results derived from this analysis, and demonstrate the use of these measures for assessing significant structures in PPI networks. Experiments performed on a rich collection of PPI networks show that the proposed model provides a reliable means of evaluating statistical significance of dense patterns in these networks.

Jeff Drocco: Knowledge-Graph Guided Bayesian Active Learning for Discovering Top-K Genetic Interactions

In silico methods for predicting the effects of multi-gene perturbations hold great promise for advancing functional genomics, computational drug discovery, and disease modeling. However, the development of these predictive algorithms for mammalian systems has been hampered by limited datasets and high experimental costs. Recently, researchers introduced an active learning strategy aimed at identifying dual gene knockdowns that significantly impact morphological phenotypes from a set of 50 selected human genes. However, limitations of this work included (1) the lack of integration of prior biological information from external data sources and (2) the absence of efficient batch diversification strategies.

In this study, we present a Bayesian active learning framework designed to discover pairwise host gene knockdowns that effectively inhibit viral proliferation in an in vitro HIV- 1 infection model. Our method leverages a biological knowledge graph as side information and employs a computationally efficient batch diversification approach. We evaluated this framework using a dataset of viral load measurements obtained from multi- day dual-gene depletion experiments, encompassing all possible pairwise knockdowns of over 350 host genes associated with HIV infection. We demonstrate that our framework rapidly identifies the most effective gene knockdown pairs for reducing viral load. Furthermore, we show that incorporating side information enhances performance during the early stages of active learning (low data regime), while our batch diversification strategy

significantly boosts performance in later stages (high data regime). This framework is general and can be adapted to explore gene interactions in other contexts, such as synthetic lethality prediction and mapping epistatic effects across quantitative trait loci.

Betty Liu: A systematic investigation of transcription factor dose effects on the open chromatin reveals sequence determinants of dose response

Transcription factors (TFs) regulate gene expression by binding to accessible regions of the genome. TF dosage remains highly conserved throughout evolution, and any perturbations are often associated with haploinsufficient diseases. However, a quantitative understanding of the effects of TF dosage on chromatin accessibility and gene expression remain unclear for most TFs. Conducting a systematic investigation of TF dose effects is challenging because current accessibility assays are time-consuming, labor-intensive and costly, often lacking reproducibility across small manual batches. To address this challenge, we developed RoboATAC - an automated, scalable and plate- based ATAC-seq platform validated against the gold-standard OmniATAC. Using RoboATAC and RNAseq, we quantified genome-wide accessibility and gene expression changes induced by variable overexpression of 22 different TFs in HEK293T and K562 cells across a total of 500 samples. Our analysis revealed significant enrichment of the overexpressed TF motifs within accessible regions, dose-dependent TF footprints, and significant overexpression of genes relevant to the TF function. Additionally, we characterized regulatory elements into sensitivity groups based on diverse dose response behaviors. We also observed that peaks with lower motif affinity reached accessibility saturation at higher TF doses, consistent with previous studies suggesting low affinity motifs play a significant role in dose sensitivity. Furthermore, we trained interpretative machine learning models (ChromBPNet) on the accessibility data for each TF dose and identified the role of motif orientations and flank sequences in regulating dose responses. Despite observed correlations between dose sensitivity and ChromHMM states, we compared the performance of logistic regression models to predict TF sensitivity groups based on sequence-only (ChromBPNet-derived) features, chromatin state-only (ChromHMM-derived) features, and combined features. Our results demonstrated that sequence alone is sufficient to predict TF dose sensitivity and that the addition of chromatin state information does not improve predictive power. Overall, this dataset provides a valuable resource for generating mechanistic hypotheses on how TF level regulates open chromatin and gene expression, as well as for developing strategies to leverage those mechanisms for reversing disease states.

Andrew Savinov: High-throughput discovery of inhibitory protein fragments with AlphaFold

Peptides can bind to specific sites on larger proteins and thereby function as inhibitors and regulatory elements. Peptide fragments of larger proteins are particularly attractive for achieving these functions due to their inherent potential to form native-like binding interactions. We have recently developed experimental approaches that enable high-throughput measurement of protein fragment inhibitory activity in living cells (Savinov et al., PNAS 2022). However, it has thus far not been possible to predict de novo which of the many possible protein fragments bind to protein targets, let alone act as inhibitors. We have now developed a computational method, FragFold, that employs AlphaFold to
predict protein fragment binding to full-length proteins in a high-throughput manner. Applying FragFold to thousands of fragments tiling across diverse proteins revealed peaks of predicted binding along each protein sequence. Comparing these predictions with experimentally measured in vivo inhibitory activity of protein fragments in E. coli, we establish that our approach is a sensitive predictor of fragment function.
Evaluating inhibitory fragments from known protein-protein interaction interfaces, we find 87% are predicted by FragFold to bind in a native-like mode. Across full protein sequences, 68% of FragFold-predicted binding peaks match experimentally measured inhibitory peaks. Further, FragFold is able to predict previously unknown protein binding modes, explaining prior genetic and biochemical data. We also perform deep mutational scanning experiments on protein fragments, uncovering a rich sequence-function landscape that supports predicted binding modes – and discovering superior inhibitory peptides in high throughput. The success rate of FragFold demonstrates that this computational approach should be broadly applicable for discovering inhibitory protein fragments across proteomes.

Neda Bagheri: Computational modeling of emergent spatiotemporal cell population dynamics
Computational models are essential tools that can be used to simultaneously explain and guide biological intuition. My lab employs agent-based modeling, machine learning, and dynamical systems to explain biological observations and interrogate multi-lateral regulatory networks that drive individual cellular decisions as well as cell population dynamics. We are interested in the inherent multiscale nature of biology, with a specific focus on system-level dynamics that emerge from interactions of simpler individual-level modules.
In this presentation, I introduce a multiscale agent-based model of a generic solid tumor microenvironment that integrates subcellular signaling and metabolism, cell-level decision processes, and dynamic vascular architecture and function. We use this modeling framework to understand decision processes among heterogeneous cell agents in changing microenvironments. The model is open-source and flexible/adaptable (it can characterize countless cell population dynamics!), but it is computationally costly to simulate and analyze at large scales. I highlight these challenges along with strategies to mitigate them, and showcase successes that derive from our model development process. I also describe how the model can be used to inform the design of experiments, interventions, and hypotheses that modulate population level responses.

Max Wilson: Optogenetic engineering of cellular signaling for discovery and therapy
The complexity of human cellular behavior presents both immense potential for biomedical innovation and substantial challenges for control. Our lab pioneers a unique approach to dissect and program these behaviors using cellular optogenetics– integrating hardware, software, and wetware to deliver precise, "virtual" signals to signaling pathways. This platform enables us to explore and engineer cellular memories, with the goal of uncovering cellular design principles that govern stress resilience and immune function. We engineered an optogenetically controlled Fc receptor (opto-FCR) expressed in macrophages and discovered that specific stimulation patterns dramatically increase phagocytosis, enhancing their ability to engulf cancer cells. Similar optogenetic programming of Protein Kinase R (opto-PKR) showed that specific temporal inputs at one time point enhance cellular resilience to stress in the future, providing a mechanistic explanation for a cellular memory and laying the groundwork for engineering adaptive cell states. These fundamental insights into how signaling pathways govern cell function provided the foundation for developing a high- throughput, all-optical screening platform to test the effects of small molecules on signaling pathway dynamics. This platform identified a novel class of "potentiators," compounds that selectively amplify critical stress-response pathways in already-threatened cells. These potentiators exhibit broad-spectrum antiviral activity, demonstrating efficacy against over 10 viruses in vitro and significantly reducing viral titers in an in vivo herpes model. Overall, this platform enables drug discovery in context, allowing us to find modulators of ongoing signaling dynamics rather than indiscriminate activators or inhibitors. These breakthroughs exemplify the power of programmable biology to bridge the gap between fundamental discovery and translation.

Lingyan Shi: Metabolic Nanoscopy for Studying Aging and Diseases
Understanding metabolism in living organisms is crucial for uncovering the fundamental mechanisms underlying various biological processes. My Lab has developed a multimodal microscopy technique that integrates heavy water-probed stimulated Raman scattering (DO-SRS), multiphoton fluorescence (MPF), fluorescence lifetime imaging (FLIM), and second harmonic generation (SHG) into a unified nanoscopy platform. By developing and implementing A-PoD and PRM algorithms, we have transformed this metabolic imaging platform into a super-resolution multiplex nanoscopy, achieving 59 nm resolution for volumetric imaging.
This approach utilizes various deuterated molecules—including glucose, amino acids, fatty acids, and water—as bioorthogonal metabolic probes. The enzymatic incorporation of deuterium results in carbon-deuterium (C-D) bonds in newly synthesized molecules, which can be detected by DO-SRS in the Raman spectrum's spectral cell-silent region, distinguishing them from older molecules. This nanoscopy provides novel insights into metabolic heterogeneity across different cell types and organ tissues under both physiological and pathological conditions.
One key discovery, made through collaboration, revealed that overexpressed tau proteins significantly disrupt lipid metabolism in aged and Alzheimer's-affected brains, leading to an excessive accumulation of newly formed lipid droplets in glial cells—an effect that can be mitigated by AMPK activation. This advanced nanoscopy

imaging platform holds significant potential for disease detection, diagnosis, drug discovery, and evaluating drug efficacy or resistance. Additionally, it serves as a powerful tool for understanding the fundamental mechanisms of aging and disease progression.

Ilka Bischofs: Variability in the entry and exit from bacterial dormancy
Bacteria of the genus Bacillus, known for their beneficial and industrial significance, have developed sophisticated strategies to adapt to environmental changes. A particularly fascinating trait is their ability to switch between a vegetative state and an endospore state in response to nutrient availability fluctuations. The spores can remain dormant for many years, potentially thousands or more, withstand adverse conditions and can revive within minutes. Spore-based applications of Bacillus, such as in green agriculture, leverage these properties. Understanding the formation and revival of spores is therefore of both fundamental and applied interest. Here, I will introduce our group's efforts to quantitatively study the life cycle of Bacillus subtilis using microscopy-based assays. I will specifically focus on how phenotypic population heterogeneity arises during both the entry into and exit from dormancy, provide evidence of their linkage, and discuss implications for spore quality control in ecological and industrial contexts.

Sabrina Spencer: Quiescence and senescence as graded states of cell-cycle withdrawal
A key determinant of tumor recurrence is whether residual cancer cells that remain after treatment are in a reversible state of cell-cycle withdrawal (quiescence) or an irreversible state of withdrawal (senescence). This distinction has been difficult to make because quiescent and senescent cells are defined by overlapping biomarkers and because cell-cycle withdrawal is much more poorly characterized than the cell cycle itself. While the reigning model in the field is that senescence and quiescence are entirely distinct states, we find that popular senescence markers (such as SA-β-Gal) are expressed in a graded fashion, with signals that increase gradually as a function of time spent withdrawn from the cell cycle. Moreover, our single-cell RNA-seq dataset (containing a population of bona fide senescent cells) reveals a bridge between quiescence and senescence. We therefore favor an alternative model in which quiescence and senescence exist on a continuum of cell-cycle withdrawal, where the probability of cell-cycle re-entry approaches zero at senescence. We are currently working toward quantifying a cell's depth-of-withdrawal to predict the likelihood of cell-cycle re-entry.

Arjun Raj | Can a cell learn?
Do cells remember their past experiences? DNA is one way to encode these past experiences, but cannot generally be altered within a lifetime. How, then, can cells learn? We present data showing that cancer cells, in the context of therapy resistance, remember their past, and this past can influence their future. We find the molecular mechanism by which such regulatory flexibility is achieved. We think this mechanism may be responsible for many forms of cellular adaptation.

Matthew Deyell: SatSeq: a scalable platform to map protein sequence to cellular and in vivo functions
Predicting phenotype directly from genomic sequences remains one of the most challenging problems in biology. This challenge is complicated by widespread pleiotropy, the observation that mutations in a single gene can affect multiple phenotypic traits. Here we present SatSeq: a modular method for combining single cell transcriptomics with saturation mutagenesis and DNA barcoding for mapping cellular and in vivo functions of a protein to individual amino acid residues at scale. We demonstrate SatSeq by mapping the functional landscape of an immune-related protein called Stimulator of Interferon Genes (STING) in the context of metastatic breast cancer. STING has been shown to be a highly pleiotropic regulator with both anti-tumor and pro-tumor effects at different stages of disease progression. By mapping phenotypic and drug effects across the entire mutagenesis library, SatSeq reveals how STING variants impact cancer progression, where emerging STING-directed therapies bind, and specific patient populations most likely to respond. Loss- and gain-of-function mutations were identified by comparing gene expression in each variant relative to wildtype STING. Notably, we identified multiple residues that uniquely map to IFN, ER stress or NFKb functions of STING; suggesting independent functional elements that may be targeted through allosteric effects. Next, we demonstrate that we can capture

known resistance and sensitizing mutations to DMXAA, a STING agonist that failed in clinical trials due to failed target binding. This approach has the potential to transform pre-clinical drug validation by both verifying the on-target effects of candidate compounds and revealing their mechanisms of action.

P. C. Dave Dingal: Co-evolution models predict specificity of protease-substrate interactions
Proteases from the family Potyviridae are widely popular choices for cleaving peptide epitopes off recombinant proteins and for engineering synthetic biological circuits. More than 3,800 Potyviral proteases are predicted to cleave peptides in a sequence-specific manner, making characterization of the specificity and efficiency of each protease experimentally challenging. To accelerate functional determination of substrate specificity, we built statistical models via direct coupling analysis (DCA) to predict and developed cell-based assays to validate protease performance. DCA modeling revealed several features of the Potyviral protease family: First, we identified several proteases that performed better in cell-based assays than the commercially available tobacco etch virus protease. Second, DCA predictions correlated well with experimental data on crosstalk interactions between multiple proteases. Third, the predictive power of DCA is sufficient to resolve the effect of single amino-acid changes on protease specificity and efficiency. We leveraged this substrate-specificity resolution to demonstrate the ability of model- specified proteases in detecting mutations that may arise in the cellular proteome. Current efforts are focused on directed evolution of de novo protease-substrate pairs and targeting substrates in vivo that are of biomedical relevance.

Negin Farzad: Human Lymph Node Cellular Senescence Atlas Reveals Age-Dependent Alteration in Germinal Center B Cell Function and Niches
The accumulation of senescent cells within organs and tissues is recognized as a fundamental hallmark of aging and is implicated in various age-related diseases. Despite the identification of several aging-associated gene markers contributing to this phenomenon, it is yet to further pinpoint the specific cell types that develop senescence and the impact on the tissue environment in an organ-specific manner aging is yet to be delineated. In this study, we employed a comprehensive approach integrating single-cell and spatial multi-omics methodologies to systematically investigate the proteome, transcriptome, and epigenome changes in the human lymph node (LN) associated with senescence during the aging process. Single-cell transcriptomes of human lymphoid tissues identified 34 cell types and comparison between young and old donors identified the major cell types with upregulated senescent signature (i.e., SenMayo) or specific genes with age. We performed spatial analysis of 30 fresh-frozen and formalin-fixed paraffin-embedded whole lymph node samples from donors spanning ages from 1 to 86 years. Spatial cell typing using high-plex immunofluorescence imaging of 12 whole lyph nodes via CODEX detected a total of ~20 million single cells that were successfully mapped to all cell types identified by single-cell transcriptomics. More importantly, we imaged a panel of senescence markers including P16 and P21 for cell cycle arrest, HMGB1 for senescence-associated secretory phenotype (SASP)-related processes, and gH2AX for DNA damage response (DDR), representing critical hallmarks of cellular senescence. Spatially linking the expression of distinct senescence markers to cell types defined by the fusion of single-cell transcriptome and CODEX revealed cell-type-specific senescence-like cells (SnCs) and their distinct distributions within the whole lymph nodes. Specially, we observed a shift from extrafollicular to germinal center (GC) enrichment of SnCs from young to old lymph nodes and highly focal clonal-like development of senescence in GC B cells in old donors, highlighting both cell type and spatial heterogeneity of SnC in aging. Global proteomics profiling expands the senescent markers to identify SnC proteomic signature in senescent stromal and GC B cells. We also performed spatial transcriptome sequencing of all 30 human lymph nodes that were integrated with spatial cell type atlas to elucidate SnC transcriptional heterogeneity and revealed the distinct pathways and functional impairment associated with senescent GC B cells in aging LNs. Spatial-ATAC-seq further revealed the gene regulation mechanism and transcription factor motifs presumably driving the development of senescence in a cell type specific manner. This study presents the first comprehensive spatial cellular senescence atlas of whole human lymph nodes in aging and revealed a distinct shift of cell-type senescence and focal development of GC B cell senescence, which has significant implication in human immune responses.

Keisuke Ishihara: Hydrogel swelling drives cavity formation in multicellular spheroids
Mechanical forces play a fundamental role in shaping the complex architecture of organs that compose our body. While molecular motors are a primary source of biological forces, forces can also emerge without direct energy consumption. Entropic forces, such as those generated by macromolecular crowding, have emerged as a novel principle for cellular morphogenesis. However, the potential of entropic forces to drive tissue-scale morphogenesis remains unknown. Here we report that the swelling forces of a cell- synthesized extracellular hydrogel drives cavity formation in multicellular tissues. We induce the synthesis of hyaluronic acid by overexpressing the hyaluronic acid synthase HAS2 in HEK cell spheroids. Within 24 hours, spheroids harbor acellular cavities that are up to ~100 micron in diameter and enriched with hyaluronic acid. Upon enzymatic digestion of hyaluronic acid, the cavities shrink and disappear rapidly. This indicates that the swelling forces of hyaluronic acid support acellular cavities by counteracting cell contraction. Altogether our results demonstrate a new physical mechanism to generate acellular space within multicellular tissues, leading to a new strategy to engineer the shape and topology of synthetic tissues.

Andrew Lu: Therapeutic protein circuits for cancer therapy
Targeted inhibition of oncogenic signaling has improved specificity and outcomes in cancer treatment, but can be susceptible to resistance, provoking an urgent need for new therapeutic modalities. Advances in synthetic biology now enable an alternative approach in which engineered proteins interact with cellular pathways to discriminate tumor and normal cells. In principle, these therapeutic protein circuits could be delivered to cells as mRNAs in lipid nanoparticles (LNPs), sense oncogenic signals, conditionally trigger cell death, and modulate the immune system to enable tumor clearance. In contrast to targeted small molecules that rely on oncogenic addiction, therapeutic protein circuits contain all components necessary to directly and immediately induce apoptosis or pyroptosis, and thus have the potential to be more effective and less prone to resistance. However, this paradigm has not yet been demonstrated end-to-end. Here, we show that therapeutic circuits can be engineered to sensitively and specifically eliminate Ras-mutant tumor cells while minimizing resistance in vitro and in vivo. We used rational engineering and computational protein design to build sensors that can selectively respond to mutant Ras in diverse cancer cell contexts, and signal amplifiers that can further increase circuit sensitivity. Encoding circuit components as mRNAs and delivering them to cells in LNPs, we demonstrated that sensors can be coupled to conditional cell death execution proteins to effectively kill Ras-mutant cancer cell lines without harming healthy cells expressing wild-type Ras. Preliminary in vivo results show that sense-kill circuits have potent anti-tumor activity against Ras-driven mouse liver tumors. Finally, data suggests sense-kill circuits could be more effective and less prone to resistance than state-of-the-art Ras-targeted small molecules Sotorasib and RMC-7977 in various cell culture models. We anticipate that therapeutic circuits could be a versatile and programmable platform, enabling context-dependent selection of sensors, signal processing units, cell death modes, and delivery vehicles. These results establish the foundation for engineering therapeutic circuits to treat human cancers.

Shirin Shivaei: Deep-brain imaging of neuronal activity-dependent gene expression
Despite significant advances in neurotechnology, methods to capture whole-brain activity in live, intact animals remain limited. Achieving this goal requires non-invasive imaging tools capable of penetrating deep inside the tissue. Light-based imaging techniques, while offering high spatiotemporal resolution, are hindered by poor penetration into the intact brain. In contrast, ultrasound is a widely accessible biomedical imaging modality that enables deep tissue visualization. Recently, gas vesicles (GVs)—microbially derived, gas-filled protein nanostructures—have been established as genetically encodable acoustic reporters, allowing cellular-level imaging. However, their use in mammalian systems has been restricted to ex vivo engineering of cell lines. Here, we extend the utility of acoustic reporters to in situ imaging of gene expression deep within the brain by engineering an AAV-based system that enables robust GV expression in neurons. Following intracranial injection in mice, we demonstrate, for the first time, the ability to use ultrasound to image in situ gene expression in the intact brain of a live animal through a cranial window. To track neuronal activity, we linked GV expression to activity-dependent promoters that track immediate early gene expression in the hippocampus after seizure episodes. Leveraging ultrasound's capability to image deep structures, this reporter system allowed us to

repeatedly image brain-wide gene expression in the same animal over time, a significant advancement over terminal histological methods and optical imaging techniques with limited tissue penetration. This platform represents the first genetic toolset for whole-brain imaging of gene expression in live animals with ultrasound and paves the way for longitudinal studies of whole-brain function and activity dynamics in their native context.

Aika Toyama: Quantitative Analysis of Cyclin-CDK Interactions in Living Cells Using FCCS with Green and Near-infrared Fluorescent Protein

Epidermal growth factor (EGF) plays diverse roles in a wide range of physiological and pathological processes, including wound healing and cancer invasion. However, accurate quantification of EGF gradients in vivo remains elusive. To address this challenge, we have tried to develop an EGF biosensor based on a circularly permuted fluorescent protein (cpFP). Binding of EGF to its receptor (EGFR) induces significant conformational changes in the structure. cpFP is uniquely suited for biosensor applications as its fluorescence intensity is highly sensitive to structural changes in the proteins attached to its N- and C-termini. By inserting cpFP into a site of EGFR that undergoes conformational changes upon EGF binding, we designed a biosensor capable of detecting EGF-induced structural dynamics based on the change in fluorescence intensity. This sensor enables the real-time observation of the spatiotemporal dynamics of EGF in the extracellular space of living cells and organisms. A key challenge in developing functional biosensors is the labor-intensive process of screening sensor variants. Due to its nature as a membrane protein, our biosensor cannot be screened using *E. coli*-based systems. To overcome this limitation, we utilized the Landing Pad cell system (Matreyek, *Nucleic Acids Res*, 2017; 2020), which allows for fluorescence-activated cell sorting (FACS)-based screening in mammalian cells. We have constructed a mutant library by introducing random mutations into the linker sequences between the extracellular domain of EGFR and cpEGFP. Our next step is to sort this library based on fluorescence intensity changes in the presence and absence of EGF. In this presentation, we will discuss the results of our screening and highlight the utility of this novel biosensor for visualizing EGF.

Toshimichi Yamada: Synthetic Organizer Cells Guide Development via Spatial and Biochemical Instructions

In vitro development often depends on treating progenitor cells with media-borne morphogens, lacking the spatial information present in native tissues. To address this, we engineered morphogen-secreting organizer cells that are programmed to self-assemble around mouse embryonic stem (ES) cells into defined architectures through engineered cell adhesion. By inducing the morphogen WNT3A and its antagonist DKK1 from organizer cells, we generated range of morphogen gradients varying in range and steepness. These gradients were closely correlated with specific morphogenetic outcomes: the range of minimum-maximum WNT activity determined the resulting range of Anterior to Posterior (A- P) axis cell lineages. Notably, while shallow WNT activity gradients produced truncated A-P lineages, they generated higher-resolution tissue morphologies, such as a beating, chambered cardiac-like structure associated with an endothelial network. This study demonstrates that synthetic organizer cells, capable of integrating spatial, temporal, and biochemical cues, offer a versatile tool to guide ES or other progenitor cells through various paths in the morphogenetic landscape.

Martin Tran: Lineage motifs as developmental modules for control of cell type proportions

In multicellular organisms, cell types must be produced and maintained in appropriate proportions. One way this is achieved is through committed progenitor cells or extrinsic interactions that produce specific patterns of descendant cell types on lineage trees. However, cell fate commitment is probabilistic in most contexts, making it difficult to infer these dynamics and understand how they establish overall cell type proportions. Here, we introduce Lineage Motif Analysis (LMA), a method that recursively identifies statistically overrepresented patterns of cell fates on lineage trees as potential signatures of committed progenitor states or extrinsic interactions. Applying LMA to published datasets reveals spatial and temporal organization of cell fate commitment in zebrafish and rat retina and early mouse embryonic development. Comparative analysis of vertebrate species suggests that lineage motifs facilitate adaptive evolutionary variation of retinal cell type proportions. LMA thus provides insight into complex developmental processes by decomposing them into simpler underlying modules.

Ksenia Sokolova: AI-Driven Framework for Integrating Genomic Variants into Individual Representations

Decoding the relationship between whole genome sequencing and clinical phenotypes remains a significant challenge in advancing precision medicine. To address this, we developed a novel AI-driven framework that integrates millions of genetic variants into a unified and functional patient-level representation, capturing both coding and non-coding effects. We demonstrated the utility of this method in the longitudinal CureGN study of glomerular disease patients, showing that these patient embeddings are biologically informative, predictive of clinical outcomes, and capable of stratifying patients into unique genome-centered subgroups.

A key feature of our model is the transition from traditional allele frequency-based approaches to a patient-level representation that integrates the functional effects of genetic variants, including rare or previously unseen, into a biologically interpretable format. This framework relies solely on sequence information: coding variant effects are prioritized based on predicted severity (e.g. AlphaMissense, VEP scores), while non-coding variant effects are inferred from cell-type-specific gene expression dysregulation predictions (ExpectoSC). These signals are further refined through contextual normalization and aggregation steps, enabling the creation of robust and consistent representations. This approach embeds individual genetic profiles into an informative and compact space, allowing for meaningful comparisons across patients and uncovering biologically relevant patterns.

We applied the model to the Cure Glomerulonephropathy Network (CureGN), a consortium studying four rare glomerular diseases. While these diseases share clinical similarities, they exhibit highly variable mechanisms and trajectories. Identifying genetic factors that predict critical clinical outcomes, such as progression to end-stage renal disease (ESRD) or 40% reduction in estimated glomerular filtration rate (eGFR), is essential for improving patient management and developing interventions tailored to specific disease mechanisms. Using our patient embeddings, we demonstrated the ability to stratify patients based on survival and predict key outcomes from genetic data alone. Additionally, clustering analyses identified distinct patient subgroups, providing insights into the genetic heterogeneity of these diseases and their implications for underlying mechanisms.

This study introduces a powerful AI-driven framework for integrating whole genome sequencing data into a unified, biologically meaningful representation at the patient level. By bridging the gap between genomic data and clinical phenotypes, this framework offers a pathway toward more targeted and personalized strategies for disease management and therapeutic development.

Tina Subic: Differentiable Loop Extrusion Model (dLEM) Provides a Physical, Genome-wide Description of Chromatin Contacts

3D chromatin organization is central to gene regulation and is commonly studied using contact maps from techniques like Hi-C and Micro-C. These maps reveal chromatin's hierarchical organization, with mechanisms operating across different length scales. On the scale of hundreds of kilobases to megabases, Topologically Associating Domains (TADs) are formed through cohesin-driven loop extrusion, a dynamic process influenced by sequence, chromatin state, and interactions with transcription.

How can we predict and interpret the effect of the sequence and underlying processes on the chromatin structure? Black-box machine learning models excel in mapping nonlinear sequence-to-contact map relationships, but lack interpretability and fail to encode biophysical processes like loop extrusion. Conversely, molecular dynamics simulations of loop extrusion offer mechanistic insights but are challenging to scale for large genomic regions or whole genomes. Thus, a hybrid approach is needed to combine biophysical principles with machine learning for interpretable and scalable modeling.

We propose the Differentiable Loop Extrusion Model (dLEM), a minimal mechanistic model using one-dimensional positional parameters with clear physical meaning. dLEM captures key features of TADs and fits experimental Micro-C data to recapitulate contact maps. Comparing its positional parameters to genomic tracks reveals elements of the loop extrusion machinery and confirms interactions between transcription and loop extrusion. Notably, dLEM uncovers potential roles for chromatin remodelers, such as CHD7 and PHF8, in loop extrusion regulation.

dLEM bridges biophysical modeling and machine learning by integrating physically meaningful parameters, achieving significant dimensionality reduction. We leverage the biophysical model to design a highly parameter and memory efficient deep-learning architecture, deep dLEM, . This model predicts high-resolution contact maps from 1D chromatin and sequence data, generalizing effectively to unseen cell types. This synergy of

mechanistic modeling and machine learning yields accurate predictions, novel biological insights, and an efficient, interpretable framework for chromatin organization studies.

Arjuna Subramanian: Designing novel-to-nature protein building blocks for synthetic cells with foldtuned language models

Nature has likely sampled only a fraction of all protein sequences and structures allowed by the laws of biophysics. This sparse sampling suggests untapped reservoirs of protein diversity to leverage for bottom-up assembly of synthetic cells with novel genotypes and phenotypes. However, the combinatorial scale of amino-acid sequence- space complicates attempts to traverse the full protein sequence-structure map and locate novel versions of essential machinery desired for synthetic cells, such as signaling domains, transcription factors, and cytoskeletal components. In this work, we transform artificial intelligence (AI) protein language models (pLMs) into probes to explore far-from-natural regions of sequence-space and discover sequences which code for these crucial targets while exhibiting zero or near-zero detectable homology to wild-type examples. Specifically, we incorporate pLM, AI structure prediction, and structure-based search techniques into a novel “foldtuning” strategy that pushes a pretrained pLM to simultaneously maximize structural similarity to a target protein domain and minimize sequence similarity to the natural protein universe. We scale this approach to build a library of foldtuned pLMs covering >700 modular domains relevant for synthetic biology applications, including GPCRs/G-proteins, cytokine families, kinesin-like motor heads, antimicrobial peptides, and composable DNA-binding and cell surface receptor domains. Candidate proteins generated by foldtuned pLMs reflect distinctive new “rules of language” for innovating around critical minimal structural motifs, preserving familiar structures (median model performance: 56.5% of generated proteins match the target fold) while bearing little-to-no sequence resemblance to known proteins (median model performance: 21.1% of structural matches exhibit no detectable sequence homology against UniRef50). Preliminary mass-spectrometry experimental validation on one target of interest for designing synthetic signal transduction pathways, the tyrosine-kinase- and small-GTPase-regulating SH3 domain, shows detectable *in vitro* expression of 1619/2593 (62.4%) of variants proposed by an SH3-foldtuned pLM – including 505/804 (62.8%) of zero-sequence-homology variants – in a cell-free transcription-translation system. Separately, high-throughput screens reveal specific binders for the insulin receptor among variants from an insulin-foldtuned pLM and additionally suggest novel ligand-receptor interaction motifs. Taken overall, foldtuning presents the capacity to readily access viable areas of protein sequence-space untouched by evolution, enabling redesign and reconstitution of novel-to-nature synthetic biological systems.

Razeen Shaikh: Optimal performance objectives in the highly conserved bone morphogenetic protein signaling pathway

Throughout development, complex networks of cell signaling pathways drive cellular decision-making across different tissues and contexts. The transforming growth factor β (TGF-β) pathways, including the BMP/Smad pathway, play crucial roles in determining cellular responses. However, as the Smad pathway is used reiteratively throughout the life cycle of all animals, its systems-level behavior varies from one species or context to another, despite protein sequences and pathway connectivity remaining almost perfectly conserved. For instance, some cellular systems require a rapid response, while others require high noise filtering. Our work examines how the BMP-Smad pathway balances trade-offs among three such systems-level behaviors, or “Performance Objectives (POs)”: response speed, noise amplification, and the sensitivity of pathway output to receptor input. Using a mathematical model of the Smad pathway fit to human cell data, we show that varying non-conserved parameters (NCPs), such as protein concentrations, the Smad pathway can be tuned to emphasize any of the three POs and that the concentration of nuclear phosphatase has the greatest effect on tuning the POs. However, due to competition among the POs, the pathway cannot simultaneously optimize all three, but at best must balance trade-offs among the POs. We applied the multi-objective optimization concept of the Pareto Front, a widely used concept in economics, to identify optimal trade-offs among the various requirements. We show that the BMP pathway efficiently balances competing POs across species and is largely Pareto optimal. Finally, we validate the relationship between relative phosphatase levels and approximate BMP signaling response time for three biological systems: Human Aorta, Zebrafish embryo and *Drosophila* embryo.

Our findings reveal that varying the concentration of NCPs allows the Smad signaling pathway to generate a diverse range of POs. This insight identifies how signaling pathways can be optimally tuned for each context.

Robyn Shuttleworth: Exploring the role of inflammatory cells and the extracellular matrix during tissue reprogramming

Cellular reprogramming, a process in which cells are rejuvenated, has increased in popularity in the last 20 years. Reprogramming has been explored in several contexts, from reversing signs of aging to reprogramming diseased cells to their previously healthy state. Although extensive work has been done in optimizing the biological process of cellular reprogramming, there have been fewer advances in tissue reprogramming. Tissue reprogramming involves additional complications with respect to its cellular counterpart, like the presence of an external environment, known as the extracellular matrix (ECM), and the cellular diversity in tissue composition. The ECM provides an underlying biomechanical and biochemical network, facilitating both cell-cell and cell-ECM communication. It is known that the ECM plays a large role during many normal physiological processes. Specifically, due to its role in both damaged (aged) and recovered (wound healing) tissue, we hypothesize that the ECM will also contribute to the rejuvenation of tissues. A key aspect of these systems is the role of inflammation and the associated cell types, for example myofibroblasts and senescent cells. We have constructed a computational model that looks to investigate the contribution of such cells to tissue rejuvenation. We discern these two cell types by their behaviors and interaction rules with the ECM and we look to uncover what properties of inflammatory cells play the most vital roles during whole tissue rejuvenation. The model includes cell- cell and cell-ECM interactions. We use this model to explore not only the key properties of cells and ECM required for optimal tissue rejuvenation, but also the spatial dynamics that naturally arise inside cell-dense tissues. The structure of a tissue is highly dependent on its location and organ-type, and we account for this by modeling rejuvenation over a range of different ECM structures, highlighting the adaptability of such a model. This model aims to emphasize the crucial role played by the ECM and inflammatory cells by providing predictions of tissue-specific inflammation-triggered rejuvenation, highlighting the key parameters and behaviors for optimal tissue reprogramming.

Shunzhi Wang: De novo design of quasi symmetric protein particles

The quasi-equivalence geometric principle elucidates how certain viral capsids tessellate icosahedral-like shells with a combination of pentagons and hexagons. Identical protein subunits will have to undergo subtle conformational changes, in order to assume multiple non-equivalent local environments that break perfect icosahedral symmetry. Therefore, the rational design of quasi-symmetric capsids through self-assembly remains to be an unsolved challenge. Here, we describe a computational design strategy to generate two-component quasi-symmetric protein capsids through positively curved hexagonal motifs. Electron microscopy characterizations reveal that capsid sizes can be tailored from 40 nm to over 200 nm, and reveal that particle closure is facilitated by pentagonal defects formation as intended. These de novo particles show robust assembly and genetic fusion both in vitro and within living cells, paving the way for potential applications such as de novo virus-like particles and genetically encoded bioprobes. Together, our results lay the groundwork for programmable quasi-symmetric architectures as a new class of protein nanomaterials for a wide range of biomedical applications.

Vivek Behera: Conserved Principles of Spatial Biology Define Tumor Heterogeneity and Response to Immunotherapy

Immune checkpoint blockade (ICB) has transformed cancer therapy and is now approved in every type of cancer, yet it remains difficult to identify the ~ 20% of patients that will benefit from such therapy. Explainable models that can predict response have the potential to spare ‘responders’ from toxic chemotherapies while identifying alternative options for ‘non-responders’. The only widely approved ICB biomarker of response, PD-L1, treats TME cells as individual components and is only weakly predictive of outcomes. Recent studies have suggested that biomarkers built on cell-to-cell spatial proximity, such as tumor cell to CD8+ T cell distance, can improve upon PD-L1 in predicting ICB efficacy. Yet there is a lack of generalized approaches for identifying tumor spatial biology relationships relevant to ICB response, especially when considering that high- dimensional

spatial biology techniques capture information across many physical length scales, ranging from tissue-level to nearly single-cell-level.
Through statistical covariation analysis of spatial transcriptomics data from 96 tumors across twelve tumor types, we have established a general approach for studying tumor spatial biology. The basis of this approach is our identification of a conserved unit of pan- tumor spatial organization that we termed 'Spatial Groups'. Spatial Groups (SGs) are multi-cellular units demonstrating coordinated transcriptional programs contained within spatial domains ranging in size from tens of microns (dozens of cells) to mm-scale (large regions of a tumor). SGs are hierarchically organized such that large-scale SGs exert contextual field effects on their nested small-scale SGs. For example, we found that T cell presence in large-scale SGs impacted metabolic pathway shifts occurring within nested medium-scale SGs, which in turn impacted cell adhesion programs within nested small-scale SGs.
We found that SGs revealed immune-related spatial biology patterns that we hypothesized would serve as a genome-wide spatial biomarker of ICB response. To test this idea, we performed a retrospective study of 16 patients with metastatic non-small cell lung cancer treated with ICB or chemotherapy-ICB in the frontline setting. We found that our SG-derived biomarker, unlike PD-L1 status, was highly significant in predicting progression-free survival after therapy initiation. Additionally, analysis of SGs in non- responder patients provided insights into patient-specific spatial biology architectures limiting ICB efficacy. In summary, our discovery of Spatial Groups provides a general framework for using spatial biology to identify pan-tumor ICB response.

Anat Bren: Tradeoffs in bacterial physiology determine the efficiency of antibiotic killing
Antibiotic effectiveness depends on a variety of factors. While many mechanistic details of antibiotic action are known, the connection between death rate and bacterial physiology is poorly understood. A common observation is that death rate in antibiotics rises linearly with growth rate; however, it remains unclear how other factors, such as environmental conditions and whole-cell physiological properties, affect bactericidal activity. To address this, we developed a high-throughput assay to precisely measure antibiotic-mediated death. We found that death rate is linear in growth rate, but the slope depends on environmental conditions. Growth under stress lowers death rate compared to nonstressed environments with similar growth rate. To understand stress's role, we developed a mathematical model of bacterial death based on resource allocation that includes a stress-response sector; we identify this sector using RNA-seq. Our model accurately predicts the minimal inhibitory concentration (MIC) with zero free parameters across a wide range of growth conditions. The model also quantitatively predicts death and MIC when sectors are experimentally modulated using cyclic adenosine monophosphate (cAMP), including protection from death at very low cAMP levels. The present study shows that different conditions with equal growth rate can have different death rates and establishes a quantitative relation between growth, death, and MIC that suggests approaches to improve antibiotic efficacy.

Purushottam Dixit: The ability to sense the environment is heterogeneously distributed in cell populations
Channel capacity of signaling networks quantifies their fidelity in sensing extracellular inputs. Low estimates of channel capacities for several mammalian signaling networks suggest that cells can barely detect the presence/absence of environmental signals. However, given the extensive heterogeneity and temporal stability of cell state variables, we hypothesize that the sensing ability itself may depend on the state of the cells. In this work, we present an information-theoretic framework to quantify the distribution of sensing abilities from single-cell data. Using data on two mammalian pathways, we show that sensing abilities are widely distributed in the population and most cells achieve better resolution of inputs compared to an 'average cell'. We verify these predictions using live- cell imaging data on the IGFR/FoxO pathway. Importantly, we identify cell state variables that correlate with cells' sensing abilities. This information-theoretic framework will significantly improve our understanding of how cells sense in their environment.

Julia Salzman: Dual phenotype and functional prediction with genome-free statistical systems genomics
Genomic data is now acquired at a scale providing an unprecedented opportunity to link systems-level gene and variant expression to phenotypes across the tree of life: from single cell RNA-seq to microbial DNA. Today,

analysis of this rich data are primarily conducted through multistep analytic procedures that entail assembly and, and (pseudo)alignment. This results in rich raw data filtered through the lens of assembly algorithms and or alignment to annotated gene models, attenuating, and sometimes distorting the potential to discover links between genotype and phenotype. We have recently introduced SPLASH, a unified, highly efficient, statistical, reference-free, meta-data free framework to analyze biologically regulated genetic variation (RNA or DNA). SPLASH operates directly on raw sequencing data to discover myriad forms of functional genetic variation, from new CRISPR repeats and associated Cas proteins, previously unknown transposable elements, to examples in eukaryotes such as genes and new immune-regulated gene families including those lacking homology to any annotated in current NCBI databases. Here, we extend SPLASH from a completely de-novo, meta-data free algorithm to metaSPLASH, which performs dual prediction of how genotype is linked to phenotype. metaSPLASH uses a sparse selection of SPLASH-identified sequences to predict phenotype, such as but not limited to virulence, serotype, antifungal resistance or antibiotic resistance. metaSPLASH uses modern machine learning models to identify sequences, such as but not limited to genes and mobile genetic elements, that predict phenotypes, such as anti-microbial resistance, strains and serotypes, viral host range, and in eukaryotes, cell type. Dually, metaSPLASH assigns functional prediction to the features driving predictions. metaSPLASH performs supervised prediction and unsupervised clustering using a sparse set of sequence features with no human guidance or reference genome. A snapshot of examples include prediction of capsular serotypes in S. pneumoniae bypassing time-consuming experimental tests, rediscovering known resistance targets in microbial isolates. For example, In Candida tropicalis, metaSPLASH predicts fluconazole resistance accuracy improved over state-of-the-art and recovers, de novo, the target of azole resistance to be ERG11 (Sterol 14-demethylase, the known target of fluconazole), and predicting new targets that modify this resistance. Without tuning, metaSPLASH obtains similar results in many other bacterial species; E. Faecium, metaSPLASH improves prediction of resistance including predicted synergistic function of several efflux pumps. In E. Coli, metaSPLASH models generalize resistance prediction in wild isolates with, to our knowledge, unmatched precision. In summary, metaSPLASH is a general framework can be applied to discovery of systems level genotype-phenotype relationships, including multifactorial genetic interactions, an area of current work. We anticipate metaSPLASH will contribute to de-orphanizing gene functions by direct, statistical models for systems biology that link massive- scale genomic and phenotypic measurements across the tree of life.

Jacob Hanna: Synthetic Ex Utero Embryogenesis: from Naive Pluripotent Cells to Complete Developmental Models

The identity of somatic and pluripotent cells can be epigenetically reprogrammed and forced to adapt a new functional cell state by different methods and distinct combinations of exogenous factors. The aspiration to utilize such in vitro reprogrammed pluripotent and somatic cells for therapeutic purposes necessitates understanding of the mechanisms of reprogramming and differentiation and elucidating the extent of equivalence of the in vitro derived cells to their in vivo counterparts. In my presentation, I will present my group's recent advances toward understanding these fundamental questions and further detail our ongoing efforts to generate developmentally unrestricted human naive pluripotent cells with embryonic and extra-embryonic developmental potential. I will expand on new avenues for utilizing custom made electronically controlled ex utero platforms and optimized conditions for growing natural mammalian embryos ex utero for extended periods capturing development from pre-gastrulation until advanced organogenesis, for better studying of stem cell transitions during embryogenesis and organogenesis. I will detail how the latter platforms offered an exclusive technical platform to demonstrate and unleash the self-organizing capacity of mouse naïve PSCs to generate post-gastrulation synthetic Bona Fide synthetic whole developmental models with both embryonic and extraembryonic compartment ex utero, as well as our ability to extend these findings with naïve human PSCs and generate complete structured day 14 human developmental models and beyond. Collectively, I will be highlighting prospects for new platforms for advancing human disease and embryogenesis developmental modelling.

Kate Galloway: Proliferation history and transcription factor levels drive direct conversion

The sparse and stochastic nature of reprogramming has obscured our understanding of how transcription factors drive cells to new identities. To overcome this limit, we developed a compact, portable reprogramming system that increases direct conversion of fibroblasts to motor neurons by two orders of magnitude. We show that subpopulations with different reprogramming potentials are distinguishable by proliferation history. We show that specific subpopulations with lower levels of transcription factor expression reprogram at higher rates, highlighting the importance of cell state in guiding the activity of transcription factors. Importantly, we do NOT rely on transcriptional profiling to make claims about how transcription factor levels influence reprogramming events. Instead, we used synthetic biology tools to develop live reporters of transcription factor protein levels, allowing us to directly examine how individual transcription factors influence conversion events. By controlling for proliferation history and titrating each transcription factor, we find that conversion correlates with levels of the pioneer transcription factor Ngn2, whereas conversion shows a biphasic response to Lhx3. Overall, our results show that proliferation history and transcription factor expression combine to drive cell-fate transitions. Harnessing these insights, we design improved reprogramming cocktails for direct conversion of adult human fibroblasts to motor neurons. Increasing the proliferation rate of adult human fibroblasts generates morphologically mature, human induced motor neurons at high rates. Low rates of direct conversion have previously limited the potential for central nervous system (CNS) applications. Using compact, optimized, polycistronic cassettes, we generate motor neurons that graft with the murine central nervous system, demonstrating the potential for *in vivo* therapies.

Gavin Schlissel: Single-molecule analysis of morphogen diffusion revealed a novel mechanism of evolutionary diversification

Phenotypic differences among individuals and between species often involve quantitative variation in conserved anatomical features. Whereas sharp evolutionary differences in animal anatomy are often driven by large-effect alleles that control major signaling pathways, quantitative variation in traits like body size or anatomical shape is genetically complex, involving dozens of causative alleles that lack an obvious mechanistic connection. Using single-molecule microscopy to directly observe the diffusion of extracellular signaling proteins, we recently discovered a surprising mechanism for morphogen patterning in which morphogens transiently “slide” along cell surfaces, and form long-range gradients by “jumping” from membrane-to-membrane. Under this model, the kon and koff for receptors, co-receptors and the extracellular matrix can all affect a morphogen’s extracellular diffusion rate, revealing novel evolutionary strategies to regulate developmental signaling gradients in tissue- or organism-specific ways. Now, we have identified Hedgehog- interacting proteins and extracellular matrix glycoforms that can modulate dynamic interactions between Hedgehog and the extracellular matrix and affect the range of Hedgehog signaling gradients. These Hedgehog-interacting proteins are associated with a range of vertebrate phenotypes, including human bone length and craniofacial morphology in humans and dogs. We anticipate that similar mechanisms affect the diffusion range of all extracellular signaling molecules, pointing to a broad new class of biochemical mechanisms that could regulate the emergence of quantitative phenotypic variation across evolution.

Jennifer Oyler-Yaniv: Between TNFα and Proteostatic Stress Drives Cell Death and Guard Immunity

The production and sensing of type I interferons (IFN-I) are critical for antiviral defense, yet most virus-infected cells do not produce IFN-I or upregulate IFN-stimulated genes. Using quantitative proteomics and global protein synthesis measurements, we show that productive viral infection globally down-regulates protein synthesis, restricting the IFN response. Guard immunity, which responds to disruptions in essential cellular processes, might compensate for the lack of IFN-I response by rapidly killing infected cells. However, non-pathological stressors can also disrupt proteostasis, making it unclear how cells decide to trigger guard immunity. We hypothesized that TNFα, produced by macrophages, provides a contextual signal allowing specificity. Using live-cell fluorescence microscopy and mathematical modeling, we showed that TNFα synergizes with the decay of the anti-apoptotic protein c-FLIP to induce cell death and prevent viral spread. Our findings demonstrate that TNFα contextualizes the loss of proteostasis as non-sterile, enabling the activation of guard immunity to counteract viral infection.

Zev Gartner: MAGIC matrices: freeform bioprinting materials to support complex and reproducible organoid morphogenesis

Organoids are powerful models of tissue physiology, yet their applications remain limited due to their relatively simple morphology and high organoid-to-organoid structural variability. To address these limitations we developed a soft, composite yield-stress extracellular matrix that supports optimal organoid morphogenesis following freeform 3D bioprinting of cell slurries at tissue-like densities. The material is designed with two temperature regimes: at 4 ÀöC it exhibits reversible yield-stress behavior to support long printing times without compromising cell viability. When transferred to cell culture at 37 ÀöC, the material cross-links and exhibits similar viscoelasticity and plasticity to basement membrane extracts such as Matrigel. We first characterize the rheological properties of MAGIC matrices that optimize organoid morphogenesis, including low stiffness and high stress relaxation. Next, we combine this material with a custom piezoelectric printhead that allows more reproducible and robust self-organization from uniform and spatially organized tissue ‚Äúseeds.‚Äù We apply MAGIC matrix bioprinting for high-throughput generation of intestinal, mammary, vascular, salivary gland, and brain organoid arrays that are structurally similar to those grown in pure Matrigel, but exhibit dramatically improved homogeneity in organoid size, shape, maturation time, and efficiency of morphogenesis. The flexibility of this method and material enabled fabrication of fully 3D microphysiological systems, including perfusable organoid tubes that experience cyclic 3D strain in response to pressurization. Furthermore, the reproducibility of organoid structure increased the statistical power of a drug response assay by up to 8 orders-of-magnitude for a given number of comparisons. Combined, these advances lay the foundation for the efficient fabrication of complex tissue morphologies by canalizing their self-organization in both space and time.

Keynote: Sydney Shaffer | The unseen history of cells in cancer evolution and treatment

Understanding how cells remember and respond to past experiences is crucial for treating diseases like cancer, where cellular adaptations can lead to therapy resistance. Here, we develop tools to track cellular memory and lineage relationships across multiple scales - from individual cells responding to drugs in vitro to complex tissue organization in vivo. Using single-cell RNA sequencing combined with cellular barcoding, we first demonstrate that resistance states can persist across multiple cell divisions, revealing how initial drug exposure creates lasting cellular memory that influences responses to subsequent treatments. We then extend these lineage tracing principles to human tissues, where we find unexpected relationships between different cell types at the gastroesophageal junction and during cancer development. By linking cellular memory to clonal dynamics across both systems, we provide a framework for understanding how cell state inheritance shapes disease progression and treatment response.

Keynote: John Albeck | Visualizing metabolic heterogeneity and its integration with growth factor signaling in living cells

Cellular energetic status depends on a balance of nutrient catabolism and biosynthesis, orchestrated by a regulatory network centered on the kinases mTOR and AMPK. This network is integrated with growth factor signaling, including RAS/ERK and PI3K/AKT, which operate dynamically as they adjust to a cell’s changing nutrient availability and regulatory microenvironment. We have developed a live-cell imaging approach to probe metabolic and signaling states in parallel, combining biosensors for ATP concentration, ADP/ATP ratio, and glycolytic intermediates with reporters for AMPK, AKT, mTOR and ERK kinase activity. With this approach, we observe a range of dynamic activities, which vary depending on the level of metabolic stress. Our results indicate that proliferative signaling programs can be temporally gated by phases of catabolic activity and suggest that a loss of intracellular pathway coordination may contribute to cellular dysfunction under pathological conditions.

Ashley Laughney: Mapping the Emergent Functions of Proteins in Multicellular Systems

Our lab addresses the critical challenge of understanding *genotype-to-phenotype* relationships in evolving multicellular systems such as cancer. A major focus is on pleiotropic proteins that exhibit distinct and sometimes opposing functions, influencing both tumor evolution and immune dynamics. One key area of our research explores chromosomal instability (CIN), a hallmark of metastatic cancers, which perpetuates activation of innate

immune signaling pathways and drives cancer progression and metastasis. To investigate these phenomena, we developed *ContactTracing*, an innovative computational tool (*Nature*, 2023) that predicts the influence of cell-cell interactions across tumor microenvironments in their native, in vivo contexts. Using this tool, we identified ligands emerging from Endoplasmic Reticulum (ER) stress responses as mediators of immune suppression in CIN-associated tumors. Targeting these pathways significantly suppressed metastasis in syngeneic models of melanoma, breast, and colorectal cancers, validating *ContactTracing* as a transformative methodology and identifying a targetable mediator of metastasis.

Building on these findings, we demonstrated that STING, a key immune signaling protein, exhibits dichotomous roles in cancer depending on the chronicity of its activation—acting as a tumor suppressor early in disease progression but later aiding metastasis. To address the complexity of this multi-functional protein, we developed *SatSeq*, a cutting-edge platform combining deep saturation mutagenesis with single-cell transcriptomics. *SatSeq* systematically maps sequence-function relationships of proteins, enabling us to dissect their roles across diverse biological contexts and evolutionary processes. By analyzing how specific protein regions influence cellular functions, *SatSeq* provides a detailed blueprint of protein activity and drug interactions, offering unprecedented insights into how multifunctional proteins can be precisely targeted.

We applied SatSeq to STING, revealing specific regions that govern its tumor-suppressive and tumor-promoting functions. This has identified druggable targets to inhibit STING's harmful roles while preserving its beneficial effects. By integrating computational and experimental tools, our work advances the understanding of pleiotropic protein functions, providing a scalable framework for targeting complex biological processes in cancer and beyond.

Maria Chikina: Towards interpretable sequence to function models

Deep learning approaches have demonstrated strong predictive power in modeling functional genomic readouts, such as chromatin accessibility and gene expression, directly from DNA sequences. However, existing methods often lack intrinsic interpretability, requiring computationally intensive post hoc analyses to understand the contributions of sequence features.

We have recently introduced tiSFM (totally interpretable sequence-to-function model), a novel neural network architecture that combines high predictive accuracy with built-in interpretability. Unlike traditional deep learning models, tiSFM works directly on the transcription factor (TF) binding layer through a series of interpretable transformations. The model incorporates trainable pooling and interaction layers, allowing for nuanced modeling of motif aggregation and TF-TF interactions, while maintaining biological interpretability throughout the entire network.

We apply tiSFM to a hematopoietic differentiation dataset and show that it outperforms state-of-the-art convolutional neural networks (CNNs) designed for this task, while using significantly fewer parameters. TiSFM accurately predicts open chromatin regions across different immune cell types and identifies context-specific activities of key transcription factors. In addition, tiSFM's internal parameters—such as pooling weights and interaction matrices—offer biologically meaningful insights into regulatory mechanisms. For example, the model reveals differential TF binding dependencies, reflecting underlying biochemical processes like cooperativity and saturation.

We extend the tiSFM approach to model flexible grammar where TF-TF interactions depend on position and orientation with entirely interpretable model parameters. We demonstrate that our approach can match the performance of power residual CNN networks while capturing complex interaction features such as dependence on DNA helical periodicity that are apparent without any post hoc analysis. We highlight some challenges of interpretable models including controlling optimization to obtain reproducible and sparse representations.

Mohamad Abedi: Expanding the cell signaling space with more than a thousand de novo designed agonists

Cytokines and growth factors modulate cell behavior by bringing together specific cell surface receptor subunits, triggering phosphorylation of the receptor intracellular domains and initiating downstream signaling cascades. New signaling molecules that bring together novel combinations of receptor subunits in defined clustering geometries could elicit new cell type-specific responses, but this area remains largely unexplored. Here, we

present a de novo protein design approach for systematically generating a wide range of potential novel agonists by de novo designing and fusing individually designed receptor-binding modules either with flexible linkers or rigid protein-protein interfaces. We combined twenty-eight binding modules to generate over a thousand potential agonists. Among these, 151 demonstrated signaling activity, and we conducted an in-depth analysis of eighteen in peripheral blood mononuclear cells (PBMCs). These produce downstream activation patterns distinct from those of known endogenous ligands by pairing receptor subunits that are not naturally associated. We find that the STAT signaling signature is a much stronger indication of agonist functional activity than the targeted receptor identity. These novel agonists, along with the insights they provide into the signaling potential of human cells, should contribute to applications ranging from enhancing immunotherapy strategies to advancing stem cell differentiation for regenerative medicine.

Alexander Davies: Live-cell whole tissue models reveal sources of dynamic signaling heterogeneity and single cell drug response variation in the metastatic niche

It is well established that cancer cell behavior results from the complex interplay of mutated or otherwise phenotypically reprogrammed cells with the surrounding host microenvironments. However, our current understanding of this process is often inferred by collecting data at snapshots in time, limiting our view of underly tissue and molecular-level dynamics at the root of disease pathogenesis and drug response. Our objective was to address this limitation by combining ex vivo tissue culture, fluorescent biosensors (e.g., ERK and AKT), and serial imaging. Jointly, these elements form a live-cell dynamic model of cancer metastasis in the lung, the lungSITE (Serial Imaging of Tumor and microEnvironment) model, that enables the extraction of quantitative spatiotemporal relationships and the identification of causal links between tumor-host and tumor-tumor interactions that govern cell fate. Using this novel model, we found that metastatic osteosarcoma signaling dynamics were highly sensitive to positional and temporal variation in the physical microenvironment of the lung. Tumor cells at the host interface frequently displayed elevated ERK and AKT, whereas cells in the tumor core displayed baseline low level signaling activity. Initial single cell drug responses strongly correlated with position, whereby signaling ‘high’ tumor-host interface cells persisted longer than cells in the tumor core. Strikingly, a rapid adaptive response was observed in a fraction of the tumor-host interface cells following death of cells in tumor core, eliciting a drug resistance phenotype. We discovered that initiation of cell death in the core released tumor-derived growth factors into the local microenvironment. These signals, combined with those derived from the host, upregulated signaling in the cells at tumor-host interface and induced a pro-survival/ resistance response that could be overcome by addition of receptor tyrosine kinase inhibitors. Together, our results demonstrated the utility of a novel dynamic live-cell tissue model, the lungSITE model, to quantitatively measure and understand tumor signaling dynamics and behaviors within the context of the lung metastatic niche. Data obtained from this model provided new insights into how spatial position and temporal response can influence signaling dynamics, specifically in osteosarcoma lung metastasis, to create intratumoral signaling heterogeneity and consequently single- cell drug response variation.

Tiffany Zhou: Modeling Synchronized Lysis of Bacterial Populations in Spatially Extended Environments

Synthetic gene circuits are engineered networks of genes that allow precise control of gene expression to produce a desired function in a cell. They have many promising practical applications such as targeted drug delivery, biosensing, industrial chemical production, and environmental remediation. Mathematical modeling of gene circuits is crucial for the forward-engineering of increasingly sophisticated circuit behaviors necessary for these applications. However, many existing models assume circuits operate in small (micron-scale) cell populations and homogeneous, high-nutrient environments. Predictions from these models may not translate well into practical applications because real-world environments are often highly complex and heterogeneous in time and space. Therefore, it is important to characterize gene circuits in more diverse conditions to understand how they function outside of an ideal setting. In our work, we have revised the model of an existing synthetic gene circuit, the Synchronized Lysis Circuit, to better resemble the conditions seen with cell populations growing in spatially extended (millimeter-scale) microfluidic chambers. Specifically, we implemented the presence of dead cell debris, multiple interacting colonies, and larger growth chamber dynamics into the model. As a result, we observed distinct behaviors that were not seen in previous models of the Synchronized Lysis Circuit.

Bryan Duoto: Quantitative RNA-Templated Size Control of Plant Virus-Like Particles
Potato Virus X (PVX) is a well-characterized filamentous plant virus known for its unique biomolecular properties and here we explore size and shape engineering in response to genomic RNA secondary structures. PVX virions are flexible, thread-like bodies measuring 500 nanometers in length and 15 nanometers in diameter; comprised of 1300-1350 helically folded identical coat proteins enclosing an m7g-capped and poly(a)-tailed 6.4 knt viral RNA. The coat protein-binding stem-loop element, stem-loop 1 (SL-1), at the 5' end of PVX's genome has proven to be the minimal origin of assembly (OAS) signal necessary for nucleation of the coat protein to form viral particles. This feature can be exploited to drive virus-like particle (VLP) formation around partial genomic RNA or foreign RNA's containing the OAS sequence. To this end, we have developed an RNA-templated PVX VLP platform that is capable of forming truncated or elongated particles based on input RNA size that retain the flexible morphology of the original virus. By controlling the size of VLPs through different sizes of RNA, we effectively demonstrate the feasibility of using PVX as a "payload agnostic" nanocarrier for gene delivery. Our preliminary studies indicate that the encapsulation of single RNA molecules of any tested size is possible, as well as the formation of doublet, triplet, and multiplet VLPs exhibiting higher-order proportions that exceed the dimensions of the largest naturally occurring viruses. Lastly, protein-coding mRNA can be packaged enabling protein expression in mammalian cells. All together the PVX platform opens a new modality for gene delivery and editing.

Rikki Garner: Tissue fluidity: a double-edged sword for multicellular patterning
The organization of cells into spatial patterns is a fundamental aspect of multicellularity. One major mechanism underlying tissue patterning is adhesion-based cell sorting, in which a heterogeneous mixture of different cell types spontaneously separates into distinct domains based on cell type-specific differences in adhesion protein expression. While there is ample evidence that cells can sort by adhesive compatibility, much less is known about how the biomechanical properties of the tissue control the proficiency of cells to sort. Here, we identify tissue fluidity—the extent to which cells can move freely within a tissue—as a critical regulator of adhesion-based sorting. First, we develop a first-principles biophysical model integrating both tissue fluidity and adhesion-based sorting, and demonstrate that this model can quantitatively reproduce experimentally-measured sorting dynamics in a fibroblast cell culture assay. Strikingly, we go on to show that altering tissue fluidity by any mechanism in the model (i.e., by varying cell migration, cell–cell adhesion, or tissue viscosity) leads to substantial changes in the rate or accuracy of sorting (or both). We further demonstrate that the balance between cell migration, which acts the fluidize the tissue, and homotypic cell–cell adhesion, which acts to solidify the tissue, sensitively tunes a fundamental trade-off between the rate and accuracy of sorting – such that sorting can only occur when migration and adhesion are tightly coupled. To interrogate the consequences of this coupling requirement experimentally, we varied adhesion protein expression levels in the cell sorting assay, and fit the observed sorting dynamics to infer the energies of migration and adhesion across experimental conditions. Intriguingly, we find that cells seem to naturally scale their motility strength with their adhesion protein expression – thereby maintaining a nearly optimal fluidity for sorting. Overall, our results indicate that tissue fluidity must be precisely regulated in order for sorting to occur, and that cells may have evolved a mechanism to naturally co-regulate their mechanical properties in order to maintain this optimal fluidity.

Stephanie Hartel: Quantifying Nutrition: A Metabolomic Perspective on Food Diversity and Classification
Nutrition is fundamental to organismal health, yet our understanding of the compounds comprising foods is surprisingly minimal. Currently, less than 1% of the chemical constituents in foods have been identified, a gap that severely constrains our ability to design diets, provide personalized dietary recommendations (e.g., targeting the gut microbiome or specific health conditions), and predict the physiological effects of individual or combinations of foods. Existing dietary frameworks, like the United States Department of Agriculture's (USDA) MyPlate, primarily classify foods based on phenotypic traits—such as grouping all "leafy greens" together—despite possessing quantitative data on a limited subset of food compounds. But do these broad, phenotypic-based groupings adequately capture food diversity? Recent evidence highlights the limitations of such outdated food groupings, particularly in regard to predicting outcomes. For instance, dietary fibers, a nutrient category

including non-digestible polysaccharides, can elicit both diverse responses from engineered microbial consortia and varying levels of gut inflammation in vivo. Thus, while dietary guidelines broadly recommend increasing fiber intake, not all fiber sources yield equivalent effects. To address these gaps, we leveraged untargeted metabolomics to profile 94 whole foods and develop a novel, quantitative food classification system based on metabolite composition. Using spectral inference, we uncovered a hierarchical landscape of food relationships defined by their metabolite profiles, offering a data-driven alternative to phenotypic categorization. This approach allows us to identify biochemical similarities among foods that phenotypic traits obscure. To contextualize these findings, we applied the same analysis to the USDA's quantitative dataset. Surprisingly, our analysis revealed that MyPlate's current phenotype-based categorizations fail to reflect the diversity implied by their own quantitative dataset. For instance, half of the 10 most diverse foods from their dataset identified through our analyses were classified as "leafy greens" in MyPlate, highlighting the limitations of phenotype-based groupings. Our findings demonstrate profound differences in food categorization when using enriched metabolomic data compared to the USDA's limited profiles. This novel approach provides a more nuanced understanding of food classification and its functional implications, establishing a foundation for future studies exploring the relationship between diet and health.

Oliver Inge: Combinatorial BMP4 and Activin direct choice between alternate routes to endoderm during human gastrulation

Lineage specification requires the proper interpretation of multiple signalling cues which vary in concentration and duration. How the combination and history of signals a cell is exposed to influences a fate decision remains poorly understood. In this study we use hESCs as a tractable model system to explore how combinations of cues guide state transitions during gastrulation. Using single-cell transcriptomics and live-cell imaging of engineered hESCs expressing endogenous cell state reporters, we reconstructed developmental lineages and obtained single-cell measurements of fate specification dynamics during gastrulation. We found that definitive endoderm, one of the three germ layers, arises from two distinct developmental trajectories: a direct route from pluripotency, and an indirect route via a mesoderm progenitor state. Furthermore, by modulating the signalling input we found that the relative concentration of Activin and BMP4 controls the choice between alternate trajectories to endoderm. Our findings reveal a lineage convergence event during human gastrulation with multiple routes existing to definitive endoderm dictated by the combination of signalling cues presented. This work shows that the combination of cues a cell is exposed to not only directs the final fate outcome it assumes, but the developmental route taken.

Tavis Reed: Enhancing Context-Specific Interactome Mapping: Spatially Resolved Global Protein Interactome Networks and Interactome Homology in Viral Infections

Dynamic protein-protein interactions (PPIs) regulate cellular activity during both homeostatic and perturbed states. During viral infections, viral proteins remodel these networks by introducing virus-virus and virus-host interactions and altering host-host interactions to enable viral replication within host cells. This remodeling occurs across both time and space, with infection-induced protein translocation and subcellular location- specific interactome changes playing important roles in infection outcomes. Mapping the spatiotemporal regulation of PPI networks at a global level is an important challenge for identifying key proteins and pathways that play critical roles during viral infection and can serve as potential antiviral targets. To address this, we extended our recently developed Tapioca framework, which integrates TPCA data with protein physical properties and functional networks derived from Bayesian integration of thousands of -omics datasets to predict context-specific global PPI networks. This framework has already identified a novel broad-spectrum proviral factor for several herpesviruses. Building on this, we developed two new methodologies to overcome key challenges in mapping spatially resolved PPI networks and translating them into biological insights. First, we integrated subcellular fractionation into the Thermal Proximity Coaggregation (TPCA) mass spectrometry workflow, enabling the generation of spatially resolved PPI networks. Applying this to herpes simplex virus type 1 (HSV-1) infection revealed a network of nuclear-translocating host chaperone proteins interacting with viral proteins throughout infection, as well as temporally regulated interactions with ribosomal subunits and splicing factors.

We demonstrated that this network is recruited to HSV-1-induced VICE domains—structures of previously unknown function—which we now show may play a role in modulating translation during infection. Second, we developed a deep learning framework to systematically compare global PPI networks by introducing the concept of “interactome homology,” which assesses whether proteins share interaction partners and similar graph structures within the same context. Using this framework, we analyzed temporally resolved global PPI networks of measles virus infection across seven cell types, identifying both conserved and cell-type-specific host proteins and biological processes affected by infection. These tools offer powerful new approaches for uncovering critical host-virus interactions and informing antiviral development.

David Van Valen: Understanding Kinase Substrate Interactions with Phospho-PCA and Deep Mutational Scanning

Protein kinases are crucial regulators of cellular behavior, but a comprehensive understanding of how kinases select their substrate remains a significant challenge in molecular biology. Understanding kinase-substrate specificity is essential for understanding cell signaling and developing specific kinase inhibitors as therapeutics. In this talk, we describe a new method, phospho-PCA, for profiling kinase-substrate interactions (KSIs) at scale using growth-based biosensors in yeast. This approach enables library-x-library screening of KSIs with library sizes of ~104 per experiment. Critically, the library size can be allocated as we see fit. For example, in a single experiment, we can measure KSIs for a diverse library of kinases (20) against a representative library of substrates (2000). We demonstrate that in this mode, phospho-PCA enables high-throughput engineering of kinase biosensors for live-cell imaging. Alternatively, we can perform mutational scanning for a single kinase (40000) against a library of substrates (3) to identify specificity-determining residues. We show that combining phospho-PCA with deep mutational scanning allows us to uncover allosteric regulatory sites and the sequence determinants of substrate specificity for c-Src, Lck, and Fyn kinases.

Kristen Naegle: Computational and synthetic experimental tools to understand domain-based tyrosine phosphorylation

Domains, evolutionarily conserved sequences and structures, are a central building block of human proteins. These conserved structural units are also the target of 25% of the 46,000 known tyrosine phosphorylation sites in the human proteome. Relatively little is understood about the role of tyrosine phosphorylation in regulating domain function or interactions, but we hypothesize that conserved structural locations of phosphorylation within a domain family can be used to annotate the function of tyrosine phosphorylation at a broader scale. We developed a comprehensive approach to extracting the interaction interfaces of domains from available experimental and predicted structures, within and between proteins, to predict the conserved functional effects of tyrosine phosphorylation. Using this approach, called CoDIAC, we identified and predicted that tyrosine phosphorylation likely globally regulates the interaction of SH2 domains by preventing interactions with phosphotyrosine-ligands or changing the specificity for binding partners. We have additionally developed a synthetic toolkit to produce recombinant phosphorylated proteins, which we have used to produce SH2 domain targets and phosphorylated SH2 domains for functional testing. Both our computational and experimental approaches have the capability of being useful across a broad range of scientific pursuits.

Benjamin Doran: Hierarchical organization across 60 million protein sequences reflects evolutionary constraints and differences of emergent properties

Natural selection creates systems with seemingly incomprehensible complexity. The advent of Protein Language Models (PLMs) such as Alpha-Fold and ESMfold have greatly increased our ability to predict protein’s emergent properties—physical structures, binding affinity, functional class—solely from the protein’s sequence of amino-acids. While these models have shown that an inferable relationship exists between protein sequence and emergent properties, we still don’t know the extent to which these relationships are memorized as opposed to inherent to how information is organized across the massive corpus of protein sequences on which these models are trained. To better understand the inherent organization of information across large ensembles of protein sequences, we developed a wavelet-based alignment technique in order to statistically decompose 60 million proteins spread across ~20 thousand protein families and all kingdoms of life. We find that information of

protein sequences is inherently organized in a hierarchical manner—the evolutionary history of proteins is nested within the context of distinct protein families. For individual proteins, their evolutionary history can be traced to reveal the patterns of constraint on the amino-acid sequence associated to distinct timescales of evolutionary selection. Moreover, these patterns of constraint hold intrinsic information relating to emergent properties of the protein including direct contacts and fluorescence. Our work explores the intrinsic information relating protein sequences to emergent properties and highlights the deep relationship between individual proteins and how their evolutionary history helps to describe those properties.

Ivy Liu: Scalable, compressed phenotypic screening using pooled perturbations
High-throughput phenotypic screens using biochemical perturbations and high-content readouts are constrained by limitations of scale. To address this, we establish a method of pooling exogenous perturbations followed by computational deconvolution to reduce required sample size, labor and cost. We demonstrate the increased efficiency of compressed experimental designs compared to conventional approaches through benchmarking with a bioactive small-molecule library and a high-content imaging readout. We then apply compressed screening in two biological discovery campaigns. In the first, we use early-passage pancreatic cancer organoids to map transcriptional responses to a library of recombinant tumor microenvironment protein ligands, uncovering reproducible phenotypic shifts induced by specific ligands distinct from canonical reference signatures and correlated with clinical outcome. In the second, we identify the pleotropic modulatory effects of a chemical compound library with known mechanisms of action on primary human peripheral blood mononuclear cell immune responses. In sum, our approach empowers phenotypic screens with information-rich readouts to advance drug discovery efforts and basic biological inquiry.

Cordelia McGehee: An Application of Filippov Systems to Mathematical Modeling of Adaptive Chemotherapy Administration

Acquired chemotherapy resistance remains a barrier to curative therapy in advanced cancers. Typical clinical dosing paradigms are chosen based on the maximum tolerated dose principle where drugs that are more toxic to normal tissue are given in on-off cycles and those with little toxicity are dosed continuously. Intra-tumoral cell-cell competition between sensitive and resistant cells has been hypothesized to drive chemotherapy resistance development under traditional chemotherapy dosing strategies. This is thought to be due to competitive release of the chemotherapy resistant cells after depletion of the chemotherapy sensitive cells. Mathematical modeling of this phenomenon using Lotka Volterra competition models has suggested that alternative dosing schemes aimed at maintaining a subpopulation of sensitive cells may lengthen time to tumor progression and even provide curative strategies under certain conditions. Two such adaptive therapy dosing regimens are continuous dosing (where a drug is given to reach a continuous steady state concentration with dose adjustments to maintain a constant total tumor volume) and intermittent dosing (where dosing is given in cycles to maintain a tumor volume between an upper and lower boundary). We previously explored direct mathematical comparison of these two adaptive therapy approaches in a modified Lotka Volterra model in order to prove that continuous adaptive therapy is superior to intermittent adaptive therapy across a variety of relevant metrics. In this work, we will describe how adaptive dosing can be conceptualized within the framework of Filippov Systems to extend this analytical framework to other systems and control problems.

Sadia Siddika Dima: Aggregation and DNA binding of Dorsal/NF-kappaB in early Drosophila embryos
The mechanism by which transcription factors (TFs) regulate gene expression, a process crucial for the development and maintenance of an organism, remains puzzling despite decades of research. Recent advances propose that TFs form dynamic, possibly phase- separated hubs that contribute to transcriptional regulation. However, the mechanisms connecting hub formation with transcriptional regulation remain poorly understood. To address this gap, we measured biophysical properties of the TF Dorsal (Dl), which forms a gradient to pattern the dorsal-ventral (DV) axis of the early Drosophila embryo.
Previous work using fluorescent imaging of live and fixed embryos indicated that the Dl gradient has a Gaussian-like shape with the highest level of Dl present on the ventral side of the embryo. The level of Dl

oscillates in a saw-tooth pattern with a progressively increasing amplitude during blastoderm nuclear cycles (ncs) 10 to 14. The spatiotemporal variation of Dl levels ensures the accurate spatial borders and dynamic pattern of target gene expression. Here we use Raster Image Correlation Spectroscopy (RICS), a type of scanning Fluorescence Correlation Spectroscopy, and single particle tracking in live embryos to quantify absolute concentrations of three sub-populations of Dl, each with distinct diffusivities and binding properties, along the DV axis from nc 10 to 14. We show that, while the absolute nuclear concentration of Dl increases over time, the fraction of Dl bound to DNA decreases over time. The DNA binding cannot be explained by a simple dose response relationship between free and bound concentrations, implying the existence of a mechanism beyond simple TF/DNA binding. The results from all the experiments encompassing different length and time scales suggest the presence of slowly moving clusters of Dl in addition to the expected populations of freely mobile Dl and DNA bound Dl. The data suggest that the clusters form only above a threshold concentration, a phenomenon typically associated with phase-separation. The formation of similar mobile clusters has been previously observed in other TFs, such as the anterior- posterior patterning factor Bicoid, and has been proposed to provide an efficient search strategy for regulatory regions. We also found that the clusters bind the DNA transiently. These findings enhance our understanding of the mechanism of transcriptional interpretation of the Dl gradient and are expected to generalize to other TF/DNA interactions.

Diep Nguyen: A spatially tiered design of threat-sensing mechanism in barrier organs
Barrier organs detect diverse immunogenic challenges through a small set of pattern recognition receptor (PRR) pathways. How organs assess threats and adjust the immune response to balance tolerance and resistance remains unclear. Here, by analyzing influenza infection in the lung using single-molecule imaging and spatial transcriptomics, we discovered a tiered threat-sensing strategy, where the probability for detecting and responding to viral infection is lowest in the outermost epithelia and highest in the inner stroma. Multiplexed quantification of signaling responses in human lung cells revealed that the levels of PRRs set the differential intrinsic probability of threat-sensing across cell types. Consequently, overexpressing RIG-I, an intracellular PRRs, specifically in lung epithelia in vivo exacerbated tissue damage upon pathogen-free inflammation. This spatially tiered design for assessing threats offers a strategy for tolerating challenges that are limited to the barrier cells, and still enabling potent immune responses when the threat invades the tissue.

Chenlei Hu: Scalable imaging-free spatial genomics through computational reconstruction
Tissue organization arises from the coordinated molecular programs of cells. Spatial genomics maps cells and their molecular programs within the spatial context of tissues. However, current methods measure spatial information through imaging or direct registration, which often require specialized equipment and are limited in scale. Here, we developed an imaging-free spatial transcriptomics method that uses molecular diffusion patterns to computationally reconstruct spatial data. To do so, we utilize a simple experimental protocol on two dimensional barcode arrays to establish an interaction network between barcodes via molecular diffusion. Sequencing these interactions generates a high dimensional matrix of interactions between different spatial barcodes. Then, we perform dimensionality reduction to regenerate a two-dimensional manifold, which represents the spatial locations of the barcode arrays. Surprisingly, we found that the UMAP algorithm, with minimal modifications can faithfully successfully reconstruct the arrays. We demonstrated that this method is compatible with capture array based spatial transcriptomics/genomics methods, Slide-seq and Slide-tags, with high fidelity. We systematically explore the fidelity of the reconstruction through comparisons with experimentally derived ground truth data, and demonstrate that reconstruction generates high quality spatial genomics data. We also scaled this technique to reconstruct high- resolution spatial information over areas up to 1.2 centimeters. This computational reconstruction method effectively converts spatial genomics measurements to molecular biology, enabling spatial transcriptomics with high accessibility, and scalability.

Nik Kovinich: MYB SG2 and WRKY33 KEEP Motifs Mediate Essential Interactions of a Conserved Transcription Factor Network that Regulates Divergent Pathogen- Induced Biochemical Defenses
Phytoalexins are plant-specialized metabolites that are biosynthesized *de novo* in response to pathogens. Their chemical structures are highly distinct among plant lineages, yet recent studies suggest that their divergent

biosynthetic pathways may be directly regulated by a conserved network of transcription factors. Here we identify the conserved transcription factors of the phytoalexin gene regulatory network. Through cross-species complementation, we show that orthologous transcrption factors can fully regulate non-host phytoalexin biosynthetic pathways. Specifically, we demonstrate that transcription factors for the biosynthesis of phenylalanine-derived phytoalexins in soybean can regulate the expression of tyrosine-derived phytoalexins in Arabidopsis,and vice versa. This is despite that the pathways consist of ~25 distinct biosynthetic genes. DNA-binding assays show that orthologous transcription factors directly bind non-host phytoalexin gene promoters, suggesting that phytoalexin transcription factors have coopted distinct gene sets into phytoalexin biosynthesis. Further, protein-protein interactions among the transcription factors are maintained during cross-species complementation, suggesting that conserved interactions are important for regulating phytoalexin biosynthesis. Finally, we identify two conserved motifs, the SG2-2 motif of MYBs and the KEEP motif of WRKYs that is essential for the interaction of those transcription factors and for activation of phytoalexin biosynthesis. Our results open the door to enhancing accessibility to valuable phytoalexin pharmaceuticals through cross-species engineering of the phytoalexin transcription factor network.

Tess Marvin: Developing a Transfer Learning Model for Variant Pathogenicity Prediction
Lyme disease, caused by Borrelia burgdorferi infection, is the most common vector-borne disease in the United States, yet the complex host-pathogen interactions governing immune response and tissue repair remain poorly elucidated. Furthermore, the disease often progresses undiagnosed due to the low sensitivity of early-stage serologic tests. Here, we leverage paired single-cell RNA-seq and ATAC-seq profiling of peripheral blood mononuclear samples from infected patients and uninfected controls to identify exposure- specific biomarkers and regulatory circuits associated with acute Lyme disease. We resolve chromatin-remodeling-linked gene expression alterations in patients with acute Lyme disease. Our analysis identified HES1 as a key inflammation suppressor downregulated in Lyme disease, alongside IGFB4 and LMNA, which are involved in tissue repair and structural integrity and demonstrated transcriptional shifts consistent with impaired repair mechanisms. We also identify Lyme disease induced shifts in alternative splicing programs. These significant transcriptional and epigenetic shifts underpin immune dysregulation and tissue repair deficits associated with disease progression. The identified signatures demonstrate high specificity for Lyme disease, distinguishing it from other bacterial and viral exposures. By integrating regulatory circuit mapping with biomarker prioritization, this work advances our understanding of Lyme disease pathogenesis while providing a foundation for more accurate and early diagnostic assays.

Zakary Singer: Engineered bacteria launch and control an oncolytic virus
The ability of bacteria and viruses to selectively replicate in tumors has led to synthetic engineering of new microbial therapies. Here we design a cooperative strategy whereby S. typhimurium bacteria transcribe and deliver the Senecavirus A RNA genome inside host cells, launching a potent oncolytic viral infection. “Encapsidated” by bacteria, the viral genome can further bypass circulating antiviral antibodies when delivered systemically to reach the tumor and initiate replication and spread within immune mice. Finally, we engineer the virus to require a bacterially delivered protease to achieve virion maturation, demonstrating bacterial control over the virus. This work extends bacterially delivered therapeutics to viral genomes, and shows how a consortium of microbes can achieve a synergistic aim.

Marcella M. Gomez: A data-driven approach to modeling and control of wound state progression and healing outcomes
Wound care remains a significant global challenge. Although many wounds heal with minimal intervention, improper management can lead to severe complications, including infection and even death. And wound care costs are expected to increase; for instance, chronic wounds are expected to rise by 8.4% annually. Accurate wound stage measurement and reliable tracking of healing trends are crucial for assessing the effectiveness of treatment strategies and enabling personalized decision-making for effective risk management in healthcare. The advancement of technology in bioelectronic sensors and actuators provides a platform for automating custom treatment strategies through feedback control. Here, we discuss the development and sense and

treatment strategies in wound healing. The timing of treatments is as critical to consider as the choice of drug or therapy due to the highly dynamic process of wound healing. In this work, we propose the first closed-loop control framework that takes in a time series of wound images in real-time and continuously adapts the treatment strategy based on the estimated and projected state of the wound. This work leverages a deep neural network architecture called DeepMapper, which analyzes wound images, to implement a novel deep reinforcement learning architecture that optimizes wound healing treatments towards accelerated wound closure using only images. This method directs accelerated wound closure without the need for a mechanistic model but lends to interpretability to understand the effects of drug or external stimuli on the wound. Finally, we introduce a recent extension of our work to chronic wounds and transcriptomic based models.

Brian Munsky: How crappy can our experiments be, without sacrificing uncertainty
Recent biochemical efforts have made outstanding progress to elucidate mechanisms for how cells sense and respond to external stimuli. Despite this growing body of insight, uncertainties in model parameters make it difficult to predict the full spatial, temporal, and heterogeneous responses of signal-activated gene expression. To overcome this limitation, stochastic models introduce random noise as a simplifying abstraction to capture the subtle effects of unknown or uncertain reactions or dynamics. When inferred from appropriate single-cell experiments, such as single-molecule Fluorescence in situ Hybridization (smFISH) or immunocytochemistry (ICC), the resulting semi-mechanistic models can achieve quantitatively accurate predictions for the statistics of complex biological responses in new environments. However, a nearly infinite number of different smFISH/ICC experiment designs could be proposed at different induction levels, for different measurement times, or considering different observed biological species. Moreover, each experiment can be time consuming or expensive to perform and will result in labeling, imaging, or data processing errors. To determine which experiments are best suited to identify predictive models, we adopt the chemical master equation framework to define likelihood functions, and we calculate the Finite State Projection based Fisher Information Matrix (FSP-FIM) to compare the information content inherent to different experiment designs. We extend the FSP-FIM with an empirically determined probabilistic distortion operator to estimate how measurement errors affect model identification. By analyzing different combinations of models, experiment designs, and image distortions, we discover practical working principles to simplify single-cell experiments to allow for the use of inexact (e.g., ‘crappy’) imaging approaches. We validate our FSP-FIM approach in HeLa cells using ICC data for glucocorticoid receptor transport and smFISH data for DUSP1 gene regulation upon Dexamethasone stimulation.