

## **Mitotic waves start from nuclei**

Oshri Afanзар<sup>1</sup>, Garrison K. Buss<sup>2</sup>, Tim Stearns<sup>2</sup> and James E. Ferrell, Jr<sup>1</sup>.

1. Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA

2. Department of Biology, Stanford University, Stanford, CA 94305, USA

Mitosis is the process by which one eukaryotic cell employs an array of measures to become two eukaryotic cells. It is fairly well understood when and how mitosis starts; CDK1, the universal activator of M-phase, is activated, and its activity spreads as a wave through the cytoplasm. Where mitosis starts from - we know little of. To determine whether specific parts of the cytoplasm can initiate mitosis, we observed the wave-spread of mitotic activity in large 2D round cytoplasm sheets, ~4 mm in diameter and ~200 micron in height. (The cytoplasm was made of *Xenopus* eggs extracts, a model system for the study of the cell cycle, where cytoplasmic constituents can be added or subtracted.) We found the mitotic activity in the sheet to spread in circular waves. When we back-traced the circular waves to their origin, we found that it coincided with small cytoplasmic volumes that contained nuclei. The wave-originating nuclei corresponded to locations where the lag time for mitotic initiation was shorter than in other parts of the cytoplasm. The shortening of mitotic lag time can imply the build-up of high local concentration of active CDK1 at the nucleus. When nuclei were not present, it was very common to observe that the circular waves originated from the edge of the sheet. As was the case with nuclei, shorter lag time were measured at the edge wave origins. Because the edge can mirror incoming molecules, the generation of circular edge-waves may be caused by high local concentration of CDK1 and its regulators, which is the sum of the freely diffusing and edge-deflected molecules. Curiously, although centrosomes can concentrate cytoplasmic elements by active transport, we found that they did not nucleate mitotic waves, nor reduced mitotic lag time. We conclude that mitosis initiates from the nucleus, and further propose that it does so by confining the diffusion CDK1 and its regulators to make high local concentrations.

## The bacterial flagellar motor got gears

Oshri Afanar<sup>1†</sup>, Diana Di Paolo<sup>2</sup>, Miriam Eisenstein<sup>3</sup>, Kohava Levi<sup>1</sup>, Anne Plochowitz<sup>2</sup>,  
Achillefs N. Kapanidis<sup>2</sup>, Richard Michael Berry<sup>2</sup> and Michael Eisenbach<sup>1\*</sup>

1. Department of Biomolecular Sciences, Weizmann Institute of Science, 7610001 Rehovot, Israel

2. Biological Physics Research Group, Clarendon Laboratory, Department of Physics, University of Oxford, Oxford, UK

3. Department of Chem. Res. Support, Weizmann Institute of Science, 7610001 Rehovot, Israel

† Present address: Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA

The bacterial flagellar motor can rotate in two directions; counterclockwise for the bacterium to swim in straight lines and clockwise to abruptly change the direction of swimming. But, can the motor show an intermediate state of rotation for gentle maneuvers? We found that it does, and it was much by serendipity. It is known that a single protein named CheY binds to the motor to switch its direction of rotation. It is also known that CheY can be activated by phosphorylation or by acetylation to enable motor switching. When we initiated this study, we focused on the mechanism by which phosphorylation and acetylation affects the interaction of CheY with the motor. As part of this work, we removed from the motor a peptide domain that strongly binds CheY. When we measured the motor response to activated CheY in the absence of the peptide, we found something completely unexpected. Before it got fully committed to clockwise rotation upon CheY activation, the motor was stuck in an indecisive limbo of rapid back-and-forth switching from counterclockwise to clockwise directions. When we investigated the interaction of CheY with these motors we found that CheY binds sequentially to two motor sites. When we lowered the affinity of CheY to one site, the motor generated mostly stable clockwise rotation. When the other site was mutated, the motor generated mostly instable intervals of clockwise rotation. So it seems, that the presence of the peptide has masked so far an underlying two-step interaction of CheY with the motor, where one step initiates an unstable state of clockwise rotation, and the other stabilizes the clockwise state. To determine what advantage it is for bacteria to have two “gears” for clockwise generation, we observed the swim behavior of a strain in which the first step, but not the second, was mostly active. We found that these bacteria swim in a way that is neither just straight lines nor abrupt change in swimming direction. Instead, it was most common for bacteria to just a slightly deflect from the original trajectory of swimming. We conclude that CheY sequentially binds to two different motor sites, one initiates an unstable clockwise rotation and the other stabilizes the clockwise state. This two-step mechanism can result in two modes of swimming behavior, trajectory deflection and abrupt change in swimming direction, respectively.

## Quantitative live cell imaging of PolII activity in plants uncovers the cellular basis of transcriptional dynamics.

Simon Alamos<sup>1,2</sup>, Armando Reimer<sup>3</sup>, Krishna K. Niyogi<sup>1,2,4</sup> and Hernan G. Garcia<sup>3,5,6</sup>

<sup>1</sup> Department of Plant and Microbial Biology, University of California at Berkeley. Berkeley, CA.

<sup>2</sup> Howard Hughes Medical Institute.

<sup>3</sup> Biophysics Graduate Group, University of California at Berkeley, Berkeley, CA.

<sup>4</sup> Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA.

<sup>5</sup> Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, CA.

<sup>6</sup> Department of Physics, University of California at Berkeley. Berkeley, CA.

The regulation RNA PolII activity at specific genes in particular cells enables multicellular systems to adjust their physiology and development to their environment. This is particularly important for sessile complex organisms such as plants. Despite this being a dynamic phenomenon operating at the cellular level at timescales as fast as seconds, current methods to assess transcriptional activity in plants rely on bulk-sampled ‘snapshots’. To bridge this gap we implemented a quantitative imaging strategy to count the number of actively transcribing RNA Polymerases at the single locus level in live plant cells within their tissue context. Using this system in the context of development and stress responses we characterized how temporal patterns of mRNA accumulation arise from gene activity in individual *Arabidopsis thaliana* leaf cells. These experiments revealed that the time-dependent fraction of cells engaged in transcription and not their level of activity explains the response dynamics at the tissue level. To understand the source of cell-to-cell variability in this binary regulatory mode we generated plants where transcription of two alleles can be measured at the same time in each cell. Preliminary results from these experiments point to a major role of extrinsic (i.e. correlated) noise. Because plant cells are connected and do not move this result predicts spatial clustering in gene activity, a hypothesis that we confirmed. We hope that the methods and results presented here will open new questions in the quantitative study of gene regulation in plants.



## ***De novo* assembly and chromosome-level scaffolding of two *Anopheles* mosquitoes genomes**

Pavel Avdeyev<sup>1,\*</sup>, Anton Zamyatin<sup>2,\*</sup>, Varvara Lukyanchikova<sup>3</sup>, Igor Sharakhov<sup>3</sup>,  
and Max A. Alekseyev<sup>1</sup>

<sup>1</sup>The George Washington University, Washington, DC, USA

<sup>2</sup>ITMO University, St. Petersburg, Russia

<sup>3</sup>Virginia Tech University, Blacksburg, VA, USA

Malaria has a devastating global impact on public health and welfare, with the majority of the world's malaria cases occurring in sub-Saharan Africa. *Anopheles* mosquitoes are exclusive vectors of malaria, with species from the *An. gambiae* complex being the most important African vectors. *An. coluzzii* and *An. arabiensis*, along with *An. gambiae*, are the malaria vectors of most widespread importance in sub-Saharan Africa.

While population genomics studies can uncover both challenges and opportunities for mosquito control in Africa, until recently chromosome-level genomic sequences were available only for one strain of *An. gambiae*. In the present project, we have successfully generated chromosome-level *de novo* assemblies for *An. coluzzii* and *An. Arabiensis* genomes using Oxford Nanopore sequencing and Hi-C proximity ligation data.

\* Equal contribution

## Linearization of Median Genomes under the DCJ-Indel Model

Pavel Avdeyev<sup>1</sup>, Max A. Alekseyev<sup>1,2</sup>

<sup>1</sup> Department of Mathematics, George Washington University, 801 22<sup>nd</sup> St NW, Washington, DC 20052

<sup>2</sup> Department of Biostatistics & Bioinformatics, George Washington University, Washington, DC

In the course of evolution genomes become a subject to a number of large-scale evolutionary events such as genome rearrangements that shuffle genomic architectures, and gene insertions and deletions (*indels*) that insert or remove continuous intervals of genes. Since these evolutionary events are rare, they often are used in phylogenomic studies to measure the evolutionary distance between the genomes. Such measurement is usually based on the *maximum parsimony assumption*, implying that the evolutionary distance can be estimated as the minimum number of events between genomes. A convenient model for the most common genome rearrangements is given by the *Double-Cut-and-Join* (DCJ) operations, which make two “cuts” in a genome and “glues” the resulting genomic fragments in a new order. Namely, DCJs mimic *reversals* (that inverse contiguous segments of chromosomes), *translocations* (that exchange tails of the two chromosomes), and *fissions / fusions* (that split / join chromosomes), while indels can be modeled by the DCJs on certain artificial circular chromosomes called *prosthetic*.

The maximum parsimony assumption enables addressing the *ancestral genome reconstruction problem*, which asks to reconstruct ancestral genomes from given extant genomes, by minimizing the total distance between genomes along the branches of the phylogenetic tree. The basic case of this problem with just three given genomes is known as the *genome median problem* (GMP), which asks for a single ancestral genome (*median genome*) at the minimum total distance from the given genomes.

The GMP is NP-hard under a number of models of genome rearrangements, including reversals-only and DCJ. While these problems can be posed for both *circular* genomes (consisting of circular chromosomes) and *linear* genomes (consisting of linear chromosomes), the DCJ model allows appearance of circular chromosomes in transformations between linear genomes. Correspondingly, a solution to the GMP under the DCJ model may contain circular chromosomes even if the given genomes are linear. We will therefore distinguish between the GMP and the *linear genome median problem* (L-GMP), where the latter is restricted to linear genomes only.

To the best of our knowledge, there exist no solvers for the L-GMP, while there are some advanced GMP solvers, which allow the median genome to contain circular chromosomes. This deficiency inspired us to pose the problem of using the solution for the GMP to obtain a linear genome approximating the solution to the L-GMP. In the present study, we propose an algorithm that linearizes chromosomes of a given GMP solution in a certain optimal way. Our approach also provides insights into the combinatorial structure of genome transformations by DCJs and indels with respect to appearance of circular chromosomes.

## **Exploring IL-10 negative regulation of cell-to-cell heterogeneity in the TLR4 inflammatory response**

Amanda Alexander<sup>1</sup>, Kathryn Miller-Jensen<sup>1,2</sup>

<sup>1</sup> Department of Biomedical Engineering, Yale University, 55 Prospect St, New Haven, CT 06511

<sup>2</sup> Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06511

Macrophages are critical mediators of the innate immune response and act as the first line of defense against foreign pathogens. To rapidly mount an inflammatory response without tissue damage, macrophages use both positive and negative feedback motifs mediated through paracrine signals, including the pro-inflammatory cytokine TNF- $\alpha$  and the anti-inflammatory cytokine IL-10. We have recently showed that upon activation, macrophages exhibit extensive cell-to-cell heterogeneity in these secreted paracrine signals, raising questions about how this heterogeneity affects positive and negative feedback regulation that enables rapid but controlled inflammatory responses. Here we studied cell-to-cell heterogeneity in the TLR4 stimulated response in primary bone-marrow derived macrophages (BMDMs), with a focus on secretion of TNF- $\alpha$  and IL-10. Population-level experiments are consistent with the interpretation that IL-10 is secreted after pro-inflammatory cytokines and acts to inhibit pro-inflammatory secretion at high LPS doses. In contrast, measuring multiplexed single-cell cytokine secretion in macrophages isolated in nano-wells (which enables analysis of autocrine signaling without paracrine signaling), we found that IL-10 is secreted concomitantly with pro-inflammatory signals. Upon inhibiting IL-10 signaling via IL10R blocking, we found that significantly more cells secreted low levels of proinflammatory cytokines, suggesting that IL-10 is acting to constrain the fraction of cells that activate the pro-inflammatory program at all LPS doses. Overall, we conclude that IL-10 has an important role in controlling cell-to-cell functional heterogeneity of the pro-inflammatory secretion program and could provide insight into higher order immune response regulation and enable more rational targeting of immunotherapies.

**Gregoire Altan-Bonnet**, National Institutes of Health

*Robotic mapping & machine learning to improve tumor immunotherapies*

Assessing the antigenicity of tumors is a critical task when designing fine-tuned cancer immunotherapy. Antigen quality, or the affinity of an antigen to a T cell receptor, and antigen quantity, the density of antigen presented on the surface of immune cell targets, both play crucial roles in determining the efficacy of an immune response. Their roles in T cell activation are intimately linked, as low quantities of a high quality antigen and high quantities of a low quality antigen can produce similar immune responses, making it difficult to disentangle the true binding strength of an antigen from the amount of the antigen encountered by a T cell. Because measuring the quality of T cell response to neoantigens is critical to cancer immunotherapy, and because the quantity of an tumor neoantigen is more difficult to measure *in vivo* and less functionally relevant than its quality, a robust metric for quantifying T cell response to tumor neoantigens should be able to quantify both the quality and quantity of the neoantigen separately. Conventional metrics for assessing neoantigen quality often rely on only a single measurement of T cell activation (e.g. IFN-g ELISPOT), and therefore have difficulties deconvolving antigen quality and quantity. We developed a TECAN robotic platform to track the *ex vivo* dynamics of T cell differentiation, proliferation and cytokine secretion simultaneously in response to antigens of different binding affinities and concentrations. We introduced a simple neural network to classify “kinetic features” (derivatives/integrals of different observables over different time periods), and deconvolve antigen quality/quantity. We show that this machine-learning-based method indeed does allow for more accurate prediction of antigen quality/quantity than any single measurement of an activation marker, and can therefore represent a promising new metric to use for clinical assessment of the immunotherapeutic potential of tumor neoantigens

**Nihal Altan-Bonnet**, National Institutes of Health

*All aboard: Vesicles, viruses and strength in numbers*

Individual viral particles have historically been considered the optimal units of transmission. This is in large part due to the idea that viruses, by moving independently from one another, have a greater probability of spreading to as many hosts as possible. Our studies over the last 5 years have revealed another form of viral transmission where by viruses move from host to host as populations inside extracellular vesicles, and infect en bloc. These extracellular vesicles can be derived from autophagosomes, multivesicular bodies and the plasma membrane. They can carry both RNA and DNA viruses and even naked infectious genomes. We have found that en bloc transmission of viral cargo inside extracellular vesicles has significant advantages: increasing viral multiplicity of infection; protecting viral cargo against environmental assaults; and suppressing host innate immunity by perturbing the hosts' ability to distinguish self from non-self. Our findings show that vesicle-cloaked viruses are highly virulent and highlight a need for novel methods of viral therapeutics that will target vesicles and disrupt the transmission of viruses en bloc.

## **Manipulating p53 expression dynamics to control target promoter activation and regulate cell fate decisions**

Marie D. Harton<sup>1</sup>, Woo Seuk Koh<sup>1</sup>, Amie D. Bunker<sup>1</sup>, Ryan L. Hanson<sup>1</sup>, Zhilin Yang<sup>1</sup>,  
Abhyudai Singh<sup>2</sup>, and Eric Batchelor<sup>1</sup>

<sup>1</sup> Laboratory of Cell Biology, CCR, NCI, NIH, Bethesda, MD

<sup>2</sup> Department of Electrical and Computer Engineering, University of Delaware, DE

The p53 tumor suppressor regulates distinct responses to cellular stresses. Although different stresses generate different p53 dynamics, the mechanisms by which cells decode p53 dynamics to differentially regulate target genes are not well understood. Previously, we showed that p53 target expression dynamics are shaped by the relationship between the target mRNA and protein decay rates and the oscillation frequency of p53 dynamics. Here, we focused on aspects of target gene synthesis. We determined in individual cells how canonical p53 target gene promoters vary in responsiveness to features of p53 dynamics. Employing both a chemical perturbation approach and an optogenetic approach, we independently modulated p53 pulse amplitude, duration, or frequency, and we then monitored p53 levels and target promoter activation in individual cells. We identified distinct signal processing features—thresholding in response to amplitude modulation, a refractory period in response to duration modulation, and dynamic filtering in response to frequency modulation. We then showed that the signal processing features not only affect p53 target promoter activation, they also affect downstream cellular functions. Our study shows how different promoters can differentially decode features of p53 dynamics to generate distinct responses, providing insight into how perturbing p53 dynamics can be used to generate distinct cell fates. Such insight may help to inform novel cancer therapy strategies based on optimized timing of therapeutic genotoxic agents.

## **A simple hierarchical regulatory design decouples induction decision from level**

Chiara Ricci-Tam<sup>1,\*</sup>, Jue Wang<sup>2,\*</sup>, Ishay Ben-Zion<sup>1,\*</sup>, Julius Palme<sup>1</sup>, Ang Li<sup>1</sup>, Yonatan Savir<sup>3</sup> and Michael Springer<sup>1</sup>

<sup>1</sup> Harvard Medical School, Dept. of Systems Biology, Boston, MA

<sup>2</sup> University of Washington, Dept. of Chemical Engineering, Seattle, WA

<sup>3</sup> Technion, Dept. of Physiology, Biophysics & Systems Biology, Haifa, Israel

Networks of transcription factors control complex biological responses. One aspect of response complexity is the ability to decouple the decision to express a gene from its expression level. This is thought to be achieved by independent control of chromatin accessibility and transcriptional output, however, direct tests of this hypothesis have been very limited. Here we find that the *S. cerevisiae* galactose utilization (GAL) pathway displays such a behavior - the decision to express the pathway is controlled by one input and the level of expression is controlled by another input. Surprisingly, this behavior is not achieved through independent control of chromatin accessibility and transcription, which are strongly correlated in the GAL system. Instead, decoupling of decision from level is achieved through a hierarchy between transcription and signaling. This design provides a physiological advantage allowing cells to tune the investment made in preparing for a future response. Remarkably, the decision and level controllers are genetically decoupled, allowing evolution to independently act on both axes. The simplicity and flexibility of this design suggest that it will be found in many biological systems.

\* Equal contributions

## Using synthetic biology to explore bacterial tolerance, persistence, and antibiotic resistance

Heather S. Deter<sup>1</sup>, Alawiah H. Abualrahi<sup>1</sup>, Tahmina Hossain<sup>1</sup>, and Nicholas C. Butzin<sup>1</sup>

1. Department of Biology and Microbiology. South Dakota State University. 1224 Medary Ave Brookings, SD 57007-0493, USA.

Antibiotic tolerance is a widespread phenomenon that renders antibiotic treatments less effective and facilitates antibiotic resistance. Recently, we explored the role of proteases in antibiotic tolerance, short-term population survival of antibiotics, using queueing theory (i.e. the study of waiting lines), computational models, and a synthetic biology approach. Proteases are essential cellular components that degrade proteins and play a key role in a multi-drug tolerant subpopulation of cells, called persisters. We found that queueing at the protease ClpXP increases antibiotic tolerance ~80 and ~60 fold in an *E. coli* population (called queueing-tolerance) treated with ampicillin and ciprofloxacin, respectively. There does not appear to be an effect on antibiotic persistence, which we distinguish from tolerance based on population decay. These results demonstrate that proteolytic queueing is a practical method to probe bacterial tolerance and related genes, while limiting the unintended consequences frequently caused by gene knockout and overexpression. We are using several different strategies to gain a better understanding of queueing-tolerance and probe persistence. We are using microfluidic devices and our custom pipeline to quantify 1000s of single-cell response over time. This pipeline takes advantage of recent advances in imaging and machine learning, and allows us to identify, track, and quantify single cells before and after antibiotic treatment. We have also moved beyond the *E. coli* “chassis” to study antibiotic survival in the synthetically developed minimal genome *Mycoplasma mycoides* JCVI-syn3B (Syn3B). Syn3B allows us to take a bottom-up approach to study persistence because it contains less than 500 genes and was constructed to only contain essential genes (and a few non-essential genes for genetic manipulation and speed of growth). In addition, we are using transcriptional analysis along with several other techniques to identify key drug targets that bacteria utilize to survive antibiotics.

## **Single-cell heterogeneity in cell cycle commitment is mediated by slow p21 induction and cell cycle inertia during G1 phase**

Jenny Nathans, James A. Cornwell, Marwa Afifi, Debasish Paul, and Steven D. Cappell

Center for Cancer Research, National Cancer Institute, NIH, 37 Convent Drive, Bethesda, MD 20892

Mammalian cells integrate growth and stress signals to make the decision to either enter or exit the cell cycle, and this decision is thought to be executed at a discreet point during G1 phase called the restriction point. However, we recently showed that if cells encounter genotoxic stress even after they have passed the restriction point, they can still exit the cell cycle and return to quiescence. Surprisingly, there is considerable single-cell variability in this decision; 40% of cells exit to quiescence after DNA damage while 60% of cells seemingly ignore the DNA damage and continue into S phase. These results prompt the questions of why and how cells choose these opposite fates and what the biological implications are of choosing to exit rather than commit to the cell cycle in response to DNA damage. Here we show using time-lapse microscopy and fluorescent biosensors for cyclin-dependent kinase 2 (CDK2) and the anaphase promoting complex/cyclosome (APC/C) that cells are more likely to exit to quiescence in response to DNA damage early in G1 phase and become gradually less likely to exit to quiescence as they progress to S phase. Furthermore, the fate choices of genetically identical sister cells are highly correlated, indicating that the response to DNA damage is largely deterministic rather than stochastic. We show that single-cell variability in the response to DNA damage is due to a relatively slow induction of the cell cycle inhibitor p21 compared to the length of G1 phase. Cells that produce sufficient p21 prior to entering S phase exit to quiescence, whereas cells that fail to produce enough p21 will enter S phase, where the S-phase specific ubiquitin ligase CRL4<sup>Cdt2</sup> will degrade p21 and allow cells to proceed through the cell cycle with damaged DNA. We observe through long-term tracking of cells that progression to S phase despite DNA damage results in enlarged nuclei, which is often associated with re-replication and aneuploidy. Taken together, our study demonstrates that cells will continue progressing through the cell cycle for at least 2 hours after DNA damage due to slow p21 induction, which is often sufficient to progress erroneously through cell cycle checkpoints and alter their long-term fate outcomes.

## **HeartENN: developing a tissue-specific deep learning model that detects the noncoding *de novo* variant contribution to congenital heart disease**

Kathleen M. Chen<sup>1,\*</sup>, Felix Richter<sup>2,3,4,\*</sup>, Evan M. Cofer<sup>5,6</sup>, Jian Zhou<sup>1,5</sup>, Bruce D. Gelb<sup>3,4,7</sup>,  
Olga G. Troyanskaya<sup>5,8</sup>

<sup>1</sup> Flatiron Institute, Simons Foundation, New York, NY

<sup>2</sup> Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

<sup>3</sup> Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY

<sup>4</sup> Departments of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

<sup>5</sup> Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ

<sup>6</sup> Graduate Program in Quantitative and Computational Biology, Princeton University, Princeton, NJ

<sup>7</sup> Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY

<sup>8</sup> Department of Computer Science, Princeton University, Princeton, NJ

\* Equal contributions

Noncoding changes in the genome can significantly alter phenotype, yet their contributions to human disease remain poorly characterized. Recently, new techniques in deep learning have been developed to model the regulatory activity encoded in genomic sequences; such models predict mutation effects directly from sequence, even for previously unseen mutations. Furthermore, rapid increases in the amount of genome-wide regulatory data generated across various cell types and conditions newly enables the development of specialized sequence-based deep learning models that learn regulatory patterns particular to different biological contexts. These models can be used to study the impact of noncoding mutations in specific diseases.

We developed the first such specialized model, HeartENN, to understand the contribution of *de novo* noncoding variation to congenital heart disease (CHD), a highly tissue-specific disease known to be primarily genetic but not fully explained by coding variants. HeartENN, a model that predicts chromatin profiles measured only in cardiac cells, was used to characterize the heart-specific noncoding contribution to CHD. Based on a comparison of whole-genome sequencing data from a cohort of 750 CHD trios to 1616 Simons Simplex Collection controls, we observed significant causal effects at the noncoding level. Notably, significant differences between cases and controls were observed for *de novo* variants (DNVs) associated with CHD genes and genes highly expressed during heart development. This work illustrates the effect of regulatory variants

in CHD and demonstrates the utility of using tissue-specific regulatory features to predict mutation effects.

While we focus on congenital heart disease here, our approach for creating specialized, context-specific models of transcriptional regulation is general. We envision that the development of sequence models specialized to different cell type, organ, and disease contexts will continue to elucidate the molecular basis of metazoan cell/tissue specificity and of human diseases. To support such work, we created Selene (<https://selene.flatironinstitute.org/>), a PyTorch-based library for fast and easy development and application of sequence-based deep learning models to delineate genomic sequence regulation and mutation effects. Selene enables a wide range of users to develop custom deep learning sequence models.

# Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor

Haiqi Chen<sup>1</sup>, Sophia Liu<sup>1,2,3\*</sup>, Samuel Padula<sup>1\*</sup>, Daniel Lesman<sup>1</sup>, Kettner Griswold<sup>4,5,6,7</sup>, Allen Lin<sup>8,9</sup>, Tongtong Zhao<sup>1,10</sup>, Jamie L. Marshall<sup>1</sup>, Fei Chen<sup>1</sup>

<sup>1</sup> Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA;

<sup>2</sup> Biophysics Program, Harvard University, Boston, MA 02115, USA;

<sup>3</sup> Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

<sup>4</sup> MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

<sup>5</sup> Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA;

<sup>6</sup> Department of Genetics, Harvard Medical School, Boston, MA 02115, USA;

<sup>7</sup> Charles Stark Draper Laboratory, Cambridge, MA 02139, USA;

<sup>8</sup> Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA;

<sup>9</sup> Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA;

<sup>10</sup> Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.

\* These authors contributed equally to this work.

## Abstract

Due to a recent convergence of molecular and synthetic biology protocols (e.g. CRISPR, next generation sequencing), there has been immense interest in encoding information such as lineage and stimuli, into the sequence of nucleic acids. These and other applications which rely on nucleotide diversification, such as barcoding and directed evolution, are fundamentally reliant on methods which enable continuous diversification of genomic sequences *in vivo*. Here, we describe TRACE, a system which enables continuous, targeted mutagenesis in eukaryotic cells by combining the somatic hypermutation capability of cytidine deaminases with the DNA processivity of an orthologous bacteriophage T7 RNA polymerase. We demonstrate that TRACE is an engineerable tool for inducing high rates of nucleotide diversification in a targeted region within the human genome across multiple cell generations. It enables ~500,000 fold higher mutation rates than the basal somatic mutation rate, and its modularity ensures that even higher editing rates can be achieved by engineering either the cytidine deaminase or T7 RNA polymerase. Moreover, TRACE mutations occur over an order-of-magnitude larger editing window than previously published technologies (at least 2000 bp). We applied TRACE by performing a MEK1 inhibitor resistance screen via directed mutagenesis within mammalian cells and identified novel functionally correlated mutations. Thus, TRACE is a powerful tool to generate continuous diversification of user-defined genomic loci in the native context of mammalian cells, which will be useful to a broad range of biological applications.

# F5-peptide enhances the efficacy of the non-hormonal male contraceptive adjuvin

Haiqi Chen<sup>a,d</sup>, Dolores Mruk<sup>a</sup>, Chris K.C.Wong<sup>b</sup>, Bruno Silvestrini<sup>c</sup>, C. Yan Cheng<sup>a</sup>

<sup>a</sup> The Mary M. Wohlford Laboratory for Male Contraceptive Research, Center for Biomedical Research, Population Council, 1230 York Ave, New York, NY

<sup>b</sup> Department of Biology, Hong Kong Baptist University, Hong Kong, China

<sup>c</sup> SBM S.r.l. Pharmaceuticals, Rome, Italy

<sup>d</sup> Broad Institute of MIT and Harvard, Cambridge, MA, USA

## Objective

The bioavailability of the non-hormonal male contraceptive adjuvin is low in due to the blood-testis barrier (BTB). This study was designed to examine if F5-peptide, an endogenously produced reversible BTB modifier, could enhance the bioavailability of adjuvin to affect spermatogenesis and provide a contraceptive effect while reducing systemic toxicity.

## Study Design

We overexpressed F5-peptide in adult male rats (n=10 rats; with 3 or 4 rats for each of the three different experiments noted in the three regimens) by intratesticular injection of a mammalian expression vector pCI-neo (pCI-neo/F5-peptide) vs. empty vector alone (pCI-neo/Ctrl) to be followed by treatment with adjuvin by oral gavage at a dose of 10 or 20 mg/kg. The status of spermatogenesis was assessed by histological analysis and dual-labeled immunofluorescence analysis on Day 16. To assess fertility, we allowed treated males (n=3–4 rats) to mate with mature female rats (n=3–4) individually, and assessed the number of pups on Days 23, 36 and 82 to assess fertility and reversibility.

## Results

All 4 treated rats overexpressed with F5-peptide and low-dose adjuvin were infertile by Day 36, and half of these rats were fertile by Day 82, illustrating reversibility. However, overexpression of F5-peptide alone (or low-dose adjuvin alone) had no effects on fertility in n=3 rats. These findings were consistent with the histology data that illustrated the BTB modifier F5-peptide promoted the action of adjuvin to induce germ cell exfoliation, mediated by changes in cytoskeletal organization of F-actin and microtubules across the epithelium, thereby reducing the systemic toxicity of adjuvin.

## Conclusion

In this proof-of-concept study, it was shown that overexpression of the F5-peptide prior to administration of adjuvin to rats at a low (and ineffective dose by itself) was found to induce reversible male infertility.

## Implications

Overexpression of F5-peptide, an endogenously produced biomolecule in the testis known to induce BTB remodeling, enhanced the contraceptive effect of adjuvin in rats. This combinatorial treatment holds the potential to become a novel non-hormonal contraceptive for men.

# Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor

Haiqi Chen<sup>1</sup>, Sophia Liu<sup>1,2,3\*</sup>, Samuel Padula<sup>1\*</sup>, Daniel Lesman<sup>1</sup>, Kettner Griswold<sup>4,5,6,7</sup>, Allen Lin<sup>8,9</sup>, Tongtong Zhao<sup>1,10</sup>, Jamie L. Marshall<sup>1</sup>, Fei Chen<sup>1</sup>

<sup>1</sup> Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA;

<sup>2</sup> Biophysics Program, Harvard University, Boston, MA 02115, USA;

<sup>3</sup> Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

<sup>4</sup> MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA;

<sup>5</sup> Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA;

<sup>6</sup> Department of Genetics, Harvard Medical School, Boston, MA 02115, USA;

<sup>7</sup> Charles Stark Draper Laboratory, Cambridge, MA 02139, USA;

<sup>8</sup> Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA;

<sup>9</sup> Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA;

<sup>10</sup> Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA.

\* These authors contributed equally to this work.

## Abstract

Due to a recent convergence of molecular and synthetic biology protocols (e.g. CRISPR, next generation sequencing), there has been immense interest in encoding information such as lineage and stimuli, into the sequence of nucleic acids. These and other applications which rely on nucleotide diversification, such as barcoding and directed evolution, are fundamentally reliant on methods which enable continuous diversification of genomic sequences *in vivo*. Here, we describe TRACE, a system which enables continuous, targeted mutagenesis in eukaryotic cells by combining the somatic hypermutation capability of cytidine deaminases with the DNA processivity of an orthologous bacteriophage T7 RNA polymerase. We demonstrate that TRACE is an engineerable tool for inducing high rates of nucleotide diversification in a targeted region within the human genome across multiple cell generations. It enables ~500,000 fold higher mutation rates than the basal somatic mutation rate, and its modularity ensures that even higher editing rates can be achieved by engineering either the cytidine deaminase or T7 RNA polymerase. Moreover, TRACE mutations occur over an order-of-magnitude larger editing window than previously published technologies (at least 2000 bp). We applied TRACE by performing a MEK1 inhibitor resistance screen via directed mutagenesis within mammalian cells and identified novel functionally correlated mutations. Thus, TRACE is a powerful tool to generate continuous diversification of user-defined genomic loci in the native context of mammalian cells, which will be useful to a broad range of biological applications.

**Talk title:** *Structural Insights from De novo Beta-Sheet Proteins Promise Novel Therapeutics*  
<sup>1</sup>Tamuka Chidyausiku Ph.D, <sup>2</sup>Enrique Marcos, Ph.D, <sup>1</sup>Jason Klima, Ph.D, <sup>1</sup>Hugh Hadox, Ph.D  
and <sup>1</sup>David Baker, Ph.D

1. *Institute for Proteins Design, University of Washington, Seattle, WA*
2. *Institute for Research in Biomedicine, Barcelona, Spain*

### **Abstract**

$\beta$ -sheet protein domains are ubiquitous in nature, carrying out a wide range of functions including transporting hydrophobic molecules, recognition and enzymatic processing of carbohydrates, and scaffolding of virus capsids and antibodies. Although  $\beta$ -sheet protein scaffolds are well suited for incorporating new functions, their design from first principles have remained an outstanding challenge. We made recent progress in de novo protein design of mixed- $\alpha\beta$  folds which subsequently enabled the accurate design of many hyperstable and structurally diverse proteins suitable for ligand-binding, catalytically active or protein-protein mediating functions. Toward this end, we investigated the mechanisms controlling  $\beta$ -sheet curvature by studying the geometry of  $\beta$  sheets in naturally occurring protein structures and folding simulations. The principles emerging from this initial work with NTF2-like designs enabled the accurate and precise control in the design of the 1st ever de novo double-stranded ALL  $\beta$ -helix protein with the most non-local interactions possible. This topology is known as the jelly roll fold and can be classified as a  $\beta$ -sandwich. Lastly, after developing a set of design principles for these jelly roll proteins, we then were able to use our previously developed high-throughput protein design and characterization method to iteratively generate a library of hyperstable, monomeric and cooperatively folding jelly roll scaffolds suitable for ligand binding. This thesis work set the stage for the successful design of other de novo  $\beta$ -sheet topologies like immunoglobulin (Ig) domain of antibodies, where two sandwiched  $\beta$ -sheets anchor the variable loops (i.e. complementary determining regions (CDRs)) responsible for antigen binding.

**Poster Title:** *Hip Hop Biochemistry: Lessons from Designing Diverse De novo Designed Proteins without Sacrificing Excellence*  
Tamuka Chidyausiku, Ph.D

YouTube & Instagram: @TamukaInvestments #HipHopBiochemistry

### **Abstract**

The lack of diverse media representation of scientists is NOT ONLY a reflection of the lack of diversity in science but also the skewed and biased perception that society has about what a scientist can look like, act like or what they do. As the 1st UW black student to graduate with a PhD in biochemistry from the Institute for Protein Design, I am intimately familiar with the problems associated with the struggle of coming from an underrepresented community in science. However, having graduated with 3 high impact papers, 11 de novo PDB structures and 1000s of designed diverse topologies; I have proof that it is possible to have Diversity without Sacrificing Excellence. So it's time we all contribute to the social fight for diversity and inclusion. Towards that end I launched an experimental course on my Youtube channel called, "Hip Hop Biochemistry". The goal of this series of short ~10mins videos is to (aim 1) diversify the set of analogies that we pull from when teaching science to include more pop culturally relevant and exciting examples to help students understand complex concepts. (Aim 2) Dispel the myths of science being an elitist subject by meeting & reaching students where they are - on the internet and (aim 3) recruit URM students more enthusiastically particularly those who otherwise would have thought themselves as "not belonging" in this field. Finally, using the statistics gained from this experience (aim 4) Propose a new syllabus for in-class teaching 1st at predominantly URM institutions like Seattle's Garfield High and then turn this informal but very data centered Youtube course into a more structured and formally backed curriculum for all schools internationally.



## Abstract

$\beta$ -sheet protein domains are ubiquitous in nature, carrying out a wide range of functions including transporting hydrophobic molecules, recognition and enzymatic processing of carbohydrates, and scaffolding of virus capsids and antibodies. Although  $\beta$ -sheet protein scaffolds are well suited for incorporating new functions, their design from first principles have remained an outstanding challenge. We made recent progress in de novo protein design of mixed- $\alpha\beta$  folds which subsequently enabled the accurate design of many hyperstable and structurally diverse proteins suitable for ligand-binding, catalytically active or protein-protein mediating functions. Toward this end, we investigated the mechanisms controlling  $\beta$ -sheet curvature by studying the geometry of  $\beta$  sheets in naturally occurring protein structures and folding simulations. The principles emerging from this initial work with NTF2-like designs enabled the accurate and precise control in the design of the 1st ever de novo double-stranded ALL  $\beta$ -helix protein with the most non-local interactions possible. This topology is known as the jelly roll fold and can be classified as a  $\beta$ -sandwich. Lastly, after developing a set of design principles for these jelly roll proteins, we then were able to use our previously developed high-throughput protein design and characterization method to iteratively generate a library of hyperstable, monomeric and cooperatively folding jelly roll scaffolds suitable for ligand binding. This thesis work set the stage for the successful design of other de novo  $\beta$ -sheet topologies like immunoglobulin (Ig) domain of antibodies, where two sandwiched  $\beta$ -sheets anchor the variable loops (i.e. complementary determining regions (CDRs)) responsible for antigen binding.

# Engineering therapeutic T cell circuits that harness combinatorial antigen recognition to overcome tumor heterogeneity in glioblastoma

Joseph H. Choe<sup>1\*</sup>, Payal B. Watchmaker<sup>2\*</sup>, Diego A. Carrera<sup>2</sup>, Wei Yu<sup>1</sup>, Kira M. Downey<sup>2</sup>, Nira A. Krasnow<sup>1</sup>, Ryan D. Gilbert<sup>2</sup>, Kole T. Roybal<sup>5-8#</sup>, Hideho Okada<sup>2,5,8</sup>  
# and Wendell A. Lim<sup>1,3,4,8#</sup>

<sup>1</sup> Department of Cellular & Molecular Pharmacology, University of California, San Francisco, CA 94158, USA

<sup>2</sup> Department of Neurological Surgery, University of California, San Francisco, CA 94158, USA

<sup>3</sup> UCSF Center for Systems and Synthetic Biology, San Francisco, CA 94158, USA

<sup>4</sup> Howard Hughes Medical Institute, San Francisco, CA 94158, USA

<sup>5</sup> Parker Institute for Cancer Immunotherapy

<sup>6</sup> Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, CA 94158, USA

<sup>7</sup> Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

<sup>8</sup> Helen Diller Cancer Center, University of California, San Francisco, San Francisco, CA 94158, USA

\* These two authors contributed equally as the co-first authors

# These three authors contributed equally as the co-corresponding senior authors

Treatment of solid cancers with chimeric antigen receptor (CAR) T cell is challenging because of a lack of target antigens that are both tumor-specific and homogeneously expressed in the cancer cells. In glioblastoma (GBM), the epidermal growth factor receptor variant III (EGFRvIII) antigen is highly tumor-specific but is a non-ideal target because it is heterogeneously expressed. In contrast, several more homogeneously expressed antigens, while associated with GBM, are non-ideal targets because they are also expressed in other normal organs. We show that “prime-and-kill” dual antigen recognition circuits can overcome this challenge *in vitro* and *in vivo*. Where a synthetic Notch (synNotch) receptor that recognizes a specific but heterogeneous antigen (EGFRvIII) is engineered to locally induce expression of a CAR that kills neighboring cells expressing tumor-associated antigens. This type of circuit spatially integrates recognition of multiple antigen targets across the tumor, yielding T cells with improved effectiveness in treating heterogeneous GBM.

## Analyzing insect wing shapes using quasi-conformal mappings

Gary Pui-Tung Choi<sup>1</sup> and L. Mahadevan<sup>1,2,3,4</sup>

<sup>1</sup> John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

<sup>2</sup> Departments of Physics, Harvard University, Cambridge, MA

<sup>3</sup> Departments of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA

<sup>4</sup> Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA

In this work, we propose a computational approach for analyzing insect wing shapes. We first establish a correspondence between the boundaries of two insect wing shapes with boundary landmarks using geometric functional data analysis, and then compute a landmark-matching curvature-guided Teichmüller mapping with uniform quasi-conformal distortion in the bulk. This allows us to quantify the pair-wise difference between the wing shapes and construct a similarity matrix on which we deploy methods from network analysis to cluster shapes. We deploy our method to study a variety of *Drosophila* wings across species to highlight the phenotypic variation between them, and *Lepidoptera* wings over time to study the developmental progression of wings. Our approach of combining complex analysis, computation and statistics to quantify, compare and classify insect wing shapes may be usefully deployed in other biological and physical systems.

## **An unexpected role for CDK4/6 activity after DNA damage in G2**

James A. Cornwell<sup>1</sup> and Steven D. Cappell<sup>1</sup>

1. Center for Cancer Research, National Cancer Institute, NIH, 37 Convent Drive, Bethesda, MD 20892

Ordered and unidirectional progression through the stages of the cell cycle is required to maintain genome integrity and ensure faithful production of two daughter cells. We recently showed that inactivation of the anaphase promoting complex (APC) acts as a bistable and irreversible switch to ensure a proper transition from G1 to S phase. This switch is triggered at the end of G1 phase by cyclin-dependent kinase 2 (CDK2) and rendered irreversible by a double negative feedback loop between early mitotic inhibitor 1 (Emi1) and the APC, which generates hysteresis. Following the G1/S transition, the APC remains inactive until mitosis when APC activation ensures ordered progression from metaphase to anaphase. Interestingly, DNA damage during G2 phase can lead to premature re-activation of the APC which is posited to result in permanent exit from the cell cycle to a senescent state. Previous studies have shown that the re-activation of APC in G2 is mediated by p53, p21, and downregulation of EMI1 mRNA through inhibition of the CDK2-pRb-E2F1 axis. However, these studies have relied upon synchronization of cells and population-level measurements using standard biochemical assays that lack cellular and temporal resolution. Therefore, questions regarding the heterogeneity of the response and the exact timing of these events remain largely unaddressed. Using live-cell fluorescent biosensors for CDK2 and APC activity coupled with long-term, quantitative, single-cell tracking, we show that even in the presence of very high amounts of DNA damage not all cells activate the APC, demonstrating significant amounts of single-cell heterogeneity. Furthermore, we show that EMI1 downregulation prior to DNA damage allows APC activation in the presence of high levels of CDK2 activity, demonstrating the importance of EMI1 in conveying hysteresis to APC activation in G2 phase. However, most surprisingly, we have uncovered a new pathway leading to premature APC re-activation that is independent of p21. Our data show that, unexpectedly, in response to DNA damage the G1-specific kinase CDK4/6 is required to maintain activity of the CDK2-pRb-E2F1 axis in G2 phase. Concurrent treatment with neocarzinostatin (which induces double-stranded breaks) and Palbociclib (a potent and selective inhibitor of CDK4/6 activity) results in total loss of CDK2 activity, Rb phosphorylation, and the transcription of E2F1 target genes, even in the absence of p21. Furthermore, in contrast with previous studies we find that neither downregulation of EMI1 transcription, or EMI1 degradation, are pre-requisites for APC re-activation. Thus, our data provide new insights into the mechanism of APC bi-stability which implicate a novel role for CDK4/6 in S/G2 that regulates the decision to progress from S/G2 to mitosis.

## **‘Mother’ tongues? A quantitative comparison of genetic and linguistic transmission worldwide**

Worldwide patterns of genetic variation are driven by human history. To test whether this demographic history has left similar signatures on languages to those it has left on genes, we analyzed sounds (phonemes) from 2,082 languages and microsatellite polymorphisms from 246 populations. Globally, both genetic distance and phonemic distance between populations were significantly correlated with geographic distance; populations that were closer to one another tended to be more similar, genetically and linguistically. Close examination of this pattern suggests the influence of two processes: parent-to-offspring transmission of both genes and languages during the peopling of the world, and linguistic borrowing (often coupled with genetic admixture) when neighboring populations speak very different languages. To understand whether sex-biased patterns in human history affect the associations between genetic and linguistic variation, we merged genetic, linguistic, and ethnographic data to perform two new studies. First, we address the hypothesis that languages are likely to be transmitted from mother to child by comparing linguistic variation to maternally inherited (mitochondrial) and paternally inherited (Y-chromosome) genetic distance separately. In addition, we use ethnographic data to annotate genotyped populations with information about their kinship system and post-marital residence pattern. With this annotated database, we observe measurable effects of matrilineal kinship systems and matrilineal residence patterns on the evolution of genes and languages; our results suggest that sex-biased demographic processes influence genetic variation more than linguistic. By integrating data types and quantitative approaches across disciplines, these analyses shed light on the similarities and differences in genetic and cultural signatures of human population history.

## Proteolytic queue formation at ClpXP increases antibiotic tolerance

Heather S. Deter<sup>1</sup>, and Nicholas C. Butzin<sup>1</sup>

1. Department of Biology and Microbiology. South Dakota State University. Brookings, SD. 57006. USA.

Antibiotics have saved millions of lives, but antibiotic resistance has become a global crisis as treatments become more expensive and less effective. An important facet of resistance is the role of antibiotic tolerance, which increases and precedes the development of antibiotic resistance. Persistence is a subtype of tolerance that is distinguished by a distinct death rate during antibiotic treatment. The mechanisms underlying tolerance and persistence are varied and poorly defined, but their presence is ubiquitous among bacterial populations. We have chosen to focus on systems associated with antibiotic persistence and tolerance that are also found throughout bacteria, specifically proteolytic systems and proteases. The precise role of proteases in antibiotic tolerance and persistence has been difficult to identify, as protease knockouts have mixed results, likely because knocking out individual proteases and their chaperones “break” cellular networks and affects cell growth (cell growth is directly related to antibiotic survival). To address these concerns, we have taken a synthetic biology approach to interfere with proteolytic degradation in natural systems. Proteolytic queueing occurs when proteases become overloaded by proteins targeted for degradation. The effects of queueing have been demonstrated in the context synthetic circuits (primarily oscillators) that rely on the LAA tag for rapid protein degradation by ClpXP, wherein overproduction of LAA tagged proteins results in a proteolytic queue that coordinates otherwise unrelated processes and slows the degradation of proteins by a specific protease complex. We found that queue formation in *E. coli* at ClpXP increases antibiotic tolerance ~80- and ~60-fold after three hours of ampicillin and ciprofloxacin treatment, respectively. Additionally, we have shown that tolerance occurs as an active response to the proteolytic queue, and computational modeling supports that this change is due to alterations in tolerance rather than the number of persister cells. We are currently in the process of identifying which proteins are responsible for the queueing-tolerance effect, which includes obtaining and analyzing transcriptional data on queueing-tolerance. Our results demonstrate that synthetic, proteolytic queues impact natural phenomenon and queueing provides an alternative method for studying proteolytic degradation.

## **A phase-separation based positive feedback loop between a substrate and its kinase allows for effective T Cell Receptor Signaling**

Elliot Dine and Jared E. Toettcher

Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

T Cell Receptor (TCR) signaling differs from many other signaling pathways in that the receptor does not directly activate downstream signaling pathways. The TCR instead activates the kinase ZAP70, which then phosphorylates a membrane localized adaptor protein called LAT. It is ZAP70's phosphorylation of LAT that leads to activation of downstream pathways. Recently, it was shown that ZAP70 and LAT undergo liquid-liquid phase separation in vitro and form dynamic clusters in cell lines. However, the necessity and purpose of ZAP70-LAT cluster formation, as compared to other types of protein-protein interactions, remains unclear. To study this question, we developed a suite of optogenetic tools that leads to either the phase separation and clustering of LAT and ZAP70 or just to their heterodimerization. In non-T cells, inducing phase separation of LAT and ZAP70 with blue light activates downstream pathways associated with TCR signaling. However, inducing heterodimers of ZAP70 and LAT with blue light does not lead to productive signaling. Our work suggests that clusters of ZAP70-LAT turn on downstream signaling pathways due to the full phosphorylation and activation of ZAP70 that only occurs inside the clusters. Surprisingly, this activation of ZAP70 requires the presence of phosphorylated LAT. Thus, the initial phosphorylation and clustering of LAT positively feeds back onto the previous step of T Cell receptor signaling, the phosphorylation and activation of ZAP70. We determine that this feedback loop contains at least three components: ZAP70, LAT and Src Family Kinases. Overall, this work reveals a novel consequence of protein phase separation: the ability of clustered proteins to act not just on their immediate binding partner, but on other members of the cluster allowing for the amplification of weak initial stimuli.

## Genome Replication as Origin of Regulatory Variation and Epigenetic Drift

Julien Morival<sup>1,2,4</sup>, Honglei Ren<sup>2,3</sup>, Navied Akhtar<sup>1,4</sup>, Qing Nie<sup>1,2,5,6</sup>, Elizabeth L. Read<sup>2,7</sup>,  
and Timothy L. Downing<sup>1,2,4,8</sup>

<sup>1</sup> Department of Biomedical Engineering, UC Irvine

<sup>2</sup> NSF-Simons Center for Multiscale Cell Fate Research, UC Irvine

<sup>3</sup> Graduate Program in Mathematical, Computational and Systems Biology, UC Irvine

<sup>4</sup> The Edwards Lifesciences Center for Advanced Cardiovascular Technology, UC Irvine

<sup>5</sup> Department of Mathematics, UC Irvine

<sup>6</sup> Department of Developmental & Cell Biology, UC Irvine

<sup>7</sup> Department of Chemical & Biomolecular Engineering, UC Irvine

<sup>8</sup> Department of Microbiology & Molecular Genetics, UC Irvine

Molecular heterogeneity is emerging as a critical feature of multicellular life. While single-cell analyses have revealed the existence of cell-to-cell variation in the levels and activities of the molecules responsible for gene regulation, the source of such variation is still poorly understood. Cytosine methylation is a highly conserved epigenetic modification that plays an important role in mammalian development and its occurrence within phenotypically uniform cell populations is often variable even at the same genomic location. We recently developed a new sequencing method (Repli-BS) that enables analysis of methylation heterogeneity across cytosine residues within newly replicated strands of DNA over time. Using this method, we discovered that much of the methylation heterogeneity observed within HUES64 human embryonic stem cells (hESCs) is temporal in nature and associated with DNA replication (Charlton et al., *Nat. Struc. Mol. Bio.* 2018). More recently, we used our hESC Repli-BS dataset to establish kinetic rate parameters that numerically reflect the speed at which individual cytosines achieve steady-state methylation levels after being replicated (Busto-Moner et al., *BioRxiv* 2019). Here, we employ bioinformatic analyses to explore how properties of post-replication DNA methylation dynamics relate to well-established features of the genome and the broader chromatin landscape. Our preliminary findings reveal that unique patterns of methylome replication associate with distal regulatory regions throughout the genome, enrich for cytosine residues dynamically methylated across cell types, and coincide with the location of stem cell-specific transcription factor binding and chromatin architectures. We also find correlations between sub-cell cycle kinetics in DNA methylation and the divergence of bulk methylation patterns observed during multiple cell generations and natural aging. Taken together, our studies suggest that (epi)genome replication may act as an important source of (temporal) regulatory variation in hESCs while, simultaneously, conferring susceptibility to epigenetic drift throughout the human lifespan.

**Maitreya Dunham**, University of Washington

*High throughput functional annotation of pharmacogene variants using yeast*

The field of pharmacogenomics is currently overwhelmed by the huge amount of genetic variation being discovered by new sequencing efforts. One key limitation is the lack of corresponding functional annotation of these gene variants that would allow the field to link them to clinically actionable drug responses. We are addressing this problem in a particularly important family of pharmacogenes: cytochrome P450s. In particular, I'll discuss our work on CYP2C9, an enzyme responsible for metabolizing many different drugs including warfarin, phenytoin, and flurbiprofen. Genetic variants in this gene are known to affect the efficacy of these and other drugs. We have developed a yeast-based activity assay to test thousands of variants in a pooled fashion using a deep mutational scanning approach. Yeast has been used as a model system for recombinant P450 expression for over 30 years and can be engineered to express highly active human P450 enzyme. Our yeast assay, which uses activity-based protein profiling, is able to recapitulate the activity of known variants in both individual and pooled tests. Briefly, humanized yeast cells expressing a single CYP2C9 variant are bound in an activity-dependent manner by a modified CYP2C9 inhibitor that is then labeled via click chemistry with a fluorophore for cell sorting and sequencing. This is done in a massively parallel manner, such that the entire library of variants is sorted into bins based on activity level, and each bin is deep sequenced to determine variant frequency. We have created a barcoded library of CYP2C9 single amino acid variants and have determined variant activity scores using our yeast-based assay and deep sequencing. So far, we have generated activity scores for ~70% of the 9,800 possible CYP2C9 single amino acid variants. Preliminary analysis shows that roughly 60% of missense variants tested have significantly decreased activity and may have altered drug metabolism. We have validated individual variants using gold standard in vitro metabolic assays with clinically relevant substrates to demonstrate the reliability of our data. Our approach will lead to advances in adverse drug response prevention by providing clinical guidance for patients carrying both currently known and yet-to-be discovered alleles of CYP2C9 and other CYPs.

**Miki Ebisuya**, The European Molecular Biology Laboratory (EMBL), Barcelona  
*Human Time vs. Mouse Time: in vitro segmentation clock as a model system*

Different species have different tempos of development: larger animals tend to grow more slowly than smaller animals. My group has been trying to understand the molecular basis of this interspecies difference in developmental time, using the segmentation clock as a model system. The segmentation clock is the oscillatory gene expressions that regulate the timing of somite formation from presomitic mesoderm (PSM) during embryogenesis. We have recently succeeded in inducing PSM from both human iPS cells and mouse ES cells, detecting the oscillation and traveling wave of segmentation clock in vitro. Interestingly, the oscillation period of human segmentation clock was 5-6 hours while that of mouse was 2-3 hours. Taking advantage of our in vitro system and simple mathematical models, we have been comparing the genome sequences and molecular processes of the segmentation clock between human and mouse to explain the interspecies difference in the oscillation period



# Non-Equilibrium Regulation of Chromatin Accessibility and Transcription in Development

Elizabeth Eck<sup>1,\*</sup>, Jonathan Liu<sup>2,\*</sup>, Shelby Blyther<sup>3</sup>, Hernan G. Garcia<sup>1,2,4,5</sup>

<sup>1</sup> Biophysics Graduate Group, University of California at Berkeley, Berkeley CA

<sup>2</sup> Department of Physics, University of California at Berkeley, Berkeley CA

<sup>3</sup> Department of Molecular Biosciences, Northwestern University, Evanston IL

<sup>4</sup> Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley CA

<sup>5</sup> Institute for Quantitative Biosciences-QB3, University of California at Berkeley, Berkeley CA

An open challenge in developmental biology is the prediction of gene expression patterns from knowledge of the spatiotemporal concentration dynamics of input transcription factors and their binding site arrangement on regulatory DNA. Following successes in bacteria, most descriptions of transcriptional regulation in development have been built under the assumption of equilibrium, where no energy is expended during transcription factor binding or interaction with the transcriptional machinery. These models have been particularly successful in predicting transcriptional regulation in bacteria, but have seldom been rigorously examined in eukaryotes. A mounting body of evidence suggests that processes such as ATP expenditure for chromatin modification, the very short-lived binding times of transcription factors on DNA, and the dynamic nature of the general transcriptional machinery conspire to keep the system out of equilibrium.

To put this widespread equilibrium assumption to a stringent test, we use the widely studied formation of the step-like gene expression pattern of *hunchback* mediated by the Bicoid activator and the pioneer transcription factor Zelda in the early embryo of the fruit fly *Drosophila*. Using a combination of quantitative single-cell live imaging, theoretical models, and computational simulations, we demonstrate that no equilibrium processes can recapitulate how Bicoid and Zelda dictate *hunchback* expression. Instead, we show that *hunchback* regulation can only be described by models where energy is expended in transcriptional regulation, which we speculate stems from the process of making chromatin accessible to transcription factors. Our results suggest that, in order to reach a predictive and quantitative understanding of transcriptional programs in development and, more generally, in eukaryotes, a non-equilibrium view of gene regulation must be considered. Further, the steps that operate outside of equilibrium in the transcriptional cascade, where energy is expended, need to be identified and characterized.

\* Equal Contributions

**Hana El-Samad**, University of California, San Francisco

*Why do cells arrest in cell cycle when stressed? Not what you think!*

We explore what happens when yeast cells don't slow/arrest in their cell cycle upon hyperosmotic stress. Our explorations establish links between stress, cell cycle progression and metabolic control. These links underlie a tradeoff that cells might be navigating between fast response to stress and resilience to repeated stress.

# Structural Analysis and Robust Design of Biomolecular Circuits with Fast Binding and Slow Catalysis

Fangzhou Xiao<sup>1</sup> and John C. Doyle<sup>1,2</sup>

<sup>1</sup> Department of Bioengineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA.

<sup>2</sup> Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

Natural biomolecular systems in a cell are robust to uncertain parameters and reaction rates, yet they can perform complex and versatile behaviors such as tracking in chemotaxis, multistability in development, and oscillations in circadian rhythms. However, our existing theoretical tools are incapable of analyzing this behavior. On the one hand, intuitive Hill function type models implicitly assume some species' concentrations are much larger than others. Un-modeled disturbances *in vivo* often break such assumptions. On the other hand, chemical reaction networks (CRNs) are used to describe the full behavior of a circuit making minimal physical assumptions. But CRNs are often so complicated by conservation laws and boundary cases that simple biological questions cannot be answered unless un-biological assumptions, e.g. detailed-balance, are made.

Inspired by the observation that biological reactions are mostly fast binding and slow catalysis, we propose every biomolecular species' dynamics naturally consist of a production term and a degradation term, each operating at several structural regimes (e.g. linear or saturating). Catalysis corresponds to production and degradation terms, while binding reactions govern the order of the catalysis dynamics. For example, a binding reaction forming inactive complexes results in a suppressive regime where more binding factors causes less catalytic activity. Compared to Hill-function type models, this allows us to explicitly take into account the "undesirable" regimes that could happen due to disturbances. Compared to full-blown CRNs, this is simpler to analyze and more intuitive. Furthermore, as regimes are structural, i.e. they only describe the order of the production and degradation terms, this framework naturally calls for structural analysis and design that are robust to rate variations. Hence this framework has the combination of structural robustness and versatile behavior baked into its foundation. As a demonstration, we have analyzed and designed circuit structures exhibiting multistability together with parameter regimes guaranteeing it. We are also working on structural conditions for oscillations.

## **Development of math models and a tumorigenic index (TI) for quantitative analysis of liver cancer**

Gaowei Wang, Xiaolin Luo, Yan Liang, Kota Kaneko, Hairi Li, Xiang-Dong Fu, and Gen-Sheng Feng.  
Department of Pathology, Division of Biological Sciences, Department of Cellular and Molecular Medicine,  
Moore's Cancer Center, University of California San Diego, La Jolla, CA 92093, USA

The incidence and mortality of liver cancer, mainly hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC), are increasing rapidly worldwide. Diverse risk factors for primary liver cancer have been identified, including infection of hepatitis B virus (HBV) and hepatitis C virus (HCV), alcohol abuse and non-alcoholic steatohepatitis (NASH) as well as intake of aflatoxin B1. Consistent with the complex and multifactorial etiologies, multi-omics analyses of human HCC and ICC samples have identified vast genomic heterogeneity, molecular and cellular defects, metabolic reprogramming, and subtypes of tumors as well as altered tumor microenvironment in the liver.

However, it remains to be determined if any common molecular signatures in the transcriptomes exist for liver cancer, despite their considerable genomic heterogeneity. Furthermore, little is known about the kinetics and fashions, either gradual accumulation or dramatic transition, in generation of cell-intrinsic and -extrinsic signals that are intertwined to drive malignant transformation of hepatocytes and tumor initiation. To dissect the stepwise tumorigenic signals and mechanisms at the pre-cancer stages, we chose to work on mouse models that recapitulate key pathogenic features in human liver cancer. By pathological examination, transcriptomic profiling, and bioinformatic analysis, we identified a sudden switch in transcription factor (TF) clusters at a critical pre-cancer stage of hepato-oncogenesis. Based on a new multi-layer analysis of the TF clusters and mathematic modeling, we established a tumorigenic index (TI) for quantitative measuring of liver tumorigenesis. Using the TI as a tool developed from mouse tumor models, we analyzed human patients' data deposited in the public datasets, and demonstrated applicable effectiveness of the TI value in prognosis of liver cancer patients with diverse etiologies and also in diagnosis of pre-cancer patients with chronic liver diseases.

# Learning developmental trajectories from CRISPR lineage tracing and single-cell gene expression

Aden Farrow<sup>1</sup> and Geoffrey Schiebinger<sup>2</sup>

<sup>1</sup> Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Oxford, OX2 6GG, UK

<sup>2</sup> Department of Mathematics, University of Vancouver, 1984 Mathematics Road, Vancouver, BC, V6T 1Z2, Canada

Recent research has shown that the mathematical theory of optimal transport can effectively reconstruct developmental trajectories from time courses of single cell gene expression; however, this approach requires expensive experiments with fine time resolution. In this work, we present a novel framework that leverages new types of lineage-tracing information measured simultaneously with gene expression. These experimental techniques use heritable CRISPR-induced genetic barcodes to trace the history of cell divisions over the course of development. Crucially, these barcodes can be measured in the same cells as gene expression. Our method, designed for lineage-tracing time courses, learns from both kinds of information together using mathematical tools from graphical models, structural equation models, and optimal transport. We find that lineage data improves optimal transport's effectiveness in disentangling complex state transitions with lower temporal resolution, thereby reducing experimental cost.

## **Towards clinical systems pathology: spatially resolved in situ characterization of proliferation and quiescence in human tissue resection samples.**

Giorgio Gaglia<sup>1,3,\*</sup>, Sheheryar Kabraji<sup>4,\*</sup>, Danae Argyropoulou<sup>1,3</sup>, Yang Dai<sup>1,3</sup>, Jean Zhao<sup>4</sup>, Peter Sorger<sup>1,2</sup> and Sandro Santagata<sup>1,3</sup>

<sup>1</sup>Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115

<sup>2</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115

<sup>3</sup>Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115

<sup>4</sup>Department of Cancer Biology, Dana Farber Cancer Institute, Boston, MA 02115

Advances in both genomic and transcriptomic techniques have greatly improved our ability to systematically characterize the molecular processes that contribute to disease pathogenesis. However, the classification of human tumors and the choice of therapeutic strategy is still mainly based on the pathological evaluation of tissue sections by using light microscopy to capture critical morphological features that lack overt molecular details. Technologies that will allow diagnosticians to integrate morphologic features with complex molecular information at cellular and subcellular resolution within intact tissue sections will dramatically expand current models of diagnosis and prognosis as well as the ability to predict response to treatment. Our team has developed a method for routinely capturing highly multiplexed images of biomarkers (30-60 markers) from archival human resection specimens. Here, we apply this method called tissue cyclic immunofluorescence (t-CyCIF) to acquire highly multiplexed protein measurements with single-cell granularity and subcellular resolution in whole sections of breast cancer tissue resections. We combine markers of proliferation, cell cycle phase, cell-cycle arrest and quiescence to fully capture the ability of breast cancer cells to grow and their response to cell-cycle inhibitor therapy. The highly multidimensional feature space allows us to discriminate the cell types that populate the tumor micro-environment, tumor cell proliferation potential and multiple types of quiescent states previously observed only in pre-clinical models. In single breast tumors, we identify pockets of rapidly growing tumor cells that co-exist with areas of transiently and terminally arrested cells, and we disentangle phenotypic diversity and spatial heterogeneity. Moreover, we reveal that single markers that have been widely used clinically to assess tumor growth, define tumor grade and determine the potential for response to therapy are either imprecise, excessively conservative or biased toward a subset of cell-cycle phases. Our results demonstrate an important need for the development of spatially resolved multiplexed in situ technologies for application in biological discovery, in pre-clinical work, in human clinical trials and eventually in clinical practice.

\*Equal contributors

# Cooperation-Driven Phase Transitions in Two-Species Microbial System

Jin-Hyeon Kim and Kwang-Il Goh

*Department of Physics, Korea University, Seoul 02841, Korea*

Cooperative effect by means of division of labor is believed to underlie many important evolutionary transitions including the emergence and proliferation of multicellularity. Here we study the emergence of cooperation-driven phase transitions in an interacting particle system-type model of two-species microbial system incorporating the effect of cooperative reproduction. We perform extensive stochastic simulations to construct the phase diagram of the model in one dimension, accompanied by supporting mean-field-type numerical calculations. The model exhibits two different kinds of nonequilibrium phase transitions. The first is the transition between active phase (in which the microbial species can survive in the long run) and absorbing state (in which both species are sure to get extinct). This transition is primarily governed by the effective reproduction rate and belongs to the well-known directed percolation universality class of nonequilibrium phase transitions. Within the active phase, there emerges another transition between two macroscopically-difference patterns of microbial configurations. In one phase (corresponding to low cooperation rate), each of the two species is living separated from the other, forming segregated single-species domains with well-defined boundary. In the other phase (corresponding to large cooperation rate), the two species can live and mingle together, forming macroscopic mixed-species domains inside which different species cells tend to locate adjacent to each other and effectively behave like bi-cellular organisms. This secondary transition is governed by the effective cooperation rate. We characterize these transition properties from the perspective of nonequilibrium phase transition.

## CoRa – An approach for quantifying feedback control in biomolecular systems

Mariana Gómez-Schiavon<sup>1</sup> & Hana El-Samad<sup>1,2,3</sup>

<sup>1</sup> Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA.

<sup>2</sup> Cell Design Initiative, University of California, San Francisco, CA, USA.

<sup>3</sup> Chan–Zuckerberg Biohub, San Francisco, CA, USA.

Feedback control is ubiquitous in biomolecular networks, being the mechanism behind homeostasis and biochemical adaptation. The ability to rigorously evaluate the contribution and limitations of feedback control mechanisms is a critical step to meaningfully understand and ultimately design feedback control systems in biomolecular contexts. However, there is no general approach to quantify biological feedback control and compare different control strategies. Here, we introduce **CoRa** –or **Control Ratio**–, a metric that quantifies the advantage of a feedback control system compared to a series of “locally analogous” systems without feedback. Each “locally analogous” system is constructed by substituting the feedback with a constant input that guarantees that the internal states of the two systems with and without control are identical without perturbations. CoRa is then computed by perturbing both systems and scoring the difference in their behaviors. In this way, CoRa effectively isolates and measures the contribution of the feedback control, while considering the impact of all the intrinsic biomolecular constraints of the system. CoRa can be applied to any given feedback control system, regardless of the underlying complexity of the biomolecular network, producing a readily interpretable value. We demonstrate through examples that CoRa is a powerful tool to characterize control systems, design biomolecular controllers, and compare controller motifs in biomolecular contexts. As an example, we demonstrate how CoRa recapitulates established knowledge and provides novel insights in the analysis of an antithetic feedback control strategy. In this case, CoRa clearly highlights how the molecular complex dynamics in the antithetic motif can have dramatic and unexpected effects on the controller performance. Additionally, we show how CoRa can be productively used to optimize a synthetic feedback control strategy. Finally, we demonstrate that CoRa provides a unifying framework that allows for the comparison of different control strategies. We illustrate this point by using CoRa to extract principles about the operation of a large swath of synthetic feedback control strategies reported in the literature. In summary, CoRa is a simple, generalizable and informative metric that can guide efforts for dissecting and designing biomolecular feedback control.

# Single-cell quantification of the concentrations and dissociation constants of endogenous proteins

Yuhei Goto<sup>1,2,3\*</sup>, Akira T. Komatsubara<sup>1,4\*</sup>, Yohei Kondo<sup>2,3</sup>, Michiyuki Matsuda<sup>4</sup>,  
and Kazuhiro Aoki<sup>1,2,3</sup>

<sup>1</sup> National Institute for Basic Biology, NINS, Okazaki, Aichi, Japan

<sup>2</sup> Exploratory Research Center on Life and Living Systems, NINS, Okazaki, Aichi, Japan

<sup>3</sup> Department of Basic Biology, Faculty of Life Science, SOKENDAI (Graduate University for Advanced Studies), Okazaki, Aichi, Japan

<sup>4</sup> Department of Pathology and Biology of Diseases, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Kinetic simulation is a useful approach for elucidating complex cell-signaling systems. The numerical simulations critically require parameters such as protein concentrations and dissociation constants (Kd). Especially, Kd value is affected by various intracellular conditions, such as competitive binding, molecular crowding, pH, and/or salt concentration, so that Kd in living cells (hereafter *in vivo* Kd) should be different from Kd measured *in vitro*. However, only a limited number of Kd values have been measured experimentally in living cells.

Here we describe an approach for quantifying the concentration and *in vivo* Kd of endogenous proteins at the single-cell level with CRISPR/Cas9-mediated knock-in and fluorescence cross-correlation spectroscopy (FCCS). First, the mEGFP gene was knocked in at the gene encoding extracellular signal-regulated kinase 2 (ERK2) through microhomology-mediated end joining (MMEJ). Next, the HaloTag gene was knocked in at the end of the ribosomal S6 kinase 2 (RSK2) gene. We then used fluorescence correlation spectroscopy (FCS) to measure the protein concentrations of endogenous ERK2-mEGFP and RSK2-HaloTag proteins in living cells, revealing substantial heterogeneities. Moreover, FCCS analyses showed temporal changes in the *in vivo* Kd values of the binding between ERK2-mEGFP and RSK2-HaloTag in response to epidermal growth factor stimulation. Our approach provides a robust and efficient method for quantifying endogenous protein concentrations and *in vivo* Kd in living cells.

\*Equal contributions

## **Alignment of the PAR polarity axis orientation to the cell shape through active nematic line-tension in the *C. elegans* zygote**

Peter Gross<sup>1</sup>, Michael Nestler<sup>2</sup>, Archit Bhatnagar<sup>1</sup>, Axel Voigt<sup>2</sup>, Stephan Grill<sup>1</sup>  
<sup>1</sup>MPI-CBG, Dresden, Germany, <sup>2</sup>Technical University Dresden, Germany

The anterior-posterior (A-P) body axis of *C. elegans* single-cell embryos is encoded in the orientation of the partitioning-defective (PAR) polarity domains. The initially unpolarized egg receives a symmetry-breaking signal from the cytoplasmic centrosome, which initializes the polarity establishment phase [1]. Here, a posterior PAR domain grows to about half the cell size, assisted by large-scale actomyosin flows. While it is observed that the position of the symmetry-breaking centrosome, and thus the posterior PAR domain is initially only loosely aligned to the ellipsoidal body axis of the zygote, at later stages of the polarity establishment phase this misalignment is corrected and the polarity axis and the body axis converged.

We here investigate the mechanism by which the mechanochemically self-organized PAR-actomyosin system can correct for initial misalignments in the PAR polarity axis. First we show that this process is driven by the local curvature, through manipulations of the naturally ellipsoidal shape. Interestingly, this alignment is absent when actomyosin flows are inhibited, indicating that active actomyosin flows have a central role in aligning polarity domains to the cell shape. We then test two hypothesized mechanisms for the actomyosin-driven alignment process. First, a feedback loop between actomyosin surface flows, intracellular cytoplasmic flows and the resulting re-positioning of the centrosomal polarity trigger is explored. Second, we investigated the role of emergent line-tensions, created through compressive actomyosin flows, and the role of this effective purse-string for the alignment of the PAR domains to the cell shape. Through quantitative image analysis, genetic perturbations and theoretical modeling, we show that both mechanisms operate in the cell, even though the mechanism that relies on nematic line-tension dominates the alignment dynamics by about an order of magnitude.

These results allow for a deeper understanding of how self-organized active processes allow for the convergence of a biochemical cellular pattern to the shape of a cell.

[1] P. Gross, K. Vijay Kumar, J. Bois, N. Goehring, C. Hoege, F. Jülicher, S. Grill, Guiding Self-Organized Pattern Formation in Cell Polarity Establishment Nature Physics 15, 293-300 (2019).

## **Phenotypic plasticity influences the success of collective invasion in lung cancer**

Seth Haney<sup>1</sup>, Emily Summerbell<sup>2</sup>, Yi Jiang<sup>3</sup>, Adam I. Marcus<sup>2,4</sup>, Maxim Bazhenov<sup>1</sup>

<sup>1</sup> Department of Medicine, University of California, San Diego, La Jolla, CA

<sup>2</sup> Winship Cancer Institute, Emory University, Atlanta, GA

<sup>3</sup> Department of Mathematics and Statistics, Georgia State University, Atlanta, GA

<sup>4</sup> Department of Hematology and Medical Oncology, Emory University, Atlanta, GA

Collective invasion in cancer is a critical step on the path to metastasis. We have shown in lung cancer that two phenotypically distinct cell types, leader and follower, play complementary roles during invasion. Highly invasive but less proliferative leader cells drive invasive behavior, whereas, highly proliferative and supportive follower cells succeed these leaders and stabilize the mitotic defects of the leader cells. Here, we demonstrate that plastic transitions occur between the leader/follower cell types. However, there is a large asymmetry between the transition rates, with follower to leader transition occurring much faster than leader to follower transitions. We used computational modeling to investigate how the plastic changes between leader and follower phenotypes might impact invasive success. During local invasion at the primary tumor site, we found that plasticity can be beneficial for invasion when signaling deficits exist (e.g. VEGF block). To determine how these effects may impact the establishment of metastatic colonies, we replanted successful invading packs in new, low density, environments with variable environmental conditions (e.g. ECM fiber density, length, leader/follower cell fitness, etc.). We found that both collective invasion and follower to leader plasticity shows significant benefits in establishing new metastatic colonies. Surprisingly, leader to follower plasticity was detrimental in establishing new colonies. This finding is supported by our earlier work showing that leader to follower plasticity occurs 100-fold slower than follower to leader plasticity. Our results demonstrate that: a) collective invasion plays an important role both at primary and metastatic tumor sites; b) plasticity producing new leaders (i.e. follower to leader transition) is critical in the early stages of establishing new colonies but plasticity producing new followers can be detrimental.

## **Regulation of the oxidative stress response via dynamic regulation of p53 and MAPK activity**

Ryan L. Hanson<sup>1</sup> and Eric Batchelor<sup>1</sup>

<sup>1</sup> Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 10 Center Drive, MSC 1500, Bethesda, MD 20892

Reactive oxygen species (ROS) naturally occur within cells as a byproduct of normal metabolic processes. Prolonged oxidative stress can lead to DNA damage, growth arrest, and cell death. Many therapeutic treatments for cancer rely on induction of ROS to drive cell death within tumors. Paradoxically, many cancers exhibit elevated levels of ROS and oxidative stress that favor tumor progression and hinder chemotherapeutic treatment. We sought to determine how individual cells regulate cell death and survival via oxidative stress in comparison to the DNA damage response (DDR) induced by DNA double strand breaks, a common mechanism of many chemotherapies.

Using a fluorescently tagged p53 fusion protein, we examined single cell dynamics of p53 by time lapse microscopy in response to DNA double strand breaks (DSBs) induced by neocarzinostatin (NCS) or oxidative stress induced by hydrogen peroxide treatment. Prior single cell studies had observed that temporal dynamics of p53 can specify distinct stress responses and dictate cell fate. Interestingly, we found that single cell p53 dynamics were largely similar in response to DSBs or hydrogen peroxide treatment over 24 hours. Despite similar p53 dynamics, cells treated with hydrogen peroxide exhibited significant cell death suggesting alternative pathways likely cooperate with p53 to encode stress specificity. Focusing on the MAPK signaling networks parallel stress response pathways, we paired our fluorescent p53 with kinase biosensors for ERK, JNK, and p38. We found that these kinases exhibit distinct temporal activation patterns in response to NCS or hydrogen peroxide.

To determine the biological function of these kinase activation patterns, we perturbed kinase activity with specific pharmacological inhibitors and characterized the impact on the oxidative stress response. We observed that stress-specific JNK activation patterns rely on activation of unique upstream kinases, ATM for NCS-induced JNK pulsatile activity and Src for H<sub>2</sub>O<sub>2</sub>-induced JNK biphasic activity. For the biphasic response to H<sub>2</sub>O<sub>2</sub>, inhibition of the initial phase increased cell death, suggesting that early JNK activation plays a pro-survival role in the oxidative stress response. The initial phase of JNK activity appeared to act through a mechanism independent of p53, as inhibition of JNK had minimal impact on p53 dynamics. Rather, we found that hydrogen peroxide treatment led to alteration in gene expression patterns of p53 targets, suggesting that p53, JNK, and JNK-regulated transcription factors form a network of coherent feedforward loops to coordinate expression of pro-survival and pro-apoptotic transcripts.

These findings suggest that by integrating temporally distinct signals from p53 and MAPKs, cells can increase the specificity of stress responses through regulation of key cell fate regulators. Additionally, by manipulating the dynamics of these pathways in conjunction with specific chemotherapies, we may be able to enhance the induction of cell death within tumors.

**Hao Nan**, University of California, San Diego  
*Divergent trajectories of single-cell aging*

Cellular aging is a complex process that involves many interwoven molecular processes. Studies in model organisms have identified many individual genes and factors that have profound effects on lifespan. However, how these genes and factors interact and function collectively to drive the aging process remains unclear. We investigated single-cell aging dynamics throughout the replicative lifespans of *S. cerevisiae* and found that isogenic cells diverge towards two aging paths, with distinct phenotypic changes and death forms. We further identified specific molecular pathways driving each aging fate and revealed that these pathways interact and operate dynamically to enable an early-life switch that governs the aging fate decision and the progression towards death. Our work uncovers the interconnected molecular pathways that drives the aging process and opens up the possibility of designing interventions that simultaneously target multiple network nodes, instead of single genes, to more effectively extend the healthspan.

## Precision medicine is advanced by precision microsystems

Amy E. Herr, Ph.D.,  
Professor, University of California, Berkeley  
Investigator, Chan Zuckerberg Biohub  
herrlab.berkeley.edu

### Abstract

Refined tumor classification underpins advances in cancer diagnostic and treatment strategies. The integration of morphological information (from pathology) with protein signaling at the single-cell level is a missing capability. In fact, oncoproteins and their proteoforms are implicated in tumor progression, metastasis, and drug resistance across different cancer types. For example, human epidermal growth factor receptor 2 (HER2, a.k.a. erbB2) is one oncoprotein that manifests as truncated isoforms (t-erbB2s). Full-length HER2 is amplified in 15-20% of invasive breast cancers (BCa) and is targeted by trastuzumab, a humanized monoclonal antibody therapy developed in 1998. Trastuzumab – arguably one of the most effective targeted cancer therapies ever developed – cannot target the truncated HER2 isoforms, as the t-erbB2 isoforms lack the extracellular trastuzumab-binding epitopes. Consequently, while the paradigm-shifting drug decreased mortality in HER2+ patients, ~50% of women with HER2+ tumors do not respond. While patients with t-erbB2 have worse progression-free survival with standard trastuzumab therapy, these patients do benefit from other treatments, such as tyrosine kinase inhibitors. A next-generation of cancer subtype classification tools are urgently needed.

Quantitative, protein-specific measurements play an important role in cancer sub-type classification and, thus, offer diagnostic, prognostic, and treatment value to patients. Immunohistochemistry (IHC) is an important example. In fact, immunoassays (including IHC) are the de facto standard for direct measurement of endogenous, unmodified oncoproteins. Unfortunately, immunoassays lack the specificity needed for quantitation and even detection of important proteins, including truncated isoforms like t-erbB2s. Consequently, this presentation will detail a suite of high-specificity, protein analysis tools – with single-cell and sub-cellular resolution – that a profile protein isoform expression. The precision microfluidic tools are designed to augment classic IHC and single-cell genomics and transcriptomics – shedding light on ‘blind spots’ in pathology.

We will describe microfluidic systems engineered for precise cellular and molecular manipulation and measurement, centered around a single-cell immunoblotting (native, western, complexes, and isoelectric focusing). Sample preparation including isolation and handling of single cell and sub-cellular compartments will be examined, with attention to the time scales of critical processes (e.g., cell membrane lysis, nuclear membrane lysis, and protein solubilization). Design for high sample throughput from one cell to thousands will be discussed. We will also discuss the single-cell assay design for molecular selectivity, including resolving protein targets by differences in key physicochemical properties (size, complexation, isoelectric point). Design, fabrication, and integration of standards to quantify and control technical variation will be presented. Both analytical variability (lack of isoform-specific antibody probes) and biological variability (small cell subpopulations diluted in bulk analysis) can render oncoproteins & their isoforms undetectable.

From an engineering vantage point, we will consider the dominant transport and reaction processes (sensitivities, time scales) during selective target detection by immunoassay, including the important role of thermodynamic partitioning of immunoprobe into an immunoassay scaffold, and informed design of new hydrogel metrology tools and materials to overcome transport limitations. Lastly, we look to new projects focused on affording high selectivity for protein targets like the t-erbB2's – with single-cell resolution – yet, at the scale of whole tissue slices that are central to tissue pathology.

Title: Pattern formation by multi-species bacterial colonies with programmed population-to-population interactions

A. I. Curatolo,<sup>1, 2</sup>, N. Zhou,<sup>3</sup>, Y. Zhao,<sup>1, 3</sup>, C. Liu,<sup>4</sup> A. Daerr,<sup>1</sup> J. Tailleur,<sup>1</sup>, and J. Huang<sup>3, 4, 5</sup>

<sup>1</sup> Université Paris Diderot, Sorbonne Paris Cité, MSC, UMR 7057 CNRS, 75205 Paris, France

<sup>2</sup> John A. Paulson School of Engineering and Applied Sciences and Kavli Institute for Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA

<sup>3</sup> School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, University of Hong Kong, Pok Fu Lam, Hong Kong, PR China

<sup>4</sup> Institute of Synthetic Biology Shenzhen, Institutes of Advanced Technology, Shenzhen 518055, PR China

<sup>5</sup> Shenzhen Institute of Research and Innovation, University of Hong Kong, Shenzhen 518057, PR China

The ability of living organisms to form patterns has been traditionally studied by genetics. We aim to generate unique patterns by rewiring the genetic circuitry controlling cell motility and cell-cell interactions. First, a single population of *E. coli* cells are programmed to regulate their movement by sensing local cell density. Furthermore, different modes of interactions are designed to mutually control the movements of two interacting populations of *E. coli* cells. Interesting patterns are formed by newly engineered cells. An engineered low-density mover strain spreads outwards and autonomously forms a sequential and periodic pattern with the single population programming. More complicated patterns are formed by programming two interacting populations. Moreover, we build theoretical models that satisfactorily fit our current experimental data, and also predicts some parameters which may significantly affect the pattern formation. The study of this self-organized spatial distribution of cells may help us to probe the principles underlying the formation of natural biological patterns, and to prepare for future engineering of biological structures.

# Improving Molecular Dynamics Simulations to Study the Assembly and Phase Separation of Disordered Proteins

Zongyang Qiu, Jian Zhu, Xiang Yu, and [Jing Huang](#)

School of Life Sciences, Westlake University, 18 Shilongshan Road, Hangzhou, 310024, Zhejiang, China

The abundance of intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) in eukaryotic proteomes has been widely recognized. The functional significance of disordered proteins is typically related to their assembly as well as their engagement in liquid-liquid phase separation (LLPS). Computer simulations, especially molecular dynamics (MD) simulations, are powerful tools to study these processes [1], given that the underlying computational models are accurate enough and the simulation techniques allows accessing relevant space and time scales. Here we present our efforts in improving MD simulations on both aspects.

Following our previous work in optimizing additive protein force fields for IDPs [2], we have further improved the Drude polarizable protein force fields. The optimization details and a few applications on the self-assembly of model peptides in a variety of complicated electrostatics environments will be presented. With respect to algorithm development, we will present our efforts to develop an implicit LLPS solvent model, and to couple atomistic MD with lattice boltzmann methods (LBMs) to include large-scale hydrodynamics. These methods will allow us to study phase separation that happens in crowded environment using MD simulations.

[1] Jing Huang and Alexander D. MacKerell, Force Field Development and Simulations of Intrinsically Disordered Proteins, *Curr. Opin. Struct. Biol.*, **48**, 40 (2018)

[2] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L. de Groot, Helmut Grubmueller, and Alexander D. MacKerell, CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins, *Nature Methods*, **14**, 71 (2017)

## Mapping proteins in living cells

Bo Huang <sup>1,2</sup>

<sup>1</sup> Department of Pharmaceutical Chemistry, UCSF, 1700 4<sup>th</sup> St, San Francisco, CA 94143

<sup>2</sup> Chan Zuckerberg Biohub, San Francisco, CA 94158

Cellular processes are orchestrated by a large number of biomolecules in a spatially and temporally coordinated manner within a tiny volume. To uncover the underlying organizational principles and their functional relevance, we are developing new fluorescent labeling methods and microscopy techniques to systematically map the spatial localization, temporal dynamics and activity profiles of proteins. In particular, we have developed the split fluorescent protein tagging method that allows large-scale generation of cell lines with endogenously-labeled proteins by CRISPR/Cas9-mediated gene editing. Correspondingly, we have also build a single-objective high-resolution light-sheet microscope (eSPIM) that enables high-throughput imaging of these cell lines. These tools have paved the way to the generation of an atlas of proteins in living cells.

## Characterization of FOXO Dynamics Following Inhibition of the EGFR/PI3K/Akt Signaling Pathway

Julie M. Huynh<sup>1,2</sup>, Kayenat S. Aryeh<sup>1</sup>, Lisa M. Shanks<sup>1</sup>, Andrew L. Paek<sup>1,3</sup>

<sup>1</sup>Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA

<sup>2</sup>College of Medicine-Tucson, University of Arizona, Tucson, AZ, USA

<sup>3</sup>Cancer Biology GIDP, University of Arizona, Tucson, AZ, USA

The FOXO transcription factor family (Foxo1, Foxo3a, Foxo4) is one of 19 subgroups in the large Forkhead family of proteins found in humans and mice. Members of the FOXO subgroup are activated by oxidative stress, growth factor depletion and metabolic starvation and play roles in apoptosis, cell cycle regulation and metabolism. Because of their diversity in response to environmental cues in addition to their involvement in multiple cellular processes, FOXOs are thought of as homeostasis regulators.

Activated FOXO results in transcription of genes involved in a variety of diverse and opposing responses. Subcellular localization of Foxo1, 3a and 4 are regulated through EGFR/PI3K/Akt signaling. Those three isoforms have three Akt phosphorylation sites that when phosphorylated, result in inactive and cytoplasmic FOXO. In the absence of EGFR/PI3K/Akt signaling, Foxo1, 3a and 4 are dephosphorylated, enter the nucleus and bind to promoters with its DNA consensus recognition site where it can activate or repress transcription of its target genes which include apoptotic, arrest and oxidative stress programs. It is unknown how FOXOs determine which gene program to activate.

Protein dynamics, meaning how the levels or location of a protein changes over time, is a mechanism cells use to encode information about which transcriptional gene program to activate. The p53 transcription system is an established example of a dynamically regulated system—it responds to multiple inputs and is capable of enacting multiple outputs (gamma radiation causes oscillations of p53 levels that results in cell-cycle arrest while UV causes a single burst of p53 can lead to cell death. FOXO's network is similar to the p53 system (multiple input, multiple output), making it a good candidate for dynamic regulation.

We have tagged Foxo1 and Foxo3a with a fluorescent marker (Venus) at its endogenous locus in MCF7 Breast Cancer cells and PC9 NSCLC cells, respectively, using CRISPR/Cas9 to characterize FOXO dynamic patterns following inhibition of different kinases in the EGFR/PI3K/Akt signaling network. We have acquired FOXO nuclear cytoplasmic shuttling dynamics using live cell microscopy in single cells. Through our studies, we hope to gain a better understanding of the dynamic profiles exhibited following inhibition of these kinases. Our long-term goal is to identify dynamic patterns that push cancer cells to more apoptotic gene programs and thus terminal fates following treatment with targeted therapies for cancer.

## **Modelling the mutation and selection of synthetic constructs in *E. coli***

Duncan Ingram<sup>1,3</sup>, Mark Isalan<sup>2,3</sup> and Guy-Bart Stan<sup>1,3</sup>

<sup>1</sup> Department of Bioengineering, Imperial College London

<sup>2</sup> Department of Life Sciences, Imperial College London

<sup>3</sup> Imperial College Centre for Synthetic Biology, Imperial College London

In synthetic biology, a recurrent challenge is to engineer cells with high protein yield while minimising gene expression burden. This is often tackled by tuning the regulatory parts of synthetic constructs (e.g. promoters and RBSs), however a frequently overlooked problem is mutation events that reduce protein output. Many factors influence the rate at which mutations propagate that are well-understood, but two yet to be studied in combination are (i) the construct's sequence, and (ii) the burden caused by its expression.

To better understand these phenomena in *E. coli*, we developed a mathematical model that explores the interplay between sequence, burden, mutation and selection at the population level. At its core, our model considers cells in continuous culture that transition between different mutation states over time. Mutation probability is affected by the sequence composition of the synthetic constructs, while growth and selection of mutant cells is governed by the synthetic regulatory parts due to their impact on resource allocation, burden and growth rate. We show how these features can be used to predict sequence and part designs that reduce mutation spread while maintaining stable protein expression, resulting in engineered cells with increased genetic stability and higher protein yield.

## Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments

Christopher A Jackson<sup>1,2</sup>, Dayanne M Castro<sup>2</sup>, Giuseppe-Antonio Saldi<sup>2</sup>, Richard Bonneau<sup>1,2,3,4,5</sup>, David Gresham<sup>1,2</sup>

<sup>1</sup>Center For Genomics and Systems Biology, NYU, New York, NY 10008, USA

<sup>2</sup>Department of Biology, NYU, New York, NY 10008, USA

<sup>3</sup>Courant Institute of Mathematical Sciences, Computer Science Department, New York University, New York, NY 10003, USA

<sup>4</sup>Center For Data Science, NYU, New York, NY 10008, USA

<sup>5</sup>Flatiron Institute, Center for Computational Biology, Simons Foundation, New York, NY 10010, USA

All living things regulate how they express their genes in response to the challenges they face in their environment. We need to know how genes are connected to their regulators in a network in order to understand how an organism functions and consequently how to modify or engineer it.

Single-cell RNA sequencing is a powerful new tool to measure the gene expression of individual cells out of a large population of cells. This technique has been applied to many organisms, and there are data sets with millions of single-cell expression profiles for mice and humans. We don't know how well computational methods for determining gene regulatory networks from traditional expression measurements like RNA sequencing or microarrays will apply to single-cell data. Each observation (cell) in a single-cell gene expression experiment is noisier and carries less information than a traditional experiment, but there are many more observations overall. In order to benchmark computational methods for determining gene regulatory networks, we need data where the underlying network is known. Simulating data from a defined correct answer is one way to test computational tools; another way is to use data from a simpler organism where we have more literature-derived knowledge. Mice and humans have lots of data, but it's harder to tell if computational tools work because we just don't know what the right output is supposed to be.

Yeast (*Saccharomyces cerevisiae*) has been a model organism for a hundred years, and we know a lot about how it works at a molecular biology level. We would like to benchmark our computational tools on single-cell data from yeast, instead of on simulated data. Yeast has been harder to do single-cell sequencing on than mammalian cells, and the few single-cell *Saccharomyces* data sets are only a few hundred cells total. We have adapted a commercial system for single-cell sequencing to *Saccharomyces*, and generated a yeast single-cell gene expression data set with 40,000 cells. We used this data to benchmark our regularized linear regression-based gene regulatory network inference tool (the Inferelator), and found that it performs well with no changes to the underlying model assumptions. We also find that single-cell data

lets us understand regulatory relationships between things that are heterogeneous within most samples (like the cell cycle) and things that are generally similar within most samples (like metabolism). Bulk RNA sequencing often hides the heterogeneous aspects of biology.

The data set that we have generated for yeast will be immediately useful to benchmark single cell computational tools against in general. The ability to study heterogeneous processes without having to perturb them will eventually let us figure out how things like cell cycle and metabolism are connected.

## Quantifying the Central Dogma in the p53 pathway in live single cells

Antonina Hafner<sup>1,2</sup>, José Reyes<sup>1,2</sup>, Jacob Stewart-Ornstein<sup>1</sup>, Ashwini Jambhekar<sup>1</sup>, Galit Lahav<sup>\*1</sup>

<sup>1</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>2</sup>These authors contributed equally to this work

Transcription factors integrate signals to turn target genes on or off to orchestrate appropriate cellular responses. Although the target genes of many transcription factors have been identified, the quantitative relationships between transcription factor levels, transcription of the target gene, and the protein levels of the target remain poorly understood. We have used fluorescent protein reporters to monitor how changes in the transcription factor p53 control the transcription and protein levels of its target gene p21 in live single cells. In response to ionizing radiation, p53 exhibits oscillating protein levels, allowing us to monitor p21 expression at different p53 levels. We found that p21 transcription tracked with p53 oscillations, whereas p21 protein steadily accumulated. p53 levels regulated the probability, but not magnitude, of p21 transcription activation. Variations in p53 levels between cells contributed to heterogeneous p21 transcription, whereas independent p21 alleles within single cells exhibited highly-correlated behaviors. Within single p53 pulses, transcription showed bursts of mRNA production that were not highly correlated between independent p21 alleles in the same cell. Pharmacologically elevating p53 had minor effects on the magnitude of p21 transcription but increased the probability of transcriptional activation, which led to a strong increase in p21 protein levels. Our results reveal quantitative mechanisms by which TF dynamics can regulate the protein levels of its targets.

## A New Approach to Metabolism: The Power of a Good ‘RACIPE’

Dongya Jia<sup>1</sup>, Mingyang Lu<sup>2</sup>, Kwang Hwa Jung<sup>3</sup>, Jun Hyoun Park<sup>3</sup>, B. Bishal Paudel<sup>4,5,6</sup>, Vito Quaranta<sup>4,5</sup>, Herbert Levine<sup>1,7,8</sup>, Benny Abraham Kaiparettu<sup>3,9</sup>, and José N. Onuchic<sup>1,10,11,12</sup>

<sup>1</sup>Center for Theoretical Biological Physics, Rice University, Houston, TX 77005

<sup>2</sup>The Jackson Laboratory, Bar Harbor, ME 04609

<sup>3</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030

<sup>4</sup>Department of Biochemistry, Vanderbilt University, Nashville, TN 37232

<sup>5</sup>Quantitative Systems Biology Center (QSBC), Vanderbilt University, Nashville, TN 37232

<sup>6</sup>Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908

<sup>7</sup>Department of Bioengineering, Northeastern University, Boston, MA 02115

<sup>8</sup>Department of Physics, Northeastern University, Boston, MA 02115

<sup>9</sup>Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030

<sup>10</sup>Department of Biosciences, Rice University, Houston, TX 77005

<sup>11</sup>Department of Physics and Astronomy, Rice University, Houston, TX 77005

<sup>12</sup>Department of Chemistry, Rice University, Houston, TX 77005

Metabolic plasticity allows cells to adjust their metabolic phenotypes. Both glycolysis and oxidative phosphorylation (OXPHOS) can be adapted by cells to meet their bioenergetic and biosynthetic requirements in a context-dependent manner. Despite the advance in studies focusing only on glycolysis or OXPHOS, it remains largely unknown how cells orchestrate different metabolic phenotypes. To address this question, there is an urgent need to develop systemic approaches to quantitatively study the interplay between glycolysis and OXPHOS.

Mathematical modeling approaches have been employed to elucidate metabolic plasticity. Constraint-based models including flux-balance analysis based on conservation of mass have been the most widely used approaches to simulate metabolism. In addition, modeling efforts have also been made to identify gene activity involved in metabolism. These computational studies offer a quantitative and dynamical perspective of metabolism mostly focusing on either metabolic pathways or gene activities. However, the alteration of the metabolic activity is often coupled with the change in gene activity, and *vice versa*. Thus, to comprehensively characterize metabolism, a modeling framework integrating gene regulation with metabolic pathways is needed.

Here, we establish a theoretical framework to elucidate metabolic decision-making by coupling gene regulation with metabolic pathways. After an extensive literature search, we construct a regulatory network of metabolism featuring regulations by both genes and metabolites. To identify the robust dynamical features of the regulatory network, we utilize a variation of our previously developed computational method called RAndom Circuit PErturbation (RACIPE). The overall strategy involves randomizing the modeling parameters for each simulation and collecting all stable steady solutions for statistical analysis, by which the most significant solution patterns can be identified.

Our modeling results demonstrate a direct association between the activities of AMPK and HIF-1, master regulators of OXPHOS and glycolysis, respectively, with the activities of three major metabolic pathways: glucose oxidation, glycolysis and fatty acid oxidation (FAO). Our modeling results indicate that in addition to glycolysis and OXPHOS, cells can acquire two additional metabolic phenotypes - a hybrid metabolic phenotype where cells actively use both glycolysis and OXPHOS and a metabolically inactive phenotype where cells exhibit low activity of both glycolysis and OXPHOS. We verify the model prediction using metabolomics and transcriptomics data from paired tumor and adjacent benign tissue samples from a cohort of breast cancer patients and RNA-sequencing data from The Cancer Genome Atlas and Gene Expression Omnibus. We further validate the model prediction by in vitro studies of aggressive triple-negative breast cancer (TNBC) cells and BRAF-mutated melanoma cells. The experimental results confirm that TNBC cells can maintain a hybrid metabolic phenotype and targeting both glycolysis and OXPHOS is necessary to eliminate their metabolic plasticity. The experimental results confirm that drug-tolerant BRAF-mutated melanoma cells can acquire a metabolically inactive phenotype upon long-term MAPK inhibition. In summary, our work serves as a platform to symmetrically study cellular metabolic plasticity by modulating both genes and metabolic pathways, through integrating mathematical modeling, data analysis with experiments.

## Reconstructing Molecular Histories using DNA Records

Mike Jin<sup>1,2</sup>, Charleston Noble<sup>1,3</sup>, Jeffrey Gerold<sup>3</sup>, Jocelyn Kishi<sup>1,2</sup>, Thomas Schaus<sup>1,2</sup>,  
and Peng Yin<sup>1,2</sup>

<sup>1</sup> Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA

<sup>2</sup> Department of Systems Biology, Harvard University, Boston, MA

<sup>3</sup> Program for Evolutionary Dynamics, Harvard University

At present, recording and storing time course data such as concentration or temperature remains largely within the domain of traditional computer-based recording. Using molecules to sense and record information, however, offers the advantage of parallel and distributed data acquisition. Current molecular recording methods can detect the occurrence of an individual event or determine the temporal order of multiple events, but a continuous, multiplexed data recorder remains elusive. The major challenge lies in the stochasticity of molecular events: changes in system properties translate only to changes in average molecular event rates, but each individual molecular recording is subject to random interactions and therefore generates very noisy data. By examining an ensemble of many parallel molecular event recordings over the same time period, however, time course data for that period can be reconstructed with high fidelity. We take advantage of the programmability of DNA, using Primer Exchange Reaction (PER) to implement a molecular recorder which documents concentration time course data in single-stranded DNA records. We show that time course data can be reconstructed with as few as 500 DNA records: we can recover up to three concentration changes happening within two hours by using a MCMC algorithm on short records obtained by Illumina sequencing of the PER concatemers produced, and for long records, we show that given simulated PER data we can reconstruct concentration time courses of arbitrary complexity with high accuracy. Theory suggests that the upper limit of reconstruction precision is unlimited. If adapted for use in biological systems, similar molecular recorders could provide a robust method for non-destructive, continuous recording of molecular time course information. Grant acknowledgments: Wyss Institute, ONR N000141310593, NSF CCF1317291.

## **The interplay between the FOXO and p53 signaling pathways in response to stress.**

Elizabeth Jose<sup>1</sup>, Lisa Shanks<sup>1</sup>, Andrew Paek<sup>1</sup>

<sup>1</sup>Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ.

The p53 and Forkhead Box O (FOXO) proteins are transcription factors that have a striking number of similarities despite their distinct evolutionary origins. Both p53 and FOXO are activated in response to a range of cellular stresses and upon activation upregulate genes in overlapping pathways including cell-cycle arrest and apoptosis. Both FOXO and p53 are tumor suppressors that are frequently downregulated in cancer. Finally, stresses that are canonical activators of FOXO (serum starvation) and p53 (DNA damage) have been reported to activate both pathways though the evidence for this is conflicting. Here, we want to determine the relationship between FOXOs and p53 after treatment with DNA damaging agents, serum starvation and a combination of each stimuli. We measure activity of each transcription factor by using Immunofluorescence and by tracking the dynamics of each protein in single cells. Since both these proteins can respond to multiple external stimuli and integrate these stimuli into cell fate decisions, determining which protein responds to which stimuli and how this response changes over time will help in understanding as to how cells make fate decisions. We want to define if there is any correlation between these pathways or if these pathways are isolated in terms of their upstream stimuli. We want to understand how cells coordinate the activation of these two pathways to determine if they are to live or die. Ultimately, we wish to use this information to design combination treatments that selectively target cancer cells.

## **TRACE – the power of AI and citizen science in the analysis of microscopy time-lapse data**

Katarzyna M. Kedziora<sup>1</sup>, David Borland<sup>2</sup>, Hong Yi<sup>2</sup>, Jonnah Holloway<sup>1</sup>, Wayne Stallaert<sup>1</sup>,  
Chi Pham<sup>1</sup>, Jeremy Purvis<sup>1</sup>

<sup>1</sup> Department of Genetics, Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

<sup>2</sup> Renaissance Computing Institute, University of North Carolina, Chapel Hill, 100 Europa Drive, Suite 540, Chapel Hill, NC 27517, USA.

Time-lapse microscopy experiments provide overwhelming amounts of dynamic data on the behaviors of single cells. In order to take a full advantage of all the collected information, however, individual cells must be accurately tracked over time. Although deep learning is currently revolutionizing microscopy image analysis with major breakthroughs in the fields of classification, segmentation, and super resolution, the use of AI in the domain of cell tracking has been challenging because of the insufficient number of well-annotated datasets available for training. Here, we present a solution to this problem by turning the task of cell tracking into a video game. TRACE is a gamified, user friendly web interface that makes it possible for anyone to contribute time and effort towards research projects in quantitative single-cell biology. TRACE enables both experts and citizen scientists to manually curate annotations that empower state-of-the-art convolutional neural network architectures.

Deep learning and TRACE made it possible to obtain fully tracked data sets of cell colonies growing under a variety of experimental conditions. We will present the logic of TRACE's crowdsourcing software design and give specific examples of the insights into single-cell behaviors revealed through enhanced image annotation. Our goal is to introduce TRACE to the quantitative biology community as a useful tool for collecting near-perfect annotations for dynamic single-cell data that will facilitate further development of deep learning algorithms in the field of cell tracking.

## Single-cell characterization of macrophage polarization plasticity and mechanisms of regulation

Ilana Kelsey<sup>1</sup>, Kathryn Miller-Jensen<sup>1,2</sup>

<sup>1</sup> Biomedical Engineering Department, Yale University, New Haven, CT 06511

<sup>2</sup> Department of Molecular, Cellular, and Developmental Biology, Yale University

Macrophages perform a wide variety of immune functions, and they respond to environmental signals by altering their phenotype (i.e., polarization). Although oversimplified, these phenotypes have been described as two polarization states on opposite ends of a spectrum: M1 (pro-inflammatory) and M2 (wound healing, or “alternative”). A key to macrophage function is their plasticity, allowing them to transition between phenotypes in response to environmental cues, but this plasticity is also coopted or dysregulated in a variety of disease states. One challenge in designing immunotherapies is a lack of understanding about the extent of polarization plasticity and how it is regulated. This challenge is compounded by the fact that macrophages exhibit high cell-to-cell variability within a population, necessitating profiling at single-cell resolution to fully characterize the underlying mechanisms regulating the plasticity and function of macrophages. Here we explore macrophage polarization in response to varied dose and temporal combinations of polarization cues. Our findings indicate that a commitment to an M2-like polarization state can be made upon exposure to either continuous IL-4 exposure at low doses (1 ng/mL) or with a short, high-dose pulse of IL-4 (100 ng/mL). In contrast, high-dose pulses of M1 cues (10 ng/mL LPS + 10 ng/mL IFN $\gamma$ ) are insufficient to stimulate a similar fraction of the population to commit to an M1-like state when compared to continuously stimulated macrophages. While all cells in the population are initially able to respond to either M1 or M2 stimuli as assessed by phosphorylation of STAT1 (M1) or STAT6 (M2), differences emerge further downstream and at later timepoints. By 24 hours, the M1 fraction of the population approaches 90% in continuously stimulated conditions only, while the M2 response is capped between 30-50% of the population in both continuous and pulsed conditions, as assessed by protein marker levels. A 15-minute pulse of IL-4 (M2) stimulates the expression of a variety of M2-marker genes up to 48 hours after stimulation, while only continuous stimulation with LPS+IFN $\gamma$  (M1) results in sustained mRNA transcription. Our data suggest that the M2 polarization state is maintained by alterations in the chromatin environment, as either continual or pulsed stimulation with IL-4 is sufficient to increase chromatin accessibility at M2-associated genes, including *Arg1*. Interestingly, a fraction of IL-4-stimulated cells are capable of remaining committed to an M2-like polarization state even if subsequently exposed to M1 stimuli, indicating that a stable commitment to polarization status is made by a fraction of the overall macrophage population. These findings show that macrophages require continuous exposure to pro-inflammatory cues in order to maintain an M1 response, while even short exposure to M2 cues is sufficient to stimulate a sustained alternative polarization state. We speculate that differences in response to M1 versus M2 polarization cues enables macrophages to respond efficiently to the complex environments found in the body, with a carefully regulated response to inflammatory cues

that requires continuous exposure, and a fractional M2 response that leaves a reservoir of cells capable of responding to a second challenge.

Blood clot contraction plays an important role in prevention of bleeding and in thrombotic disorders. We unveiled and quantified the structural mechanisms of clot contraction at the level of single platelets. We found that activated platelets bend and shorten individual fibrin fibers via their filopodia that undergo sequential extension and retraction, as if pulling hand-over-hand on a rope. Platelets also induce compaction of fibrin fibers into platelet-attached agglomerates. As a result of simultaneous pulling on multiple, closely spaced fibrin fibers, platelets pull themselves closer to each other and form secondary clusters larger than the initial aggregates. Contracting platelets actively remodel the fibrin network by increasing its density followed by enhanced clot stiffness. Kinetic analysis of the time course of structural and mechanical transitions revealed a multiphasic behavior with at least three distinct phases that differ in duration and rate constants. We also showed that the active phase of platelet-induced clot contraction is followed by platelet dysfunction and disintegration. After ~30 minutes of incubation with thrombin, human platelets disintegrated into cellular fragments containing organelles, such as mitochondria, glycogen granules, and vacuoles. This platelet fragmentation was preceded by  $\text{Ca}^{2+}$  influx, integrin  $\alpha_{\text{IIb}}\beta_3$  activation and phosphatidylserine exposure (activation phase), followed by mitochondrial depolarization, generation of reactive oxygen species, metabolic ATP depletion and impairment of platelet contractility along with dramatic cytoskeletal rearrangements, concomitant with platelet disintegration (death phase).

## Diverse ligands of the BMP pathway have flexible redundancy relationships

Heidi Klumpe<sup>1</sup>, Yaron Antebi<sup>2</sup>, Matthew Langley<sup>3</sup>, James Linton<sup>3</sup>, Christina Su<sup>3</sup>, and Michael Elowitz<sup>3,4</sup>

<sup>1</sup> Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup> Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>3</sup> Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>4</sup> Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

Combinations of Bone Morphogenetic Proteins (BMPs) activate the BMP pathway to guide multiple cell fates. However, BMPs have very similar protein sequences and activate the same transcription factors, so the usefulness of combining very similar ligands rather than using a single ligand remains unclear. Previously, we showed that cells can distinguish some BMP combinations from single BMPs, as two different BMPs can combine non-additively to activate the downstream pathway. Therefore, we hypothesized that the diversity of BMP ligands includes unique inputs with non-additive combinatorial interactions as well as redundant inputs to compensate for potential mutations. Because the response to combinations cannot be predicted from responses to individual BMPs, determining which of the many BMPs have unique pairwise interactions requires exhaustive measurement of responses to BMP pairs. Therefore, we combined mammalian cell engineering with robotic liquid handling to measure responses to all possible pairs of 10 BMPs in a BMP reporter cell line. While responses to most BMP pairs were additive or ratiometric, rare instances of synergy break the ligands into four groups of BMPs that have non-additive interactions with other groups. Members of the same group appear as perfect replacements, eliciting the same responses individually and in combination with any other BMP, whereas members of different groups have unique combinatorial properties. Surprisingly, these functional classifications cannot be predicted from similarity of amino acid sequence. Moreover, a mathematical model of BMP signaling shows that random biochemical parameter sets for 10 ligands rarely generate as many as four groups of equivalent signals. Finally, these redundancy relationships change with semi-rational perturbation of BMP receptor profile, giving insight into the hierarchical nature of BMP ligand similarity across all possible combinations of BMP ligands and receptors. Together, these results show how the redundancy of BMP ligands depends on cell context, with changes in BMP receptor expression masking or discriminating differences between BMPs. Overall, this framework helps describe how responses to BMP combinations depend on the context of BMPs and BMP receptors, and could guide the understanding or construction of BMP combinations used in development or therapies.

## Ultra-fast Cycling for Multiplexed Cellular Fluorescence Imaging

Jina Ko<sup>1</sup>, Juhyun Oh<sup>1</sup>, Maaz S. Ahmed<sup>1</sup>, Jonathan C. T. Carlson<sup>1,2</sup>, Ralph Weissleder<sup>1,3</sup>

<sup>1</sup> Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114

<sup>2</sup> Cancer Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114

<sup>3</sup> Department of Systems Biology, Harvard Medical School, Boston, MA 02115

We have developed an ultra-fast (< 1 sec), highly efficient (>95%) quenching method for multiplexed protein profiling from single cells. We implemented tetrazine (Tz) / *trans*-cyclooctene (TCO) click chemistry for staining, quenching, and cycling using clickable fluorophores and quenchers. This gentle and rapid cycling method was used to profile different protein targets from single cells obtained using fine needle aspiration (FNA), a sample type that does not work well with existing cycling technologies due to its fragility. With a site-specific delivery of fluorescence quenchers and remarkable acceleration of chemical reaction kinetics (3 orders of magnitude faster than predicted), we identified different immune cell types (N=7) by profiling 12 markers within an hour. This ultra-fast, highly efficient quenching technology opens the possibility for the development of diagnostics via multiplexed single cell profiling.

## Effect of synaptic cell-to-cell transmission and recombination on the evolution of double mutants in HIV

Jesse Kreger<sup>1</sup>, Natalia L. Komarova<sup>1</sup>, and Dominik Wodarz<sup>2,1</sup>

<sup>1</sup>Department of Mathematics, University of California Irvine, Irvine CA 92697

<sup>2</sup>Department of Population Health and Disease Prevention Program in Public Health, University of California Irvine, Irvine CA 92697

Recombination in HIV infection can impact virus evolution *in vivo* in complex ways, as has been shown both experimentally and mathematically. The effect of free virus versus synaptic, cell-to-cell transmission on the evolution of double mutants, however, has not been investigated. In this talk we investigate this using a stochastic agent-based model. Consistent with data, we assume spatial constraints for synaptic but not for free-virus transmission. Two important effects of the viral spread mode are observed: (i) For disadvantageous mutants, synaptic transmission protects against detrimental effects of recombination on double mutant persistence. Under free virus transmission, recombination increases double mutant levels for negative epistasis, but reduces them for positive epistasis. This reduction for positive epistasis is much diminished under predominantly synaptic transmission, and recombination can in fact lead to increased mutant levels. (ii) The mode of virus spread also directly influences the evolutionary fate of double mutants. For disadvantageous mutants, double mutant production is the predominant driving force, and hence synaptic transmission leads to highest double mutant levels due to increased transmission efficiency. For advantageous mutants, double mutant spread is the most important force, and hence free virus transmission leads to fastest invasion due to better mixing. For neutral mutants, both production and spread of double mutants are important, and hence an optimal mixture of free virus and synaptic transmission maximizes double mutant fractions. Therefore, both free virus and synaptic transmission can enhance or delay double mutant evolution. Implications for drug resistance in HIV are also discussed.

## Co-Transcriptional RNA Secondary Structure and Alternative Splicing

Tammy C.T. Lan<sup>1</sup>, Harish Swaminathan<sup>1</sup>, Silvi Rouskin<sup>1</sup>

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, MA

Alternative splicing (AS) occurs in nearly all human genes and plays a central role in gene expression regulation. AS expands proteomic diversity by an order of magnitude and is critical in determining cell fate and development; errors in splicing are associated with numerous genetic diseases and cancers. Co-transcriptional RNA secondary structures at splice junctions have long been proposed to play a role in regulating pre-mRNA splicing: by 1) affecting spliceosome accessibility to the pre-mRNA, 2) interacting with splicing factors and 3) promoting long-range RNA interactions in a protein-independent fashion. However, we lack a mechanistic and predictive understanding of how RNA structures affect AS decisions

A well-characterized example is the human microtubule-associated protein Tau (MAPT), where a predicted stem-loop structure at the 5' splice site of exon 10 affects the exon inclusion rate. Mutations that destabilize the 5' splice site structure promote exon inclusion and are strongly linked to several neurodegenerative diseases. However, it is still unclear whether MAPT represents a singular case of structure affecting splicing or is an example of a common mechanism to regulate splicing.

Currently, limited approaches exist to probe co-transcriptional RNA secondary structures *in vivo*. We have developed a method based on chemical probing coupled with nascent-RNA labelling and next-generation sequencing to identify both gene-specific and genome-wide co-transcriptional RNA structures. Using this method, we revealed a direct relationship between the stability of co-transcriptional structures measured in cells and exon skipping. In addition, genome-wide analysis of our approach confirmed enrichment of nascent RNA sequences, enabling us to globally assess the role of pre-mRNA structure in splicing.

## **The Dynamics of FOXO Transcription Factors in Response to Cellular Stress**

Kathleen Lasick<sup>1</sup>, Sneha Srinivasan<sup>1</sup>, Lisa Shanks<sup>1</sup>, Andrew Paek<sup>1</sup>

<sup>1</sup> Molecular and Cellular Biology, UArizona, 1007 E Lowell St, Tucson, AZ 85719

The Forkhead Box O, or FOXO transcription factors have been implicated in lifespan extension in a diverse set of species including worms, flies and mice. FOXO's are likely to play a role in human aging as well, as alleles of FOXO3a have been consistently associated with increased human lifespan in longevity studies. FOXO's function as homeostatic regulators at both the cellular and tissue level. They are activated in response to different stresses, including reactive oxygen species (ROS), starvation, hypoxia and in some instances are upregulated in non-stress conditions. FOXO's are mainly controlled by nuclear/cytoplasmic shuttling. Serine/Threonine kinases, predominantly in the insulin/PI3K pathway, phosphorylate FOXOs which inhibits their NLS and sequesters them in the cytoplasm. Stresses typically inhibit phosphorylation which leads to FOXO entering the nucleus where they can upregulate and downregulate hundreds of genes. These include a diverse set of genes including cytoprotective programs like ROS scavenger genes but also cytotoxic apoptotic genes. How FOXO's determine different fates is not known. Due to the nuclear/cytoplasmic shuttling of FOXO's in response to different stimuli, we hypothesize that the dynamic patterns of nuclear localization are responsible for FOXO's ability to induce different outcomes. I am testing these hypotheses using live cell microscopy of fluorescently tagged Foxo1 in response to different stresses. I have determined that Foxo1 responds to ROS and serum starvation via different patterns of activation. This research will add to our understanding of how FOXO's regulate cellular response to a variety of stresses.

**Deconstructing Network Inference:  
Filling in the gaps between chromatin accessibility and gene expression.**

Danxun Li<sup>1</sup>, Ashley Hine<sup>2</sup>, Mei San Tang<sup>3</sup>, P'ng Loke<sup>2</sup>, and Rich Bonneau<sup>1</sup>

<sup>1</sup> Systems Biology, Center for Computational Biology, Flatiron Institute, 162 5<sup>th</sup> Avenue, New York, NY 10011 <sup>2</sup> Department of Microbiology, NYU Grossman School of Medicine, 550 First Avenue, New York NY 10016 <sup>3</sup> Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, 660. S Euclid Ave., St. Louis, MO 63110

Synthesizing biological insight by integrating multi-omic datasets continues to be nontrivial. We can feed them into a variety of inference schemes to generate gene regulatory networks or we can directly correlate gene accessibility with gene expression. The former approach often feels abstracted from the data whereas the latter is too coarse-grained to capture the biology that occurs between chromatin opening and transcription. Here, we describe a framework in which we build up a plausible disease network from mucosal immune cells taken from IBD patients by stepping the datasets through the biological system that we are investigating. Starting with differentially accessible chromatin regions in samples taken from the inflamed mucosa, we probe for enriched motifs, examine the expression of candidate transcription factors and verify the transcription levels of their downstream targets. We show that we can recover aspects of previous statistically inferred networks in our patient dataset and discuss implications of both novel interactions and absence of prominent previously inferred interactions in our approach. We also describe a potential makeup of cell types involved in the inflammatory response by attributing parts of the network to regulatory mechanisms characteristic of various cell types. With this approach we hope to introduce a paradigm in which the biological system directs the analysis and various proxies are developed to expose the machinery at different stages of the pipeline.

# **Ocean wave model: organization principles of nonribosomal peptide synthetase pathways revealed by intensive recombination of a marine microbe**

Zhiyuan Li<sup>1</sup>, Jindong Zan<sup>2</sup>, Mohamed S. Donia<sup>2</sup>

<sup>1</sup>Center for Quantitative Biology and Center for Life Science, Peking University, 100871, China

<sup>2</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey, 08544, USA

A marine bacteria with special symbiotic lifestyle, recently discovered in the Black Point Bay of Hawaii, may reveal general organization principles of a class of natural products named nonribosomal peptide synthetases (NRPSs). NRPS is a class of multimodular enzymes widely exist in the microbial world, assemble diverse peptides that have tremendous pharmaceutical implications. The modular structure of NRPS has inspired the de novo engineering by recombining together modules or domains of existing NRPSs together, yet efforts following this direction are largely unsuccessful. In our newly-discovered marine bacteria *Ca. E. kahalalide* (cEK) which lives intracellularly to marine alga *Bryopsis* sp. and provides chemical defense for its host, 20 NRPSs pathways with 120 modules exchange gene fragment in an unprecedented frequency. Such intensive and indiscriminative recombination acts like an "ocean wave", which erases the trivial sequence association induced by sharing phylogeny, and leaves the important functional constraints in NRPS observable. By characterizing these sequence covariations, we revealed a non-modular and even non-domain based organization principle of NRPSs in cEK. Moreover, some of the functional constraints we observed in cEK can be recovered in a large database of NRPS, if we computationally restricting the interference from sharing phylogeny. Our analysis presented a bacteria NRPS system with unusual capacities of recombination, provided insights into the organization principles of NRPSs, and suggested computational methods in uncovering functional constraints in general.

Advances in synthetic biology have led to an arsenal of proof-of-principle bacterial circuits that can be leveraged for applications ranging from therapeutics to bioproduction. A unifying challenge for most applications is the presence of selective pressures that lead to high mutation rates for engineered bacteria. A common strategy is to develop cloning technologies aimed at increasing the fixation time for deleterious mutations in single cells. We adopt a complementary approach that is guided by ecological interactions, whereby cyclical population control is engineered to stabilize the functionality of intracellular gene circuits. Three strains of *Escherichia coli* were designed such that each strain could kill or be killed by one of the other two strains. The resulting “rock-paper-scissors” dynamic demonstrates rapid cycling of strains in microfluidic devices and leads to an increase in the stability of gene circuit functionality in cell culture.

# The Circadian Clock Generates Rhythmicity in DNA Replication to Protect the Cyanobacterial Genome

Yi Liao<sup>1, 2</sup>, and Michael J. Rust<sup>1,2,3</sup>,

<sup>1</sup> Department of Molecular Genetics and Cell Biology

<sup>2</sup> Institute for Genomics and Systems Biology

<sup>3</sup> Department of Physics, University of Chicago, Chicago, IL 60637

To cope with daily cycling of the external environment imposed by Earth's rotation, organisms from bacteria to humans have evolved circadian clock to coordinate their physiology in anticipation of the rising and setting of the sun. The simplest organism to possess a circadian clock is the cyanobacterium *Synechococcus elongatus*, a photoautotroph that strictly depends on light for photosynthesis, metabolism and growth. To maximize fitness in the day-night environment, the *S. elongatus* circadian clock exhibits controls over a diverse range of cellular processes, including transcription, energy storage, and cell division. However, it is unknown whether DNA replication, a fundamental process in all domains of life, is under the clock control. In addition, it has been shown among many species, including *S. elongatus*, that misalignment between the circadian clock and the external environments often results in reduced fitness and viability at the individual as well as the population level, but the causes for such deleterious effects remain elusive. Here, we show that in *S. elongatus* DNA replication is regulated by the circadian clock, which schedules rhythmic assembly of the replication machinery, ensuring that replication mostly occurs early in the day, and any remaining replication after sunset is able to complete as well. Mathematical modeling and experimental data both demonstrate that when the clock and environment are misaligned, the stability of the replisome is compromised, and ongoing replications aborted, leaving cells with incomplete chromosomes that persist throughout the night. Our study thus reveals that a major function of the circadian clock is to safeguard the integrity of the cyanobacterial genome in the diurnal cycle.

# **Mathematical modeling of metabolic cross-feeding constrained by laboratory experiments delineates limits of robustness of mixed microbial communities**

Chen Liao<sup>1</sup>, Tong Wang<sup>2,3</sup>, Sergei Maslov<sup>3,4</sup>, Joao B. Xavier<sup>1,\*</sup>

<sup>1</sup>Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, USA,

<sup>2</sup>Department of Physics, University of Illinois at Urbana-Champaign, IL 61801, USA,

<sup>3</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, IL 61801, USA

<sup>4</sup>Department of Bioengineering, University of Illinois at Urbana-Champaign, IL 61801, USA

Mathematical models of microbial communities span a wide range of scales: from models that exhaustively detail the biochemistry of cell-cell interactions, to coarse-grained population dynamics models that abstract most molecular details into species-species interaction matrices. Deciding the appropriate modeling scale to extract from empirical data principles governing microbial communities remains a challenge. Here we propose a model of cross-feeding interactions that lies at an intermediate scale between the complexity of genome-scale metabolic models and consumer-resource models, and we show how that the model can be constrained by experimental data to reveal the limits of robustness of mixed communities. We apply our method to three published examples of mixed laboratory communities of increasing complexity all consisting of cross-feeding *Escherichia coli* strains: uni-, bi-, and multi-directional cross-feeding of by-products or essential nutrients. Models constrained by data from each of these three systems reveal the limits of community robustness to perturbations, such as nutrient downshifts, antibiotic treatments and invasion by cheaters. Our results supports the emerging view in the literature that modeling microbial processes at an intermediate scale provides a practical framework for applications to real world multi-species communities such as the human microbiome.

## Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis

Akihide Yoshimi<sup>1,\*</sup>, **Kuan-Ting Lin**<sup>2,\*</sup>, Daniel H. Wiseman<sup>3,4,\*</sup>, Mohammad Alinoor Rahman<sup>2</sup>, Alessandro Pastore<sup>1</sup>, Andrew M. Intlekofer<sup>1</sup>, Bo Wang<sup>1</sup>, Stanley Chun-Wei Lee<sup>1</sup>, Jean-Baptiste Micol<sup>5</sup>, Xiao Jing Zhang<sup>1</sup>, Stephane de Botton<sup>5</sup>, Virginie Penard-Lacronique<sup>5</sup>, Eytan Stein<sup>6</sup>, Hana Cho<sup>1</sup>, Rachel E. Miles<sup>1</sup>, Daichi Inoue<sup>1</sup>, Todd R. Albrecht<sup>7</sup>, Tim C.P. Somervaille<sup>3</sup>, Kiran Batta<sup>4</sup>, Fabio Amaral<sup>3</sup>, Fabrizio Simeoni<sup>3</sup>, Deepti P. Wilks<sup>8</sup>, Catherine Cargo<sup>9</sup>, Craig B. Thompson<sup>10</sup>, Ross L. Levine<sup>1,6</sup>, Heidi Dvinge<sup>11</sup>, Robert K. Bradley<sup>12</sup>, Eric J. Wagner<sup>7</sup>, Adrian R. Krainer<sup>2</sup>, Omar Abdel-Wahab<sup>1,6\*</sup>

- <sup>1</sup> Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
- <sup>2</sup> Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
- <sup>3</sup> Leukaemia Biology Laboratory, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, UK
- <sup>4</sup> Division of Cancer Sciences, The University of Manchester, Manchester, UK
- <sup>5</sup> Gustave Roussy, Université Paris-Saclay, Villejuif, France
- <sup>6</sup> Leukemia Service, Dept. of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA
- <sup>7</sup> Department of Biochemistry & Molecular Biology, The University of Texas Medical Branch at Galveston, Galveston, Texas, 77550, USA
- <sup>8</sup> Manchester Cancer Research Centre Biobank, The University of Manchester, Manchester, UK.
- <sup>9</sup> Haematological Malignancy Diagnostic Service, St James's University Hospital, Leeds LS29 7TF, UK
- <sup>10</sup> Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA.
- <sup>11</sup> Department of Biomolecular Chemistry, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, 53706, USA
- <sup>12</sup> Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Transcription and pre-mRNA splicing are key steps in the control of gene expression and mutations in genes regulating each of these processes are common in leukaemia. Despite the frequent overlap of mutations affecting epigenetic regulation and splicing in leukaemia, how these processes influence one another to promote leukaemogenesis is not understood and, to our knowledge, there is no functional evidence that mutations in RNA splicing factors initiate leukaemia. Here, through analyses of transcriptomes from 982 patients with acute myeloid leukaemia, we identified frequent overlap of mutations in IDH2 and SRSF2 that together promote leukaemogenesis through coordinated effects on the epigenome and RNA splicing. Whereas mutations in either IDH2 or SRSF2 imparted distinct splicing changes, co-expression of mutant IDH2 altered the splicing effects of mutant SRSF2 and resulted in more profound splicing changes than either mutation alone. Consistent with this, co-expression of mutant IDH2 and SRSF2 resulted in lethal myelodysplasia with proliferative features *in vivo* and enhanced self-renewal in a manner not observed with either mutation alone. IDH2 and SRSF2 double-mutant cells exhibited aberrant splicing and reduced expression of INTS3, a member of the integrator complex, concordant with increased stalling of RNA polymerase II (RNAPII). Aberrant INTS3 splicing contributed to leukaemogenesis in concert with mutant IDH2 and was dependent on mutant SRSF2 binding to cis elements in INTS3 mRNA and increased DNA methylation of INTS3. These

data identify a pathogenic crosstalk between altered epigenetic state and splicing in a subset of leukaemias, provide functional evidence that mutations in splicing factors drive myeloid malignancy development, and identify spliceosomal changes as a mediator of IDH2-mutant leukaemogenesis.

- The talk will focus on the computational analysis (mutation and splicing) in this work
- The work may inspire a new field in quantitative biology to study the interactions between epigenetic regulation and RNA splicing

\* Equal contribution, co-first authors

# Dynamic Modeling of the Eukaryotic Transcription Cycle

Jonathan Liu<sup>1</sup>, Hernan Garcia<sup>1,2</sup>

<sup>1</sup> Department of Physics, UC Berkeley, Berkeley, CA 94720

<sup>2</sup> Department of Molecular and Cell Biology, UC Berkeley, Berkeley, CA 94720

The eukaryotic transcription cycle is a complex process by which an RNA polymerase molecule loads onto a gene, initiates transcription, elongates through the body of the gene, produces an mRNA transcript, and then terminates to be recycled for future usage. Although the biochemical aspects of this process are well-known, dynamic investigations have been difficult to conduct, mainly due to the challenges involved in imaging transcription in real-time *in vivo*. Here, we describe a novel experimental technique to visualize the transcription cycle in developing fruit fly embryos in real time. With a simple model, we develop a statistical methodology to simultaneously infer all of the effective parameters of the transcription cycle at single-cell resolution. Initial results agree with previous measurements and additionally suggest a biophysical coupling between transcriptional activity and polymerase termination rates.

# **An Integrative Circuit-Host Modeling Framework for Describing Synthetic Gene Networks**

Chen Liao, Andrew Blanchard, and [Ting Lu](#)

Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL61801

## **Abstract**

One fundamental challenge in synthetic biology is the lack of quantitative tools that accurately describe and predict the behaviors of engineered gene circuits. This challenge arises from multiple factors, among which the complex interdependence of circuits and their host is a leading cause. Here we present a gene circuit modelling framework that explicitly integrates circuit behaviors with host physiology through bidirectional circuit-host coupling. The framework consists of a coarse-grained but mechanistic description of host physiology that involves dynamic resource partitioning, multilayered circuit-host coupling including both generic and system-specific interactions, and a detailed kinetic module of exogenous circuits. We showed that, following training, the framework was able to capture and predict a large set of experimental data concerning the host and its foreign gene overexpression. To demonstrate its utility, we applied the framework to examine a growth-modulating feedback circuit whose dynamics is qualitatively altered by circuit-host interactions. Using an extended version of the framework, we further systematically revealed the dynamics of a toggle switch across scales from single-cell dynamics to population structure and to spatial ecology. This work advances our quantitative understanding of gene circuit behaviors and also benefits the rational design of synthetic gene networks.

## **The Jonckheere-Terpstra-Kendall based nonparametric algorithm to detect differentially expressed genes in time-series omics data**

Hitoshi Iuchi<sup>1, 2</sup> and Michiaki Hamada<sup>1, 2\*</sup>

<sup>1</sup> Computational Bio Big-Data Open Innovation Laboratory, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

<sup>2</sup> Department of Electrical Engineering and Bioscience, Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Much biological research, especially in developmental or physiological biology focuses on the gene expression pattern changing over time. Despite the decreasing experimental cost of omics analysis, genome-wide analysis has continued to have formidable challenges. In contrast to imaging, sampling for sequencing analysis and mass spectrometry is destructive, requiring many samples. This experimental limitation causes time-resolved datasets are sparse, with limited replicates, and low signal/noise ratio. The restriction prevents accurate identification of differently expressed genes. Here we present Jonckheere-Terpstra-Kendall (JTK) based nonparametric algorithm to detect differentially expressed genes in time-series omics data. The Jonckheere-Terpstra test is a nonparametric test for detecting orderings on independent groups. Kendall's tau is a rank correlation statistics to measure the ordinal association between two quantities. We aimed to identify time-resolved differentially expressed genes from limited data with statistical confidence by a combination of two statistics. In the poster session, we present a fundamental idea, an application for simulation data and discuss the difference between JTK-based algorithms and other approaches. Also, we discuss scalability and statistical justification of JTK based algorithm through application for actual sequence data of developmental biology.

# Optimal tuning of weighted kNN and diffusion-based methods for denoising single cell RNA-seq data

Andreas Tjärnberg <sup>4</sup>, Omar Mahmood <sup>1</sup>, Giuseppe Saldi <sup>3</sup>, Chris Jackson <sup>2</sup>, Lionel A Christiaen <sup>4</sup>, Richard A Bonneau <sup>1, 2, 3</sup>

<sup>1</sup> NYU Center for Data Science, New York, NY

<sup>2</sup> Flatiron Institute Center for Computational Biology, New York, NY

<sup>3</sup> NYU Department of Biology, New York, NY

<sup>4</sup> NYU Center for Developmental Genetics, New York, NY

The analysis of single cell genomics data presents several statistical challenges, for example it is often noisy and sparse. Extensive efforts have been made to impute values thought to be missing due to experimental limitations, sampling issues and noise, while maintaining cell individuality in the expression space. In spite of such efforts no consensus on best practice has been established and all current approaches vary based on the available data and empirical tests.

Existing imputation-like methods such as the MAGIC algorithm and kNN-impute make use of the k Nearest Neighbour Graph (kNN-G). The kNN-G is often used to infer the identities of and relationships between cells; it is the basis of widely used dimensionality reduction and projection methods. However, due to the lack of an optimal objective function for choosing hyperparameters, these imputation-like methods tend to over-smooth data thereby resulting in a loss of information with regard to cell identity and the specific gene-to-gene patterns underlying regulatory mechanisms.

In our work, we investigate the tuning of kNN and diffusion-based denoising methods with a novel non-stochastic method for optimally preserving biologically relevant variance in single cell data. The framework, *Denoising Expression data with a Weighted Affinity Kernel and Self Supervision* (DEWÄKSS), uses a self-supervised objective function to tune its parameters, in accordance with the principles behind the noise2self [1] algorithm. We demonstrate that denoising with optimal parameters selected by our objective function (i) is robust to preprocessing methods using data from established benchmarks, (ii) disentangles cellular identity and maintains robust clusters over dimensionality reduction methods, (iii) maintains variance along several expression dimensions, unlike previous heuristic-based methods that tend to over-smooth data variance, and (iv) rarely involves diffusion but rather uses a fixed weighted kNN graph only once. Together, these findings provide a new understanding of kNN and diffusion-based denoising methods that serve as a foundation for future research.

[1] Joshua D. Batson and Loïc Royer. Noise2self: Blind denoising by self-supervision. CoRR, abs/1901.11365, 2019.

**Suliana Manley, EPFL**

*Super-resolution microscopy reveals mitochondrial organization and dynamics*

Super-resolution microscopy enables imaging beyond the diffraction limit, while maintaining the specificity and live-cell imaging capability of fluorescence microscopy. However, it has been technically challenging to collect large datasets required to capture or account for the variability of biological systems. Previously, we created automated and large field of view single molecule localization microscopy (PALM/STORM), which we now extend to instant structured illumination microscopy (iSIM). We demonstrate the power of this approach for studying the architecture of supramolecular protein complexes as well as the organizing principles of mitochondrial division and transcription

# **A geometric and structural approach to the analysis and design of biological circuit dynamics — a theory tailored for synthetic biology**

John P. Marken<sup>1\*</sup>, Fangzhou Xiao<sup>1\*</sup>, Richard M. Murray<sup>1,2</sup>

<sup>1</sup> Department of Bioengineering, Caltech, 1200 E California Blvd, Pasadena, CA 91125

<sup>2</sup> Department of Control and Dynamical Systems, Caltech, Pasadena, CA

Much of the progress in developing our ability to successfully design genetic circuits with predictable dynamics has followed the strategy of molding biological systems to fit into conceptual frameworks used in other disciplines, most notably the engineering sciences. Because biological systems have fundamental differences from systems in these other disciplines, this approach is challenging, and the insights obtained from such analyses are often not framed in a biologically intuitive way. We have recently developed a new theoretical framework for analyzing the dynamics of genetic circuits that is tailored towards the unique properties associated with biological systems and experiments. The premise of the framework is to conceptualize these circuits into simpler structural forms that correspond to the saturation states of the system's various components. Much of the system's dynamics can then be represented by the transitions between these saturation regimes. The mathematical details of the theory are presented in the accompanying poster by Xiao *et al*— here, we apply this framework to the analysis of two case studies, the leaky positive autoactivation motif and the Repressilator. We illustrate how our framework can conceptualize multistability and oscillations within a saturation-based mindset, and highlight how our framework enables a geometric visualization of the system dynamics that allows an intuitive understanding of the structural sources of these dynamics.

\* Equal contributions

# Theoretical Design of Paradoxical Signaling-Based Synthetic Population Control Circuit in *E. coli*

Michaëlle N. Mayalu<sup>1</sup>, and Richard M. Murray<sup>1,2</sup>

<sup>1</sup>Department of Computing and Mathematical Sciences, California Institute of Technology, 1200 E California Blvd, Pasadena, CA, 91125

<sup>2</sup>Division of Biology and Biological Engineering, California Institute of Technology, 1200 E California Blvd, Pasadena, CA, 91125

We have developed a mathematical framework to analyze the cooperative control of cell population homeostasis via paradoxical signaling in synthetic contexts. Paradoxical signaling functions through quorum sensing (where cells produce and release a chemical signal as a function of cell density). Precisely, the same quorum sensing signal provides both positive (proliferation) and negative (death) feedback in different signal concentration regimes. As a consequence, the relationship between intercellular quorum sensing signal concentration and net growth rate (cell proliferation minus death rates) can be non-monotonic. This relationship is a condition for robustness to certain cell mutational overgrowths and allows for increased stability in the presence of environmental perturbations. Here, we explore stability and robustness of a conceptualized paradoxical feedback population control synthetic circuit. Furthermore, we assess possible design principles that could exist among a subset of paradoxical circuit implementations. This analysis sparks the development for a bio-molecular theory to identify ideal underlying characteristics for paradoxical signaling control systems.

# Modeling Adaptive Chemotherapy Dosing for Competing Drug Sensitive and Drug Resistant Cancer Cell Populations

Cordelia McGehee<sup>1,2</sup> and Yoichiro Mori<sup>2,3,4</sup>

<sup>1</sup>Medical Scientist Training Program, Mayo Clinic, Rochester, MN

<sup>2</sup>School of Mathematics, University of Minnesota, Minneapolis, MN

<sup>3</sup>Department of Mathematics, University of Pennsylvania, Philadelphia, PA

<sup>4</sup>Department of Biology, University of Pennsylvania, Philadelphia, PA

Despite significant therapeutic advances in the treatment of metastatic cancers, acquired chemotherapy resistance remains common. Recently, adaptive dosing strategies where a chemotherapy is given in increments based on ongoing measurements of tumor burden have been proposed to increase progression free survival by exploiting intratumoral cell-cell competition. These therapy regimens are based on underlying ecological models of competition and are designed to control emerging chemotherapy resistant cells within a tumor by maintaining a population of chemotherapy sensitive cells to outcompete the resistant ones. However, rigorous mathematical exploration of the dynamics of the accompanying mathematical and conceptual models has been limited. Exploiting techniques from dynamical systems and control theory, we explore the dynamics of a two species intratumoral competition model where one cell type is drug sensitive and the other is drug resistant. We show that under certain sets of competition parameters and initial conditions, the tumor size can always be controlled and that the resistant cell population can be decreased to near zero. We are further able to compute lower limits on the total dose of drug administered to decrease the number of resistant cells to near zero as well as the minimum time necessary to achieve this disease control. Finally, we explore disease treatment strategies when control of the resistant cell population is not achievable in order to compare adaptive versus non-adaptive treatment regimens.

## **A new experimental platform facilitates assessment of the chromatin and transcriptional landscapes of aging yeast**

David G Hendrickson, Ilya Soifer, Bernd J Wranik, Griffin Kim, Michael Robles, Patrick A Gibney, R Scott McIsaac

Calico Life Sciences LLC

Replicative aging of *Saccharomyces cerevisiae* is an established model system for eukaryotic cellular aging. A limitation in yeast lifespan studies has been the difficulty of separating old cells from young cells in large quantities. We engineered a new platform, the Miniature-chemostat Aging Device (MAD), that enables purification of aged cells at sufficient quantities for genomic and biochemical characterization of aging yeast populations. Using MAD, we measured DNA accessibility and gene expression changes in aging cells. Our data highlight an intimate connection between aging, growth rate, and stress. Stress-independent genes that change with age are highly enriched for targets of the signal recognition particle (SRP). Combining MAD with an improved ATAC-seq method, we find that increasing proteasome activity reduces rDNA instability usually observed in aging cells and, contrary to published findings, provide evidence that global nucleosome occupancy does not change significantly with age.

**Kathryn Miller-Jensen**, Yale University  
*Single-cell analysis of macrophage heterogeneity*

Macrophages are innate immune cells that contribute to fighting infections, regulating development and tissue repair, and maintaining tissue homeostasis. To enable such functional diversity, individual macrophages must mount precise and coordinated responses to environmental cues, yet macrophages display extensive cell-to-cell heterogeneity. Using a combination of population and single-cell measurements, we are exploring how macrophage heterogeneity is regulated in response to inflammatory stimuli alone and in combination with other environmental cues. In one study, we discovered that the anti-inflammatory cytokine IL-10, secreted by a subset of macrophages even at low LPS doses, acts to enforce a threshold for macrophage pro-inflammatory activation that strongly contributes to intercellular heterogeneity. In another study of macrophages co-stimulated with LPS+IFN $\gamma$  and the resolving cytokine IL-4, we found that variable negative cross-regulation between a subset of LPS+IFN- $\gamma$ - and IL-4-specific gene programs resulted in significant heterogeneity, such that some co-stimulated macrophages were skewed towards one of the two transcriptional programs and exhibited specialized functions even when faced with multiple opposing stimuli.

Overall, we find that cell-to-cell heterogeneity provides strategies for macrophages to integrate signals and tune functional responses in complex tissue microenvironments.

## The dynamics of different FOXO isoforms in response to cellular stress

Allison Moreno Samayoa<sup>1</sup>, Lisa Shanks<sup>1</sup>, and Andrew Paek<sup>1</sup>

<sup>1</sup>Department of Molecular and Cellular Biology, UA, 1007 E. Lowell St., Tucson, AZ 85721

The Forkhead Box O (FOXO) transcription factor network is implicated in slowing down aging and extending lifespan in a diverse set of organisms. The FOXO family consists of four isoforms, Foxo1, Foxo3a, Foxo4, and Foxo6. Each isoform responds to multiple upstream stress signals such as nutrient or growth factor starvation, reactive oxygen species (ROS), and DNA damage. When activated, FOXO's translocate to the nucleus to activate downstream transcription to coordinate distinct cell fates. *How FOXO's can interpret a variety of stimuli and differentiate between different outcomes remains unknown. Further, how this is coordinated between the FOXO isoforms is not well understood.* We investigated whether differences in FOXO localization to and from the nucleus, or their dynamics, differ based on stress and differentiate cell outcomes. We performed Immunofluorescence (IF) on breast cancer cells to measure the nuclear to cytoplasmic ratio of each FOXO isoform in individual cells in response to different stresses. To measure dynamics, we used CRISPR-Cas9 gene editing to endogenously tag Foxo1, Foxo3a, and Foxo4 in each cell line. We measured nuclear to cytoplasmic shuttling dynamics of each FOXO isoform in single cells over time in response to different upstream stresses. Taken together, we identified different dynamic patterns of FOXO based on stress and identified correlations in activation between isoforms in response to stress which has not previously been reported.

## **Non-synonymous to synonymous substitutions suggest that orthologs tend to keep their functions better than paralogs**

Juan Miguel Escorcia-Rodríguez<sup>2</sup>, Mario Esposito<sup>1</sup>, Julio Augusto Freyre-González<sup>2</sup>, Gabriel Moreno-Hagelsieb<sup>1,\*</sup>

<sup>1</sup> Dept of Biology, Wilfrid Laurier University, Waterloo, Ontario, Canada N2L 3C5

<sup>2</sup> Regulatory Systems Biology Research Group, Laboratory of Systems and Synthetic Biology, Center for Genomic Sciences, Universidad Nacional Autónoma de México (UNAM), Av. Universidad s/n, Col. Chamilpa, Cuernavaca, Morelos. Mexico 62210

\* Corresponding author: gmorenohagelsieb@wlu.ca

Orthologs diverge after speciation events and paralogs after gene duplication. It is thus expected that orthologs would tend to keep their functions, while paralogs would tend to be a source of new functions. Because protein functional divergence follows from non-synonymous substitutions, we performed an analysis based on the ratio of non-synonymous to synonymous substitutions ( $dN/dS$ ) as proxy for functional divergence. We used four working definitions of orthology, including reciprocal best hits (RBH), among more accurate definitions based on network analyses and clustering. The results showed that orthologs by all definitions tested had lower values of  $dN/dS$  than paralogs, not only suggesting that orthologs keep their functions better, but also that paralogs are a readily source of functional novelty. The differences in  $dN/dS$  ratios remained favouring the functional stability of orthologs after deletion of genes with potential problems, such as having a high codon usage bias, low coverage of either of the aligned sequences, or sequences with very high similarities. While we also found that the ratio of  $dN/dS$  tends to increase with overall divergence, the tendency was not enough to explain the  $dN/dS$  differences between orthologs and paralogs.

## Multiplexed chemical control of signaling pathways in living cells by orthogonal self-localizing ligand (SL)-induced protein translocation (SLIPT).

Akinobu Nakamura<sup>1,6</sup>, Choji Oki<sup>2</sup>, Gembu Maryu<sup>3,7</sup>, Tatsuyuki Yoshii<sup>1,4</sup>, Michiyuki Matsuda<sup>3,5</sup>,  
Kazuhiro Aoki<sup>6,7,8</sup> and Shinya Tsukiji<sup>1,2,9,\*</sup>

<sup>1</sup> Department of Life Science and Applied Chemistry, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

<sup>2</sup> Department of Nanopharmaceutical Sciences, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

<sup>3</sup> Laboratory of Bioimaging and Cell Signaling, Graduate School of Biostudies, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

<sup>4</sup> PRESTO, Japan Science and Technology (JST), 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan

<sup>5</sup> Department of Pathology and Biology of Diseases, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

<sup>6</sup> Exploratory Research Center on Life and Living Systems (ExCELLS), National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji-cho, Okazaki, Aichi 444-8787, Japan

<sup>7</sup> National Institute for Basic Biology, National Institutes of Natural Sciences, 5-1 Higashiyama, Myodaiji-cho, Okazaki, Aichi 444-8787, Japan

<sup>8</sup> Department of Basic Biology, Faculty of Life Science, SOKENDAI (The Graduate University for Advanced Studies), 5-1 Higashiyama, Myodaiji-cho, Okazaki, Aichi 444-8787, Japan

<sup>9</sup> Frontier Research Institute for Materials Science (FRIMS), Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

\*Corresponding author: [stsukiji@nitech.ac.jp](mailto:stsukiji@nitech.ac.jp)

Most cell behaviors are the outcome of processing information from multiple signals generated upon cell stimulation. Thus, a systematic understanding of cellular systems requires methods that allow the activation of more than one specific signaling molecule or pathway within a cell. A systematic understanding of cellular systems requires methods that allow the activation of more than one specific signaling molecule or pathway within a cell. However, the construction of tools suitable for such multiplexed signal control remains challenging. Here we present orthogonal chemogenetic systems that allow control of multiple signaling pathways in living mammalian cells based on self-localizing ligand (SL)-induced protein translocation (SLIPT). We have previously reported a SLIPT system by which proteins fused to *E. coli* dihydrofolate reductase (eDHFR) can be relocated from the cytoplasm to the inner-leaflet of the plasma membrane (PM) by a trimethoprim-based SL (mgcTMP)<sup>[1]</sup>. The mgcTMP-induced translocation of signal proteins could activate the target signaling pathway. In recently, we succeeded in developing a new SLIPT system which based on (engineered) SNAP-tag and SL for SNAP-tag (mgcBCP)<sup>[2]</sup>. These systems could rapidly induce translocation of eDHFR/SNAP-tag-fusion proteins relocate to the PM upon addition of the corresponding SL in a time scale of minutes in the same cell. In addition, by combining the orthogonal SLIPT systems with fluorescent reporters, we achieved simultaneous multiplexed activation and fluorescence imaging of endogenous ERK and Akt activities in a single cell. Here, we will present the details of the molecular design of SLs (mgcTMP and mgcBCP) and the multiplexed signaling control by the eDHFR/SNAP-tag-based SLIPT.

[1] Ishida M. *et al.*, *J. Am. Chem. Soc.*, **2013**, *135*, 12684–12689.

[2] Nakamura A. *et al.*, *ChemRxiv*, **2019**, DOI: 10.26434/chemrxiv.8222456

## **Robustness and parameter geography in post-translational modification systems**

Kee-Myoung Nam<sup>1</sup>, Benjamin M. Gyori<sup>2</sup>, Silviana V. Amethyst<sup>3</sup>, Daniel J. Bates<sup>4</sup>,  
Jeremy Gunawardena<sup>1</sup>

<sup>1</sup> Department of Systems Biology, Harvard Medical School, Boston, MA

<sup>2</sup> Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

<sup>3</sup> Department of Mathematics, University of Wisconsin—Eau Claire, Eau Claire, WI

<sup>4</sup> Department of Mathematics, United States Naval Academy, Annapolis, MD

Biological systems are often said to be “robust to perturbations,” but a quantitative understanding of this has been elusive. In a mathematical model, perturbations usually exert their effects through the model's parameters, and so the robustness of a particular model behavior can be quantified through the size and shape of the parametric region in which that behavior arises. In this talk, I will discuss recent efforts to explore this “parameter geography” of bistability in post-translational modification (PTM) systems. Combining the “linear framework” for timescale separation and recent advances in numerical algebraic geometry, we were able to both describe the steady-state behavior of a two-site PTM system with arbitrary mechanistic complexity and survey this behavior at as many as  $\sim 10^9$  parameter points. In following this approach, we were able to make several interesting observations regarding the robustness of bistability, of which I shall focus on two. First, we found that bistability is much rarer for mechanisms that permit product rebinding than for those that forbid it, like the Michaelis–Menten mechanism, and we uncovered a previously unsuspected tradeoff between bistability and product rebinding that underlies this discrepancy. Second, we found that, while the volume of the bistable region appears to increase monotonically with the total amount of substrate, the region itself does not increase monotonically, as parameter points can move “back and forth” between monostability and bistability as substrate grows more abundant relative to enzyme. I will discuss these insights within the broader context of cultivating a mathematically rigorous understanding of a system's dependence on its parameters.

### **References**

Kee-Myoung Nam, Benjamin M. Gyori, Silviana V. Amethyst, Daniel J. Bates, Jeremy Gunawardena. Robustness and parameter geography in post-translational modification systems, submitted, 2019; bioRxiv:10.1101/862003.

## **Nuclear shuttling dynamics of the transcription factor TFEB following starvation reveal a regulatory feedback loop to fine tune its activity.**

Sara Napolitano<sup>1,2,\*</sup>, Iacopo Ruolo<sup>1,2,\*</sup>, Giuliano De Carluccio<sup>1,2</sup>, Diego di Bernardo<sup>1,2</sup>

<sup>1</sup> Department of Chemical, Materials and Production Engineering, University of Naples Federico II, Piazzale Tecchio 80, 80125 Naples, Italy.

<sup>2</sup> Telethon Institute of Genetics and Medicine, Via Campi Flegrei 34, 80078 Pozzuoli, Italy.

During amino-acid starvation the transcriptional activation of catabolic processes is induced by nuclear translocation and activation of Transcription Factor EB and E3 (TFEB/TFE3), master modulators of autophagy and lysosomal biogenesis. Autophagy is, in fact, one of the major pathways to recycle cellular constituents for energy production and protein building, thus allowing cell survival during nutrient deficiency.

In resting cells, under nutrient-rich conditions, TFEB is phosphorylated by the amino-acid sensitive mTORC1 complex, thus becoming inactive and cytosolic. Upon starvation, mTORC1 is inhibited and TFEB is dephosphorylated by the phosphatase Calcineurin and rapidly translocates to the nucleus activating the transcription of its target genes.

We first developed a dynamical model of TFEB nuclear shuttling to elucidate the biological mechanisms driving its regulation based on the literature. We considered a two-compartmental model (nucleus and cytoplasm) with two different species (phosphorylated and de-phosphorylated TFEB) for each compartment. Both de/phosphorylation and transport are modeled as first order kinetics and the input (the lack of nutrients) acts by changing the de/phosphorylation rates but not the transport rates.

We then generated a clonal HeLa cell population with constitutive expression of a TFEB-GFP fusion protein to follow TFEB nuclear translocation by fluorescence microscopy. We quantitatively measured nuclear shuttling dynamics in single cells in a microfluidics device that allows long-term culture under the microscope and alternating between complete medium and medium lacking amino-acids. We observed that upon starvation TFEB quickly enters the nucleus, this phase is then followed by a slower “overshoot” where TFEB partially exits the nucleus, despite the cell being still in starved conditions. Upon refeeding, TFEB quickly exits the nucleus. This behaviour suggests that TFEB levels in the nucleus during starvation are adjusted to just the right level via some sort of feedback mechanisms yet to be identified.

Since the model was not able to explain the presence of this overshoot, we modified it to add a negative feedback loop, which we hypothesise to be mediated by either a TFEB

direct target or by amino-acid accumulation following TFEB-mediated autophagy activation, which partly re-activate mTORC1.

We performed a series of microfluidics experiments by starving cells in the presence of small molecule inhibitors of different pathways and biological processes to distinguish between alternative hypotheses and demonstrated that the feedback loop is indeed caused by a TFEB direct target.

We are developing a fluorescence reporter of TFEB transcriptional activity to evaluate the impact of TFEB dynamics on gene expression. This is not a trivial as TFEB levels are low and cells reduce protein synthesis in starvation. To overcome these limitations, we designed a transducer system to convert the transcriptional activity of TFEB in an amplified translocation of a fluorescent protein from the membrane to the nucleus, through the enzymatic activity of an orthogonal protease, the Tobacco Etch Virus protease (TEVp).

Cellular response to starvation is an ubiquitous and evolutionary conserved feature which impacts ageing and neurodegenerative disorders, hence any new knowledge gained in this field may have long lasting impact on human health.

\* Equal contributions

## Accurate prediction of genetic circuit behavior requires modeling part expression level

Galen Dods<sup>1,\*</sup>, Mariana Gomez-Schiavon<sup>1,\*</sup>, Hana El-Samad<sup>1,2,3\*</sup>, Andrew H. Ng<sup>2,\*</sup>

<sup>1</sup>Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA

<sup>2</sup>Cell Design Initiative, University of California, San Francisco, San Francisco, CA

<sup>3</sup>Chan-Zuckerberg Biohub, San Francisco, CA

\*Equal Contributions

Tuning the input-output transfer function of genetic circuits is essential for tailoring their design to a given application. These transfer functions can be determined experimentally, but as circuits grow larger and more complex, it becomes impractical to screen all possible configurations. Mathematical models can be a powerful tool to predict the behavior of these circuits and may reduce the amount of experimental screening required. Here we investigate a genetic circuit consisting of a cascade of two inducible transcription factors to determine the optimal strategy for building a model using experimental results. Typically, models of individual circuit components are composed together to predict overall circuit behavior. We show that circuit models constructed in this manner fail to recapitulate overall circuit behavior when transcription factor models do not consider the effect of part expression level in addition to drug input. By fitting our part models to experimental data that includes the dimension of expression level, we are able to more accurately predict the transfer functions of our transcriptional cascades. Using our improved models, we computationally screened all possible circuit designs to explore the effects of expression level and ordering on the circuit dose response for both inputs. Our simulations revealed a variety of circuit behaviors that are suitable for different applications, and we validated several of these predictions experimentally. These results illustrate the importance of considering part expression level when designing and fitting models for genetic circuit design.

## A Hysteresis-inspired Approach to Investigate Intergenerational Stability of Heterochromatin Domains in the *S. pombe* Genome

Henry Ng<sup>1,2</sup> and Bassem Al-Sady<sup>2</sup>

<sup>1</sup>TETRAD Graduate Program, University of California, San Francisco

<sup>2</sup>Department of Microbiology and Immunology, George Williams Hooper Foundation, UCSF, San Francisco, CA, 94143, USA.

The partition of eukaryotic genome into euchromatin and heterochromatin domains plays a crucial role in gene regulation. The epigenetic behavior of heterochromatin structure, meaning the progeny's ability to inherit the repressive structure from previous generations even in the absence of nucleation signals, remains poorly studied. Heterochromatin epigenetic heritability is proposed to be fundamental in developmental programs, due to the need of temporal control of gene expressions at different developmental loci. However, the inheritance of some heterochromatin loci appeared to be unstable, and the precise property and cause of local inheritance stability remain poorly understood. Hence, to quantitatively measure the stability of heterochromatin inheritance in different loci, I measured the local hysteretic capacity using a fluorescent reporter system and a hormone dosage control system. Hysteresis is the dependence of the state of a system on its history, representing a system's stability and resistance to change. By monitoring the change of expression of a fluorescent protein gene integrated in a heterochromatin region while controlling the expression of the structural scaffold heterochromatin protein Swi6, I probed the hysteresis and the stability of heterochromatin loci, including wildtype and mutant versions of the mating type locus (*MAT*). We found that a deletion of *cenH*, a strong RNAi-dependent nucleator DNA sequence within *MAT* homologous to the centromeric *dg/dh* nucleator, increased the locus hysteretic capacity. On the contrary, a mutation in Atf1/Pcr1 binding site *REIII*, a weak RNAi-independent nucleator, decreased the hysteretic capacity. This approach was able to probe hysteresis and inheritance stability of different heterochromatin loci in our model system.

## **Integration of single cell DNA sequencing and mathematical modeling indicates transient genomic instability in TNBC**

Michael D. Nicholson<sup>1,2</sup>, Darlan Conterno Minussi<sup>3,4</sup>, Alexander Davis<sup>3,4</sup>, Thomas O. McDonald<sup>1</sup>, Nicholas Navin<sup>3,4,5</sup>, Franziska Michor<sup>1,2,6</sup>

<sup>1</sup>Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA.

<sup>2</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA.

<sup>3</sup>Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>4</sup>Graduate School of Biomedical Sciences, University of Texas MD Anderson Cancer Center, Houston, TX

<sup>5</sup>Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX

<sup>6</sup>Department of Stem Cell and Regenerative Biology, Harvard University, Boston, MA.

Many cancer genomes exhibit aneuploidy however the principles underlying the generation and maintenance of this pervasive phenomenon are still unclear. In this study we examine the accumulation of copy number alterations (CNAs) during tumor evolution. A novel acoustic-cell-tagmentation single cell DNA sequencing approach was developed and used to obtain copy number profiles of 4055 single cells from four TNBC breast tumors. Integrating this data with mathematical modeling and rigorous statistical inference, we demonstrate that the majority of observed CNAs are likely to be acquired during a period of transient instability in early tumorigenesis. We inferred that the per cell division rate of acquiring new breakpoints in the copy number profiles is increased by up to 2 orders of magnitude (fold change for individual patients; 115, 45, 44, 38) during this period of instability. Our study provides evidence of elevated genomic turbulence at the initial transition to cancer, which shapes the copy number landscape of the growing tumor.

## Revealing lineage-related signals in single-cell gene expression using random matrix theory

Mor Nitzan and Michael Brenner

John A. Paulson School of Engineering and Applied Sciences, Harvard University, 29 Oxford St, Cambridge, Massachusetts 02138, USA

The expression profiles of a population of single cells, generated for example by single cell RNA sequencing, contains rich information about the biological state of each cell, including cell type, location along the cell cycle, gene regulatory patterns and relative location in the tissue from which the cell was dissociated. A major challenge is to disentangle information about these different biological states from each other, including distinguishing them from cell lineage, since the correlation of expression patterns between different cells is necessarily contaminated by ancestry. Based on a recent advance in random matrix theory, we have shown that ancestral relationships between cells are expected to create a power law signature in the eigenvalue distribution of the covariance matrix of single-cell data. We demonstrate the existence of such signatures in single cell RNA sequencing data, and show that the genes driving them are indeed related to differentiation and developmental pathways. We illustrate these principles in diverse tissues and organisms, including the mammalian epidermis and lung, *Drosophila* whole-embryo, whole adult *Hydra*, yeast colonies, and cells undergoing iPSCs reprogramming. Based on these results, we can disentangle ancestry-related signals from other biological processes in single cell data. We illustrate the potential for such disentanglement for yeast single cell data, where we substantially improve the quality of reconstruction of gene regulatory networks by removing lineage-related effects.

**Yoshi Ohya**, University of Tokyo  
*From Genome to Morphology*

Morphology is the basic phenotypic characteristic that can be affected by genetic and environmental perturbations. Consequently, living organisms have their own morphologies which evolve through natural selection. In addition to the individual morphology, morphology can also be defined in cell level. While cell morphology reflects intra-cellular communication of cells in multicellular organism, it relates with genotype and genetic network in unicellular organisms. Therefore, morphological phenotyping of unicellular organisms has been carried out to achieve a global understanding of the cell system as well as to answer to the specific questions in cell biology.

Microscopic image data provide a rich and high-dimensional representation of cell morphology, and therefore handcrafted morphological features are easily extracted from the digital images. The budding yeast *Saccharomyces cerevisiae*, a unicellular model organism, provides an invaluable system for dissecting complex cellular processes using high-resolution phenotyping. We developed image-processing system, CalMorph specialized for the budding yeast and have been using it for long time<sup>1)</sup>. Our unique approach is assignment of unimodal probability distributions to most of the morphological features of yeast cells. By considering statistical behaviors and population modality of the morphological features, the true value of the morphological measurements can be well

described with probability distribution. We propose that higher statistical specificity can be obtained by applying the simple distribution model to the morphological data. By using this system, a wealth of biologically meaningful information was provided. The analysis of morphological phenotypes of numerous sets of mutants is a powerful tool to extract genetically and biologically important information). In this symposium I address some of our findings to which high-dimensional morphometric analysis has contributed significantly.

# Engineering immune cell sensors and network with advanced split CAR system

Atsushi Okuma<sup>1,2,3,10</sup>, Jang Hwan Cho<sup>1,2,10</sup>, Katri Sofjan<sup>1,2</sup>, Seunghee Lee<sup>1,2</sup>,  
James J. Collins<sup>4,5,6,7,8,9</sup>, and Wilson W. Wong<sup>1,2</sup>

<sup>1</sup>Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.

<sup>2</sup>Biological Design Center, Boston University, Boston, MA 02215, USA.

<sup>3</sup>Current affiliation; Center for Exploratory Research, Hitachi, Ltd., Kobe, Japan

<sup>4</sup>Synthetic Biology Center, MIT, Cambridge, MA 02139, USA

<sup>5</sup>Institute for Medical Engineering and Science, MIT, Cambridge, MA 02139, USA

<sup>6</sup>Department of Biological Engineering, MIT, Cambridge, MA 02139, USA

<sup>7</sup>Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA 02139, USA

<sup>8</sup>Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA

<sup>9</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, 3 Blackfan Circle, Boston, MA 02115, USA

<sup>10</sup>These authors contributed equally

The immune system is a sophisticated network of diverse immune cell types. Each cell type has mechanisms, in which multi-input is integrated and a set of outputs are executed, to distinguish “nonself from self”, “malignants from normal cells”, or “inflammation from steady-state”. Some of immune cells interact with each other beyond the cell types and collaboratively perform various functions such as redundancies, feedbacks, and 2-step verifications. In most cases, possible adverse events are solved by those immune cell consortia rather than a single immune cell. Engineering an immune cell sensor has been achieved with chimeric antigen receptors (CARs) which can redirect T cells to kill cancer cells. However, a single conventional CAR cannot control T cell reaction computationally nor interaction with other immune cells. We leverage our split, universal, and programmable (SUPRA) CAR system to expand the functionality and programmability of CAR T cells. We develop an inhibitory feature into the SUPRA CAR system, achieving an A AND NOT B logic and a three-input (A AND B) AND NOT C logic in a cell. We also demonstrate that our programmable system is functional in 7 different cell types including innate immune cell types. We engineer synthetic intercellular communications between different cell types; especially multi-cellular A AND NOT (B AND C) by redirected regulatory T cells, multi-cellular kill switch by NK cells, and a novel cell-cell communication channel based on a synthetic adaptor secretion system. Our work highlights that a simple split CAR design can generate diverse and complex phenotypes, including advanced logic, and provide a foundation of engineering an immune cell consortium with user-defined functionalities.

**Banu Ozkan**, Arizona State University  
*Protein Evolution and Monod's Second Secret of Life*

Proteins are the most efficient nano-machines and perform a broad range of functions. All of the information necessary for function is encoded in 1-D sequences. Proteins exquisitely translate this code to fold and function, yet deciphering this encoded information remains an open challenge. Inferring evolutionary record of extant proteins offers a tractable and highly effective solution to better understand the relation between sequence and protein function in order to decipher the 1-D sequence code. This is because evolution in itself has been a single massive ongoing experiment in diversification and optimization of protein sequence-structure-function relation occurring over billions of years. We have developed a physics-based metric called the Dynamic Flexibility Index (DFI) to study protein evolution. DFI quantifies the resilience of a given position to the perturbations occurring at various parts of a protein using linear response theory, mimicking the multidimensional response when the protein's conformational space is probed upon interaction with small molecules or other cellular constituents. DFI provides us with an opportunity to retrace evolutionary steps which, in turn, have led to structural dynamics analysis of resurrected ancestral proteins. We demonstrated that protein static structures do not need to be modified in order for new function or molecular adaptation to emerge. Proteins may evolve and adapt new function by fine tuning their native state conformational dynamics. These studies provide us a molecular mechanism: Nature utilizes minimum perturbation-maximum response as a principle through the allosteric alteration of the dynamics of the active/catalytic sites by mutating distal positions, rather than introducing mutations on active sites. We also showed allosteric regulation has been utilized in cell signaling to adapt/evolve its signaling pathways in response to external stimuli warranting exploratory (high information/low signal) or exploitative (low information/high signal) strategies.

## **The dynamics of FOXO transcription factor shuttling**

Andrew Paek, apaek@email.arizona.edu

Assistant Professor, Molecular and Cellular Biology, University of Arizona

The FOXO transcription factors (TFs) are critical regulators of cellular homeostasis that have an evolutionarily conserved role in aging as increased FOXO activity can slow the aging process and increase the lifespan of worms, flies and mice. FOXOs are activated in response to a range of stress signals including nutrient/growth factor depletion, oxidative stress and DNA damage. These stresses activate FOXO activity through a network of post-translational modifications that control FOXO nuclear-cytoplasmic shuttling. In the nucleus, FOXO TFs upregulate hundreds of genes in diverse and often conflicting cellular processes. For example, during quiescence, Foxo3a upregulates ROS scavenger and autophagy genes required for long-term viability of stem cell populations. In contrast, Foxo3A activity following DNA damage or oncogene inhibition can lead to apoptosis. To determine how FOXO TFs dictate different cell fates we used time-lapse microscopy to measure the dynamics FOXO nuclear/cytoplasmic shuttling in single mammalian cells. We have found broad similarities between the dynamics of FOXO shuttling and other multiple-input multiple-output TFs like Msn2 and p53. Specifically, we have found different stresses lead to different patterns of FOXO shuttling dynamics. For example, serum starvation leads to rapid pulses of FOXO in and out of the nucleus lasting 30 minutes to one hour, while Hydrogen peroxide causes sustained nuclear FOXO over several hours. In addition, we see correlations between dynamic patterns and different cell fates. Taken together our work suggests that the FOXO network uses different dynamic patterns to upregulate the proper target genes and ensure that the outcome of FOXO activation is commensurate with the upstream stress.

# **Pakman: a modular and efficient software tool for approximate Bayesian inference**

Thomas Pak<sup>1</sup>, Ruth E. Baker<sup>1</sup>, and Joe M. Pitt-Francis<sup>2</sup>

<sup>1</sup> Mathematical Institute, University of Oxford, Oxford, UK

<sup>2</sup> Department of Computer Science, University of Oxford, Oxford, UK

The amount of data generated in the biological sciences has risen sharply due to the development of high-throughput experimental techniques. Bayesian statistics provides a framework for data-driven parameter estimation and model selection, even when dealing with complicated mathematical models. In particular, inferring model parameters from observed data is possible with approximate Bayesian computation (ABC) methods, even when the likelihood function is analytically or computationally intractable.

Certain ABC methods, such as the ABC rejection and sequential Monte Carlo (SMC) method, involve a simulation workload that can be parallelised across many computational nodes. A parallel implementation can thus dramatically reduce the time needed to run these methods. In this talk, we introduce Pakman, a tool for parallel ABC that is designed to be modular, efficient and portable. Pakman is modular at the systems-level, which means that any executable application can be used with Pakman in a “plug and play” manner. This modular framework allows researchers to use their existing software in ABC workloads, without the need to rewrite their code. Moreover, the use of the Message Passing Interface (MPI) standard for parallelisation means that Pakman can be built and used on virtually any parallel computing platform. Thus, Pakman enables researchers to leverage all the computational resources at their disposal to fit mathematical models to experimental data in a convenient and efficient manner.

## **Model reduction tools for phenomenological modeling of input controlled biological circuit**

Ayush Pandey<sup>1</sup> and Richard M. Murray<sup>1</sup>

<sup>1</sup> Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA, 91125

We present a Python based software package to automatically obtain phenomenological models of input controlled synthetic biological circuits that guide the design using the chemical reaction level descriptive models. From the parts and mechanism description of a synthetic biological circuit, it is easy to obtain a chemical reaction level model of the circuit under the assumptions of mass-action kinetics using various existing tools. However, using these models to guide design decisions during an experiment is difficult due to a large number of reaction rate parameters and species in the model. Hence, phenomenological models are often developed that describe the effective relationships among the circuit inputs, outputs, and only the key states and parameters. In this paper, we present an algorithm to obtain these phenomenological models in an automated manner using a Python package for circuits with inputs that control the desired outputs. This model reduction approach combines the common assumptions of time-scale separation, conservation laws, and species' abundance to obtain the reduced models that can be used for design of synthetic biological circuits. We consider an example of a simple gene expression circuit and another example of a layered genetic feedback control circuit to demonstrate the use of our model reduction procedure.

## **QuartataWeb: integrated chemical-protein-pathway mapping for polypharmacology and chemogenomics**

Fen Pei<sup>1,2‡</sup>, Hongchun Li<sup>1,3‡</sup>, D. Lansing Taylor<sup>1,2</sup>, Ivet Bahar<sup>1,2\*</sup>

<sup>1</sup> Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA, 15213, USA

<sup>2</sup> The University of Pittsburgh Drug Discovery Institute, Pittsburgh, PA 15213, USA

<sup>3</sup> Research Center for Computer-Aided Drug Discovery, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

QuartataWeb is a user-friendly server developed for polypharmacological and chemogenomics analyses. Users can easily obtain information on experimentally verified (known) and computationally predicted (new) interactions between 5,494 drugs and 2,807 human proteins in DrugBank, and between 315,514 chemicals and 9,457 human proteins in the STITCH database. In addition, QuartataWeb links targets to KEGG pathways and GO annotations, completing the bridge from drugs/chemicals to function via protein targets and cellular pathways. It allows users to query a series of chemicals, drug combinations, or multiple targets, to enable multi-drug, multi-target, multi-pathway analyses. QuartataWeb is expected to serve as a first filter toward designing more effective phenotypic screens and polypharmacological strategies for complex diseases.

QuartataWeb is freely accessible at <http://quartata.csb.pitt.edu>.

‡equal contribution

\*To whom correspondence should be addressed (bahar@pitt.edu).

## **DropSynth 2.0: high-fidelity multiplexed gene synthesis in emulsions**

Calin Plesa<sup>1\*</sup>, Angus M. Sidore<sup>2\*</sup>, Joyce A. Samson<sup>1</sup>, and Sriram Kosuri<sup>1,4\*</sup>

<sup>†</sup>Phil and Penny Knight Campus for Accelerating Scientific Impact, University of Oregon, Eugene, OR

<sup>1</sup>Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA

<sup>2</sup>Department of Chemical and Biomolecular Engineering, UCLA, Los Angeles, CA

<sup>4</sup>UCLA-DOE Institute for Genomics and Proteomics, Molecular Biology Institute, Quantitative and Computational Biology Institute, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, UCLA, Los Angeles, CA

Our ability to build and refine predictive models of biological systems and processes is often dependent on the scale at which we can build data sets of functionally characterized DNA-encoded hypotheses. Multiplexed assays now allow functional testing of large synthetic libraries of genetic elements, using sequencing as a readout. These libraries are currently limited by the designability, length, fidelity, and scale of the input DNA. We previously introduced DropSynth, a scalable, low-cost method to build thousands of defined gene-length constructs in a multiplexed manner. Briefly, this approach builds gene libraries by compartmentalizing and assembling barcoded microarray-derived oligos in vortexed emulsions. Here we systematically optimize and improve the DropSynth process. By optimizing the polymerase, enzyme choice, algorithmic oligo design, adding enzymatic error correction, and increasing barcoding scale, we demonstrate an increase in fidelity from 4% to over 20%, while simultaneously increasing the scale of a single assembly reaction from 384 to 1,536 genes at once. These improvements pave the way for larger and longer gene libraries to be synthesized, significantly increasing the amount of sequence-function relationships which can be tested using designed hypotheses.

\*Equal contribution

## **A predictive model of gene expression reveals the role of regulatory motifs in the mating response of yeast**

Amy E. Pomeroy<sup>1,4</sup>, Matthew I. Pena<sup>2</sup>, John R. Houser<sup>2</sup>, Gauri Dixit<sup>3</sup>, Henrik G. Dohlman<sup>3</sup>, Beverly Errede<sup>2</sup>, and Timothy C. Elston<sup>1,3,4</sup>

<sup>1</sup> Curriculum in Bioinformatics and Computational Biology, UNC, Chapel Hill, NC 27599

<sup>2</sup> Department of Pharmacology, UNC, Chapel Hill, NC

<sup>3</sup> Department of Biochemistry and Biophysics, UNC, Chapel Hill, NC

<sup>4</sup> Computational Medicine Program, UNC, Chapel Hill, NC

Cells use signaling pathways to receive and process information about their environment. These systems are nonlinear, relying on feedback and feedforward regulation to respond appropriately to changing environmental conditions. Mathematical models developed to describe signaling pathways often fail to show predictive power, because the models are not trained on data that probe the diverse time scales on which feedforward and feedback regulation operate. We addressed this limitation using microfluidics to expose cells to a broad range of dynamic environmental conditions. In particular, we focus on the well-characterized mating response pathway of *S. cerevisiae* (yeast). This pathway is activated by mating pheromone and initiates the transcriptional changes required for mating. Although much is known about the molecular components of the mating response pathway, less is known about how these components function as a dynamical system. Our experimental data revealed that pheromone-induced transcription persists following removal of pheromone and that long-term adaptation of the transcriptional response occurs when pheromone exposure is sustained. We developed a model of the regulatory network that captured both persistence and long-term adaptation of the mating response. We fit this model to experimental data using an evolutionary algorithm and used those fits to predict scenarios the model was not trained on including different temporal stimulus profiles and genetic perturbations to pathway components. From this model we were able to gain new insights into how four regulatory motifs coordinate to control of the pathway response to both persistent and dynamic stimulation.

## **Functional Plasticity and Evolutionary Adaptation of Allosteric Regulation**

Megan Leander<sup>1</sup>, Yuchen Yuan<sup>3</sup>, Anthony Meger<sup>1</sup>, Qiang Cui<sup>3</sup> and Srivatsan Raman<sup>1,2,\*</sup>

<sup>1</sup> Department of Biochemistry, University of Wisconsin-Madison, Madison, WI – 53706,

<sup>2</sup> Department of Bacteriology, University of Wisconsin-Madison, Madison, WI – 53706

<sup>3</sup> Department of Chemistry, Boston University, Boston, MA – 02215, United States

Allostery is a fundamental regulatory mechanism of protein function. Despite notable advances, understanding the molecular determinants of allostery remains an elusive goal. Our current knowledge of allostery is principally shaped by a structure-centric view which makes it difficult to understand the decentralized character of allostery. We present a function-centric approach to elucidate the molecular basis and underlying functional landscape of allostery. We show that allosteric signaling exhibits a high-degree of functional plasticity and redundancy through myriad mutational pathways. Residues critical for allosteric signaling are surprisingly poorly conserved while those required for structural integrity are highly conserved, suggesting evolutionary pressure to preserve fold over function. Our results contradict the common view of a finely-tuned allosteric residue network maintained under selection, instead suggesting multiple solutions to the thermodynamic conditions of cooperativity.

## **Layered Feedback Improves Robust Functionality across Heterogeneous Cell Populations**

Xinying Ren<sup>1</sup>, Richard M. Murray<sup>1,2</sup>

<sup>1</sup> Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125

<sup>2</sup> Department of Bioengineering, California Institute of Technology, Pasadena, CA 91125

Realizing homeostatic control of metabolites or proteins is one of the key goals of synthetic circuits. Feedback control is often used so individual cells can sense and adapt to environmental disturbances. However, cell-cell heterogeneity is common in bacteria and it may break homeostasis at population level with only internal feedback control. For example, cells don't contribute equally to production or regulation in bet hedging or if mutation occurs. Quorum sensing (QS) serves as a collective mechanism by releasing and sensing small and diffusible signaling molecules for group decision-making. To overcome the disturbance to population level homeostasis caused by heterogeneity, we propose a layered feedback control structure. It includes a global controller using quorum sensing and a local controller via internal signal-receptor system. We demonstrate with modeling and simulation that the global controller drives contributing cells to compensate for disturbance while the local controller governs non-contributing cells. The layered controller can tolerate higher portion of non-contributing cells or longer generation of mutant cells within small expression error, compared with only internal feedback control. We further discuss the potential of such layered structures in controlling cell population size, population fraction and other population-dependent functions.

**Kim Reynolds**, University of Texas Southwestern

*Mapping and modeling functional constraints on an essential metabolic enzyme*

Understanding how molecular scale variation in metabolic enzyme activity relates to changes in cellular growth rate presents a well-defined instantiation of the genotype-phenotype mapping problem. Practically, a deep understanding of the relationship between enzyme velocity and growth rate is critical to engineer metabolic pathways with novel functions, relate mutations to disease progression, and discover new therapeutic opportunities. However, current efforts at relating enzyme activity to growth are hindered by three complexities: 1) the mapping between gene activity and phenotype is non-linear and generally unknown, 2) this mapping is shaped by epistasis between genes, and 3) this mapping is influenced by environmental conditions. Here, we use a combination of comparative genomics, deep mutational scanning, and laboratory evolution to understand how perturbations to the essential metabolic enzyme Dihydrofolate Reductase (DHFR) influence growth rate, and are adaptively compensated. We find that reducing the activity of a single other folate metabolic enzyme - Thymidylate synthase (TYMS) – is sufficient to compensate for loss of activity in DHFR. The epistatic relationship between these two enzymes can be described by a relatively simple phenomenological model, suggesting a possible path forward for understanding dependencies in metabolism more generally.

## **A dynamic variant effect map for human methylenetetrahydrofolate reductase across genetic and environmental backgrounds**

Jochen Weile<sup>1,2,3,4</sup>, Nishka Kishore<sup>1,2,3</sup>, Song Sun<sup>1,2,3,4,5</sup>, Marinella Gebbia<sup>1,2,3</sup>, Marta Verby<sup>1,2,3</sup>, Roujia Li<sup>1,2,3,4</sup>, Atina Cote<sup>1,2,3</sup>, Yingzhou Wu<sup>1,2,3,4</sup>, Rima Rozen<sup>6</sup>, David Watkins<sup>6</sup>, Sean Froese<sup>7</sup>, Wyatt Yue<sup>8</sup>, Gustavo Bezerra<sup>8</sup>, Frederick P. Roth<sup>1,2,3,4</sup>

<sup>1</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; <sup>2</sup>The Donnelly Centre, University of Toronto, Toronto, ON, Canada; <sup>3</sup>Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; <sup>4</sup>Department of Computer Science, University of Toronto, Toronto, ON, Canada; <sup>5</sup>Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; <sup>6</sup>Department of Human Genetics, McGill University, Montreal, Canada; <sup>7</sup>University Children's Hospital, Zurich, Switzerland; <sup>8</sup>Structural Genomics Consortium, Nuffield Department of Medicine, University of Oxford, Oxford, UK.

It has recently become feasible to experimentally test the functionality of all possible missense variants of particular target genes. However, the landscape of variant functionality may depend on changes in genetic background or environment. Methylenetetrahydrofolate reductase (MTHFR) is a key component of the human one-carbon metabolic pathway, with some variants causing homocysteinuria. Clinical interpretation of MTHFR missense variants is challenging, and most are currently classified as variants of uncertain significance (VUS). One variant, A222V, is both common (global allele frequency ~30%) and known to have diminished enzymatic activity. Although this variant is associated with neural tube defects, the association seems to depend on low dietary folate, while other disease associations for A222V are controversial. We systematically mapped MTHFR missense variant effects by testing the ability of each variant to functionally complement loss of the *S. cerevisiae met13Δ* deletion. Mapping was carried out in wild-type and A222V genetic backgrounds, and at four concentrations of folinic acid (which controls the level of the MTHFR substrate) for each background. The resulting variant effect scores correlated with pathogenicity annotations, and with the age of onset of symptoms associated with classic homocystinuria. The environment- and genetic background-dependent landscape of variant effects for MTHFR showed that some missense effects varied substantially as a function of A222V status and folinic acid levels. Strikingly, many variants were damaging only in the presence of the A222V variant, especially at low folinic acid levels. The resulting maps promises to be a valuable resource in clinical variant interpretation.

## **A new algorithm for detecting and quantifying alternative RNA structures reveals HIV-1 conformational heterogeneity and splicing regulation.**

Phillip Tomezsko<sup>1,2\*</sup>, Vincent Corbin<sup>1\*</sup>, Paromita Gupta<sup>1\*</sup>, Harish Swaminathan<sup>1</sup>, Tammy Lan and Silvi Rouskin<sup>1</sup>

<sup>1</sup>Whitehead Institute, Rouskin lab, Cambridge, MA

<sup>2</sup>Harvard University, Department of Virology, Cambridge, MA

RNA is unique among all biomolecules as it can be both information-storing and enzymatic. These features are tightly linked to its structure, in which base pairing interactions give rise to a highly folded macromolecule. Decades of research in model organisms have shown that specific RNA structures are critical for gene expression, and that altered RNA structures are sufficient to cause a variety of pathologies. Recent genome-wide association and biochemical studies have linked multiple diseases, including cancer and neurodegenerative diseases, to changes in RNA structure. Despite the clear importance of RNA structure in regulating gene expression, the catalog of known RNA structures is limited. Moreover, our understanding of the functions of RNA structure is derived largely from a handful of cases that have been examined through years of painstaking molecular biology and *in vitro* biochemistry. Although high throughput RNA structure probing methods were recently developed, their application is limited to detection of structures from a population average, which can be grossly misleading. We have developed and applied quantitative approaches from statistics and computer science in order to radically transform our understanding of RNA structure *in vivo*, not as a population average, but as a heterogeneous ensemble of RNA structures with different functional potential.

We used dimethyl sulfate probing with sequencing (DMS-MaPseq) to probe the structure of RNAs in cells, and developed an algorithm called Detection of RNA folding Ensembles using Expectation-Maximization (DREEM), which identifies and quantifies alternative conformations assumed by the same RNA sequence. We applied this approach to study the genome-wide structure of human immunodeficiency virus-1 (HIV-1). Contrary to previous models, which analyzed population averages, our results reveal the widespread heterogeneous nature of HIV-1 RNA structure. Importantly, we discovered alternative conformations at critical splice sites, which influences the ratio of transcript isoforms. Our simultaneous measurement of splicing and intracellular RNA structure provides evidence for the long-standing hypothesis that RNA folding regulates splice site usage, and that there is a major role for RNA conformational heterogeneity in regulating viral gene expression. The methods we developed open the door to studying RNA structure ensembles at single nucleotide resolution in living cells. We plan to apply the DREEM approach has wide-range of applications, and we are also applying it to elucidate to the role of RNA structure in human alternative splicing, where as little as 2-fold changes in splice site usage are associated with multiple diseases.

\*Equal contributions

## Modelling of the *atoSC* two-component system to rationally design an engineered bacterial biosensor

Jack Rutter<sup>1</sup>, Emma Donovan<sup>2</sup>, Geraint Thomas<sup>1</sup> and Chris Barnes<sup>1,3</sup>

<sup>1</sup> Department of Cell and Developmental Biology, UCL, Gower Street, WC1E 6BT, London, UK.

<sup>2</sup> Department of Mathematical and Physical Sciences, UCL, Gower Street, WC1E 6BT, London, UK.

<sup>3</sup> Department of Genetics, Evolution and Environment, UCL, Gower Street, WC1E 6BT, London, UK.

Two-component systems (TCSs) are a common form of sensing mechanism, that can respond to wide a range of input signals. The *atoSC* TCS detects and responds to levels of acetoacetate within the environment, regulating genes involved in short-chain fatty acid metabolism, acetoacetate metabolism, and cell motility. Acetoacetate itself is a ketone body, which can serve as an alternative energy source during periods of low glucose and accumulation of which can lead to a condition known as ketoacidosis after periods of prolonged starvation or excessive alcohol consumption. Here we report the experimental characterisation of two forms of plasmid-based acetoacetate inducible whole-cell biosensor, within knockout strains of the *atoSC* TCS: (1) pAtoL-amp, which contains the GFP expression under control of the *pato* promoter and relies on genomic expression of *atoSC*, and (2) pAtoL-AtoSC, which contains *pato* promoter controlled GFP and plasmid-based expression of the *atoSC* genes. We also developed an ODE model to describe the mechanisms underpinning these biosensors, with the goal of using this model to guide rational design strategies for improving final biosensor behaviour. Since this model contains a large number of unknown parameters, we developed a sampling scheme for sensitivity analysis. Our preliminary results suggest that the levels of *atoS* and *atoC* expression have a large impact on a range of final biosensor characteristics, including the sensitivity, uninduced expression (leakiness) and dynamic range. We will use our model to design improved biosensor behaviour, through tailoring the levels of *pato*, *atoS* and *atoC* expressed within the signaling system.

## Abstract

### Decoding cellular signals during embryonic transitions - from 2D to 3D

Borzo Gharibi and [Silvia Santos](#)  
The Francis Crick Institute, London, UK

Naïve pluripotency is a transient state during mammalian development that can be recapitulated indefinitely *in vitro* by inhibition of the mitogen-activated protein kinase (MAPK/Erk) signalling and activation of STAT and Wnt pathways. How Erk is inhibited *in vivo* to promote naïve pluripotency remains largely unknown. By combining live cell imaging of 3D models of mouse development, quantitative proteomics and mathematical modelling we found that FGF2, a known Erk activator, induces instead long-term Erk inhibition in both ES cells and mouse embryos. We show that Erk inhibition is mediated by an incoherent feedforward loop. Importantly, we found that FGF2 induces up-regulation of specific naïve pluripotency factors and down-regulation of DNA methylation by suppression of *de novo* DNA methylases and thereby supports naïve pluripotency *in vivo* and *in vivo*. We propose an *in vivo* context-specific role for FGF2 in regulating Erk dynamics and maintaining naïve pluripotency and suggest that feedforward regulation might play a key role in regulating the naïve-to-prime transition during early mammalian development.

# Logarithmic sensing is optimal under broad assumptions

Anish A. Sarma<sup>1</sup> and John C. Doyle<sup>2</sup>

<sup>1</sup>*Computation and Neural Systems*, <sup>2</sup>*Control and Dynamical Systems*,  
California Institute of Technology, Pasadena, CA

Logarithmic sensing has been observed in biology across organisms and implementations. Examples include chemotaxis [1], inflammation [2], developmental patterning [3], and vision [4]. In each of these systems, the sensitivity to change in a quantity scales with the baseline magnitude of the quantity: the change between magnitudes of 10 and 20 elicits a response more like the change between 100 and 200 (both  $\times 2$  changes) than the change between 100 and 110 (both  $+10$  changes).

The breadth of convergently evolved logarithmic sensing systems suggests fundamental advantages. In this theoretical work, taking bacterial chemotaxis as the main case study, we draw on results in information-limited control [5–7] to show that logarithmic sensing is optimal under three broad, biologically plausible assumptions:

- (1) Limits on the exactness of actions (e.g., inertia, stochasticity, discretization, all-or-nothing decisions).
- (2) Costs on fast or accurate acquisition of sensory information.
- (3) Tasks that reward graded response (informally, both too much and too little action are possible).

Gradient sensing in bacterial chemotaxis can be shown to approximate an optimal solution to a navigation problem satisfying these assumptions. The intuition can be developed with a simple example: suppose you want to travel from California to a conference in Hawaii. If accurate estimates of positional information are difficult to obtain, it might initially suffice to travel to the correct island; however, once on the island, more accurate positional information will be necessary. Tolerance to error scales with the distance left to travel.

The mechanisms of logarithmic sensing in chemotaxis can be interpreted as consequences of this scaling. Allosteric feedback to a dense sensor array has speed and energy advantages over counterfactual feedforward or parallelized implementations. Several other well-studied cases of logarithmic sensing can also be expressed, somewhat more abstractly, as information-limited control problems satisfying the key assumptions, motivating new interpretations of widely varying mechanistic implementations in immunology, development, and neuroscience.

- [1] Milena D. Lazova, Tanvir Ahmed, Domenico Bellomo, Roman Stocker, and Thomas S. Shimizu. Response rescaling in bacterial chemotaxis. *PNAS*, 108(33):13870–13875, Aug 2011.
- [2] Robin E.C. Lee, Sarah R. Walker, Kate Savery, David A. Frank, and Suzanne Gaudet. Fold change of nuclear  $\text{nf-}\kappa\text{b}$  determines  $\text{tnf}$ -induced transcription in single cells. *Molecular Cell*, 53(6):867–879, Mar 2014.
- [3] Lea Goentoro and Marc W. Kirschner. Evidence that fold-change, and not absolute level, of  $\beta$ -catenin dictates  $\text{wnt}$  signaling. *Molecular Cell*, 36(5):872–884, Dec 2009.
- [4] Gordon L. Fain, Hugh R. Matthews, M. Carter Cornwall, and Yiannis Koutalos. Adaptation in vertebrate photoreceptors. *Physiological Reviews*, 81(1):117–151, Jan 2001.
- [5] Anish A. Sarma and John C. Doyle. Flexibility and cost-dependence in quantized control. In *2019 American Control Conference (ACC)*. IEEE, Jul 2019.
- [6] Yorie Nakahira, Quanying Liu, Terrence J. Sejnowski, and John C. Doyle. Diversity-enabled sweet spots in layered architectures and speed-accuracy trade-offs in sensorimotor control. *arXiv*, 2019.
- [7] Nicola Elia and Sanjoy K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400, 2001.

## How Simple is the Cell? Signals, Switches and Gates that Control Reproduction

Frederique Ruf-Zamojski<sup>1</sup>, Michel Zamojski<sup>1</sup>, German Nudelman<sup>1</sup>, Yongchao Ge<sup>1</sup>, Natalia Mendeleev<sup>1</sup>, Gregory R Smith<sup>1</sup>, Xiang Zhou<sup>2</sup>, Chirine Toufaily<sup>2</sup>, Gauthier Schang<sup>2</sup>, Luisina Ongaro<sup>2</sup>, Hanqing Liu<sup>3</sup>, Rosa Gomez Castanon<sup>3</sup>, Mika Moriwaki<sup>6</sup>, Venugopalan Nair<sup>1</sup>, Hanna Pincas<sup>1</sup>, Joseph R. Nery<sup>3</sup>, Anna Bartlett<sup>3</sup>, Andrew Aldridge<sup>3</sup>, Nitish Seenarine<sup>1</sup>, Mary Anne Amper<sup>1</sup>, Nimisha Jain<sup>1</sup>, Angela K. Odle<sup>4</sup>, Gwen Childs<sup>4</sup>, Judith L. Turgeon<sup>5</sup>, Corrine Welt<sup>6</sup>, Joseph R. Ecker<sup>3</sup>, Daniel J. Bernard<sup>2</sup> and Stuart C. Sealfon<sup>1</sup>

<sup>1</sup>Icahn School of Medicine at Mount Sinai, New York, NY

<sup>2</sup>McGill University, Quebec, Canada

<sup>3</sup>Salk Institute, La Jolla, CA

<sup>4</sup>University of Arkansas for Medical Sciences, Little Rock, AR

<sup>5</sup>UC Davis, Davis, CA

<sup>6</sup>University of Utah, Salt Lake City, UT

The regulation of LH and FSH biosynthesis and release in the pituitary gonadotrope cell mediates the control of reproduction by the brain. The major regulator, GnRH, is released in brief pulses at a period extending to hours in humans. FSH gene induction responds to GnRH frequency domain signals whereas the LH gene is regulated only by average GnRH concentration (PMID: 28385888). An AND logic gate requiring activation of two MAP kinase pathways has been identified (PMID: 2146256) and high accuracy single cell assays has shown that gene responses to GnRH constitute all-or-none binary switches (PMID: 30357357). Here we identify a potential gate underlying the co-dependence of FSH expression and reproduction on both GnRH and activin signaling that has been found *in vivo* using mouse models. We used bulk and single cell gene expression and epigenetic assays in a gonadotrope cell line and in mouse pituitary. We identify a gene locus corresponding to a human polymorphism that alters reproduction that is regulated by activin and GnRH and that is accessible uniquely in gonadotropes *in vivo*. We propose this locus as a component of the molecular AND gate mediating the need for both GnRH and activin signaling for reproductive competence. As we generate more accurate single cell resolution datasets, Boolean switch and logic gate models of gene regulatory processes appear to be surprisingly faithful representations of the behaviors observed.

# Orthogonal tuning of gene expression noise using CRISPR-Cas

Fan Wu<sup>1+</sup>, Jiyoung Shim<sup>1+</sup>, Ting Gong<sup>1</sup>, and Cheemeng Tan<sup>1\*</sup>

<sup>1</sup>Department of Biomedical Engineering, University of California Davis, Davis, CA 95616

\*corresponding author ([cmtan@ucdavis.edu](mailto:cmtan@ucdavis.edu)), <sup>+</sup>Equal contribution.

The noise in gene expression of cells can cause cellular heterogeneity that leads to the failure of antibiotic- and chemo-therapy, as well as synthetic circuits. To study or control the heterogeneity, previous work has tuned gene expression noise by changing the rate of transcription initiation, mRNA degradation, and mRNA translation. However, these methods are invasive: they require changes to the target genetic components. Here, we propose an orthogonal system based on CRISPR-dCas9 to tune gene expression noise and mean independently in a single cell. Specifically, we modulate the protein expression of a reporter gene in *Escherichia coli* by incorporating CRISPR activation and repression (CRISPRar) simultaneously competing to the same binding site. The CRISPRar uses a single dCas9 that recognizes two different single guide RNAs (sgRNA). We build a library of sgRNA variants with different expression activation and repression strengths. We find that expression noise and mean of a reporter gene can be tuned independently by CRISPRar. Our results suggest that the expression noise is tuned by the competition between two sgRNAs that modulate the binding of the RNA polymerase to a promoter. Our work provides a systematic, quantitative foundation towards designing CRISPR-dCas9 for modulating gene expression noise. The CRISPRar may also change how we tune expression noise at the genomic level. Our work has broad impacts on the study of gene functions, phenotypical heterogeneity, and genetic circuit control.

# Single cell gene regulatory network inference of the *Drosophila* optic lobe using the Inferelator

Claudia Skok Gibbs<sup>1</sup>, Neset Ozel<sup>2</sup>, Claude Desplan<sup>2</sup>, Richard Bonneau<sup>1,2,3</sup>

<sup>1</sup>Flatiron Institute for Computational Biology, New York, New York, 10010, USA.

<sup>2</sup>Department of Biology, New York University.

<sup>3</sup>Center for Data Science, New York University.

## Abstract

Cells in multicellular organisms must control how they express genes based on developmental cues. Understanding developmental patterns is a central challenge in systems biology. Learning gene regulatory networks is necessary to understand mechanisms that drive development. Computational methods for inferring and modeling regulatory networks must allow for complexity at the systems level, while also adjusting for limitations in available data.

*Drosophila melanogaster* (the fruit fly) is an excellent model organism for studying neuronal development. We have acquired single-cell RNA sequencing experimental data from the optical lobe of *Drosophila* during larval development, which has a quarter of a million individual cell gene expression profiles. We have also acquired single-cell ATAC chromatin accessibility data from the same optical lobe samples. We are applying a computational method for gene regulatory network inference based on regularized linear regression (the Inferelator [1-4]) to this *Drosophila* data. We constrain the network model using chromatin accessibility in combination with transcription factor motifs and then learn the gene regulatory network using the single-cell gene expression data.

## References

1. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, et al. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* 2006;7: R36.
2. Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. DREAM3: Network Inference Using Dynamic Context Likelihood of Relatedness and the Inferelator. *PLoS One.* 2010;5: e9803.
3. Greenfield A, Hafemeister C, Bonneau R. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics.* 2013;29: 1060–1067.
4. Greenfield A, Madar A, Ostrer H, Bonneau R. DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. *PLoS One.* 2010;5: e13397.

## Sorting signal from noise in signal transduction networks

Whitney E. Heavner<sup>1</sup>, Stephen E.P. Smith<sup>1,2,3</sup>

<sup>1</sup> Center for Integrative Brain Research, Seattle Children's Research Institute; Seattle, WA

<sup>2</sup> Department of Pediatrics, University of Washington School of Medicine, Seattle WA.

<sup>3</sup> Interdisciplinary Program in Neuroscience, University of Washington, Seattle WA.

Signal transduction experiments are inherently noisy. But is this noise a technical artifact, or does it reflect a stochasticity in the underlying biology? During signal transduction events, individual proteins join or dissociate from larger complexes, which changes the function of the protein complexes, alters cellular homeostasis, and ultimately instructs the cellular response to the signaling input. However, due to extensive cross-talk among signal transduction pathways and multiple inputs into individual pathways, isolating an experimental response from uncontrolled biological noise can be difficult. Here, we present a combined experimental and statistical approach to isolate experimental signal transduction network activation from biological or technical noise, using quantitative multiplex co-immunoprecipitation followed by two independent statistical analyses: ANC (adaptive, non-parametric, corrected for multiple comparisons) and wCNA (weighted correlation network analysis). We analyzed the process of synaptic homeostatic scaling, a well-characterized response of cultured cortical neurons to large-scale changes in synaptic activity. We performed three separate QMI experiments, each consisting of an N of 4-5 biological replicates. QMI identified proteins in shared complexes (PiSCES) that change in response to stimulation. We noticed that while a core set of PiSCES were identified in every experiment, there were several PiSCES in each dataset that were unique to that set. To distinguish if these unique PiSCES were stimulus-associated changes that only rose above signal-to-noise in one of three experiments, or if they were uncontrolled biological phenomena, or if they were technical errors, we ran a weighted correlation network analysis of all 13 individual experimental Ns. We identified a large module of co-varying PiSCES that strongly correlated with synaptic scaling. However, we also identified a second module of PiSCES that co-varied with each other, implying a biological function, but that did not correlate with experimental treatment or any other measured variable. Several PiSCES in this module were significantly different between control and treatment groups in individual N-of-4 experiments, but did not show a consistent pattern across all experiments. Finally, we identified a small module consisting of PiSCES involving a single probe antibody that only correlated with one of three experiments, implying a technical error. Thus, our approach allowed us to both: 1) identify a set of PiSCES that changed consistently across multiple experiments, even though they were often below the limits of detection; and 2) identify a second set of 'noise' PiSCES that often appeared in individual N-of-4 datasets, but did not show consistent patterns across datasets. We will discuss the neurobiological significance of the individual protein networks identified. More broadly, we believe that the ability to distinguish between modules of interacting proteins that respond to a stimulus vs. modules of interacting proteins that fluctuate in an apparently random fashion yields an important insight into the experimental noise that is apparent in all signal transduction experiments. Our results demonstrate that some experimental noise reflects patterned biological fluctuations, not technical artifacts.

## **A manifold model of the human cell cycle after oncogenic KRAS induction identifies a potential decision point for irreversible arrest**

Wayne Stallaert<sup>1</sup>, Jamshaid Shahir<sup>1</sup>, Katarzyna M. Kedziora<sup>1</sup>, Juanita C. Limas<sup>1</sup>, and Jeremy E. Purvis<sup>1</sup>

<sup>1</sup> Department of Genetics, Computational Medicine Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.

The role of oncogenic KRAS in the initiation of tumorigenesis remains poorly understood due to its paradoxical effects on cell cycle fate: while some cells exhibit a hyperproliferative phenotype that drives tumor initiation and growth, other cells succumb to cell cycle arrest and eventual senescence. Determining how individual cells decide between these two fates is critical for understanding the initiation, progression and treatment of KRAS-driven cancers. By combining multiplexed, single cell imaging of >50 signaling and cell cycle effectors with a geometrically-inspired model of the “cell cycle manifold”, we demonstrate that acquisition of a KRAS-G12V mutation generates multiple cell cycle states within a single isogenic cell population. Manifold learning identified a bifurcation point in G1 that determines whether a given cell re-enters the cell cycle or yields to cell cycle arrest. From this “decision” point, we observe at least two divergent cell cycle trajectories corresponding to the hyperproliferative and arrested cellular states. We also found that growth factors in the cellular microenvironment protect against oncogenic KRAS-induced arrest through increased AKT signaling, indicating that both intrinsic and extrinsic factors collaborate to control this cell fate decision. Our model identifies key nodes in the KRAS signaling network that can be targeted to divert tumorigenic cells from a hyperproliferative to an arrested state, revealing potential therapeutic strategies for the treatment of KRAS-driven cancers.

## Quantifying the dose dependent landscape of DNA repair

Lyubov Kublo<sup>1</sup>, Shayla Goller<sup>1</sup>, Jacob Stewart-Ornstein<sup>1</sup>

<sup>1</sup>Dept. of Comp. and Sys. Biology & Hillman Cancer Center, University of Pittsburgh, Pittsburgh PA.

Many cancer cells are uniquely vulnerable to DNA damage compared to proliferative cells in our normal tissues and this differential sensitivity rests at the heart of the efficacy of chemotherapy. This defect in cancer cell survival when challenged with chemotherapy or radiation is due to ubiquitous defects in DNA damage signaling and repair in cancer; DNA damage pathways are disrupted in >80% of cancers. The genotype of a cancer therefore determines its sensitivity to specific DNA damaging agents and potentially to other aspects of treatment such as dose and duration. However, we currently lack a framework to connect cancer genotype to optimal DNA damaging therapeutic regimes. Here we use targeted genetic screens at multiple doses and durations of chemotherapy treatment to characterize the landscape of DNA damage sensitivity. We find that dose specific sensitivity clusters genes by function, that many genotypes lead to non-monotonic sensitivities to DNA damage, and that these patterns are reproducible across cell lines. For example, homologous recombination is most critical at low damage doses and apoptotic signaling at higher doses. Experiments comparing high dose low duration and long duration low dose treatment likewise identify duration specific pathways. Simple models of DNA damage repair and cell death are used to interpret these results and suggest optimal therapeutic strategies for cancers with defects in specific repair pathways based on dose and duration rather than treatment agent.

# Ligand-Receptor Promiscuity Enables Specific Targeting of Cell Types Through Combinatorial Addressing

Christina Su<sup>1</sup>, Yaron Antebi<sup>2</sup>, James Linton<sup>1</sup>, Akshay Yeluri<sup>1</sup>, Justin Bois<sup>1</sup>, Arvind Murugan<sup>3</sup>, and Michael Elowitz<sup>1,4</sup>

<sup>1</sup> Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA

<sup>2</sup> Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>3</sup> Department of Physics and James Franck Institute, University of Chicago, Chicago, IL 60637, USA

<sup>4</sup> Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

Communication systems enable information transfer from sender(s) to specifically designated receiver(s). In biological systems, core intercellular signaling pathways allow cells to communicate through secretion and sensing of signaling molecules, or ligands, across diverse developmental and physiological contexts. Many of these pathways, such as the bone morphogenetic protein (BMP) pathway, exhibit promiscuous, many-to-many interactions between multiple ligand and receptor variants. Indeed, multiple ligands are generally present in developmental contexts, and cells typically express multiple relevant receptors. These observations provoke the question of how promiscuous communication systems can specifically activate, or address, the appropriate cell type(s) with the spatial and temporal precision required for proper development. Previous work has demonstrated that this promiscuous ligand-receptor architecture enables cells to perform computations on ligand combinations. However, it has remained unclear whether and how such interactions allow addressing of signals to particular targets. Here, we reveal that promiscuous ligand-receptor interactions can, counterintuitively, enable different combinations of relatively few ligands to selectively activate a larger number of target cell types. Experimentally, analysis of cell lines differing in receptor expression illustrates the potential for such combinatorial addressing in cell culture. Mathematically, we develop a general modeling framework and demonstrate that combinations of as few as two ligands can enable flexible, independent addressing of many more individual cell types or groups of cell types. These results provide a foundation for understanding how the BMP pathway, and others that exhibit ligand-receptor promiscuity, can allow specificity at the level of cellular populations.

## **Model-driven experimental design identifies counter-acting feedback regulation driving MAPK dynamics**

Kimiko Suzuki<sup>1</sup>, Beverly E. Errede<sup>2</sup>, Henrik G. Dohlman<sup>3</sup>, Timothy C. Elston<sup>1,3,4</sup>

<sup>1</sup> Curriculum in Bioinformatics and Computational Biology, UNC Chapel Hill, NC 27514

<sup>2</sup> Department of Biochemistry and Biophysics, UNC Chapel Hill, NC

<sup>3</sup> Department of Pharmacology, UNC Chapel Hill, NC

<sup>4</sup> Computational Medicine Program, UNC Chapel Hill, NC

Cells rely on mitogen-activated protein kinases (MAPKs) to survive environmental stress. In yeast, activation of the MAPK Hog1 is well known to mediate the response to high osmotic conditions. Recent studies of Hog1 revealed that its temporal activity is subject to both negative and positive feedback regulation, yet the mechanisms of feedback are unknown. By designing mathematical models of increasing complexity for the Hog1 MAPK cascade, we identified pathway circuitry sufficient to capture Hog1 dynamics observed *in vivo*. We then used these models to optimize experimental designs for distinguishing potential feedback loops. Performing experiments based on these models revealed that fast positive feedback at the level of the MAPK is necessary for switch-like (dose-independent) activation, while a delayed negative feedback underlies the dose-dependent timing of MAPK deactivation. We validated this model by experimentally screening the candidate regulators implicated by the circuit. Our results suggest that negative feedback acts through the accumulation of osmolytes while positive feedback results from mutual inhibition of the MAPK and its phosphatases. We predict that the duration of phosphatase inactivation is dictated by the strength of the stimulus. Our findings reveal a new and important positive signaling function for MAPK phosphatases. More broadly, they demonstrate the value of methodically adding network motifs to mechanistic models to infer targets of feedback regulation in signaling pathways. We postulate that similar counter-acting feedback loops are common in MAPK signaling systems conserved throughout all eukaryotic organisms.

# How old are proteins? (You need to first know PPI parameters)

Jithin Sreedharan<sup>1,2</sup>, Krzysztof Turowski<sup>1</sup>, and Wojciech Szpankowski<sup>1,2</sup>

<sup>1</sup> Center for Science of Information, Purdue University, W. Lafayette, IN 47907.

<sup>2</sup> Dept. of Computer Science, Purdue University, 305 N. University St., W. Lafayette, IN 47907.

Complex systems in life sciences are comprised of interacting entities; e.g., cellular processes involve interacting genes, proteins, and other biomolecules; microbial signaling is at the core of complex macro environments such as the human gut; and brain connectomes are composed of neuronal networks. These systems are modeled as *dynamic networks*, evolving over time, with nodes representing entities and edges representing their interactions. Typical systems continually evolve to optimize various criteria, including function (e.g., the evolution of brain connectomes to specialize function), structure (e.g., the evolution of protein interaction networks to guard against point mutations in coding genes), and survivability (e.g., redundant pathways in genic interactions as evidenced by synthetic lethality screens). The problem of inferring the evolution of a dynamic network from a small number of snapshots is of considerable significance: In networks of biochemical interactions (e.g., protein interaction networks -- PPI) one can identify early biomolecules that are known to be differentially implicated in diseases. Our ultimate aim is to infer the evolution of a PPI network from a single snapshot of the network.

Here, however, we first have to deal with estimating parameters for fitting the PPI data into graph models since inferring the evolution of a network with rigorous statistical guarantees is intrinsically linked to the assumed model. Many of the existing parameter estimation techniques for the PPI networks overlook the critical property of *network symmetry* (also known formally as graph automorphisms), which is known to be significant for the PPI networks of many species, thus giving statistically insignificant results concerning the observed network. To demonstrate it and to develop accurate estimation procedures, we focus on the biologically inspired *duplication-divergence model*, and the up-to-date data of the PPI of seven species including human and yeast. Using exact recurrence relations of some prominent graph statistics, we devise a parameter estimation technique that provides the right order of symmetries and uses phylogenetically old proteins as the choice of seed graph nodes. We also find that our results are consistent with the ones obtained from maximum likelihood estimation (MLE). However, the MLE approach is significantly slower than our methods in practice.



## Quantitative analysis of GPCR Downstream Signaling by Multiplexed Fluorescence Imaging

Ryosuke Tany<sup>1,2,3</sup>, Yohei Kondo<sup>1,2,3</sup>, Yuhei Goto<sup>1,2,3</sup>, and Kazuhiro Aoki<sup>1,2,3</sup>

<sup>1</sup>Department of Basic Biology, School of Life Science, SOKENDAI, Aichi, Japan

<sup>2</sup>Division of Quantitative Biology, NIBB, NINS, Aichi, Japan

<sup>3</sup>Quantitative Biology Research Group, ExCELLS, NINS, Aichi, Japan

G protein-coupled receptors (GPCRs) are seven-transmembrane receptors recognizing a wide range of ligands like neurotransmitters, hormones, tastants, and odorants. Despite the fact that more than 800 GPCRs exist in the human genome, most GPCRs are known to couple only 4 types of heterotrimeric G proteins, Gs, Gi, Gq, and G12/13, to transmit extracellular stimuli. It is still largely unknown how the ligand information such as ligand types, dose, dynamics, and their combinations, is encoded by GPCR-mediated activation of four types of heterotrimeric G proteins. We hypothesized that the information of GPCR ligands are encoded by the dynamics of the heterotrimeric G protein activation. Here, we demonstrate the multiplexed fluorescence imaging system for the quantification of GPCR downstream signaling in living cells. This imaging system employs 4 fluorescence probes for as a proxy for heterotrimeric G protein activation; cAMP probe (EpacSensor), RhoA probe (DORA-RhoA), Ca<sup>2+</sup> probe (R-GECO), and ERK probe (ERK-KTR) were used. First, we established two types of stable HeLa cell lines expressing a pair of EpacSensor and R-GECO or DORA-RhoA and ERK-KTR. For time-lapse imaging, these two cell lines were co-cultured, followed by the introduction of a GPCR gene. Next, the cells were stimulated with its GPCR ligand, and time-lapse imaged to quantify the dynamics of cAMP, Ca<sup>2+</sup>, RhoA activity, and ERK activity. At present, we examined the dynamics of cAMP, Ca<sup>2+</sup>, RhoA, and ERK elicited by over 50 GPCRs by this imaging system. showing distinct and divergent dynamics of cAMP, Ca<sup>2+</sup>, RhoA, and ERK by GPCR activations. For instance, Gq-coupling GPCRs exhibit sustained or transient Ca<sup>2+</sup> increase in a ligand-dependent manner. In this presentation, we will introduce the recent progress in this study.

## **Temporal perturbations of stress reveal proline accumulation as an evolutionary conserved protective mechanism in the response to stress in human cells.**

Alexander Thiemicke and Gregor Neuert

*Department of Molecular Physiology and Biophysics, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA*

Cells are exposed to temporal changes of stresses in their environment. These temporal changes may cause differential regulation of cell signaling, metabolism, phenotype and cell fate. A common stressor in many human tissues is extracellular hypertonicity, critically involved in physiological and pathophysiological processes. We demonstrate that temporal osmolyte concentration gradients have a differentiating effect on signaling and cell fate. To investigate molecular mechanisms causing this phenotype, we adapted Fluorescent Cell Barcoding for flow cytometry to time course experiments to screen for dynamic changes in apoptosis/caspase signaling, stress signaling, growth/proliferation and DNA damage. This temporal screen reveals a differential regulation of several caspase and stress pathways between instant and slowly changing NaCl application. Other pathways (growth/inflammation and DNA damage) are mostly independent of the method of NaCl application and are coregulated with the cumulative exposure to hypertonic stress. Only when NaCl is applied rapidly, caspases are activated and trigger apoptosis by a caspase 8 and 9 independent pathway. Gradual application of NaCl instead is characterized by an accumulation of proline, a protective mechanism conserved in bacteria and plants.

# Dynamic modeling of gene regulation aided by *pseudoveloc*ity

Andreas Tjärnberg<sup>1,2</sup>, Lionel A. Christiaen<sup>1</sup> and Richard A. Bonneau<sup>2</sup>

<sup>1</sup>Center for Developmental Genetics at NYU Biology, New York University.

<sup>2</sup>Center for Genomics and Systems Biology, New York University.

## Abstract

Single-cell RNA sequencing methods have been applied to problems like cell type characterization of whole animals, embryonic developmental cell specification, and developmental trajectory inference. Techniques that analyze single-cell RNA sequencing expression data for dynamic modeling and gene regulatory network inference are still in development, largely due to the difficulties in observing dynamics of genes directly.

One computational method that has shown promise is the RNA velocity approach which uses the ratios between exons and non-coding introns, which are removed by pre-mRNA splicing, to infer which genes are actively being transcribed and at what rate. This approach has some experimental limitations though, and is therefore difficult to apply to genome wide velocity estimation.

We propose a new technique for inferring dynamic rates of change based on a concept of pseudovelocity. With the large number of cells in a single-cell data set, a local gradient of gene expression can be estimated for each cell based on its relationship to other nearby cells, using a k-nearest neighbor graph. The rate of change of every gene for each cell can then be determined relative to a pseudotime axis.

We have applied this pseudovelocity estimation to existing published data, to the dynamic modeling and interpretation of gene regulation that describes embryonic heart development in the model organism *Ciona intestinalis* (the sea squirt), and to the dynamic modeling of the cell cycle in the model organism *Saccharomyces cerevisiae* (the budding yeast).

# **A quantitative approach to benchmark ancestral sequence reconstruction methods**

Lily Tran<sup>1</sup>, Ryan N. Randall<sup>1</sup>, Eric A. Gaucher<sup>1</sup>

<sup>1</sup> Department of Biology, Georgia State University, Atlanta, GA 30303

Ancestral sequence reconstruction (ASR) is a still-burgeoning method that has revealed many key mechanisms of molecular evolution. One criticism of the approach is an inability to validate its algorithms within a biological context as opposed to a computer simulation. Here we build an experimental phylogeny using the gene of a single red fluorescent protein to address this criticism. The evolved phylogeny consists of 19 operational taxonomic units (leaves) and 17 ancestral bifurcations (nodes) that display a wide variety of fluorescent phenotypes. The 19 leaves then serve as ‘modern’ sequences that we subject to ASR analyses using various algorithms and to benchmark against the known ancestral genotypes and ancestral phenotypes. We confirm computer simulations that show all algorithms infer ancient sequences with high accuracy, yet we also reveal wide variation in the phenotypes encoded by incorrectly inferred sequences. Specifically, Bayesian methods incorporating rate variation significantly outperform the maximum parsimony criterion in phenotypic accuracy. Subsampling of extant sequences had minor effect on the inference of ancestral sequences.

## Live imaging of single gene loci reveals surge in transcription factor concentration at onset of transcriptional bursts

Meghan A. Turner<sup>1\*</sup>, Nicholas C. Lammers<sup>1\*</sup>, Emma Luu<sup>2</sup>, Simon Alamos<sup>3</sup>, Thomas Graham<sup>2</sup>, Mustafa Mir<sup>2</sup>, Michael R. Stadler<sup>2</sup>, Xavier Darzacq<sup>1,2</sup>, Michael B. Eisen<sup>1,2</sup>, and Hernan G. Garcia<sup>1,2,4</sup>

<sup>1</sup>Biophysics Graduate Group, University of California, Berkeley, Berkeley, CA

<sup>2</sup>Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA

<sup>3</sup>Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA

<sup>4</sup>Department of Physics, University of California, Berkeley, Berkeley, CA

Over the last few decades *in vitro* and *in situ* approaches have revealed the identity of the molecular players driving transcription in eukaryotes. Yet, these studies are virtually silent on the precise timing of the recruitment of each of these players to the promoter, and on how this recruitment determines output transcriptional dynamics *in vivo*. To fill in this gap, we have developed a new method for simultaneously measuring local input transcription factor concentration at target loci and the resulting output transcriptional activity of these loci in single living cells. Specifically, we studied how the Dorsal activator, a member of the NFκB family and a key transcription factor in the development of the fruit fly *Drosophila melanogaster*, is recruited to the promoter of its target gene *snail* in order to drive transcriptional bursting. We found that transient surges in Dorsal concentration coincide with, but do not precede, the onset of transcriptional bursts. Interestingly, these surges are not maintained throughout the duration of the bursts but exist for a fixed length of time no matter the duration of the associated burst. Instead, we discovered that the amplitude of the transient Dorsal concentration surge at the start of a transcriptional burst, and not surge duration, dictates transcriptional burst size. We speculate that Dorsal delivers a “package” of downstream transcription factors (e.g. P-TEFb) to the promoter that sustains the transcriptional burst until this package is exhausted. Thus, our method provides a tool to uncover the precise *in vivo* timing and ordering of the diverse molecular players that drive the transcriptional process.

\*Equal contributions

**Shyni Varghese, Duke University**

*Extracellular matrix: a potential therapeutic target for cancer and fibrosis*

Reciprocal interactions of cells with their microenvironment are fundamental to multiple cellular processes necessary for tissue development, homeostasis, and regeneration. It is becoming increasingly apparent that while the extracellular environment normally maintains tissue homeostasis, it may contribute to disease progression and age-dependent pathologies when negatively perturbed. In this talk, I will discuss our efforts to delineate the role of the extracellular matrix (ECM) on various cellular responses contributing to cancer metastasis and fibrosis. Our findings show that the cells transition from a proteolytic-independent mode of invasion to a proteolytic-dependent mode upon an increase in the mechanical resistance from the extracellular environment. By employing a cutaneous fibrosis model, we unraveled the role of elastic fibers and their components, which lie at the interface of tissue stiffness and inflammation, on fibrosis progression. Surprisingly, interfering with the ECM organization to alter the elastin content and tissue stiffness to levels comparable to normal skin diminished the inflammatory response and abrogated the fibrotic phenotype. I will end by briefly discussing our recent efforts in understanding the role of physicochemical cues of the ECM on lung fibrosis by using in vitro (biomechanically active lung-on-chip) and in vivo injury models.

## **Machine Vision based morphological profiling: Charting a time course for morphology changes**

Jagadish Venkataraman<sup>1</sup>, Ian Foe<sup>1</sup>, Scott McIsaac<sup>1</sup>

<sup>1</sup> Calico Life Sciences LLC, 1170 Veterans Blvd, South San Francisco, CA 94080

Over sixty years ago, Mortimer & Johnston discovered that “mother” cells in *Saccharomyces cerevesiae* (budding yeast) divide a finite number of times. Due to asymmetric cell division, the overwhelming majority of cells in a yeast culture is young. A recently developed technology, the Miniature-chemostat Aging Device (or MAD), allows us to purify millions of aged cells at a time.

By combining MAD with live cell imaging, we are beginning to uncover new age-associated biological phenotypes, such as an age-dependent expansion of the endoplasmic reticulum that doesn't depend on the Unfolded Protein Response (UPR). Additionally, proteins targeted to the same cellular compartment in young cells can have different localization patterns than in old cells. How do we quantify these age-dependent patterns? How do they vary between mutants? Can we reconstruct time courses from static images?

To address these questions, we used machine vision techniques to create an automated inference pipeline to segment out individual cells from a field of view, quantify the age of each segmented cell, map the localization pattern of each segmented cell to a computational representation, and finally, associate the representations with their respective replicative ages. This pipeline is highly scalable and allows for quickly and efficiently exploring the aging process at the single-cell level. In this poster, we present the architecture and methods underlying the yeast machine vision pipeline.

## Engineering of Predictable Degradation-tuning RNAs

Qi Zhang<sup>1</sup>, Duo Ma<sup>2</sup>, Fuqing Wu<sup>1</sup>, Kylie Standage-Beier<sup>1</sup>, Xingwen Chen<sup>1</sup>, Alexander A. Green<sup>2</sup>,  
Xiao Wang<sup>1</sup>

<sup>1</sup>School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287, USA.

<sup>2</sup>Biodesign Center for Molecular Design and Biomimetics, The Biodesign Institute and School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA

The ability to tune RNA dynamics is greatly needed for biotechnological applications, including therapeutics and diagnostics. Here, we systematically design a library of RNA modules named degradation-tuning RNAs (dtRNAs). These dtRNAs can form specific upstream secondary structures prior to ribosome binding site (RBS) to modulate transcript stability while having a minimal influence on translation initiation. We optimize the structural features of these RNA modules so that gene expression can be tuned by over 40-fold. We then confirm the viability of using dtRNAs to modulate gene circuit dynamics as well as noncoding sgRNA levels. We further expand the usage of dtRNA library into cell-free system and demonstrate its effectiveness of gene expression enhancement *in vitro*. Finally, we integrate our dtRNA modules with synthetic toehold sensor to enable rapid paper-based norovirus diagnostics even without the need for degradation inhibitor, illustrating the potential of synthetic dtRNA design for biotechnological applications.

## **The Human Blood Metabolome as a Readout of the Gut Microbiome**

**Tomasz Wilmanski<sup>1\*</sup>, Noa Rappaport<sup>1\*</sup>, John C. Earls<sup>1</sup>, Andrew T. Magis<sup>1</sup>, Ohad Manor<sup>2</sup>, Jennifer Lovejoy<sup>1</sup>, Gilbert S. Omenn<sup>3</sup>, Leroy Hood<sup>1</sup>, Sean M. Gibbons<sup>1,4</sup> and Nathan D. Price<sup>1</sup>**

<sup>1</sup> **Institute for Systems Biology, Seattle, WA**

<sup>2</sup> **Arivale, Seattle, WA**

<sup>3</sup> **Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA**

<sup>4</sup> **eScience Institute, University of Washington, Seattle WA**

The gut microbiome has been implicated in a number of human diseases, however the extent to which gut microbial composition is reflected in host physiology remains unclear. Using machine learning based approaches, we attempted to predict gut microbiome alpha-diversity from nearly 1000 measured blood analytes in individuals self-enrolled in a Scientific Wellness program (Discovery N=399, Validation N=540). Although a wide panel of common clinical laboratory tests and over 250 plasma proteins were not predictive of gut microbiome alpha-diversity, we identified a subset of 40 plasma metabolites (~33% of microbial origin) that collectively explained up to 50% of the variance in alpha-diversity measures. An even smaller subset of 11 metabolites could successfully classify individuals with low, and potentially problematic, alpha-diversity. The predictive capacity of our metabolite panel was consistent across validation cohorts and disease states, demonstrating the robustness of the identified relationships. The reflection of gut microbial alpha-diversity in the host metabolome also remained consistent across majority of the Body Mass Index (BMI) spectrum, but was modified in extreme obesity (class II/III, BMI $\geq$ 35), suggesting significant metabolic perturbation. Several of the microbial metabolites identified in our analysis have been previously shown to exert potentially detrimental effects on human tissues, including the heart (trimethylamine oxide), liver (imidazole propionate) and kidneys (p-cresol sulfate). Additional identified metabolic predictors have been previously shown to originate in the gut microbiome, but their impact on host physiology remains to be investigated. Collectively, our findings reveal a strong coupling between the human plasma metabolome and gut microbial structure, which opens new possibilities for studying the gut microbiome in the context of human health.

\* Equal contributions

Aaron Wong, awong@flatironinstitute.org, Flatiron Institute (Presenter)  
Julien Funk, jfunk@flatironinstitute.org, Flatiron Institute  
Kevin Yao, kyao@flatironinstitute.org, Flatiron Institute  
Chandra Theesfeld, chandra.theesfeld@princeton.edu, Princeton University  
Olga Troyanskaya, ogt@genomics.princeton.edu, Princeton University

HumanBase: A portal for data-driven predictions of gene expression, function, regulation and interactions

Biologists using modern experimental methods are generating massive amounts of genome-scale data in diverse cellular conditions. These large data collections are noisy, highly heterogeneous and must be interpreted in the context of tissue- and cell-type specificity — key aspects to understanding the precise actions of genes and their role in human disease. Integrative methods can prioritize relevant experiments and extract meaningful signal from large functional genomics data collections. However, access to and visualization of these results are critical to gleaning biological insights and developing follow-up experiments.

HumanBase ([hb.flatironinstitute.org](http://hb.flatironinstitute.org)) is a comprehensive online resource for biomedical researchers interested in exploring expression, function, regulation and interactions of human genes, particularly in the context of specific tissues/cell-types and human disease. Data-driven integrative analyses underlying HumanBase are especially powerful because they reach beyond existing “biological knowledge” represented in the literature to identify novel associations that are not biased toward well-studied areas of biomedical research. HumanBase uses machine learning to combine data from more than 38,000 genomic experiments and more than 14,000 scientific publications to uncover genes’ tissue-specific function and roles in disease (Krishnan et al., *Nat. Neurosci.*, 2016), inter-relationships between genes (Greene et al., *Nat. Genet.*, 2015) and the gene expression effects of genetic variants (Zhou et al., *Nat. Genet.*, 2018).

Biomedical researchers can query the web-based system, or programmatically through an API, with genes or genetic variants of interest and explore predictions with dynamic, interactive visualizations and analyses. HumanBase offers biologists sophisticated, data-driven tools to both interpret experimental results and drive new hypotheses and follow-up experiments.

## **Division of Labor Redirects Aberrant Cascading Cell Fate Decisions Perturbed by Resource Competition**

Xiao-Jun Tian

School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85281, USA,

Circuit-host interactions add an additional hidden layer to synthetic gene circuits and could significantly perturb the performance of the circuits. However, it is unclear how the resource competition within one gene circuit affects the whole circuit's function. It is very challenging to engineer one gene circuit to achieve accurate multiple cell fate transitions in one host cell due to the nonlinearity associated with the finite pool of available resources. Here, we first built a "synthetic cascading bistable switches (Syn-CBS)" circuit in one single strain with two coupled self-activation submodules to achieve two successive cell fate decisions. Interestingly, we found that the two switches cannot be activated at the same time, i.e., the activation of the second switch turns the first switch off. The loss of modularity results from resources competition between two submodules, which creates additional indirect inhibitive links between two modules. To decouple the resource competition between the two submodules and minimize unfavorable effects, we constructed a two-strain Syn-CBS circuit, which can achieve a stable co-activation of two coupled bistable switches. Thus, the effect of the resource competition within the Syn-CBS circuit is minimized through a division of labor using microbial consortia.

# **Development and application of a new optogenetic system to manipulate intracellular contractile force**

Kei Yamamoto<sup>1,2,3,\*</sup>, Haruko Miura<sup>2,3</sup>, and Kazuhiro Aoki<sup>1,2,3</sup>

<sup>1</sup> Department of Basic Biology, School of Life Science, SOKENDAI, Aichi, Japan

<sup>2</sup> Division of Quantitative Biology, NIBB, NINS, Aichi, Japan

<sup>3</sup> Quantitative Biology Research Group, ExCELLS, NINS, Aichi, Japan

Intracellular contractile force generated cooperatively by nonmuscle myosin II (NMII) and actin filaments is of critical importance for a wide range of physiological processes such as cell motility, cytokinesis, and morphogenesis of embryonic development. Although several chemical inhibitors for NMII and its regulators are available, there is still a need for a tool for controlling contractile force with high spatial and temporal resolution. Recently, several groups reported optogenetic tools capable of regulating local RhoA activity, leading to an increase in the intracellular contractile force. However, no reports have demonstrated decrease in contractile force with high spatial and temporal resolution.

To this end, we aimed to develop an optogenetic tool to manipulate contractile force. We focused on myosin light chain phosphatase (MLCP), which directly dephosphorylates myosin light chains (MLCs), leading to decrease the contractile force. The MLCP is composed of three subunits; a catalytic subunit of type I phosphatase (PP1c), a regulatory subunit (MYPT), and a smaller subunit (M20), a protein of unknown function. To control MLCP function by light, we employed CRY2-CIB1 light-induced heterodimerization system, and designed that endogenous PP1c is co-recruited to the plasma membrane through the membrane translocation of CRY2-fused PP1c binding domain of MYPT. Although overexpression of the PP1c binding domain of MYPT (1-296 aa) reduced MLC phosphorylation with or without the membrane translocation, the shorter variant of PP1c binding domain of MYPT showed nice response to blue light to induce MLC dephosphorylation. We further confirmed that PP1c is also recruited to plasma membrane along with the membrane translocation of PP1c binding domain of MYPT. Interestingly, local membrane translocation of the PP1 binding domain of MYPT evoked local membrane protrusion, indicating that local MLC dephosphorylation suffices to induce membrane protrusion. Therefore, this system would be a useful optogenetic tool to decrease the intracellular contractile force. Here, we will introduce the recent optimization and application of this newly developed optogenetic system.

# Kinetic Uncertainty Relations for the Control of Stochastic Reaction Networks

Jiawei Yan<sup>1\*</sup>, Andreas Hilfinger<sup>2,3\*</sup>, Glenn Vinnicombe<sup>4</sup>, and Johan Paulsson<sup>1</sup>

<sup>1</sup> Department of Systems Biology, Harvard University, 200 Longwood Avenue, Boston, MA 02115, USA

<sup>2</sup> Department of Chemical & Physical Sciences, University of Toronto, Mississauga, Ontario L5L 1C6, Canada

<sup>3</sup> Department of Mathematics, University of Toronto, 40 St. George St., Toronto, Ontario M5S 2E4, Canada

<sup>4</sup> Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom

Many biological systems form stochastic interaction networks where components present in low numbers affect each other's production or degradation rates. One particularly important type of interactions are feedback loops. In general, such stochastic networks are hard to analyze and small differences of parameters or functions in the model may lead to significant differences of dynamical behaviours. However, despite of various dynamics created by feedback loop, there could exist qualitative rules constraining the range of dynamical behaviours regardless of their detailed feedback mechanism, e.g., the Bendixson criterion in deterministic systems. Here we asked whether there exists such kind of broad rules on noise suppression for stochastic feedback systems. We started from analyzing the two-component feedback loop, allowing arbitrary rate functions and parameters. Intriguingly, we found that no matter what the rate functions are, such two-component feedback loop cannot suppress noise in both components simultaneously below its uncontrolled levels. This inaccessible "Poisson Square" forms a hard bound that describes a striking trade-off inherent to all systems in which two components directly control each other. We then extended our systems to which many components control each other in a ring-like structure but again allowing arbitrary rate functions. The results suggest that for such network topology, only one component can exhibit a sub-Poisson fluctuation. In other words, when setting up ring-like feedback control structure one has to decide which component it is whose fluctuation should be suppressed, and all other components become sacrificial ones to reduce noise in the chosen one. We also considered more general systems which  $N$ -components control each other's production rates in arbitrary topologies, and found a similar principle that the noise suppression is bounded by a  $N$ -dimensional "Poisson Hypercube"—meaning that at least one component has to display a fluctuation above Poisson noise. Our results suggest that it is impossible to ensure every component in any molecular regulatory network to display sub-Poisson fluctuations, reducing fluctuations in one components comes with expenses of increasing variations of some others.

\* Equal contributions

# ATP Demand Drives Metabolic Remodeling During Antibiotic Stress and Lethality

Jason H. Yang<sup>1,2</sup>, James J. Collins<sup>3,4,5</sup>

<sup>1</sup> Department of Microbiology, Biochemistry & Molecular Genetics, Rutgers New Jersey Medical School, Newark, NJ, 07103

<sup>2</sup> Center for Emerging and Re-Emerging Pathogens, Rutgers New Jersey Medical School, Newark, NJ, 07103

<sup>3</sup> Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, 02139

<sup>4</sup> Broad Institute of MIT and Harvard, Cambridge, MA, 02142

<sup>5</sup> Wyss Institute for Biologically Inspired Engineering, Boston, MA, 02115

Antibiotic stress induces significant metabolic remodeling in bacteria to fundamental processes such as central carbon metabolism as part of their lethality. However, the mechanisms driving such changes are poorly understood as it is difficult to decipher which metabolic pathways are relevant to antibiotic killing and how they are coupled. To address these, we developed a “white-box machine learning” approach integrating biochemical screening with metabolic network modeling and machine learning to rapidly identify metabolic pathways capable of altering antibiotic lethality. Applying this approach to three different bactericidal antibiotics in *E. coli*, we discovered the novel contribution of purine biosynthesis to antibiotic-induced death physiology. Model analysis revealed that purine biosynthesis imposes a significant burden on ATP utilization and demand, creating a large driving force for central carbon metabolism activity. Using a combination of genetic, biochemical, and phenotypic assays, we demonstrated that purine biosynthesis activity significantly contributes to both antibiotic lethality and antibiotic-induced changes in central metabolism through its effects on ATP demand. We further demonstrated that pyrimidine biosynthesis exerts opposite effects via network crosstalk with purine biosynthesis.

## Transient Laplacian Dynamics

Pencho Yordanov and Jeremy Gunawardena

Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA

A remarkable characteristic of some biological systems is their capability to quickly, yet accurately and robustly, perform information processing tasks. One prominent example is the switch-like transcriptional response of the gene hunchback observed in early *Drosophila* development, which is established in only a few minutes. The mechanisms giving rise to the steep response have recently been investigated using a linear modeling framework. The “linear framework” formalizes continuous-time Laplacian and Markovian dynamics on graphs in the context of biology and allows steady-state analyses that are analytic and parametric in the graph’s edge labels. However, the framework does not offer such a symbolic approach for the potentially important temporal aspect of the dynamics. Here we extend the scope of the linear framework to enable the parametric study of transient phenomena. Namely, we adopt a graph-theoretical interpretation of the Laplace transform, which represents the temporal dynamics on a graph in the time domain as the steady state on an augmented graph in the Laplace domain. The transformation permits the application of the parametric steady-state analysis toolbox, including prime factorization of Kirchhoff polynomials, to characterize the dynamics on general graphs. The developed theory allows us to study relaxation and first-passage processes, which we illustrate with a model of gene regulation.

## **Synthetic helper T cells drive local proliferation of therapeutic immune cells**

Nicholas W. Frankel<sup>1-3</sup>, Greg M. Allen<sup>1-4</sup>, Maia A. Yoshida<sup>1-3</sup>, Wei Yu<sup>1-3</sup>, Aileen W. Li<sup>1-3</sup>,  
K. Christopher Garcia<sup>8</sup>, Kole T. Roybal,<sup>5-7</sup> Wendell A. Lim<sup>1-3,6</sup>

<sup>1</sup>Department of Cellular and Molecular Pharmacology, <sup>2</sup>Cell Design Institute, <sup>3</sup>Howard Hughes Medical Institute, <sup>4</sup>Department of Hematology and Oncology, <sup>5</sup>Department of Microbiology and Immunology, <sup>6</sup>Helen Diller Family Comprehensive Cancer Center, <sup>7</sup>Parker Institute for Cancer Immunotherapy, University of California, San Francisco, CA 94158, USA

<sup>8</sup>Departments of Molecular and Cellular Physiology and Structural Biology, Howard Hughes Medical Institute, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

Cell proliferation is a central part of a strong immune response, either by native or engineered T cells. While engineered T cell therapies such as chimeric antigen receptor (CAR)-T cells have recently exhibited substantial success in fighting blood cancers, extending T cell therapies to solid cancers will require tools to drive immune cell proliferation locally to achieve a sufficient therapeutic effect within the tumor microenvironment. Cytokines such as IL-2, IL-7, and IL-15, produced by immune cells like helper T cells are known to provide activation, survival, and/or proliferation signals to modulate lymphoid homeostasis and effector cell function. Inspired by natural signaling systems, we designed synthetic helper T cells that inducibly secrete proliferative cytokines upon recognition of locally specific tumor antigen. Remarkably, these cells expanded locally *in vivo* without directly relying on T cell activation through a T cell receptor (TCR) or CAR. Synthetic helper T cells could therefore be a powerful tool to locally target expansion of endogenous or adoptively transferred immune cells for many disease applications.

## **Topology-Dependent Interference of Circuit Function by Growth Feedback**

Rong Zhang, Xiao-Jun Tian

School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85281, USA,

Growth-mediated feedback between synthetic gene circuits and host organisms leads to diverse emerged behaviors, including growth bistability and enhanced ultrasensitivity. However, the range of possible impacts of growth feedback on different gene circuits remains underexplored. Here, we mathematically and experimentally demonstrated that growth feedback affects the functions of memory gene circuits in a network topology-dependent way. Specifically, the memory of the self-activation circuit is quickly lost due to the fast growth-mediated dilution of the circuit products. Decoupling of growth feedback reveals its memory, manifested by its hysteresis property across a broad range of stimulus. On the contrary, the toggle switch is more refractory to the growth-mediated dilution and can retrieve its memory after the fast-growth phase. The underlying principle lies in the different dependence of active and repressive regulations in these circuits on the growth-mediated dilution. Our results unveil the topology-dependent mechanism on how growth-mediated feedback influences the behaviors of gene circuits.

## **Integrating mathematical modeling and computational genomics reveals design features of a gene regulatory network controlling activated B cell dynamics**

Ankur S. Saini<sup>1</sup>, Yin Zhang<sup>1</sup>, Guan-Sheng Liu<sup>1</sup>, Jeremy Riddell<sup>1</sup>, Krista Dienger-Stambaugh<sup>1</sup>, Kellen Andrienas<sup>2</sup>, Masashi Yukawa<sup>1</sup>, Artem Barski<sup>1</sup>, Trevor Siggers<sup>2</sup>, Matthew T. Weirauch<sup>1</sup>, Tongli Zhang<sup>1</sup> and Harinder Singh<sup>1</sup>

<sup>1</sup>Cincinnati Children's Hospital Medical Center and the University of Cincinnati, Cincinnati OH 45229, <sup>2</sup>Boston University, Boston MA, 02215

Gene regulatory networks (GRNs) are widely utilized in systems biology but their explanatory power remains to be fully harnessed by integration of mathematical modeling and computational genomics with experimental testing. We undertook such analysis of a GRN that controls exceptional B cell fate dynamics. Upon sensing pathogens, B cells undergo bifurcating trajectories initiated by the transcription factors (TFs) IRF4 and IRF8 that converge to generate germinal center (GC) independent or GC-dependent plasma cells. Combining mechanistic modeling with machine learning, we uncover a dominant parameter in the GRN that underlies this unusual emergent property. This prediction led to the identification of NFATc2 as a selective signaling induced modulator of IRF8. As predicted, loss of NFATc2 promoted the GC-independent trajectory. A novel computational genomic strategy further revealed a stereospecific NFAT-IRF composite element (NICE) that promoted cooperative binding of NFATc2 with IRF8, thereby uncovering a feedforward loop that reinforces the GC trajectory. Collectively, this generalizable strategy illustrates the power of combining mechanistic modeling, machine learning and computational genomics with experimental testing to uncover key design principles underlying GRNs.

## **reCod: Challenges in genome-scale metabolic reconstruction of a non-model species**

Eileen Marie Hanna<sup>1</sup>, Xiaokang Zhang<sup>1</sup>, Marta Eide<sup>2</sup>, Shirin Fallahi<sup>3</sup>, Fekadu Yadetie<sup>2</sup>, Inge Jonassen<sup>1</sup>, Anders Goksøyr<sup>2</sup>

<sup>1</sup>Computational Biology Unit, Department of Informatics, University of Bergen, Norway

<sup>2</sup>Department of Biological Sciences, University of Bergen, Norway

<sup>3</sup>Department of Mathematics, University of Bergen, Norway

The availability of genome sequences, annotations and knowledge of the biochemistry underlying metabolic transformations has led to the generation of metabolic network reconstructions for a wide range of organisms including bacteria, archaea, and eukaryotes. Such networks include the form and function of genes in a target organism. When modeled using mathematical representations, a reconstruction can simulate the underlying genotype-phenotype relationships. Accordingly, genome-scale models (GEMs) can be used to predict the reactions of organisms to genetic and environmental variations. A bottom-up reconstruction procedure typically starts by generating a draft model from existing annotation data on a target organism. For model species, this part of the process seems to be the least labor-intensive, due to the likely abundant organism-specific biochemical data. However, the process becomes intricate for non-model less-annotated species. In this work, we present our case study on Atlantic cod (*Gadus morhua*), a non-model teleost fish. The liver is the most important organ in metabolism of nutrients and toxicants. For example, the highly lipid-rich liver of cod accumulates lipophilic environmental pollutants and therefore becomes an interesting tissue to study the toxicant biotransformation system of cod. Using three available cod genome assemblies and considering human liver model as template, we drafted the cod liver model and manually curated it. By this work, we showed a practicable guide in terms of steps, methods, choice of tools, and challenges, for cases with comparable resources.

*This study is part of the dCod 1.0 project funded by the Research Council of Norway (project no. 248840).*

# **An analytical framework for interpretable ‘quasilinear’ single-cell omics analysis**

Jian Zhou<sup>1,2</sup> and Olga G. Troyanskaya<sup>2,3,4</sup>

<sup>1</sup> Lyda Hill Department of Bioinformatics, UT Southwestern Medical Center, Dallas, TX, USA

<sup>2</sup> Flatiron Institute, Simons Foundation, New York, NY

<sup>3</sup> Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ

<sup>4</sup> Department of Computer Science, Princeton University, Princeton, NJ

Scaling single-cell data exploratory analysis with the rapidly growing landscape of single-cell omics data demands interpretable and robust data representation that is generalizable across datasets. Unlike linear representations that are easily interpretable via linear mapping, transferable across datasets, and amenable to statistical inference, nonlinear representations that are widely used for single-cell data do not generally possess those desirable properties. To address this challenge, here we present a novel ‘quasilinear’ framework that combines the interpretability and transferability of linear methods and the representational power of nonlinear methods. Specifically, we introduce a fast and scalable quasilinear data representation method, GraphDR, and a general structure discovery framework, StructDR, that unifies cluster, trajectory, and surface estimation and allows their confidence set inference. Both methods produce representations interpretable in linear space yet preserve or even outperform the cell state representation qualities of common nonlinear methods. These methods also facilitate easy comparison and generalization of knowledge across datasets. In addition, GraphDR is at least an order of magnitude faster than commonly used nonlinear visualization methods. These methods thus contribute to a basis for interpretable single-cell data exploratory analysis and statistical inference, as we demonstrate with extensive evaluations including achieving top performance in a large public benchmark dataset and with applications to diverse datasets in several organisms. An open-source python library and a user-friendly graphical interface for 3D data visualization and analysis with these methods is available at [github.com/jzthree/quasildir](https://github.com/jzthree/quasildir).

## The geometry of blood chemistries and wellness states

Anat Zimmer<sup>1</sup>, Yael Korem<sup>2</sup>, Noa Rappaport<sup>1</sup>, Thomas Wilmanski<sup>1</sup>, Priyanka Baloni<sup>1</sup>,  
Leeroy Hood<sup>1</sup>, Nathan Price<sup>1</sup>

<sup>1</sup>Institute for Systems Biology, Seattle, WA, USA

<sup>2</sup>Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100

It is a challenge to answer questions like: why some people develop a disease, react to a specific treatment and/or develop severe side-effects while others don't. In order to explain these occurrences, one has to take a holistic approach and study the body physiology from a systems level perspective. Longitudinal multi-omics measurements together with genetics, on a large population, can serve such a purpose and help in predicting, reasoning, and preventing diseases.

In partnership with Arivale Inc., we have developed infrastructure to collect longitudinal Personalized Dense Dynamic Data clouds (PD3 clouds) on thousands of individuals, which include genetics and longitudinal measurements of clinical labs, microbiome, metabolome, proteome, and self-reported data.

The value of these extremely high-dimensional data clouds is clear; however, it also comprises a challenge in data analysis and interpretation.

One way to reduce data dimensionality is called Pareto Task Inference (PARTI, Hart *et al.* 2015). We used this method to analyze the clinical labs and found that the data falls on a significant tetrahedron. The 4 vertices are archetypes that specialize in a certain task. Using all other datatypes, we identified enriched traits next to every archetype and revealed the underline tradeoffs that shape the data.

This distinctive analysis uncovers unexpected relationships between datasets such as metabolomics, proteomics and clinical labs, and helps in interconnecting the different datatypes to characterize different states of human health.