Integrated circuit design and cellular engineering: Probing and reshaping the genome to control cell fate

As a chemical engineer specializing in molecular systems biology, my ultimate goal is to leverage synthetic biology principles to transform how we understand cellular transitions and engineer cellular therapies.

I. Summary. Native biological circuits (i.e. connected networks of genes) utilize data within the cell to coordinate cellular behaviors in space and time. Synthetic biology aims to harness the power of biological systems to dynamically access information in the cell, enabling synthetic biomedical tasks such as tumor surveillance, pathogen identification, or metabolite manipulation. My research focuses on integrating synthetic circuitry to interrogate and drive cellular behaviors. As a postdoc, I discovered that topoisomerases—the enzymes responsible for curating DNA topology—act in a rate-limiting step to regulate direct lineage reprogramming. My discovery suggests that topological stress across the genome represents a primary barrier to cellular reprogramming (Fig. 1). In fact, increasing topoisomerase concentration increases the speed, efficiency, and efficacy of reprogramming. This finding opens completely new questions around how the structure of the human genome stabilizes cellular identity and buffers cells against transitions to pathological states. My goal is to establish a leading research program focused on designing and constructing integrated synthetic circuits to probe and actuate changes in DNA topology that drive changes in cell fate.

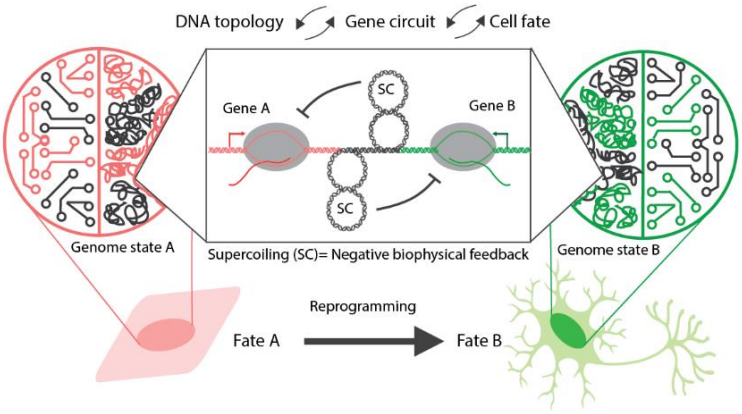

Figure 1. Gene circuits integrated into the genome drive cellular reprogramming and the transition from one genomic state to another. By providing biophysical feedback, local DNA topology may be harnessed to buffer or accelerate cellular transitions mediated through gene circuits.

II. Research Accomplishments. Spanning a range of model organisms (e.g. yeast, mouse, and human cells), I have engineered systems for dynamic behaviors across multiple scales, from the molecular design of noncoding RNA devices to the optimization of large transcriptional networks.

Graduate: Working with Christina Smolke at Caltech, I constructed synthetic gene circuits and developed a class of genetic control systems called "molecular network diverters." By interfacing these molecular network diverters with the native MAPK signaling pathway, I could temporally and spatially control cellular decision-making events in yeast cells (Galloway, KE et al. *Science*. 2013). Beyond controlling cell fate, my work highlighted the principle that integrated negative regulators can buffer a system against noise amplification, mediated through positive feedback loops by providing resistance to amplification (Franco, E and Galloway, KE. *Computational Methods in Synthetic Biology*, 2015).

Postdoc: Working with Justin Ichida at USC Stem Cell, I have focused on elucidating and overcoming the reprogramming roadblocks to the robust generation of mature neural cell types from fibroblasts, a readily available cell source. Accurately modeling neurological disorders with *in vitro* cellular models requires reliable methods of generating the distinct neural subpopulations affected by the disease. The conversion process was extremely inefficient when I began my project; large-scale efforts yielded only a few hundred cells. Moreover, the central mechanistic rules for direct lineage conversion were undefined. Today, as a direct result of my work to improve the reprogramming process, I can robustly generate tens of thousands of cells with signatures of enhanced maturity (Babos, KN*, Galloway, KE*, et al. Balancing dynamic tradeoffs to drive cellular reprogramming. *In revision*.). In addition to improving the reprogramming process, my work uncovers a previously unrecognized explanation for why only rare cells successfully undergo reprogramming (Fig. 2). I discovered that cellular reprogramming is limited to a small population of cells equipped to process the dual, competing demands of hyperproliferation and hypertranscription. High rates of transcription and replication accelerate the rate of DNA tangling (e.g. supercoiling). Only cells with high expression of topoisomerases—which relax DNA supercoils—are capable of processing the massive genomic and transcriptional realignment to convert from one state to another. My findings suggest that topological stress impacts the function of gene networks (e.g. native or synthetic circuits), limiting access to some transcriptional states. By more precisely defining the impact of topological stress on gene circuits, I am in the process of elucidating principles for buffering (or alternatively, harnessing) topological stress to enhance the performance and capabilities of genome-integrated synthetic circuits.

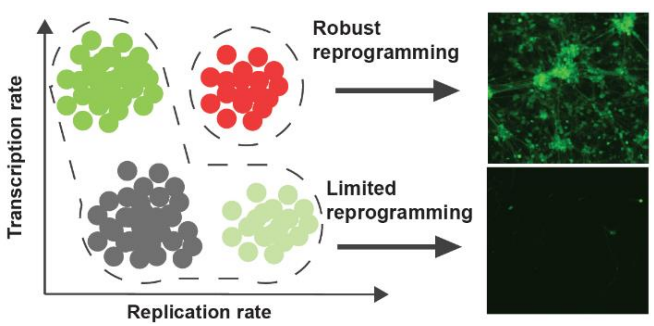

Figure 2. Cells capable of high transcription and high replication rates represent a rare population capable of reprogramming at high efficiency. Expanding this population results in robust, complete reprogramming.

III. Research Directions. Prokaryotic systems and single-cell eukaryotic organisms have dominated the field of synthetic biology and revealed important paradigms in cellular biology, including the roles of feedback, noise, and cooperativity (*1*). However, translation of synthetic biology to mammalian systems has been slow. With the advent of improved genetic tools for mammalian cells (e.g. CRISPR technologies and AAVs), synthetic circuits will reshape how we study and treat diseases. Elucidating the principles of mammalian circuit design offers the opportunity to engineer cellular behaviors and to identify and target diseased states. Further, understanding how cell types differentially process classes of synthetic circuits will improve our ability to predict the response of native transcriptional networks and design systems that are optimally wired for their function and context. My lab will harness the insight that emerges from basic research questions (e.g. how DNA topology influences gene circuits and cell fate (Fig. 1)) to develop tools for biomedical applications (Fig. 3).

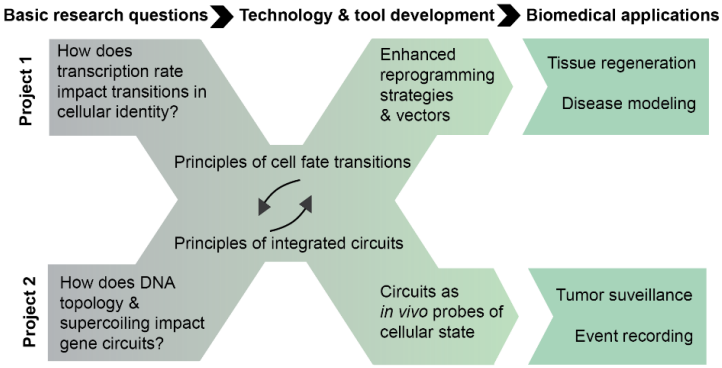

Figure 3. My lab will connect basic research questions in gene circuits, genome architecture, and cell fate transitions to tool development for biomedical applications.

Project 1: Tailoring transcription rate to scale with cellular capacity during reprogramming. While transcription drives changes in cellular state, my own research indicates that high transcription rates early in reprogramming actually inhibit proliferation and stall reprogramming. Developing strategies to 1) prohibit unnecessary transcription from competing gene regulatory networks and 2) tune expression of reprogramming factors may facilitate more rapid, efficient reprogramming and enable applications such as tissue regeneration.

Aim 1.1. Examining the role of premature hypertranscription in reprogramming. My previous work established that hypertranscribing, hyperproliferating cells (HHCs) preferentially undergo direct lineage reprogramming. Inclusion of DDRR, a cocktail of genes and small molecules, increases the population of HHCs. However, DDRR globally reduced transcription (Fig. 4A). Given the generally antagonistic relationship between transcription and proliferation, DDRR may increase proliferation by transiently reducing the transcription rate. But how? Via single-cell RNAseq, I discovered that DDRR increased expression of inhibitor of differentiation (Id) genes (e.g. Id1, Id2, Id3) (Fig. 4B). As helix-loop-helix (HLH) proteins, Id proteins resemble basic HLH (bHLH) factors, a large class of transcription factors. Id genes are expressed as a cell transitions identity in development. Effectively acting as competitive inhibitors, Id proteins block dimerization of bHLH factors, preventing the factors from binding to DNA and inducing transcription. Id expression promotes both iPSC reprogramming and cancer (*2, 3*). Together with my data, this suggests that Id genes may facilitate cellular transitions by transiently reducing transcription. By overexpression and knockdown of Id genes, my lab will examine the role of Id genes and define their impact on transcription and reprogramming. Harnessing Id genes may facilitate a transcriptional "reset" during cellular transitions, enabling more efficient and effective reprogramming strategies.

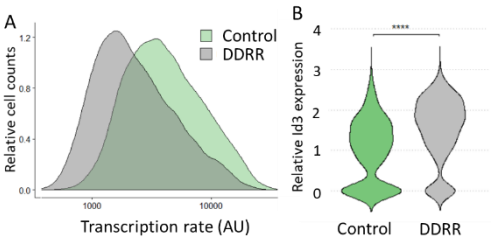

Figure 4. Inclusion of DDRR in reprogramming globally reduces the transcription rate. Reduced transcription rate coincides with increased expression of Id genes (e.g. Id3).

Aim 1.2. Scaling circuit output to match cellular capacity. Hypertranscription early in reprogramming limits cellular division, reducing cellular reprogramming. However, hypertranscription in hyperproliferative cells drives cellular reprogramming. Consequently, hypertranscription and hyperproliferation represent dynamic objectives in reprogramming. To effectively balance both processes, reprogramming vectors tailored to scale transcription with the capacity of individual cells may improve reprogramming strategies. Circuits that scale with cellular capacity may more efficiently guide cells through reprogramming by limiting transcriptional strain on the genome. Simple selection of promoters to regulate transcription factor expression may be sufficient to improve expression-scaling from transgenic constructs. Alternatively, more complex control systems, such as those I developed in my graduate work, may enable proportional control of expression linked to cellular metabolism. Using a multi-cistronic construct of all three motor neuron factors to control for stoichiometry, my lab will examine how an array of promoters affects reprogramming of fibroblasts. I will assay how promoter selection impacts cellular proliferation and reprogramming efficiency. Improved designs may be rationally iterated through additional design schemes to tailor expression to a variety of cell types.

Aim 1.3 Long-term vision: Towards tissue regeneration: Developing tools and insights to enable *in vivo* reprogramming. Tissue regeneration to replace damaged or diseased cells is a prime objective of regenerative medicine. Reprogramming native cell populations *in vivo* may provide the most rapid and least invasive approach to regeneration. In addition to guaranteeing patients an immunological match, *in vivo* reprogramming avoids the time-consuming and labor-intensive *ex vivo* processing of cells. With improved *in vivo* delivery vehicles, conversion efficiency in human cells remains the primary barrier to *in vivo* reprogramming therapies. While my previous work improved reprogramming of human cells ten-fold, reprogramming efficiency remains low (15%). Comparison of mouse

and human reprogramming suggests mechanisms that limit the human system. Additionally, translating the findings from Aims 1.1 and 1.2 into methods that promote *in vivo* reprogramming will be an important long-term goal of my lab.

Project 2: Integrated gene circuit design. One challenge in synthetic biology is integrating synthetic circuitry into larger transcriptional networks to mediate predictable cellular behaviors. While significant efforts have been devoted to the logical design of enhanced synthetic circuitry (e.g. circuits for synchronized quorum sensing, edge-detection), less is understood regarding how cellular hardware and the emergent three-dimensional structure of genetic elements affect circuits. Understanding how transcription reshapes DNA and impacts gene circuits will provide the insight for building circuits capable of monitoring changes in DNA topology in living cells.

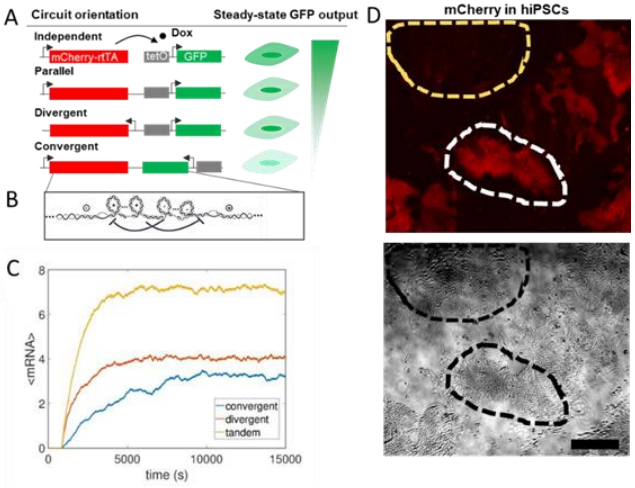

Figure 5. Supercoiling in synthetic gene circuits. A. Schematic of tet-ON inducible circuit with constitutively expressed red (mCherry-, rtTA) inducing GFP from a tetO promoter in different orientations produce varying steady-state GFP expression (**B and C)**. Theoretical modeling predicts that supercoiling barriers impact dynamics and steady-state gene expression from convergent genes more than other orientations. C is adapted from (22). **D.** Silencing of integrated constitutive mCherry in differentiating human iPSCs (yellow dashes) compared to maintained colony (white dashes).

Aim 2.1. Investigating the impact of compositional context on circuit performance. In a post-genomics era, the three-dimensional structure of the mammalian genome has emerged as a powerful mediator of cellular behavior (*4*). The three-dimensional structure of DNA, influenced by local DNA topology (e.g. supercoiling) and the epigenome, modulates gene expression (*5*). Consequently, constructing robust transcriptional circuits requires understanding and accounting for the influence of DNA topology (Fig. 1 and Fig. 5A-B). As recently detailed through computational modeling, compositional context (i.e. the relative arrangement of genetic elements in direction and spacing) impacts gene expression (Fig. 5C) (*6*). In bacteria, compositional context within a circuit impacts the behavior of bistable switches with topological barriers forming between convergently-oriented genes (*7*). Topological barriers introduce a layer of biophysical feedback, increasing the ultrasensitivity of state switching. In the mammalian context, circuits display varying functionality in different cell types, often losing function as they transition from states with high topoisomerase expression (e.g. iPSCs) to states with lower expression (e.g. differentiated cells) (Fig. 5D) (*8*). To examine the phenomena of circuit silencing, I will compare performance of an inducible circuit (tet-ON) in convergent, parallel, and divergent orientations integrated at one of several "safe-harbor" sites (Fig. 5A). Altering the circuit logic (e.g. toggle, pulse-generator, oscillator) will enable me to define how compositional context affects different classes of circuits. Further, I will examine how various integration sites topologically influence circuit performance. Defining endogenous topological influence on integrated circuits is essential for guiding rational designs and enabling cellular engineering. From these data, I will exact principles of three-dimensional circuit design for predictable functions of integrated circuits.

Aim 2.2. Measuring transition speed: Characterizing the dynamics of proliferation-mediated transitions between cell states. The synthetic context of reprogramming provides a model of cellular state switching that is unconstrained by development. Thus, reprogramming provides a unique opportunity to parse the contribution of developmentally-entangled systems-level phenomena (e.g. transcriptional activity, proliferation) to cellular transitions. State transitions from one cell fate to another represent transitions in large transcriptional networks. Transitions in large networks can be modeled by transitions in small transcriptional networks (e.g. bistable switches), which can be more easily measured and modified. Consequently, insights derived from a simple switch model can be directed to facilitate desirable transitions (e.g. reprogramming) or inhibit pathological transformations (e.g. cancer). Using a synthetic circuit as a probe, I will characterize how direct conversion and hyperproliferation affect the dynamics of bistable transcriptional networks. Given that hyperproliferating cells contribute to the majority of reprogramming events (*9*), proliferation may be a general motif to promote cell fate transitions. To examine the generality of this phenomena in the context of conversion, I will use a simple synthetic circuit (i.e. an inducible switch) controlling GFP to measure how proliferation affects the dynamics (e.g. the rate of GFP induction) and penetrance of induced transitions between two states (e.g. fraction of GFP+ cells).

Aim 2.3 Long-term vision: TANGLES as probes: Constructing circuits to assay for topological influence in the genome. Insight into how DNA supercoiling directly impacts cellular events is limited. Methods for examining DNA topology require harvesting millions of cells for biochemical analysis (*10*). Assaying topological states in living cells via circuits constructed as probes represents an opportunity to examine how the structure of the genome influences cellular decision-making at the single-cell level. From my work examining how compositional context influences circuit output (section 2.1), I will identify candidate TANGLES (topologically-affected networks of genes linking expression to state). TANGLES link circuit activity to topological state, which is a function of topoisomerase activity. Circuits that are silenced in cells with low topoisomerase activity represent candidate TANGLES. My lab will construct and characterize the performance of an array of TANGLES across several cell types, including healthy primary cells with low

topoisomerase expression (e.g. fibroblasts) and high expression (e.g. iPSCs, cancer cell lines). TANGLES that function in "topologically privileged" contexts (i.e. where highly active topoisomerases unwrap the DNA to maintain circuit function) are expected to be silenced in non-privileged states. Given the role of topoisomerase in cancer (*11*), the long-term vision of developing TANGLES is to deploy these probes as surveillance circuits, allowing us to dynamically monitor and potentially target diseased cell states. More immediately, TANGLES may be used to live-monitor how cellular events (e.g. reprogramming, division, signaling) impact circuit expression.

Project 3. Modeling of biological systems with systems-identification techniques. Myriad molecular species interact across diverse time-scales within the cell to generate expression of a single protein. Consequently, precise molecular modeling of large transcriptional networks across a heterogeneous cellular population becomes unfeasible. Alternatively, system identification techniques may facilitate extraction of principles governing molecular and phenotypic relationships in the cell. For example, in generating a model of the yeast pheromone-responsive MAPK pathway, I experimentally generated transfer functions for the relationship between the pathway output and the expression of an array of pathway components. From this data and a set of known molecular interactions, I constructed a phenomenological model of signaling through the pathway. I subsequently built open- and closed-loop genetic controllers called "molecular network diverters." By interfacing these molecular network diverters with the native MAPK signaling pathway, I could temporally and spatially control cellular decision-making (Fig 6A). Beyond controlling fate, the design and construction of molecular network diverters allowed me to explore fundamental principles of integrating native and synthetic circuitry. Circuits can be structured to buffer a system against circuit leakage and noise amplification (Fig. 6B). Tradeoffs in the system indicate limits in maximizing pathway output while minimizing noise (e.g. buffer capacity). Currently, I am constructing a model of cellular reprogramming focused on examining how topoisomerase activity modulates DNA supercoiling to set transcription and replication rates. I plan to develop this model to examine the function of synthetic circuits constructed in project 2 in my independent lab. By pairing modeling with the construction of synthetic circuitry, I will illuminate paradigms in biological control and apply those principles to enhance cellular reprogramming.

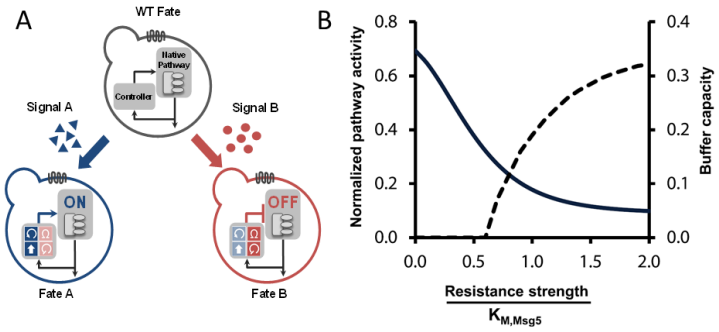

Figure 6. Dynamically programming cell fate with synthetic genetic controllers. A. Small molecule addition induces activity of regulatory modules to route cells to divergent fates. **B.** Buffer capacity (dashed line) increases with increasing resistance strength and is inversely related to pathway activation (solid line).

IV. Impact. Within mammalian systems, there exists an enormous opportunity to use synthetic circuitry to dynamically probe transitions within and across normal and pathological states. My research indicates that topological effects, such as DNA supercoiling, impact the activity of native and synthetic gene circuits. However, methods to examine these phenomena are limited to large-scale biochemical assays precluding live single-cell tracking. To address this issue, my lab will develop and characterize circuits capable of translating global changes in DNA supercoiling to changes in circuit output. Developing these tools will enable my laboratory to utilize synthetic circuits as sensors to probe the transcriptional and topological dynamics of cell fate transitions. The ultimate goal of my research is to transform how we understand cellular transitions and to develop cellular therapies by engineering circuits to leverage their three-dimensional context. The principles I develop for integrated circuit design will guide the next generation of integrated genomic circuits to enable biomedical applications, such as tissue regeneration and surveillance of pathological states.

V. Funding. To fund my research, I anticipate support from NIH (NINDS, NCI), the ALS Association (ALSA), and NSF. As my NRSA Fellowship through the National Institute of Neurological Disease and Stroke (NINDS) indicates, my projects to improve cellular reprogramming for disease modeling and to develop *in vivo* reprogramming strategies align with NINDS mission to "translate basic and clinical discoveries into better ways to prevent and treat neurological disorders." Development of a synthetic circuit probe within the model of reprogramming fits with the NSF Systems and Synthetic Biology Cluster's stated objectives of funding "systems-level, theory-driven analysis of regulatory, signaling and metabolic networks" and "tool development to facilitate systems and synthetic biology studies."

VI. References.

1. Y. Y. Chen, K. E. Galloway, C. D. Smolke, Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome Biol* **13**, 240 (2012).
2. Y. Hayashi *et al.*, BMP-SMAD-ID promotes reprogramming to pluripotency by inhibiting p16/INK4A-dependent senescence. *PNAS* **113**, 13057-13062 (2016).
3. H. F. Yuen *et al.*, Id-1 and Id-2 are markers for metastasis and prognosis in oesophageal squamous cell carcinoma. *Br J Cancer* **97**, 1409-1415 (2007).
4. D. U. Gorkin, D. Leung, B. Ren, The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762-775 (2014).
5. F. Kouzine, D. Levens, L. Baranello, DNA topology and transcription. *Nucleus* **5**, 195-202 (2014).
6. S. A. Sevier, H. Levine, Properties of gene expression and chromatin structure with mechanically regulated elongation. *Nucleic Acids Res* **46**, 5924-5934 (2018).
7. E. Yeung *et al.*, Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst* **5**, 11-24 e12 (2017).
8. E. P. Papapetrou, A. Schambach, Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol Ther* **24**, 678-684 (2016).
9. S. Guo *et al.*, Nonstochastic reprogramming from a privileged somatic cell state. *Cell* **156**, 649-662 (2014).
10. S. S. Teves, S. Henikoff, Transcription-generated torsional stress destabilizes nucleosomes. *Nat Struct Mol Biol* **21**, 88-94 (2014).
11. J. L. Nitiss, Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer* **9**, 338-350 (2009).