

Cellular Reprogramming: Mathematics Meets Medicine

Gabrielle A. Dotson¹, Charles W. Ryan^{1,2,3}, Can Chen^{4,5}, Lindsey Muir¹, and Indika Rajapakse^{1,4,*}

¹Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, 48109, USA

²Program in Cellular and Molecular Biology, University of Michigan, Ann Arbor, Michigan, 48109, USA

³Medical Scientist Training Program, University of Michigan, Ann Arbor, Michigan, 48109, USA

⁴Department of Mathematics, University of Michigan, Ann Arbor, Michigan, 48109, USA

⁵Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, 48109, USA

*To whom correspondence should be addressed; E-mail: indikar@umich.edu.

Abstract

Generating needed cell types using cellular reprogramming is a promising strategy for restoring tissue function in injury or disease. A common method for reprogramming is addition of one or more transcription factors that confer a new function or identity. Advancements in transcription factor selection and delivery have culminated in successful grafting of autologous reprogrammed cells, an early demonstration of their clinical utility. Though cellular reprogramming has been successful in a number of settings, identification of appropriate transcription factors for a particular transformation has been challenging. Computational methods enable more sophisticated prediction of relevant transcription factors for reprogramming by leveraging gene expression data of initial and target cell types, and are built on mathematical frameworks ranging from information theory to control theory. This review highlights the utility and impact of these mathematical frameworks in the field of cellular reprogramming.

1 Introduction

In 1989, pioneering work by Hal Weintraub demonstrated conversion of human skin cells into muscle cells through overexpression of a single transcription factor, *MYOD1*. In 2007, Shinya Yamanaka reprogrammed human skin cells into induced pluripotent stem cells (iPSCs) using four transcription factors, a discovery that would earn him a Nobel Prize in 2012. These remarkable findings demonstrated that the genome is a system that can be controlled via an external input of transcription factors. Following these experimental discoveries, engineering, statistical and mathematical methods predicting candidate transcription factors for cellular reprogramming began to emerge.

In this review, we show how experimental methodologies for cellular reprogramming and mathematics have come together to enhance our understanding and control of cell fate conversions. We start with a comprehensive overview of key milestones in cellular reprogramming before touching on the translational impact in the field thus far. We then offer a mathematical perspective of cellular reprogramming, covering the use of mathematical principles to describe cellular states and their dynamics. Additionally, we explore how we can harness mathematics to improve the specificity and efficacy of cellular reprogramming by summarizing existing computational frameworks designed to facilitate and validate reprogramming methodologies. Finally, we discuss future applications and implications of engaging mathematics in the domain of cellular control.

2 Background

The premise of cellular reprogramming arose following the discovery of DNA as the "genetic code", with all cells in an organism possessing the same sequences of DNA. From this point forward, biologists sought to understand how cells with the same underlying code take on different phenotypes. From neurons and cardiomyocytes to dermal epithelia, cells with identical DNA somehow differentiate into subsets of cells with complementary functions in a beautifully coordinated process. These specialized cells coalesce to form tissues with emergent functions, which in turn form organs and ultimately organisms. While this process is tightly controlled with awe-inspiring fidelity, errors do occur leading to a wide spectrum of human pathology. Errors along the developmental process, either due to germline mutations or environmental exposures, may lead to developmental malformations with high morbidity and mortality. In addition, tumorigenesis is often considered the de-differentiation of a cell, wherein a cell loses its identity and fails to conform to its intended function (Yamada et al. 2014). Finally, tissues begin to fail with age as the epigenetic factors that define cell identity wane, rendering a need for restorative cellular reprogramming techniques (Kimmel et al. 2019; Kanherkar et al. 2014). For these reasons and more, it is necessary to develop our understanding of the underlying mechanisms of cell fate decision making, and to use that knowledge to improve methods for control of cellular differentiation.

2.1 Cellular Differentiation

Typically, cells are envisioned as having a fixed identity capable of performing specialized functions. Indeed, fully differentiated cells show remarkable functional stability over a range of physiological conditions. However, cells do not start off this way but rather develop from precursor cells called stem cells. The earliest stem cells, embryonic stem cells (ESCs), form following conception of a zygote and are totipotent, meaning they can give rise to all cell types. These totipotent cells proliferate, and some of their progeny take on specialized identities through a process called differentiation. As a cell differentiates, the spectrum of possible progeny cells that it can give rise to narrows and it begins to express only those genes needed for its subsequent role. As a zygote develops, gradients of cell signals across the principle axes of the organism lead to differential expression of transcription factors, which modify the expression levels of genes in such a way as to give rise to distinct phenotypes, thus differentiating cells from one another (Figure 1A). Over time, this difference in gene expression caused by transcription factor availability is thought to be reinforced by epigenetic markers, which package the DNA depending on the needs of the cell. Actively expressed genes are left relatively unpacked (euchromatin) while silenced genes are tightly stowed away (heterochromatin) (Allshire and Madhani 2018).

The mechanism by which cells suppress unused genes as they differentiate and specialize their

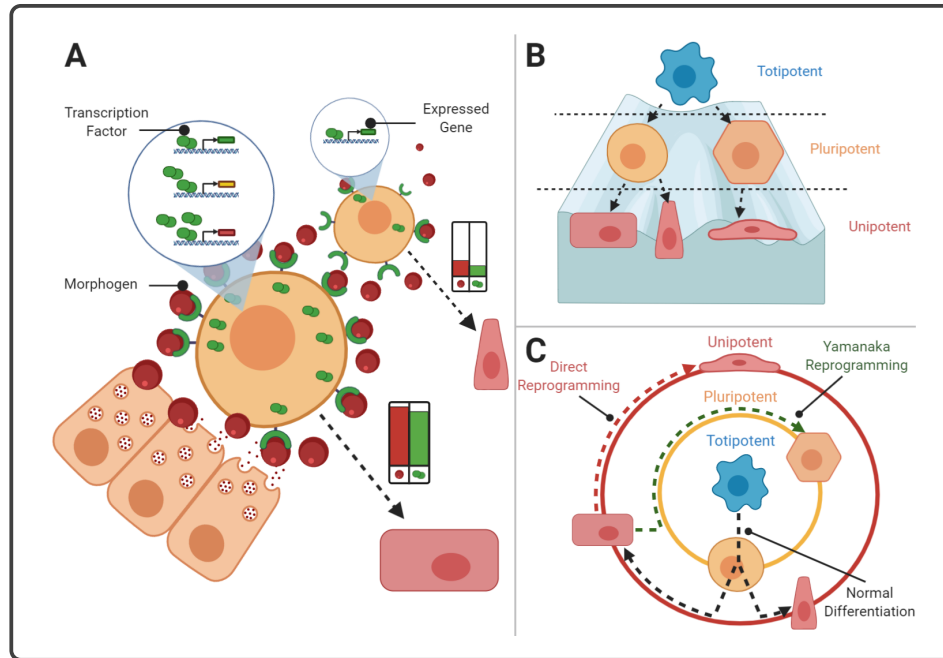


Figure 1: (A) Cellular differentiation in developmental biology. During normal development, concentration gradients of morphogens (red spheres) lead to differing levels of transcription factor (dimeric green spheres) activation in a distance-dependent manner. This leads to differing transcription profiles of cells as a function of spatial location, allowing for body patterning. (B,C) Re-imagining of Waddington's epigenetic landscape. Waddington's original model (B) best describes the process of cellular differentiation and presents an intuitive illustration of non-specialized cell types traversing down peaks and settling in valleys once specialized. In a re-imagined version of Waddington's landscape (C), the essence of cellular reprogramming is captured, where a cell is unimpeded by the gravity of a hierarchical model and can flow between any potency, germ layer, and cell state. The common center of the concentric circles represents the totipotent state while the subsequent inverted circles represent decreasing levels of cellular potency moving outwards. Here, direct reprogramming is analogous to 'direct conversion' or 'transdifferentiation', and refers to a change in cell fate that does not incorporate a pluripotent or progenitor state.

function remains an open question, though biologists have proposed theories since the mid-twentieth century. A popular model of unidirectional differentiation was postulated by Conrad Waddington in 1957 suggesting that stem cells can be thought of as a high-energy state in which all genes are active (Waddington 1957). As cells differentiate, genes unnecessary for the cell's subsequent role were thought to be silenced as the cell enters a lower-energy stable state, analogous to a ball rolling down a hill. While an insightful model for its time, this analogy of a ball rolling down a hill implies a unidirectional fatalistic process. Cellular reprogramming challenges this notion of cell state permanence, and allows a cell's fate to be redirected. This requires us to level the hierarchical epigenetic landscape (Ladewig et al. 2013) envisioned by Waddington and instead adopt a flattened landscape in which cells are free to shift from one cell potency and state to another, under the right conditions (Figure 1B-C). That is, one or more

core genetic elements may be manipulated in a way that forces a cell to take on a new form and function.

Typically, this cellular transformation is facilitated by an exogenous stimulus like the addition or removal of one or more transcription factors. Transcription factors (TFs) are DNA-binding proteins that modulate how often a gene is transcribed into messenger RNA (mRNA) for protein production. Of the roughly 20,000 known proteins in the genome, around 1,500 are known to function as TFs (Ignatieva et al. 2015). Therefore, focusing on TFs as target inputs significantly reduces the dimensionality of the problem. Notably, the exogenous expression of only one to three TFs is often sufficient to achieve the acquisition of specialized lineages from other cell types (Takahashi 2012).

Table 1: Glossary of Terms

Epigenetics	The study of heritable changes in gene activity caused by mechanisms other than changes in the underlying DNA sequence. These can include biochemical modifications of genomic DNA or histones, the proteins that help package genomic DNA
Differentiation	The unidirectional progression of a cell towards a specialized cell type
Reprogramming	The directed transformation of one cell type into a less or differently specialized cell type, typically mediated by exogenous TF expression
Cell Potency	The potential for a cell to develop into a more specialized cell type. The more lineages the cell can give rise to, the higher the potency and the closer it is to a stem cell-like state
Reprogramming Factor	One or more transcription factors capable of driving a cell from one state to another when exogenously overexpressed
Gene Regulatory Network	Cell-type specific set of interactions between genes, governed by the regulatory influence they exert on each other

2.2 Cellular Reprogramming

This notion of cellular reprogramming is far from a theoretical whim. Rather, it has been established over the last four decades in several contexts. Reprogramming has been achieved by introducing critical TFs to susceptible cell types and by transferring the nucleus of a somatic cell into an oocyte to "rejuvenate" the nucleus into a totipotent zygote.

2.2.1 Transcription Factor Mediated Reprogramming

In 1989, Harold Weintraub *et al.* demonstrated that non-muscle cells could be driven to express muscle-specific genes by *MYOD1* activation (Weintraub et al. 1989). This was the first success in directly reprogramming one differentiated cell to another and demonstrated that TFs such as *MYOD1* possess the capability to override expression programs. This inspired others to search for additional TFs capable of driving differentiation. In 2006, Shinya Yamanaka's group redefined what was possible when they demonstrated that murine adult fibroblasts could return to a pluripotent state following lentiviral transduction of just four TFs: *Oct3/4*, *Sox2*, *Klf4*, and *c-Myc* (Takahashi and Yamanaka 2006). These factors would be coined the "Yamanaka factors" and Yamanaka would be awarded the Nobel Prize in Physiology or Medicine in 2012 for this groundbreaking discovery. Not only were these cells transcriptionally similar to embryonic stem cells, but they successfully differentiated into all three germ layers - endoderm, ectoderm, and mesoderm - when introduced into murine blastocysts, indicating true pluripotency. The follow-

ing year, the Yamanaka factors were again used to convert human fibroblasts to iPSCs whose differentiation to neurons and cardiomyocytes could be subsequently induced, and that could form teratomas with all germ layers present when injected into nude mice (Takahashi, Tanabe, et al. 2007). These results indicate that cell identity remains malleable through adulthood, and when the correct TFs are applied, cells may be driven towards identities that meet experimental or therapeutic needs.

2.2.2 Somatic Cell Nuclear Transfer

Another route to reprogram cells is by transferring the nucleus of one cell type into the cytoplasm of another. The cell-signaling molecules in the recipient cytoplasm, including TFs, act to reprogram the donor nucleus to exhibit an expression profile more reminiscent of the recipient cell. The oocyte is an ideal recipient cell - its large cytoplasmic volume buffers it against the donor signaling molecules, as it contains sufficient quantities of TFs to overpower those contained in the donor nucleus. In addition, the oocyte's role in generating the embryo and resetting the sperm nucleus suggests that it contains factors necessary for nuclear reprogramming to a totipotent state. These ideas became reality when in 1952, it was demonstrated that a nucleus from a blastula cell may be introduced into an enucleated frog egg to give rise to a normal embryo (Briggs and King 1952). Following this discovery, John Gurdon demonstrated that even nuclei of intestinal epithelial cells were capable of giving rise to an entire embryo when implanted into an oocyte (Gurdon 1962). In 1996, the first mammalian clone was produced by transferring sheep mammary nuclei into an enucleated oocyte (WA and Wilmut 1996). Finally, in 2013, the first hESC cell line was produced using somatic cell nuclear transfer techniques (Tachibana et al. 2013).

These nuclear transfer experiments highlight a potential application to regenerative medicine, where a nucleus derived from a patient skin cell could be introduced into an oocyte for reprogramming into embryonic stem cells, which may in turn be differentiated into cell types of clinical interest. Since the generated cells would contain DNA of the graft recipient, this would provide an avenue for autologous tissue grafting without concern for rejection.

2.2.3 Translational Success

Recently, cellular reprogramming has made the journey from bench to bedside as efficiency and safety have improved. The therapeutic potential of autologous grafts derived from reprogrammed cells is broad. Easily accessible yet malleable cells such as dermal fibroblasts, for example, may be reprogrammed to various cell types for a range of downstream medical applications (Figure 2). One such example is the application of reprogrammed cells to treat age-related macular degeneration. As described in a 2017 report, retinal pigment epithelium (RPE) cells were derived from fibroblasts induced with non-integrating episomal vectors carrying TFs to treat a 77-year old patient with wet macular degeneration (Mandai et al. 2017). A skin sample was taken and reprogrammed to an iPSC state, and subsequently differentiated into RPE cells. These reprogrammed cells exhibited DNA methylation and gene expression patterns consistent with RPE cells. Additionally, genome sequencing indicated that no mutagenesis had occurred during the process, mitigating the possibility for oncogenesis. Since these cells are autologous and therefore carry the DNA of the patient, there is no risk of graft rejection upon re-introduction. The reprogrammed cells were surgically grafted into the patient's retina and found to have engrafted successfully after one year without evidence of vision loss. This early success demonstrates the clinical potential of cellular reprogramming using non-integrating TF delivery methods, as an autologous graft was generated from reprogrammed cells without introducing potentially harmful mutations.

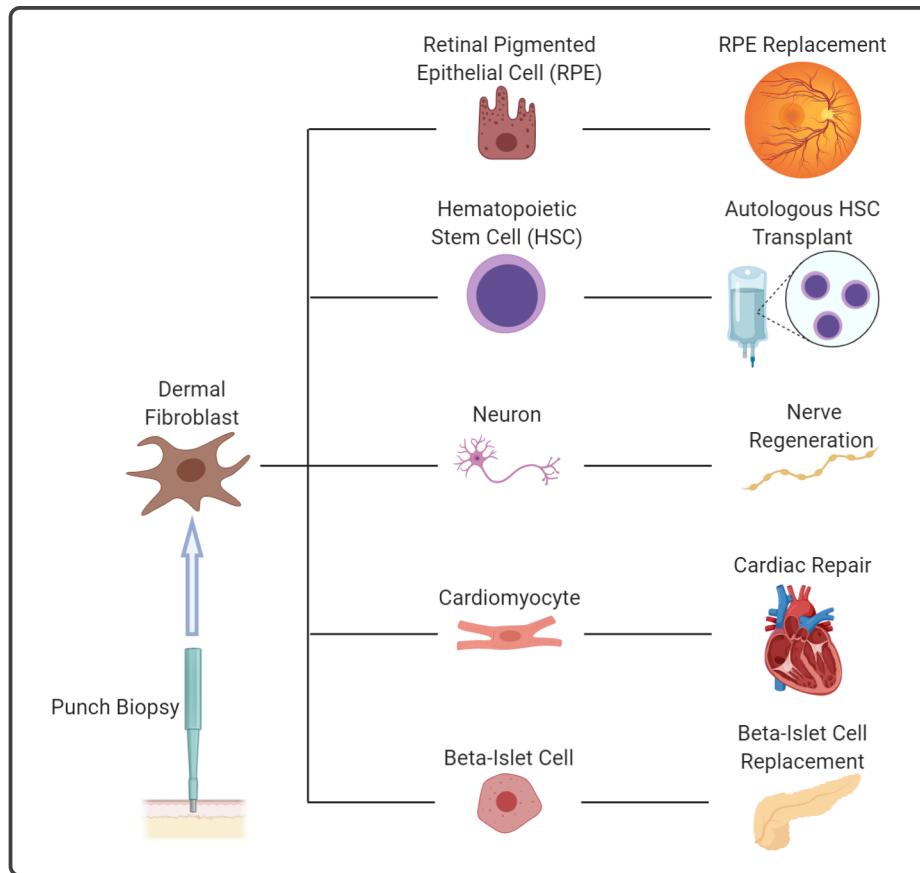


Figure 2: Medical applications of cellular reprogramming. Dermal fibroblasts can be acquired via a minimally invasive punch biopsy, and subsequently differentiated into a desired cell type via the addition of select TFs. Finally, autologous cultured cells of the desired type can be reintroduced into the patient without concern for graft rejection.

Clinically-relevant reprogramming successes in laboratory settings have presented exciting potential as well. For example, critical steps have been made towards autologous hematopoietic stem cell therapies for post-chemotherapy leukemia patients, as murine committed lymphoid and myeloid progenitors (Riddell et al. 2014) as well as human fibroblasts (Silvério-Alves et al. 2019) were converted into hemogenic cells. This success extends to the field of nerve regeneration therapy, as fibroblasts were reprogrammed into glutaminergic neurons (Yang et al. 2019). Yet another clinically-relevant success is the differentiation of human fibroblasts to cardiomyocytes (N. Cao et al. 2016; Fu et al. 2013), as the ability to cultivate autologous cardiomyocytes may allow replacement therapy following myocardial injury in the future. Finally, with type-I and late-stage type-II diabetes mellitus characterized by the loss of insulin-producing beta-islet cells, success in generating autologous beta cells has curative potential. The feasibility of this solution was recently demonstrated with the conversion of pancreatic acinar cells to beta-islet cells (Cavelti-Weder et al. 2015) as well as the conversion of fibroblasts to beta-islet-like cells via directed differentiation through endodermal intermediates (Zhu et al. 2016). Taken together,

these results indicate that an easily accessible and malleable cell type such as the fibroblast may be converted into a range of clinically applicable cell types for autologous grafting, potentially challenging the permanence of diseases ranging from macular degeneration to diabetes mellitus.

3 Mathematics of Cellular Reprogramming

Cellular reprogramming is a very calculated process involving inputs, transition states, and outputs - a control theory problem at its core. Long before control theory concepts would be implemented in cellular reprogramming studies, however, mathematicians were already conceptualizing how biological systems could be systematically controlled (Liu 2011). From observing state-dependent network entropy during differentiation (Rajapakse et al. 2011) to modeling cells as complex networks subject to perturbations that can modulate the system's equilibrium and placement in an n -dimensional state space (Cornelius et al. 2013), it became clear that mathematics could be exploited to optimize cellular reprogramming.

3.1 Cellular States

During cellular reprogramming, cells encounter a start state and end state, occupying one or more transition states in between. These states are most commonly described in terms of their gene expression profiles or gene regulatory networks (Cahan et al. 2014; D'Alessio et al. 2015; Rackham et al. 2016; Del Vecchio et al. 2017; Ronquist et al. 2017). Derived from RNA-sequencing data, gene expression can be represented as a vector of n non-negative values, where n is the roughly 20,000 genes in the human genome. A single expression vector represents a state-dependent snapshot of a genome's transcriptional landscape in time. Collecting data as the cell evolves can then give rise to a sequence of expression profiles reflecting discrete points along a reprogramming trajectory. While gene expression has provided sufficient cellular state representation thus far, additional measurements such as those from ChIP-seq, DNase-seq, or chromosome conformation data can contribute to a more comprehensive definition of cellular states and may be considered in future studies (Trapnell 2015).

Entropy plays an important role in cellular dynamics and offers a quantitative measure of a cell's differentiation status (i.e., a cell's placement within the Waddington landscape). One frame of thought is that the specialization of cells minimizes entropy and thus controllability (Rajapakse et al. 2011; Rajapakse et al. 2012). The closer a cell is to a pluripotent state, the higher its entropy - as there exists a wider availability of signaling pathways that can be activated and conformations the cell can adopt (more uncertainty) in the lesser committed state (MacArthur and Lemischka 2013; Cover 1999). We can glean local and global entropy from data on gene expression and protein networks (Banerji et al. 2013). Local entropy indicates the susceptibility of specific signaling pathways in the cell to perturbation and is derived simply from the Shannon entropy for a particular gene or protein. Global entropy indicates how specialized the cell is and can be captured by finding the average of those local entropies.

3.2 Cellular State Perturbations

Perturbations push cells to overcome the epigenetic barriers that consign them to a single differentiated state (MacArthur, Ma'ayan, et al. 2009). In the context of cellular reprogramming, these perturbations are often the external introduction of one or more TFs that evoke changes in transcriptional activation and inhibition throughout the cell's gene regulatory network. The internal response to these external perturbations, however, is not fully deterministic. That is, perturbations may drive the cell towards the vicinity of a target state (i.e., a basin of attraction)

with the expectation that the cell will converge to the intended target state (attractor) in a stochastic manner.

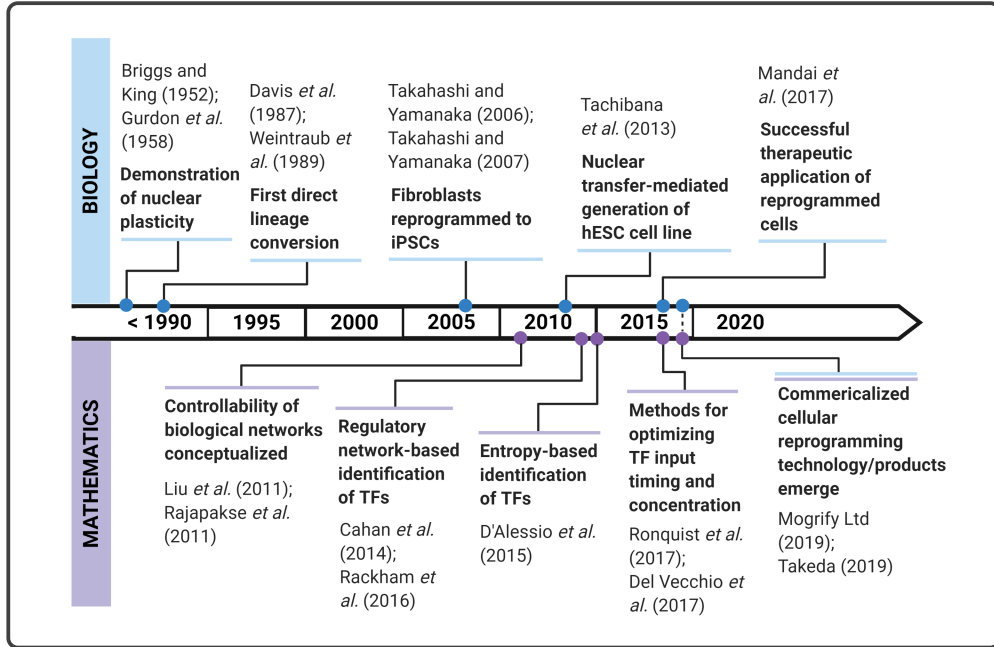


Figure 3: Timeline of key experimental and computational cellular reprogramming advancements.

In their work on perturbation patterns and network topology, Marc Santolini and Albert-László Barabási demonstrate how topological models can offer substantial insight into accurately predicting the influence of perturbations on biological networks (Santolini and Barabási 2018). Santolini and Barabási explain that the effects of external stimuli do not diffuse throughout the genome uniformly. Rather, perturbation spread is confounded by interaction properties like directedness, modularity, or whether an interaction is activating or repressive. As such, every node in a biological network will correspond to a differential equation with parameters encompassing these properties. Network topologies can then be captured from a set of differential equations by constructing the Jacobian matrix, a matrix of partial derivatives that describes the impact of node i on the activity of node j and that quantifies direct interactions between all pairwise nodes in the network. An adjacency matrix, describing the influence of the network, can subsequently be derived as $\mathbf{A} = \text{sign}(\mathbf{J}^T(\mathbf{x}^*))$, where the sign function represents the directionality of the edges between nodes and is applied element-wise, \mathbf{J}^T is the transpose of the Jacobian, and \mathbf{x}^* is a vector of magnitudes (i.e. gene expression) representing the steady state of the system. To account for indirect interactions in a network, Santolini and Barabási also describe a sensitivity matrix, \mathbf{S} , where each element is the full derivative between two nodes. The relationship between the sensitivity and Jacobian matrices becomes:

$$\mathbf{S} = (\mathbf{I} - \mathbf{J})^{-1} \mathbf{D} (\mathbf{I} / (\mathbf{I} - \mathbf{J})^{-1}),$$

where, \mathbf{I} is the identity matrix, \mathbf{D} is the diagonal operator, and the $/$ operator denotes element-wise division. By modifying (perturbing) the differential equation of nodes of interest (TFs) and

analytically deriving the sensitivity matrix, we get a sense of how that node's influence propagates throughout the network. Taken together, this predicative framework facilitates evaluation of how cells might react to a perturbation, providing an avenue for which to test the influence of candidate reprogramming factors.

4 Computational Tools

The earliest cellular reprogramming studies incorporated a systematic "guess-and-check" method for identifying TFs that would be necessary and sufficient for reprogramming in their experimental system. The four Yamanaka factors were narrowed down from an initial list of 24 candidate reprogramming factors that were selected based on their properties and suspected involvement in promoting and maintaining an embryonic stem cell-like state. It took several trials of transducing different permutations of these factors to finally converge upon the four factors capable of inducing pluripotency in differentiated cells. Several computational frameworks for streamlined data-guided prediction of reprogramming factors have since been developed (Table 2), demonstrating that we can bypass the "guess-and-check" method of validating TFs, which tends to be time consuming and cost prohibitive.

4.1 CellNet

The successful realization of a target cell identity following reprogramming can be assessed by establishing transcriptional similarity using gene expression profiling and hierarchical clustering analyses. This validation approach, however, fails to capture functional dissimilarities that may exist between the reprogrammed cell and its naturally occurring counterpart. A more quantitative and rigorous system is therefore required to verify the integrity of reprogrammed cells. To this end, in 2014, Patrick Cahan and co-author Samantha Morris *et al.* introduced CellNet, a computational tool employing network biology to assess and improve the fidelity of reprogrammed cells (Cahan *et al.* 2014). In a companion paper, Morris and Cahan put CellNet to the test with B cell to macrophage and fibroblast to hepatocyte-like cell conversions, experimentally verifying that original reprogramming schemes could not achieve full conversion but that CellNet-refined schemes could (Morris, Cahan, *et al.* 2014).

As described in the original publication, CellNet first generates a training dataset containing cell- and tissue- (C/T) specific GRNs derived from a large array of publicly sourced microarray datasets. The CellNet algorithm has since been extended for compatibility with RNA-sequencing data (Radley *et al.* 2017). Subsequently, based on input gene expression data from a reprogramming experiment, CellNet classifies the query cell by its most likely C/T-specific GRN and evaluates how close it is to the target phenotype's GRN. If the two GRNs deviate significantly, CellNet proposes TFs that could further drive the reprogrammed cell to better mimic its target identity. Therefore, CellNet's measure of reprogramming success is the extent to which a reprogrammed cell has established its target state's GRN. To evaluate the reprogrammed cell's GRN in comparison to the target cell type's GRN, the CellNet algorithm computes a GRN status score as follows:

$$\text{RGS} = \sum_{i=1}^n \text{Zscore}(\text{gene}_i)_{C/T} \times (\text{gene weight}_i)_{\text{GRN}},$$

where, n is the number of genes in the reprogrammed cell's GRN, the z-score of gene i is derived from the distribution of expression values for gene i in the C/T-specific training data, and the gene weight is the expression of gene i as determined from the reprogrammed cell's expression profile. If the reprogrammed and target GRNs are not equivalent, CellNet then investigates

how transcriptional regulators (TRs) of the target GRN can push the reprogrammed cell to fully convert. This is determined by computing a network influence score (NIS) for TRs (i.e., TFs) in the GRN:

$$\text{NIS}(\text{TR}) = \sum_{i=1}^n (\text{Zscore}(\text{target}_i)_{\text{C/T}} \times \text{weight}_{\text{target}_i}) + n \times \text{Zscore}(\text{TR})_{\text{C/T}} \times \text{weight}_{\text{TR}},$$

where, n is the number of genes in the target C/T-specific GRN, the z-score of the target is derived from the distribution of expression values for gene i in the training data, $\text{weight}_{\text{target}_i}$ is the mean expression value of gene i in the training data, the z-score of the given TR is based on the distribution of expression values for the TR in the training data, and $\text{weight}_{\text{TR}}$ is the mean expression value of the TR in the training data.

With CellNet, Cahan *et al.* uncover variations in reprogrammed cells' regulatory networks compared to the target state thereby highlighting inherent limitations in specificity that accompany current reprogramming methods. This technique is unique in that it seeks to address both how well a reprogrammed cell recapitulates its target and how the administered combination of TFs can be revised to better recapitulate that target - what Cahan later designates the 'assessment' and 'improvement' problems, respectively (Cahan 2016).

4.2 D'Alessio et al. Method

In contrast to CellNet, a later method for identifying reprogramming factors described by D'Alessio *et al.* in 2015 incorporates only the expression of TFs instead of genome-wide expression data in its prediction algorithm. (D'Alessio et al. 2015). Specifically, TFs are ranked by their specificity and uniqueness, where a given TF is highly ranked and ideal for reprogramming if it is highly expressed in the query (i.e., target cell type) dataset and not expressed in cell types comprising the background dataset. To identify such TFs, D'Alessio *et al.* define the distribution of a TF's expression level across cell types, such that the expression of a TF in a query cell type may be compared to the expression of that TF in all other cell types. The cell type specificity of each TF is evaluated using an entropy-based score adapted from previously described applications of Jensen-Shannon (JS) divergence for describing tissue-specific gene expression (Trapnell et al. 2010) and lincRNA expression (Cabili et al. 2011). Specifically, a distance metric describing the difference between a query expression pattern and the ideal expression pattern is obtained by taking the square root of the JS divergence as follows:

$$\text{JS}_{\text{dist}}(\mathbf{e}, \mathbf{e}^I) = \left[\mathbf{H}\left(\frac{\mathbf{e} + \mathbf{e}^I}{2}\right) - \frac{\mathbf{H}(\mathbf{e}) + \mathbf{H}(\mathbf{e}^I)}{2} \right]^{1/2},$$

where, \mathbf{e} is an n -dimensional vector, $\mathbf{e} = [e_1, e_2, \dots, e_n]$, consisting of the expression-derived abundance density of a specific TF across n cell types. Vector \mathbf{e}^I represents the ideal distribution in the form of a binary n -dimensional vector where $e_i^I = 1$ if $i = \text{query cell type}$ and 0 for all other cell types. Lastly, $\mathbf{H}(\cdot)$ represents the Shannon entropy. TFs with cross-cell type expression distributions closely aligned to the ideal distribution, (i.e., expressed in the query cell type but not in other cell types) will yield JS distances close to zero. D'Alessio *et al.* identified the top ten TFs with expression profiles most similar to the ideal expression profile for nearly 200 cell type and tissue pairs using this approach. Experimental validation of TFs predicted to define RPE cells demonstrated successful reprogramming from fibroblasts into RPE cells. This success lends credence to the identification of reprogramming factors based on the influence that TFs have on the identity of a desired cell type, as determined by the level of unique and relative expression.

4.3 Mogrify

Similar to CellNet, Mogrify leverages regulatory networks to predict TFs for direct conversion of cells, selecting TFs that exert regulatory influence on genes that identify most with the target cell type (Rackham et al. 2016). Using publicly available gene expression data, the Mogrify algorithm starts by accumulating TFs that are differentially expressed between the target cell type and a background set of non-target cell types. A tree-based approach is employed to ensure that the background set is not saturated with cell types that are too highly or distantly related to the target cell type. Selected TFs are subsequently ranked based on the combination of their differential expression (scored as the product between the log-transformed fold change and adjusted p-value) and local network influence based on protein-DNA and protein-protein interaction data. This network influence is derived as follows:

$$N_x = \sum_{r \in V_x} G_r \times \frac{1}{L_r} \times \frac{1}{O_r},$$

where, $r \in V_x$ are genes (r) that are nodes (V_x) in the local subnetwork of TF x , G_r is the aforementioned differential expression-based score for gene r , L_r is the number of edges between gene r and TF x in the local subnetwork, and O_r is the out-degree of gene r 's parent node in the network. This score is determined for the network of known gene protein product interactions and separately for the network of known TF-DNA interactions. With the described distance-based and connectivity-based weightings, Mogrify carefully selects for TFs with high influence and regulatory specificity. The subsequent combination of gene and network scores yields a set of TFs that are highly ranked if they are non-ubiquitous with a high fold change and low p-value. To further prune the list, TFs that regulate over 98% of the same genes are deemed redundant and the higher ranked TF preserved. Additionally, TFs in the ranked list that are already highly expressed in the starting cell type are removed from consideration.

While Mogrify's systematic identification of reprogramming factors is only as good as the fidelity of publicly available gene expression data and breadth of known TF binding interactions, it offers robust predictive power in mediating direct cell conversion, having recovered 84% of previously discovered conversion factor cocktails and validated two novel conversions at its time of publication. Mogrify-enabled predictions for TF-mediated conversions between cell types represented in the FANTOM5 gene expression atlas have been compiled into an online directory (<http://mogrify.net/>), providing a convenient one-stop shop for a wide range of direct reprogramming objectives (Ouyang et al. 2019). Additionally, since its inception, Mogrify has undergone its own transformation from technology to corporation, now leading the way in commercialized direct human cell conversion and revolutionizing cell therapy.

4.4 Del Vecchio et al. Method

The natural dynamics of a cell's GRN are not always amenable to preset overexpression of TFs (Morris and Daley 2013; Schlaeger et al. 2015; Goh et al. 2013; Xu et al. 2015). Del Vecchio *et al.* address this uncertainty in reprogramming success by employing mathematical modeling to facilitate the design of a genetic feedback controller, independent of GRN dynamics, that iteratively adjusts TF input concentrations based on the difference between current and target state TF concentrations (Del Vecchio et al. 2017). In this method, a cellular GRN (\mathbf{x}) with n TFs, $\mathbf{x} = [x_1, \dots, x_n]$ is modeled as a set of ordinary differential equations represented as:

$$\frac{dx_i}{dt} = H_i(\mathbf{x}) - \gamma_i x_i + u_i \text{ for } i \in \{1, \dots, n\},$$

where $H_i(\mathbf{x})$ is the Hill function capturing the GRN regulation of TF x_i , γ_i is the constant decay rate of TF x_i , and u_i is a non-negative scalar indicating the input perturbation corresponding to TF x_i . In contrast to open loop control which takes a constant or time-varying TF concentration as input, the authors employ closed loop feedback control which updates the input concentration for a system in real time. Since the feedback control is dependent on the error between the most current concentration, x_i , and the target concentration, x_i^* , the u_i term in the former equation can be replaced with the expression $G_i(x_i^* - x_i)$, where G_i is a large positive constant representing gain that renders H_i and $\gamma_i x_i$ negligible. The resulting form describes the simultaneous production and degradation of TF x_i mRNA that govern the behavior of Del Vecchio *et al.*'s novel synthetic genetic controller circuit. Operation of this circuit is facilitated by two inducers, one acting on a synthetic copy of gene x_i and the other on siRNA complementary to both the endogenous and synthetic mRNA, where the inducer concentrations are informed by x_i^* . Implemented for each TF in the GRN, or a predetermined subset, the circuit steers the concentration of a given TF x_i and its mRNA, m_i , towards x_i^* and m_i^* by inducing the concurrent production of the synthetic TF and degradation of the endogenous TF. Once the desired concentration is achieved, the inducers are set to zero to stop the synthetic controller and the endogenous system takes over TF production to maintain the stability of the reprogrammed state.

At its time of publication, the proposed approach had been simulated on a two-node model of the pluripotency network but not yet demonstrated on a real cellular GRN. The authors suggested, however, that the application of their controller may ensure more success in iPSC reprogramming as pluripotent states can be difficult to attain with preset TF overexpression due to the potential for the TF to dominate the network's behavior. Though highly theoretical at this time, this work offers a promising and well thought out synthetic biology approach for fine-tuning cellular reprogramming.

4.5 Data-Guided Control (DGC)

Coupling gene expression and TF binding data with control theory, Ronquist *et al.*'s universal algorithm for cellular reprogramming offers a slightly more mathematically rigorous framework than its predecessors for data-guided prediction of reprogramming factors, or data-guided control (DGC) (Ronquist *et al.* 2017). In addition to identifying candidate reprogramming factors, Ronquist *et al.* also present an optimization method for identifying the ideal time during the cell cycle for addition of TFs. The foundation of this algorithm is the following linear, control theory difference equation:

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B} \mathbf{u}_k$$

where $\mathbf{x}_k \in \mathbb{R}^N$ is a gene expression vector representing the cell state at time k . To ease the computational burden, the authors reduce the dimensions of this vector from $\sim 20,000$ to $\sim 2,000$ ($\mathbb{R}^{\tilde{N}}$) by summing the expression levels of genes that are in close proximity to one another, or more specifically, that occupy the same cell-type invariant topological domains. $\mathbf{A}_k \in \mathbb{R}^{\tilde{N} \times \tilde{N}}$ is the transition state matrix representing time-varying changes to the cellular state at time k . $\mathbf{B} \in \mathbb{R}^{\tilde{N} \times M}$ is the input matrix indicating interactions between genomic domains (\tilde{N}) and TFs (M). Elements of the matrix are weighted by the magnitude of regulatory influence, determined from publicly available TF binding site data, and signed according to whether a given TF activates or represses the genes in each domain. Finally, $\mathbf{u}_k \in \mathbb{R}^M$ is a binarized input

vector, with nonzero elements indicating which TF(s) to introduce at time k . An overview of this framework is reproduced in Figure 4.

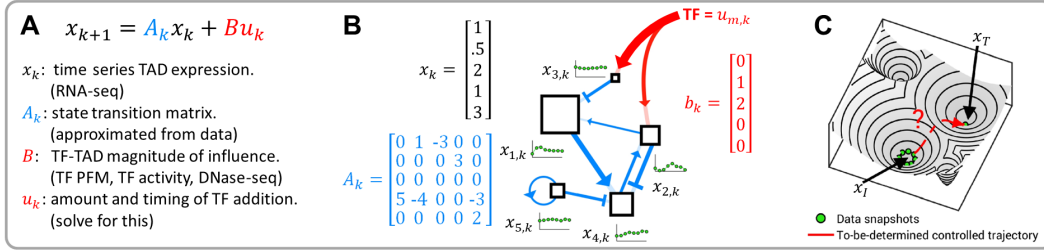


Figure 4: Data-guided control overview. (A) Summary of control equation variables. (B) Each box represents a topological domain containing several genes. The blue connections represent the edges of the network and are determined from time series RNA-seq data. The small green plots at each node represent the expression of each domain changing over time. The red arrows indicate additional regulation imposed by exogenous TFs. (C) Conceptual illustration of determining TFs to push a cell state from one basin to another. Figure reproduced with permission from Ronquist *et al.* 2017.

Once the solution to the difference equation is obtained, TFs are subsequently scored by executing the following optimization problem for all possible input signals:

$$\begin{aligned} & \text{minimize}_{\mathbf{u}} \|\mathbf{x}_T - \mathbf{z}_F(\mathbf{u})\| \\ & \text{subject to } \begin{cases} \mathbf{u}_{m,k}, & k = 1, 2, \dots, F \\ \mathbf{u}_{m,k}, & \text{if } m \notin \hat{p} \\ \mathbf{u}_{m,k+1} \geq \mathbf{u}_{m,k} \end{cases}, \end{aligned}$$

where, the objective is to minimize the distance between the final state, \mathbf{z}_F , and target state, \mathbf{x}_T , with $\mathbf{z}_F(\mathbf{u})$ denoting that the final state is dependent on the input signal, \mathbf{u} . We can evaluate this distance as the Euclidean norm ($\|\cdot\|$) of the difference between the gene expression levels at each state. The first constraint addresses the restriction that TFs only be added to the cell and not removed. The second constraint requires all elements of input signal \mathbf{u}_k to be zero if they do not correspond to the subset of TFs (\hat{p}) selected to drive the system. Lastly, the third constraint indicates that select TFs can be added consecutively or added sequentially, but that once added, the TFs must continue to be exogenously expressed until the final time point. Notably, Ronquist *et al.* demonstrate that TF scores are dependent on the time of input, with some TFs showing preference for addition towards the beginning of the cell cycle and others towards the end. Overall, Ronquist *et al.* present a convincing case for incorporating time-varying data into TF prediction frameworks and considering the time of TF input to a system. This approach mimics the natural course of differentiation and may increase the efficiency of reprogramming to achieve yields sufficient for tissue grafting, though experimental validation is needed.

Table 2: Summary of Computational Tools for Reprogramming

Method	Input ¹	Output	Approach	Validation
CellNet (Cahan et al. 2014)	Gene expression data and GRN information	Assessment of original reprogramming scheme and proposal for refined reprogramming scheme	Scores TFs based on GRN status and network influence	Experimentally validated
D'Alessio et al. Method (D'Alessio et al. 2015)	Gene expression data for known TFs	Set of core TFs specific to target cell type	Scores TFs by expression specificity to target cell type based on entropy-based measure	Experimentally validated
Mogrify (Rackham et al. 2016)	Gene expression data	Set of candidate TFs for conversion to target cell type	Ranks TFs based on influence over differentially expressed genes in the target cell type	Comparison to previously confirmed conversion cocktails and novel conversions were experimentally validated
Del Vecchio et al. Method (Del Vecchio et al. 2017)	TFs in target GRN	Desired endogenous TF concentrations for target cell type	Synthetic genetic feedback controller that steers TF concentrations based on discrepancy between actual and desired TF concentrations	Simulation on model of a cellular network
Data-Guided Control (DGC) (Ronquist et al. 2017)	Time-series gene expression and TF binding data	Set of candidate TFs for conversion to target cell type and suggested time of input	Genome dynamics for initial and target cell states modelled using control theory difference equation and TFs scored based on Euclidean distance between states under TF control	Comparison to previously confirmed conversion cocktails

¹ Data necessary to facilitate prediction of reprogramming factors. Input of initial and target cell types is assumed.

5 Experimental Realization

With the guidance of predictive models such as those mentioned above, we can narrow down the list of candidate TFs to quantities achievable with modern high-throughput methods. Rather than considering any combination of the approximate 1,500 TFs in the human genome, the problem is reduced to only those with favorable predictions. Now that the problem is of a more manageable size, experimental validation is critical.

5.1 Transcription Factor Delivery

In order to experimentally achieve reprogramming, it is necessary to determine a method of adding TFs to cells. TFs are proteins coded by genes, and thus could be introduced to the cell directly or via DNA or mRNA coding for that protein (Figure 5).

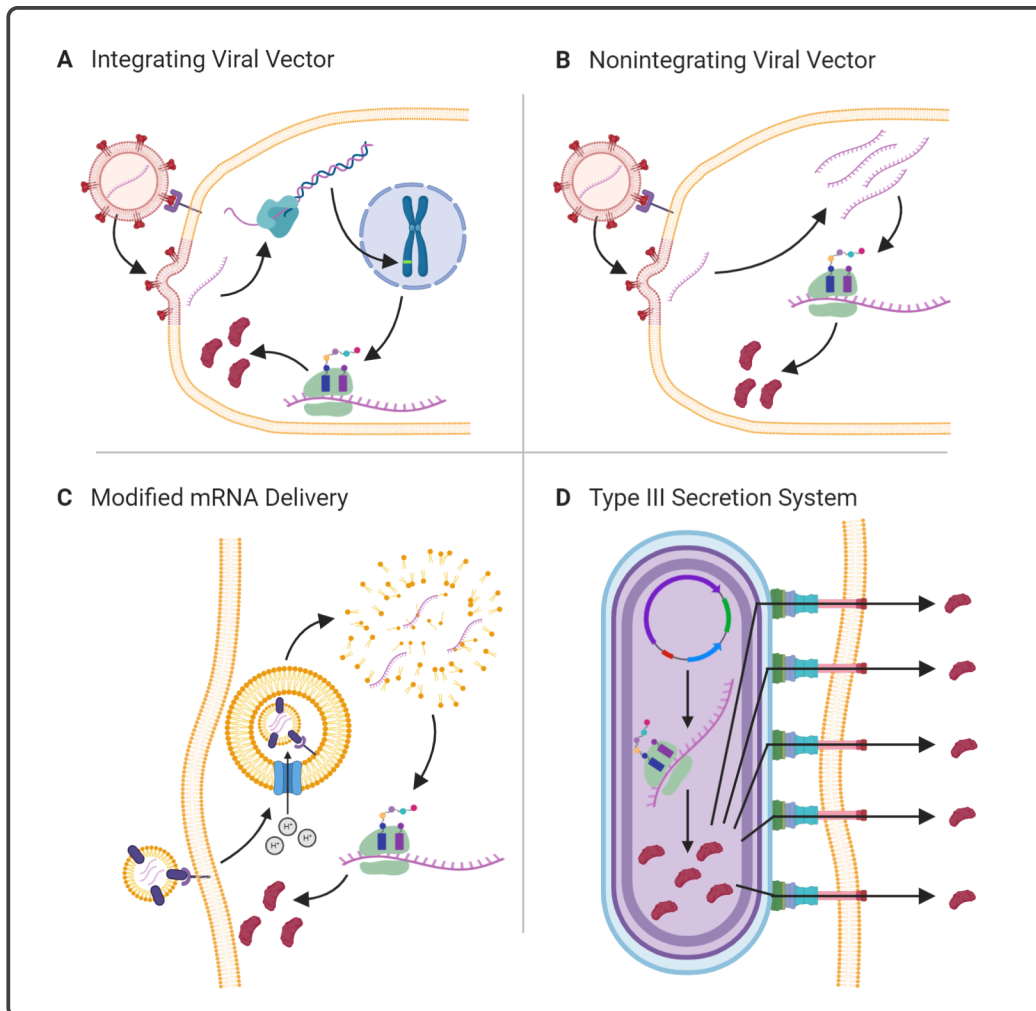


Figure 5: Transcription factor delivery. (A) Transduction of a TF-encoding gene by a DNA integrating virus such as a lentivirus. The virus first binds to the host membrane and fuses its viral envelope with the eukaryotic phospholipid membrane to enter the cell. Nucleic acids are released in the cytosol, where RNA-dependent DNA polymerase creates double stranded DNA. Viral DNA is integrated into the host genome. Host cells then express virally-delivered genes, which are translated into functional proteins. (B) Introduction of mRNA coding for a TF by a non-integrating virus such as a Sendai virus. As above, the virus docks, fuses with the cell membrane, and releases nucleic acids into the cytosol. RNA-dependent RNA polymerase then generates positive sense RNA, which is translated to functional protein. (C) Delivery of modified mRNA via a lipid nanoparticle. First, the lipid particle-embedded peptides bind target cell receptors and trigger receptor-mediated endocytosis. Next, the endosome is acidified by H⁺ pumps. The acidified endosome and lipid nanoparticle are destabilized, releasing mRNA into the cytosol. Free mRNA is translated to functional protein. (D) Bacterial type-III secretion system as a TF delivery mechanism. First, a bacterial DNA plasmid is expressed and protein is produced. Next, the bacterial T3SS delivers the protein to the eukaryotic cell through a molecular needle.

5.1.1 Viral Vectors

Perhaps the most common modality by which biologists introduce genes into cells is lentiviral transduction. Lentiviruses are a class of retroviruses commonly used in experimental procedures. Retroviruses, such as the lentivirus, are made up of an RNA-based genome, RNA-dependent DNA polymerase (reverse transcriptase), and DNA integrase surrounded by a protein capsid and a phospholipid envelope. These viruses function by fusing with the membrane of cells and releasing their RNA into the cytoplasm, which is reverse transcribed into DNA and integrated into the host genome (Ryu 2016). These viruses serve as a useful tool, as one can package a gene of interest into their RNA genome and use these viruses to integrate that gene into the genome of a cell in culture. In addition, antibiotic selection genes and fluorescent proteins can be added to aid in selecting cells that have taken up and readily express the viral genome. Despite these advantages, insertional mutagenesis may confound experimental results and increase the risk of tumorigenesis in therapeutic applications (Uren et al. 2005). To circumvent this risk, non-integrating viruses such as the Sendai virus can be used to temporarily express exogenous TFs and initialize reprogramming processes within the cell. While this solves one problem, both retroviruses and non-integrating viruses trigger host cell innate immune responses as viral RNA is sensed by pattern recognition receptors (PRRs). This leads to a wide range of cell signaling events and transcriptional changes that alter cell behavior, potentially interfering with experimental and therapeutic applications (Said et al. 2018).

5.1.2 Modified mRNA

An alternative to viral vectors is direct application of TF-encoding mRNA packaged in lipid nanoparticles (LNPs). These LNPs are taken up via receptor-mediated endocytosis, and nucleic acids are released into the cytoplasm as acidification of the endosome leads to dissolution of the lipid nano-particle and endosome (Reichmuth et al. 2016). An advantage of liposomal delivery of synthetic mRNA molecules over viral transduction is the avoidance of the host cell immune response, as synthetic mRNAs can be designed to mimic host mRNA and avoid PRR binding (Linares-Fernández et al. 2020). In addition, similar to non-integrating viruses, modified mRNA does not cause changes to the host genome and therefore is not associated with risk of mutagenesis of oncogenes or tumor suppressor genes. Much progress over the last decade has led to efficient protocols to generate iPSCs from differentiated cells by TF introduction via modified mRNA (Warren and C. Lin 2019). Nanoparticle-delivery of mRNA has also been explored as a COVID-19 vaccine candidate in human trials, highlighting the safety and stability of LNPs *in vivo* (Jackson et al. 2020).

5.1.3 Bacterial Type-III Secretion System

A disadvantage of mRNA-based gene introduction is the lag time between initiation of translation and accumulation of functional protein due to the time it takes for a cell to translate delivered mRNA molecules into proteins of interest. In addition, the dependence of the rate of translation on cellular properties may further confound experiments, as the number of available ribosomes and tRNA substrate is not universal (Ross and Orłowski 1982). As suggested in the Del Vecchio *et al.* and Ronquist *et al.* methods above, cellular reprogramming experiments may benefit from introducing high concentrations of TFs at a precise point in time which presents a challenge for mRNA-delivery methods considering the time needed for translation. An emerging technique to introduce proteins directly into cells takes advantage of the bacterial type-III secretion system (T3SS). The bacterial T3SS functions as a molecular needle, puncturing the membrane of eukaryotic cells and injecting proteins that carry an N-terminal secretion signal. Originally a mechanism of toxin delivery by pathogens such as *Salmonella*, *Shigella*, and *Yersinia* species, this system has been commandeered for use in various peptide delivery experiments. This system has

several advantages over nucleic acid-based approaches. First, the work of protein production is outsourced to bacterial cells, which generate and deliver the protein of interest to the eukaryotic cell at a rate that is independent of recipient cell state. In addition, the host response to foreign nucleic acids is avoided, as proteins are delivered directly without a nucleic acid intermediate. Finally, no changes to the genome are made, mitigating the risk of mutagenesis seen with lentiviral delivery methods. This system therefore has the potential to deliver TFs to cells without permanent genetic aberrations while minimizing confounding cell behavior caused by foreign nucleic acids.

Bacterial T3SS delivery methods have recently been employed in several contexts with promising results. A T3SS expressing strain of *Pseudomonas aeruginosa* successfully delivered *MYOD* to mouse embryonic fibroblasts (MEFs) to reprogram MEFs to myotubes (Bichsel et al. 2013), and facilitated differentiation of human ESCs and iPSCs into cardiomyocytes by sequential addition of five TFs (Jin et al. 2018). In addition to the aforementioned advantages, the versatility of bacteria allows for engineering of the system. Recently, optogenetic interaction control switches were added to the *Yersinia enterocolitica* T3SS such that protein injection occurred in a light-dependent manner, further increasing the ability to control the delivery of peptides with respect to space and time (Lindner et al. 2020). The T3SS therefore represents a novel path forward in TF delivery for cellular reprogramming, both experimentally and therapeutically.

6 Future Directions

6.1 Biology

Incredible progress has been made in the development of cellular reprogramming regimes that give rise to a broad spectrum of cell types across all germ layers. However, overcoming low cell fate conversion efficiencies is a primary hurdle facing the future of cellular reprogramming (Aydin and Mazzoni 2019; Grath and Dai 2019; Horisawa and A. Suzuki 2020). While direct reprogramming reduces some of the common side-effects of iPSC reprogramming like mutagenesis and tumorigenesis, full conversion to an intended cell identity is still not guaranteed. It is common for reprogramming perturbations to drive cells to a stable yet hybrid state where some of the starting cell's transcriptional program is retained and the target identity is only partially acquired. Such failures to fully convert have been attributed to the presence of TFs that maintain the starting cell's gene regulatory network, lineage-specific repressors, and inaccessible chromatin among other factors. More rigorous reprogramming regimens involving the silencing of endogenous genes via CRISPR/Cas9 (C. Wang et al. 2017), endogenous gene activation using CRISPR/Cas9-based methods (Liao et al. 2017; H. Huang et al. 2020), and cocktails of chemical compounds with or without TFs (S. Cao et al. 2017; Gao et al. 2017; Herdy et al. 2019) have already shown considerable success in combating these roadblocks. Recent achievements in direct reprogramming spanning these and other conversion methods are summarized in Table 3. Further improvements in determining effective combinations of source cell types, reprogramming mechanisms, delivery methods, and cultivation conditions will be necessary to achieve consistent large-scale production of reprogrammed cells.

Stable cultivation and expansion of reprogrammed cells remains a challenge (K. G. Chen et al. 2014), limiting sufficient quantities of cells to reconstitute tissues and organs. Once methods of reprogrammed cell proliferation are improved, it is theoretically possible to regenerate organs derived from a patient's somatic cells for autologous transplantation (D. Zhang and W. Jiang 2015). Since these organs would be recipient-derived, there would be little concern for trans-

plant rejection. Moreover, lifelong immunosuppression, which is currently a significant source of mortality among transplant recipients, would no longer be necessary.

Table 3: Recent Successes in Direct Reprogramming

Starting Cell Fate	Target Cell Fate	Reprogramming Factors	Delivery Method	Reference
Dermal fibroblasts	Adipocyte-like cells	PPAR γ 2	Lentivirus	J.-H. Chen et al. 2017
Dermal fibroblasts	Endothelial progenitor cells	ETV2	mRNA transfection	Van Pham et al. 2017
Embryonic fibroblasts	Antigen-presenting dendritic cells	PU.1, IRF8, and BATF3	Lentivirus	Rosa et al. 2018
Glioblastoma cells	Neuronal cells	Forskolin, ISX9, CHIR99021, I-BET 151, and DAPT	Small molecule cocktail treatment	C. Lee et al. 2018
Embryonic fibroblasts	Hepatocytes	GATA4, FOXA2, HHEX, HNF4A, HNF6A, MYC, and P53-siRNA	Lentivirus	B. Xie et al. 2019
Bone marrow-derived cells, Fibroblasts, and Keratinocytes	Neural precursor cells	MSI1, NGN2, and MBD2	Plasmid transfection	Ahlfors et al. 2019
Dermal fibroblasts	Cardiomyocyte-like cells	GATA4, MEF2C, and TBX5	Nanoparticles	Kim et al. 2020
Foreskin fibroblasts	Cardiac progenitor cells	GATA4, HAND2, MEF2C, TBX5, and MEIS1	CRRISPR-cas9 based endogenous gene activation	J. Wang et al. 2020

6.2 Mathematics

Looking forward, a vast arena of mathematical principles with the potential to further sufficient understanding and facilitation of cellular reprogramming still remains. The high dimensional nature of cellular states, for example, invites an opportunity to examine cellular data in a tensor state space (C. Chen, Surana, et al. [2019](#)). Tensors are multidimensional arrays generalized from vectors and matrices, and have wide applications in many domains such as social sciences, biology, applied mechanics, machine learning and signal processing (Cichocki et al. [2016](#); Lu et al. [2008](#); W. Wang et al. [2019](#); Williams et al. [2018](#)). Classical linear control systems, as used in Ronquist *et al.*'s work, often fail to fully capture the dynamics of cellular reprogramming because state vectors only represent gene expression, neglecting structural information. Chen *et al.* generalized the classical systems notion of controllability into multilinear systems in which the states, inputs, and outputs are preserved as tensors (C. Chen, Surana, et al. [2019](#)). Multilinear control systems can significantly relieve the difficulty of describing genome-wide structure and gene expression simultaneously, and will be beneficial in analyzing the dynamics of cellular reprogramming more comprehensively. Further, Chen *et al.* exploited the notion of hypergraphs in modeling the network dynamics of cellular reprogramming (C. Chen and Rajapakse [2020](#)). A hypergraph is a generalization of a graph in which its edges can join any number of nodes. The notion of transcription factories supports the existence of multiway interactions involving multiple genomic loci (Cook and Marenduzzo [2018](#)), which implies that the human genome configuration can be more accurately captured by hypergraphs. Chen *et al.* developed the notions of entropy and controllability for hypergraphs, which will be potentially advantageous in investigating network dynamics of cellular reprogramming (e.g. detecting cell identities transition

points and identifying minimum TF inputs) (C. Chen and Rajapakse 2020; C. Chen, Surana, et al. 2020). Ultimately, cellular reprogramming would be required to account for nonlinearity or nonlinear control in the multiway dynamical system representation and analysis framework, and is an important direction for future research.

6.3 Medicine

Reprogrammed cells have many translational applications, from their direct use in replacement therapy to use as experimental models of disease and pharmacologic screens (Srivastava and DeWitt 2016). Many human diseases are characterized by the loss of a functioning tissue, such as the loss of hematopoietic stem cells in aplastic anemia or insulin-secreting beta-islet cells in diabetes mellitus. Reprogramming methods to generate cells lost in such disease states have been described and are rapidly maturing in the laboratory setting, such as those to produce hematopoietic stem cells, neurons, cardiomyocytes and pancreatic beta-islet cells (Y. Shi et al. 2017). Once refined, their translation to clinical trials will pave the way for curative approaches to chronic diseases. Early clinical successes, such as the successful autologous graft of RPE cells, described above, indicate that autologous grafts using reprogrammed cells are safe and teeming with potential.

Moreover, reprogrammed human cells have the potential to make superior models of disease compared to animal models. A patient-specific disease model may be created by generating iPSCs from skin fibroblasts and differentiating them into the cell type relevant to the patient's condition. This approach allows the experimentalist to shed light on the mechanisms of pathogenesis, as cells of patients with susceptibility to disease may be closely observed from their developmental infancy to their diseased state. Such iPSC-derived models have already been developed for several conditions, from microcephaly (Lancaster et al. 2013) to autism spectrum disorder (Mariani et al. 2015). While a useful tool in several settings, iPSCs may be limited in their ability to model age-related disease such as neurodegenerative conditions, as epigenetic rejuvenation caused by transit through pluripotent states may reverse aspects of pathogenesis. Direct reprogramming techniques have been employed to circumvent this, as this approach avoids widespread reversal of age-related epigenetic changes (Traxler et al. 2019; M. Lee et al. 2019). Using these techniques, reprogrammed cell lines may be generated from patients to study the mechanisms and pharmacologic susceptibilities of their disease, enabling better understanding of the connections between patient genotype and pathologic phenotype (Cherry and Daley 2013).

7 Conclusion

This review sought to examine advancements in cellular reprogramming and computational methods that contribute to them. Many promising approaches in cellular reprogramming are under development, though discovery of new TF recipes to efficiently convert source cells to target cells remains limiting. Mathematically-based approaches such as those outlined here may facilitate their systematic discovery. Moreover, convergent ideas in the areas of biology, mathematics and medicine point to hybrid (concurrent TF addition and removal), time-dependent, and concentration-dependent reprogramming regimes as viable next steps to improve reprogramming methods. These refinements could offer increased control over cell identity of normal and abnormal cells, and their *in vivo* regenerative potential.

Funding Information

This work is supported in part by AFOSR Award No: FA9550-18-1-0028, the Smale Institute. GAD is funded by the University of Michigan NIH NIGMS Bioinformatics Training Grant (T32 GM070449). CWR is supported by the NIH Medical Scientist Training Program (T32 GM007863-41).

Conflict of Interest

The authors have declared no conflicts of interest for this article.

Author Contributions

Gabrielle Dotson: Conceptualization; Investigation; Visualization; Writing - Review & Editing. **Charles Ryan:** Investigation; Visualization; Writing - Review & Editing. **Can Chen:** Investigation; Writing - Review & Editing. **Lindsey Muir:** Writing - Review & Editing. **Indika Rajapakse:** Conceptualization; Supervision; Writing - Review & Editing.

Acknowledgments

Figures (with the exception of Figure 4) were created with [BioRender.com](https://www.biorender.com). We would also like to thank Dr. Thomas Reid, Dr. Scott Ronquist, Dr. Glenn Dotson, and Karen Dotson for their insightful feedback on the manuscript.

References

- Ahlfors, J.-E., Azimi, A., El-Ayoubi, R., Velumian, A., Vonderwalde, I., Boscher, C., Mihai, O., Mani, S., Samoilova, M., Khazaei, M. et al. (2019). Examining the fundamental biology of a novel population of directly reprogrammed human neural precursor cells. *Stem Cell Research & Therapy*, *10*(1), 1–17.
- Allshire, R. C., & Madhani, H. D. (2018). Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology*, *19*(4), 229.
- Aydin, B., & Mazzoni, E. O. (2019). Cell reprogramming: The many roads to success. *Annual Review of Cell and Developmental Biology*, *35*, 433–452.
- Banerji, C. R., Miranda-Saavedra, D., Severini, S., Widschwendter, M., Enver, T., Zhou, J. X., & Teschendorff, A. E. (2013). Cellular network entropy as the energy potential in waddington's differentiation landscape. *Scientific Reports*, *3*(1), 1–7.
- Bichsel, C., Neeld, D., Hamazaki, T., Chang, L.-J., Yang, L.-J., Terada, N., & Jin, S. (2013). Direct reprogramming of fibroblasts to myocytes via bacterial injection of myod protein. *Cellular Reprogramming*, *15*(2), 117–125.
- Briggs, R., & King, T. J. (1952). Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proceedings of the National Academy of Sciences*, *38*(5), 455–463. <https://doi.org/10.1073/pnas.38.5.455>
- Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding rnas reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927.
- Cahan, P. (2016). Enabling direct fate conversion with network biology. *Nature Genetics*, *48*(3), 226–227.

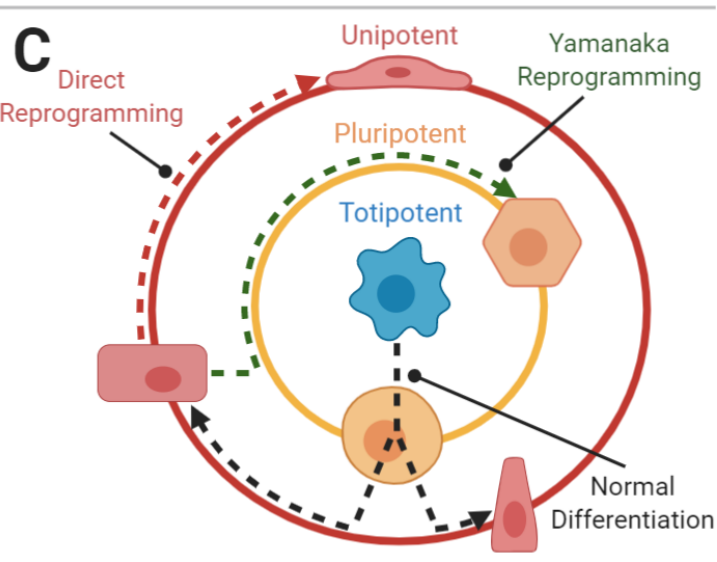
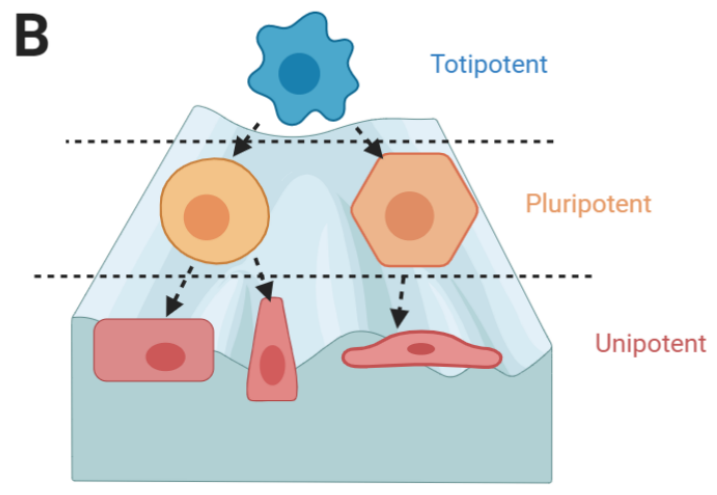
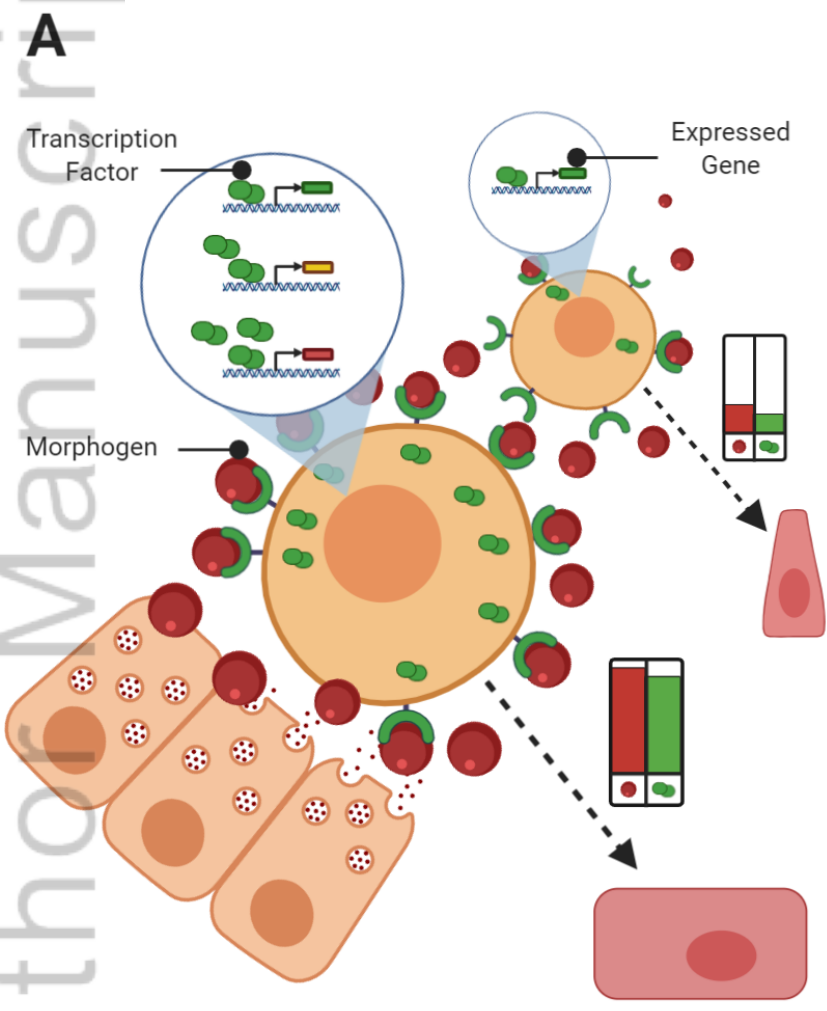
- Cahan, P., Li, H., Morris, S. A., Da Rocha, E. L., Daley, G. Q., & Collins, J. J. (2014). Cellnet: Network biology applied to stem cell engineering. *Cell*, *158*(4), 903–915.
- Cao, N., Huang, Y., Zheng, J., Spencer, C. I., Zhang, Y., Fu, J.-D., Nie, B., Xie, M., Zhang, M., Wang, H. [Haixia] et al. (2016). Conversion of human fibroblasts into functional cardiomyocytes by small molecules. *Science*, *352*(6290), 1216–1220.
- Cao, S., Yu, S., Chen, Y., Wang, X., Zhou, C., Liu, Y. [Yuting], Kuang, J., Liu, H., Li, D., Ye, J. et al. (2017). Chemical reprogramming of mouse embryonic and adult fibroblast into endoderm lineage. *Journal of Biological Chemistry*, *292*(46), 19122–19132.
- Cavelti-Weder, C., Li, W., Zumsteg, A., Stemann, M., Yamada, T., Bonner-Weir, S., Weir, G., & Zhou, Q. (2015). Direct reprogramming for pancreatic beta-cells using key developmental genes. *Current Pathobiology Reports*, *3*(1), 57–65.
- Chen, C., & Rajapakse, I. (2020). Tensor entropy for uniform hypergraphs. *IEEE Transactions on Network Science and Engineering*.
- Chen, C., Surana, A., Bloch, A., & Rajapakse, I. (2019). Multilinear time invariant system theory. *2019 Proceedings of the Conference on Control and its Applications*, 118–125.
- Chen, C., Surana, A., Bloch, A., & Rajapakse, I. (2020). Controllability of hypergraphs. *arXiv preprint arXiv:2005.12244*.
- Chen, J.-H., Goh, K. J., Rocha, N., Groeneveld, M. P., Minic, M., Barrett, T. G., Savage, D., & Semple, R. K. (2017). Evaluation of human dermal fibroblasts directly reprogrammed to adipocyte-like cells as a metabolic disease model. *Disease Models & Mechanisms*, *10*(12), 1411–1420.
- Chen, K. G., Mallon, B. S., McKay, R. D., & Robey, P. G. (2014). Human pluripotent stem cell culture: Considerations for maintenance, expansion, and therapeutics. *Cell Stem Cell*, *14*(1), 13–26.
- Cherry, A. B., & Daley, G. Q. (2013). Reprogrammed cells for disease modeling and regenerative medicine. *Annual Review of Medicine*, *64*, 277–290.
- Cichocki, A., Lee, N., Oseledets, I., Phan, A.-H., Zhao, Q., & Mandic, D. P. (2016). Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends® in Machine Learning*, *9*(4-5), 249–429.
- Cook, P., & Marenduzzo, D. (2018). Transcription-driven genome organization: A model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic acids research*, *46*.
- Cornelius, S. P., Kath, W. L., & Motter, A. E. (2013). Realistic control of network dynamics. *Nature Communications*, *4*(1), 1–9.
- Cover, T. M. (1999). *Elements of information theory*. John Wiley & Sons.
- D'Alessio, A. C., Fan, Z. P., Wert, K. J., Baranov, P., Cohen, M. A., Saini, J. S., Cohick, E., Charniga, C., Dadon, D., Hannett, N. M. et al. (2015). A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports*, *5*(5), 763–775.
- Del Vecchio, D., Abdallah, H., Qian, Y., & Collins, J. J. (2017). A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Systems*, *4*(1), 109–120.
- Fu, J.-D., Stone, N. R., Liu, L., Spencer, C. I., Qian, L., Hayashi, Y., Delgado-Olguin, P., Ding, S., Bruneau, B. G., & Srivastava, D. (2013). Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state. *Stem cell reports*, *1*(3), 235–247.
- Gao, L., Guan, W., Wang, M., Wang, H. [Huihan], Yu, J., Liu, Q., Qiu, B., Yu, Y., Ping, Y., Bian, X. et al. (2017). Direct generation of human neuronal cells from adult astrocytes by small molecules. *Stem Cell Reports*, *8*(3), 538–547.

- Goh, P. A., Caxaria, S., Casper, C., Rosales, C., Warner, T. T., Coffey, P. J., & Nathwani, A. C. (2013). A systematic evaluation of integration free reprogramming methods for deriving clinically relevant patient specific induced pluripotent stem (ips) cells. *PLoS One*, *8*(11), e81622.
- Grath, A., & Dai, G. (2019). Direct cell reprogramming for tissue engineering and regenerative medicine. *Journal of Biological Engineering*, *13*(1), 14.
- Gurdon, J. B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Development*, *10*(4), 622–640.
- Herdy, J., Schafer, S., Kim, Y., Ansari, Z., Zangwill, D., Ku, M., Paquola, A., Lee, H., Mertens, J., & Gage, F. H. (2019). Chemical modulation of transcriptionally enriched signaling pathways to optimize the conversion of fibroblasts into neurons. *Elife*, *8*, e41356.
- Horisawa, K., & Suzuki, A. (2020). Direct cell-fate conversion of somatic cells: Toward regenerative medicine and industries. *Proceedings of the Japan Academy, Series B*, *96*(4), 131–158.
- Huang, H., Zhong, L., Zhou, J., Hou, Y., Zhang, Z., Xing, X., & Sun, J. (2020). Leydig-like cells derived from reprogrammed human foreskin fibroblasts by crispr/dcas9 increase the level of serum testosterone in castrated male rats. *Journal of Cellular and Molecular Medicine*, *24*(7), 3971–3981.
- Ignatieva, E. V., Levitsky, V. G., & Kolchanov, N. A. (2015). Human genes encoding transcription factors and chromatin-modifying proteins have low levels of promoter polymorphism: A study of 1000 genomes project data. *International Journal of Genomics*, *2015*.
- Jackson, L. A., Anderson, E. J., Roupheal, N. G., Roberts, P. C., Makhene, M., Coler, R. N., McCullough, M. P., Chappell, J. D., Denison, M. R., Stevens, L. J. et al. (2020). An mrna vaccine against sars-cov-2—preliminary report. *New England Journal of Medicine*.
- Jin, Y., Liu, Y. [Ying], Li, Z., Santostefano, K., Shi, J., Zhang, X., Wu, D., Cheng, Z., Wu, W., Terada, N. et al. (2018). Enhanced differentiation of human pluripotent stem cells into cardiomyocytes by bacteria-mediated transcription factors delivery. *PLoS one*, *13*(3), e0194895.
- Kanherkar, R. R., Bhatia-Dey, N., & Csoka, A. B. (2014). Epigenetics across the human lifespan. *Frontiers in Cell and Developmental Biology*, *2*, 49.
- Kim, H. J., Oh, H. J., Park, J. S., Lee, J. S., Kim, J.-H., & Park, K.-H. (2020). Direct conversion of human dermal fibroblasts into cardiomyocyte-like cells using cicmc nanogels coupled with cardiac transcription factors and a nucleoside drug. *Advanced Science*, *7*(7), 1901818.
- Kimmel, J. C., Penland, L., Rubinstein, N. D., Hendrickson, D. G., Kelley, D. R., & Rosenthal, A. Z. (2019). Murine single-cell rna-seq reveals cell-identity-and tissue-specific trajectories of aging. *Genome Research*, *29*(12), 2088–2103.
- Ladewig, J., Koch, P., & Brüstle, O. (2013). Leveling waddington: The emergence of direct programming and the loss of cell fate hierarchies. *Nature Reviews Molecular Cell Biology*, *14*(4), 225–236.
- Lancaster, M. A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., & Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature*, *501*(7467), 373–379.
- Lee, C., Robinson, M., & Willerth, S. M. (2018). Direct reprogramming of glioblastoma cells into neurons using small molecules. *ACS Chemical Neuroscience*, *9*(12), 3175–3185.
- Lee, M., Sim, H., Ahn, H., Ha, J., Baek, A., Jeon, Y.-J., Son, M.-Y., & Kim, J. (2019). Direct reprogramming to human induced neuronal progenitors from fibroblasts of familial and sporadic parkinson's disease patients. *International journal of stem cells*, *12*(3), 474.

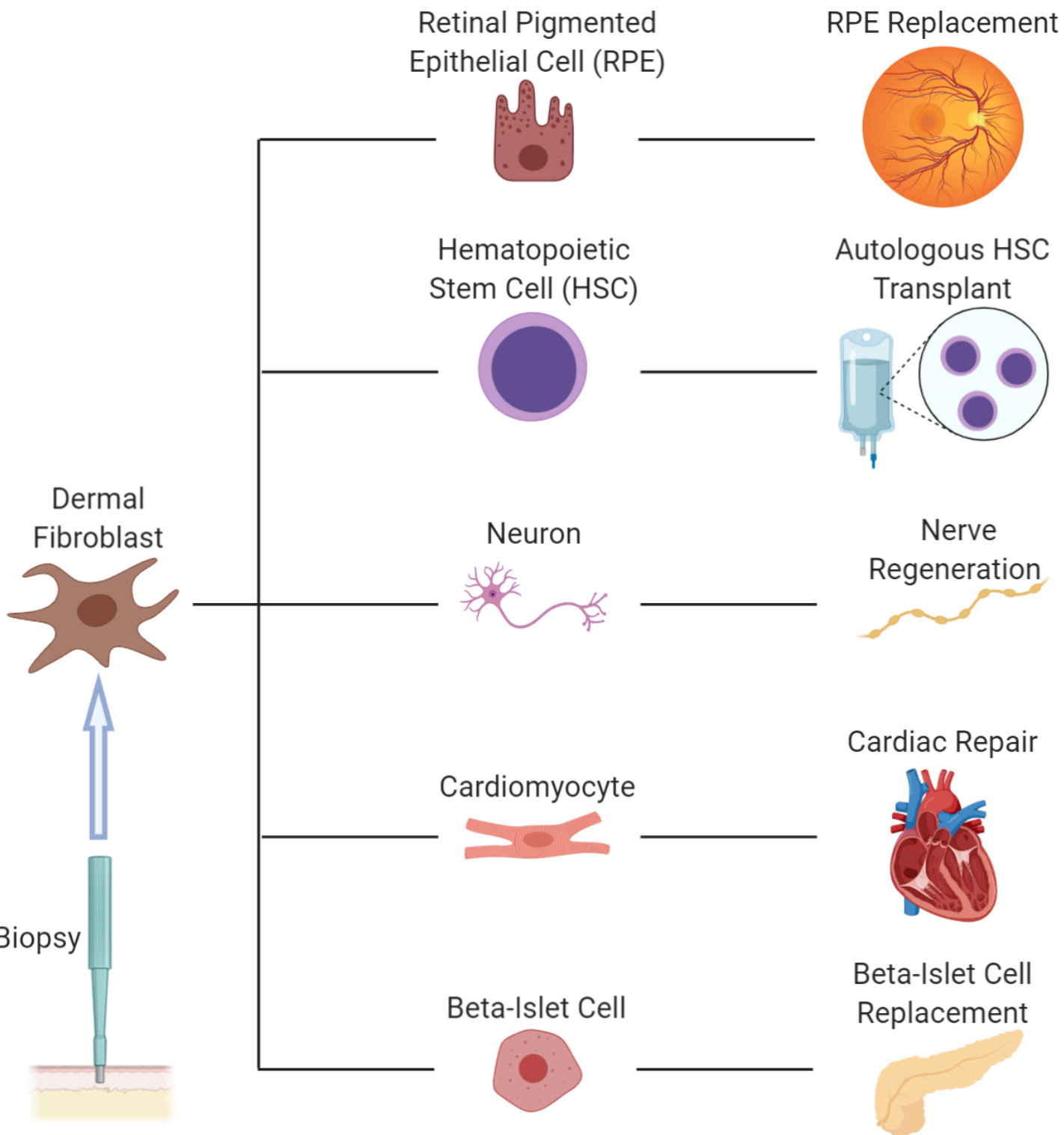
- Liao, H.-K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M.-Z., Sui, Y., Yamauchi, T., Sakurai, M., O'Keefe, D. D., Núñez-Delgado, E. et al. (2017). In vivo target gene activation via crispr/cas9-mediated trans-epigenetic modulation. *Cell*, *171*(7), 1495–1507.
- Linares-Fernández, S., Lacroix, C., Exposito, J.-Y., & Verrier, B. (2020). Tailoring mrna vaccine to balance innate/adaptive immune response. *Trends in Molecular Medicine*, *26*(3), 311–323.
- Lindner, F., Milne-Davies, B., Langenfeld, K., Stiewe, T., & Diepold, A. (2020). Litesec-t3ss-light-controlled protein delivery into eukaryotic cells with high spatial and temporal resolution. *Nature Communications*, *11*(1), 1–13.
- Liu, Y. (2011). Y., slotine. J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature*, *473*, 167–176.
- Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). MPCA: Multilinear principal component analysis of tensor objects. *IEEE transactions on Neural Networks*, *19*(1), 18–39.
- MacArthur, B. D., & Lemischka, I. R. (2013). Statistical mechanics of pluripotency. *Cell*, *154*(3), 484–489.
- MacArthur, B. D., Ma'ayan, A., & Lemischka, I. R. (2009). Systems biology of stem cell fate and cellular reprogramming. *Nature reviews Molecular cell biology*, *10*(10), 672–681.
- Mandai, M., Watanabe, A., Kurimoto, Y., Hirami, Y., Morinaga, C., Daimon, T., Fujihara, M., Akimaru, H., Sakai, N., Shibata, Y. et al. (2017). Autologous induced stem-cell-derived retinal cells for macular degeneration. *New England Journal of Medicine*, *376*(11), 1038–1046.
- Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., Amenduni, M., Szekeley, A., Palejev, D., Wilson, M. et al. (2015). Foxg1-dependent dysregulation of gaba/glutamate neuron differentiation in autism spectrum disorders. *Cell*, *162*(2), 375–390.
- Morris, S. A., Cahan, P., Li, H., Zhao, A. M., San Roman, A. K., Shivdasani, R. A., Collins, J. J., & Daley, G. Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via cellnet. *Cell*, *158*(4), 889–902.
- Morris, S. A., & Daley, G. Q. (2013). A blueprint for engineering cell fate: Current technologies to reprogram cell identity. *Cell Research*, *23*(1), 33–48.
- Ouyang, J. F., Kamaraj, U. S., Polo, J. M., Gough, J., & Rackham, O. J. (2019). Molecular interaction networks to select factors for cell conversion. *Computational stem cell biology* (pp. 333–361). Springer.
- Rackham, O. J., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., Suzuki, H., Nefzger, C. M., Daub, C. O., Shin, J. W. et al. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nature Genetics*, *48*(3), 331.
- Radley, A. H., Schwab, R. M., Tan, Y., Kim, J., Lo, E. K., & Cahan, P. (2017). Assessment of engineered cells using cellnet and rna-seq. *Nature Protocols*, *12*(5), 1089–1102.
- Rajapakse, I., Groudine, M., & Mesbahi, M. (2011). Dynamics and control of state-dependent networks for probing genomic organization. *Proceedings of the National Academy of Sciences*, *108*(42), 17257–17262.
- Rajapakse, I., Groudine, M., & Mesbahi, M. (2012). What can systems theory of networks offer to biology? *PLoS Comput Biol*, *8*(6), e1002543.
- Reichmuth, A. M., Oberli, M. A., Jaklenc, A., Langer, R., & Blankschtein, D. (2016). Mrna vaccine delivery using lipid nanoparticles. *Therapeutic delivery*, *7*(5), 319–334.

- Riddell, J., Gazit, R., Garrison, B. S., Guo, G., Saadatpour, A., Mandal, P. K., Ebina, W., Volchkov, P., Yuan, G.-C., Orkin, S. H. et al. (2014). Reprogramming committed murine blood cells to induced hematopoietic stem cells with defined factors. *Cell*, *157*(3), 549–564.
- Ronquist, S., Patterson, G., Muir, L. A., Lindsly, S., Chen, H., Brown, M., Wicha, M. S., Bloch, A., Brockett, R., & Rajapakse, I. (2017). Algorithm for cellular reprogramming. *Proceedings of the National Academy of Sciences*, *114*(45), 11832–11837.
- Rosa, F. F., Pires, C. F., Kurochkin, I., Ferreira, A. G., Gomes, A. M., Palma, L. G., Shaiv, K., Solanas, L., Azenha, C., Papatsenko, D. et al. (2018). Direct reprogramming of fibroblasts into antigen-presenting dendritic cells. *Science Immunology*, *3*(30).
- Ross, J. F., & Orlowski, M. (1982). Growth-rate-dependent adjustment of ribosome function in chemostat-grown cells of the fungus *mucor racemosus*. *Journal of Bacteriology*, *149*(2), 650–653.
- Ryu, W.-S. (2016). *Molecular virology of human pathogenic viruses*. Academic Press.
- Said, E. A., Tremblay, N., Al-Balushi, M. S., Al-Jabri, A. A., & Lamarre, D. (2018). Viruses seen by our cells: The role of viral rna sensors. *Journal of Immunology Research*, *2018*.
- Santolini, M., & Barabási, A.-L. (2018). Predicting perturbation patterns from the topology of biological networks. *Proceedings of the National Academy of Sciences*, *115*(27), E6375–E6383.
- Schlaeger, T. M., Daheron, L., Brickler, T. R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M. et al. (2015). A comparison of non-integrating reprogramming methods. *Nature Biotechnology*, *33*(1), 58–63.
- Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2017). Induced pluripotent stem cell technology: A decade of progress. *Nature reviews Drug discovery*, *16*(2), 115–130.
- Silvério-Alves, R., Gomes, A. M., Kurochkin, I., Moore, K. A., & Pereira, C.-F. (2019). Hemogenic reprogramming of human fibroblasts by enforced expression of transcription factors. *JoVE (Journal of Visualized Experiments)*, (153), e60112.
- Srivastava, D., & DeWitt, N. (2016). In vivo cellular reprogramming: The next generation. *Cell*, *166*(6), 1386–1396.
- Tachibana, M., Amato, P., Sparman, M., Gutierrez, N. M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.-S., Sritanaudomchai, H. et al. (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell*, *153*(6), 1228–1238.
- Takahashi, K. (2012). Cellular reprogramming—lowering gravity on waddington's epigenetic landscape. *Journal of Cell Science*, *125*(11), 2553–2560.
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *cell*, *131*(5), 861–872.
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676.
- Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Research*, *25*(10), 1491–1498.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., & Pachter, L. (2010). Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515.
- Traxler, L., Edenhofer, F., & Mertens, J. (2019). Next-generation disease modeling with direct conversion: A new path to old neurons. *FEBS letters*, *593*(23), 3316–3337.

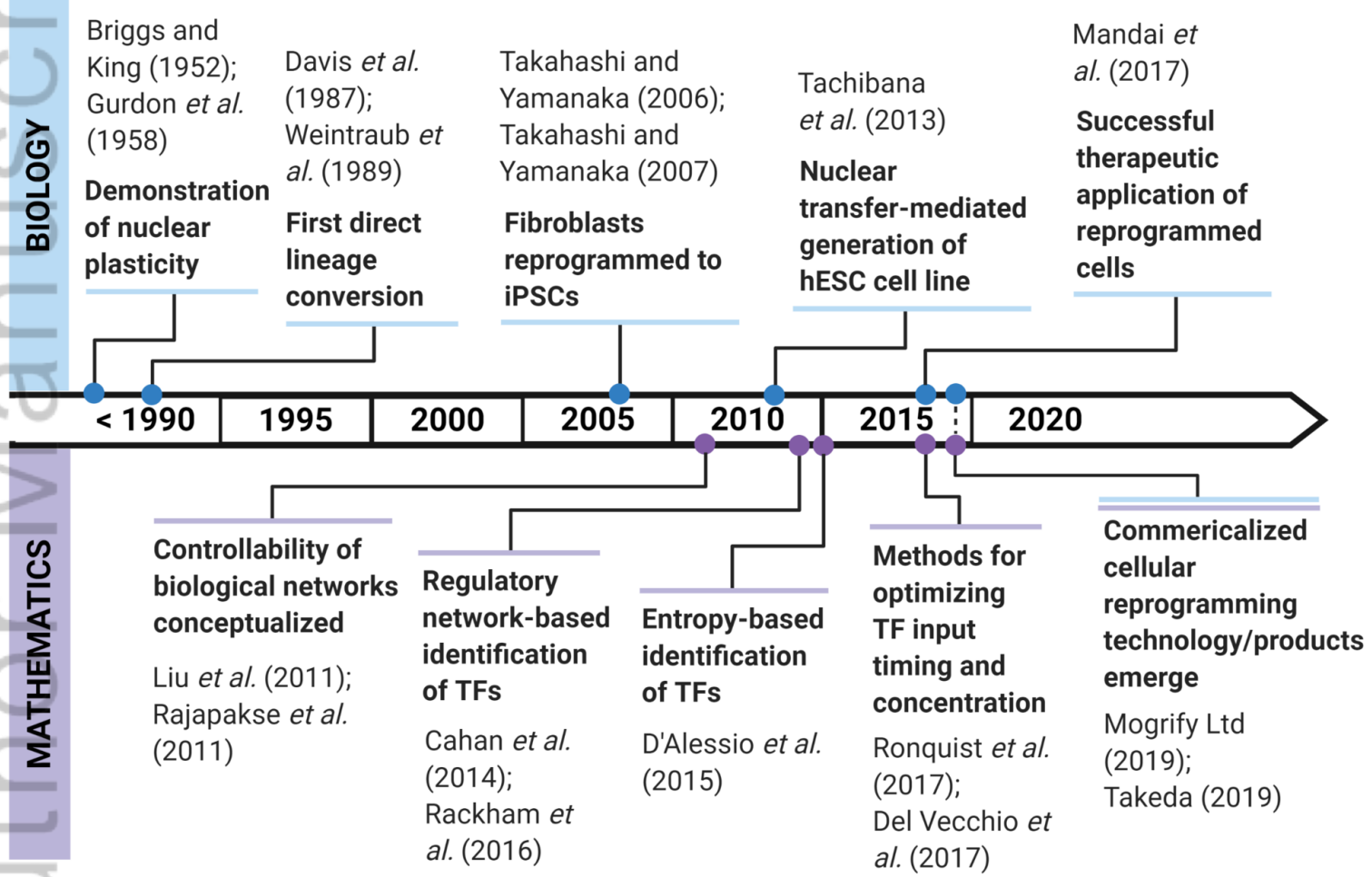
- Uren, A., Kool, J., Berns, A., & Van Lohuizen, M. (2005). Retroviral insertional mutagenesis: Past, present and future. *Oncogene*, *24*(52), 7656–7672.
- Van Pham, P., Vu, N. B., Dao, T. T.-T., Le, H. T.-N., Phi, L. T., & Phan, N. K. (2017). Production of endothelial progenitor cells from skin fibroblasts by direct reprogramming for clinical usages. *In Vitro Cellular & Developmental Biology-Animal*, *53*(3), 207–216.
- WA, C. K. M. J. R., & Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line nature 3806466. *Campbell, KH, McWhir, J., Ritchie, WA, and Wilmut, I.(1996). Sheep cloned by nuclear transfer from a cultured cell line. Nature*, *380*, 64–66.
- Waddington, C. H. (1957). *The strategy of the genes, a discussion of some aspects of theoretical biology*. G. Allen; Unwin.
- Wang, C., Liu, W., Nie, Y., Qaher, M., Horton, H. E., Yue, F., Asakura, A., & Kuang, S. (2017). Loss of myod promotes fate transdifferentiation of myoblasts into brown adipocytes. *EBioMedicine*, *16*, 212–223.
- Wang, J., Jiang, X., Zhao, L., Zuo, S., Chen, X., Zhang, L., Lin, Z., Zhao, X., Qin, Y., Zhou, X. et al. (2020). Lineage reprogramming of fibroblasts into induced cardiac progenitor cells by crispr/cas9-based transcriptional activators. *Acta Pharmaceutica Sinica B*, *10*(2), 313–326.
- Wang, W., Aggarwal, V., & Aeron, S. (2019). Principal component analysis with tensor train subspace. *Pattern Recognition Letters*, *122*, 86–91.
- Warren, L., & Lin, C. (2019). Mrna-based genetic reprogramming. *Molecular Therapy*, *27*(4), 729–734.
- Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of myod. *Proceedings of the National Academy of Sciences*, *86*(14), 5434–5438.
- Williams, A. H., Kim, T. H., Wang, F., Vyas, S., Ryu, S. I., Shenoy, K. V., Schnitzer, M., Kolda, T. G., & Ganguli, S. (2018). Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis. *Neuron*, *98*(6), 1099–1115.
- Xie, B., Sun, D., Du, Y., Jia, J., Sun, S., Xu, J., Liu, Y. [Yifang], Xiang, C., Chen, S., Xie, H. et al. (2019). A two-step lineage reprogramming strategy to generate functionally competent human hepatocytes from fibroblasts. *Cell Research*, *29*(9), 696–710.
- Xu, J., Du, Y., & Deng, H. (2015). Direct lineage reprogramming: Strategies, mechanisms, and applications. *Cell Stem Cell*, *16*(2), 119–134.
- Yamada, Y., Haga, H., & Yamada, Y. (2014). Concise review: Dedifferentiation meets cancer development: Proof of concept for epigenetic cancer. *Stem Cells Translational Medicine*, *3*(10), 1182–1187.
- Yang, Y., Chen, R., Wu, X., Zhao, Y., Fan, Y., Xiao, Z., Han, J., Sun, L., Wang, X., & Dai, J. (2019). Rapid and efficient conversion of human fibroblasts into functional neurons by small molecules. *Stem cell reports*, *13*(5), 862–876.
- Zhang, D., & Jiang, W. (2015). From one-cell to tissue: Reprogramming, cell differentiation and tissue engineering. *BioScience*, *65*(5), 468–475.
- Zhu, S., Russ, H. A., Wang, X., Zhang, M., Ma, T., Xu, T., Tang, S., Hebrok, M., & Ding, S. (2016). Human pancreatic beta-like cells converted from fibroblasts. *Nature communications*, *7*(1), 1–13.



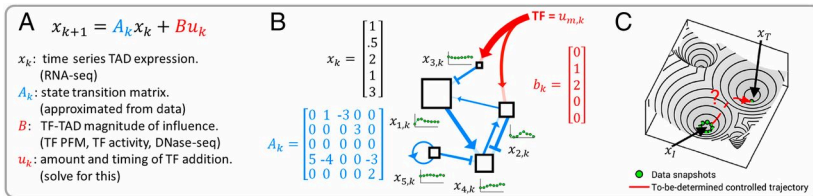
WSBM_1515_FIGURE1.tiff



WSBM_1515_FIGURE2.tiff

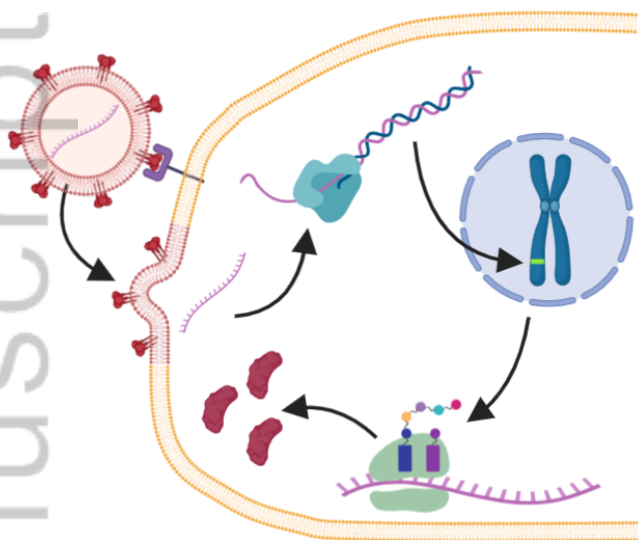


WSBM_1515_FIGURE3.tiff

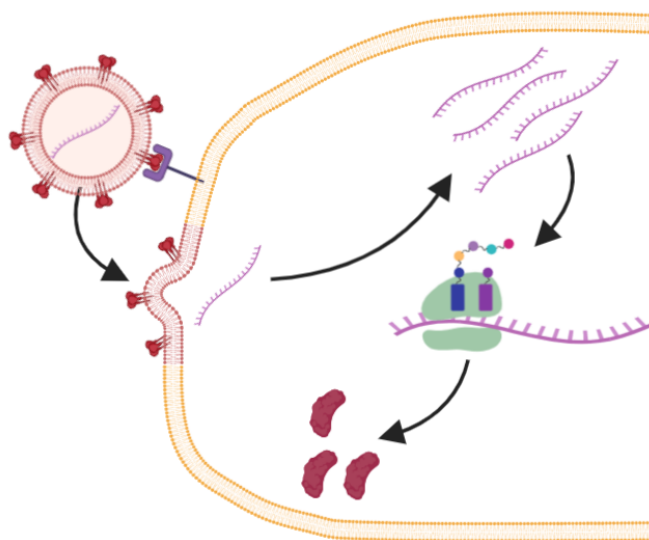


WSBM_1515_FIGURE4.tiff

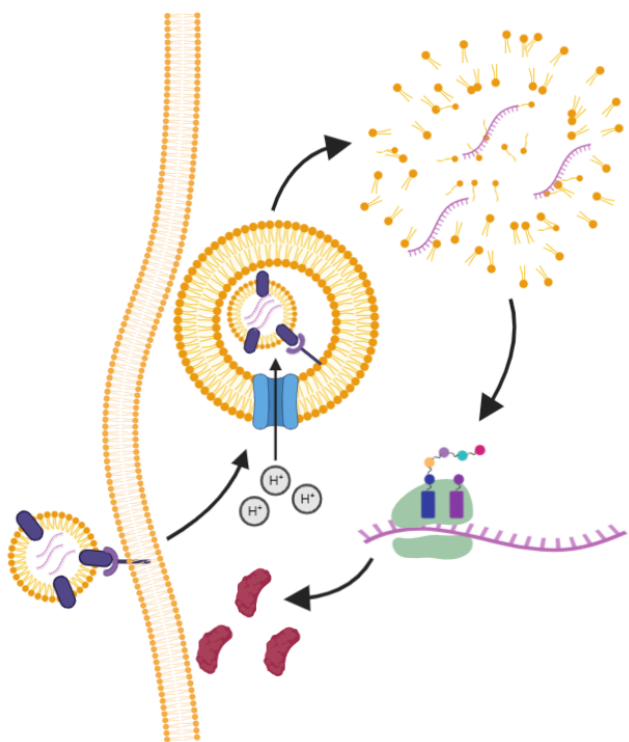
A Integrating Viral Vector



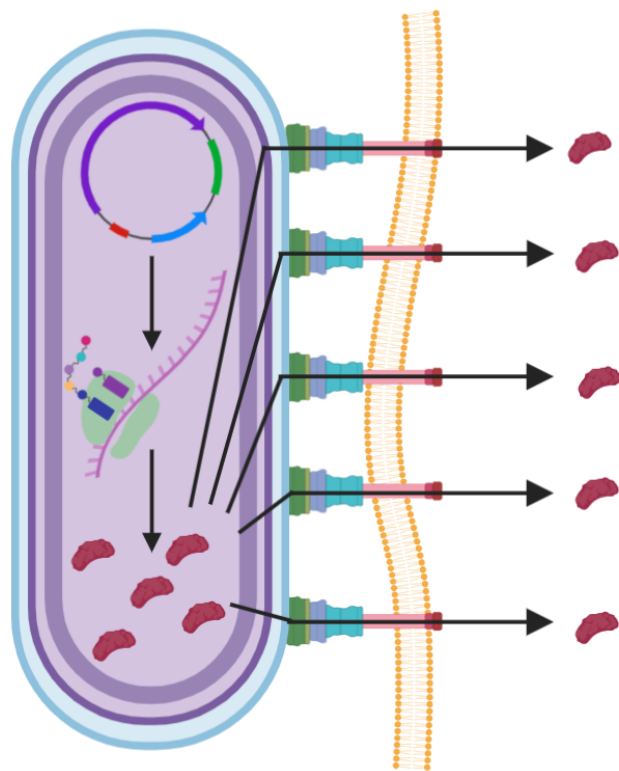
B Nonintegrating Viral Vector

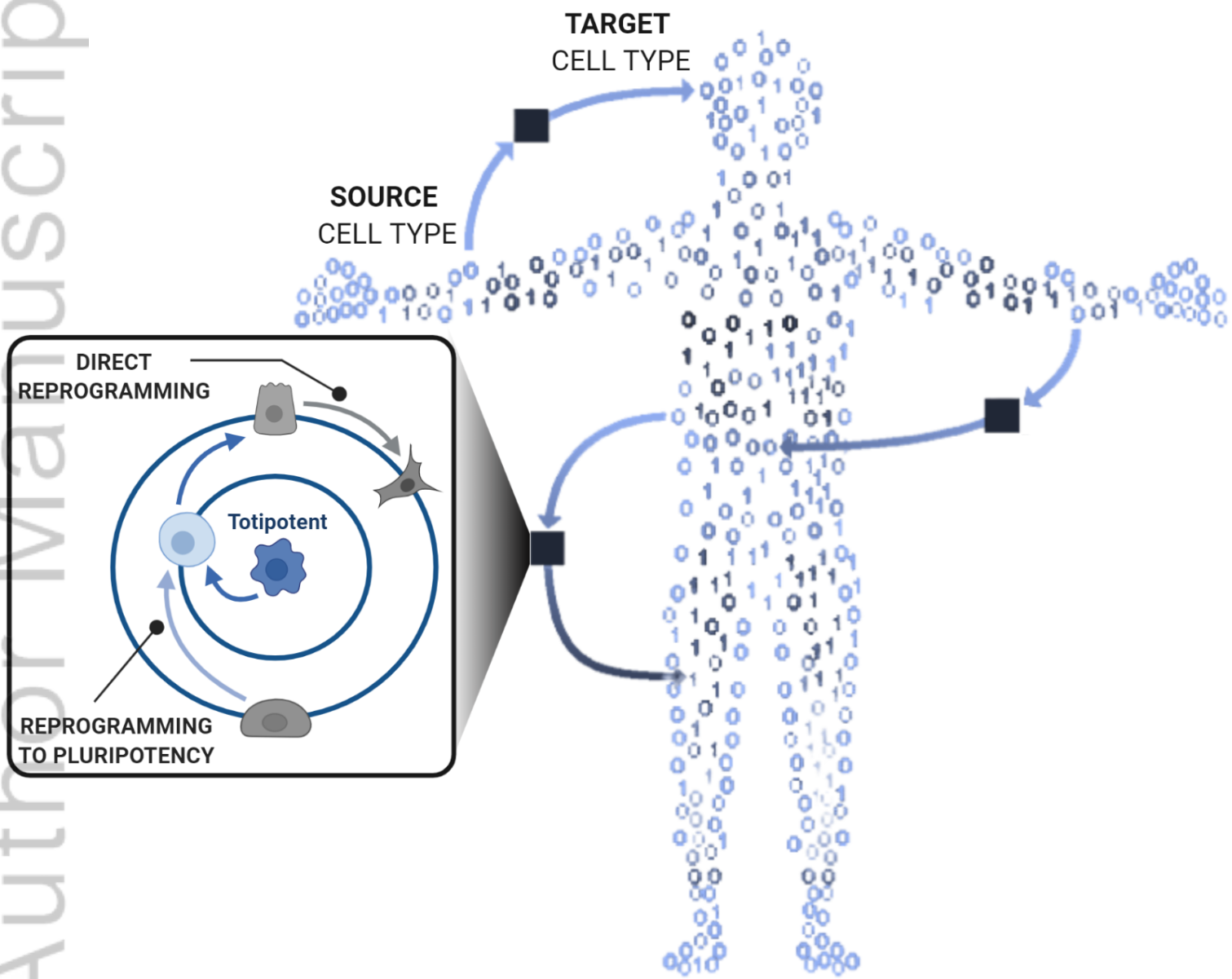


C Modified mRNA Delivery



D Type III Secretion System





WSBM_1515_GRAPHICAL_ABSTRACT2.tiff