Development of Novel Technologies for Human Pluripotent Stem Cell Biomanufacturing

by

Kaivalya Molugu

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Biophysics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 08/03/2021

The dissertation is approved by the following members of the Final Oral Committee:
Krishanu Saha, Associate Professor, Biomedical Engineering
Aseem Ansari, Professor, Chemical Biology & Therapeutics
William Murphy, Professor, Biomedical Engineering
Melissa Skala, Professor, Biomedical Engineering
Rupa Sridharan, Associate Professor, Cell and Regenerative Biology

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PROGRAMS

ABSTRACT

Development of Novel Technologies for Human
Pluripotent Stem Cell Biomanufacturing
Kaivalya Molugu
Under the supervision of Professor Krishanu Saha
at the University of Wisconsin-Madison

Somatic cell reprogramming to generate patient-specific induced pluripotent stem cells (iPSCs) and gene editing of iPSCs are important steps in many biomanufacturing applications, including disease modeling, drug discovery, and personalized cell therapies. However, the process of reprogramming is slow and inefficient leading to a heterogeneous population of cells, making it difficult to isolate high-quality iPSCs for downstream applications. Moreover, gene editing of iPSCs can typically be inefficient and laborious, involving multiple steps, requiring lengthy cell culture periods, drug selection, and several clonal events.

This thesis aims to develop new technologies to address the aforementioned challenges associated with reprogramming and gene editing of iPSCs. First, I present a platform to visually track the nucleus of reprogramming cell populations in real-time using high-content analysis, allowing the facile identification of high-quality iPSCs. Second, I combine this platform with label-free high-resolution optical metabolic imaging to build single-cell reprogramming trajectories and subsequently study reprogramming heterogeneity. Finally, I demonstrate a chromatin-modulation based CRISPR-Cas9 gene editing workflow to increase the efficiency of iPSC gene editing. Overall, these methods demonstrate an important step towards advancement in the biomanufacturing of iPSCs for clinical applications.

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Krishanu Saha, for his constant encouragement and support throughout my PhD journey. Thank you for providing me with the freedom to define my path during PhD, pushing me to think beyond my boundaries, and always trusting in my abilities.

Thank you to my committee members: Dr. Aseem Ansari, Dr. Melissa Skala, Dr. Rupa Sridharan, and Dr. William Murphy, who were instrumental in my growth as a scientist. Your feedback at every committee meeting helped me to think critically, challenge myself at every step, and helped to guide my thesis directions.

Thank you to all the Saha lab members (past and present) for great scientific discussions, incredible moral support, and creating an extremely fun and friendly work environment. Special thanks to Nicole Piscopo, Katie Mueller, Amr Abdeen for enabling my smooth transition into the Saha lab. Thank you to Nisha Iyer and Nikolai Fedorchak from Ashton lab, Brett Napiwocki from Crone lab for making Wisconsin Institute for Discovery a fun place to work at.

I am fortunate to have been a part of the student organizations, WiSCR and BADGRS, during my time at UW-Madison. Thank you to all the members of these organizations, who were instrumental in my growth as a scientist and as an individual.

Thank you to the Biophysics Program for accepting me into the program and providing a supportive environment. I would like to especially thank Allison Lynch and Kim Voss for always being there to support the Biophysics students.

I would like to thank all my Madison friends: Varshini Kamaraj, Parvathi Madathil Kooloth, Aashrith Saraswathibhatla, Amber Zhou, Dr. Zafirah Zaidan, Dr. Sachin Muley, Dr. Raunak Bardia, Dr. Vivek Saraswat, Debaditya Chatterjee, Roshan Xavier, Prasanth Kumar, Deepak Sirone, Dr. Mrugank Bhatt, Samartha Patel, Harshit Agrawal, Sneha Upadhyay, Laxmi Prasad, Sounak Mohanta, Rakesh Reddy, Dr. Shalaka Burlawar, Kshiteej Mahajan, Dr. Nitish Mathur, Kunal Bhagat, Asvin G, Sripradha Karkala, Simmi Pateriya,

Sudarsan Mahendran, Namita Khajanchi, and Vaidehi Patil. Thank you for making my time in Madison enjoyable and keeping my sanity in check when things became overwhelming. I could not have done this without all of you.

Thank you to my undergraduate friends who were always a phone call or a message away to provide me with incredible moral support: Tapan Goel, Sabareesh Ramachandran, G Mounika, Sampada Kolhatkar, Shashank H.R., Aravind Rao Karanam, Deepak Vinod, and Kishalay De.

Most importantly, I am so thankful to my family: my parents, brother, and sister-in-law. I am grateful to them for always supporting me and trusting my career choices. Thank you for all your courage and strength to send me to a far-off location 10,000 miles away to pursue an uncommon career. Thank you for all the sacrifices you have made to provide me with the best education and the best life possible, and for always being on my side through all the ups and downs of my PhD journey.

1 INTRODUCTION AND BACKGROUND

1.1 Induced Pluripotent Stem Cells

Pluripotent stem cells (PSCs) have the ability to self-renew indefinitely and develop into three primary germ layers of the early embryo and therefore into all cells of a multicellular organism other than the extra-embryonic tissues such as the placenta [1]. PSCs are typically confined to the early stages of embryonic development and represent a powerful system to study gene function and the physiological processes that occur during human development. PSCs also offer a rare route to studying the pathogenesis of the human genetic disease, especially for childhood developmental and age-related degenerative diseases, and to test new drugs [2]. The capacity of PSCs to differentiate into specific cell types could be leveraged in regenerative medicine to treat damaged or diseased tissues through cell replacement therapies [3, 4]. Accordingly, understanding the biology of pluripotency remains a pivotal aim in the fields of human development, drug discovery, and regenerative medicine.

Two commonly studied types of PSCs include embryonic stem cells (ESCs) and induced PSCs (iPSCs). Both the derivation of ESCs [5] and the generation of iPSCs [6] were awarded the Nobel Prizes in 2007 and 2012, respectively. ESCs are isolated from the inner cell mass of pre-implantation embryos [5, 7] but the use of human ESCs is greatly restricted due to inherent ethical issues, limited access to human embryos available for research, and risk of immunological rejection. Over the past decade, an alternative approach called reprogramming was developed to generate iPSCs from somatic cells, through the ectopic expression of a small set of transcription factors (Oct4, Sox2, c-Myc, Klf4; OSKM) [8], Nanog or RNA binding protein (Lin28) [9]. iPSC characterization has revealed that they are truly pluripotent, i.e., they have nearly identical gene expression profiles to ESCs; and show teratoma formation, chimera generation, and tetraploid complementation [10–12].

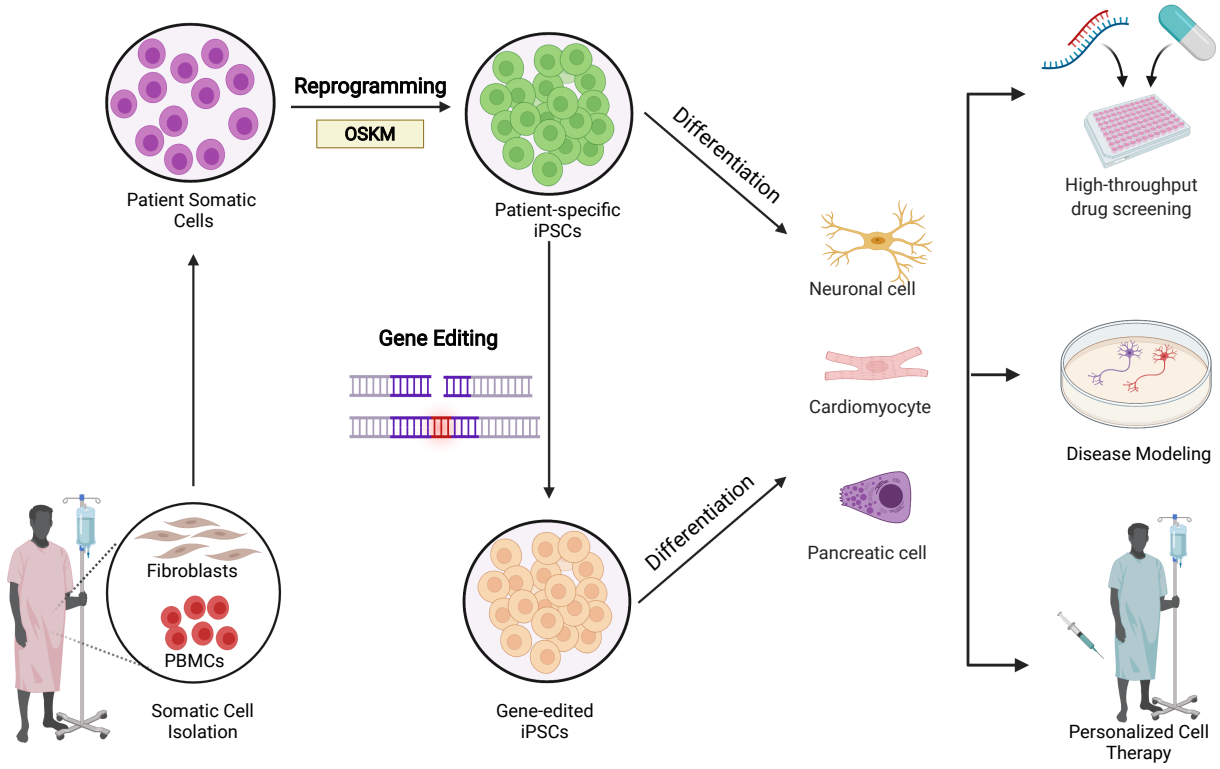


Figure 1.1: **Generation of induced pluripotent stem cells (iPSCs) and their applications.** Adult somatic cells isolated from patients can be reprogrammed to form iPSCs using reprogramming factors (OSKM). These iPSCs can then be directly differentiated to cell types from the three germ layers and used for disease modeling, drug screening, and personalized cell therapy. Alternatively, the iPSCs can be edited by gene editing technologies for gene insertion, gene disruption, and gene correction. These gene-edited iPSCs can then be used for further applications. Created with BioRender.com.

iPSCs thus represent a unique self-renewing cell source that carry the genome of the patient, facilitating dissection of the genetic causes of disease [13–15], and are immunologically matched to the patient, facilitating the engraftment of any cell therapies developed from these cells with a reduced risk of immune rejection. Moreover, iPSCs can also be genetically corrected using gene editing technologies like zinc finger nucleases (ZFN) [16], transcription activator-like effector nucleases (TALENs) [17], and clustered

regularly interspaced short palindromic repeats and CRISPR associated protein (CRISPR-Cas9) [18], re-differentiated into the desired cell type, and transplanted or transfused back into the same patient (**Figure 1.1**).

Despite the key advances of iPSCs in regenerative medicine, several challenges need to be addressed before iPSCs can be employed as viable therapeutic. One of the major challenges is the variability and heterogeneity that is associated with the processes of reprogramming and gene editing of iPSCs, thus warranting the development of technologies to enable further intensive investigations to understand the biology of these processes. Such enhanced technologies will contribute to moving iPSC-based regenerative medicine closer to clinical practice and applications in precision medicine. The following sections detail the current state-of-art technologies used for reprogramming and gene editing of iPSCs.

1.2 Cellular Reprogramming

Reprogramming to iPSCs can be achieved by administering defined exogenous transcription factors to somatic cells [6, 9]. However, this process can be slow and incomplete, resulting in highly heterogeneous populations of cells [19]. This makes it difficult to identify and isolate the cells that will undergo the necessary events for successful reprogramming. Moreover, the efficiency of this process is relatively low, ranging between 0.1% to 10% for most somatic cell types [20, 21]. There are also safety concerns associated with the overexpression of the reprogramming factors, including genetic mutations, gene insertions, epigenetic changes, incomplete reprogramming, and immunogenicity [12, 22–24].

Since the discovery of reprogramming in 2006, there have been extensive efforts to understand and engineer the reprogramming process to improve its speed, efficacy, and safety. Firstly, the consistency and reproducibility of the reprogramming process have been significantly augmented by establishing chemically defined, xeno-free culture conditions for deriving and maintaining iPSCs [25, 26], and by controlled delivery and expression

of reprogramming factors [27–29]. Secondly, a variety of alternative transcription factors and small molecules have been used to enhance cellular transitions during reprogramming, including metabolic regulators, epigenetic remodelers, cell cycle regulators, and pathway modulators [8, 30, 31]. Finally, the safety of the reprogramming process has been enhanced by use of non-integrating reprogramming methods, including adenoviruses [32], expression plasmids[33], piggyBac transposition [34], Sendai virus[35], direct delivery of reprogramming proteins[36], mRNAs [37], episomal vectors[38], chemical compounds [39] and self-replicative RNA [40], to combat the risk of transgene activation and insertional mutagenesis associated with traditional retroviral and lentiviral methods [41].

Even with these methodological and technical improvements, the efficiency of reprogramming remains low. Currently, two mechanistic theories, the elite and stochastic models, have been proposed to explain low reprogramming efficiencies [42] (**Figure 1.2**). According to the elite model, reprogramming would take place only in a few predisposed cells within a population, and according to the stochastic model most or all cells are competent for reprogramming at low probabilities. While some reports provide support for the elite model [43–48], other reports support the stochastic model [27, 49–51]. Thus, the mechanism of reprogramming continues to be debated and a comprehensive understanding of the mechanisms behind reprogramming is still lacking.

Epigenetic Changes during Reprogramming

The reprogramming process involves extensive global genetic and epigenetic changes, including major changes in chromatin and DNA methylation, and multiple players synergistically establish new transcriptional networks and remove epigenetic barriers [52]. Recent studies have demonstrated that there are two waves of transcriptional changes that occur during reprogramming, one immediately following the induction of the reprogramming factors and the other occurring later to activate the pluripotency gene network [53].

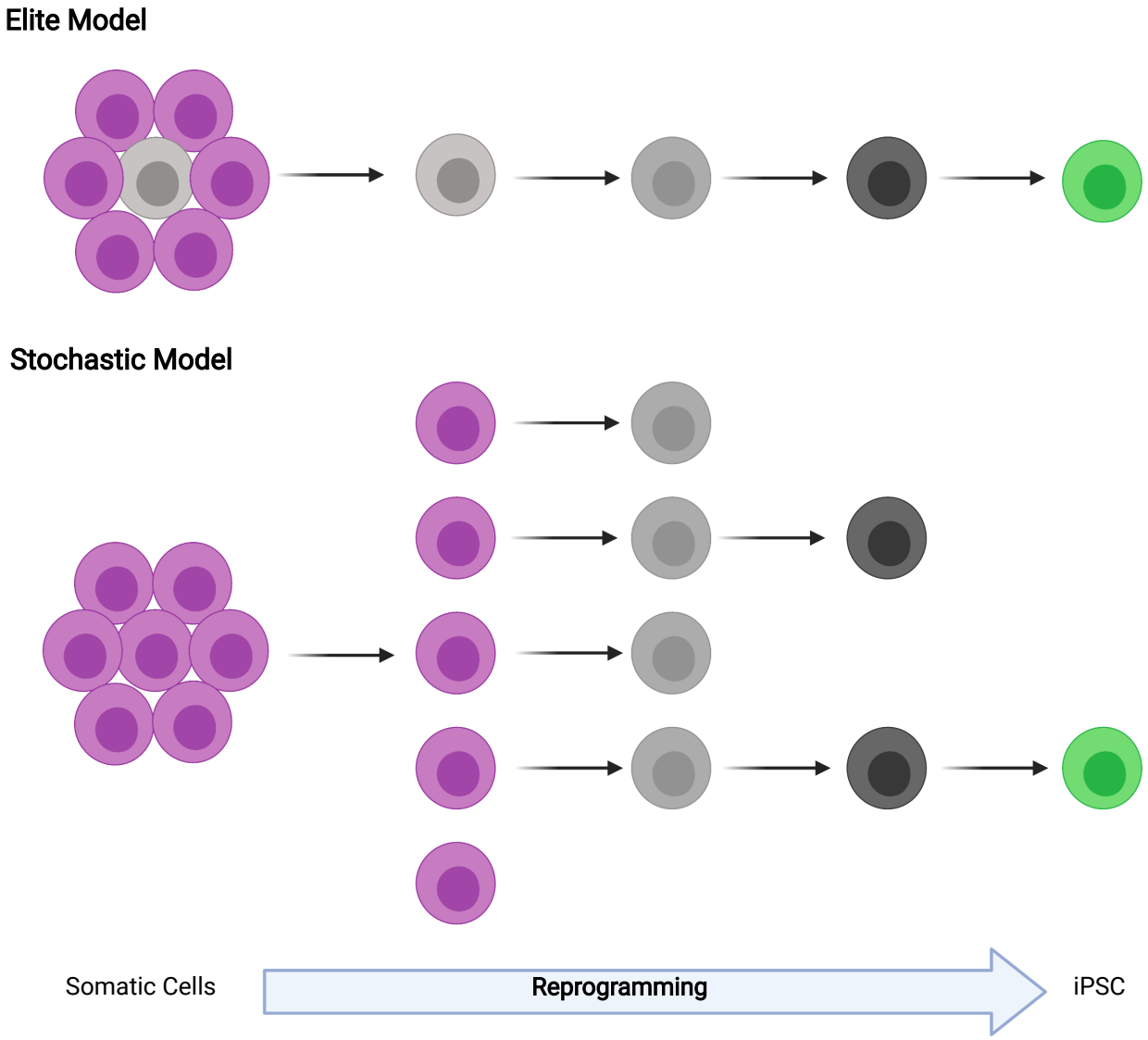


Figure 1.2: **Elite and stochastic models of reprogramming.** In the elite model, only a small number (light gray) of somatic cells (magenta) can be reprogrammed either partially or completely to iPSCs (green). In the stochastic model, most somatic cells (magenta) initiate the reprogramming process, but only a few can achieve complete reprogramming to iPSCs (green). Created with BioRender.com.

At the level of the epigenome, these changes include chromatin reorganization, DNA demethylation of promoter regions of pluripotency genes (*NANOG*, *SOX2* and *OCT4*),

reactivation of the somatically silenced X chromosome, and genome wide resetting of histone posttranslational modifications [54–56]. While some locus specific chromatin marks are established early [57], global changes in histone modification profiles [58] and DNA methylation [53] occur late in the reprogramming process, concurrent with the activation of pluripotency network.

Overall, these epigenome changes can manifest as changes in chromatin architecture during reprogramming. For instance, the global chromatin architecture has a more dispersed chromatin conformation characterized by a paucity of heterochromatin marks (H3K9me3) and an abundance of euchromatin marks (pan-H3ac, pan-H4ac, H3K36me2, H3K9ac, and H3K4me3) in iPSCs compared to somatic cells [59]. The modulation of chromatin architecture can then provide a basis for changes in nuclear morphology as well as gene expression [54, 60, 61], and has been studied in the context of reprogramming [55, 62]. High-resolution imaging of reprogramming cells has identified that nuclear geometry is dramatically altered during reprogramming [54, 63]. However, specific nuclear changes during the heterogeneous middle stages of reprogramming process have not been comprehensively characterized, especially in intact reprogramming cultures. In this thesis, I embark upon this challenge by longitudinal characterization of nuclear changes that occur during human somatic cell reprogramming and subsequently construct a predictive mathematical model for the identification of high-quality iPSCs.

Metabolic Changes during Reprogramming

Glycolysis and OXPHOS are two major metabolic pathways to provide energy for cells. Glycolysis converts glucose into 2 molecules of ATP, NADH, and pyruvate, via a series of metabolic reactions. This process is oxygen-independent and when the oxygen is available, the pyruvate is transported to mitochondria and converted to acetyl-CoA. Acetyl CoA is further oxidized in the tricarboxylic (TCA) cycle (or Krebs cycle), generating the electron carriers, NADH and $FADH_2$, to deliver the electrons to the electron transport chain (ETC).

The flow of electrons through ETC results in the pumping of protons from mitochondrial inner matrix to the outer matrix. ATP is synthesized through ATP synthase when protons flow back to the mitochondrial matrix. This process is called Oxidative Phosphorylation (OXPHOS) and produces 36 ATP molecules.

Compared to differentiated somatic cells, iPSCs are highly proliferative and have unique metabolic requirements [64]. Proliferation requires energy and significant amounts of nucleotides, lipids, and amino acids to assemble the two daughter cells that are produced with every cell division. Although OXPHOS produces far more ATP compared to glycolysis, iPSCs favor glycolysis regardless of oxygen availability (Warburg effect [64]) since it ensures that some glucose is diverted to generate precursors such as acetyl-CoA for fatty acid synthesis, glycolytic intermediates for nonessential amino acids, and ribose for nucleotides. During reprogramming, somatic cells thus undergo a metabolic shift from OXPHOS to glycolysis [65, 66]. This metabolic shift has been shown to be triggered by a transient OXPHOS burst, resulting in the initiation and progression of reprogramming to iPSCs [67, 68]. Moreover, the OXPHOS burst has been proposed to be induced by the expression of estrogen-related nuclear receptors (ERRα and ERRγ) [69], which then could result in the increase of reactive oxygen species (ROS) production, leading to activation of Hypoxia-inducible factor 1 (HIF1) and enhancement of the glycolytic rate [70].

In addition to these metabolic pathway changes, mitochondrial changes occur during the process of reprogramming. For instance, mitochondrial DNA (mtDNA) decreases during reprogramming [71], and mitochondria reverts to a more immature ESC-like state in terms of morphology, cellular distribution, and OXPHOS efficiency [71, 72].

Recent evidence also indicates that the metabolic shift during reprogramming occurs prior to changes in gene expression and that the modulation of glycolytic metabolism or OXPHOS can alter reprogramming efficiency [73–75]. This suggests that cell metabolism plays an important role in determining cell fate rather than passively responding to cell fate changes. However, there is still a lack of understanding of how dynamic cellular

metabolism changes that occur during reprogramming may impact pluripotency, notably the heterogeneous middle stage of reprogramming. In this thesis, I addressed this challenge by developing a high-content imaging platform to longitudinally track metabolic changes *in situ* during human somatic cell reprogramming and subsequently built single-cell reprogramming trajectories to gain deeper insights into reprogramming heterogeneity.

1.3 Genome Editing

Genome editing is a method by which cellular DNA sequences can be precisely manipulated to alter cell fates and traits of an organism [76]. This method allows genetic material to be added, removed, or altered at particular locations in the genome, and has a wide variety of uses in the biomedical field. Specifically, the pairing of gene editing technology with human iPSCs offers enormous potential for the development of autologous cell-based gene therapies for genetic diseases and can overcome the problem of immune rejection associated with allogeneic cell therapy. Moreover, since iPSCs have an unlimited proliferative ability, they can be used to generate an unlimited supply of syngeneic patient-derived transplantable gene-corrected cells [77]. Gene-edited iPSCs can be also used for disease modeling [78], screening potential new therapeutics [79], testing toxic side-effects of drug treatments [80], as they accurately represent the genetic background or cellular physiology of the patient [81]. Therefore, an increased basic understanding of gene editing would improve the efficiency of industrial processes of generating the aforementioned classes of precision medicine interventions.

Genome editing is typically initiated through the induction of a break within the DNA using a molecular pair of scissors, typically endonuclease enzymes, which can then be repaired by different cellular mechanisms. Most recently, engineered molecular machinery [18, 82–86] derived from bacterial immune pathways—known as clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins

(CRISPR–Cas systems) [87]—have revolutionized genome editing [88–90], overcoming many of the hurdles of previous systems. The key advancement of the CRISPR-Cas9 system is the use of an easily designed single guide RNA (sgRNA) in combination with a universal endonuclease that creates double stranded breaks (DSB) at a defined point in the genome. Use of RNA rather than protein-binding motifs imparts a higher level of specificity to gene editing, unlike the traditional engineered DNA-cleaving enzymes, i.e, meganucleases [91], zinc-finger nucleases (ZFNs) [16] and transcription activator-like effector nucleases (TALENs) [17].

Mechanisms of CRISPR-Cas9 Genome Editing

CRISPR-Cas9 exploits a ribonucleoprotein (RNP) complex consisting of two essential components: 1) a protein, SpyCas9 (referred to as "Cas9" hereafter), that localizes to the genome; and 2) a single-guide RNA (gRNA or sgRNA). sgRNA consists of 1) transactivating crispr (tracrRNA), short synthetic RNA composed of a scaffold sequence necessary for Cas9 binding; and 2) crispr RNA (crRNA), a user-defined 20 nucleotide spacer that defines the genomic target to be modified [92, 93]. The DNA recognition site must be adjacent to a short motif (the protospacer adjacent motif or PAM) that acts as a switch enabling Cas9 to create DSB at the target site. In cells of all multicellular organisms, including humans, such double-stranded DNA breaks induce DNA repair by endogenous cellular pathways that can introduce alterations to the DNA sequence, including small sequence changes or genetic insertions [94, 95].

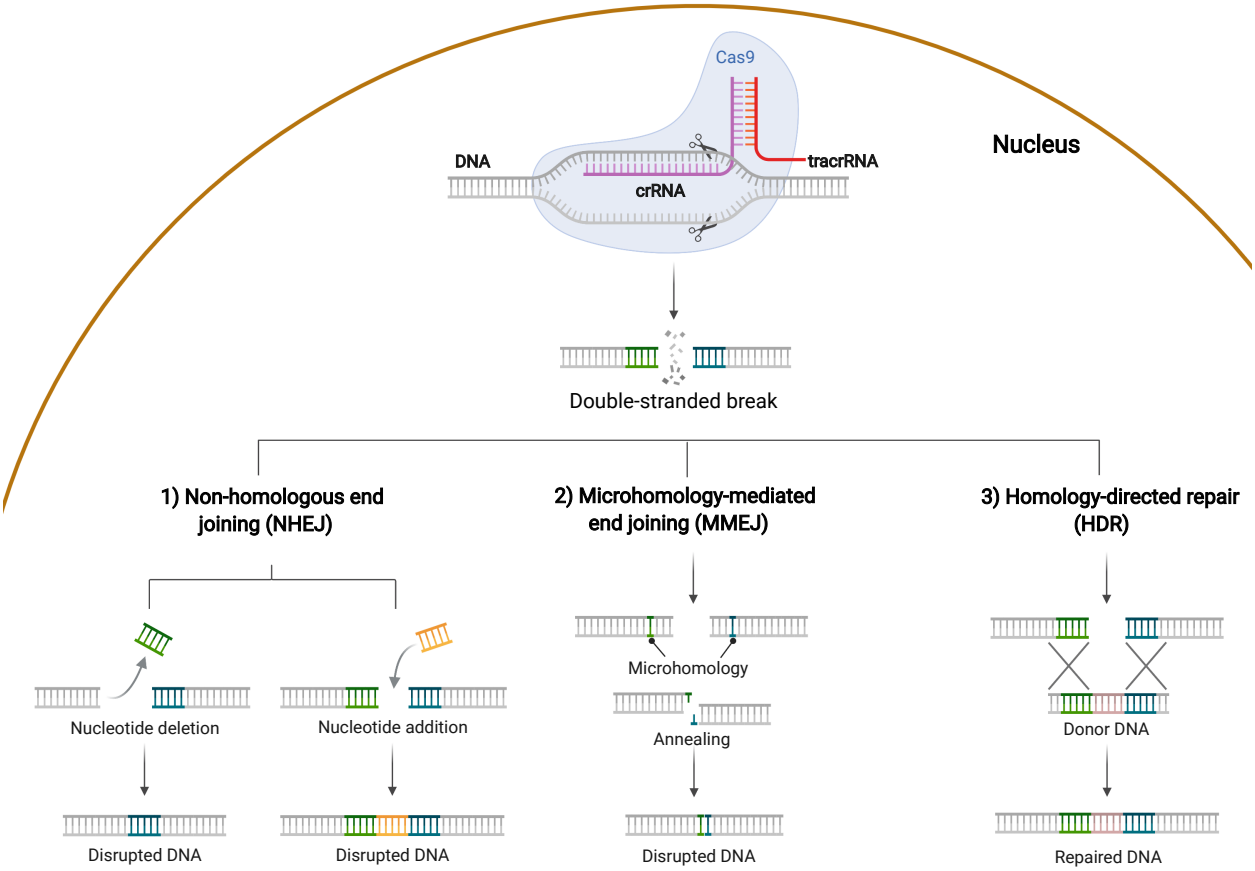


Figure 1.3: **Schematic of DSB repair mechanisms induced by CRISPR-Cas9.** DNA double stranded breaks (DSBs) induced by CRISPR-Cas9 can be repaired by: 1) non-homologous end-joining (NHEJ), 2) microhomology-mediated end joining (MMEJ), and 3) homology directed repair (HDR). NHEJ repair can produce variable-length insertion (yellow) and deletion (green) mutations at the site of the DSB. In MMEJ repair, the microhomology sequences (green and blue) exposed at the target site are annealed to each other, often leading to predictable small deletions. HDR repair can introduce precise point mutations or insertions (pink) from a single-stranded or double-stranded DNA donor template. Created with BioRender.com.

Three major DNA repair pathways resolve Cas9-induced DSBs: end joining (non-homologous or microhomology mediated, NHEJ or MMEJ) and homology-directed repair

(HDR) (**Figure 1.3**). In most cells, NHEJ is the predominant mechanism of repair that rejoins DSBs with minimal processing as quickly as 30 minutes after break induction [96]. The DSB is rejoined through the addition or subtraction of nucleotides thus resulting in new mutations in the genome known as insertion/deletion mutations (indels). NHEJ is active throughout the cell cycle and has a higher capacity for repair, as there is no requirement for a repair template or extensive DNA synthesis [97, 98]. In contrast to NHEJ, HDR is inherently precise because it proceeds via DNA 5′-3′ end resection followed by DSB repair using a homologous nucleic acid donor template (double-stranded or single-stranded) [99]. HDR is considerably slower than NHEJ and MMEJ, requiring seven or more hours to complete [96]. Moreover, HDR is primarily limited to the S and G2 phases of cell cycle, concurrent with the availability of the homologous sequence that is used as a repair template [100]. These two DNA repair pathways are currently being exploited to conduct targeted genomic changes in various clinical trials [101]: 1) gene editing through HDR used to replace a pathogenic variant or insert foreign DNA elements to restore the wild-type (WT) expression of a missing or truncated gene, and 2) NHEJ used to remove DNA elements leading to aberrant expression of genes or to gain a therapeutic function.

MMEJ is a relatively recent defined pathway that shares characteristics of both NHEJ and HDR pathways [102, 103] and is beginning to exploited as a strategy for therapeutic gene knockouts [103, 104] and corrections [102, 105, 106]. Similar to NHEJ, MMEJ joins the broken DSB ends without a repair template from an exogenous donor and requires initial DSB end resection like in the case of HDR [107]. However, resection during MMEJ is relatively short and exposes 5-25 base pairs (bp) homologies on opposite strands that anneal to one another. MMEJ repair is low in G0/G1 phase and is increased during S and G2 phases of the cell cycle [107].

The clinical utility of genome editing depends predominantly on precision (fraction of on-target edits that produce the desired genetic outcome) and accuracy (ratio of on- versus off-target genetic changes) of the process. Firstly, precision in gene editing is limited

by the high variability in genotypic outcomes of the commonly deployed NHEJ repair pathway or the low efficiency of the more precise HDR pathway. Currently, CRISPR-Cas9 gene editing efficiency varies from 2% to approximately 25% depending on the cell type [18], which is not yet up to the requirements for clinical use. This shortcoming often results in complicated, labor-intensive, and destructive selection processes for characterizing and identifying gene edits of interest from heterogeneous human cell populations [108, 109].

Specifically in iPSCs, gene editing can involve multiple steps, requiring lengthy cell culture periods, drug selection, and several clonal events (i.e., initial reprogramming, gene targeting, and subsequent genetic excision of a selection cassette [13, 110]). The selection of rare edited iPSC clones can be particularly difficult because they grow poorly as single cells [111, 112]. Low accuracy or off-target mutagenesis can also be a major concern, particularly for clinical applications. Off-target sites that were not predicted by standard bioinformatic sequence analysis were uncovered by latest techniques such as digenome-seq [113], CIRCLE-seq [114], CHANGE-seq [115], GUIDE-seq [116], DISCOVER-seq [117] and integrase-deficient lentiviral vectors [118]. Varying rates of off-target events were reported ranging from >1000 per sgRNA sequence [119] to negligible effects [120], thus warranting better evaluation of off-target sites.

Overall, there is a need to improve the precision and accuracy of genome editing and expand the throughput of current *in vitro* human culture systems where novel genome editing approaches can be evaluated within human cells and tissues. In particular, advanced iPSC genome editing capabilities could eventually expand the suite of personalized regenerative medicine applications and human preclinical model systems, ranging from patient-specific cell lines to complex human tissues derived from stem cells.

Since the inception of genome editing, there have been extensive efforts to understand the mechanisms underlying genome editing and to engineer the process to improve its specificity, efficiency, and versatility. Firstly, modified nuclease strategies, such as base [121] and prime editors [122], that use different pathways for gene modification have

been designed. Secondly, other natural and engineered Cas9 variants have been identified and developed with distinct and enhanced targeting properties [123], including Cas12a (Cpf1), Cas12b (C2c1), FokI fused to dCas9, Cas9-HF1, eSpCs9, evoCas9, and HypaCas9. Thirdly, Cas9 variants with distinct protospacer-adjacent motif (PAM) recognition sites have been generated, including VQR and VRER variants [124], and xCas9 [125]. Fourthly, sgRNAs have been modified with respect to their length, structure, and chemistry to reduce off-target properties. Finally, small molecules affecting DNA repair [126, 127] and cell cycle modulators [128] have been used to control and promote editing. Although these developments are promising, the underlying genomic context, particularly the chromatin state of the target locus, can also significantly influence the CRISPR-Cas9 gene editing outcomes [115, 129, 130]. However, the underlying mechanisms of how Cas9 interacts with chromatinized DNA are still not clear, warranting a deeper understanding of these mechanisms.

Chromatin Structure in Cas9-mediated Genome Editing

The chromatin structure of the genomic target has been reported to have an influence on Cas9 binding and gene editing efficiencies [115, 129, 131], thus leading to variable gene editing outcomes in different cell types both *in vitro* and *in vivo*. For instance, closed chromatin (heterochromatin) can negatively affect Cas9 binding [132–136] or delay CRISPR-Cas9 mutagenesis [137], nucleosomes can block or present a hurdle for Cas9 access to PAM sites [138–141], and active transcription in open chromatin (euchromatin) states can directly stimulate DNA cleavage by influencing Cas9 release states in a strand-specific manner [129, 142]. Moreover, the off-target binding of Cas9 to "seed" sequences can directly correlate with DNase I hypersensitivity of sequences [119] and inversely correlate with CpG methylation sites, thus indicating a possible correlation between off-target sites and their chromatin structure [113, 135, 137].

Local chromatin can also impact the choice of DNA repair pathway [143–146].

While some studies showed that active histone mark H3K36me3 and inactive H3K9me2/3 heterochromatin marks, promote HDR over NHEJ [144, 147, 148]; some studies found no major change in the balance between NHEJ and HDR between closed chromatin and open chromatin states of a sequence [134, 137]. A recent study also showed that heterochromatin is more prone to MMEJ as compared to NHEJ and that the shift depends on precise heterochromatin marks present [149].

Since the chromatin state of the target site has been shown to have a strong effect on the accessibility of DNA to Cas9, recent approaches to manipulate the chromatin state have emerged to modulate gene editing efficiency. One way in which the open or closed state of chromatin structure can be controlled is by the balance of histone acetylation and deacetylation, which is regulated by two groups of enzymes called HAT (histone acetyl-transferase) and HDAC (histone deacetylase) [150, 151]. While histone acetylation leads to euchromatin, histone deacetylation leads to heterochromatin. Hence, small-molecule HDAC inhibitors, such as trichostatin A (TSA), romidepsin, valproic acid and entinostat, have been used to improve gene editing [133, 152–154].On the other hand, HAT inhibitors such as C646 (p300/CBP inhibitor) and MG149 (Tip60 and MOZ inhibitor), have been shown to decrease gene editing efficiencies [133]. However, the effect of these HATs or HDACs on genome editing of CRISPR-Cas9 is yet to be extensively characterized in iPSCs. Though the chromatin structure has been shown to impact the off-target effects [113, 135, 137, 155] of editors and the potency of edited iPSCs [62, 156], these effects have not been thoroughly investigated thus far. In this thesis, I tackle the aforementioned challenge by extensively characterizing the impact of chromatin modulation (via Trichostatin A; HDAC inhibitor) on CRISPR-Cas9 gene editing outcomes in iPSCs.

2 TRACKING AND PREDICTING HUMAN SOMATIC CELL REPROGRAMMING USING NUCLEAR CHARACTERISTICS

Work in this chapter was adapted from:

Tracking and Predicting Human Somatic Cell Reprogramming Using Nuclear Characteristics

Kaivalya Molugu*, Ty Harkness*, Jared Carlson-Stevermer, Ryan Prestil, Nicole Piscopo, Stephanie Seymour, Gavin Knight, Randolph Ashton and, Krishanu Saha

Biophysical Journal 2020 May 5;118(9):2086-2102.

*These authors contributed equally.

2.1 Abstract

Reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) generates valuable resources for disease modeling, cell therapy, and regenerative medicine. However, the reprogramming process can be stochastic and inefficient, creating many partially-reprogrammed intermediates and non-reprogrammed cells in addition to fully-reprogrammed iPSCs. Much of the work to identify, evaluate, and enrich for iPSCs during reprogramming relies on methods that fix, destroy, or singularize cell cultures, thereby disrupting each cell's microenvironment. Here we develop a micropatterned substrate that allows for dynamic live-cell microscopy of hundreds of cell subpopulations undergoing reprogramming while preserving many of the biophysical and biochemical cues within the cells' microenvironment. On this substrate, we were able to both watch and physically confine cells into discrete islands during the reprogramming of human somatic cells from skin biopsies and blood draws obtained from healthy donors. Using high-content analysis, we identified a combination of eight nuclear characteristics that can be used to generate a computational model to predict the progression of reprogramming and distinguish partially-reprogrammed cells from those that are fully-reprogrammed. This approach to track reprogramming *in situ* using micropatterned substrates could aid in biomanufacturing of therapeutically-relevant iPSCs, and be used to elucidate multiscale cellular changes (cell-cell interactions as well as subcellular changes) that accompany human cell fate transitions.

2.2 Introduction

Somatic cell reprogramming to induced pluripotent stem cells (iPSCs) is an important step in many workflows involving drug discovery, regenerative medicine, and toxicology. However, reprogramming is a stochastic process resulting in highly heterogeneous populations of cells that include fully-reprogrammed iPSCs, partially-reprogrammed intermediates,

and non-reprogrammed cells [27, 49, 157, 158]. The quality control of iPSC colony determination is extremely important for its downstream expansion and maintenance of a homogeneous culture of undifferentiated cells [159]. Failure to select high-quality (fully-reprogrammed) iPSCs during the reprogramming process may result in cell lines that require increased maintenance by manual cell culture practices, insufficient downstream differentiation into functional cells, and unreliable experimental outcomes. The reprogramming process still routinely involves significant manual handling and expert judgement in picking the high-quality iPSCs and is not typically automated. Therefore, it is essential to develop reliable quantitative methods to select high-quality iPSCs by eliminating the contamination from partially-reprogrammed intermediates and non-reprogrammed cells [160]. This could further result in significant gains in creating robust cell therapies and disease models.

Current approaches for assessing pluripotency and identifying high-quality iPSCs are time-consuming, costly, destructive, and often descriptive rather than quantitative. Cellular morphology has been used as an approach to identify high-quality iPSCs [160–162]. However, manual picking of iPSC colonies based on cellular morphology can be time consuming, tedious, and qualitative. Even recent quantitative evaluation methods based on cellular morphology have had limited success in the field and are not widely implemented [163, 164]. Immunofluorescence staining or reporter systems to detect pluripotency markers such as Oct-4, Nanog, TRA-1-60, and TRA-1-81 [165–167] have been coupled with automated fluorescence microscopy or flow cytometry for the detection of high-quality iPSCs. However, this kind of probe-based labelling can cause potential safety issues for downstream applications and might also require cell fixation. Additionally, teratoma assays, embryoid body assays, PluriTest, and TaqMan hPSC Scorecard™ assays [168–171] are time-consuming and have limited resolution, as they are unable to distinguish high-quality iPSC lines composed of a high percentage of pluripotent stem cells from those that may be more heterogeneous in nature and contain a subpopulation of cells that are pluripotent.

Moreover, such multi-week characterization methods of isolated potential iPSC subpopulations have the problem of leading to significant cell culture artifacts [172] and increased costs [173]. The high number of cells that do not undergo successful reprogramming also means that single-cell assays profile many cells that do not fully reprogram, thus limiting the throughput of such assays (e.g., RNA-seq, ChIP-seq, ATAC-seq) [49, 53, 158, 174, 175]. Such methods also disrupt the cellular microenvironment, including cell-cell junctions, and drastically perturb the cytoskeleton, thus resulting in significant changes in the biophysical properties of cells undergoing reprogramming. Microenvironment disruption is a concern during reprogramming, as the cellular microenvironment can initiate cytoskeletal signaling cascades that act to regulate the structure of the nuclear lamina [176] and in turn affect the epigenetic state of cells. Such disruption in the epigenetic state may result in low-quality iPSCs that have limited differentiation potential. Further, the nuclear lamina has been shown to play an active role in genome-wide gene expression by physically repressing genes [176–179].

Changes in nuclear morphology have been characterized in reprogramming of mouse cells [157], and in human cells [63]. High resolution imaging of reprogramming cells has identified that nuclear geometry is dramatically altered during reprogramming [54, 63]. This may be due to the expression of kinases that activate cytoskeletal remodeling processes which are critical for reprogramming [180]. These biophysical changes have traditionally been studied in the context of mesenchymal-to-epithelial transition (MET), a process that occurs relatively early during reprogramming [181–183], concurrent with epigenetic changes indicating a loss of somatic identity, known as erasure [57, 182, 184]. However, specific nuclear changes during the epigenetic reprogramming process have not been comprehensively characterized, especially in intact cultures undergoing reprogramming.

Here, we describe a microcontact printed (μCP) platform that dissects cell cultures undergoing reprogramming into hundreds of subpopulations with $\sim 10^3$ cells in each subpopulation and allows for high-content dynamic live-cell microscopy of these cell

subpopulations. This platform supports reprogramming of fibroblasts from skin biopsies and erythroid progenitor cells (EPCs) from routine blood draws, representing the two most common sources of somatic cells for reprogramming [9, 41, 185, 186]. Three-dimensional principal component maps generated using a combination of eight nuclear characteristics were used to track reprogramming progression and distinguish non-reprogrammed cells and partially-reprogrammed intermediates from those that were fully-reprogrammed. These maps are consistent with and complement the insights obtained by single cell analysis [49, 58, 158, 187–190]. Further, high-quality iPSCs were isolated within 3-4 weeks using this platform. Overall, generating a comprehensive map of nuclear characteristics as well as a quantitative predictive model using our method, helps in the identification of homogeneous populations of high-quality iPSCs in a non-destructive, quantitative, and reliable manner.

2.3 Results

Controlled adhesion of reprogramming cells

We utilized a cell culture platform, the microcontact printed well plate (μCP Well Plate) [191], to control cell adhesion of both fibroblasts and EPCs during reprogramming. The μCP Well Plate is formed by creating hydrophilic polyethylene glycol (PEG) brushes that resist protein adsorption at defined locations on a gold-coated glass sheet. This sheet is then combined with a bottomless standard tissue culture 24-well plate to form a μCP Well Plate (**Figure A.1A**). We next seeded these plates with Oct4-Sox2-Klf4-cMyc reprogrammable human fibroblasts [63] or with human erythroid progenitor cells (EPCs) electroporated using four episomal reprogramming plasmids encoding Oct4, shRNA knockdown of p53, Sox2, Klf4, L-Myc, Lin28, and miR302-367 cluster [33, 192]. Within hours after cell attachment, this platform (**Figure 2.1**) constrained cell-to-cell contacts and controlled the geometry of multicellular subpopulations. We were able to pattern cell subpopulations into various geometries and within microfeatures (μFeatures) of a defined diameter (**Figure**

A.2C).

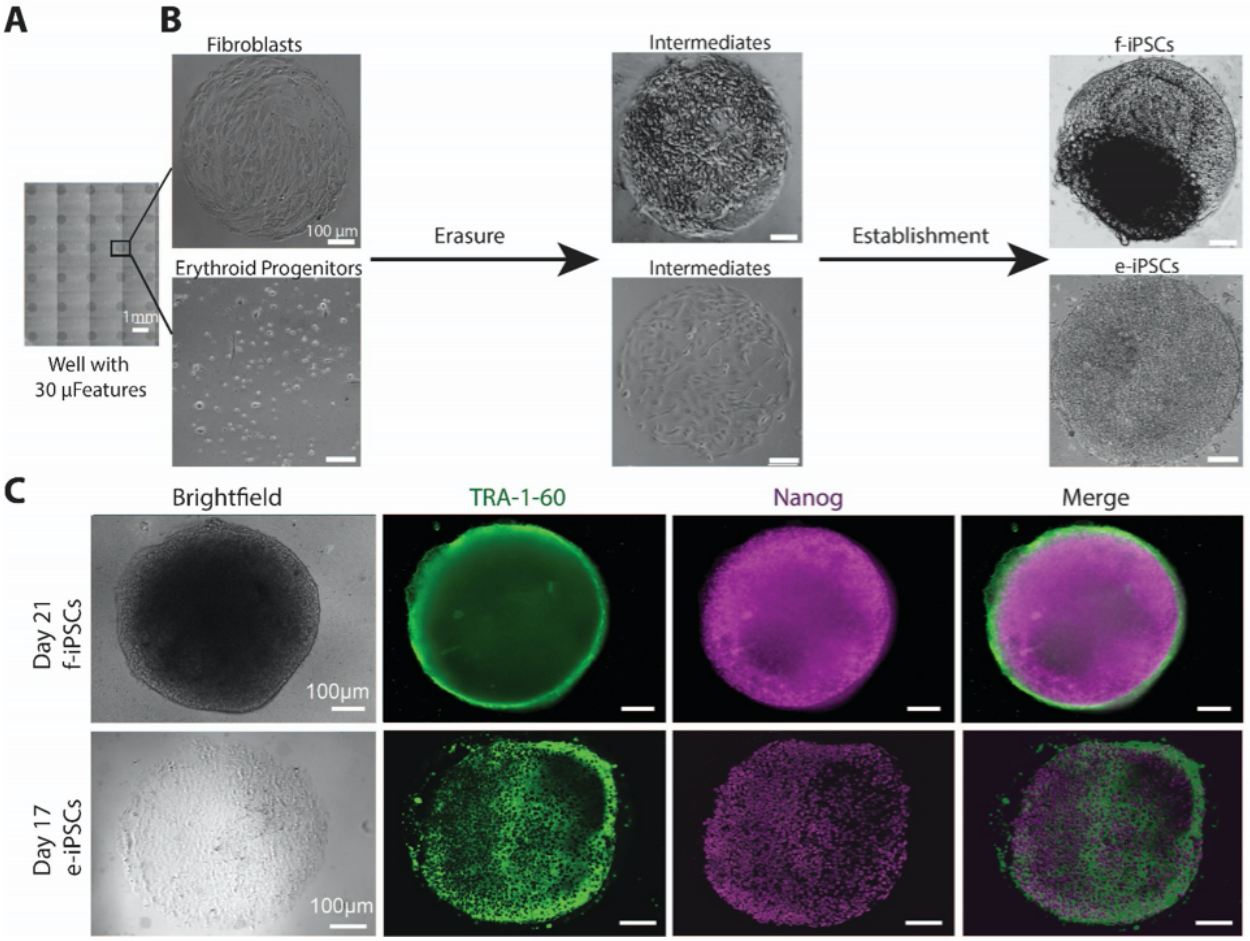


Figure 2.1: **Reprogramming fibroblasts and erythroid progenitor cells (EPCs) on micropatterned substrates.** A) Brightfield image of a portion of the µCP Well. B) Representative images of the progression of a human somatic cell subpopulation (fibroblasts or EPCs) on a single 300 µm radius circular µFeature through a reprogramming time course. C) Representative image of a 300 µm radius circular µFeature with f-iPSC (from fibroblasts) and e-iPSC (from EPCs) colonies, stained with TRA-1-60 and Nanog, markers of pluripotency.

We next assessed the ability of the µCP Well Plate to sustain long-term reprogramming studies that occur over several weeks. The high-content imaging capabilities enabled

us to track each individual cell subpopulation (>30 μFeatures per well in a 24-well plate) longitudinally at multiple time points [191, 193]. Starting fibroblasts, reprogramming intermediates, and endpoint iPSCs remained viable, attached, and confined to the desired μFeatures during the erasure of somatic cell identity and the establishment of pluripotency network over a 3-4 week reprogramming time course (**Figure A.1B**). For human EPCs undergoing reprogramming, these substrates permitted the adhesion of reprogramming intermediates and endpoint iPSCs over a 2-3 week time course, allowing for the tracking of middle to late stages of reprogramming, including the establishment stage. The day 21 and day 17 time points were picked to be the reprogramming endpoints for fibroblasts and EPCs respectively as there were several iPSC colonies per well at these time points without significant overgrowth within a μFeature, which can make it difficult to analyze single cell behavior. When μFeatures were stained at these reprogramming endpoints for the iPSC markers, Nanog and TRA-1-60 [41, 194], we found that there were several marker-positive iPSC colonies on the μFeatures (**Figure 2.1C**). The iPSC colonies obtained from fibroblasts and EPCs are termed as "f-iPSC" and "e-iPSC", respectively. In separate experiments that co-cultured human pluripotent stem cells (hPSCs) with fibroblasts, we ensured strong adhesion of each cell type for more than one month (**Figure A.1B**). Thus, one cell type does not displace or cause detachment of one cell type over another in our platform. Using a previously established protocol, we were also able to track live reprogramming cell subpopulations on μFeatures with antibodies targeting cell surface markers identifying fibroblasts and pluripotent cells [63, 160, 195] (**Figure A.1C**) and found that fibroblasts ($CD44^+$/TRA-1-60^-), f-iPSCs ($CD44^-$/TRA-1-60^+), and reprogramming intermediate ($CD44^-$/TRA-1-60^-) cells were readily detected on the same μFeature. These changes correspond to two distinct phases of reprogramming, the erasure of somatic cell identity to a reprogramming intermediate cell state followed by the establishment of pluripotency (**Figure A.1C**). These results demonstrate that the micropatterned substrate can impose physical constraints on each cell population over the entire course of reprogramming for

fibroblasts and during the establishment stages for reprogramming EPCs.

Establishment of iPSCs on micropatterned substrates

We next explored the effect of our μCP Well Plates on the establishment of pluripotency, a stage where cells are adherent and constrained on the micropatterned substrates. μFeatures with fibroblasts and EPCs undergoing reprogramming were fixed at their respective reprogramming endpoints and then immunostained for Nanog, a bona-fide marker of fully-reprogrammed iPSCs [49, 196]. We found that there was a wide range of endpoint Nanog percentages (the percentage of the total number of nuclei within a single μFeature positive for Nanog expression, termed "endpoint Nanog$^+$ percentage") within a single μFeature (**Figure 2.2**). Furthermore, due to the 3D nature of reprogrammed μFeatures, we tested whether the endpoint Nanog$^+$ percentage varied within a μFeature at several distances away from the substrate (z-axis) and found no significant changes in this z-direction (**Figure A.2A,B**). With the imaging plane parallel to the substrate, we then studied the relationship between the area of the circular μFeature and the endpoint Nanog$^+$ percentage of the μFeature. For fibroblasts undergoing reprogramming, a significant change in the median endpoint Nanog$^+$ percentage was observed only between μFeatures of 400 μm and 700 μm radii (**Figure 2.2B**). For EPCs undergoing reprogramming, the μFeatures had similar median endpoint Nanog$^+$ percentages (**Figure 2.2C**) irrespective of the μFeature radius.

It has previously been shown that confinement of cells into different geometries causes different stress patterns and subsequently differing downstream effects [197]. To explore this effect, we modulated the geometry of μFeatures on our plates without changing their surface area (**Figure A.2C**). We seeded fibroblasts, started reprogramming, and measured the endpoint Nanog$^+$ percentage on each μFeature. We found that μFeatures had similar endpoint Nanog$^+$ percentages irrespective of the μFeature geometry (**Figure 2.2D**). We next tested the effect of changing initial fibroblast seeding density on endpoint Nanog$^+$ percentages. The endpoint Nanog$^+$ percentage increased after a critical seeding

density of 50,000 cells per well (**Figure 2.2E**).

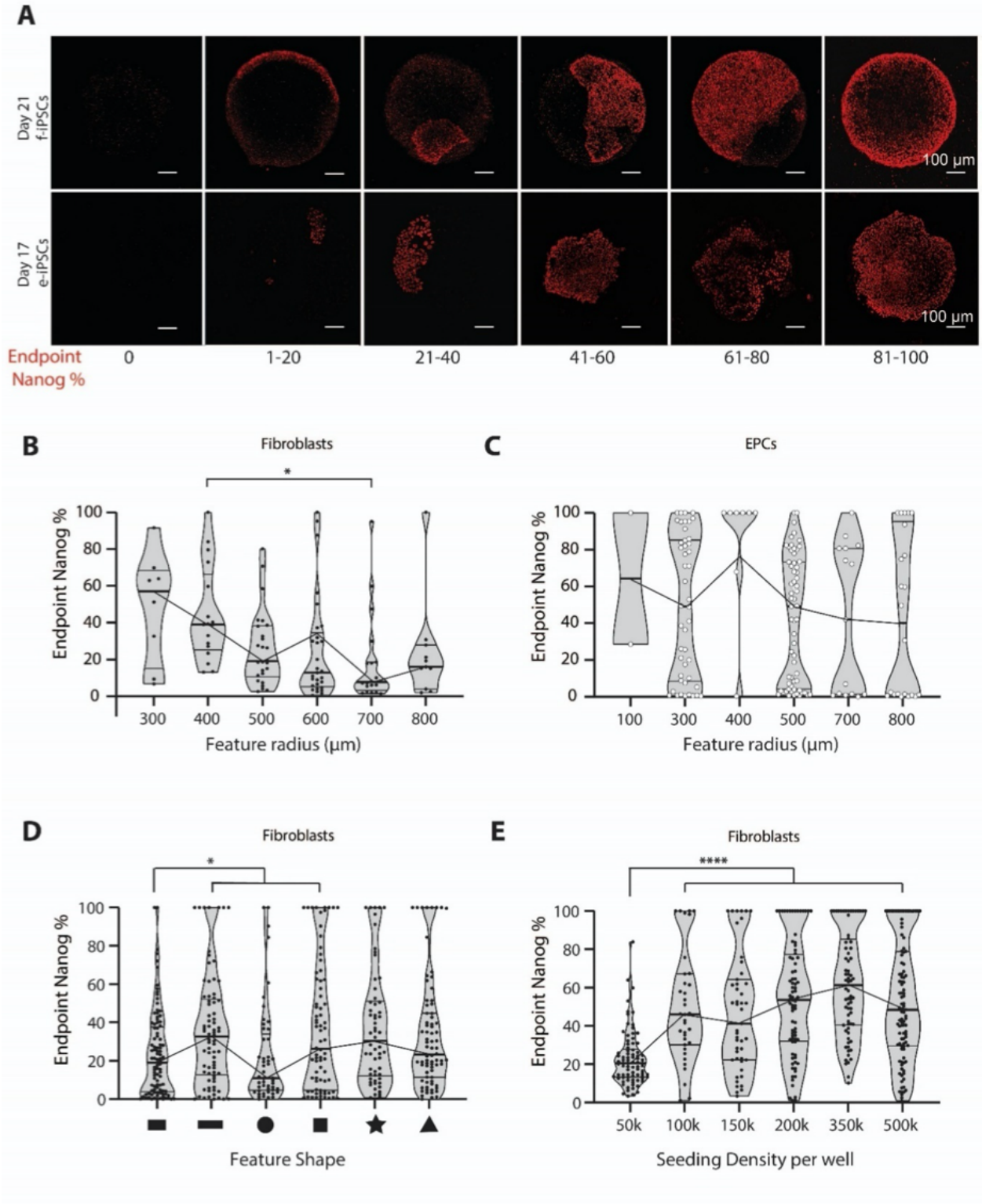


Figure 2.2: Caption next page

Figure 2.2: **A broad variety of reprogramming outcomes are captured on nearly all micropatterned substrates.** A) Representative images of 300 μm radius circular μFeatures with reprogramming fibroblasts and EPCs at their respective endpoints, classified into 20% endpoint Nanog$^+$ bins. B) Violin plots showing the distribution of endpoint Nanog$^+$ percentages as a function of the circular μFeature radius for reprogramming fibroblasts (n = 3 biological replicates, 36 technical replicates each). C) Violin plots showing the distribution of endpoint Nanog$^+$ percentages as a function of circular μFeature radius for reprogramming EPCs.μFeatures had similar median endpoint Nanog$^+$ percentages irrespective of the circular μFeature radius (n = 2 biological replicates, 36 technical replicates each). D) Violin plots showing the distribution of endpoint Nanog$^+$ percentages on differently shaped μFeatures for reprogramming fibroblasts. There was no significant difference in median endpoint Nanog$^+$ percentages between the six different shapes (n = 3 biological replicates, 100 technical replicates each). E) Violin plots showing the distribution of endpoint Nanog$^+$ percentages as a function of seeding fibroblast density (n = 3 biological replicates, 100 technical replicates each). Black and white dots represent technical replicates with non-zero endpoint Nanog$^+$ percentage in fibroblasts and EPCs, respectively. *p*-values calculated using Brown-Forsythe and Welch one-way ANOVA: ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$).

Increased cell number may help with clustering and erasure of somatic cell identity resulting in higher endpoint Nanog$^+$ percentages. While the above results suggest that there may be some cell type-dependent inherent characteristic of area available for cell attachment that influences endpoint Nanog$^+$ percentage, we did not find a strong consistent effect across the cell types by modulating substrate patterning. Therefore, we focused our subsequent studies instead at the inter-μFeature heterogeneity (**Figure 2.2A**), where cells varied greatly in morphology and reprogramming outcome.

Nuclear characteristics to track cell state

In our immunocytochemistry images of cells undergoing reprogramming, we noted dramatic changes in shape, size, and clustering of nuclei throughout the reprogramming process. For example, fibroblast nuclei are elongated and far apart from each other, whereas hPSC nuclei are circular and close together. We hypothesized that these changes could be used as a tool to track progression of cells through the reprogramming process. As a proof-of-concept for distinguishing cell state by tracking nuclei on a single μFeature, we seeded three distinct cell types on to the μCP Well Plates: human pluripotent stem cells (hPSCs), fibroblasts, and neural progenitor cells (NPCs), and used Hoechst dye to track nuclear characteristics through high-content imaging. The images were then used as inputs for a high-content analysis CellProfiler software pipeline [198] to identify nuclei within the images and output a set of geometrical, intensity, and clustering measurements (**Figure 2.3A**). We started with a large dataset of 33 nuclear characteristics (**Table A.1**) that was then filtered to a set of 8 core characteristics by evaluating correlation between measured variables. These core characteristics are area, perimeter, mean radius, nuclear shape index (NSI), extent, solidity, nearest neighbor, and number of neighbors. We initially attempted to identify different cell types by analyzing individual nuclei present in a single circular μFeature of 300 μm radius using two methods: principal components analysis (PCA) [199] and t-distributed stochastic neighbor embedding (t-SNE) [200]. However, neither method could faithfully distinguish the three different cell types from each other (**Figure 2.3C**). We next turned to analyzing nuclear characteristics on a per-μFeature basis (i.e., average nuclear characteristics of all the nuclei within a single μFeature) for each cell type using the same methods rather than analyzing the nuclear characteristics on a per-cell basis. Cell types analyzed at a μFeature level separated cleanly using both methods but were more highly clustered using PCA (**Figure 2.3D**). The multicellular nature of a μFeature includes information on cell-to-cell distances (e.g., clustering) and packing of cells that can influence nuclear shape and size. Based on the results in **Figure 2.3**, we found that these

multicellular measurements contain important information in distinguishing cell types, and thus we proceeded with this method for investigating cell types generated during reprogramming.

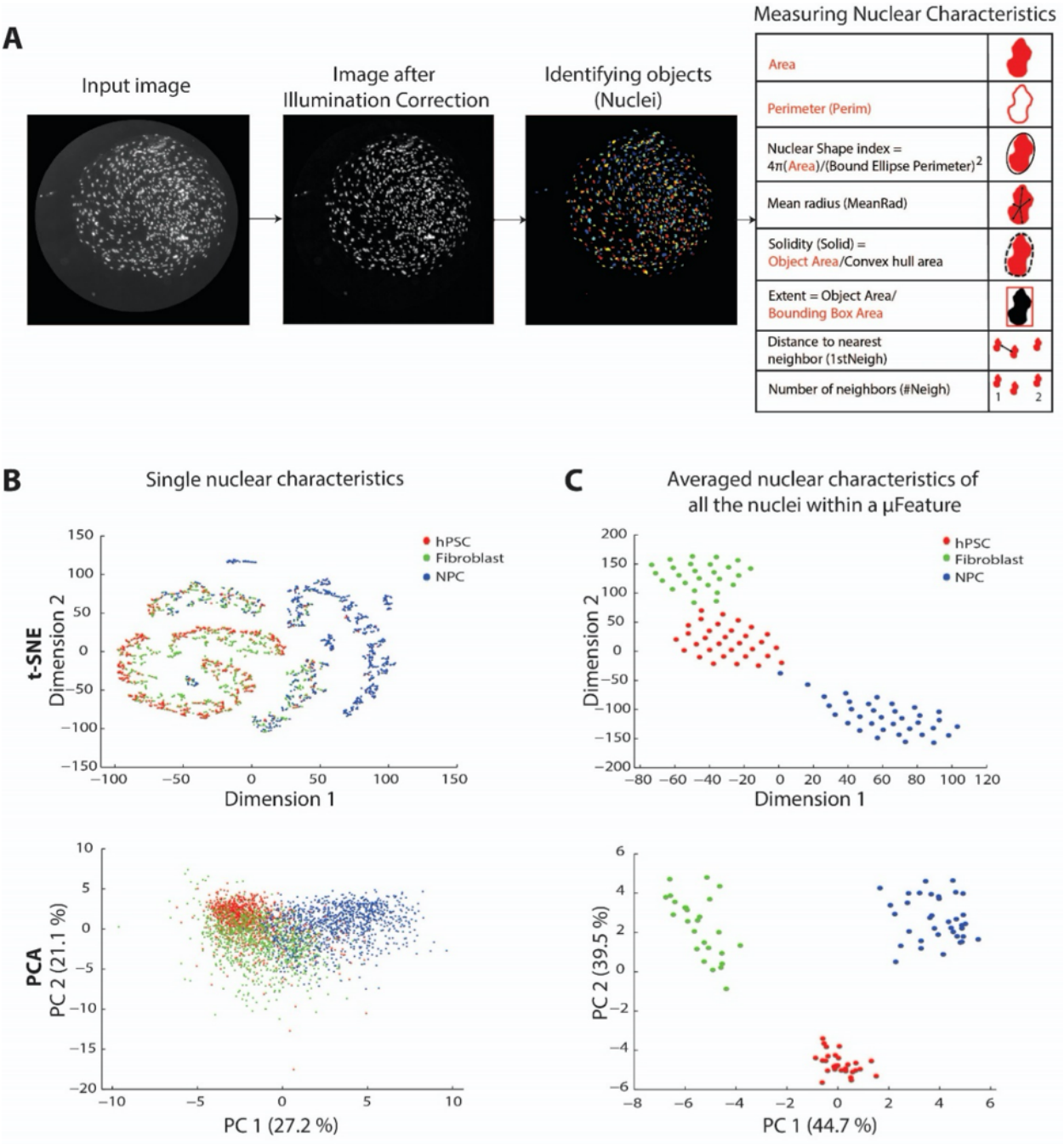


Figure 2.3: Caption next page

Figure 2.3: **Average nuclear characteristics of μFeatures can separate cell types better than single cell nuclear characteristics.** A) CellProfiler pipeline to identify eight core nuclear characteristics. B) Separation of cell types using single cell characteristics with t-SNE (top) and PCA (bottom) algorithms. No distinct clusters were identified. Each dot represents nuclear characteristics of a single nucleus within a μFeature (n = 1000 cells for each cell type). C) Separation of cell types using nuclear characteristics averaged over μFeatures. Distinct cell populations were visible using t-SNE and PCA. Each dot represents averaged nuclear characteristics of the nuclei within a single μFeature (n = 28-35 μFeatures analyzed for each cell type). Percent variance explained by the individual PC components is indicated within the parentheses.

Changes in nuclear characteristics during reprogramming

To determine whether reprogramming cells can be tracked using nuclear characteristics on a per-μFeature basis, we stained the nuclei of control hPSCs and fibroblasts undergoing reprogramming with Hoechst dye on 300 μm radius circular features at four intermediate time points (day 4, 8, 12, and 20) as well as at the beginning (day 0) and endpoint of reprogramming (day 21). When nuclear characteristics were plotted on a per-μFeature basis for each time point in a reduced three-dimensional PCA map, a progression from fibroblasts to f-iPSCs emerged when using three principal components (PCs) (**Figure 2.4A**). The three PCs captured 95% of the variance in the nuclear characteristics (**Figure 2.4C**). By analyzing the loadings of each PC, we found that PC1 is largely driven by variables describing nuclear shape (extent, perimeter, solidity, NSI; henceforth referred to as "shape-PC1") while PC2 corresponds largely to changes in nuclear size (area, mean radius; henceforth referred to as "size-PC2") (**Figure 2.4D**). In contrast, PC3 is dominated by shifts in nuclear clustering (nearest neighbor, number of neighbors; henceforth referred to as "clustering-PC3"). High values of shape-PC1, size-PC2, clustering-PC3 indicate circular,

large-sized, less clustered nuclei while low values of shape-PC1, size-PC2, clustering-PC3 indicate non-circular, small-sized, highly clustered nuclei, respectively.

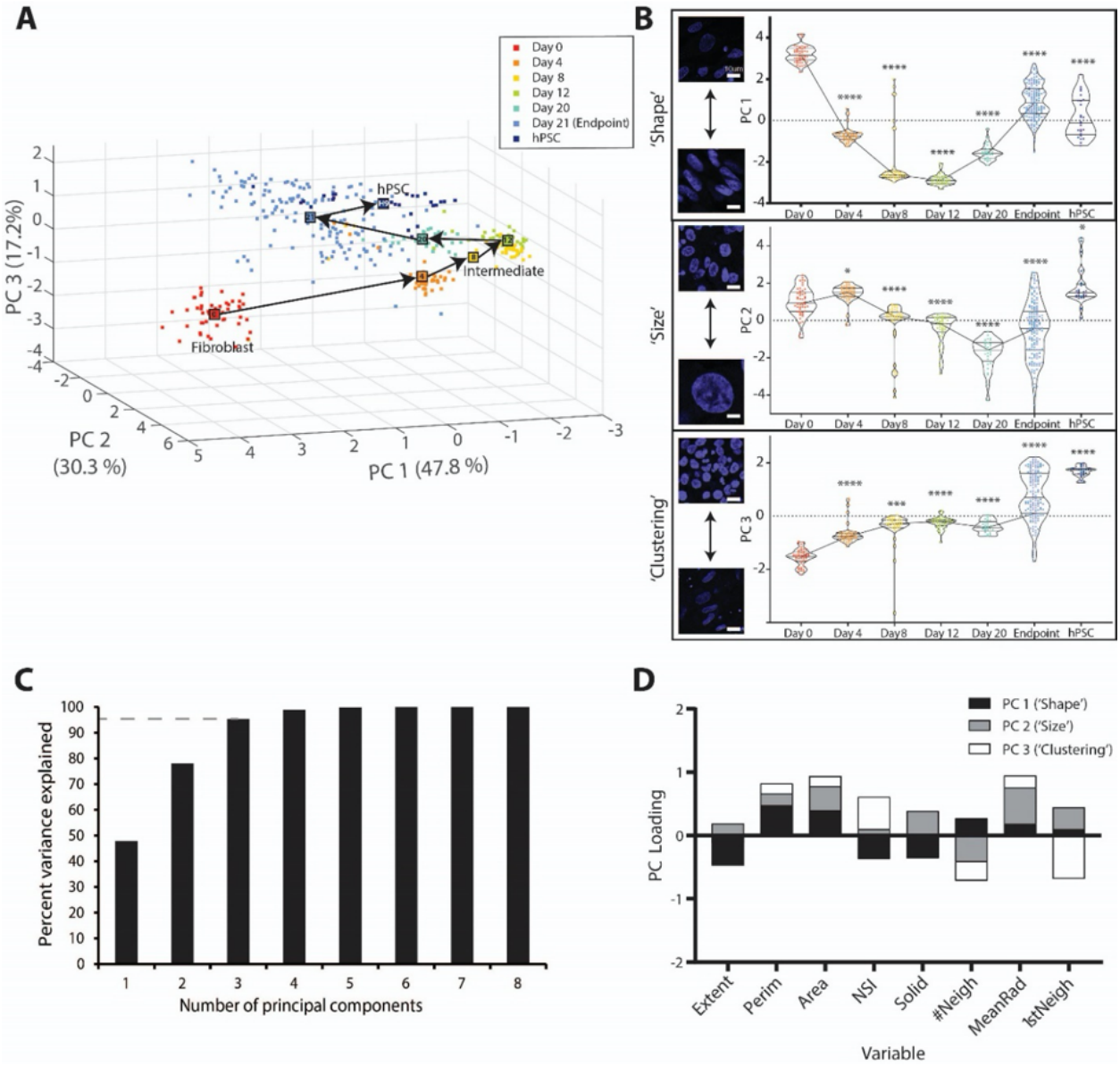


Figure 2.4: Caption next page

Figure 2.4: **Changes in nuclear characteristics of fibroblasts during reprogramming.** A) Principal component analysis of per-μFeature nuclear measurements on human fibroblasts undergoing reprogramming. Centroid values for each reprogramming time point (large boxes color coded by the time point) indicate a 3D spiral of reprogramming progression. Model was generated for circular μFeatures of 300 μm radius using 49 technical replicates (small boxes color coded by the time point) at each time point. Percent variance explained by the individual PC components is indicated within the parentheses. B) Violin plots showing the distribution of individual PCs as a function of the reprogramming time point for fibroblasts undergoing reprogramming (49 technical replicates at each time point represented by color coded boxes and 153 technical replicates at Day 21, *p*-values generated by one-way ANOVA using the Brown-Forsythe and Welch test for multiple comparisons to Day 0; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$). Representative nuclei with high and low PC values are shown next to each axis. C) Three PCs explained 95% of the variance in the PCA model while including additional components had diminishing returns. D) Stacked column bar graph showing loadings of the PCA model for each of the first three PCs. The height of the bar represents the loading value. Based on the loading, PCs can be roughly broken down to correspond to shape, size, clustering of the nuclei.

Upon focusing on the centroids of per-μFeature nuclear characteristics at each time point (represented as *large boxes*), we observed a spiral trajectory in the progression of cell state from fibroblast through erasure to reprogramming intermediate and finally establishment to f-iPSC (**Figure 2.4A**). Moreover, day 0 fibroblasts are the farthest away from any other population. Endpoint μFeatures clustered closer to hPSCs than any of the reprogramming intermediates at day 4, 8, 12, and 20. Additionally, there was a biphasic progression in shape-PC1 and size-PC2, while clustering-PC3 progressed in a monotonic fashion (**Figure 2.4B**).

The largest shifts in PCA map occur between days 0-8 corresponding with erasure of fibroblast identity. Temporally, this shift matches closely with the well-studied MET that occurs within the first 10 days of reprogramming and involves dramatic shifts in cell morphology [181–183]. Following erasure, there was a clustering of time points between Day 8-20 corresponding to the rise of a reprogramming intermediate cell state ($CD44^-$/TRA-$1\text{-}60^-$) (**Figure A.1C**). A final transition corresponding to establishment of pluripotency occurred between day 20 and the endpoint of reprogramming. Nuclei following establishment are highly circular, densely packed and cluster closely to hPSCs in the PCA map. To further confirm our PCA map, hierarchical clustering was used to determine the distance among experimental time points as well as variables. As seen in the PCA map, solidity, extent, and NSI are closely related, as are the number of neighbors and distance to the nearest neighbor (**Figure A.3**).

At the endpoint of reprogramming, there was a wide range of successfully reprogrammed μFeatures as indicated by the wide range of endpoint $Nanog^+$ percentages (**Figure 2.2A**). When nuclear characteristics of endpoint μFeatures were plotted on a per-μFeature basis in the PCA map, μFeatures that showed no reprogramming (0% $Nanog^+$) clustered the closest to the four intermediate time points and farthest from the hPSCs as compared to the other endpoint μFeatures (1-100% $Nanog^+$). Conversely, endpoint μFeatures that highly reprogrammed (81-100% $Nanog^+$) clustered the closest to hPSCs (**Figure 2.5A**). Overall, the endpoint μFeatures clustered more closely to hPSCs as their $Nanog^+$ percentage increased. When we plotted the individual PC coordinates as a function of the endpoint $Nanog^+$ percentage in the μFeature (**Figure 2.5B**), we found that shape-PC1 underwent a biphasic change, while size-PC2 and clustering-PC3 changed in more of a monotonic fashion, resembling the changes that occurred in these PCs from Day 12 onwards during the late/establishment phase of reprogramming (**Figure 2.4B**). Put together, this suggests that cells within these μFeatures may have started the establishment of pluripotency but had variable latency periods at the intermediate cell stage.

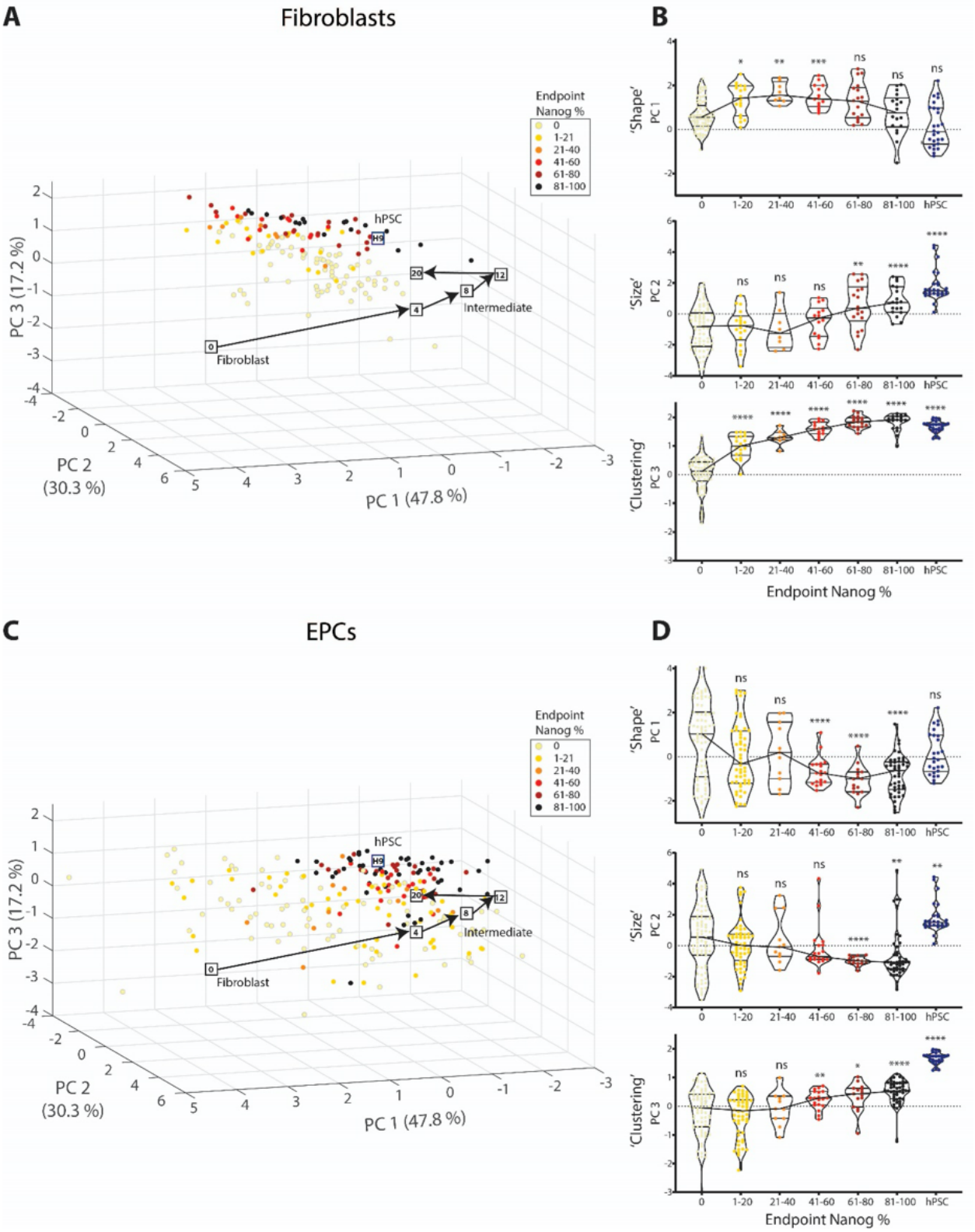


Figure 2.5: Caption next page

Figure 2.5: **Mapping endpoint heterogeneity in reprogramming across μFeatures for both fibroblasts and EPCs.** A) Fibroblast reprogramming PCA space (centroid values for each fibroblast reprogramming time point boxed as in Figure 4A) with endpoint reprogramming fibroblast data, for circular μFeatures of 300 μm radius, color coded by endpoint Nanog$^+$ percentage. Low Nanog$^+$ features clustered closely to time points of intermediate cell states while features that expressed high levels of Nanog clustered closely to hPSCs. Percent variance explained by the individual PC components is indicated within the parentheses. B) Violin plots showing the distribution of individual PCs as a function of the endpoint Nanog$^+$ percentage for reprogramming fibroblasts (n = 5 biological replicates, total number of endpoint μFeatures = 153, *p*-values generated by one-way ANOVA using the Brown-Forsythe and Welch test for multiple comparisons to 0% endpoint Nanog$^+$ percentage; ns = p $\geqslant$0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001). C) Fibroblast reprogramming PCA space (centroid values for each fibroblast reprogramming time point boxed as in Figure 4A) with endpoint reprogramming EPC data, for circular features with radius varying from 100-800 μm, color coded by endpoint Nanog$^+$ percentage. Low Nanog$^+$ features clustered closely to time points of intermediate reprogramming fibroblast states while features that expressed high levels of Nanog clustered closely to hPSCs. Percent variance explained by the individual PC components is indicated within the parentheses. D) Violin plots showing the distribution of individual PCs as a function of the endpoint Nanog$^+$ percentage for reprogramming EPCs (n = 10 biological replicates, total number of endpoint μFeatures = 229, *p*-values generated by one-way ANOVA using the Brown-Forsythe and Welch test for multiple comparisons to 0% endpoint Nanog$^+$ percentage; ns = p $\geqslant$ 0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001).

Next, nuclear characteristics of μFeatures (circular μfeatures with radius varying from 100-800 μm) were plotted on a per-μFeature basis for EPCs undergoing reprogramming in the PCA map that was previously generated for fibroblasts undergoing reprogram-

ming (**Figure 2.4A**), to investigate if the PCA map is dependent on the reprogramming system. We fixed and stained endpoint μFeatures on Day 18 for Nanog and found again that there was a wide range of outcomes (**Figure 2.2A**). We found that regardless of the reprogramming system and the μFeature radius, μFeatures that showed no reprogramming (0% $Nanog^+$) clustered the closest to the day 4 intermediate time point and the farthest to hPSCs when compared to the other endpoint μFeatures (1-100% $Nanog^+$). μFeatures that were more reprogrammed, 81-100% $Nanog^+$, clustered the closest to hPSCs (**Figure 2.5C**). Moreover, like in the case of fibroblasts undergoing reprogramming, the EPCs undergoing reprogramming clustered more closely to hPSCs as their endpoint $Nanog^+$ percentage increased. Additionally, the 0% $Nanog^+$ endpoint μFeatures in the case of EPCs undergoing reprogramming were more broadly distributed in the PCA map as compared to the 0% $Nanog^+$ endpoint μFeatures of fibroblasts undergoing reprogramming. We reasoned that the non-adherent nature of starting EPCs and the variability associated with the episomal reprogramming system used for EPCs might contribute to this phenomenon.

When we plotted the individual PCs as a function of the endpoint $Nanog^+$ percentage for EPCs undergoing reprogramming (**Figure 2.5D**), we observed that shape-PC1 and size-PC2 decreased in a roughly monotonic fashion while clustering-PC3 increased in a monotonic fashion. These PC changes are observed even when 300 μm radius μFeatures were exclusively considered (**Figure A.4B**). The decrease in shape-PC1 and increase in clustering-PC3 for EPCs undergoing reprogramming were similar to the changes that occurred with fibroblast-derived cells at higher levels of reprogramming (endpoint $Nanog^+$ percentage from 41 to 100%) (**Figure 2.5B**). In contrast, μFeatures in the 0-40% endpoint $Nanog^+$ range were more different between fibroblasts and EPCs. This could be because 1) starting/early intermediate cells are adherent versus nonadherent for fibroblasts versus EPCs, respectively, and 2) EPCs are closer to hPSCs on the Waddington epigenetic landscape than fibroblasts [201]. Size-PC2 decreased for EPCs while it increased for fibroblasts as the endpoint $Nanog^+$ percentage of the μFeatures increased (**Figure 2.5D**). The nuclear size

of starting EPCs is larger than fibroblasts (6 μm versus 3.5 μm diameter), and we find that nuclear size changes during reprogramming diverge across these two reprogramming systems, while the changes in shape and clustering during reprogramming are more similar across the two systems.

Put together, our results indicate that eight core nuclear characteristics of cells reprogramming on a μFeature can be used to determine the coordinates of that μFeature within the PCA map, and this location can further be used to estimate the percentage of $Nanog^+$ cells in that given μFeature.

Predicting reprogramming using nuclear characteristics

If analysis of nuclear characteristics can separate reprogramming intermediate cell states from one another, we reasoned that these same measurements would be able to predict which μFeatures were successfully reprogrammed at the end of the reprogramming time course. Creation of a prediction model could be used to aid in the rapid purification of high-quality iPSCs. We generated a partial least squares regression model (PLSR) using the eight minimal nuclear characteristics at the reprogramming endpoint, described previously as inputs and outputted the expected endpoint $Nanog^+$ percentage. The resulting three-component PLSR model explained 89% (**Figure 2.6A**) of the variance in endpoint reprogramming fibroblast cultures and relied heavily on NSI, distance to the nearest neighbor, and number of neighbors (**Figure 2.6B**). The three-component model was highly predictive of reprogramming with a root mean square error of prediction (RMSE) of 0.11 (**Figure 2.6C**). In the case of EPCs, the three-component PLSR model explained 70% (**Figure 2.6A**) of the variance in endpoint EPC cultures and relied heavily on neighbor characteristics, solidity, and NSI (**Figure 2.6B**). Again, the model was highly predictive of reprogramming with RMSE of 0.18 (Figure 6D). We compared these models to a third PLSR model using TRA-1-60 (a popularly used live cell surface marker for identifying iPSCs) expression levels as the inputs. This third model was not as predictive of reprogramming, with RMSE of

0.28 (**Figure A.5A**). We further confirmed that TRA-1-60 expression was not wholly predictive of Nanog expression by staining for both markers and observing significant staining differences between the two immunostained cell populations (**Figure A.5B**). TRA-1-60 expression is activated prior to Nanog during reprogramming [160], and many cells that express TRA-1-60 never transition to a fully pluripotent state [202].

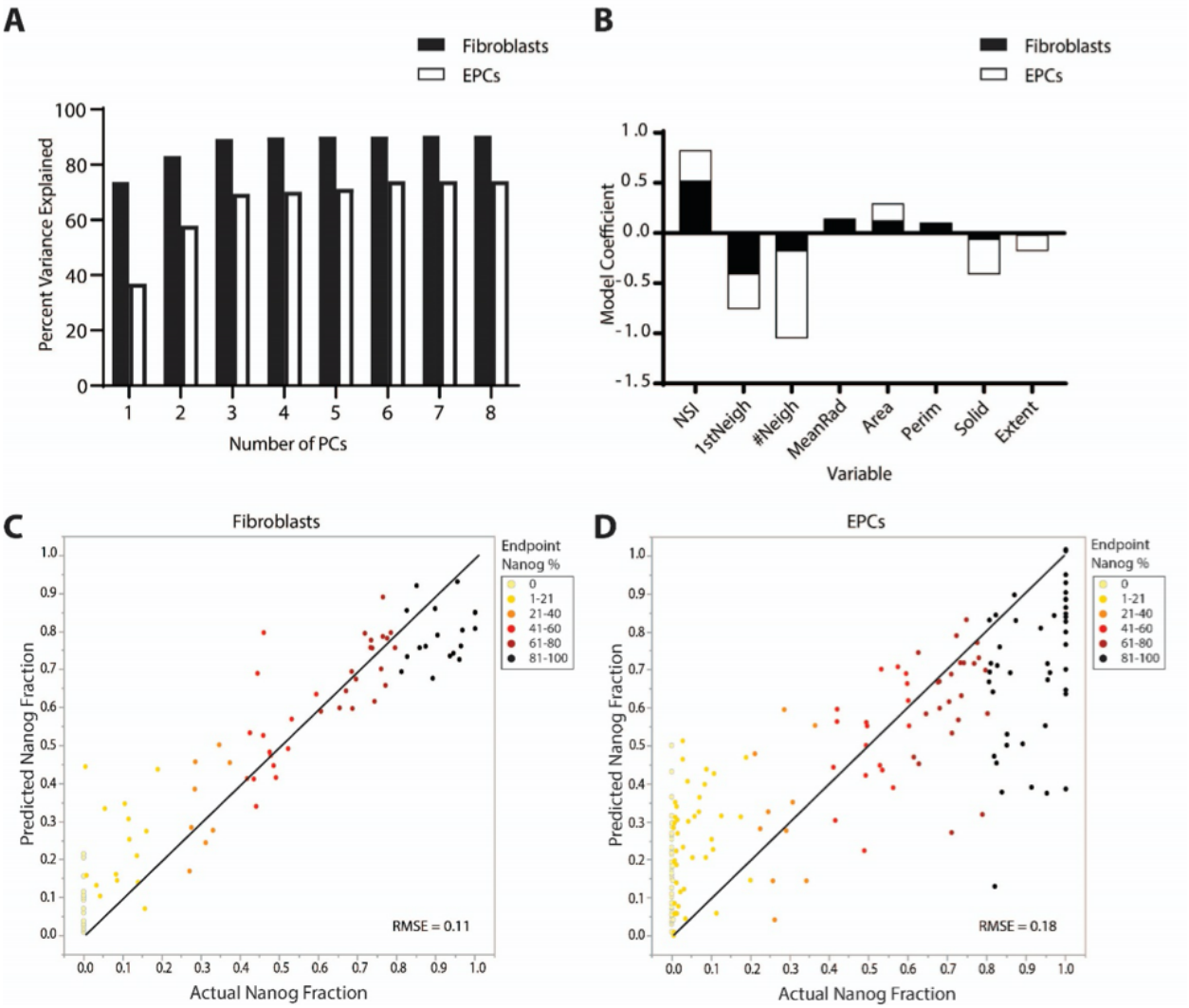


Figure 2.6: Caption next page

Figure 2.6: **Nuclear characteristics predict reprogramming outcomes at the reprogramming endpoint.** A) Percent variance explained by the principle components in the partial least squares regression (PLSR) model for reprogramming fibroblasts (black) and EPCs (white). Three principal components explained over 85% of the data for fibroblasts and over 70% of the data for EPCs while including additional PCs showed limited gains. B) Stacked column bar graph showing model coefficients of PLSR model for reprogramming fibroblasts (black) and EPCs (white). The height of the bar represents the value of the model coefficient. Nuclear shape and neighbor characteristics heavily influenced the fibroblast model. Neighbor characteristics, solidity, and NSI heavily influenced the EPC model. C) PLSR model for reprogramming fibroblasts predicting endpoint Nanog$^+$ fraction of μFeatures using only nuclear characteristics. Data points are individual μFeatures color coded by endpoint Nanog$^+$ percentage. Model was predictive of reprogramming with low error (RMSE=0.11) (n = 5 biological replicates, total number of endpoint μFeatures = 153). D) PLSR model for reprogramming EPCs predicting endpoint Nanog$^+$ fraction of μFeatures using only nuclear characteristics. Data points are individual μFeatures color coded by endpoint Nanog$^+$ percentage. Model was predictive of reprogramming with low error (RMSE=0.18) (n = 10 biological replicates, total number of endpoint μFeatures = 229).

Isolation of high-quality iPSCs

The terminal goal of any reprogramming platform is to successfully isolate iPSCs that can be used for downstream applications. Overall, reprogramming on μCP Well Plates enabled the simple isolation of high-quality iPSC lines with minimal time and effort spent on purification. The physical separation of micropatterns from one another, combined with high fraction of Nanog expressing cells, even up to 100%, throughout the μFeature resulted in easy picking and isolation of reprogrammed f-iPSCs (**Figure 2.7A**) and e-iPSCs

(**Figure 2.7B**). However, we observed that even the picking of low-quality f-iPSC colonies based on early TRA-1-60 expression (**Figure 2.7C**) resulted in the presence of Nanog$^+$ cells that could be isolated following one additional round of picking (**Figure 2.7D,E**). We confirmed that these cells expressed Nanog at the same level as control hPSCs using flow cytometry (**Figure 2.7F**). To stringently assess the pluripotency of established lines, we cultured cells isolated from two separate μFeatures for 10+ passages in fully-defined stem cell media and then formed embryoid bodies (f-EBs) using the AggreWell method. f-iPSCs and f-EBs were then subjected to the TaqMan Scorecard™ assay, a benchmarked quantitative assay for pluripotency [171]. f-iPSCs had an expression profile similar to nine control hPSC lines and f-EBs expressed genes indicative of all three germ layers (**Figures 2.7G, A.6A**). These results indicate that reprogrammed cell lines isolated from the μCP Well Plate are fully-pluripotent. Moreover, karyograms of fibroblasts, f-iPSCs, EPCs, and e-iPSCs exhibited a normal karyotype with no major chromosome abnormalities detected before and after reprogramming, indicating that this reprogramming platform can be utilized for generating genetically-stable iPSC lines (**Figure A.6B-E**). Additionally, cell cycle analysis (**Figure A.7**) shows that f-iPSCs have a higher percentage of cells in S phase than starting fibroblasts. Both e-iPSCs and EPCs have a high percentage of cells in S phase. The high percentage of cells in S phase indicates that f-iPSCs and e-iPSCs generated using our platform proliferate and divide rapidly, similar to iPSCs generated using standard methods [203].

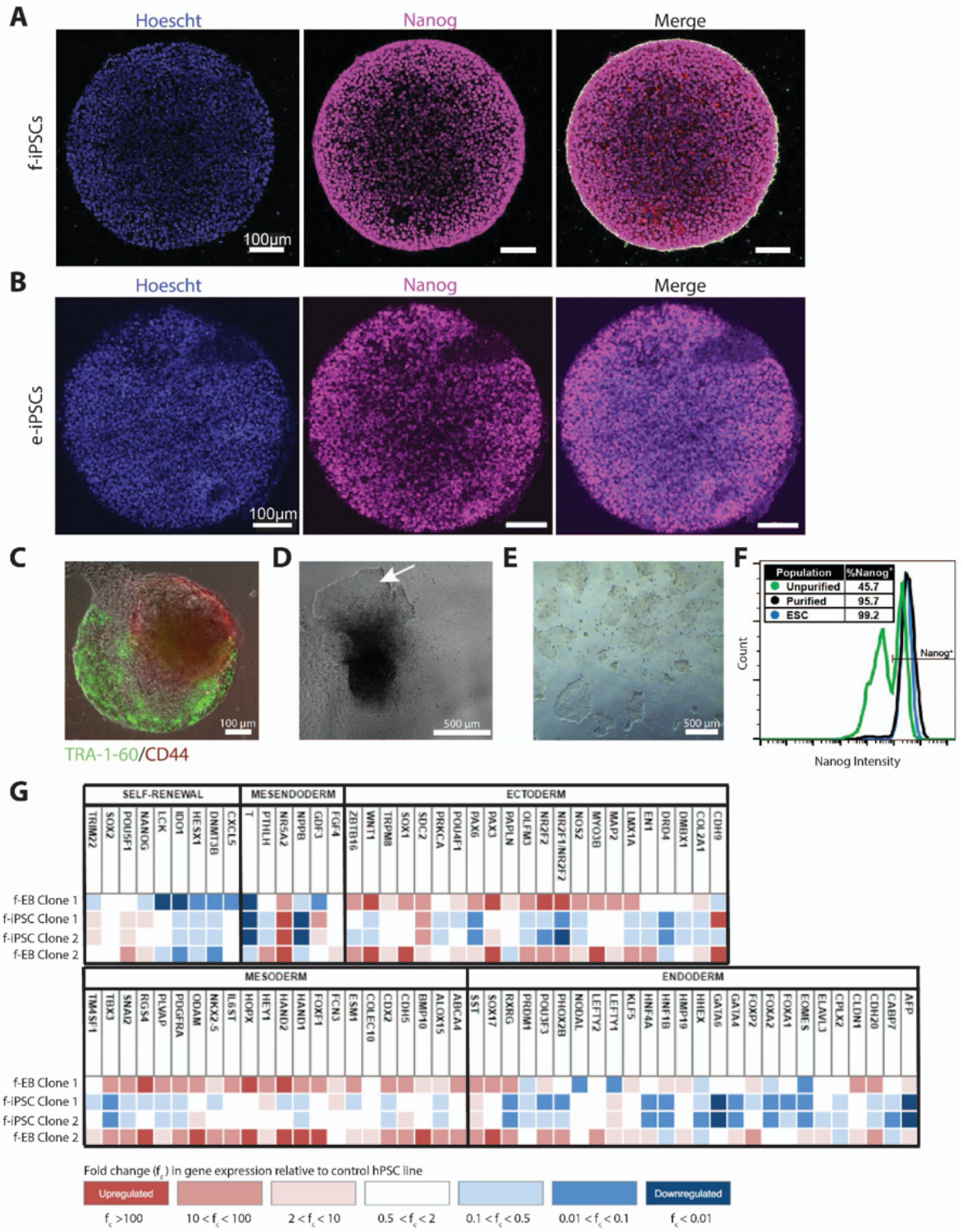


Figure 2.7: Caption next page

Figure 2.7: **μCP Well Plates enable isolation of high-quality iPSC populations.** A) Confocal image of 300 μm radius circular μFeature with reprogramming fibroblasts that underwent full reprogramming. B) Confocal image of 300 μm radius circular μFeature with EPCs that underwent full reprogramming. C-E) Representative images of the isolation of high-quality f-iPSC lines from μFeatures. C) Day 21 live cell subpopulation live-stained with CD44 (red) and TRA-1-60 (green) antibodies. D) Cell subpopulation from (C) three days after picking from a μFeature. White arrow represents expanding iPSC colony. E) f-iPSC line after one picking step from (D) and several passages in standard pluripotent culture conditions. F) Flow cytometry histogram indicating the percentage of $Nanog^+$ cells before (D) and after (E) one purification step. G) Taqman Scorecard™ of f-iPSCs and differentiated embryoid bodies (f-EB) from reprogramming. Two f-iPSC clones expressed pluripotency factors at the same level as nine benchmarked hPSC lines. EBs generated from f-iPSC lines expressed high levels of genes for all three germ layers, significantly above the benchmarked hPSC lines. Blue boxes are downregulated while red boxes are upregulated compared to 9 control hPSC lines.

2.4 Discussion

The longitudinal study of cell subpopulations within the microcontact printed substrates produced rich datasets for mapping cell fate transitions and trajectories within reprogramming cultures. The maps generated by our approach, using nuclear characteristics of reprogramming cells, are quantitative and reflect some of the complexity seen by flow and mass cytometry in mouse reprogramming cultures [158, 175]. While mouse and human reprogramming have important differences, nearly all studies describe erasure and establishment or maturation of PSCs. A large variety of intermediate cell states were also identified by flow and mass cytometry, which is consistent with the heterogeneity seen

in our maps, notably after erasure and during establishment. Stochasticity in reprogramming during this establishment stage has been attributed to variable reprogramming factor expression and degradation, proliferation rates, or the stochastic nature of remodeling various epigenetic marks. Simple expression of one to three live cell surface markers could be identified to distinguish various subpopulations within these maps, such as SSEA1 in mouse studies and TRA-1-60 in human studies. These strategies have been used to fractionate and enrich for iPSC colony identification within these heterogeneous cultures. Our regression model based on subpopulation nuclear analysis outperformed such a surface marker strategy (**Figure 2.6**), indicating that multidimensional analysis can be useful for isolating desired cell types from these cultures. Single-cell RNA-sequencing of these cell states [58, 187, 189, 190, 204, 205] are complementary and likely provide additional insights into mapping and understanding heterogeneity during reprogramming at the transcriptional level.

Microenvironmental signaling from cell-cell contact, mechanotransduction, and paracrine soluble factors within the μFeatures are likely reflected in nuclear morphology and organization. Such signaling can establish different "in vitro niches" with cell culture systems, and microenvironmental signaling is known to vary across discrete cell types described during stem cell differentiation [206–208] and somatic cell reprogramming [47, 209]. We utilized subpopulation nuclear morphology and clustering data to distinguish cell types more effectively than via single cell analysis (**Figure 2.3B,C**). Notably, clustering information of cells is lost during single cell analysis but was highly loaded in clustering-PC3 (**Figure 2.4D**) and had a high coefficient in the regression models (**Figure 2.6**). Cytoskeletal staining provided little additional information for our efforts to distinguish NPCs, fibroblasts, and hPSCs (data not shown), indicating that nuclear organization and morphology correlated with cytoskeletal changes. It is anticipated that additional stains or reporters of the cell [210–212] – plasma membrane, metabolic activity, mitochondria, focal adhesion, mitotic spindle, and chromatin mobility – could complement these maps and add addi-

tional information to distinguish each cell state. While detailed high-content analysis of each of these aspects of cells has potential to clarify the reprogramming map, our analysis sought to identify a minimal set of nuclear characteristics to distinguish iPSCs from somatic and intermediate reprogramming cell states. Because none of the measurements relied on the intensity of the nuclear stain, we anticipate a wide variety of live nuclear dyes could enable nuclear tracking to obtain these minimal measurements. These capabilities would complement other live staining strategies to mark highly dividing cells [160] and metabolic shifts [212] during reprogramming. Alternatively, we anticipate that label-free imaging [213] could also enable live nuclear tracking without the usage of any external dyes to stain the nucleus. We expect nuclear characteristics to be more reliable than visual identification of iPSCs as flat cell colonies with tightly-packed cells [41].

Nuclear geometry is dramatically altered during reprogramming [63, 191] Furthermore, the packing and mobility of chromatin shift during cell fate changes, notably at points during stem cell differentiation [59, 214] and at the beginning and end of reprogramming [54]. Watching these changes *in situ* during stem cell differentiation permits the isolation of specific cell types of defined potency [206, 215]. More recently, super-resolution imaging of nuclei has successfully forecast transitions within stem cell cultures [216]. Similar studies with our microcontact printed substrates could characterize these chromatin dynamics during reprogramming, notably during the heterogeneous establishment stage of reprogramming.

Fully-pluripotent iPSCs were readily isolated from the micropatterns. The presence of nearly 100% Nanog positive μFeatures with f-iPSCs and e-iPSCs was surprising (**Figure 2.7A,B**), given the overall low efficiencies of reprogramming when considering the entire well. This may reflect context-specific eliteness that was recently observed in murine reprogramming with barcoded cells [47]. Future work with fate mapping and clonal tracing [217–219] in our system is needed to confirm that the final cells arose from single or multiple cells that progressed through the establishment transition. Because

many cell types firmly adhered to the surface when we performed co-culture analysis (**Figure A.1**), we anticipate that some enhanced local reprogramming occurred on these μFeatures, although these studies were underpowered to make firm conclusions about such mechanisms that may include various roles for. cell adhesion, cell volume, actomyosin contractility and nuclear lamins. Utilization of the micropatterned substrates to probe these mechanisms in future work has strong potential to provide deeper insight in the biophysical and microenvironmental signaling involved in epigenetic reprogramming.

Our approach is complementary to other imaging modalities for dissecting complex reprogramming cultures [e.g., on cell surface markers [157, 173, 202], metabolic markers [220], or dye uptake [157, 221]] and could be implemented in a variety of formats [e.g., microfluidic [222], μShear [223]]. Notably, the microcontact printed substrate enables the non-destructive optical tracking of subcellular changes within reprogramming populations, in contrast to current methods to identify high-quality iPSCs which require that cells be fixed [e.g., immunocytochemistry], destroyed during the preparation of cell extracts [e.g., PluriTest [170]], or differentiated [e.g., embryoid body formation [169] and teratoma assays [168]]. Our method can also reduce the labor and time required for analyzing iPSC quality as multiple iPSC clones can be analyzed simultaneously on these microcontact printed substrates with minimal sample processing. Moreover, manufacturing micropatterned substrates is relatively easy and cheap. Only Hoechst dye and confocal microscopy are required for imaging, making it readily capable of being integrated into workflows for characterizing the quality of iPSCs in a cost-effective manner. Additionally, the starting fibroblasts, intermediate cells, and fully-reprogrammed endpoint f-iPSCs all adhere onto the micropatterned substrates with minimal detachment, thereby enabling novel control over cell adhesion during all stages of fibroblast reprogramming. For human EPCs, these substrates permit adhesion of intermediate cells and fully-reprogrammed endpoint e-iPSCs, and thus aid in the study of the establishment phase of reprogramming.

Limitations of the current micropatterning approach include 2D imaging, culture

duration, and potential issues with dye compatibility for live cell imaging. Although we did not develop significantly different predictions of reprogramming when using imaging data at different z-planes (**Figure A.2**), comprehensive 3D imaging of each μFeature could provide maps at higher resolution to further dissect the differences in morphological and clustering changes further away from the cell-substrate interface. Second, there is a limited duration of culture before cells overgrow the μFeature and potentially detach from the pattern, depending on the balance of cell-cell adhesion versus cell-substrate adhesion. For the reprogramming experiments described here, circular features with radii varying from 100-800 μm have been used for about ~20-40 days of culture, although the cell seeding density or micropatterned geometry could be easily changed [191, 224] to study other epigenetic reprogramming processes. Finally, live tracking applications may be limited for some cell lines, as imaging of live cells through uses of fluorescent dyes may cause DNA crosslinking, damage, and toxicity. Low dye concentrations and short incubation periods can potentially avoid these artifacts [205, 206, 225]. To mitigate these effects further, different dyes can also be tested [226] or label-free imaging can be used [213]. While we developed a quantitative model for predicting the Nanog positive iPSC colonies at the reprogramming endpoint, We anticipate that this work can be developed further to predict the ability of iPSC colony formation at earlier time points of reprogramming. We also envisage that this work can be utilized to study downstream cell differentiation, cell line development, or direct reprogramming processes to identify appropriate cells in a fully traceable and quantitative way for regenerative medicine and precision medicine [227–229] applications. Because skin and blood constitute the majority of cell sources for research and clinical-grade reprogramming projects and biobanks [230] and since these substrates dissect reprogramming cultures into small subpopulations, these substrates have strong potential to advance the manufacturing and isolation of high-quality iPSCs for a variety of disease modeling and regenerative medicine applications.

2.5 Materials and Methods

Cell culture and derivation of cell lines

Fibroblasts: Fibroblast reprogramming experiments were performed using the previously reported C1.2 human secondary fibroblast line (28, 45), which incorporates the stably integrated transgenes Oct4, Sox2, Klf4, and c-Myc on doxycycline (DXC)-inducible cassettes. All f-iPSCs were generated through DXC-mediated reprogramming of the C1.2 line.

Fibroblasts were maintained on gelatin-coated polystyrene tissue culture plates in fibroblast media containing DMEM-high glucose (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% Fetal Bovine Serum (Thermo Fisher Scientific), 1 mM L-glutamine (Thermo Fisher Scientific), 1% non-essential amino acids (EMD Millipore, Burlington, MA), and 1% Penicillin/Streptomycin (Thermo Fisher Scientific). They were passaged with 0.05% trypsin (Thermo Fisher Scientific) every 3-5 days.

Erythroid Progenitor Cells (EPCs): EPCs were isolated from fresh peripheral human blood that was obtained from healthy donors (Interstate Blood Bank, Memphis, TN). Blood was processed within 24 hours of collection, where hematopoietic progenitor cells were extracted from whole blood using negative selection (RosetteSep; StemCell Technologies, Vancouver, Canada) and cultured in polystyrene tissue culture plates in erythroid expansion medium (StemCell Technologies) for 9 days to enrich for EPCs. EPCs were then examined by performing flow cytometry analysis for CD71 (334107; Biolegend, San Diego, CA), erythroid cell surface marker. Human Pluripotent Stem Cells (hPSCs): Transgenic hPSCs constitutively expressing H2B-mCherry and LifeAct-GFP were generated via CRISPR/Cas9 introduction of the H2B-mCherry plasmid to the *AAVS1* locus and lentiviral transduction of LifeAct-GFP (#51010; Addgene, Watertown, MA) to the hESC WA09 line (WiCell Research Institute, Madison, WI). The H2B-mCherry plasmid was generated by cloning the H2B-mCherry sequence (#20972; Addgene) into the GFP sequence of the AAV-CAGGS-EGFP plasmid (#22212; Addgene) and was electroporated with Cas9 plasmids [18,

231]. Clonal isolation of homogeneous cell lines constitutively expressing H2B-mCherry and LifeAct-GFP was then performed [63, 191]. The resultant transgenic hPSCs were maintained on Matrigel-coated polystyrene tissue culture plates in E8 medium, formulated in-house according to an established recipe [25, 26]. E8 medium was fed daily and the cells were typically passaged using Versene EDTA (Thermo Fisher Scientific, Waltham, MA) every 3-5 days. During passaging, the culture media was often supplemented with the Rho kinase inhibitor Y-27632 (Sigma-Aldrich, St. Louis, MO) at a 10 μM concentration to encourage cell survival. All cells were maintained at 37ºC and 5% CO_2.

μCP Well Plate construction

Micropatterned substrates or μCP Well Plates were constructed as previously described [191]. In brief, large PDMS stamps were used with traditional microcontact printing techniques [224] to pattern a thin sheet of gold-coated glass (Coresix Precision Glass, Williamsburg, VA) which had been precut to the size of a standard tissue culture plate. Once patterned, the glass was then fastened to the bottom of a bottomless 24-well plate via medical-grade double sided tape (ARcare 90106; Adhesives Research, Glen Rock, PA). Bottomless well plates were made in-house by removing well bottoms of standard tissue culture plates (Fisher Scientific, Hampton, NH) using a laser cutter (Universal Lasers Systems) or purchased directly (Greiner Bio-One, Monroe, NC).

Reprogramming

Fibroblasts: C1.2 secondary fibroblasts were seeded onto micropatterned substrates, at a seeding density of 100k cells/well, in fibroblast media one day prior to reprogramming initiation. The following day, media was switched to E7 (E8 without TGF-β) supplemented with hydrocortisone (49), and DXC was added at 2 μg/mL (5 μM) to activate expression of the reprogramming factors. Media was changed every other day, and cells were passaged

before reaching full confluence. After the first passage, DXC was removed from E8 media and iPSCs were maintained in standard pluripotent culture as described above.

Erythroid Progenitor Cells (EPCs): Day 9 EPCs were electroporated with four episomal reprogramming plasmids encoding Oct4, shRNA knockdown of p53 (#27077; Addgene, Watertown, MA); Sox2, Klf4 (#27078; Addgene); L-Myc, Lin28 (#27080; Addgene); miR302-367 cluster (#98748; Addgene), using the P3 Primary Cell 4D-Nucleofector Kit (Lonza) and the EO-100 program. Electroporated EPCs were seeded onto micropatterned substrates with erythroid expansion medium (StemCell Technologies) at a seeding density of 1000k cells/well. Cells were supplemented with ReproTeSR (StemCell Technologies, Vancouver, Canada) on alternate days starting from Day 3 without removing any medium from the well. On Day 9, the medium was entirely switched to ReproTeSR, and the ReproTeSR medium was changed daily starting from Day 10.

Isolation of iPSCs

To isolate high-quality iPSC lines, candidate colonies were picked from micropatterns using a 200 μL micropipette tip and transferred to Matrigel-coated polystyrene tissue culture plates in mTeSR1 media (WiCell Research Institute, Madison, WI). If additional purification was required, one additional manual picking step with a 200 μL micropipette tip was performed. During picking and subsequent passaging, the culture media was often supplemented with the Rho kinase inhibitor Y-27632 (Sigma-Aldrich, St. Louis, MO) at a 10 μM concentration to encourage cell survival and establish clonal lines. iPSCs obtained from EPCs were maintained in mTeSR1 media on Matrigel-coated polystyrene tissue culture plates and passaged with ReLeSR (StemCell Technologies, Vancouver, Canada) every 3-5 days. All cells were maintained at 37ºC and 5% CO_2.

Antibodies and Staining

All cells were fixed for 15 minutes with 4% paraformaldehyde in PBS (Sigma-Aldrich, St. Louis, MO) and permeabilized with 0.5% Triton-X (Sigma-Aldrich) for >4 hours at room temperature before staining. Hoechst (H1399; Thermo Fisher Scientific, Waltham, MA) was used at 5 μg/mL with 15 min incubation at room temperature to stain nuclei. Primary antibodies were applied overnight at 4°C in a blocking buffer of 5% donkey serum (Sigma-Aldrich) at the following concentrations: CD44-PE (555479; BD Biosciences, San Jose, CA) 1:200; TRA-1-60 (MAB4360; EMD Millipore, Burlington, MA) 1:100; Nanog (AF1997; R&D Systems, Minneapolis, MN) 1:200; CD71 (334107; Biolegend, San Diego, CA) 1:100. Secondary antibodies were obtained from Thermo Fisher Scientific and applied in a blocking buffer of 5% donkey serum for one hour at room temperature at concentrations of 1:400 – 1:800. For live-cell stains, antibodies were added to the appropriate cell culture media at equivalent dilutions to those used for fixed cells. Two-hour incubations were used for primary antibodies, followed by 30-minute incubations for secondary antibodies.

High-content analysis

High-content image analysis was performed similarly to previously published methods [191]. A Nikon Eclipse Ti epifluorescence microscope was used to acquire single 10x images of each micropattern, and a Nikon AR1 confocal microscope was used to acquire 60x stitched images of each micropattern using the z-plane closest to the micropatterned substrate for reprogramming studies. Images were processed using image analysis software CellProfiler [198], which analyzed images as described in **Figure 2.3**. Objects $< 75\mu m^2$ in area were filtered out of the data set to exclude apoptotic or other debris, and neighbors were identified by expanding objects until all pixels on the object boundaries were touching one another. Two objects are neighbors if any of their boundary pixels are adjacent after expansion.

PCA and PLSR

CellProfiler nuclear measurements were averaged across each μFeature and fed as mean values into PCA analysis via MATLAB software and into PLSR analysis via JMP Pro software. PLSR was performed using the NIPALS method and predictions were tested and validated using K-Fold validation with 7 folds. Endpoint $Nanog^+$ percentages were obtained by calculating the total area of the nuclei in each micropattern expressing Nanog, multiplying by 100, and then dividing by the total area of the nuclei stained by Hoechst.

Statistics

p-values were calculated using Welch and Brown-Forsythe one-way ANOVA (considering unequal variances across different groups) or two-way ANOVA with GraphPad Prism software. Statistical tests were deemed significant at $\alpha \leqslant 0.05$. Technical replicates are defined as distinct μFeatures within an experiment. Biological replicates are experiments performed at distinct time points during reprogramming. Outliers were identified as 1.5*IQR and excluded from statistical test. Outlier data points are still shown for illustration. No a priori power calculations were performed.

Karyotype Analysis

Cells cultured for atleast 5 passages were grown to 60-80% confluence and shipped for karyotype analysis to WiCell Research Institute, Madison, WI. G-banded karyotyping was performed using standard cytogenetic protocols [232]. Metaphase preparations were digitally captured with Applied Spectral Imaging software and hardware. For each cell line, 20 GTL-banded metaphases were counted, of which a minimum of 5 were analyzed and karyotyped. Results were reported in accordance with guidelines established by the International System for Cytogenetic Nomenclature 2016 [233].

Cell cycle analysis

Cell cycle phases were obtained by propidium iodide (PI) staining followed by analysis with flow cytometry as previously described [234]. In brief, cells were grown to 60-80% confluence and fixed for two hours in 70% ethanol at 4°C. After ethanol fixation, cells were centrifuged at 900 x g for 5 minutes and ethanol was removed. Cells were washed twice with D-PBS buffer (Thermo Fisher Scientific), re-suspended in 1 mL D-PBS containing 0.5 µg/mL DNase-free, protease-free RNase A (Thermo Fisher Scientific) and 50µg/mL propidium iodide (Thermo Fisher Scientific), and then incubated for overnight at 4°C. Data collection was performed using Attune™ NxT Flow Cytometer. Cell cycle analysis was performed using ModFit LT software (Verity Software House, Topsham, ME) after excluding debris, doublets and clumps by gating on the BL2-H (PI-H) versus BL2-A (PI-A) plot.

3 LABEL-FREE OPTICAL METABOLIC IMAGING TO TRACK HUMAN SOMATIC CELL REPROGRAMMING

Work in this chapter was adapted from:

Label-free Optical Metabolic Imaging to Track Human Somatic Cell Reprogramming

Kaivalya Molugu, Giovanni A. Battistini, Tiffany M. Heaster, Jacob Rouw, Emmanuel C. Guzman, Melissa C. Skala, Krishanu Saha

Article in preparation.

3.1 Abstract

The process of reprogramming patient samples to human induced pluripotent stem cells (iPSCs) is stochastic, asynchronous, and inefficient leading to a heterogeneous population of cells. Hence, identifying and evaluating the quality of cells during reprogramming in real-time could provide new strategies to develop high-quality iPSC cultures and lines. Here, we developed a micropatterned platform for non-destructive, label-free autofluorescence live-cell imaging of NAD(P)H and FAD to track the reprogramming status of single cells. Erythroid progenitor cells(EPCs) isolated from human peripheral blood showed distinct patterns of autofluorescence lifetime and nuclear morphometry at different timepoints during reprogramming. Random forest models classified starting EPCs, partially-reprogrammed intermediate cells, and iPSCs with 95% accuracy and single-cell reprogramming trajectories provided insights into reprogramming heterogeneity. This combination of micropatterning, autofluorescence imaging, and machine learning provides a unique non-destructive method to assess the quality of iPSCs in real-time for various applications in regenerative medicine, cell therapy biomanufacturing, drug development, and disease modeling.

3.2 Introduction

Derivation of patient-specific induced pluripotent stem cells (iPSCs) from their somatic cells via reprogramming generates a unique self-renewing cell source for disease modeling, drug discovery, toxicology, and personalized cell therapies. These cells carry the genome of the patient, facilitating elucidation of the genetic causes of disease, and are immunologically matched to the patient, facilitating the engraftment of any cell therapies developed from these cells [235–237].

With several clinical trials planned and underway [238], there has been a significant progress in developing iPSC-based cell therapies in the recent years. However, several chal-

lenges remain before successful commercial translation and wide usage of iPSC technology [239]. Firstly, derivation of high quality iPSCs must be efficient, rapid, and cost-effective for ensuring that patients receive their treatments in a timely fashion. Secondly, ensuring the safety and quality of iPSCs is extremely important. Finally, iPSC-derived cell therapies require scalable and standardized manufacturing processes to overcome the inconsistencies arising from variability in starting materials, process reagents, microenvironmental fluctuations, and stochasticity of the reprogramming process [240]. Current typical assays used for quality control of GMP (good manufacturing practices)-grade iPSCs include testing for cell line identity (STR analysis, SNP analysis, genomic sequencing), genomic instability (G-banding, chromosomal microarray, Nanostring technology), pluripotency (marker expression analysis via flow cytometry or immunochemistry, embryoid body analysis, teratoma assays, Pluritest, TaqMan Scorecard Assay) and residual reprogramming factors (PCR or immunochemistry) [229, 241–243]. However, these methods are low-throughput, labor-intensive, time-consuming, and require destructive processing. While recent studies have indicated that automated machine learning can be used identify cell structures from label-free brightfield images that cannot be manually identified [244–248], non-destructive automated deep learning methods to identify iPSCs based on cellular morphology have had limited success in the field [160, 163, 164]. Hence, it is important to develop new standardized platforms for biomanufacturing of iPSCs that are integration-free, fast, efficient, scalable, and easily transferrable to GMP-compliant conditions.

Previous studies indicate that somatic cells undergo dramatic metabolic changes throughout reprogramming. Somatic cells primarily utilize mitochondrial oxidative phosphorylation (OXPHOS) for proliferation. However, pluripotent cells favor glycolysis, in a manner reminiscent of the Warburg effect in cancer cells [64, 73]. During reprogramming, somatic cells thus undergo a metabolic shift from OXPHOS to glycolysis [65, 66], triggered by a transient OXPHOS burst, resulting in initiation and progression of reprogramming to iPSCs [67–69]. Recent evidence also indicates that this metabolic shift occurs

prior to changes in gene expression and that the modulation of glycolytic metabolism or OXPHOS alters reprogramming efficiency [73–75]. This suggests that cell metabolism plays an important role in determining cell fate rather than passively responding to cell fate changes. High-resolution imaging of reprogramming cells has identified that nuclear geometry is also dramatically altered during reprogramming [54, 63, 249]. Therefore, monitoring metabolic and nuclear changes during reprogramming could reveal insights into the reprogramming process and aid in the subsequent identification of iPSCs.

Optical Metabolic Imaging (OMI) is a non-invasive and label-free two-photon microscopy technique that can provide dynamic measurements of cellular metabolism at a single-cell level. It is based on the quantification of endogenous fluorescence of metabolic coenzymes, NADH and FAD [250], which are both used across several cellular metabolic processes. NADH and NADPH have overlapping fluorescence properties and are collectively referred to as NAD(P)H [251]. The optical redox ratio, defined as the ratio of NAD(P)H intensity to total NAD(P)H and FAD intensity, provides a measure of the relative oxidation-reduction state of the cell [252, 253]. Fluorescence lifetime imaging microscopy (FLIM) of NAD(P)H and FAD provides additional information specific to protein binding activity. The two-component decays of NAD(P)H and FAD measure the short (τ_1) and long (τ_2) fluorescence lifetimes that correspond to the free or bound states of these coenzymes [254–256], along with the fractional contributions of short (α_1) and long lifetimes (α_2). Since NAD(P)H and FAD are found predominantly in the cytoplasm, the lack of fluorescence signal in the images can also be used to identify cell borders and nuclei [257]. Thus, OMI provides multiple readouts for cell metabolism and nuclear morphometry to track metabolic and nuclear changes of cells undergoing reprogramming.

Here, we address some of the challenges associated with the biomanufacturing of iPSCs by developing a facile microcontact printed (μCP) platform [191, 249, 258] to non-invasively monitor metabolic and nuclear changes during 25-day reprogramming of erythroid progenitor cells (EPCs) to iPSCs using the combination of OMI with quantitative

image analysis. With this study, we demonstrated that OMI is sensitive to the metabolic and nuclear differences during reprogramming, performed accurate classification to determine the reprogramming status of cells using machine learning models, and subsequently built single-cell reprogramming trajectories [259]. Our label-free, non-destructive, rapid, scalable method to track reprogramming could provide a valuable resource for the scientific community towards developing a standardized procedure to derive iPSCs that are putatively GMP compliant.

3.3 Results

Establishment of reprogramming on microcontact printed substrates

We first designed a microcontact printed (μCP) substrate to spatially control the adhesion of EPCs undergoing reprogramming [191, 193, 249]. The μCP substrate is formed by Matrigel coating of 300μm radius circular regions, referred to as μFeatures, on a 35mm ibiTreat dish that allows for cell adhesion. Remaining regions of the dish are then backfilled with polycationic graft copolymer, PLL-g-PEG, that resists protein adsorption and prevents cell adhesion in these regions [260, 261] (**Figure B.1A**).The ibiTreat dishes are made of gas-permeable material, enabling maintenance of carbon dioxide or oxygen exchange during cell culture, and have high optical quality. These properties make the dishes suitable for two-photon microscopy during reprogramming. To verify the stability of the Matrigel-coated circular regions, we immunofluorescence labeled for laminin which is a major component of Matrigel [**hughes_Matrigel_2010**]. Fluorescence imaging showed laminin consistently within the circular μFeatures indicating uniform patterning of Matrigel (**Figure 3.1A**). We next assessed the ability of the μCP substrates to enable cell attachment by seeding two different cell types, i.e., normal human dermal fibroblasts (NHDFs) and H9 human embryonic stem cells (H9 ESCs). We observed that both NHDFs and ESCs remained viable, attached, and confined to the circular μFeatures indicating that the μCP substrates enable

spatial control of cell attachment (**Figure B.1B**).

Next, we isolated peripheral blood mononuclear cells (PBMCs) from peripheral blood of healthy human donors and further enriched them for EPCs. We examined the enrichment of EPCs by flow cytometry with erythroid cell surface marker CD71 [262]. Flow cytometry confirmed the presence of enriched EPCs showing that >98% of the cells expressed CD71 on day 10 of culture (**Figure B.1C**).

We electroporated the EPCs with four episomal reprogramming plasmids, encoding Oct4, shRNA knockdown of p53, Sox2, Klf4, L-Myc, Lin28, and miR302-367 cluster; and seeded them onto μCP substrates [33, 192]. We assessed the ability of the μCP substrates to sustain long-term reprogramming studies by performing high-content imaging to track individual μFeatures (>30 μFeatures per 35mm dish) longitudinally at multiple timepoints over the 3-week reprogramming time course. Day 22 was picked as the reprogramming endpoint because there were several iPSC colonies at this timepoint without significant outgrowth within a μFeature, enabling analysis at single-cell resolution. While starting EPCs are non-adherent, reprogramming intermediates (IMs) and endpoint iPSCs adhere to the circular μFeatures within the μCP substrates (**Figure 3.1A**) indicating that μCP substrates can support reprogramming of EPCs. Overall, the μCP platform provides unique spatial control over reprogramming cells and enables high-content quantitative imaging of reprogramming.

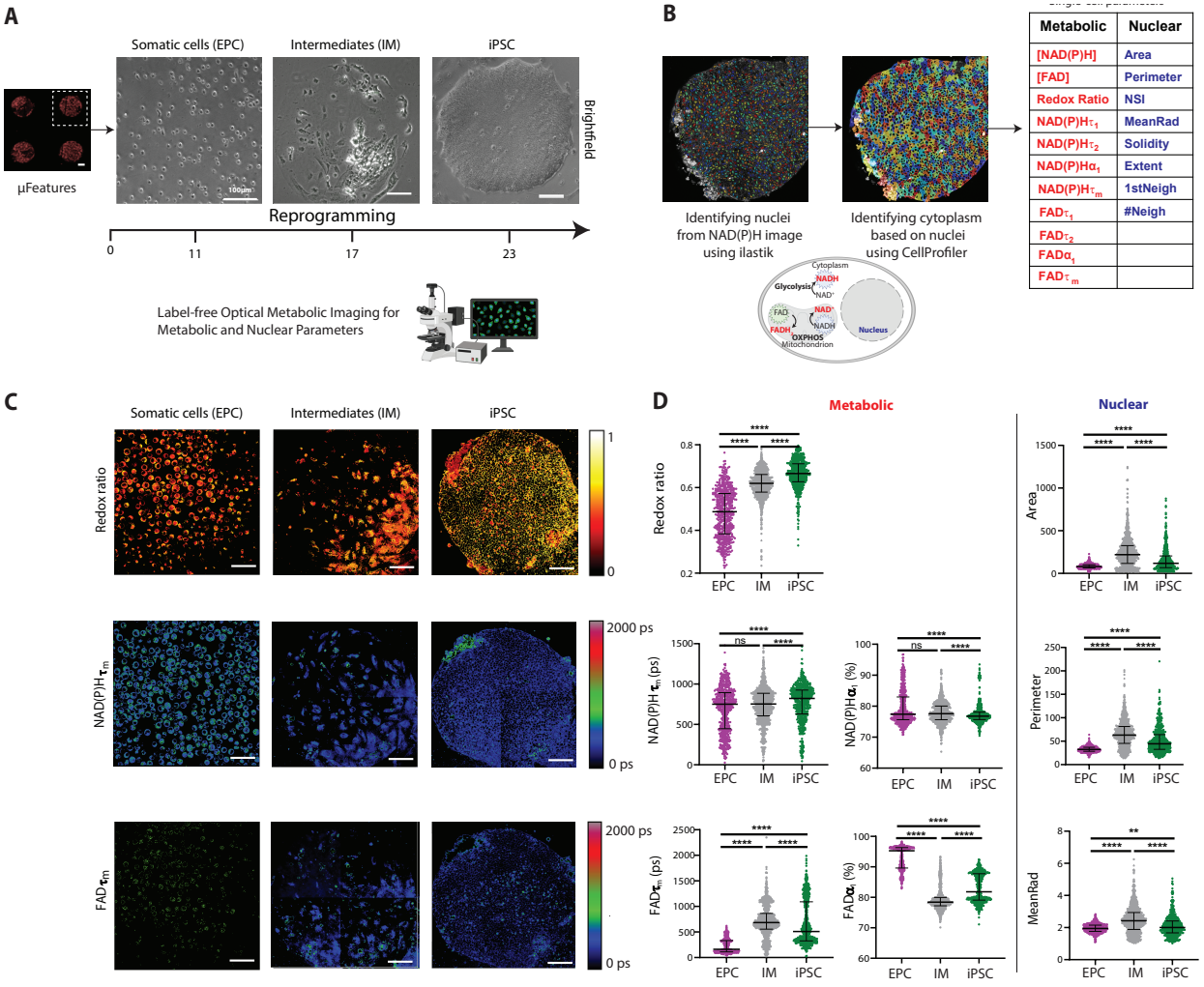


Figure 3.1: Caption next page

Figure 3.1: **NAD(P)H and FAD autofluorescence imaging revealed metabolic differences between EPCs, IMs and iPSCs.** A) Left: Matrigel-coated μFeatures on ibiTreat dish visualized with anti-laminin antibody (red) shows good fidelity in the transfer from the Matrigel-coated PDMS mold. Scale bar, 100 μm. Right: Representative images of the progression of EPCs on a single 300 μm radius circular μFeature through a reprogramming time course. B) Top: Image analysis pipeline to identify 11 metabolic (red) and 8 nuclear (blue) parameters using ilastik and CellProfiler software. Bottom: Schematic representation of cell metabolism with a focus on NADH and FAD. NADH and FAD are highlighted as the fluorescent molecules in the diagram, and molecules in bold indicate the net direction of the reaction. C) Representtive optical redox ratio, NAD(P)Hτ_m and FADτ_m images (3 μFeatures selected from 36 μFeatures acquired from 3 different donors) for EPC, IM, and iPSC. Color bars are indicated on the right. Scale bar, 100 μm. D) Single-cell quantitative analysis of metabolic parameters: optical redox ratio, NAD(P)Hτ_m, NAD(P)Hα_1, FADτ_m, FADα_1; and nuclear parameters: area, perimeter and mean radius (n = 561, 990, 586 for EPC, IM, and iPSC respectively). Data are presented as median with interquartile range for each cell type. Statistical significance was determined by one-way analysis of variance (ANOVA) using the Kruskal-Wallis test for multiple comparisons; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$).

OMI reveals distinct metabolic changes during reprogramming

Metabolic state plays an important role in regulating reprogramming and pluripotency of iPSCs [263–267] and can be non-invasively monitored via OMI. NADH is an electron donor and FAD is an electron acceptor, both present in all cells as coenzymes for metabolic reactions. For example, glycolysis in the cytoplasm generates NADH and pyruvate, while OXPHOS consumes NADH and produces FAD (**Figure 3.1B**). Autofluorescence imaging of NADH and FAD is thus dynamically responsive to the oxidation-reduction state of a

cell and is influenced by many reactions [250, 268].

We tracked the autofluorescence dynamics of NAD(P)H and FAD by performing OMI on μCP substrates at different timepoints during EPC reprogramming. In these autofluorescent images, the nucleus remains dark as NAD(P)H is primarily located in cytosol and mitochondria, and FAD is primarily located in mitochondria. The NAD(P)H images were used as inputs for ilastik software [269] to identify the nuclei. The identified nuclei were then used as an input for high-content CellProfiler software [198] pipeline to segment the cytoplasm, and measure various metabolic and nuclear parameters (**Figure 3.1B**). Overall, a total of 11 metabolic parameters (Optical Redox Ratio, [NAD(P)H], NAD(P)Hα_1, NAD(P)Hτ_1, NAD(P)Hτ_2 , NAD(P)Hτ_m , [FAD], FADα_1, FADτ_1 , FADτ_2 , FADτ_m), and 8 nuclear parameters [249] (Area, Perimeter, MeanRad, NSI, Solidity, Extent, #Neigh, 1stNeigh) were measured by the analysis pipeline (**Table B.1**). Additionally, immunofluorescence labelling verified the cell type at these different timepoints, i.e., EPCs ($CD71^+$, $Nanog^-$), IMs ($CD71^-$, $Nanog^-$), and iPSCs ($CD71^-$, $Nanog^+$). NAD(P)H and FAD autofluorescence imaging revealed metabolic differences between starting EPCs, IMs, and iPSCs (**Figures 3.1,B.2, B.3**).

We observed a significant increase in the optical redox ratio (iPSC>IM>EPC) during the process of reprogramming (**Figure 3.1D**), indicating that EPCs are more oxidized than IMs and iPSCs. Additionally, we noted that patterned IMs and iPSCs have significantly higher optical redox ratios as compared to their non-patterned counterparts (**Figure B.2K**). This observation is consistent with previous studies which show that mechanical cues can regulate the relative use of glycolysis [270–273] and may require further investigation.

Next, we observed that NAD(P)H and FAD lifetime components undergo biphasic changes during the progress of reprogramming. FAD lifetime components undergo a more significant change relative to the NAD(P)H components (**Figures 3.1D, B.2A-J**). The fraction of protein-bound FAD (FADα_1) undergoes a decrease from EPCs to IMs and then an increase from IMs to iPSCs, which could be reflective of the OXPHOS burst

[67–69]. FADτ_m is inversely related to FADα_1 and therefore changes accordingly. Similar biphasic changes occur in nuclear parameters during reprogramming, consistent with our previously study [249] (**Figures B.1D, B.3**).

We noted that H9 ESCs have metabolic and nuclear parameters similar to iPSCs, as expected. Fibroblasts had metabolic parameters significantly different from EPCs (**Figure B.2**). This could be because 1) fibroblasts are adherent while starting EPCs are non-adherent, 2) fibroblasts and EPCs occupy distinct locations in Waddington landscape [201] with different proliferation rates and energy needs.

Taken together, autofluorescence imaging of NAD(P)H and FAD showed significant changes during reprogramming.

OMI enables the classification of reprogramming cells with high accuracy

Uniform Manifold Approximation and Projection (UMAP) [274], a dimension reduction technique, was used to visualize how cells cluster from metabolic and nuclear parameters onto 2D space. Neighbors were defined through Jaccard similarity coefficient computed across the metabolic parameters and nuclear parameters. UMAP was chosen over t-distributed stochastic neighbor embedding (t-SNE) [200] because UMAP (**Figure 3.2**) separated EPCs, IMs, and iPSCs better than t-SNE (**Figure B.4A**). Moreover, UMAP has a higher speed, includes non-metric distance functions, and preserves the global structure of the data.

We used UMAP to visualize how cells cluster exclusively from the 11 metabolic parameters (**Figure 3.2B**) and exclusively from the 8 nuclear parameters (**Figure 3.2C**). While these UMAP representations revealed separation of EPCs, IMs, and iPSCs; UMAP generated using both metabolic and nuclear parameters provided a cleaner separation between EPCs, IMs, and iPSCs (Figure 3.2A). We also plotted a heatmap representation of the z-score of metabolic and nuclear parameters at the donor level (each row is the

mean data of a single donor and cell type) to examine donor heterogeneity. While EPCs and iPSCs obtained from different donors clustered together on the heatmap, IMs showed interdonor heterogeneity (**Figure B.4B**). In summary, clustering of 11 metabolic and 8 nuclear parameters by using UMAP and z-score heatmap clustering showed that EPCs, IMs, and iPSCs can be distinguished based on these parameters.

Next, classification models were developed based on 11 metabolic and 8 nuclear parameters to predict the reprogramming status of cells, i.e., EPCs, IMs, or iPSCs. Supervised machine learning classification (Naïve Bayes, K-nearest neighbor) and regression algorithms (logistic regression, and random forest) [275] were implemented to test the prediction accuracy for iPSCs when all the metabolic and nuclear parameters are used. To protect against over-fitting, the classification models were trained using 15-fold cross-validation on single-cell data from three different donors with reprogramming status assigned from morphological characteristics and tested on data with same cell CD71 and Nanog staining validation from three donors (completely independent and non-overlapping observations). Receiver operator characteristic (ROC) curves of the test data revealed highest classification accuracy for predicting iPSCs (area under the curve AUC = 0.993), IMs (AUC = 0.993) and EPCs (AUC = 0.999) when random forest classification model was used (**Figures 3.2D, B.4C,D**). We thus used the random forest classification model for further analysis in this study.

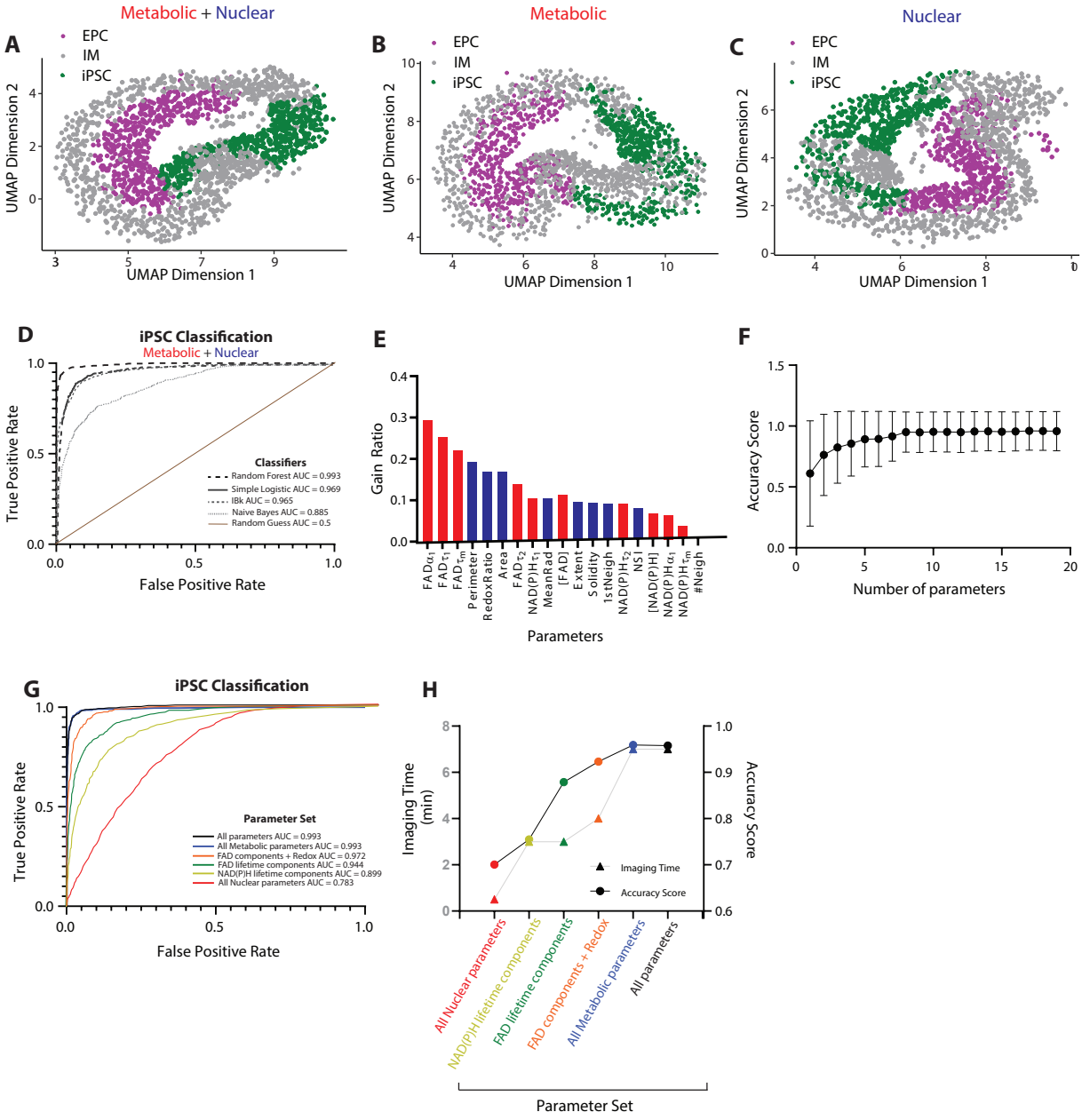


Figure 3.2: Caption next page

Figure 3.2: **Optical metabolic imaging enables classification of cells based on their reprogramming status.** Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction was performed on A) all 11 metabolic and 8 nuclear parameters, B) only 11 metabolic parameters, and C) only 8 nuclear parameters for each cell, projected onto 2D space. UMAP shows separation of different cell types (EPCs, IMs, and iPSCs). Each color corresponds to a different cell type. Data are from three different donors. Each dot represents a single cell, and n =561, 990, and 586 cells for EPC, IM, and iPSC, respectively. D) Model performance of the different classifiers (random forest, simple logistic, k-nearest neighbor (IBk), naïve bayes) for iPSCs was evaluated by receiver operating characteristic (ROC) curves using all 11 metabolic and 8 nuclear parameters. The area under the curve (AUC) is provided for each classifier as indicated in the legend. E) Parameter weights for random forest classification of EPCs, IMs, and iPSCs using gain ratio method. Analysis was performed at a single-cell level using three different donors. F) Classification accuracy with respect to number of parameters was evaluated based on the gain ratio parameter selection with the random forest model. The number of parameters included in the random forest model is indicated on the x-axis. G) Model performance of the random forest classifier for iPSCs was evaluated by ROC curves using different metabolic and nuclear parameter combinations as labelled. AUC is provided for each parameter combination as indicated in the legend. H) Imaging time (left y-axis) and accuracy score (right y-axis) evaluation of the random forest classifier for different metabolic and nuclear parameter combinations as labeled.

Gain ratio analysis revealed that FAD lifetime components, FADα_1, FADτ_1 and FADτ_m, are the most important parameters for classifying the reprogramming status of cells (**Figure 3.2E**). This is consistent with the observation FAD lifetime components are significantly different among EPCs, IMs, and iPSCs (**Figures 3.1, B.2**). We then plotted the accuracy score as a function of the number of parameters (chosen based on the gain

ratio values for random forest classifier) used for classification. This plot revealed that the accuracy score increases with the number of parameters until 8 parameters and plateaus thereafter (**Figure 3.2F**). We additionally noted that using FAD lifetime variables exclusively (FAD τ_m, τ_1, τ_2, α_1; collected in the FAD channel alone), high classification accuracy can be achieved for predicting iPSCs (AUC = 0.944), IMs (AUC = 0.968) and EPCs (AUC = 0.998) (**Figures 3.2G, B.4E,F**). Using only FAD lifetime parameters ensures minimal imaging time of 2.5 min per μFeature (**Figure 3.2H**), and no additional reliance on intensity parameters which are associated with higher variability due to the confounding factors of intensity levels (throughput due to laser power, detector gain and inner filter effects). Hence, FAD lifetime parameters alone are sufficient to predict the reprogramming status of cells.

Pseudotemporal ordering of single cells reveals heterogeneous cell populations

To study the heterogeneity of reprogramming μFeatures, we used 11 metabolic and 8 nuclear parameters to construct pseudotime single-cell trajectories of cellular reprogramming using Monocle3 program [259, 276]. The resultant trajectories consisted of EPCs, IMs, and iPSCs distributed across 10 clusters, 4 branching events, and a disconnected branch (**Figure 3.3A-C**). Notably, pseudotime highly correlated with the actual reprogramming time points (**Figures 3.3B, B.5A**).

The starting EPCs were heterogeneous and occupied three clusters (Clusters: 1, 2, 3). While cluster 2 consists of starting EPCs that undergo reprogramming, clusters 1 and 3 constituted the disconnected branch with EPCs that may not easily be permissive to reprogramming. iPSCs predominantly occupied two clusters (clusters 7, 10) irrespective of the reprogramming timepoint, while IMs belonged to several clusters (clusters 4, 5, 6, 8, 9) with clusters 6 and 8 concentrated at the unsuccessful reprogramming branches (**Figure 3.3C**). Overall, this single-cell trajectory map is indicative of reprogramming heterogeneity, i.e., cells that advance right at branch points 1, 2, and 3 completely reprogram to iPSCs

within 25 days of reprogramming initiation, while cells that proceed left at branch points 1 and 3 remain at the intermediate stage (**Figure 3.3A**).

Subsequent heatmap analysis on the clusters in the single-cell reprogramming trajectory map revealed that the clusters exhibited correlation patterns based on their reprogramming status, i.e., EPCs (cluster 2) have a high correlation to early IMs (cluster 4), while late IMs (cluster 5, 6, 8, 9) demonstrate high correlation to iPSCs (cluster 10) (**Figure 3.3D**). We then compared IMs that undergo reprogramming (cluster 9) and the IMs (cluster 6 and 8) that do not reprogram to iPSCs, and noted differences in their NAD(P)H lifetime components, indicating that these parameters might play a role in determining reprogramming cell fate.

To further examine the parameters that distinguished the cell clusters (**Figure B.5B**), we performed spatial correlation analysis using Moran's I [277]. When the parameters were ranked by their effect size (Moran's I), FAD lifetime parameters (FAD τ_1, τ_2, τ_m) were found to be the most important in distinguishing clusters followed by NAD(P)H lifetime parameters (NAD(P)H τ_2, α_1, τ_1) (**Figures B.5B**). This result is consistent with high gain ratio values for FAD lifetime parameters (**Figure 3.2D**) and the observation that FAD lifetime parameters are significantly different among EPCs, IMs, and iPSCs (**Figure 3.1D**).

Overall, while FAD parameters are important in distinguishing EPCs, IMs, and iPSCs; NAD(P)H parameters are key for determining the eventual reprogramming fate of cells. When we plotted the identified important metabolic parameters as a function of pseudotime, we observed that they undergo biphasic changes during reprogramming potentially representative of the OXPHOS burst (**Figure 3.4E-J**). These pseudotime trajectory plots provided increased temporal resolution as compared to our previous plots (**Figure B.2**). Taken together, our reprogramming trajectory analyses provided insights into reprogramming heterogeneity at a high temporal resolution and unprecedented single-cell resolution.

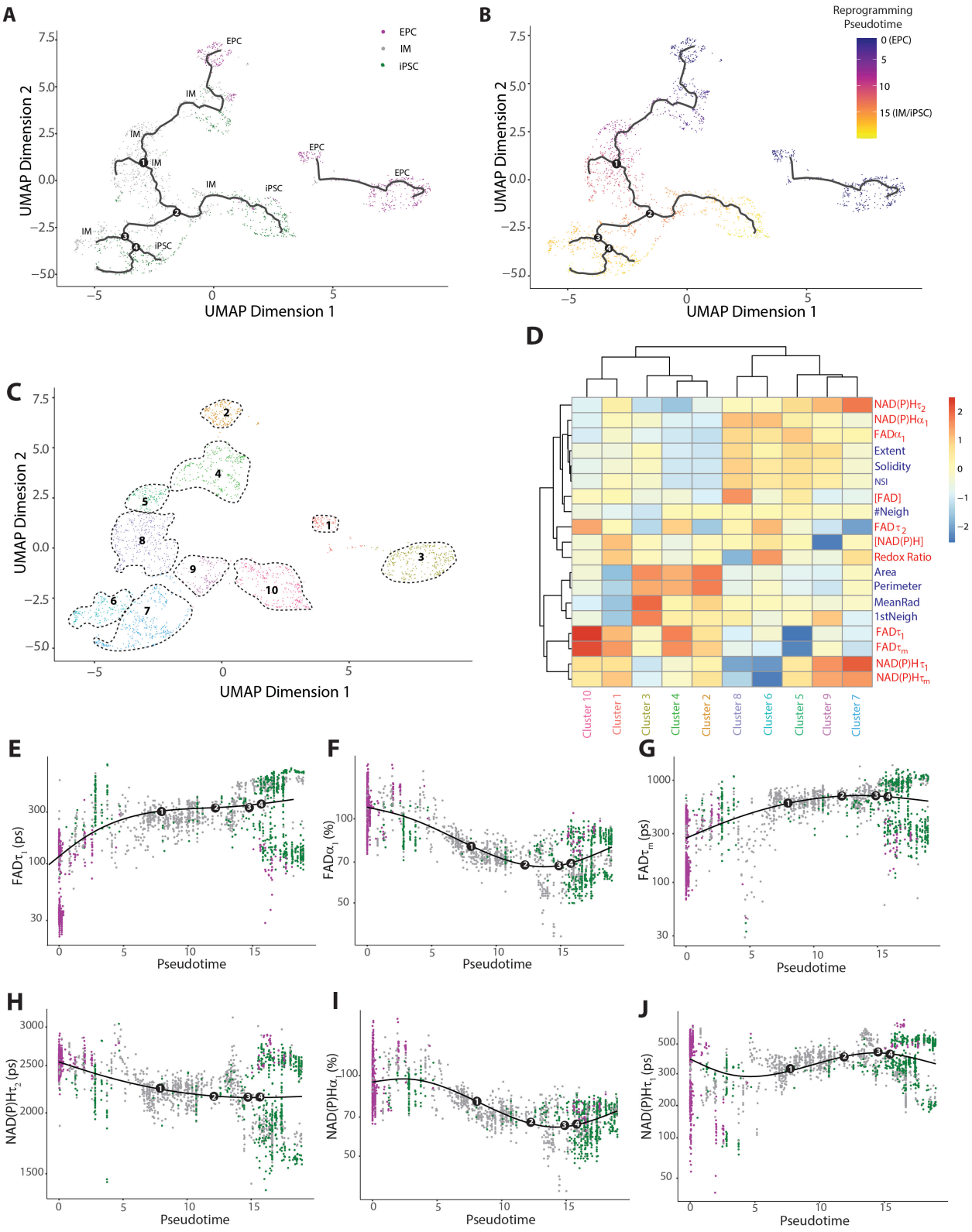


Figure 3.3: Caption next page

Figure 3.3: **Construction of single-cell reprogramming trajectories using metabolic and nuclear parameters reveals heterogeneity during reprogramming.** Trajectory of reprogramming EPCs constructed from the metabolic and nuclear parameters based on UMAP dimension reduction using Monocle (Code 1) showed four branch points, colored by A) cell type and B) pseudotime. C) Monocle UMAP plots showing clustering of reprogramming EPCs. Samples were grouped into 10 clusters. Cells colored by cluster. D) Heatmap representing the metabolic and nuclear parameters of 10 clusters. Each column is a separate cell group based on the generated clusters and each row represents a single metabolic or nuclear parameter. Dotplots indicating the expression of E) FADτ_1, F) FADα_1, G) FADτ_m, H) NAD(P)Hτ_2, I) NAD(P)Hα_1 and, J) NAD(P)Hτ_2 along the pseudotime. Smooth lines are composed of multiple dots representing the mean expression level at each pseudotime, regardless of the cell type. Four branch points are labelled on the smooth lines.

Isolation of high-quality iPSCs

The terminal goal of any reprogramming platform is to successfully isolate iPSCs that can be used for downstream applications. Using the combination of OMI, μCP platform, and machine learning models developed in this study, we were able to successfully isolate high-quality iPSCs (**Figure 3.4**).

First, we tracked the metabolic and nuclear parameters of μFeatures throughout the reprogramming timecourse using OMI (**Figure 3.4A**). Second, we employed our random forest classification model to predict the reprogramming status of the tracked μFeatures (**Figure 3.4B**). Third, we inferred the pseudotimes during the reprogramming timecourse to monitor the progress of the μFeatures along the reprogramming trajectory (**Figure 3.4C**). Finally, we performed immunostaining on the μFeatures, which showed that the reprogramming status predictions made by the machine learning models correlated well with the actual staining (**Figure 3.4D**).

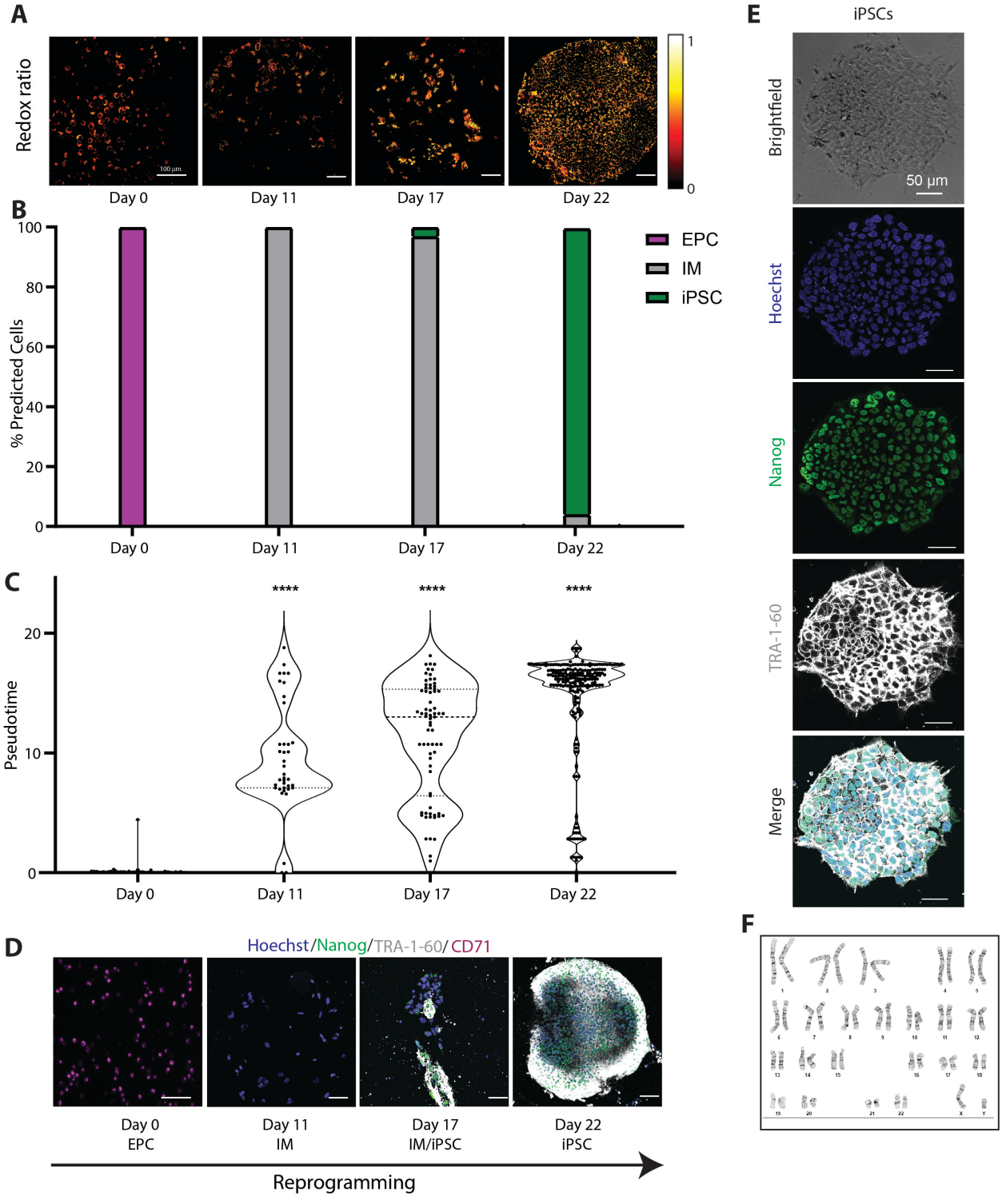


Figure 3.4: Caption next page

Figure 3.4: **Optical metabolic imaging of μFeatures aids in the identification and isolation of iPSC populations.** A) Representative optical redox ratio images of a single μFeature at different days through the reprogramming time course. The color bar is indicated on the right. Scale bar, 100 μm. B) Stacked column bar graph showing variation in the distribution of cell types during reprogramming as predicted by random forest classifier using all metabolic and nuclear parameters. The color of the bar corresponds to the cell type and the height of the bar represents the percentage of cell types. C) Violin plots showing the distribution of reprogramming pseudotime of single cells within a μFeature as a function of the actual reprogramming timepoint. Statistical significance was determined by one-way analysis of variance (ANOVA) using the Kruskal-Wallis test for multiple comparisons to Day 0; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$). D) Representative images of cell subpopulations on μFeatures at different days through the reprogramming time course, stained using antibodies against Hoechst (blue), TRA-1-60 (white), Nanog (green) ,and CD71 (magenta). E) Representative image of iPSC colony isolated from a μFeature, stained with Hoechst (nuclear dye), TRA-1-60, and Nanog (pluripotency markers). Scale bar, 50 μm. F) iPSCs derived from μCP substrates show normal karyotype suggesting that no major chromosome abnormality was present within cells after reprogramming.

We then went on to isolating iPSCs from the μCP platform based on the predictions made by the random forest classification model. The physical separation of micropatterns from one another, combined with high fraction of predicted iPSC cells, even up to 100%, throughout the μFeature resulted in easy picking and isolation of reprogrammed iPSCs. We further confirmed that the isolated iPSCs expressed pluripotency markers (**Figure 3.4E**) and showed no genomic abnormalities (**Figure 3.4F**), indicating that our reprogramming platform can be used to generate genetically-stable iPSC lines.

3.4 Discussion

Here, we report a non-invasive, high-throughput, quantitative, and label-free imaging platform to predict the reprogramming status of EPCs by combining micropatterning, live-cell autofluorescence imaging and automated machine learning. We are able to predict the reprogramming status of EPCs at any timepoint during reprogramming with a prediction accuracy of 95% and model performance of 0.99 (AUC of ROC) using random forest classification model with 11 metabolic parameters and 8 nuclear parameters. Additionally, we presented the single-cell roadmap of EPC reprogramming, which reveals diverse cell fate trajectories of individual reprogramming cells.

Recent evidence indicates that metabolic changes during reprogramming include decreasing OXPHOS and increasing glycolysis [65, 72], along with a transient hyper-energetic metabolic state, called as OXPHOS burst. This OXPHOS burst occurs at an early stage of reprogramming and shows characteristics of both high OXPHOS and high glycolysis, which could be a regulatory cue for the overall shift of reprogramming [68–70, 278]. These changes are accompanied by alterations in the amounts of corresponding metabolites and have been confirmed by genome-wide analyses of gene expression, protein levels and metabolomic profiling [53, 66, 279, 280]. The shifts in cellular metabolism affect enzymes that control epigenetic configuration [281], which can impact chromatin reorganization and provide a basis for changes in nuclear morphology as well as gene expression during reprogramming [55, 60, 62, 249]. Consistent with these studies, the redox ratio increases during reprogramming (**Figure 3.1D**), which could be indicative of increased glycolysis during reprogramming.

The changes in NAD(P)H and FAD lifetime parameters that occur during reprogramming (**Figures 3.1, B.2**) could reflect changes in quencher concentrations, such as oxygen, tyrosine, or tryptophan, or changes in local temperature and pH [250, 282, 283]. Specifically, the biphasic changes in the metabolic and nuclear parameters could be transient due to the increased production of ROS by mitochondria [250, 268, 284, 285] during

OXPHOS burst. The generated ROS further serves as a signal to activate Nuclear Factor (erythroid derived 2)-like-2 (NRF-2), which then induces hypoxia-inducible factors (HIFs) that could promote glycolysis during reprogramming by increasing the expression levels of the glycolysis-related genes [65, 66, 70]. Moreover, the importance of FAD parameters for distinguishing various reprogramming cell types (**Figures 3.1, B.2 3.2E, B.5B**) could point to the significant changes in the mitochondrial microenvironment during reprogramming. The differences in NAD(P)H lifetime parameters between IMs that successfully undergo reprogramming and the ones that don't, may suggest the role of NAD(P)H in impacting reprogramming barriers and warrants further investigation.

The classification analysis revealed that models trained on all 11 metabolic and 8 nuclear parameters yielded the highest accuracy for classification of reprogramming status of cells. Random forest classification using only FAD lifetime parameters yielded comparatively high ROC AUC values (**Figures 3.2G, B.4E,F**). Additionally, FAD lifetime parameters were more accurate for predicting reprogramming status than using nuclear parameters alone, which can be obtained using widefield or confocal fluorescence microscopy. Imaging only FAD lifetime parameters significantly reduces the time of imaging especially when multiple μFeatures are trying to be assessed for iPSC quality at a manufacturing scale and also eliminates the variability associated with intensity measurements.

Our single-cell reprogramming trajectory maps built based on metabolic and nuclear parameters (**Figure 3.3**) could indicate that the reprogramming process could be proceeded by a combination of elite and stochastic models [42]. While there seems to be a fraction of starting EPCs that are refractory towards reprogramming supporting of the elite model of reprogramming [47], there is also a fraction of intermediate cells at various stages of reprogramming that do not completely reprogram to iPSCs corroborating the stochastic model of reprogramming [27].

Much of the current work to understand the heterogeneity during reprogramming relies on bulk analysis [175, 181, 183] or single-cell analysis techniques [49, 58, 158, 187–190,

204, 205]. While bulk samples obscure variability in both the starting cell population and during fate conversion, owing to the variable kinetics and low efficiency of reprogramming; single-cell techniques disrupt the cells' microenvironment, resulting in significant changes in the biophysical properties of cells undergoing reprogramming. Our reprogramming platform overcomes these challenges of the current methods by using the combination of μCP platform, OMI, and Monocle algorithm. Firstly, μCP platform ensures an intact microenvironment for reprogramming cells while also allowing for analysis at the single-cell level. Secondly, OMI has several spatial and temporal resolution advantages compared with traditional assays enabling greater insights into reprogramming heterogeneity. OMI can be performed at high resolution to enable measurements at the single-cell level and is non-destructive allowing for spatial integrity measurements of neighboring cells. OMI also has a high temporal resolution enabling time-course study of reprogramming. Finally, Monocle algorithm [259] is a trajectory inference method that learns combinatorial changes which each cell must go through as a part of the reprogramming process, and subsequently places each cell at its proper location in the trajectory, thus overcoming the problems of reprogramming trajectories built based on absolute time points which disregard the asynchrony of the reprogramming process. Overall, we hope that the generated reprogramming trajectories will help deepen our understanding of the mechanisms of the reprogramming process. In addition, understanding these mechanisms can contribute to the identification of somatic cells or early reprogramming cells which are refractory towards reprogramming, and thus be used to increase the success rate of iPSC generation from patient-derived primary cells or cell lines.

Our reprogramming platform could contribute to the long-term commercial success of iPSC-derived therapies by migrating the iPSC manufacturing process from laborious, time-consuming, error-prone high-risk lab bench protocol to an industrial-scale, GMP-compliant manufacturing system. Firstly, μCP platform involves direct ECM printing onto optically clear substrates (**Figure B.1**) and does not involve any gold coating unlike

traditional microcontact printing methods [231, 286], making the process cost-effective and simpler by eliminating the need for cleanroom access. We were also able to isolate fully pluripotent iPSCs without any genomic abnormalities (**Figure 3.4**) using this μCP platform, indicating that this platform could be adapted for biomanufacturing of GMP-grade iPSCs. Secondly, we used erythroid progenitor cells isolated from peripheral blood as the starting cell type for reprogramming due to their lack of genomic rearrangements and demonstrated reprogramming ability [186, 287]. Moreover, blood collection is a minimally invasive procedure, and collected cells are naturally replaced as the tissue is self-renewing, making it suitable for generating iPSCs. Thirdly, we used a combination of oriP/EBNA-based viral-free non-integrating episomal reprogramming plasmids described previously [192, 288], that avoid the safety concerns surrounding the use of integrating viral vectors. Moreover, these episomes are lost at around 5% per cell generation due to defects in plasmid synthesis and partitioning and thus iPSCs devoid of plasmids can easily be isolated for clinical applications [38, 289]. We also used xeno-free components for feeder-free reprogramming and maintenance of iPSCs to eliminate the inconsistencies arising from the undefined nature of xeno-components and ensures GMP compliance. Finally, the autofluorescence imaging technique is label-free unlike other methods to study metabolism like electron microscopy, immunocytochemistry, and colorimetric metabolic assays. It also enables non-destructive real-time monitoring of live cells with lower sample phototoxicity compared to single-photon excitation [250]. Taken together, the processes of μCP platform fabrication, reprogramming, autofluorescence imaging, iPSC identification based on machine learning models and iPSC isolation can all be automated, and be extended to different reprogramming methods [290]); to other starting cell types (fibroblasts, keratinocytes); to other parameters (cell morphology [160, 163, 164] and mitochondrial structure [72, 279, 291]), and to other processes (direct reprogramming, differentiation [292–295]); making it an attractive platform for biomanufacturing of industrial-scale iPSCs and iPSC-derived cells.

3.5 Materials and Methods

EPC isolation and cell culture

EPCs were isolated from fresh peripheral human blood that was obtained from healthy donors (Interstate Blood Bank, Memphis, TN). Blood was processed within 24 hours of collection, where hematopoietic progenitor cells were extracted from whole blood using negative selection (RosetteSep; STEMCELL Technologies) and cultured in polystyrene tissue culture plates in erythroid expansion medium (STEMCELL Technologies) for 10 days to enrich for EPCs. Enriched EPCs from Day 10 were examined by staining with APC Anti-Human CD71 antibody (334107; Biolegend; 1:100) and incubating for 1 hour at room temperature. Data were collected on Attune Nxt flow cytometer and analyzed with FlowJo.

Micropattern Design and PDMS stamp production

First, a template with the feature designs was created in AutoCAD (Autodesk). The template was then sent to the Advance Reproductions Corporation, MA for the fabrication of a photomask, and a 6-inch patterned Si wafer was fabricated by the Microtechnology Core, University of Wisconsin-Madison, WI [296]. Using soft photolithography techniques, the Si wafer was spin-coated with a SU-8 negative photoresist (Y13273 1000L 1GL, MICRO CHEM) and exposed to UV light. The Si mold was then developed for 45 minutes in SU-8 developer (PGMEA, 537543, Sigma) which yielded features with a height of 150 μm. The Si mold was then washed with acetone and isopropyl alcohol.

Elastomeric stamps used for microcontact printing were generated by standard soft lithographic techniques. Briefly, the silicon mold was rendered inert by overnight exposure in vapors of (tridecafluoro-1, 1, 2, 2-tetrahydrooctyl) trichlorosilane. Polydimethylsiloxane (Sylgard 184 silicon elastomer base, 3097366-1004, Dow Corning) (PDMS) was prepared at a ratio of 1:10 curing agent (Sylgard 184 silicone elastomer curing agent, 3097358-1004, Dow Corning) and degassed in a vacuum for 30 minutes. The PDMS was then poured

over the SU-8 silicon mold on a hot plate and baked at 60 °C overnight to create the PDMS stamp.

μCP Well Plate construction

Microcontact patterned (μCP) substrates were constructed based on previous studies [260, 261, 297]. In brief, polydimethylsiloxane (PDMS) stamps with 300 μm radius circular features were coated with Matrigel (WiCell Research Institute) for 24 h. After 24 h, the Matrigel-coated PDMS stamp was dried with N_2 and placed onto 35 mm cell culture treated ibiTreat dishes (81156; Ibidi). A 50 g weight was added on top of the PDMS stamps to ensure even pattern transfer from the Matrigel-coated PDMS stamp to the ibiTreat dish. This setup was incubated for 2 h at 37°C. The 35 mm ibiTreat dish was then backfilled for 30 min at RT with PLL (20 kDa)-g-(3.5)-PEG (0.1 mg/mL solution) (Susos AG), a graft polymer solution with a 20 kDa PLL backbone with 2 kDa PEG side chains and a grafting ratio of 3.5 (mean PLL monomer units per PEG side chain). The ibiTreat dish was then washed with PBS and exposed to UV light for 15 min for sterilization to yield the micropatterned substrate.

Reprogramming

Day 10 EPCs were electroporated with four episomal reprogramming plasmids encoding Oct4, shRNA knockdown of p53 (#27077; Addgene, Watertown, MA); Sox2, Klf4 (#27078; Addgene); L-Myc, Lin28 (#27080; Addgene); miR302-367 cluster (#98748; Addgene), using the P3 Primary Cell 4D-Nucleofector Kit (Lonza) and the EO-100 program [33, 192]. Electroporated EPCs were seeded onto micropatterned substrates with erythroid expansion medium (STEMCELL Technologies) at a seeding density of 2000k cells/dish. Cells were supplemented with ReproTeSR (STEMCELL Technologies) on alternate days starting from Day 3 without removing any medium from the well. On Day 9, the medium was entirely

switched to ReproTeSR, and the ReproTeSR medium was changed daily starting from Day 10.

Isolation of iPSCs

To isolate high-quality iPSC lines, candidate colonies were picked from micropatterns using a 200 μL micropipette tip and transferred to Matrigel-coated polystyrene tissue culture plates in mTeSR1 media (WiCell Research Institute, Madison, WI). If additional purification was required, one additional manual picking step with a 200 μL micropipette tip was performed. During picking and subsequent passaging, the culture media was often supplemented with the Rho kinase inhibitor Y-27632 (Sigma-Aldrich) at a 10 μM concentration to encourage cell survival and establish clonal lines. iPSCs obtained from EPCs were maintained in mTeSR1 media on Matrigel-coated polystyrene tissue culture plates and passaged with ReLeSR (STEMCELL Technologies) every 3-5 days. All cells were maintained at 37ºC and 5% CO_2.

Cell Culture

H9 Embryonic Stem Cell (ESC) line was obtained from WiCell Research Institute, Madison, WI and was maintained in mTeSR1 medium on Matrigel-coated tissue culture polystyrene plates. Cells were passaged every 4–5 days at a ratio of 1:8 using ReLeSR solution (STEMCELL Technologies). Human Dermal Fibroblast (HDF) line was a gift from Dr. Bikash Pattnaik (UW-Madison) which was obtained from biopsy skin sample of an anonymous donor through the University of Wisconsin Foundation. HDFs were cultured in DMEM (11965092, ThermoFisher Scientific) supplemented with 10% FBS (WiCell Research Institute) and 1% Penicillin/Streptomycin (15140122, ThermoFisher Scientific). All cells were maintained at 37°C in 5% CO_2, and tested monthly for possible mycoplasma contamination (WiCell Research Institute).

Antibodies and Staining

All cells were fixed for 15 minutes with 4% paraformaldehyde in PBS (Sigma-Aldrich, St. Louis, MO) and permeabilized with 0.5% Triton-X (Sigma-Aldrich) for >4 hours at room temperature before staining. Hoechst (H1399; Thermo Fisher Scientific) was used at 5 μg/mL with 15 min incubation at room temperature to stain nuclei. Primary antibodies were applied overnight at 4°C in a blocking buffer of 5% donkey serum (Sigma-Aldrich) at the following concentrations: Anti-Laminin (L9393; Sigma-Alrich) 1:500; TRA-1-60 (MAB4360; EMD Millipore) 1:100; Nanog (AF1997; R&D Systems) 1:200; CD71 (334107; Biolegend) 1:100. Secondary antibodies were obtained from Thermo Fisher Scientific and applied in a blocking buffer of 5% donkey serum for one hour at room temperature at concentrations of 1:400 – 1:800. A Nikon Eclipse Ti epifluorescence microscope was used to acquire single 10x images of each micropattern, and a Nikon AR1 confocal microscope was used to acquire 60x stitched images of each micropattern using the *z*-plane closest to the micropatterned substrate for reprogramming studies.

Autofluorescence Imaging of NAD(P)H and FAD

Fluorescence lifetime imaging (FLIM) was performed at different timepoints during reprogramming by an Ultima two-photon microscope (Bruker) composed of an ultrafast tunable excitation laser source (Insight DS+, Spectra Physics) coupled to a Nikon Ti-E inverted microscope with time-correlated single-photon counting electronics (SPC-150, Becker & Hickl, Berlin, Germany). The laser source enables sequential excitation of NAD(P)H at 750 nm and FAD at 890 nm. NAD(P)H and FAD images were acquired through 440/80 nm and 550/100 nm bandpass filters (Chroma), respectively, using Gallium arsenide phosphide (GaAsP) photomultiplier tubes (PMTs; H7422, Hamamatsu). The laser power at the sample was approximately 3.5 mW for NAD(P)H and 6 mW for FAD. Lifetime imaging using time-correlated single-photon counting electronics (SPC-150, Becker & Hickl) was performed within Prairie View Atlas Mosaic Imaging (Bruker Fluorescence Microscopy) to

capture the entire μFeature. Fluorescence lifetime decays with 512-time bins were acquired across 512 × 512 pixel images with a pixel dwell time of 4.8 μs and an integration period of 60 seconds. Photon count rates were 1-5 x 10^5 and monitored during image acquisition to ensure that no photobleaching occurred. All samples were placed on a stage-top incubator and illuminated through a 40×/1.15 NA objective (Nikon). The short lifetime of red-blood-cell fluorescence at 890 nm was used as the instrument response function and had a full-width half maximum of 240 ps. A YG fluorescent bead ($\tau = 2.13 \pm 0.03$ ns, n= 6) was imaged daily as a fluorescence lifetime standard [212, 250].

Image Analysis

Fluorescence lifetime decays were analyzed to extract fluorescence lifetime components via SPCImage software (Becker & Hickl). A threshold was used to exclude pixels with low fluorescence signal (that is, background). Fluorescence lifetime decays were deconvolved from the instrument response function and fit to a two-component exponential decay model, $I(t) = \alpha_1 e^{-t/\tau_1} + \alpha_2 e^{-t/\tau_2} + C$, where I(t) is the fluorescence intensity as a function of time t after the laser pulse, α_1 and α_2 are the fractional contributions of the short and long lifetime components, respectively (that is, $\alpha_1 + \alpha_2 = 1$), τ_1 and τ_2 are the short and long lifetime components, respectively, and C accounts for background light. Both NAD(P)H and FAD can exist in quenched (short lifetime) and unquenched (long lifetime) configurations; the fluorescence decays of NAD(P)H and FAD are therefore fit to two components. Fluorescence intensity images were generated by integrating photon counts over the per-pixel fluorescence decays. A pixel classifier was trained on 15 images using ilastik [269] software to identify the pixels within the nuclei in NAD(P)H images. An object classifier was then used to identify the nuclei in NAD(P)H images using the pixel classifier along with the following parameters: Method = Simple, Threshold = 0.3, Smooth = 1, Size Filter Min = 15 pixels, Size Filter Max = 500 pixels. A customized CellProfiler [198] pipeline was then used to obtain metabolic and nuclear parameters. The CellProfiler

pipeline applied the following steps: Primary objects (nuclei) were inputted from ilastik. Secondary objects (cells) were then identified in the NAD(P)H intensity image by outward propagation of the primary objects. Cytoplasm masks were determined by subtracting the nucleus mask from the cell mask. Cytoplasm masks were applied to all images to determine single-cell redox ratio and NAD(P)H and FAD lifetime parameters. A total of 11 metabolic parameters were analyzed for each cell cytoplasm: Optical Redox Ratio, [NAD(P)H], NAD(P)Hα_1, NAD(P)Hτ_1, NAD(P)Hτ_2, NAD(P)Hτ_m, [FAD], FADα_1, FADτ_1, FADτ_2, FADτ_m. A total of 8 nuclear parameters were analyzed for each nucleus: Area, Perimeter, MeanRad, NSI, Solidity, Extent, #Neigh, 1stNeigh. Representative images of the optical redox ratio (fluorescence intensity of NAD(P)H divided by the summed intensity of NAD(P)H and FAD) and mean fluorescence lifetimes ($\tau_m = \alpha_1\tau_1 + \alpha_2\tau_2$) of NAD(P)H and FAD were computed using the Fiji software.

UMAP clustering

Clustering of cells across EPCs, IMs, and iPSCs was represented using Uniform Manifold Approximation and Projection (UMAP). UMAP dimensionality reduction was implemented using R on all 11 OMI parameters (optical redox ratio, NAD(P)H τ_m, τ_1, τ_2, α_1, α_2; FAD τ_m, τ_1, τ_2, α_1, α_2) and/or all 8 nuclear parameters (Area, Perimeter, MeanRad, NSI, Solidity, Extent, #Neigh, 1stNeigh) for projection in 2D space. The following parameters were used for UMAP visualizations: “n_neighbors”: 20; “min_dist”: 0.3, “metric”: jaccard, ”n_components” : 2.

Z-score hierarchical clustering

Z-score of each metabolic and nuclear parameter for each cell was calculated. *Z*-score $= (\mu_{observed} - \mu_{row}/\sigma_{row})$ where $\mu_{observed}$ is the mean value of each parameter for each cell, μ_{row} is the mean value of each parameter for all cells together, and σ_{row} is the standard deviation of each parameter across all cells. Heatmaps of *z*-scores for all OMI variables were

generated to visualize differences in each parameter between different cells. Heatmaps were generated in Python.

Classification Methods

Random forest, simple logistic, k-nearest neighbor (IBk), and naïve bayes classification methods were trained to classify reprogramming cells into EPCs, IMs, and iPSCs using Weka software [298]. All data were randomly partitioned into training and test datasets using 15-fold cross-validation for training and test proportions of 93.3% and 6.7%, respectively. Each model was replicated 100 times; new training and test data were generated before each iteration. Parameter weights for metabolic and nuclear parameters were extracted using the GainRatioAttributeEval function in Weka to determine the contribution of each variable to the trained classification models. Receiver operating characteristic (ROC) curves were generated to evaluate the classification model performance on classification of test set data and are the average of 100 iterations of data that was randomly selected from training and test sets. All of the ROC curves displayed were constructed from the test datasets using the model generated from the training data sets.

Karyotype Analysis

Cells cultured for atleast 5 passages were grown to 60-80% confluence and shipped for karyotype analysis to WiCell Research Institute, Madison, WI. G-banded karyotyping was performed using standard cytogenetic protocols [232]. Metaphase preparations were digitally captured with Applied Spectral Imaging software and hardware. For each cell line, 20 GTL-banded metaphases were counted, of which a minimum of 5 were analyzed and karyotyped. Results were reported in accordance with guidelines established by the International System for Cytogenetic Nomenclature 2016 [233].

Statistics

p-values were calculated using non-parametric Kruskal-Wallis test for multiple unmatched comparisons with GraphPad Prism software. Statistical tests were deemed significant at $\alpha \leqslant 0.05$. Technical replicates are defined as distinct μFeatures within an experiment. Biological replicates are experiments performed with different donors. No a priori power calculations were performed.

4 CHROMATIN MODULATION FOR EFFICIENT CRISPR-CAS9 GENE EDITING OF HUMAN PLURIPOTENT STEM CELLS

Work in this chapter was adapted from:

Trichostatin A for Efficient CRISPR-Cas9 Gene Editing of Human Pluripotent Stem Cells

Kaivalya Molugu, Namita Khajanchi, Cicera R. Lazzarotto, Kirstan Gimse, Amritava Das, Shengdar Q. Tsai, Krishanu Saha

Article in preparation.

4.1 Abstract

Genome-edited human induced pluripotent stem cells (iPSCs) have broad applications in disease modeling, drug discovery, and regenerative medicine. Despite the development of clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 system, the gene editing process is inefficient and can take several weeks to months to generate edited iPSC clones. One of the major challenges is that chromatin compaction limits the Cas9 protein access to the target DNA. To overcome these challenges, we developed a strategy to improve the speed and efficiency of the iPSC gene editing process via the application of histone modifier, Trichostatin A (TSA; Class I and II histone deacetylase inhibitor or HDAC inhibitor). We observed that TSA increased the gene-editing efficiency of iPSCs by ~3.5 fold and shifted the DNA repair pathway balance slightly towards non-homologous end-joining repair pathway (NHEJ) over microhomology-mediated end joining pathway (MMEJ); while concurrently ensuring no increased off-target effects, retaining pluripotency and genomic integrity. We also developed an *in situ* imaging-based pipeline to quantify the TSA-induced change in chromatin condensation of iPSCs to rapidly enable the identification of iPSCs that are amenable to gene editing. Overall, we developed a method to generate edited iPSCs rapidly and efficiently, which could further aid in the biomanufacturing of therapeutically relevant gene-edited iPSCs.

4.2 Introduction

The pairing of gene editing technologies with human iPSCs for disease modeling overcomes the problem of animal models and human immortalized cell line models which do not accurately represent the genetic background or cellular physiology of the patient [81, 88, 299, 300]. Human iPSC-based models are thus a valuable resource for studying disease mechanisms [78, 301, 302], screening potential new therapeutics and testing toxic side-effects of drug treatments [79, 80, 303–305]. Moreover, performing gene editing on patient-

derived iPSCs prior to differentiation enables the generation of isogenic iPSC lines which can function as a control for studying the genetic disease model of interest [306–308]. Additionally, gene-corrected patient-specific iPSCs can be used for cellular therapies [309–314] and could overcome the problem of immune rejection.

However, the process of gene-editing iPSCs is typically laborious and inefficient involving multiple steps, requiring lengthy cell culture periods, drug selection, and several clonal events (i.e., initial reprogramming, gene targeting, and subsequent genetic excision of a selection cassette [110, 192, 315, 316]). Also, the selection of rare targeted clones in iPSCs can be particularly difficult because they grow poorly as single cells [111, 317]. Several recent approaches have been developed to improve the specificity, efficiency, and versatility of the CRISPR-Cas9 gene editing process by optimizing the structure of gRNAs [318–320], or by using modified nucleases strategies [121, 122], natural and engineered Cas9 variants [124, 125, 321], and small molecules [126–128]. However, these techniques still generate heterogeneous human cell populations that require time-consuming and laborious steps significant subsequent characterization steps for on-target edits, off-target effects, and unknown disease-causing mutations or risk variants [110]. Therefore, there is a need to overcome these limitations for enabling the use of gene-edited iPSC for cell therapies.

Recent work indicates that chromatin structure of the target can have significant effects on Cas9 binding and gene editing efficiencies [115, 129, 131, 301], thus leading to variation in targeting efficiency and choice of DNA repair pathway [143, 145, 146, 148, 149]. Studies have shown that closed chromatin can negatively affect Cas9 binding [130, 133–136] or delay CRISPR-Cas9 mutagenesis [137], nucleosomes can block or present a hurdle for Cas9 access to DNA [138–140], and active transcription in open chromatin state can directly stimulate DNA cleavage by influencing Cas9 release states in a strand-specific manner [129, 142]. Moreover, off-target binding of Cas9 to "seed" sequences has been shown to correlate with DNase I hypersensitivity sequences and inversely correlate with

CpG methylation sites [113, 135, 137]. Although these studies have shown chromatin structure can play a key role in gene editing and strategies have emerged to manipulate chromatin state to modulate gene editing efficiency [133, 152–154], they are primarily based on immortalized cell lines or cancer cell lines which may not be clinically relevant. Moreover, though chromatin structure has been shown to impact off-target effects [113, 135, 137, 155] and the potency of iPSCs [62, 156], the impact of chromatin structure on off-target effects and the potency of iPSCs in the context of gene editing has not been well characterized. Here we address these challenges using a two-part strategy to study the impact of chromatin structure on gene editing outcomes in iPSCs via the application of a small molecule known to promote open chromatin state, called Trichostatin A (TSA; Class I and II histone deacetylase inhibitor or HDAC inhibitor) [322]. First, we used live, *in situ* nuclear imaging to quantify the TSA-induced change in global chromatin state of iPSC nuclei, called chromatin condensation [323]. Second, we performed extensive characterization to assess the impact of TSA on gene editing outcomes i.e., on-target efficiencies, DNA repair pathways, off-target effects, pluripotency, and genomic integrity of the iPSCs. We observed that chromatin decondensation via TSA, increased the gene editing efficiency of iPSCs by ~3.5 fold and shifted the DNA repair pathway balance slightly towards non-homologous end-joining repair pathway (NHEJ) over microhomology-mediated end joining pathway (MMEJ), while simultaneously ensuring genomic integrity and no increased off-target effects. Overall, we developed a strategy for rapid and *in situ* imaging-based identification of iPSCs amenable to gene editing, and also engineered the gene editing process to generate edited iPSCs efficiently.

4.3 Results

Optimization of Trichostatin A treatment for iPSC gene editing

We sought to develop a simple, yet robust, workflow to introduce mutations in iPSCs without extensive use of deep sequencing (**Figure 4.1A**). We thus utilized a mono-allelic mEGFP tagged *HIST1H2BJ* WTC-11 human iPSC line (tag at C-terminus) [210] as it provides two-fold advantages. First, targeting the mEGFP locus allows for easy assessment of iPSC editing efficiency via fluorescence imaging or flow cytometry, since the percentage of GFP^- cells showed a strong linear correlation with the indel percentage determined by deep sequencing of genomic DNA (**Figure C.1A**; $R^2 = 0.9307$). Second, these iPSCs enable *in situ* live imaging of their cell nuclei to monitor chromatin changes during the process of CRISPR-Cas9 gene editing. We employed the Alt-R® CRISPR-Cas9 2-part single guide RNA (sgRNA) system (IDT), which consists of 1) crRNA sequence to target the mEGFP locus in iPSCs (**Figure 4.1B**), and 2) ATTO 550 fluorescent dye labelled tracrRNA that binds to Cas9. We tested a range of TSA concentrations (0-200 ng/mL) on iPSCs to determine the effect of TSA on cell viability. We observed that cell viability decreased with the increase in TSA concentration (**Figure C.1B**), with a considerable decrease for TSA concentration $\geqslant$ 25 ng/mL. We thus used TSA concentrations of 0, 3.13, 6.25, and 12.5 ng/mL for our further studies. Next, we assessed a range of TSA treatment durations (0-24 hours) and two methods for delivery for the Cas9 RNP complex (lipofection and electroporation) to determine the optimal conditions for gene editing. Since TSA treatment duration of 20 hours (**Figure C.1C**) and lipofection (**Figure C.1D**) yielded the highest gene editing efficiencies, we implemented these conditions for our subsequent experiments.

Trichostatin A enhances gene editing efficiency of iPSCs

After streamlining the gene editing strategy, we next sought to examine the effect of TSA treatment on the gene editing efficiency of iPSCs. We first pre-treated the iPSCs with TSA

(0, 3.13, 6.25, and 12.5 ng/mL) and subsequently imaged their nuclei. The images were then used as inputs for a CellProfiler software [198] pipeline to output the chromatin condensation% of the nuclei (**Figure C.2A**), defined as the ratio of total heterochromatin intensity to the total nuclear intensity [323]. Immunofluorescence labeling validated the heterochromatin foci identified in the GFP images by the CellProfiler pipeline i.e., the H3K9Ac euchromatin histone mark was excluded from the heterochromatin foci, while the H3K9Me3 heterochromatin histone mark overlapped with the heterochromatin foci (**Figure C.2B**). Thus, chromatin condensation % is representative of the global chromatin state of cells.

The nuclear image analysis pipeline revealed a decrease in chromatin condensation% upon TSA treatment (**Figure 4.1C,D**), indicating that our nuclear imaging pipeline is sensitive to the chromatin decondensation induced by TSA. We then assessed the dose-dependence of Cas9 RNP transfection efficiency (% ATTO 550^{+} cells/viable cells)) by performing flow cytometry analysis on day 2 after Cas9 RNP delivery. We noted high transfection efficiencies >95% (**Figure 4.1F**), indicating successful delivery of Cas9 RNP independent of the TSA concentration. Furthermore, flow cytometry analysis on day 6 after Cas9 RNP delivery revealed a positive correlation between that the gene editing efficiency (% GFP^{-} cells/viable cells) and the dose of TSA for concentrations up to 6.25 ng/mL (**Figure 4.1G**). While increasing the concentration of TSA to 6.25 ng/mL showed ~3.5 fold increase in gene editing efficiency, an additional increase in the TSA concentration to 12.5 ng/mL did not increase the editing efficiency further. We reasoned that this could be due to the higher cell toxicity that occurred at 12.5 ng/mL TSA concentration (**Figure 4.1E**). Taken together, these results indicate that nuclear imaging is sensitive to the chromatin condensation changes induced by TSA and that TSA enhances CRISPR-Cas9 mediated gene editing in a dose-dependent manner.

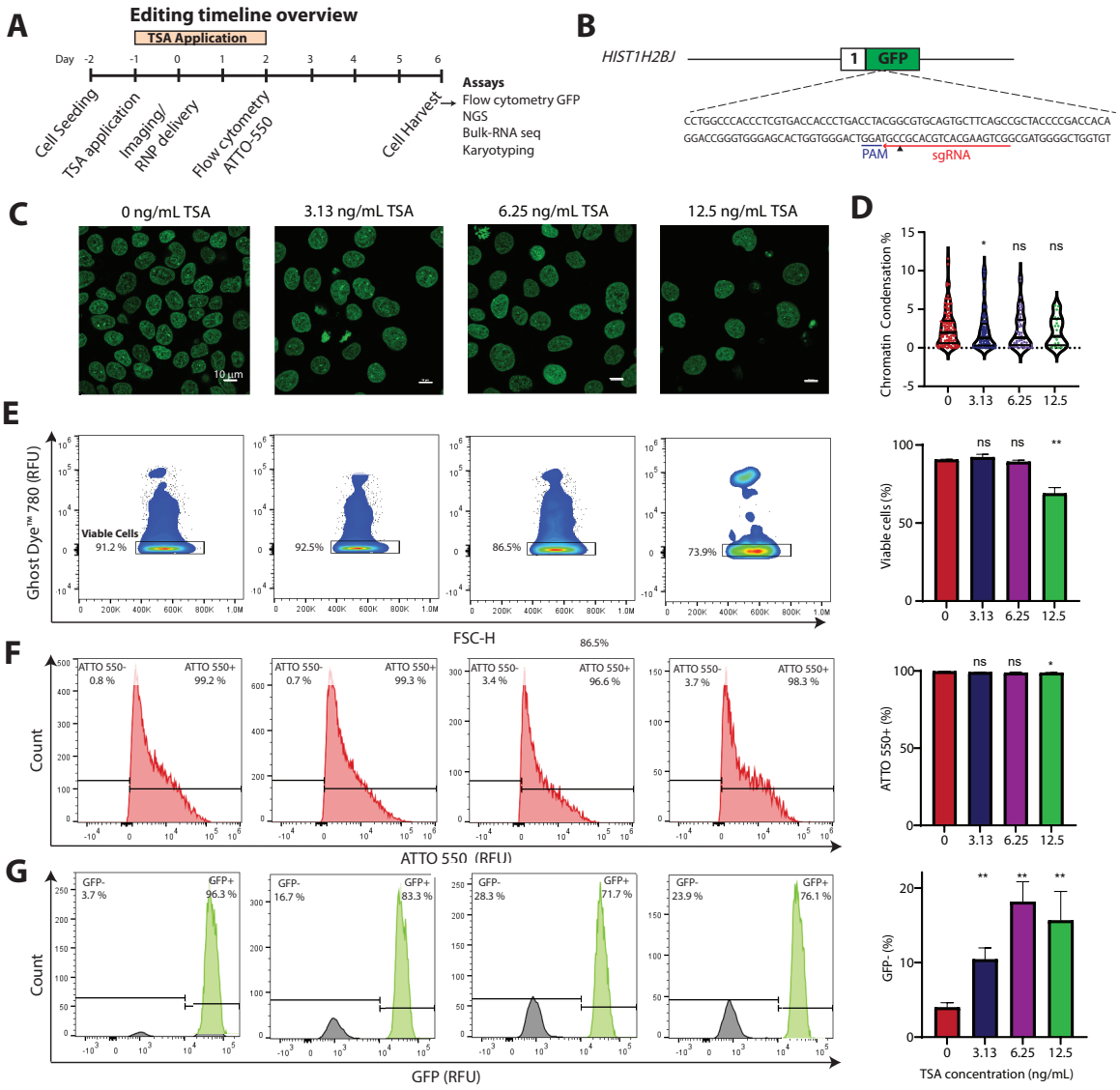


Figure 4.1: (Caption next page.)

Figure 4.1: **Trichostatin A increases CRISPR-Cas9 mediated gene editing efficiency of iPSCs.** A) Schematic showing the TSA-based gene editing strategy and analyses. B) Schematic of the *HIST1H2BJ*-mEGFP locus with gRNA target sequence, PAM sequence, and cut-site (black arrow) labelled. C) Representative images of *HIST1H2BJ*-GFP reporter iPSC nuclei after TSA treatment (0, 3.13, 6.25, 12.5 ng/mL). Bright green spots indicate heterochromatin foci; Scale bar: 10 /mum. D) Box plots showing the distribution of chromatin condensation %. Chromatin condensation % decreases with the application of TSA (n = 20 nuclei, 3 technical replicates per condition). E) Representative density flow cytometry plots showing Ghost Dye™ Red 780 viability dye levels on iPSCs treated with TSA. The quantification bar graph on the right indicates that cell viability (% GhostDye$^-$ cells/total cells) decreases upon TSA treatment. F) Histograms showing ATTO 550 expression on day 2 after Cas9 RNP delivery. The quantification bar graph on the right indicates that transfection efficiency (% ATTO 550$^+$ cells/viable cells) does not change significantly upon TSA treatment. G) Histograms showing GFP expression on day 6 after Cas9 RNP delivery. The quantification bar graph on the right indicates that gene editing efficiency (% GFP$^-$ cells/viable cells) increases upon TSA treatment. Data represented in bar graphs are represented as mean $\pm$ SEM, n = 6 technical replicates per condition from 2 independent experiments, *p*-values generated by Mann-Whitney non-parametric t-test for multiple comparisons to 0 ng/mL TSA; ns = p $\geqslant$0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001.

Deep sequencing of edited iPSCs reveals changes in indel profiles

To gain a more detailed analysis of gene-edited events in iPSCs, we performed deep sequencing on edited cells obtained after TSA treatment. We harvested the genomic DNA on day 6 after Cas9 RNP delivery. PCR using primers flanking the gRNA target sites was then performed and prepared for next-generation sequencing (NGS) using the Illumina MiniSeq

system. Sequencing data was then analyzed using the CRISPAltRations tool [324], which allowed for quantitative analysis of observed indel mutations and their spatial distribution in the target region (**Figure 4.2A, Table C.4**). We found the highest frequency of indels three to four nucleotides upstream from the protospacer adjacent motif (PAM) sequence (**Figure 4.2A**), consistent with reports of type II CRISPR systems. Next, we observed that the percentage of edited reads containing insertion and deletion events increased upon TSA treatment (**Figure 4.2B,D**), while the insertion and deletion profiles around the cut site were similar (**Figure 4.2C,E**). Furthermore, we noted that edited reads contained higher deletion events (~50-70%) than insertion events (~5-15%). Similar observations were reported in deep sequencing analysis of human cells edited by *S. Pyogenes* Cas9 (SpyCas9) [18]. We also observed that the percentage of edited reads with Single Nucleotide Polymorphisms (SNPs) decreased upon chromatin decondensation via TSA treatment (**Figure 4.2F**), while the relative distribution of SNPs around the cut site remained similar(**Figure C.2A**). This is consistent with reports that indicate nucleosomes are enriched in SNPs [325, 326]. We next analyzed the variation in MMEJ and NHEJ upon TSA treatment. While %indels with NHEJ increased, %indels with MMEJ decreased upon chromatin decondensation via TSA treatment indicating DNA repair pathway balance shift towards NHEJ(**Figure 4.2G,H**). This is consistent with the recent study that demonstrated that NHEJ is biased towards euchromatin, while MMEJ is more efficient in heterochromatin contexts [149]. Taken together, these results indicate that TSA treatment leads to shifts in indel profiles and the DNA repair pathway balance of edited iPSCs.

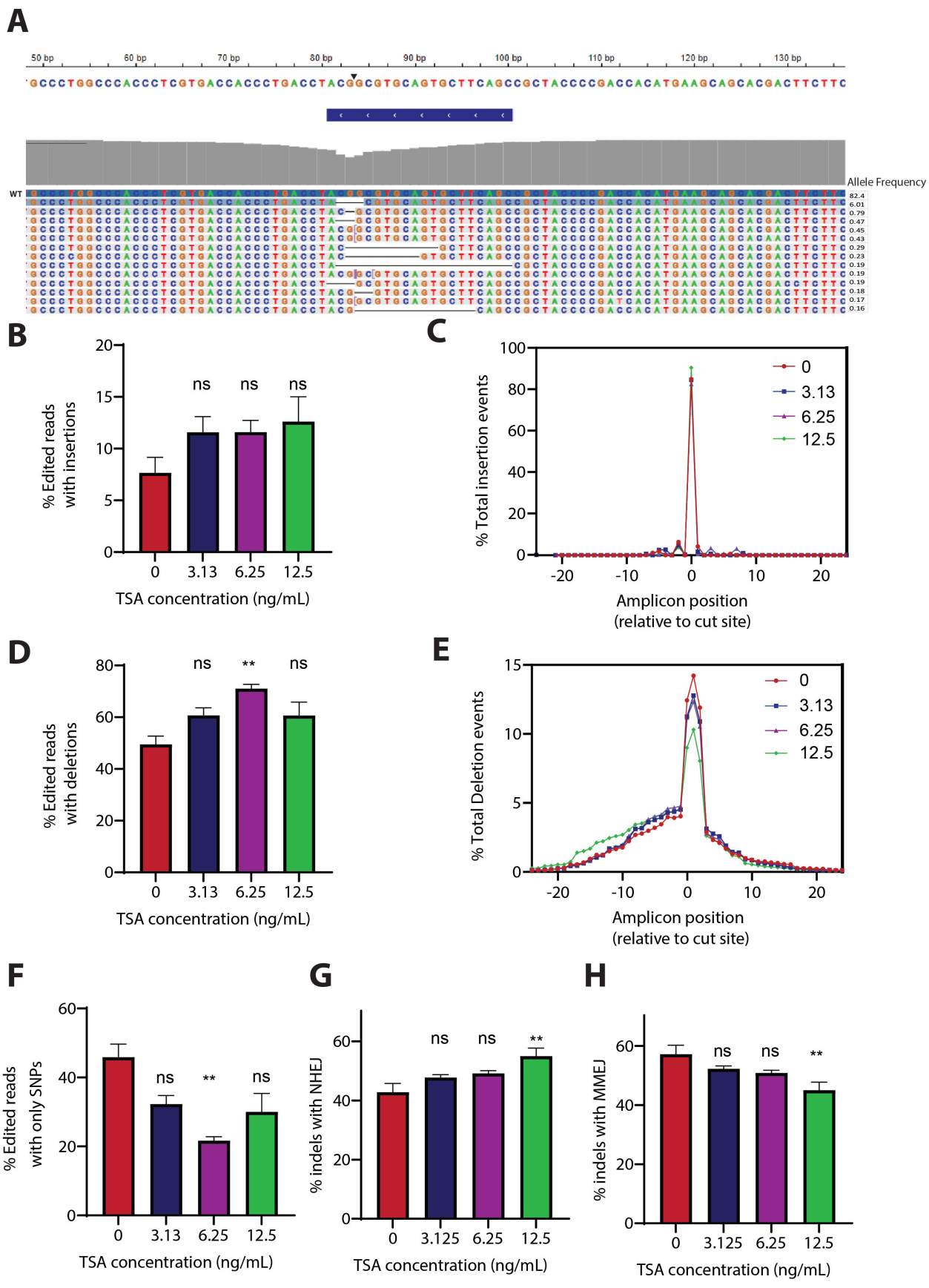


Figure 4.2: (Caption next page.)

Figure 4.2: **Deep sequencing reveals shift in indel profiles upon TSA treatment.** A) Representative indel profile for iPSCs edited with mEGFP targeting gRNA, TSA treatment: 6.25 ng/mL. Allele frequencies are indicated on the right. B) Bar graphs showing increased % edited reads with insertions upon TSA treatment. C) Representative insertion profiles 20 bp around the gRNA cut site. D) Bar graph showing increased % edited reads with deletions upon TSA treatment. E) Representative deletion profiles 20 bp around the gRNA cut site. Bar graphs showing F) increased % edited reads with only SNPs, G) increased %indels with NHEJ, and H) decreased % indels with MMEJ, upon TSA treatment. Data represented in bar graphs are represented as mean ± SEM, n = 3 technical replicates per condition , *p*-values generated by non-parametric Kruskal-Wallis test for multiple unmatched comparisons for multiple comparisons to 0 ng/mL TSA; ns = p $\geqslant$0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001.

Off-target profiling and karyotypic analysis of edited iPSCs

Highly sensitive genome-wide, off-target analysis for our TSA-based editing strategy was assayed by CHANGE-seq [115], which yielded a frequency distribution of the potential off-target sites(**Figure 4.3A,B, Table C.2**). Top 12 off-target sites were amplified from genomic DNA using the rhAmp-Seq system (IDT) and subsequently sequenced using the Illumina MiniSeq system. Sequencing data was then analyzed using the CRISPAltRations tool [324], which allowed for quantitative analysis of observed indel mutations (**Table C.5**). We found that indel reads at off-target sites primarily constituted of SNPs and indel% at off-target sites did not increase upon TSA treatment (**Figure4.3C-F**). Furthermore, we found that the highest normalized on-target edit ratio was obtained for TSA concentration of 6.25 ng/mL, representative of the optimal TSA concentration for generating gene editing iPSCs. Corroborating with this observation, we did not note any significant increase in local chromatin accessibility at the off-target sites upon TSA treatment (**Figure C.3**).

Multiple gene-edited iPSC clones were isolated from the 6.25 ng/mL TSA treatment condition and expanded for subsequent characterization for pluripotency and any genomic abnormalities. Bulk-RNA sequencing (**Table C.6**) indicated that TSA treatment did not decrease gene expression of pluripotency marker genes (*OCT4*, *NANOG*, *SOX2*, *TRA-1-60*), indicating that TSA treatment does not affect the pluripotency of iPSCs. Besides, we observed no changes in iPSC morphology during long-term culture of the edited cells, indicating that TSA treatment does not cause aberrations in iPSC morphology. Karyotyping analysis indicated no genomic abnormalities in four of the five clones, while one clone showed an interstitial duplication in the long (q) arm of chromosome 20 in five of the twenty cells examined. This is a known recurrent acquired duplication at this location in human pluripotent stem cell cultures.

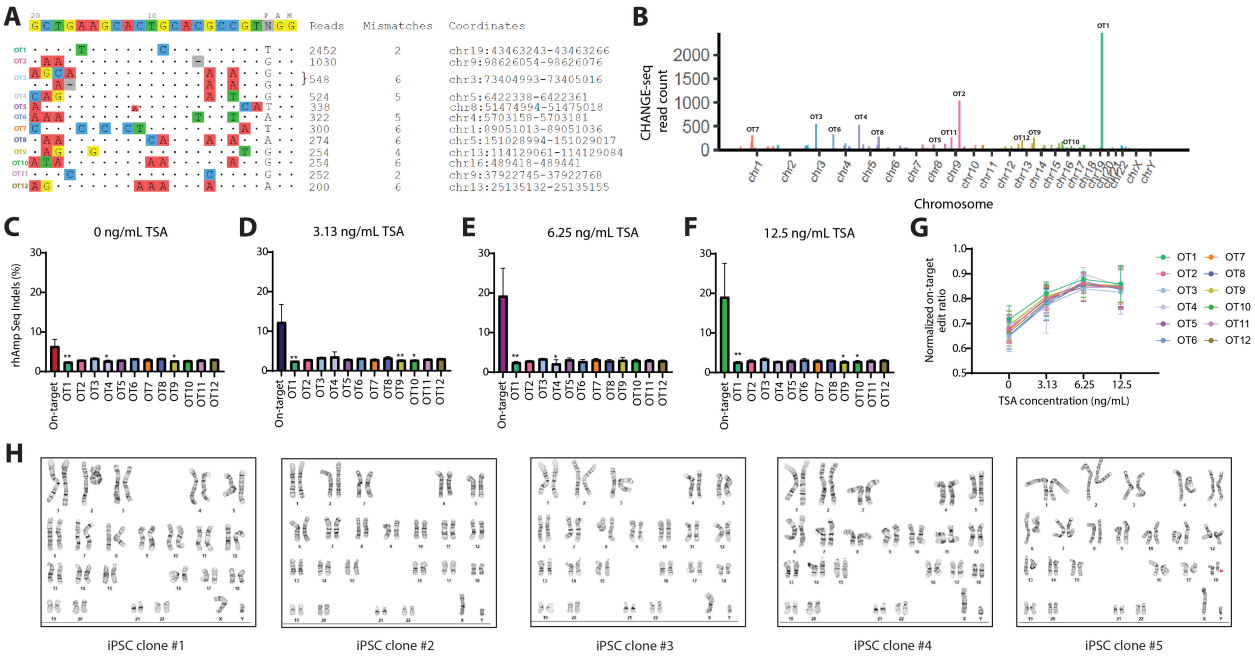


Figure 4.3: (Caption next page.)

Figure 4.3: **Off-target and karyotypic analysis of edited iPSCs.** A) Visualization of sites detected by CHANGE-seq. The intended target mEGFP sequence is shown in the top line. Cleaved sites (off-target) are shown below and are ordered top to bottom by CHANGE-seq read count, with mismatches to the intended target sequence indicated by colored nucleotides. Insertions are shown in smaller lettering between genomic positions, deletions are shown by (-). Note that output is truncated to the top 12 sites. B) Manhattan plot of CHANGE-seq-detected off-target sites organized by chromosomal position with bar heights representing CHANGE-seq read count. C-F) Indel% at mEGFP on-target and top 12 off-target sites detected by CHANGE-seq, assayed by rhAmpSeq system. While the on-target indel % increases with TSA concentration, the off-target indel % remains the same. G) Normalized on-target edit ratio for each of the 12 off-target sites plotted as a function of TSA concentration. TSA concentration of 6.25 ng/mL yields the highest normalized on-target edit ratio. H) Four of the five edited isolated iPSC clones (TSA treatment; 6.25 ng/mL) showed normal karyotype suggesting that no major chromosome abnormalities after TSA-induced gene editing. Data represented in bar graphs are represented as mean $\pm$ SEM, n = 3 technical replicates per condition, *p*-values generated by non-parametric Kruskal-Wallis test for multiple unmatched comparisons to on-target site; ns = p $\geqslant$0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001.

TSA promotes gene editing at multiple open and closed chromatin loci

Finally, we selected multiple gRNA sequences targeting different sites of open (*HIST1H2BJ* and *AAVS1*) and closed chromatin (*TRAC*) to investigate if the observed TSA-induced increase in gene-editing efficiency is specific to the gRNA sequence or the chromatin state of the target locus. We noted that TSA treatment increased gene-editing efficiency at all the sites independent of the gRNA sequence and initial chromatin state of the loci (**Figure 4.4A**). Corroborating with this observation, we noted an increase in the local chromatin

accessibility at the on-target sites upon TSA treatment (**Figure C.4A**).

To further study the impact of TSA on iPSCs, we performed bulk-RNA sequencing on TSA-treated iPSCs, and preliminary analysis showed that there were several upregulated and downregulated genes upon TSA treatment (**Table C.6**), which are involved in cell cycle regulation (*MAGEA4*, *DHRS2*), cell signaling (*STMN2*) and cell growth regulation (ERBB2, *AHRR*). A more detailed investigation with more replicates can allude to pathways involved in TSA-induced increase in gene editing efficiency.

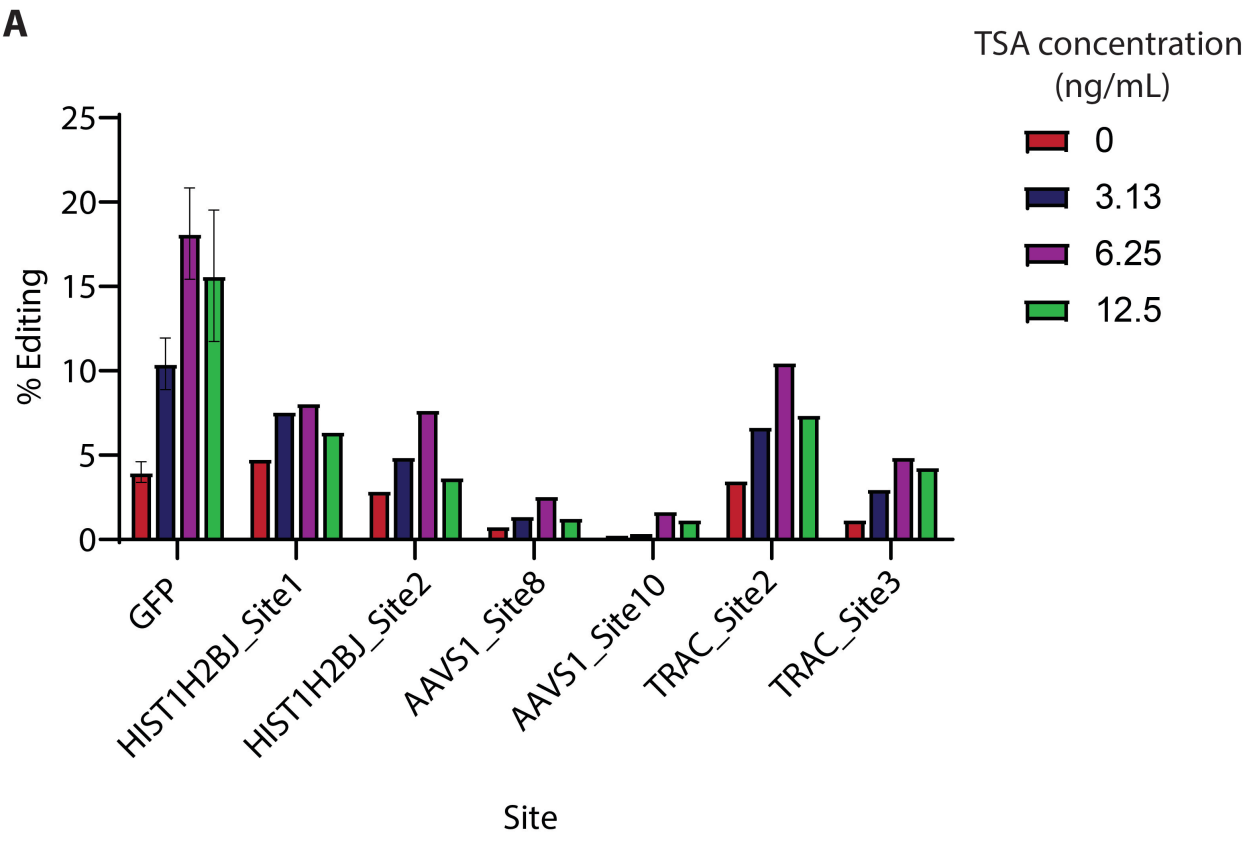


Figure 4.4: **TSA increases gene editing efficiency at multiple open and closed chromatin loci.** Bar graphs showing increased % editing efficiencies upon TSA treatment for 2 sites each at *HIST1H2BJ* (open), *AAVS1* (open), and *TRAC* (closed) loci (n=1 technical replicate).

4.4 Discussion

Overall, we described an easy and rapid chromatin modulation-based CRISPR-Cas9 workflow to manufacture gene-edited iPSCs efficiently. We used TSA (HDAC inhibitor) as the chromatin modulator, quantified TSA-induced chromatin decondensation using nuclear imaging, and subsequently edited multiple open and closed chromatin loci in *HIST1H2BJ* tagged mEGFP reporter iPSCs.

Relative to the traditional CRISPR-Cas9 gene editing workflow, our streamlined manufacturing workflow can: 1) increase the gene-editing efficiency up to ~3.5 fold while ensuring no increased off-target mutations, and maintenance of pluripotency and genomic integrity; 2) increase DNA repair by NHEJ relative to MMEJ; and 3) identify iPSCs more amenable to gene editing based on nuclear imaging outputs, thus enabling the reduction in cost and times associated with culturing and passaging to isolate rare edited iPSC clones.

To gain deeper insights into the mechanisms involved in TSA-induced gene editing and push capabilities of this strategy, we performed scATAC and bulk-RNA sequencing experiments to identify genome-wide changes in chromatin accessibility and gene expression. While our imaging pipeline revealed TSA-induced chromatin condensation changes at the global level, scATAC-seq analysis revealed local changes in chromatin accessibility at different regions in the genome (data not shown). Moreover, preliminary bulk-RNA sequencing revealed several differentially regulated genes (**Table C.6**) across the genome. However, further experiments and analyses need to be performed to identify the relevant differentially regulated genes and pathways in a statistically significant manner. The knowledge obtained from these sequencing studies can then be utilized to engineer the global epigenome using small molecules or the local epigenome using strategies like siRNA [327] and nuclease dead Cas9 (dCas9) [328], to further increase the gene editing efficiency of iPSCs.

Additionally, our workflow can also enable the assessment of other biological processes like cell proliferation, apoptosis, and DNA damage via imaging of GFP, CellEvent

Caspase 3/7 Detection reagent [329], and immunostained histone H2AX phosphorylation [330], respectively. Thus, the action of small molecules or any other perturbation and the impacted processes could be tracked in parallel in each well via high-content nuclear imaging. This study can also be extended in the future to study NHEJ vs HDR (using an HDR activated GFP to BFP switch), to other natural [87] or engineered [331] Cas enzyme variants, to other epigenetic modifiers using GFP tagging approach, and to other cell types including primary cells, cardiomyocytes, T-cells, and neural progenitors. Finally, since iPSCs are a heterogeneous population, studying changes at the single-cell level may help elucidate genome editing mechanisms more than looking at bulk level data. scATAC-seq data generated from transfected iPSCs can be coupled to genome editing outcome (via genotyping of transcriptomes [332]) for a single cell level look at the differences in the transcriptional program between edited and unedited cells. The use of single-cell sequencing-based techniques may also be useful where GFP tagging or confocal imaging is not facile (e.g. primary T-cells).

4.5 Materials and Methods

Cell culture

Mono-allelic mEGFP-tagged *HIST1H2BJ* human iPSCs (AICS-0061-036) were obtained from Allen Institute for Cell Science. This cell line was derived from the WTC parental line (GM25256) released by the Conklin Laboratory at the J. David Gladstone Institute. iPSCs were maintained in mTeSR1 medium on Matrigel (WiCell) coated tissue culture polystyrene plates (BD Falcon). Cells were passaged every 4–5 days at a ratio of 1:8 using ReLeSR solution (STEMCELL Technologies). All cells were maintained at 37°C in 5% CO_2 and tested monthly for possible mycoplasma contamination.

Chemical reagents

Trichostatin A (Millipore Sigma) was resuspended in DMSO at stock concentrations of 2 mg/mL. Aliquots were then stored at -20° C. The purity of the inhibitor was assessed by Millipore Sigma (>99%).

Cell viability

Flow cytometry was performed using Ghost Dye™ Red 780 viability dye (Tonbo Biosciences) to determine the dose range of the TSA that can be administrated to the cells without affecting cell viability. iPSCs (15,000 per well) were cultured overnight and incubated with TSA in 96-well plates for 24 h. The next day, cells were singularized using Accutase (STEMCELL technologies), washed with PBS, centrifuged at 300g for 5 minutes, and Ghost Dye™ Red 780 viability dye was added at 1:1000 concentration for 20 minutes at room temperature. Cells were then washed with PBS, spun down at 300g for 5 minutes, and resuspended in 300 µL of PBS. Cells were run on an Attune NxT™ flow cytometer (Thermo Fisher Scientific) and subsequent analysis was performed using Flowjo software (BD).

SpyCas9 RNP preparation

RNPs were produced by complexing a two-component gRNA to SpyCas9. In brief, tracrRNA and crRNA were ordered from IDT, suspended in nuclease-free duplex buffer at 100 µM, and stored in single-use aliquots at -20°C. tracrRNA and crRNA were thawed, and 0.0625 µL of each component was mixed 1:1 by volume and annealed by incubation at room temperature for 5 minutes to form a 50 µM sgRNA solution for each well of a 96 well plate. Recombinant sNLS-SpCas9-sNLS Cas9 (Aldevron, 10mg/mL) was added to the complexed gRNA at a 1:1 molar ratio (1000 ng/well, Total 0.1 µL) and incubated for 5 minutes at room temperature to form RNP.

RNP delivery

iPSCs were singularized using Accutase and counted using a Countess II FL Automated Cell Counter (Thermo Fisher Scientific) with 0.4% Trypan Blue viability stain (Thermo Fisher Scientific). iPSCs were then seeded at 15,000 cells/well on 96 well glass bottom plates (Cellvis) in mTeSR1 (WiCell) and 10 μM ROCK inhibitor (Y27632, Selleckchem), two days before lipofection or electroporation. On the following day, iPSCs were treated with TSA (0-200 ng/mL) for 16 – 24 hours.

iPSC lipofections were performed using 0.5 μL Lipofectamine Stem Cell Reagent/well (1000 ng Cas9/well and sgRNA at 1:1 molar ratio). Cells remained undisturbed for 48 hours and were then passaged 1:4 using ReLeSR solution followed by daily mTeSR1 media changes for 4 additional days before downstream analysis.

iPSC electroporations were performed using the 4D-Nucleofector System (Lonza) as per the manufacturer's instructions. Briefly, iPSCs were harvested using Accutase (STEMCELL Technologies) and counted. 2×10^5 cells per electroporation were then centrifuged at 300g for 5 min. Media was aspirated and cells were resuspended using 20 μl of P3 solution (Lonza) with 3μg of Cas9 and sgRNA at a 1:1 molar ratio. iPSCs were then electroporated using protocol CB-150. After nucleofection, samples were incubated in nucleocuvettes at room temperature for 15 min before plating into 6×10^4 cells per well on 96 well glass bottom plate in mTeSR1 media +10μM ROCK inhibitor. Media was changed 24 hours post-transfection and replaced with mTeSR1 medium.

Flow cytometry and fluorescence activated cell sorting

Flow cytometry was performed on singularized iPSCs using AttuneNxt flow cytometer (ThermoFisher Scientific) and analyzed using the FlowJo Software. Ghost Dye™ Red 780, ATTO 550, and GFP fluorescence were detected using 780/60, 585/16, and 530/30 filters in BL1, YL1, and RL3 positions, respectively. Gates were established by running singularized untransfected iPSCs. The percentage of viable cells was calculated as the ratio

of Ghost Dye™ Red 780- cells to the total number of single cells. Transfection efficiency was calculated as the percentage of ATTO 550^{+} cells to the total number of viable cells on day 2 post-transfection. Gene editing efficiency was calculated as the percentage of GFP^{-} cells to the total number of viable cells on day 6 post-transfection. For the sorting experiments, iPSCs were singularized 6 days post-transfection using Accutase, washed with PBS + 20 μM ROCK inhibitor, centrifuged at 300g for 5 minutes. Ghost Dye™ Red 780 viability dye (Tonbo Biosciences) was then added at 1:1000 concentration for 20 minutes at room temperature. Cells were washed again with PBS + 20 μM ROCK inhibitor, spun down at 300g for 5 minutes, and resuspended in 300 μL of FACS buffer (PBS + 2% BSA + 20 μM ROCK inhibitor). GFP^{-} Cells were then sorted into tubes containing mTeSR1 + 20 μM ROCK inhibitor, and seeded onto Matrigel-coated 6 well polystyrene plates at 100 cells/well to obtain single-cell iPSC clones.

Next Generation Sequencing of genomic DNA

DNA was isolated from iPSCs by adding 50 μL of DNA QuickExtract/well (Epicentre) following treatment by Accutase and centrifugation. The DNA extract solution was incubated at 65°C for 15min, 68°C for 15min, and finally 98°C for 10min. Genomic PCR was performed according to the manufacturer's instructions using Q5 Hot Start polymerase (NEB); primers are listed in **Table C.1**. Sequencing indices were added with a second round of PCR using indexing primers (IDT), followed by a purification using AMPure XP magnetic bead purification kit (Beckman Coulter). Samples were pooled and sequenced on an Illumina MiniSeq at a run length of 1 × 150 bp or 2 × 150 bp according to the manufacturer's instructions. Analysis was performed using CRISPR RGEN (rgenome.net).

Genome-wide, off-target analysis

Genomic DNA from iPSCs was isolated using Gentra Puregene Kit (Qiagen) according to the manufacturer's instructions. CHANGE-seq was performed as previously described

[115]. Briefly, purified genomic DNA was tagmented with a custom Tn5-transposome to an average length of 400 bp, followed by gap repair with Kapa HiFi HotStart Uracil + DNA Polymerase (KAPA Biosystems) and Taq DNA ligase (NEB). Gap-repaired tagmented DNA was treated with USER enzyme (NEB) and T4 polynucleotide kinase (NEB). Intramolecular circularization of the DNA was performed with T4 DNA ligase (NEB) and residual linear DNA was degraded by a cocktail of exonucleases containing Plasmid-Safe ATP-dependent DNase (Lucigen), Lambda exonuclease (NEB), and Exonuclease I (NEB). *In vitro* cleavage reactions were performed with 125 ng of exonuclease-treated circularized DNA, 90 nM of SpCas9 protein (NEB), NEB buffer 3.1 (NEB) and 270 nM of sgRNA, in a 50 µL volume. Cleaved products were A-tailed, ligated with a hairpin adaptor (NEB), treated with USER enzyme (NEB), and amplified by PCR with barcoded universal primers NEBNext Multiplex Oligos for Illumina (NEB), using Kapa HiFi Polymerase (KAPA Biosystems). Libraries were quantified by qPCR (KAPA Biosystems) and sequenced with 151 bp paired-end reads on an Illumina NextSeq instrument. CHANGE-seq data analyses were performed using open-source CHANGE-seq analysis software [115].

To determine the indel frequency at CHANGE-seq-identified off-target sites, on- and off-target sites for were amplified from iPSC genomic DNA obtained using rhAmpSeq system (IDT), with primers listed in **Table C.3**, and sequencing libraries were generated according to the manufacturer's instructions. Sequencing was then performed with 150-bp paired-end reads on an Illumina Miniseq instrument. Analysis was performed using CRISPAltRations: IDT rhAmpSeq CRISPR analysis tool [324].

Nuclei isolation and single-cell ATAC sequencing

Isolation, washing, and counting of nuclei suspensions were performed according to the Demonstrated Protocol: Nuclei Isolation for Single Cell ATAC Sequencing (10x Genomics; CG000169 Rev D). Briefly, 1,50,000 cells were added to a 2 mL microcentrifuge tube and centrifuged at 300g for 45 min at 4°C. The supernatant was removed without disrupting

the cell pellet, and 100 µL chilled Lysis Buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% Nonidet P40 Substitute, 0.01% digitonin and 1% BSA) was added and mixed 10 times. The microcentrifuge tube was then incubated on ice for 5 min. Following lysis, 1 ml chilled Wash Buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20 and 1% BSA) was added and the resulting solution was mixed 5 times. Nuclei were centrifuged at 500g for 5 min at 4°C and the supernatant was removed without disrupting the nuclei pellet. Nuclei were resuspended in chilled Diluted Nuclei Buffer (10x Genomics; 2000153) at approximately 6,000 nuclei per µL. The resulting nuclei concentration was then determined with a Countess II FL Automated Cell Counter using trypan blue inclusion. Nuclei were then immediately used to generate scATAC-seq libraries according to the Chromium Single Cell ATAC Reagent Kits User Guide v1.1 (10x Genomics; 000209 Rev D). Libraries were sequenced using the Illumina NovaSeq 6000 system using S1 flow cell and target depth of 30k-37k reads/nuclei. FASTQ files were aligned with Cellranger ATAC 2.0 to custom reference Human GRCh38 genome. Downstream analyses were performed using the ArchR software package [333] in R (**Code 2**).

RNA extraction and Bulk RNA sequencing

RNA was extracted from iPSCs in one well of 6-well plate using the GenElute Total RNA Purification Kit (Millipore Sigma) according to the manufacturer's instructions. RNA was then quantified using Nanodrop 2000 (ThermoFisher Scientific) and the integrity of samples were confirmed with the following three criteria for inclusion: 1) a concentration$>$ 50ng/ml, 2) an A260/A280 rating of 1.8 and 2.1, and 3) an A260/A230 ratio$>$ 1.8. Total RNA was submitted to the University of Wisconsin-Madison Biotechnology Center and was verified for purity and integrity via the NanoDrop One Spectrophotometer and Agilent 2100 BioAnalyzer, respectively. Samples that met the Illumina sample input guidelines were prepared according to the TruSeq® Stranded Total Sample Preparation Guide (1000000040499 v00) using the Illumina® TruSeq® Stranded Total Sample Preparation kits (Illumina Inc., San

Diego, California, USA) and sequenced on an Illumina NovaSeq 6000 system at a run length 2 × 150 bp and target read depth of 30 million reads/sample at the University of Wisconsin–Madison Biotechnology Center. For analysis, alignments and gene counts were generated against GENCODE human transcriptome version 37 using Salmon [334]. Transcripts were annotated to the gene level using the tximport package in [335] and exported to CSV format (**Code 3**). Normalized counts were then extracted and genes with read counts < 10 were filtered out. Fold change was then calculated as the ratio of the reads of the TSA-treated iPSCs to the untreated iPSCs.

Antibodies and Staining

All cells were fixed for 15 minutes with 4% paraformaldehyde in PBS (Sigma-Aldrich, St. Louis, MO) and permeabilized with 0.5% Triton-X (Sigma-Aldrich) for >4 hours at room temperature before staining. Hoechst (H1399; Thermo Fisher Scientific, Waltham, MA) was used at 5 µg/mL with 15 min incubation at room temperature to stain nuclei. Primary antibodies were applied overnight at 4°C in a blocking buffer of 5% donkey serum (Sigma-Aldrich) at the following concentrations: H3K9Me3 (ab8898; Abcam) 1:500; H3K9Ac (39918; Active Motif) 1:100; H3K27Me3 (C36B11; Cell Signaling Technology) 1:200. Secondary antibodies were obtained from Thermo Fisher Scientific and applied in a blocking buffer of 5% donkey serum for one hour at room temperature at concentrations of 1:400 – 1:800. A Nikon Eclipse Ti epifluorescence microscope was used to acquire 100x images. Images were processed using image analysis software CellProfiler [198] to calculate the chromatin condensation values [323], as described in **Figure C.2A**.

Karyotyping

Cells cultured for atleast 5 passages were grown to 60-80% confluence and shipped for karyotype analysis to WiCell Research Institute, Madison, WI. G-banded karyotyping was performed using standard cytogenetic protocols [232]. Metaphase preparations were

digitally captured with Applied Spectral Imaging software and hardware. For each cell line, 20 GTL-banded metaphases were counted, of which a minimum of 5 were analyzed and karyotyped. Results were reported in accordance with guidelines established by the International System for Cytogenetic Nomenclature 2016 [233].

Statistical Analysis

Unless otherwise specified, p-values were calculated using a non-parametric Kruskal-Wallis test for multiple unmatched comparisons with GraphPad Prism software. Statistical tests were deemed significant at $\alpha \leqslant 0.05$. Technical replicates are defined as distinct wells within an experiment. Biological replicates are experiments performed with different passages of iPSCs. No a priori power calculations were performed.

5 CONCLUSION AND FUTURE DIRECTIONS

This thesis aimed to develop new technologies for reprogramming and gene editing of iPSCs to address the challenges in iPSC biomanufacturing. First, I presented microcontact printed platform as a method to visually track the nucleus of reprogramming cell populations in real-time using high-content analysis in a non-destruction fashion. Use of this platform aided in the development of a predictive model to identify and isolate high-quality iPSCs based on nuclear morphometrics. Second, I additionally tracked cellular metabolism during reprogramming on the microcontact printed platform using label-free, high-resolution autofluorescence live-cell imaging. Complementing nuclear morphometrics with cell metabolic measurements resulted in predictive models for iPSC identification with higher accuracy. Finally, I designed a new CRISPR-Cas9 gene editing platform based on chromatin modulation. This platform increased the efficiency of generating gene-edited iPSCs. Overall, these methods represent an important step towards the biomanufacturing of iPSCs for clinical applications.

Efficient generation and identification of reprogrammed iPSCs

In chapter two, we described a new microcontact printed platform that allows for non-destructive dynamic live-cell microscopy of hundreds of microscale cell subpopulations undergoing reprogramming while also preserving many of the biophysical and biochemical cues within the cells' microenvironment. On this substrate, we were able to both watch and physically confine cells into discrete micron-sized islands during the reprogramming of human somatic cells from skin biopsies and blood draws obtained from healthy donors. Using high-content analysis, we identified a combination of eight nuclear morphometric characteristics that can be used to track the progression of reprogramming and distinguish partially reprogrammed cells from those that are fully reprogrammed to iPSCs. These eight nuclear characteristics were then used to construct a regression model (RMSE $=$ 0.15) to

identify and isolate high-quality iPSCs. In chapter three, we complemented these eight nuclear measurements with eleven other cell metabolism measurements using label-free, live cell autofluorescence imaging, called Optical Metabolic Imaging (OMI). Machine learning models were then built to determine the reprogramming status of cells, namely non-reprogrammed somatic cells, partially reprogrammed intermediates, or fully reprogrammed iPSCs. We showed that adding eleven metabolic measurements to the previously identified eight nuclear measurements increased the accuracy of iPSC identification from 70% to 95%. Furthermore, we constructed single-cell reprogramming trajectories based on the metabolic and nuclear measurements and identified four branch points that could determine whether the starting somatic cells become completely reprogrammed to iPSCs.

These studies are among the first in the reprogramming literature to use live-cell imaging-based nuclear and metabolic outputs to predict reprogramming. The use of μCP technology along with machine learning models represented a step forward from the laborious, time-consuming, error-prone lab bench protocol of identifying and isolating iPSCs to a more rapid, automated, and standardized method, which has immense potential to meet the industry standards for iPSC biomanufacturing [239]. Further, this platform could be used to image additional stains or reporters of the cell [47, 191, 210, 336, 337], such as plasma membrane, mitochondria, focal adhesion, mitotic spindle, and chromatin mobility, to complement our studies and add additional information to distinguish each reprogramming state. Such a novel imaging-based analysis can offer an orthogonal approach to investigating mechanisms of human somatic cell reprogramming that will be complementary to studies of transcript and protein expression. The imaging-based approach can also further be used in conjunction with trajectory inference methods [338] to build single-cell reprogramming maps, which could subsequently aid in the identification of novel alternate cell states and additional "roadblocks" during the reprogramming process [180, 202]. Another interesting observation made during our study was the occurrence of nuclear and metabolic measurement shifts that favor reprogramming, upon cell confine-

ment to micron-sized islands on microcontact printed substrates. This was in conjunction with other recent studies that indicated that biophysical microenvironment can affect nuclear [191, 197, 323, 339] and metabolic state of cells [271–273, 340], indicating that cell confinement using our platform could be used as a potential strategy for improving the efficiencies of the reprogramming process. Overall, our approach to track reprogramming *in situ* using micropatterned substrates could aid in the biomanufacturing of therapeutically relevant iPSCs and be used to elucidate multiscale cellular and sub-cellular changes that accompany human cell fate transitions.

Efficient generation of gene edited iPSCs

The microcontact printed platform in chapters two and three focused on the identification and isolation of high-quality iPSCs, an approach to streamlining the manufacturing process of iPSCs. Next, we sought to develop an approach to streamline the process of iPSC gene editing, another major bottleneck in the manufacturing of therapeutically relevant iPSCs. To this end, in chapter four, we developed a novel strategy to modulate the chromatin state of iPSCs to promote gene editing of iPSCs and additionally performed nuclear imaging to identify iPSCs more amenable to gene editing.

We used TSA (HDAC inhibitor) as the chromatin modulator in our study, which is known to promote an open chromatin state. We subsequently edited multiple open and closed chromatin loci in *HIST1H2BJ* tagged mEGFP reporter iPSCs and noted that the gene editing efficiency of iPSCs increased up to ~3.5 fold upon TSA treatment. We additionally demonstrated a shift towards NHEJ over MMEJ upon TSA treatment, indicating that our strategy could potentially be used to specifically engineer gene-edited iPSCs geared towards the DNA repair pathway of user's choice. Further interrogation of the edited iPSC populations showed that our strategy did not compromise with the pluripotency or genomic integrity of edited cell lines and that minimal off-target editing occurred at predicted sites. We also developed a nuclear imaging-based pipeline that is sensitive

to the chromatin changes induced by TSA treatment, thus paving the way for rapid *in situ* identification of iPSCs that enable higher gene editing efficiencies. To gain deeper insights into the mechanisms involved in TSA-induced gene editing and subsequently push capabilities of this strategy, we performed preliminary bulk-RNA sequencing experiments that showed several differentially regulated genes. However, further experiments and analyses need to be performed to identify the differentially regulated genes and pathways in a statistically significant manner.

Our study was supported by several other strategies that showed that chromatin modulation impacts gene editing efficiency [130, 133, 137, 138, 141, 149, 154]. However, our study was the first to perform a thorough on-target indel profiling and off-target profiling of the edited cells. Additionally, we were also the first to develop an imaging-based strategy to identify iPSCs more suitable for gene editing. To broaden the adoption of this strategy, several adaptations could be implemented. Firstly, this study could be extended to study TSA-induced DNA repair pathway balance shift between NHEJ and HDR by implementing a GFP to BFP switch [341], and to other classes of chromatin modulators such as HAT inhibitors [342] and DNMT inhibitors [343]. Secondly, our strategy can also be easily extended to non-GFP reporter cell lines and used in conjunction with live nuclear dyes [226] or label-free nuclear imaging methods [213, 344] to ensure compliance with clinical standards. Thirdly, this strategy can be extended to other natural [87] or engineered [331] Cas enzyme variants (e.g., high fidelity enzymes, base editors) and to other cell types, including iPSC-derivatives, adult stem cells, and primary cells for disease-specific applications. Finally, this study could be implemented on previously established microcontact printed substrates [193] to obtain edited clones on micron-sized islands enabling easier isolation of clonal edited iPSC populations.

Towards more efficient biomanufacturing of iPSCs

Overall, integrating the technologies in this thesis can enable nuclear imaging-based identification of the reprogramming stage most amenable to CRISPR-Cas9 gene editing. This knowledge could further aid in the implementation of simultaneous reprogramming and gene editing [192, 345] on our microcontact printed platform to additionally enhance the efficiency and reduce the timeline of the iPSC biomanufacturing process. Although several challenges remain before the clinical use of iPSCs, we believe that the development of rapid, efficient, and safe reprogramming and genome editing technologies will lead to the advancement of iPSC-based from bench to the bedside. We hope that this thesis contributes to the adaptation of reprogramming and gene editing technologies into the clinic and advances human health.

A SUPPLEMENTAL FIGURES AND TABLES FOR CHAPTER 2

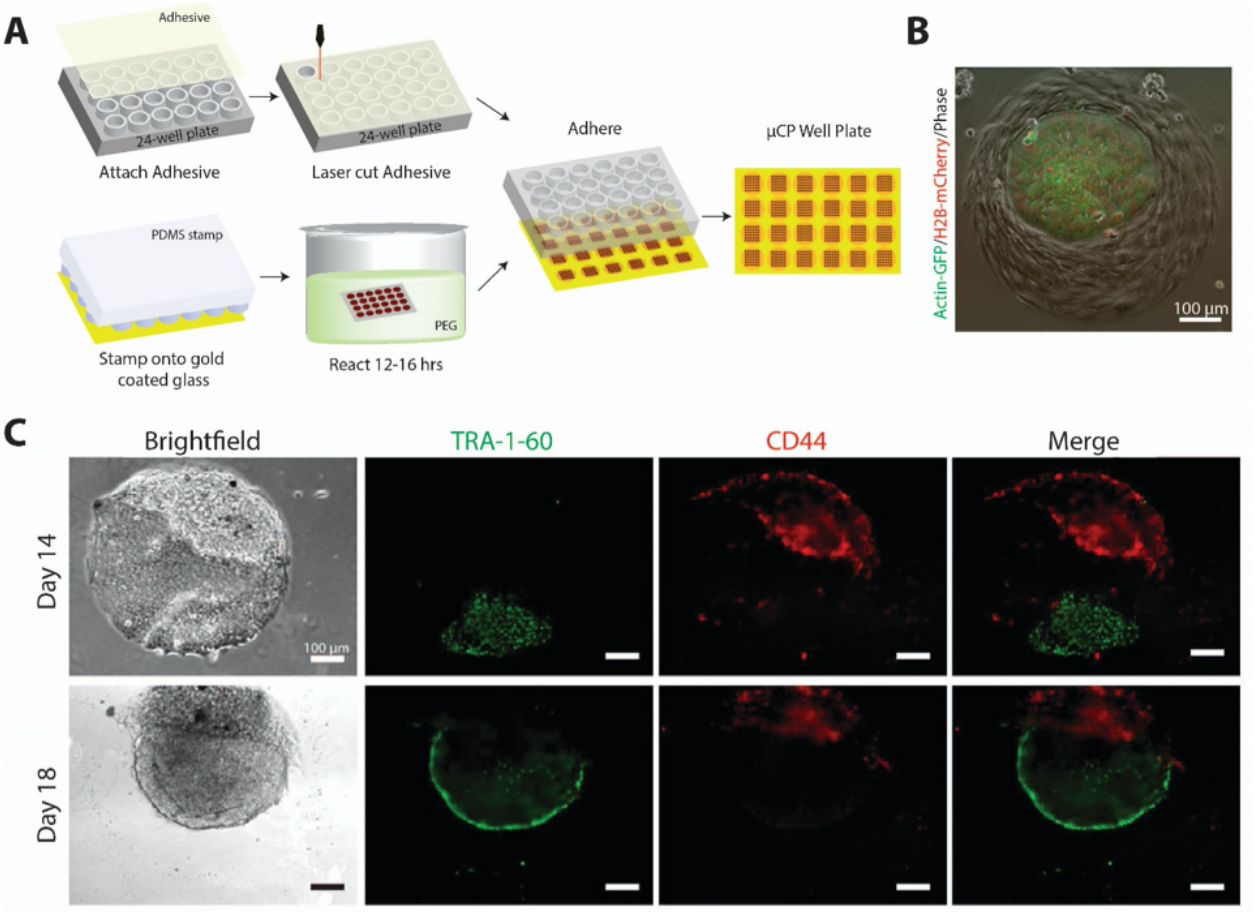


Figure A.1: **Micropatterned plates enable live tracking of reprogramming.** A) Schematic showing the generation of μCP Well Plates. Gold-coated glass is stamped with initiator solution using PDMS mold and then reacted in PEG solution. Glass is then combined with a standard bottomless 24-well plate that has a double side adhesive attached. B) Co-culture of two different cell types on a single μFeature. Brightfield only cells are fibroblasts. H2B-mCherry/Actin-GFP cells are hPSCs. C) Representative images of the progression of a fibroblast subpopulation on a single μFeature through establishment. Cells were live stained using antibodies against TRA-1-60 (green) and CD44 (red).

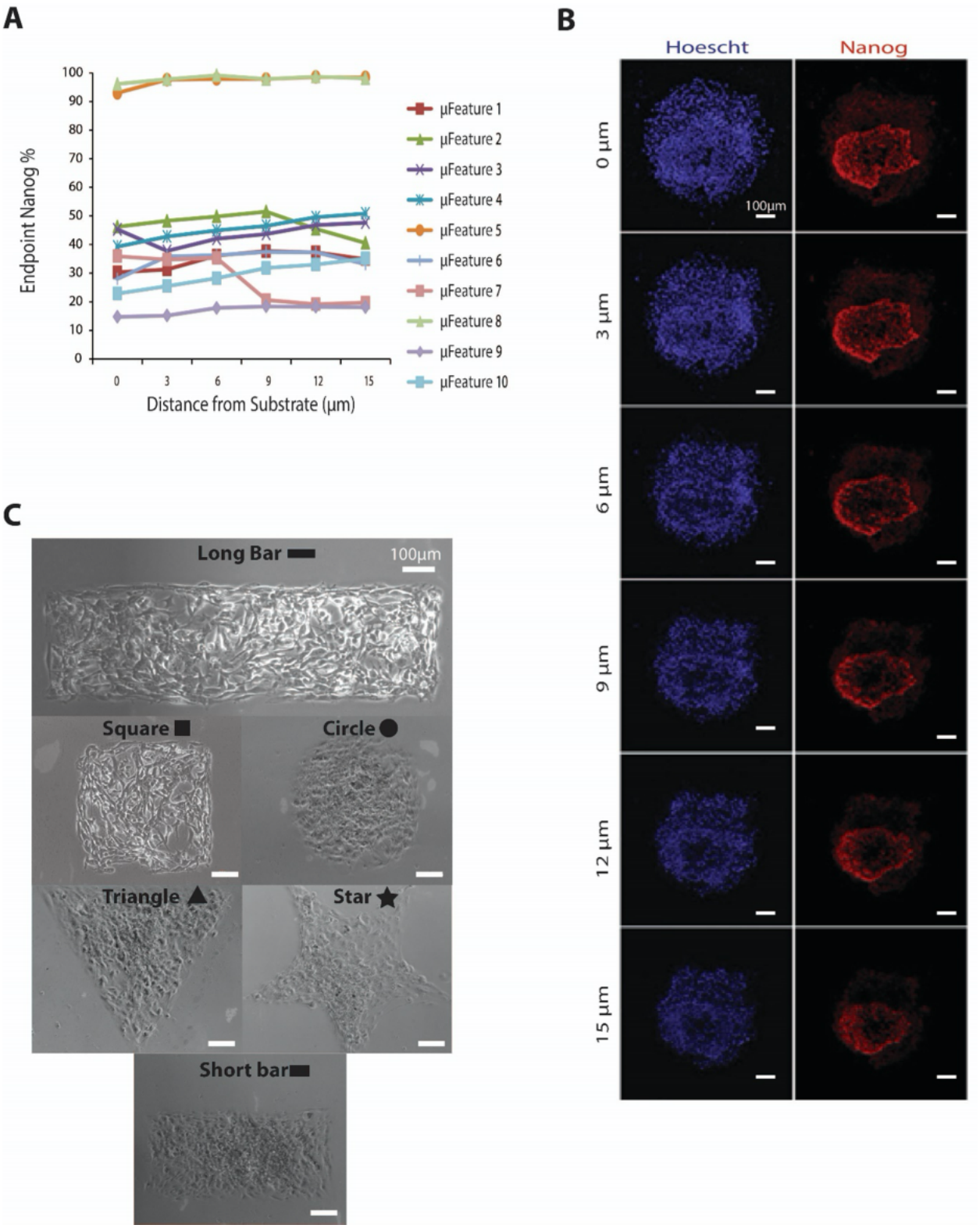


Figure A.2: Caption next page

Figure A.2: **Z-stack analysis and representative micropattern geometries.** A) Endpoint Nanog+ percentage in reprogramming fibroblasts as a function of z-distance from the substrate. Endpoint Nanog+ percentages did not vary with height. B) Representative images of Nanog expression with increasing z-distance. C) Representative bright field images of reprogramming fibroblasts on six different μFeature shapes.

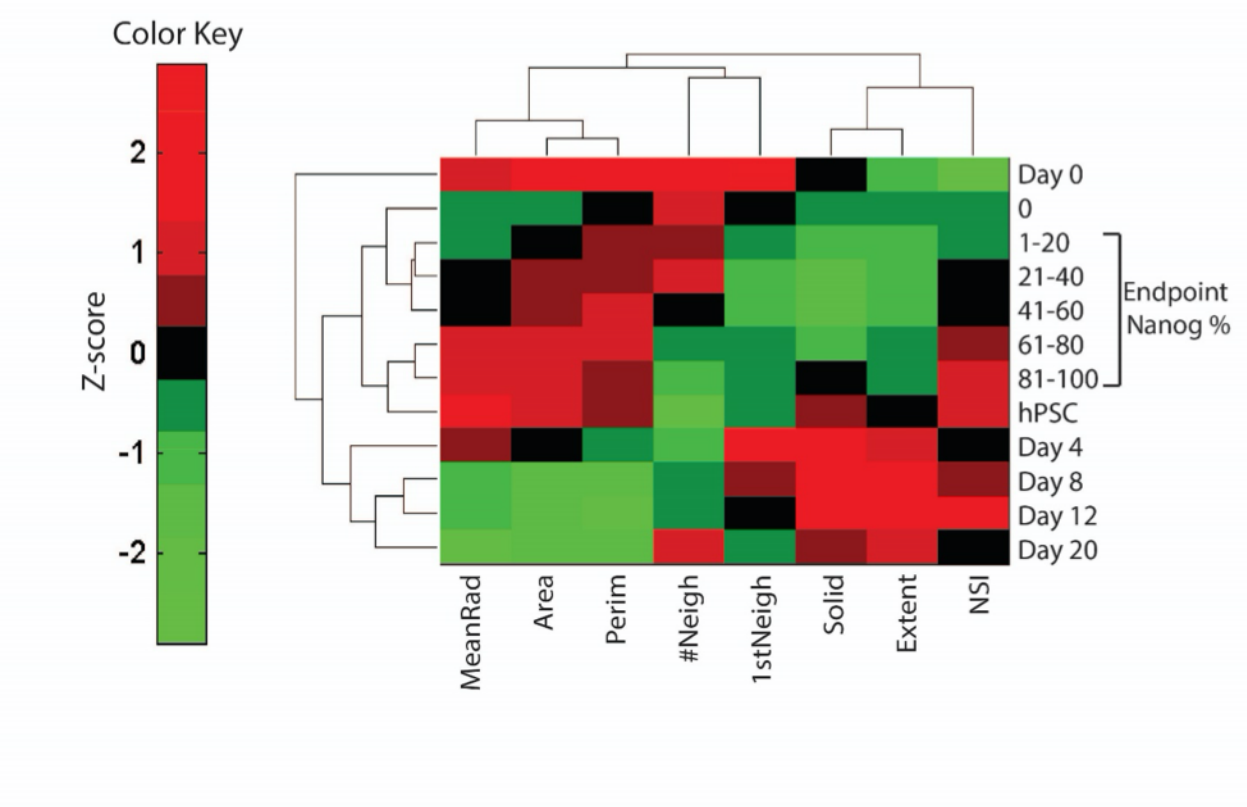


Figure A.3: **Hierarchical clustering analysis of nuclear characteristics.** Hierarchical clustering of each time point in fibroblast reprogramming and endpoint reprogramming fibroblast data (classified into 20% bins of Nanog+ μFeatures). Z-score represents the correlation distance. 60-100% positive μFeatures clustered closely to hPSCs, while reprogramming fibroblast intermediates broadly shared nuclear characteristics. Seeded fibroblasts prior to reprogramming (Day 0) were unlike any other cell type. Clustering based on nuclear characteristics is similar to what was found via PCA loadings in Figure 2.4D.

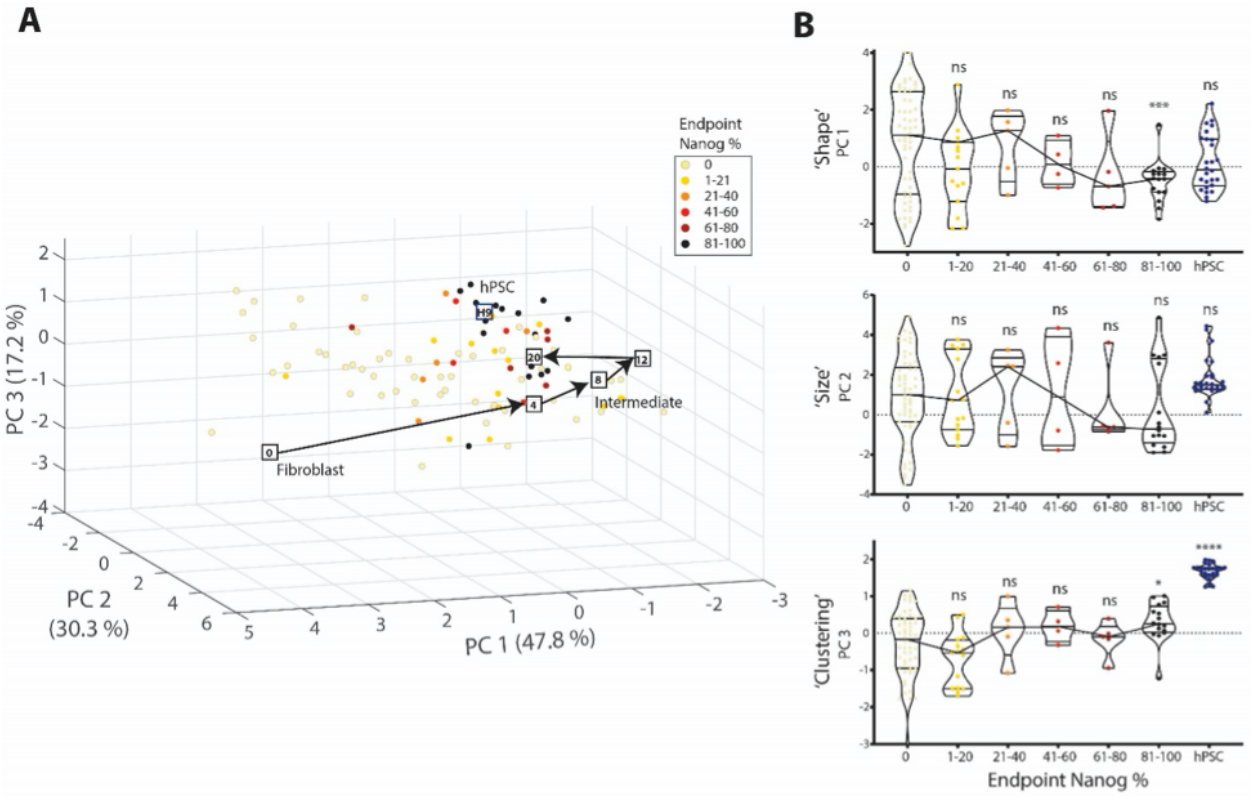


Figure A.4: **Principal component analysis of EPCs reprogramming on 300 μm radius circular features.**A) Fibroblast reprogramming PCA map (centroid values for each reprogramming time point boxed as in **Figure 2.4A**) with endpoint reprogramming EPC data, for circular μFeatures of 300 μm radius, color coded by endpoint Nanog+ percentage. Low Nanog+ features clustered closely to time points of intermediate cells states while features that expressed high levels of Nanog clustered closely to hPSCs. Percent variance explained by the individual PC components is indicated within the parentheses. B) Violin plots showing the distribution of individual PCs as a function of the endpoint Nanog+ percentage for reprogramming EPCs on 300 μm radius circular features (n = 4 biological replicates, total number of endpoint μFeatures = 103, *p*-values generated by one-way ANOVA using the Brown-Forsythe and Welch test for multiple comparisons to 0% endpoint Nanog+ percentage; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$).

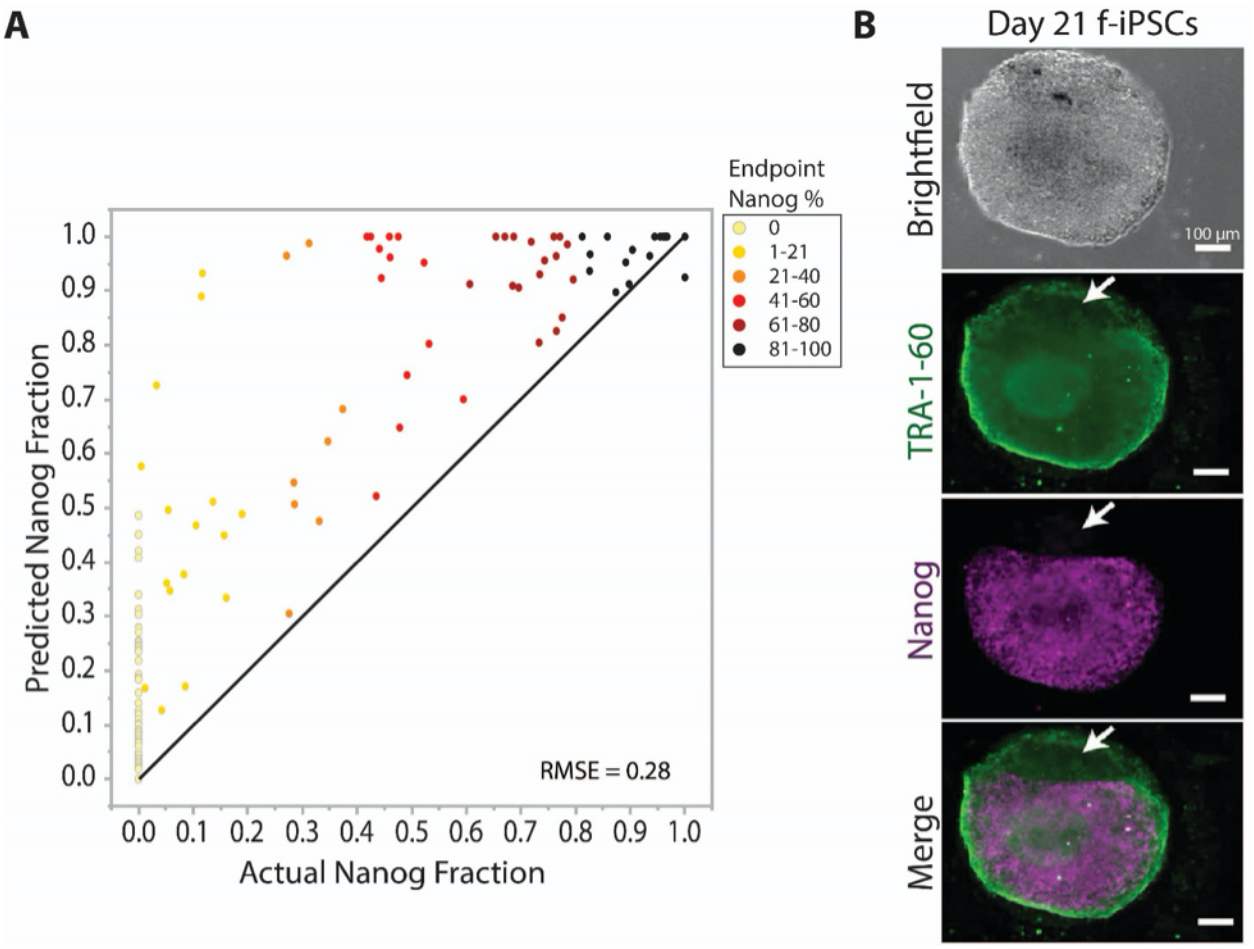


Figure A.5: **Performance of TRA-1-60 as a pluripotency marker.** A) PLSR model for reprogramming fibroblasts predicting Nanog+ fraction using TRA-1-60 staining as input variable. Model was less predictive (RMSE=0.28) than use of nuclear characteristics (n = 3 biological replicates, total number of µFeatures = 114). B) Representative image showing difference between TRA-1-60 expression and Nanog expression in f-iPSCs.

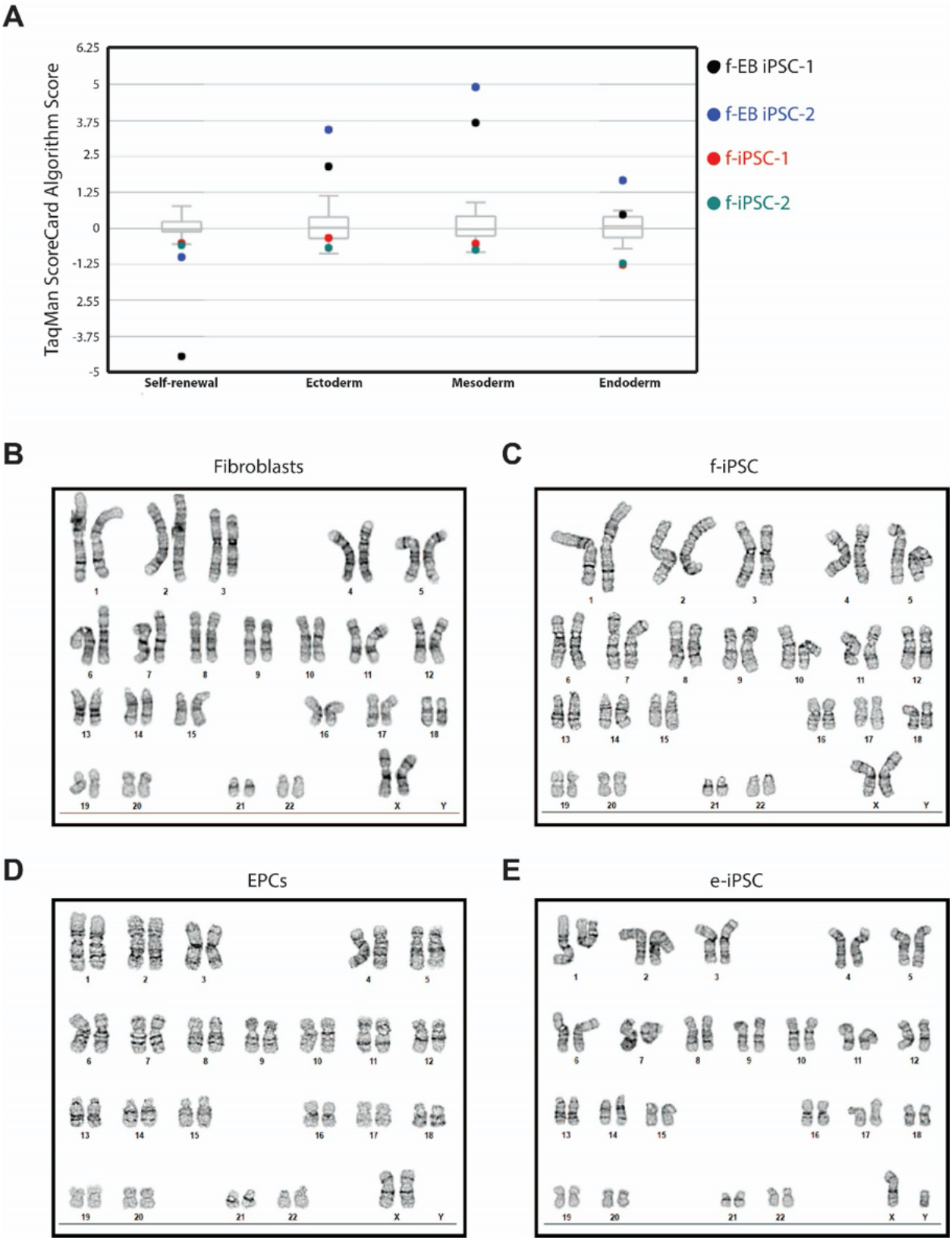


Figure A.6: Caption next page

Figure A.6: **Standardized gene expression from Scorecard™ assay and karyotype analysis.** A) Both f-iPSC clones upregulated genes associated with self-renewal. Embryoid bodies (f-EBs) formed from both clones expressed genes associated with all three germ layers. Data is compared to 9 standard hPSC lines (grey box-and-whisker plots). Karyograms of B) fibroblasts, C) f-iPSCs, D) EPCs, and E) e-iPSCs show normal karyotype suggesting that no major chromosome abnormality was present within cells before or after reprogramming.

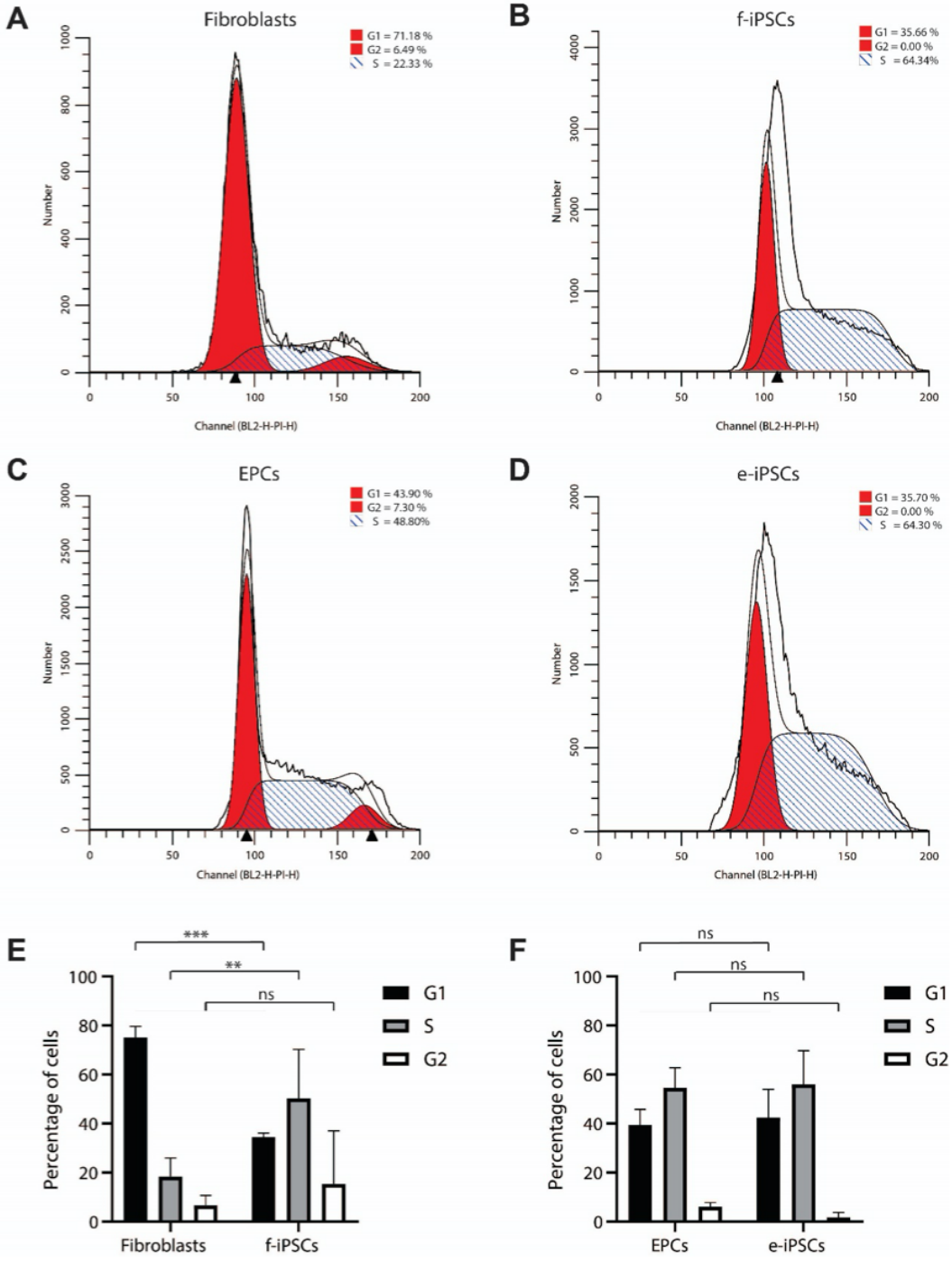


Figure A.7: Caption next page

Figure A.7: **Cell cycle distribution of cells before and after reprogramming.** Representative flow cytometry graphs for A) fibroblasts, B) f-iPSCs, C) EPCs, and D) e-iPSCs. Cells were stained with propidium iodide and analyzed 24 hours after staining by flow cytometry. E) Percentage of fibroblasts and f-iPSCs in various cell cycle phases. f-iPSCs have a higher percentage of cells in S-phase as compared to the starting fibroblasts. F) Percentage of EPCs and e-iPSCs in various cell cycle phases. Both e-iPSCs and starting EPCs have a high percentage of cells in S-phase. Data represent mean $\pm$ SD of triplicates. *p*-values generated by two-way ANOVA using Bonferroni's test for multiple comparisons to fibroblasts or EPCs; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$).

#	Measurement	Description
1	**Area**	The total number of pixels within an object.
2	**Perimeter (Perim)**	The total number of pixels around the boundary of an object.
3	**Nuclear Shape Index (NSI)**	NSI = 4πA/P2, where A is the area and P is the perimeter of an ellipse around the object's major and minor axes. NSI = 1 for a perfect circle and approaches 0 with elongation
4	**Mean Radius (MeanRad)**	The mean distance of any pixel in the object to the closest pixel outside of the object.
5	**Solidity (Solid)**	The proportion of pixels in the convex hull that are also in the object: Object Area/Convex hull Area
6	**Extent**	The proportion of pixels in the bounding box that are also in the object: Object Area/Bounding Box Area
7	**Distance to closest neighbor (1st Neigh)**	The distance to the closest object in pixels.
8	**Number of neighbors (#Neigh)**	The total number of neighbor objects.
9	Compactness	The mean squared distance of the object's pixels from the centroid divided by the area. A filled circle will have a compactness of 1, with irregular objects or objects with holes having a value greater than 1.
10	Eccentricity	The eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1.
11	Form Factor	Calculated as 4π*Area/Perimeter2. Equals 1 for a perfectly circular object.

Table A.1: **Nuclear characteristics used to create reprogramming models.** Description of the 33 nuclear characteristics that were measured. These nuclear characteristics were obtained after processing Hoechst-stained images using a CellProfiler pipeline. Eight core nuclear characteristics used to create reprogramming models are bolded.

#	Measurement	Description
12	Major Axis Length	The length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region.
13	Minor Axis Length	The length (in pixels) of the minor axis of the ellipse that has the same normalized second central moments as the region.
14	Maximum Feret Diameter	The Feret diameter is the distance between two parallel lines tangent on either side of the object. The maximum Feret diameter is the largest possible Feret diameter.
15	Integrated Intensity Edge	The sum of the edge pixel intensities of an object.
16	Integrated Intensity	The sum of the pixel intensities within an object.
17	MAD Intensity	The median absolute deviation (MAD) value of the intensities within the object.
18	Mass Displacement	The distance between the centers of gravity in the gray-level representation of the object and the binary representation of the object.
19	Maximum Intensity	The maximal pixel intensity within an object.
20	Mean Intensity Edge	The average edge pixel intensity of an object.
21	Std Intensity	The standard deviation of the pixel intensities within an object.
22	Distance to second closest neighbor	The distance to the second closest object (in units of pixels).

Table A.1: **Nuclear characteristics used to create reprogramming models.** Description of the 33 nuclear characteristics that were measured. These nuclear characteristics were obtained after processing Hoechst-stained images using a CellProfiler pipeline. Eight core nuclear characteristics used to create reprogramming models are bolded.

#	Measurement	Description
23	Percent Touching	Percent of the object's boundary pixels that touch neighbors, after the objects have been expanded to the specified distance. Note that this measurement is only available if you use the same set of objects for both objects and neighbors.
24-28	Mean Fractional Intensity	Mean fractional intensity at a given radius; calculated as fraction of total intensity normalized by fraction of pixels at a given radius. (Measured from each object's center to its boundary within a set of 5 bins, i.e., rings.)
28-33	Radial CV	Coefficient of variation of intensity within a ring, calculated across 8 slices. (Measured from each object's center to its boundary within a set of 5 bins, i.e., rings.)

Table A.1: **Nuclear characteristics used to create reprogramming models.** Description of the 33 nuclear characteristics that were measured. These nuclear characteristics were obtained after processing Hoechst-stained images using a CellProfiler pipeline. Eight core nuclear characteristics used to create reprogramming models are bolded.

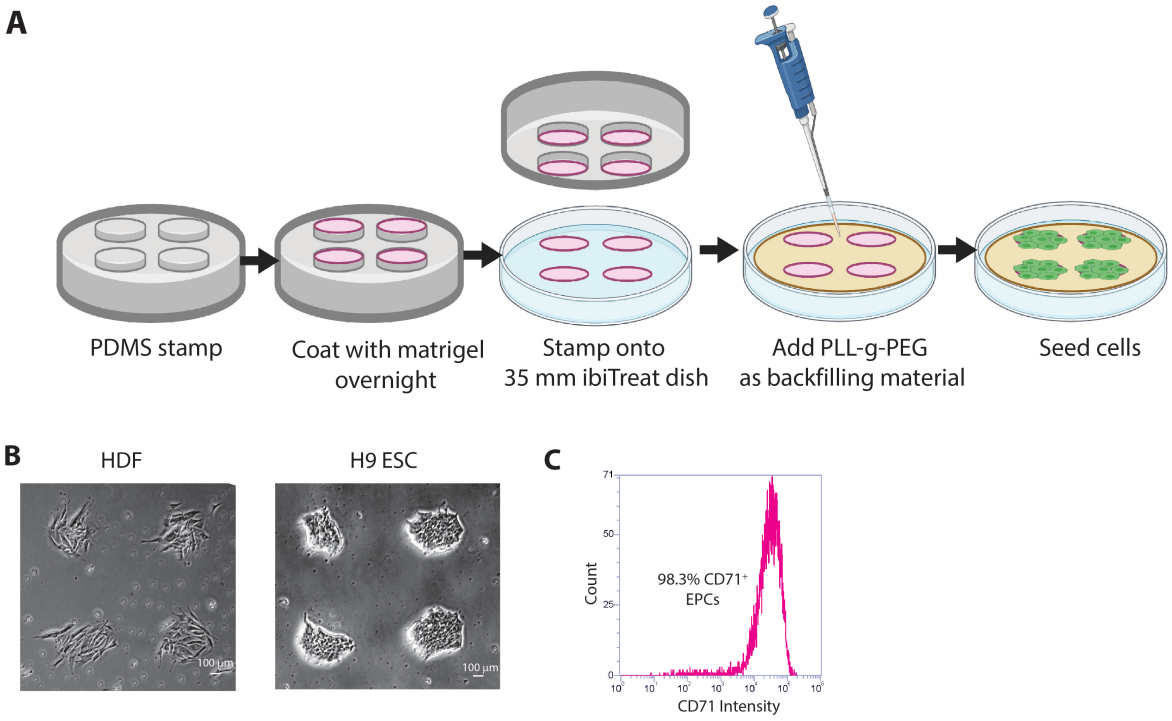


Figure B.1: **Micropatterned substrates enable controlled cell adhesion.** A) Schematic showing the fabrication of µCP substrates. PDMS mold is coated with Matrigel and stamped onto a 35 mm ibiTreat dish. PLL-g-PEG solution is then added to backfill the non-printed regions. Fabricated micropatterned substrates are then ready to be seeded with cells.B) Representative images of HDFs and H9 ESCs adhered to 300 µm radius circular µFeatures on micropatterned substrates. Scale bar, 100 µm. C) Flow cytometry histogram indicating the percentage of CD71+ cells after 10 days of EPC culture and prior to electroporation with reprogramming plasmids.

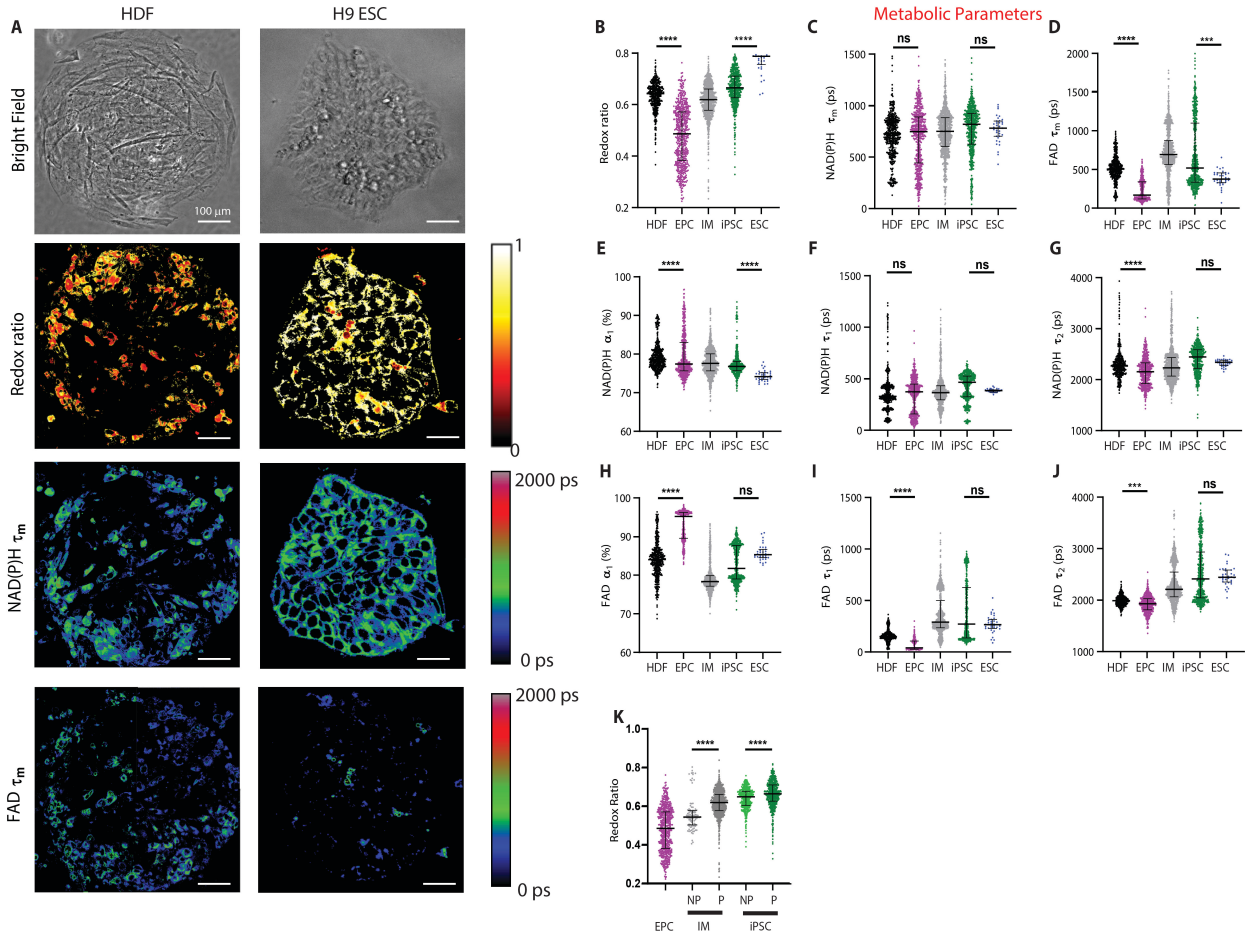


Figure B.2: **Metabolic parameter changes during reprogramming of EPCs.** A) Representative optical redox ratio, NAD(P)Hτ_m and FADτ_m images for HDFs and H9 ESCs. Color bars are indicated on the right. Scale bar, 100 μm. Single-cell quantitative analysis of B) optical redox ratio, C) NAD(P)Hτ_m, D) FADτ_m, E) NAD(P)Hα_1, F) NAD(P)Hτ_1, G) NAD(P)Hτ_2, H) FADα_1, I) FADτ_1, and J) FADτ_2 for HDFs, EPCs, IMs, iPSCs and H9 ESCs (n = 459, 561, 990, 586, 35 respectively). K) Single-cell quantitative analysis of optical redox ratio for non-patterned (NP) and patterned (P) reprogramming cells (IMs and iPSCs). Data are presented as median with interquartile range for each cell type. Statistical significance was determined by one-way analysis of variance (ANOVA) using the Kruskal-Wallis test for multiple comparisons; ns = p $\geqslant$0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001).

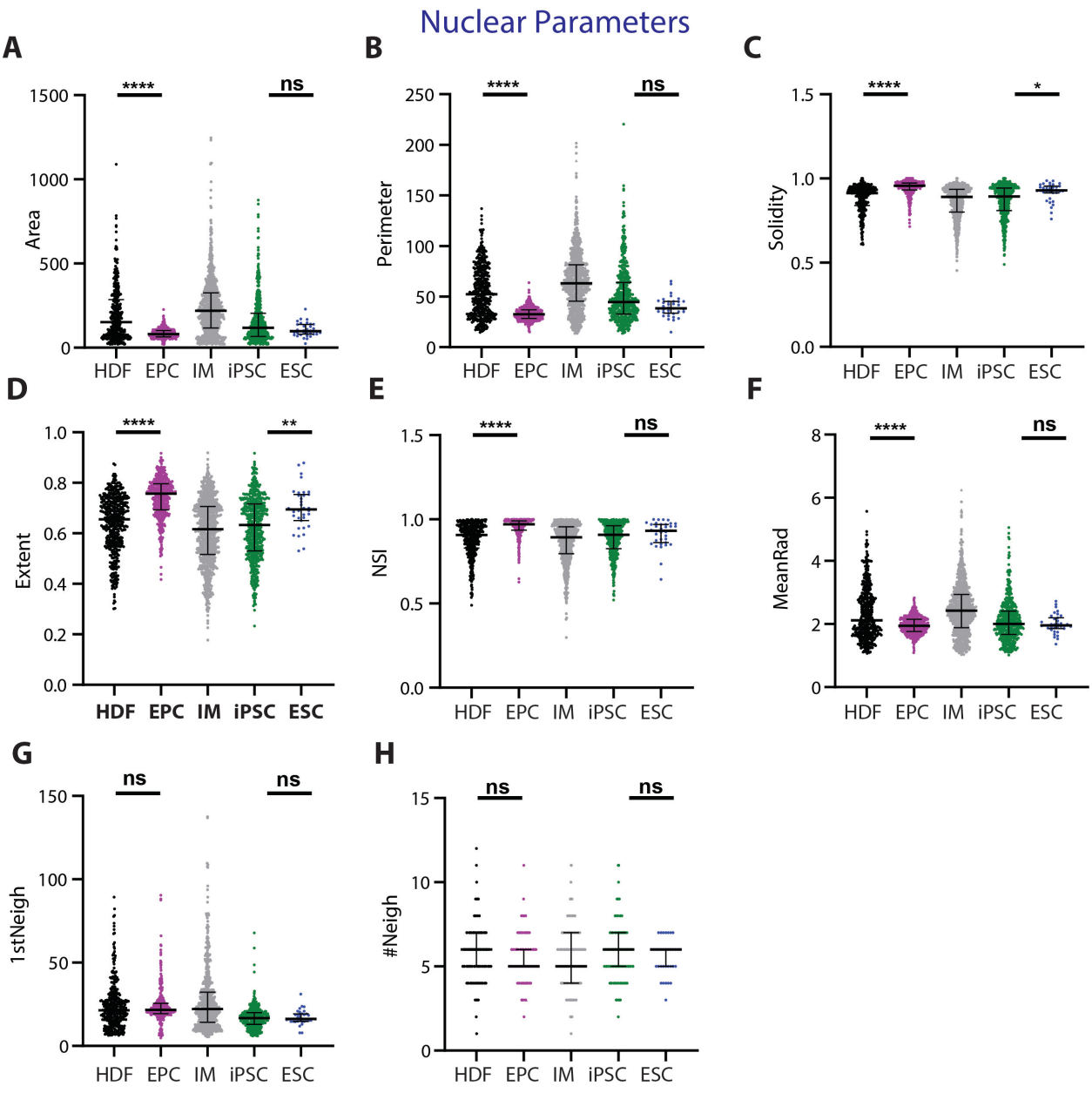


Figure B.3: **Nuclear parameter changes during reprogramming of EPCs.** Single-cell quantitative analysis of A) Area, B) Perimeter, C) Solidity, D) Extent, E) NSI, F) MeanRad, G) 1stNeigh, and H) #Neigh for HDFs, EPCs, IMs, iPSCs and H9 ESCs (n = 459, 561, 990, 586, 35 respectively). Data are presented as median with interquartile range for each cell type. Statistical significance was determined by one-way analysis of variance (ANOVA) using the Kruskal-Wallis test for multiple comparisons; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$).

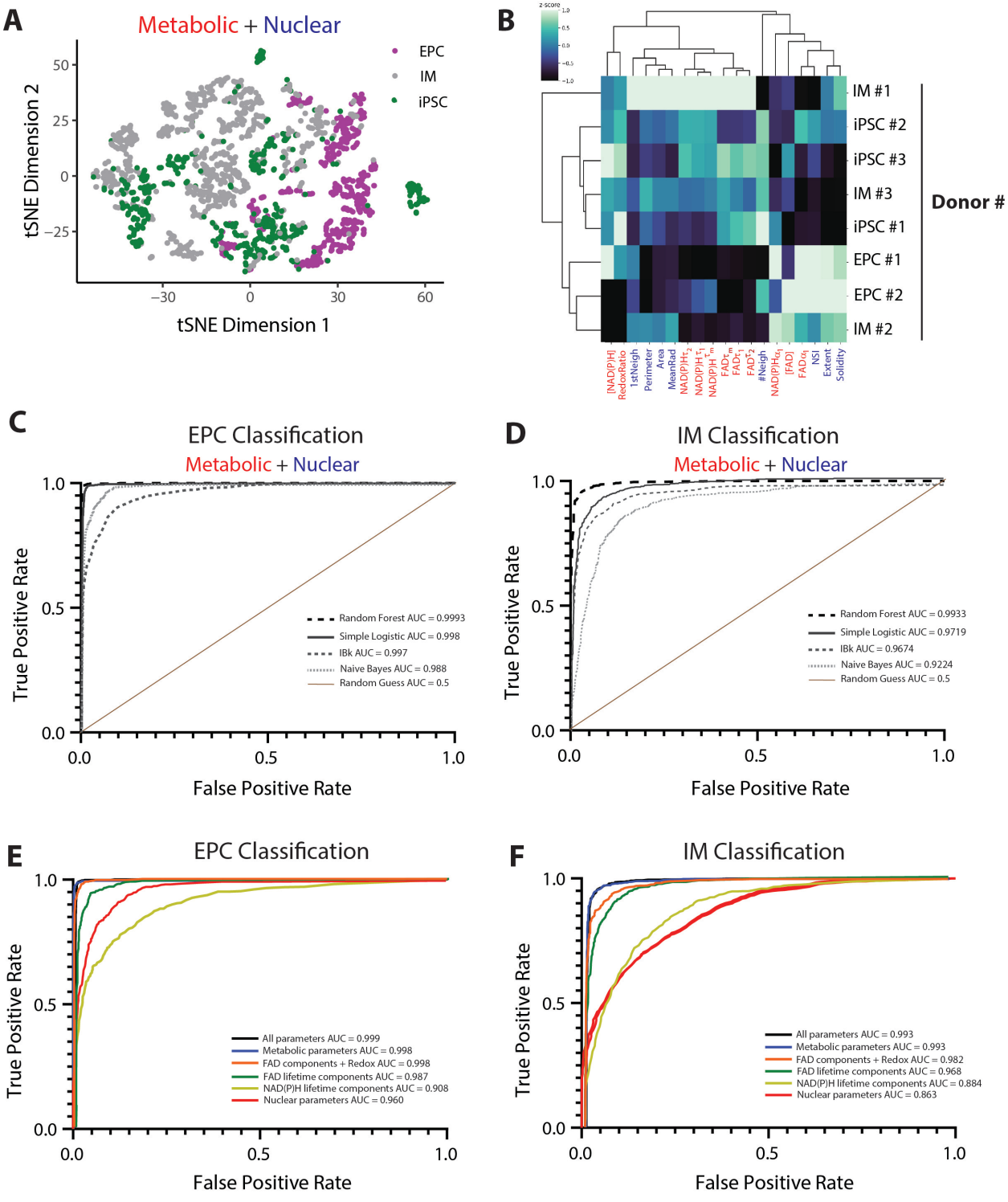


Figure B.4: Caption next page.

Figure B.4: **OMI enables accurate classification of EPCs and IMs.** A) t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction was performed on all 11 metabolic and 8 nuclear parameters for each cell, projected onto 2D space and shows poor separation of different cell types (EPCs, IMs, and iPSCs). Each color corresponds to a different cell type. Data are from three different donors. Each dot represents a single cell, and n =561, 990, and 586 cells for EPCs, IMs, and iPSCs, respectively. B) Heat map of z-scores of metabolic and nuclear parameters; each row is the mean data aggregating all cells from a single donor and cell type (EPCs, IMs, iPSCs); n = 3 biologically independent donors. ROC curves for C) EPCs and, D) iPSCs for different classifiers computed using all 11 metabolic and 8 nuclear parameters. AUC is provided for each classifier as indicated in the legend. ROC curves for E) EPCs and, F) iPSCs for different classifiers computed using different parameter combinations. AUC is provided for each parameter combination as indicated in the legend.

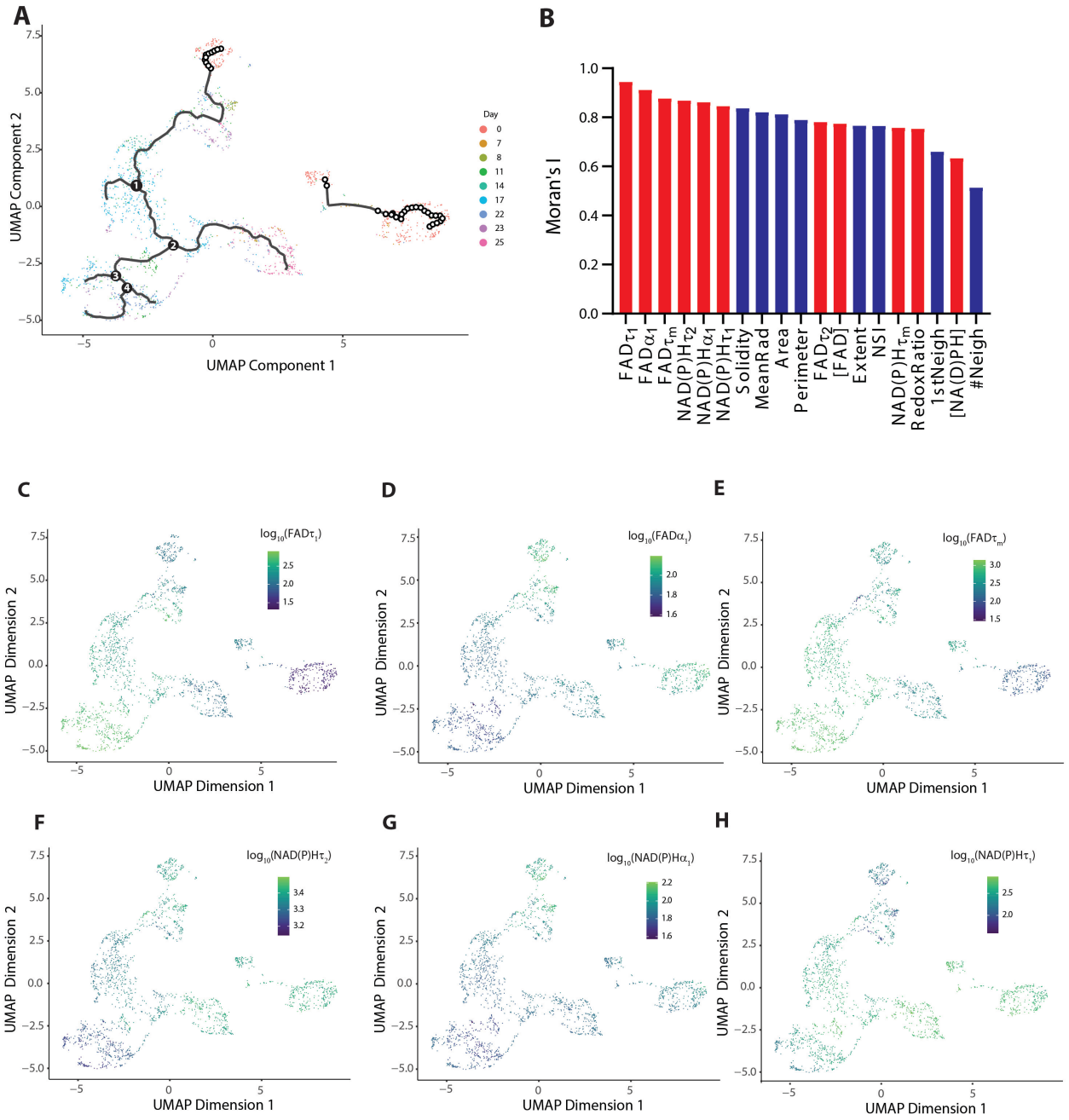


Figure B.5: **Metabolic and nuclear parameter changes along the reprogramming trajectory.** A) Trajectory of reprogramming EPCs constructed from the metabolic and nuclear parameters based on UMAP dimension reduction using Monocle, colored by day of EPC reprogramming. B) Metabolic and nuclear parameters ranked by their Moran's I (effect size) on the construction of single-cell reprogramming trajectory. UMAP plots based on Figure 5A highlighting the change of expression of top six metabolic parameters C) FADτ_1, D) FADα_1, E) FADτ_m, F) NAD(P)Hτ_2, G) NAD(P)Hα_1 and, H) NAD(P)Hτ_2 during reprogramming of EPCs.

#	Measurement	Description
1	[NAD(P)H]	NAD(P)H autofluorescence
2	[FAD]	FAD autofluorescence
3	Redox Ratio	[NAD(P)H]/([NAD(P)H + [FAD])
4	NAD(P)Hτ1	NAD(P)H short lifetime value
5	NAD(P)Hτ2	NAD(P)H long lifetime value
6	NAD(P)Hα_1	Fraction of free NAD(P)H
7	FADτ_1	FAD short lifetime value
8	FADτ_2	FAD longtime value
9	FADα_1	Fraction of bound FAD
10	NAD(P)Hτ_m	Mean NAD(P)H fluorescence lifetime $\tau_m = \alpha_1\tau_1 + \alpha_2\tau_2$
11	FADτ_m	Mean FAD fluorescence lifetime $\tau_m = \alpha_1\tau_1 + \alpha_2\tau_2$
12	Area	The total number of pixels within an object.
13	Perimeter (Perim)	The total number of pixels around the boundary of an object.
14	Nuclear Shape Index (NSI)	NSI = $4\pi A/P^2$, where A is the area and P is the perimeter of an ellipse around the object's major and minor axes. NSI = 1 for a perfect circle and approaches 0 with elongation
15	Mean Radius (MeanRad)	The mean distance of any pixel in the object to the closest pixel outside of the object.
16	Solidity (Solid)	The proportion of pixels in the convex hull that are also in the object: Object Area/Convex hull Area
17	Extent	The proportion of pixels in the bounding box that are also in the object: Object Area/Bounding Box Area
18	Distance to closest neighbor (1st Neigh)	The distance to the closest object in pixels.
19	Number of neighbors (#Neigh)	The total number of neighbor objects.

Table B.1: **Metabolic and nuclear parameters used to create reprogramming models.** Description of the 11 metabolic (red) and 8 nuclear (blue) parameters that were measured. These parameters were obtained after processing NAD(P)H and FAD images using an image analysis pipeline described in Figure 3.1B.

```r
##Install necessary packages ##
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install(version = "3.12")
BiocManager::install(c('BiocGenerics', 'DelayedArray', 'DelayedMatrixStats',
                       'limma', 'S4Vectors', 'SingleCellExperiment',
                       'SummarizedExperiment', 'batchelor', 'Matrix.utils'))
install.packages("devtools")
devtools::install_github('cole-trapnell-lab/leidenbase')
devtools::install_github('cole-trapnell-lab/monocle3')
install.packages("reshape2", "stringi", "robustbase")
library(monocle3)
library(ggplot2)
library(dplyr)
##Input raw data files##
#Need 3 matrices in .csv format. (1) expr_matrix with rows as metabolic/nuclear parameters and cells as columns.
#(2) Phenodata with cells as rows and characteristics such as reprogramming day,uFeature number, donor number,and cell type.
#(3) Featuredata with metabolic/nuclear parameters as rows and a column called gene_short_name for sure (a few functions are dependent on this).
expr_matrix <- read.csv("ControlData.csv")
expr_matrix2 <- as.matrix(expr_matrix[,-1])
rownames(expr_matrix2) <- as.matrix(expr_matrix[,1]) #setting first column as row names
sample_sheet <- read.csv("Phenodata.csv")
sample_sheet2 <- as.matrix(sample_sheet[,-1])
rownames(sample_sheet2) <- as.matrix(sample_sheet[,1]) #setting first column as row names
gene_annotation <- read.csv("featuredata.csv")
gene_annotation2 <- as.matrix(gene_annotation[,-1])
rownames(gene_annotation2) <- as.matrix(gene_annotation[,1]) #setting first column as row names
##Create cds data frame##
cds <- new_cell_data_set(as.matrix(expr_matrix2), cell_metadata = sample_sheet2, gene_metadata = gene_annotation2)
##Normalize and preprocess the data using PCA##
cds <- preprocess_cds(cds, num_dim = 19)
plot_pc_variance_explained(cds)
##Reduce dimension using UMAP##
cds <- reduce_dimension(cds, preprocess_method = c("PCA"), reduction_method = c("UMAP"))
plot_cells(cds, reduction_method="UMAP", color_cells_by="CellType") + scale_color_manual(values = c("#C000C0", "#008000", "#A0A0A4"
##Check and remove batch effects##
plot_cells(cds, reduction_method="UMAP", color_cells_by="DonorNumber")
cds = align_cds(cds, num_dim = 19, alignment_group = "DonorNumber")
##Clustercells##
cds = cluster_cells(cds, reduction_method = c("UMAP"), k = 100, cluster_method = c("louvain") )
plot_cells(cds)
##Build Trajectories##
cds <- learn_graph(
  cds,
  use_partition = TRUE,
  close_loop = FALSE,
```

```
  learn_graph_control = NULL,
  verbose = FALSE
)
##Pseudotime ordering##
plot_cells(cds,
           color_cells_by = "Dayday",
           label_cell_groups=FALSE,
           label_leaves=TRUE,
           label_branch_points=TRUE,
           graph_label_size=1.5)
cds <- order_cells(cds)
plot_cells(cds,
           color_cells_by = "pseudotime",
           label_cell_groups=FALSE,
           label_leaves=FALSE,
           label_branch_points=FALSE,
           graph_label_size=1.5)
##Plot Differential Expression of metabolic and nuclear parameters##
Parameters <- c("FADA1", "FADtm", "FADT2")
Parameters_cds <- cds[rowData(cds)$gene_short_name %in% Parameters]
plot_genes_in_pseudotime(Parameters_cds,
                         color_cells_by="CellType",
                         min_expr=0.5) + scale_color_manual(values = c("#C000C0",
                         ↪  "#008000", "#A0A0A4"))
##Heat map of differential expression by cell type##
pr_graph_test_res <- graph_test(cds, neighbor_graph="knn", cores=8)
pr_graph_test_res
pr_deg_ids <- row.names(cds)
gene_module_df <- find_gene_modules(cds[pr_deg_ids,], resolution=1e-2) #A dataframe in
↪  which the first column contains gene ids or short gene names and the second contains
↪  groups. If NULL, genes are not grouped.
cell_group_df <- tibble::tibble(cell=row.names(colData(cds)),
                                cell_group=colData(cds)$CellType)
```

Listing 1: R code used to construct single-cell reprogramming trajectories.

C SUPPLEMENTAL FIGURES AND TABLES FOR CHAPTER 4

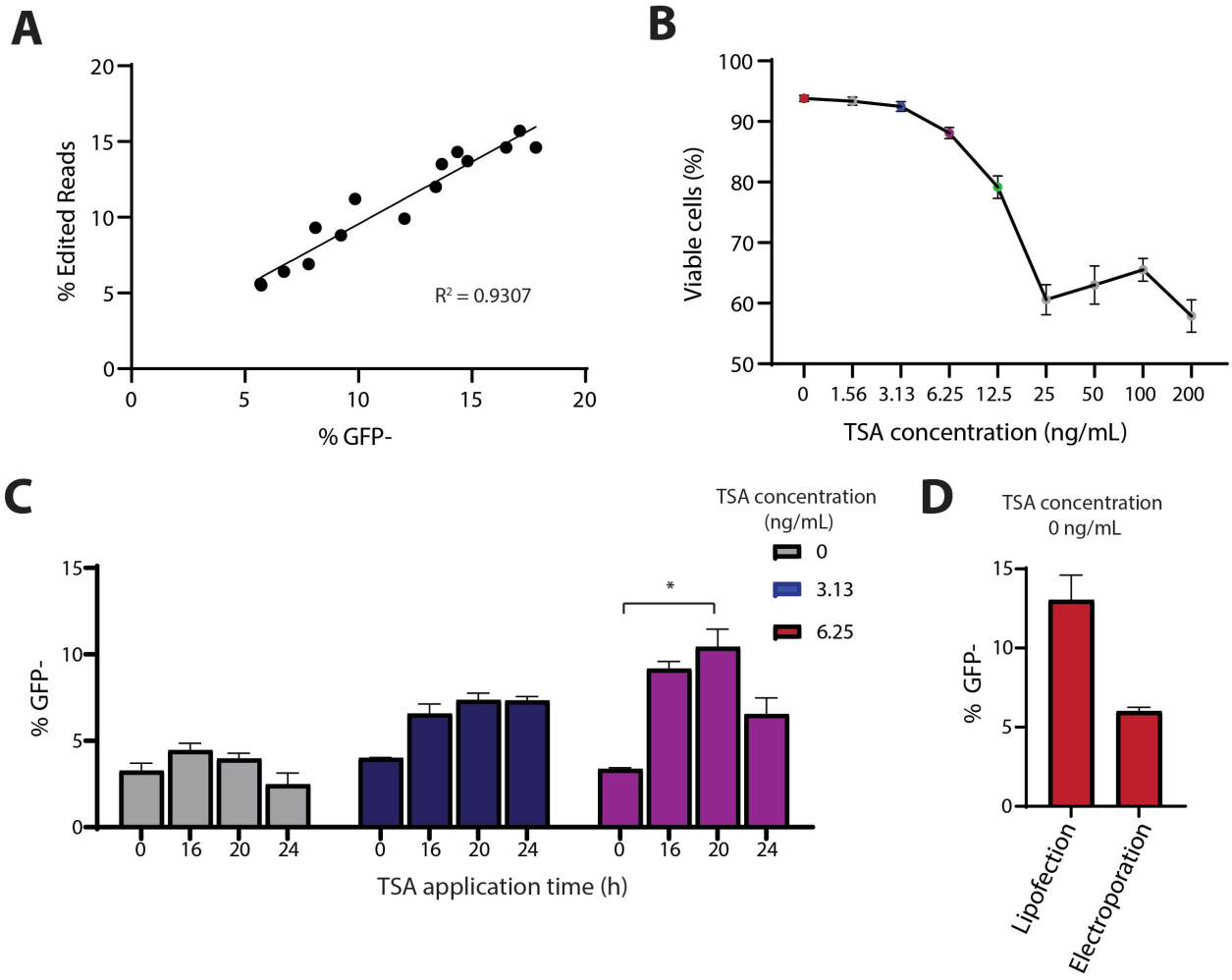
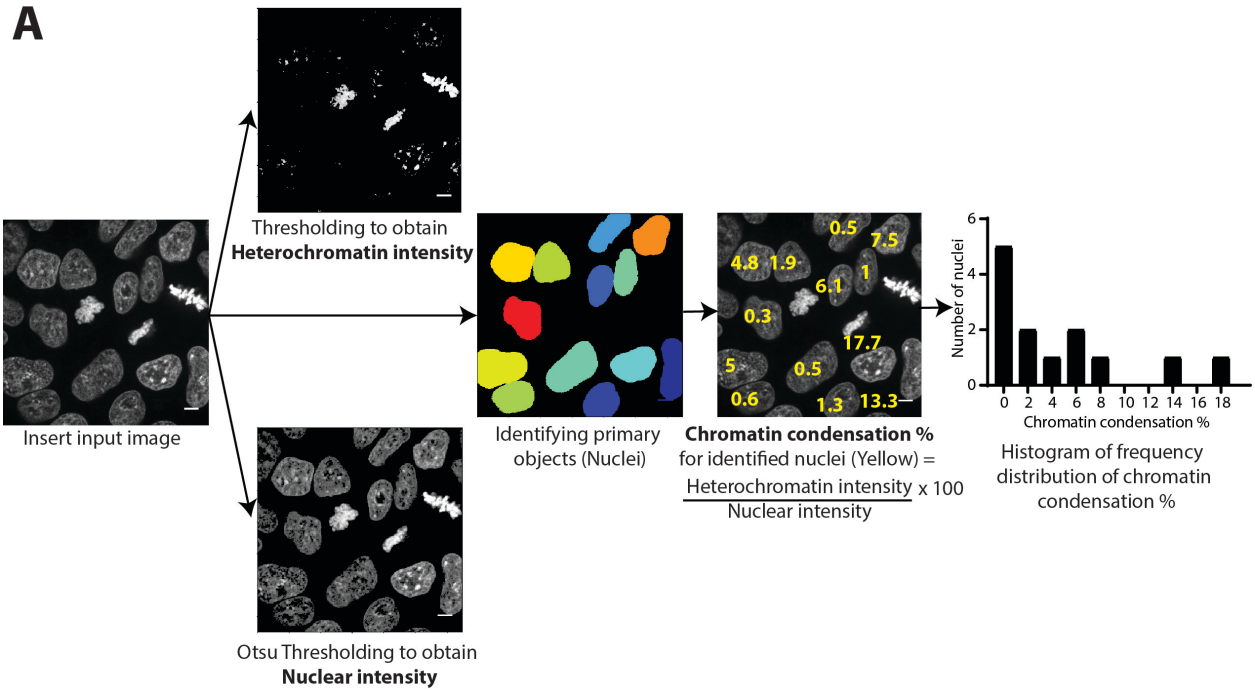


Figure C.1: Caption next page

Figure C.1: **Optimization of Trichostatin A treatment for iPSC gene editing.** A) Plot of editing indel % obtained from NGS sequencing of genomic DNA and from flow cytometry analysis ($R^2 = 0.9307$). B) Plot showing % viable cells as a function of TSA concentration. The viability (% GhostDye$^-$ cells/total cells) of iPSCs drops significantly for TSA concentrations geq 25 ng/mL. C) Bar graph showing gene editing efficiency (%GFP$^-$ cells) as a function of TSA application time. TSA application duration of 20 hours prior to Cas9 RNP delivery yielded the highest gene editing efficiencies, and was thus used as the TSA application time for the entire study. D) Bar graph showing gene editing efficiency (%GFP$^-$ cells) for different methods of Cas9 RNP delivery. Since lipofection yielded higher gene editing efficiencies, it was used as the method of Cas9 RNP delivery for the entire study. Data represented in bar graphs are represented as mean $\pm$ SEM, n = 3 technical replicates per condition, *p*-values generated by non-parametric Kruskal-Wallis test for multiple unmatched comparisons to 0 ng/mL TSA ; ns = p geq0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001.



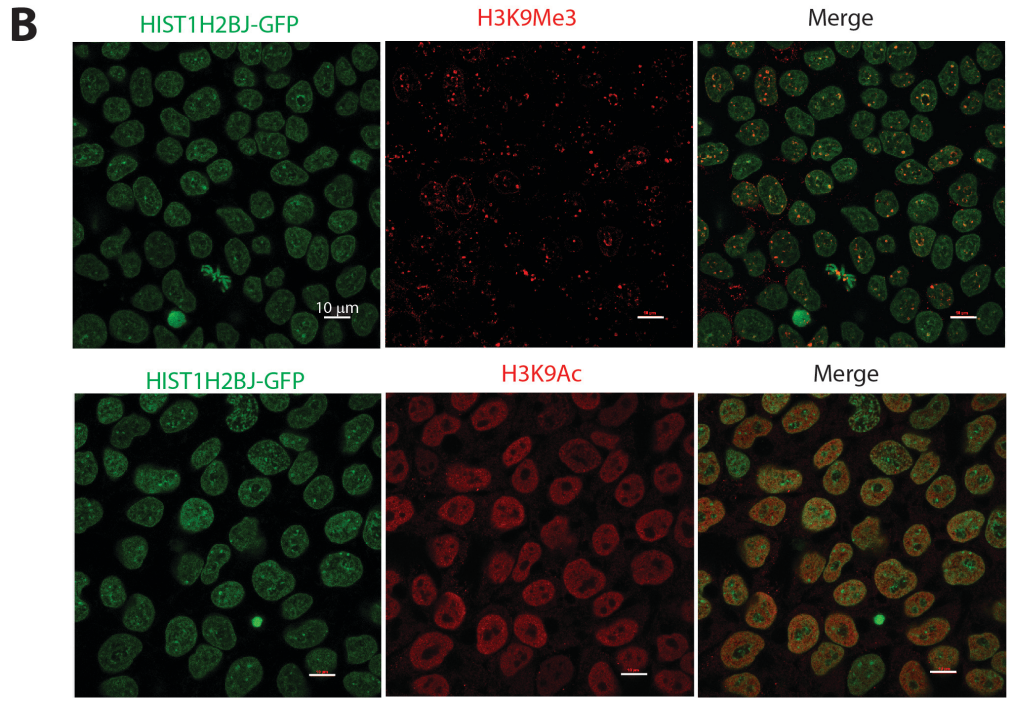


Figure C.2: Caption next page

Figure C.2: **CellProfiler pipeline to determine global chromatin condensation percentage of cell nuclei.** A) The CellProfiler pipeline calculates the global chromatin percentage of the identified nuclei as the total heterochromatin intensity to the total nuclear intensity. Chromatin condensation percentage values for the *HIST1H2BJ*-mEGFP iPSC nuclei are indicated in yellow on each nucleus and the histogram depicts the frequency distribution of the chromatin condensation percentage values. B) Representative images of iPSC nuclei stained with H3K9Me3 (heterochromatin mark) and (euchromatin mark); Scale bar: 10 μm.

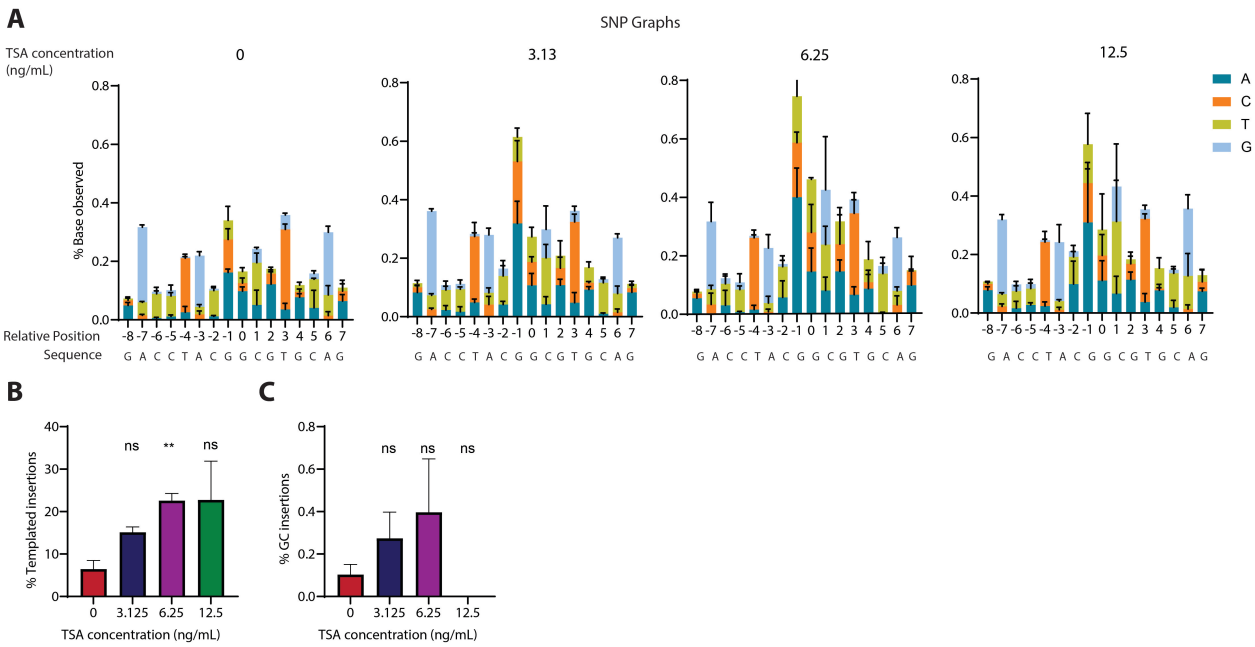


Figure C.3: **Additional on-target profiling of edited iPSCs.** A) Depiction of SNP profiles at different TSA concentrations. Bar graphs showing increased B) % templated insertions, and C) % GC insertions with the increase in TSA concentration. Data represented in bar graphs are represented as mean ± SEM, n = 3 technical replicates per condition, *p*-values generated by non-parametric Kruskal-Wallis test for multiple unmatched comparisons to 0 ng/mL TSA; ns = p geq 0.05, * for p <0.05, ** for p <0.01, *** for p <0.001, **** for p <0.0001.

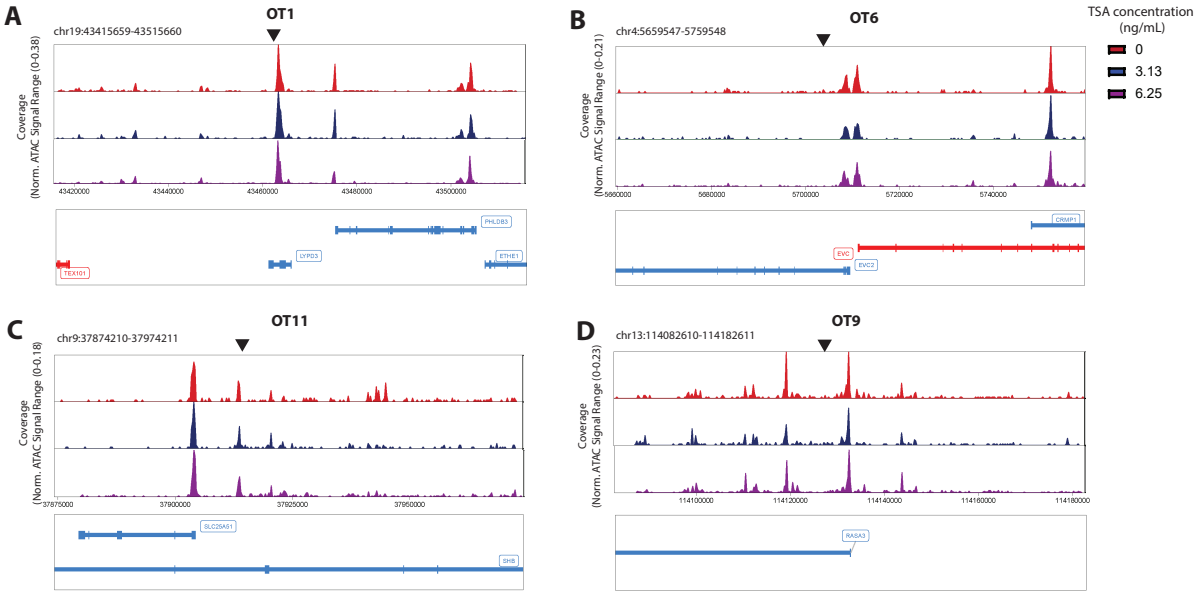


Figure C.4: **Chromatin accessibility analysis at off-target loci.** scATAC-seq plots for A) OT1, B) OT6, C) OT11, and D) OT9. Chromatin accessibility around the mEGFP gRNA off-target cut-sites (black arrow) does not increase upon TSA treatment.

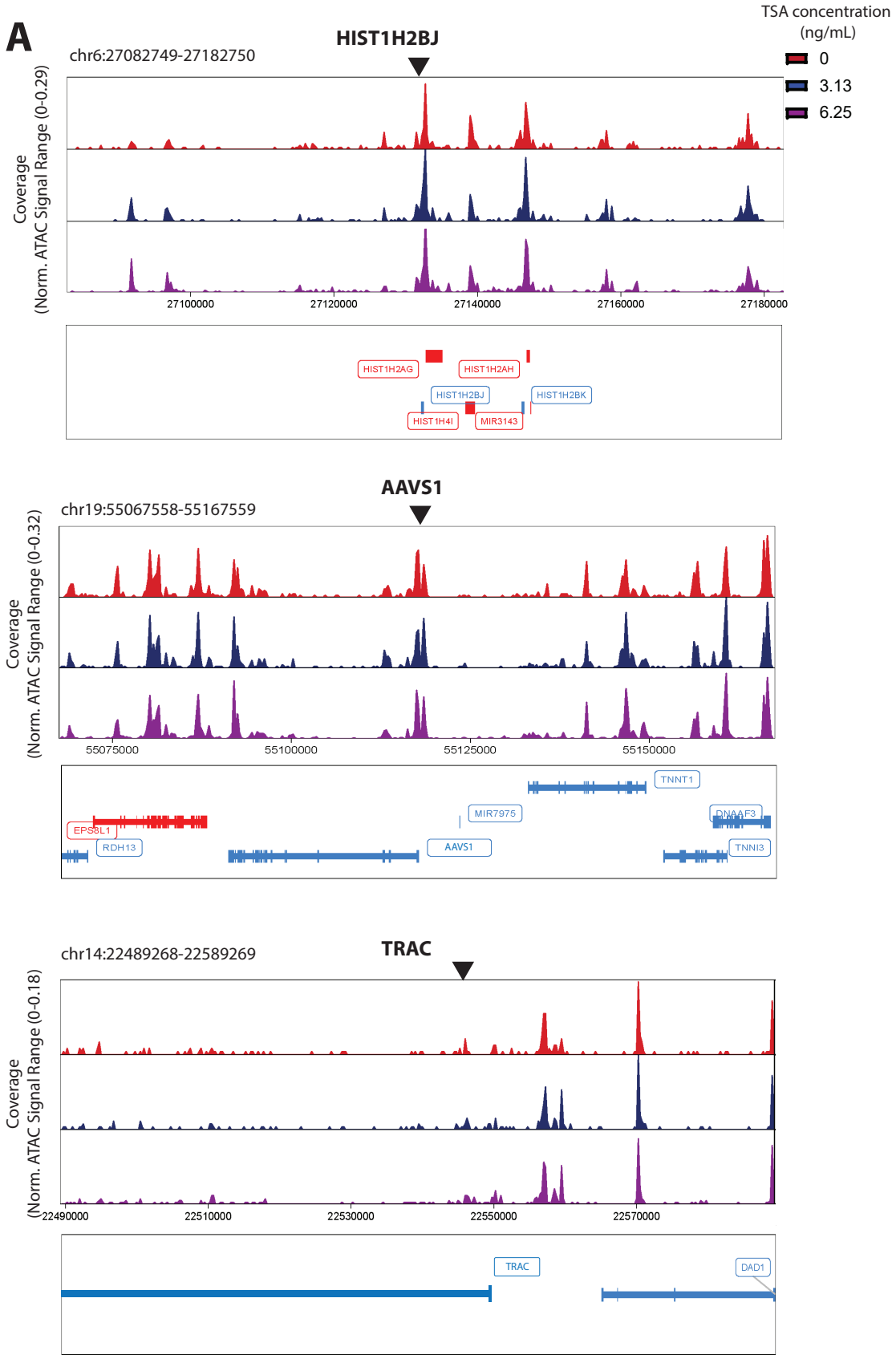

A
HIST1H2BJ
chr6:27082749-27182750
TSA concentration (ng/mL)
0
3.13
6.25
Coverage (Norm. ATAC Signal Range (0-0.29)
27100000
27120000
27140000
27160000
27180000
HIST1H2AG
HIST1H2AH
HIST1H2BJ
HIST1H2BK
HIST1H4I
MIR3143
AAVS1
chr19:55067558-55167559
Coverage (Norm. ATAC Signal Range (0-0.32)
55075000
55100000
55125000
55150000
TNNT1
MIR7975
DNAAF3
EPS8L1
RDH13
AAVS1
TNNI3
TRAC
chr14:22489268-22589269
Coverage (Norm. ATAC Signal Range (0-0.18)
22490000
22510000
22530000
22550000
22570000
TRAC
DAD1

Figure C.5

Figure C.5: **Chromatin accessibility analysis at different open and chromatin loci.** scATAC-seq plots for A) *HIST1H2BJ* (open), B) *AAVS1* (open), and C) *TRAC* (closed). Chromatin accessibility around various on-target cut-sites (black arrow) increases upon TSA treatment.

Target Locus	**gRNA (5′ to 3′)**	**NGS Forward Primer (5′ to 3′)**	**NGS Reverse Primer (5′ to 3′)**
GFP	GCTGAAGCACTGCACGCCGT	CGGCAAGCTGACCCTGAAGTTC	CGTCCAGGAGCGCACCATCTTC
HIST1H2BJ_Site_1	TCTGCTCCCGCCCCGAAAAA	CCTGCTCACGCTGTTTTTCC	TCGCTTCTTGCCGTCTTTCTT
HIST1H2BJ_Site_2	GGTGACGGCCTTAGTACCCT	TCACCTCCAGGGAGATCCAG	TGAACAGCTCCTCGCCCTTG
AAVS1_Site_8	GCTGTCCTGAAGTGGACATA	TTGCCTGGACACCCCGTTCTCCT	GCCACATTAACCGGCCCTGGGAA
AAVS1_Site_10	GGGAACCCAGCGAGTGAAGA	CTTCCTTCTCGGCGCTGCACCAC	AGCCAGGGAGACGGGGTACTTTGG
TRAC_Site_2	GCTGGTACACGGCAGGGTCA	GCCTGGGTTGGGGCAAAGAGGGA	GTTGCTCCAGGCCACAGCACTGT
TRAC_Site_3	GAGAATCAAAATCGGTGAAT	GCCTGGGTTGGGGCAAAGAGGGA	GTTGCTCCAGGCCACAGCACTGT

Table C.1: **Guide RNA and primer sequences.**

Name	#Chromosome	Start	End	Genomic Coordinate	Strand	Read Count	Site_Sequence	Site Substitution#
OT1	chr19	43463243	43463266	chr19:43463243-43463266	+	2452	GCTGTAGCACTCCACGCCGTTGG	2
OT2	chr9	98626054	98626076	chr9:98626054-98626076	+	1030		
OT3	chr3	73404993	73405016	chr3:73404993-73405016	-	548	AGCAAAGCACTGCACACAGTGGG	6
OT4	chr5	6422338	6422361	chr5:6422338-6422361	-	524	CAGGAAGCACTGCACACTGTGGG	5
OT5	chr8	51474994	51475018	chr8:51474994-51475018	+	338		
OT6	chr4	5703158	5703181	chr4:5703158-5703181	-	322	AAAGAAGCACTGCATGCTGTAGG	5
OT7	chr1	89051013	89051036	chr1:89051013-89051036	+	300	CCTGCACCCTTGCACGCCATTGG	6
OT8	chr5	151028994	151029017	chr5:151028994-151029017	-	274	GAAGAAGCACCACACACAGTAGG	6
OT9	chr13	114129061	114129084	chr13:114129061-114129084	+	254	GAGGAGGCACTGCACGCCTTGGG	4
OT10	chr16	489418	489441	chr16:489418-489441	+	254	ATAGAAGCACAACACGCAGTGGG	6
OT11	chr9	37922745	37922768	chr9:37922745-37922768	+	252	GCTCAAGCACTGCACCCCGTGGG	2
OT12	chr13	25135132	25135155	chr13:25135132-25135155	-	200	AGTGAAGCAAAACACACCGTAGG	6
OT13	chr16	489418	489441	chr16:489418-489441	+	174	ATAGAAGCACAACACGCAGTGGG	6
OT14	chr1	84346679	84346702	chr1:84346679-84346702	+	168	ATTGTAGCACAGCATGCTGTTGG	6
OT15	chr15	80849498	80849521	chr15:80849498-80849521	+	134	GCTCACCCACTTCACGCCGTGGG	4
OT16	chr4	94720214	94720237	chr4:94720214-94720237	+	132	GGGCAAGCCTTGCACACCGTTGG	6
OT17	chr13	100548625	100548648	chr13:100548625-100548648	+	130	GGAGAGGGACTGCACACAGTGGG	6
OT18	chr12	132318694	132318716	chr12:132318694-132318716	+	120		
OT19	chr8	135379753	135379776	chr8:135379753-135379776	+	118	GCACACCCACTGCACACCGTGGA	6
OT20	chr7	128351804	128351827	chr7:128351804-128351827	-	112	TATGAGGTGCTGCACGCCCTGGG	6
OT21	chr5	38456159	38456182	chr5:38456159-38456182	-	108	TAGGAAGCACTGCACATCGTAGG	5

Table C.2: **Modifications of off-target sites of the GFP locus in iPSCs detected by CHANGE-seq.** Target sequence for all analysis was GCTGAAGCACTGCACGCCGTNGG. Only top 50 loci with the highest Nuclease Read Counts are included.

Name	#Chromosome	Start	End	Genomic Coordinate	Strand	Read Count	Site_Sequence	Site Substitution#
OT22	chr8	51474994	51475018	chr8:51474994-51475018	+	108		
OT23	chr15	22272189	22272212	chr15:22272189-22272212	-	100	AATGACACACTGCACGACGTGTG	6
OT24	chr2	32648726	32648749	chr2:32648726-32648749	-	100	TGTGAGGCACTGCACCCAGTAGG	5
OT25	chr14	77261231	77261254	chr14:77261231-77261254	+	98	CCTCAGGCACTGCATGCCCTGGG	5
OT26	chr11	365544	365567	chr11:365544-365567	+	96	CCTGCAGTGTTGCACGCTGTTGG	6
OT27	chr17	73490267	73490290	chr17:73490267-73490290	-	94	CCTCCAGCCCTGCAAGCAGTGGG	6
OT28	chr5	142181271	142181294	chr5:142181271-142181294	-	94	GCACCAGCCCTGCACGCCCTGCG	6
OT29	chr3	8397815	8397838	chr3:8397815-8397838	+	90	ATTGAACTCCTGCACGCCGTCGG	5
OT30	chr10	77250682	77250705	chr10:77250682-77250705	-	88	GCTGAAGCATTGCACGCTGGTGC	4
OT31	chr22	20063932	20063955	chr22:20063932-20063955	-	88	ACATACGCAGTGCACCCCGTGGG	6
OT32	chr4	88191200	88191223	chr4:88191200-88191223	-	84	CTTGGAGCACTGCACGCAGTGAG	5
OT33	chr16	67690285	67690308	chr16:67690285-67690308	-	82	TCAGAAGTCTTGCACGCTGTTGG	6
OT34	chr2	241775589	241775612	chr2:241775589-241775612	+	76	GAGGATGCAGGGCACGCCGGTGG	6
OT35	chr6	131085171	131085193	chr6:131085171-131085193	-	76		
OT36	chr16	2526520	2526543	chr16:2526520-2526543	-	74	TCTGGGGGAGTGCACGCAGTGGG	6
OT37	chr2	235007697	235007720	chr2:235007697-235007720	-	72	TCTCAAGCACTGCACGCCTGGTC	6
OT38	chr12	32661527	32661550	chr12:32661527-32661550	+	70	GCTCAATCACTGCACGCAGTGAG	4
OT39	chr22	20569442	20569465	chr22:20569442-20569465	-	70	GCAGGCGCACTGCCAGCTGTGGG	6
OT40	chr4	122744364	122744387	chr4:122744364-122744387	+	70	GATGAAAGCCTGCACACAGTGGG	6
OT41	chr5	79831437	79831460	chr5:79831437-79831460	-	70	GCCTGGGACCTGCACGCCGTTGG	6
OT42	chr9	136432758	136432781	chr9:136432758-136432781	+	70	GGCCCAGCACGGCACGCCCTCGG	6
OT43	chr12	76923253	76923276	chr12:76923253-76923276	-	68	GATGTACCATTGCATGTCGTTGG	6

Table C.2: **Modifications of off-target sites of the GFP locus in iPSCs detected by CHANGE-seq.** Target sequence for all analysis was GCTGAAGCACTGCACGCCGTNGG. Only top 50 loci with the highest Nuclease Read Counts are included.

Name	#Chromosome	Start	End	Genomic Coordinate	Strand	Read Count	Site_Sequence	Site Substitution#
OT44	chr1	205457245	205457268	chr1:205457245-205457268	+	68	GCTGTGGCTGAGGACGCCGTCGG	6
OT45	chr1	1161377	1161400	chr1:1161377-1161400	-	66	GGTGAGCCACCGCACGCCATGGG	5
OT46	chr1	245608179	245608202	chr1:245608179-245608202	+	64	GAGAAAGCAGTGCACACCGTCGG	5
OT47	chr22	49823489	49823512	chr22:49823489-49823512	+	62	GTTGGGGCACAGCACACAGTCGG	6
OT48	chr10	131387191	131387214	chr10:131387191-131387214	-	60	GGAAAAACACAGCACACCGTGGG	6
OT49	chr16	4694177	4694201	chr16:4694177-4694201	+	60		
OT50	chr1	1355229	1355252	chr1:1355229-1355252	+	58	GGGGAAGCACTCACGGCCGTCGG	6

Table C.2: **Modifications of off-target sites of the GFP locus in iPSCs detected by CHANGE-seq.** Target sequence for all analysis was GCTGAAGCACTGCACGCCGTNGG.Only top 50 loci with the highest Nuclease Read Counts are included.

Target Name	Assay ID	rhAmp-seq Forward Primer (5′ to 3′)	rhAmp-seq Reverse Primer (5′ to 3′)
On-target	RH.3219B4F53C43413Z0Z	/rhSeq-f/TGAAGTTCATCTGCACCACrCGGCA/GT4/	/rhSeq-r/CGTCCTTGAAGAAGATGGTGrCGCTC/GT3/
OT1	RH.1987151E363D44DZ0Z	/rhSeq-f/CTTGTAGACATGATCGCTGGrCGTTG/GT1/	/rhSeq-r/ACACCTTTTCTTCCTCTCGGrCTCCC/GT3/
OT2	RH.B3BF69D7C4164C5Z0Z	/rhSeq-f/CTTAAGACAGGGAAGTCAGGrGCTCA/GT2/	/rhSeq-r/ACAATATATGGTCCTTTGCGArCAGCC/GT1/
OT3	RH.F4D7BA568E2D41FZ0Z	/rhSeq-f/GCGGCATATGGCAAGTTATTATGrGAGAA/GT2/	/rhSeq-r/CTTCATTTAAGTCATGCAGCAGArCAGTA/GT2/
OT4	RH.24F70E12F5C8482Z0Z	/rhSeq-f/AACAGCTTGATCCTGGArCTTCC/GT2/	/rhSeq-r/CCCAGCAGAACCATAACArCAGAA/GT3/
OT5	RH.6542C4B62E8248DZ0Z	/rhSeq-f/CACATTGAAATTTACGTACGGCrAATTG/GT4/	/rhSeq-r/TTGTGACTGTGATCTGATGTGrGCTGG/GT2/
OT6	RH.47BE9BC9747547BZ0Z	/rhSeq-f/TCATCAATAGGTTCATGGCAArCTGTG/GT4/	/rhSeq-r/GGAACATCAATAGGTTCTTGGrATACT/GT1/
OT7	RH.A2EC2BCC4A6B41EZ0Z	/rhSeq-f/CCACAACATCAGGAAGTGTrGACAG/GT3/	/rhSeq-r/TGCCAGAACAGAATCAAGCTrCTCTC/GT3/
OT8	RH.C3366BFE4D08452Z0Z	/rhSeq-f/CTCACCCCTCACTGGTrCCACT/GT2/	/rhSeq-r/CCTTTACCCTGCCCCArCCTTC/GT1/
OT9	RH.092E25CF3C2E47FZ0Z	/rhSeq-f/CACATGTATCAAGATAAGGCGCrCACAT/GT1/	/rhSeq-r/TATCATGCATGACTGAGCCArCACTT/GT3/
OT10	RH.32D6E4E5DF4A4FCZ0Z	/rhSeq-f/GAGGCCTCATGATCATAAGTGArGTCTT/GT3/	/rhSeq-r/TAGCCCTCAGAAGTACGTACrCCAGA/GT3/
OT11	RH.2129E4AB36B24EBZ0Z	/rhSeq-f/GGCTCCATGAACCTCTAGTrCTGTG/GT2/	/rhSeq-r/TGGATGCAGCAGGAGAAGrGGACA/GT2/
OT12	RH.0292A0CF06D3467Z0Z	/rhSeq-f/CTCCTGTCCTGTCACCTTTrCACTA/GT2/	/rhSeq-r/TGCCAGCAATAAGACCTCArGAATG/GT1/

Table C.3: **rhAmpSeq primer sequences for on-target and top 12 off-target sites for the GFP locus.**

Reference	CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
Sequence	CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC———-

Read Count	Insertion Start	Insertion Stop	Inserted Bases	Deletion Start	Deletion Stop	Deleted Bases	Snp Position	Snp base	Allele Frequency
47116	0	0	0	0	0	0	0	0	73.04919456
5885	0	0	0	81	84	CGG	0	0	9.124172468
1048	0	0	0	82	83	G	0	0	1.624831393
478	0	0	0	81	83	CG	0	0	0.741096761
472	83	83	T	0	0	0	0	0	0.731794291
429	83	83	G	0	0	0	0	0	0.665126591
251	0	0	0	82	92	GGCGTGCAGT	0	0	0.38915332
230	0	0	0	82	90	GGCGTGCA	0	0	0.356594676
217	0	0	0	81	100	CGGCGTGCAGTGCTTCAGC	0	0	0.336439325
201	83;85	83;85	T;A	0	0	0	0	0	0.311632738
190	0	0	0	80	83	ACG	0	0	0.294578211
179	0	0	0	83	85	GC	0	0	0.277523683
152	83	83	TA	0	0	0	0	0	0.235662568
143	0	0	0	83	96	GCGTGCAGTGCTT	0	0	0.221708864
142	0	0	0	77	83	CCTACG	0	0	0.220158452
134	0	0	0	0	0	0	86	C	0.207755159
128	0	0	0	0	0	0	76	G	0.198452689
126	0	0	0	75	92	GACCTACGGCGTGCAGT	0	0	0.195351866
126	83	83	A	0	0	0	0	0	0.195351866

Table C.4: **Representative indel profile for iPSCs edited with mEGFP targeting gRNA. TSA treatment: 6.25 ng/mL.**

Reference	CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
Sequence	CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC———–

Read Count	Insertion Start	Insertion Stop	Inserted Bases	Deletion Start	Deletion Stop	Deleted Bases	Snp Position	Snp base	Allele Frequency
124	0	0	0	83	90	GCGTGCA	0	0	0.192251043
124	0	0	0	83	87	GCGT	0	0	0.192251043
122	0	0	0	78	102	CTACGGCGTGCAGTGCTTCAGCCG	0	0	0.189150219
116	0	0	0	81	93	CGGCGTGCAGTG	0	0	0.17984775
108	0	0	0	78	93	CTACGGCGTGCAGTG	0	0	0.167444457
103	0	0	0	82	87	GGCGT	0	0	0.159692398
102	0	0	0	81	96	CGGCGTGCAGTGCTT	0	0	0.158141987
101	0	0	0	82	98	GGCGTGCAGTGCTTCA	0	0	0.156591575
101	0	0	0	0	0	0	89	G	0.156591575
101	0	0	0	85	86	G	0	0	0.156591575
99	0	0	0	83	92	GCGTGCAGT	0	0	0.153490752
97	0	0	0	83	97	GCGTGCAGTGCTTC	0	0	0.150389929
91	0	0	0	81	99	CGGCGTGCAGTGCTTCAG	0	0	0.141087459
87	83	83	TGCAGTGCTTCA	0	0	0	0	0	0.134885812
85	0	0	0	65	94	TGACCACCCTGACCTACGGCGTGCAGTGC	0	0	0.131784989

Table C.4: **Representative indel profile for iPSCs edited with mEGFP targeting gRNA. TSA treatment: 6.25 ng/mL.**

Reference	CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
Sequence	CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC———–

Read Count	Insertion Start	Insertion Stop	Inserted Bases	Deletion Start	Deletion Stop	Deleted Bases	Snp Position	Snp base	Allele Frequency
84	0	0	0	0	0	0	79	C	0.130234577
82	0	0	0	83	103	GCGTGCAGTGCTTCAGCCGC	0	0	0.127133754
82	0	0	0	83	89	GCGTGC	0	0	0.127133754
77	0	0	0	75	82	GACCTAC	0	0	0.119381696
76	0	0	0	78	88	CTACGGCGTG	0	0	0.117831284
73	0	0	0	0	0	0	84	G	0.113180049
73	0	0	0	81	88	CGGCGTG	0	0	0.113180049
72	0	0	0	67	88	ACCACCCTGACCTACGGCGTG	0	0	0.111629638
71	0	0	0	80	94	ACGGCGTGCAGTGC	0	0	0.110079226
68	83	83	TCA	0	0	0	84	G	0.105427991
67	0	0	0	81	82	C	82	A	0.103877579
65	0	0	0	83	94	GCGTGCAGTGC	0	0	0.100776756
61	0	0	0	0	0	0	82	C	0.09457511
61	0	0	0	75	83	GACCTACG	0	0	0.09457511
60	0	0	0	0	0	0	80	G	0.093024698
59	0	0	0	66	92	GACCACCCTGACCTACGGCGTGCAGT	0	0	0.091474286

Table C.4: **Representative indel profile for iPSCs edited with mEGFP targeting gRNA. TSA treatment: 6.25 ng/mL.**

Sample	Read Count	Insertion Start	Insertion Stop	Inserted Bases	Deletion Start	Deletion Stop	Deleted Bases	Snp Position	Snp base	Allele Frequency
OT1	79164	0	0	0	0	0	0	0	0	97.32361294
OT1	238	0	0	0	0	0	0	103	C	0.29259537
OT1	193	0	0	0	0	0	0	94	C	0.237272716
OT1	176	0	0	0	0	0	0	104	C	0.216373047
OT1	133	0	0	0	0	0	0	97	G	0.163509177
OT1	105	0	0	0	0	0	0	99	A	0.129086193
OT2	132705	0	0	0	0	0	0	0	0	97.13936448
OT2	432	0	0	0	0	0	0	105	G	0.316221736
OT2	398	0	0	0	0	0	0	96	G	0.291333914
OT2	287	0	0	0	0	0	0	101	C	0.210082496
OT2	273	0	0	0	0	0	0	91	G	0.199834569
OT2	193	0	0	0	0	0	0	99	T	0.141274988
OT2	173	0	0	0	0	0	0	93	C	0.126635093
OT2	147	0	0	0	0	0	0	100	A	0.10760323
OT2	136	0	0	0	0	0	0	104	A	0.099551287
OT2	135	0	0	0	0	0	0	95	T	0.098819292
OT2	106	0	0	0	0	0	0	102	A	0.077591444
OT2	102	0	0	0	0	0	0	103	A	0.074663465
OT2	102	0	0	0	0	0	0	94	A	0.074663465
OT2	101	0	0	0	0	0	0	94	C	0.073931471
OT2	101	0	0	0	0	0	0	97	A	0.073931471
OT2	101	0	0	0	0	0	0	92	T	0.073931471
OT2	100	0	0	0	0	0	0	98	T	0.073199476
OT3	141608	0	0	0	0	0	0	0	0	96.63700388
OT3	588	0	0	0	0	0	0	85	C	0.401266583

Table C.5: **Representative off-target profile at top 12 off-target sites for iPSCs edited with mEGFP targeting gRNA. TSA treatment: 6.25 ng/mL.**)

Sample	Read Count	Insertion Start	Insertion Stop	Inserted Bases	Deletion Start	Deletion Stop	Deleted Bases	Snp Position	Snp base	Allele Frequency
OT3	544	0	0	0	0	0	0	86	C	0.371239832
OT3	433	0	0	0	0	0	0	96	C	0.295490528
OT3	375	0	0	0	0	0	0	94	C	0.255909811
OT3	361	0	0	0	0	0	0	99	G	0.246355844
OT3	319	0	0	0	0	0	0	92	C	0.217693946
OT3	263	0	0	0	0	0	0	90	G	0.17947808
OT3	178	0	0	0	0	0	0	87	T	0.121471857
OT3	140	0	0	0	0	0	0	88	T	0.095539663
OT3	129	0	0	0	0	0	0	98	T	0.088032975
OT3	126	0	0	0	0	0	0	93	A	0.085985696
OT3	114	0	0	0	0	0	0	97	A	0.077796582
OT3	113	0	0	0	0	0	0	89	T	0.077114156
OT3	113	0	0	0	0	0	0	95	A	0.077114156
OT4	10919	0	0	0	0	0	0	0	0	97.1527716
OT5	71249	0	0	0	0	0	0	0	0	97.21119343
OT5	220	0	0	0	0	0	0	66	G	0.300165091
OT5	174	0	0	0	0	0	0	56	G	0.237403299
OT5	171	0	0	0	0	0	0	53	C	0.233310139
OT5	134	0	0	0	0	0	0	63	C	0.182827828
OT5	119	0	0	0	0	0	0	62	G	0.162362026
OT5	107	0	0	0	0	0	0	65	C	0.145989385
OT6	50427	0	0	0	0	0	0	0	0	96.85021223
OT6	179	0	0	0	0	0	0	101	C	0.343787812
OT6	151	0	0	0	0	0	0	100	C	0.290010947
OT6	121	0	0	0	0	0	0	114	G	0.232392878

Table C.5: **Representative off-target profile at top 12 off-target sites for iPSCs edited with mEGFP targeting gRNA. TSA treatment: 6.25 ng/mL.**

Ensembl ID	Gene Name	Log2 fold change
Top 10 Upregulated genes		
ENSG00000104435.14	STMN2	7.277846866
ENSG00000147381.11	MAGEA4	6.754650943
ENSG00000285958.1	NA	6.238488914
ENSG00000100867.15	DHRS2	5.767017043
ENSG00000259033.2	NA	5.42751645
ENSG00000179520.11	SLC17A8	5.065863469
ENSG00000124092.12	CTCFL	4.799176432
ENSG00000129538.14	RNASE1	4.744349224
ENSG00000173210.20	ABLIM3	4.70542518
ENSG00000164542.13	KIAA0895	4.691769828
Top 10 Downregulated Genes		
ENSG00000259040.5	BLOC1S5-TXNDC5	-2.958457066
ENSG00000270066.3	NA	-2.791940422
ENSG00000172780.17	RAB43	-2.724658478
ENSG00000285130.2	NA	-2.225404414
ENSG00000118194.20	TNNT2	-2.176186417
ENSG00000174776.11	WDR49	-2.080537629
ENSG00000063438.19	AHRR	-2.066344405
ENSG00000225783.8	MIAT	-2.016967788
ENSG00000141736.14	ERBB2	-1.962932601
ENSG00000138161.14	CUZD1	-1.811777849
Pluripotency genes		
ENSG00000128567.17	TRA-1-60	-0.178778459
ENSG00000204531.20	OCT4	-0.005607251
ENSG00000111704.11	NANOG	0.177296752
ENSG00000181449.4	SOX2	0.081692895

Table C.6: **Top 10 upregulated and downregulated genes in iPSCs after 6.25 ng/mL TSA treatment.**

```
1 library(ArchR)
2 #Create a folder with the fragment files called "Fragment_Files"
3 inputFiles <- getInputFiles("Fragment_Files")
```

```
inputFiles
#Setting default genome to hg38
addArchRGenome("hg38")
# Setting default number of parallel threads to 2
addArchRThreads(threads = 2, force = TRUE)
#Creating Arrow Files
ArrowFiles <- createArrowFiles(
  inputFiles = inputFiles,
  sampleNames = names(inputFiles),
  filterTSS = 4,
  filterFrags = 1000,
  addTileMat = TRUE,
  addGeneScoreMat = TRUE
)
#Show Arrow Files
ArrowFiles
#Inferring doublet scores
doubScores <- addDoubletScores(
  input = ArrowFiles,
  k = 10, #Refers to how many cells near a "pseudo-doublet" to count.
  knnMethod = "UMAP", #Refers to the embedding to use for nearest neighbor search.
  LSIMethod = 1
)
## Creating an ArchR Project ##
proj1 <- ArchRProject(
  ArrowFiles = ArrowFiles,
  outputDirectory = "Outputs",
  copyArrows = FALSE
)
proj1
paste0("Memory Size = ", round(object.size(proj1) / 10^6, 3), " MB")
getAvailableMatrices(proj1)
# Plotting Sample Statistics from ArchR Project
p1 <- plotGroups(
  ArchRProj = proj1,
  groupBy = "Sample",
  colorBy = "cellColData",
  name = "TSSEnrichment",
  plotAs = "ridges"
)
p1
p2 <- plotGroups(
  ArchRProj = proj1,
  groupBy = "Sample",
  colorBy = "cellColData",
  name = "TSSEnrichment",
  plotAs = "violin",
  alpha = 0.4,
  addBoxPlot = TRUE
)
p2
p3 <- plotGroups(
  ArchRProj = proj1,
  groupBy = "Sample",
```

```
  colorBy = "cellColData",
  name = "log10(nFrags)",
  plotAs = "ridges"
)
p3
p4 <- plotGroups(
  ArchRProj = proj1,
  groupBy = "Sample",
  colorBy = "cellColData",
  name = "log10(nFrags)",
  plotAs = "violin",
  alpha = 0.4,
  addBoxPlot = TRUE
)
p4
plotPDF(p1,p2,p3,p4, name = "QC-Sample-Statistics.pdf", ArchRProj = proj1, addDOC =
↪  FALSE, width = 4, height = 4)
# Plotting Sample Fragment Size Dist. and TSS Enrichment Profiles
p1 <- plotFragmentSizes(ArchRProj = proj1)
p1
p2 <- plotTSSEnrichment(ArchRProj = proj1)
p2
plotPDF(p1,p2, name = "QC-Sample-FragSizes-TSSProfile.pdf", ArchRProj = proj1, addDOC =
↪  FALSE, width = 5, height = 5)
# Filtering Doublets
proj2 <- filterDoublets(proj1)
proj2
#Saving ArchR project
saveArchRProject(ArchRProj = proj2, outputDirectory = "Save-Proj2", load = TRUE)
## Dimensionality Reduction ##
# Iterative Latent Semantic Indexing (LSI)
proj2 <- addIterativeLSI(
  ArchRProj = proj2,
  useMatrix = "TileMatrix",
  name = "IterativeLSI",
  iterations = 2,
  clusterParams = list( #See Seurat::FindClusters
    resolution = c(0.2),
    sampleCells = 10000,
    n.start = 10
  ),
  varFeatures = 25000,
  dimsToUse = 1:30
)
# Batch Effect Correction with Harmony
proj2 <- addHarmony(
  ArchRProj = proj2,
  reducedDims = "IterativeLSI",
  name = "Harmony",
  groupBy = "Sample"
)
## Clustering ##
proj2 <- addClusters(
  input = proj2,
```

```
  reducedDims = "IterativeLSI",
  method = "Seurat",
  name = "Clusters",
  resolution = 0.8
)
head(proj2$Clusters)
table(proj2$Clusters)
cM <- confusionMatrix(paste0(proj2$Clusters), paste0(proj2$Sample))
cM
library(pheatmap)
cM <- cM / Matrix::rowSums(cM)
p <- pheatmap::pheatmap(
  mat = as.matrix(cM),
  color = paletteContinuous("whiteBlue"),
  border_color = "black"
)
p
plotPDF(p, name = "HeatMap.pdf", ArchRProj = proj2, addDOC = FALSE, width = 5, height = 5)

# Clustering using scran
#projHeme2 <- addClusters(
#  input = projHeme2,
#  reducedDims = "IterativeLSI",
#  method = "scran",
#  name = "ScranClusters",
#  k = 15
#)
## Single Cell Embeddings ##
# UMAP
proj2 <- addUMAP(
  ArchRProj = proj2,
  reducedDims = "IterativeLSI",
  name = "UMAP",
  nNeighbors = 30,
  minDist = 0.5,
  metric = "cosine"
)
p1 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Sample", embedding = "UMAP")
p2 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Clusters", embedding = "UMAP")
ggAlignPlots(p1, p2, type = "h")
plotPDF(p1,p2, name = "Plot-UMAP-Sample-Clusters.pdf", ArchRProj = proj2, addDOC = FALSE, width = 5, height = 5)
#t-SNE
proj2 <- addTSNE(
  ArchRProj = proj2,
  reducedDims = "IterativeLSI",
  name = "TSNE",
  perplexity = 30
)
p1 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Sample", embedding = "TSNE")
```

```
p2 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Clusters", embedding = "TSNE")
ggAlignPlots(p1, p2, type = "h")
# Dimensional Reduction After Harmony
# UMAP
proj2 <- addUMAP(
  ArchRProj = proj2,
  reducedDims = "Harmony",
  name = "UMAPHarmony",
  nNeighbors = 30,
  minDist = 0.5,
  metric = "cosine"
)
p3 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Sample", embedding = "UMAPHarmony")
p4 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Clusters", embedding = "UMAPHarmony")
ggAlignPlots(p3, p4, type = "h")
plotPDF(p1,p2,p3,p4, name = "Plot-tSNEHarmony-Sample-Clusters.pdf", ArchRProj = proj2, addDOC = FALSE, width = 5, height = 5)
# t-SNE
proj2 <- addTSNE(
  ArchRProj = proj2,
  reducedDims = "Harmony",
  name = "TSNEHarmony",
  perplexity = 30
)
p3 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Sample", embedding = "TSNEHarmony")
p4 <- plotEmbedding(ArchRProj = proj2, colorBy = "cellColData", name = "Clusters", embedding = "TSNEHarmony")
ggAlignPlots(p3, p4, type = "h")
## Identifying Marker Genes ##
devtools::install_github("immunogenomics/presto")
markersGS <- getMarkerFeatures(
  ArchRProj = proj2,
  useMatrix = "GeneScoreMatrix",
  groupBy = "Clusters",
  bias = c("TSSEnrichment", "log10(nFrags)"),
  testMethod = "wilcoxon"
)
markerList <- getMarkers(markersGS, cutOff = "FDR <= 0.01 & Log2FC >= 1.25")
#markerList$C6
#write.table(markerList, file="markersGS.txt",sep = "\t", row.names = TRUE)
markerGenes  <- c("Nanog", "Pou5f1", "Sox2")
heatmapGS <- markerHeatmap(
  seMarker = markersGS,
  cutOff = "FDR <= 0.01 & Log2FC >= 1.25",
  labelMarkers = NULL,
  transpose = TRUE
)
ComplexHeatmap::draw(heatmapGS, heatmap_legend_side = "bot", annotation_legend_side = "bot")
```

```
plotPDF(heatmapGS, name = "GeneScores-Marker-Heatmap", width = 8, height = 6, ArchRProj = projHeme2, addDOC = FALSE)
p <- plotEmbedding(
  ArchRProj = proj2,
  colorBy = "GeneScoreMatrix",
  name = markerGenes,
  embedding = "UMAP",
  quantCut = c(0.01, 0.95),
  imputeWeights = NULL
)
p$Nanog
install.packages("Rmagic")
markerGenes  <- c(
  "Nanog", "Pou5f1", "Sox2"
)
p <- plotEmbedding(
  ArchRProj = proj2,
  colorBy = "GeneScoreMatrix",
  name = markerGenes,
  embedding = "UMAP",
  imputeWeights = getImputeWeights(proj2)
)
install.packages("rhandsontable")
plotPDF(p1,p2,p3,p4, name = "Plot-UMAP2Harmony-Sample-Clusters.pdf", ArchRProj = proj2, addDOC = FALSE, width = 5, height = 5)
## Pseudo-Bulk Replicates in ArchR ##
proj3 <- addGroupCoverages(ArchRProj = proj2, groupBy = "Clusters")
##install MACS2 by running the following command in terminal window
#pip install MACS2
pathToMacs2 <- findMacs2()
proj3 <- addReproduciblePeakSet(
  ArchRProj = proj3,
  groupBy = "Clusters",
  pathToMacs2 = pathToMacs2
)
getPeakSet(proj3)
#Calling peaks with Tile Matrix
projTmp <- addReproduciblePeakSet(
  ArchRProj = proj3,
  groupBy = "Clusters",
  peakMethod = "Tiles",
  method = "p"
)
getPeakSet(projTmp)
#Comparing methods
length(subsetByOverlaps(getPeakSet(proj3), getPeakSet(projTmp))) / length(getPeakSet(proj3)) #0.5834776
length(subsetByOverlaps(getPeakSet(projTmp), getPeakSet(proj3))) / length(getPeakSet(projTmp)) #0.9292217
length(subsetByOverlaps(resize(getPeakSet(proj3), 1000, "center"), getPeakSet(projTmp))) / length(getPeakSet(proj3)) #0.613327
length(subsetByOverlaps(getPeakSet(projTmp), resize(getPeakSet(proj3), 1000, "center"))) / length(getPeakSet(projTmp)) #0.9698539
#Save ArchR project
```

```
saveArchRProject(ArchRProj = proj3, outputDirectory = "Save-Proj3", load = FALSE)
# Add Peak Matrix
addArchRThreads(threads = 1, force = TRUE)
proj4 <- addPeakMatrix(proj3)
getAvailableMatrices(proj4)
## Identifying Marker Peaks ##
table(proj4$Clusters)
markersPeaks_cluster <- getMarkerFeatures(
  ArchRProj = proj4,
  useMatrix = "PeakMatrix",
  groupBy = "Clusters",
  bias = c("TSSEnrichment", "log10(nFrags)"),
  testMethod = "wilcoxon"
)
markersPeaks_cluster
markerList_cluster <- getMarkers(markersPeaks_cluster, cutOff = "FDR <= 0.01 & Log2FC >= 1", returnGR = TRUE)
markerList_cluster
markersPeaks_sample <- getMarkerFeatures(
  ArchRProj = proj4,
  useMatrix = "PeakMatrix",
  groupBy = "Sample",
  bias = c("TSSEnrichment", "log10(nFrags)"),
  testMethod = "wilcoxon"
)
markersPeaks_sample
markerList_sample <- getMarkers(markersPeaks_sample, cutOff = "FDR <= 0.01 & Log2FC >= 1", returnGR = TRUE)
markerList_sample
## Plotting Marker Peaks ##
# Marker Peak Heatmaps
heatmapPeaks_cluster <- markerHeatmap(
  seMarker = markersPeaks_cluster,
  cutOff = "FDR <= 0.1 & Log2FC >= 0.5",
  transpose = TRUE
)
draw(heatmapPeaks_cluster, heatmap_legend_side = "bot", annotation_legend_side = "bot")
heatmapPeaks_sample <- markerHeatmap(
  seMarker = markersPeaks_sample,
  cutOff = "FDR <= 0.1 & Log2FC >= 0.5",
  transpose = TRUE
)
draw(heatmapPeaks_sample, heatmap_legend_side = "bot", annotation_legend_side = "bot")
# Marker Peak MA and Volcano Plots
pma <- markerPlot(seMarker = markersPeaks_sample, name = "ESC", cutOff = "FDR <= 0.1 & Log2FC >= 1", plotAs = "MA")
pma
pv <- markerPlot(seMarker = markersPeaks_sample, name = "ESC", cutOff = "FDR <= 0.1 & Log2FC >= 1", plotAs = "Volcano")
pv
# Marker Peaks in Browser Tracks
set.seed(100)
p <- plotBrowserTrack(
  ArchRProj = proj4,
```

```
  groupBy = "Sample",
  geneSymbol = c("NANOG"),
  features =  getMarkers(markersPeaks_sample, cutOff = "FDR <= 0.1 & Log2FC >= 1",
  ↪  returnGR = TRUE),
  upstream = 50000,
  downstream = 50000
)
grid::grid.newpage()
grid::grid.draw(p$NANOG)
## Pairwise Testing Between Groups ##
markerTest <- getMarkerFeatures(
  ArchRProj = proj4,
  useMatrix = "PeakMatrix",
  groupBy = "Sample",
  testMethod = "wilcoxon",
  bias = c("TSSEnrichment", "log10(nFrags)"),
  useGroups = "Count_A",
  bgdGroups = "Count_B"
)
pma <- markerPlot(seMarker = markerTest, name = "Count_A", cutOff = "FDR <= 0.1 &
↪  abs(Log2FC) >= 1", plotAs = "MA")
pma
pv <- markerPlot(seMarker = markerTest, name = "Count_A", cutOff = "FDR <= 0.1 &
↪  abs(Log2FC) >= 1", plotAs = "Volcano")
pv
```

Listing 2: R code used for scATAC-seq analysis.

```r
library("tximport")
library("readr")
library("GenomicFeatures")
#Creating GFF file
gffFile <-
↪  "C:/Users/Documents/R/win-library/4.0/tximportData/extdata/gencode.v37.annotation.gtf"
#Creating TxDb object
txdb <- makeTxDbFromGFF(file=gffFile)
k <- keys(txdb, keytype = "GENEID")
df <- select(txdb, keys = k, keytype = "GENEID", columns = "TXNAME")
tx2gene <- df[, 2:1]
head(tx2gene)
#Locating the directory with  transcript abundance estimate files
dir<- "C:/Users//salmon_iPSC"
#Creating a vector for quantification files
samples <- read.table(file.path(dir, "samples.txt"), header = TRUE)
samples
files <- file.path(dir,"quants", samples$run, "quant.sf")
names(files) <- samples$run
files
#Importing transcript level estimates
txi <- tximport(files, type = "salmon", tx2gene = tx2gene)
names(txi)
head(txi$counts)
#Exporting the read counts to a csv file
write.csv(txi, file = 'C:/Users/Documents/Counts.csv')
```

Listing 3: R code used for bulk-RNA seq analysis.

D DISSECTING MOLECULAR MECHANISMS OF ASTROCYTE-NEURON INTERACTION IN RETT SYNDROME

Work in this chapter was adapted from:

Dissecting Molecular Mechanisms of Astrocyte-Neuron Interaction in Rett Syndrome Using Arrays of Gene-Edited, Patient-derived Cells

Kaivalya Molugu, Jason Kim, Greta Brown, Qiang Chang, Krishanu Saha

D.1 Introduction

Rett syndrome (RTT) is a pervasive developmental disorder of the brain that generates dysfunction in both neurons and astrocytes. While neuron-astrocyte interaction has been observed in culture and animals, the molecular mechanisms mediating this interaction are poorly understood. In this project, we utilized several human stem cell models of RTT and expertise in astrocyte-neuron co-culture as well as morphological and functional analysis of these cells. In addition, we developed a new platform to array microscale islands of astrocyte-neuron co-cultures, gene-edit them in high-throughput, and analyze them *in situ* with high content analysis without dissociating these cultures. This novel system effectively separates neurons from dense cultures into isolated 'islands' that can be easily analyzed for complexity. Overall, this project examined whether automated analysis of spatially isolated astrocyte-neuron co-cultures is feasible, and has been applying this new approach to understand neuronal biology with human RTT stem cell models. In the long term, this platform can be adapted further to advance drug discovery/toxicology for neurodevelopmental disorders.

D.2 Methodology and Results

Micropatterned culture and imaging system was established for richer classification of neurite complexity. First, a microscale system was generated in a 24-well format by creating hydrophilic polyethylene glycol (PEG) brushes that resist protein adsorption at defined locations on a gold-coated glass sheet. This sheet was then combined with a bottomless standard tissue culture 24-well plate to form the microscale system (**Figure D.1**). Initial attempts were made to pattern astrocyte-neuronal co-cultures into circles of defined radii varying from 100 to 300 μm (μFeatures) and then identified 100 μm radius circles to be the most suitable for obtaining single neuron attachment per μFeature. Next, several pilot experiments were run on 24-well plates patterned with 128 circles of 100μm radius

in each well, to identify the optimal seeding density, culture time, and image settings. Wild-type mouse astrocytes were plated on the patterned well plates. As expected, cell localized to the areas defined by the micropatterning. The astrocytes seeded uniformly on the substrates and have persisted over the course of the experiments. One week later, human stem-cell-derived neurons (H9-GFP neurons) were plated on top of the astrocytes, with a seeding density of neurons ranging from 1.0×10^4 to 6.4×10^5 cells/well. Some representative bright field (astrocytes + neurons) and GFP channel (neurons) images are shown in **Figure D.1B**. Minimal or no attachment of neurons was observed at lower seeding densities, multiple neurons attached per μFeature at higher seeding densities. An optimal neuron seeding density of 16×10^4 cells/well yielded maximum number of μFeatures with single neuron attachment. Overall, a robust protocol and methodology have been established to consistently produce 12-well and 24-well plates with two optimized micropatterns (400-600 μm diameter islands; 75-701 single neurons/well).

The accuracy and speed of the microscale system are currently being evaluated with H9-GFP human embryonic stem cell (hESC)-derived neural cells that have well-characterized differences in complexity. This methodology uses mouse astrocytes as a feeder layer, and we have generated these cells by passaging the cultures several times to remove all the mouse neurons. An *in vitro* human cell-based RTT model will be generated via *MECP2* gene knockout in H9-GFP neurons using CRISPR-Cas9 genome editors. The *MECP2* gene has four exons, with different isoforms being expressed from exons 1 and 2. As exon 3 is the first shared exon among all isoforms, exon 3 was chosen as the target exon to ablate all MeCP2 protein isoforms. Two single guide RNAs (sgRNAs) were designed within exon 3 (**Figure D.2A**) using the inDelphi and Benchling software programs to ensure high frameshift frequency, high on-target score, and low off-target score. The sgRNAs were individually cloned into the lentiCRISPRv2-mCherry plasmid, which encodes SpCas9-P2A-mCherry and the sgRNA cloning site (**Figure D.2B**). Using BsmBI digestion of the plasmid, sgRNAs were cloned into the guide RNA scaffold, which was verified via Sanger sequencing

(**Figure D.2**). After packaging these constructs, lentivirus was tested by performing viral transductions on 3T3 cells and Human Embryonic Kidney cells (HEKs) plated at a seeding density of 2500 cells/well of a 96-well plate (data shown for 3T3 cells only), with 0 to 8 units of virus per well (1 unit of virus = 1.5 μL of freshly prepared virus). 3T3 cells were imaged using an epifluorescence microscope on day 2 after transduction (Figure D.2D) and flow cytometry was performed on day 3 after transduction (**Figure D.2E,F**), where noted mCherry^{+} cells were noted indicating a successful viral infection for both the sgRNAs. These *MECP2* genome editors were then used to transduce H9-GFP stem cells and differentiated them into neurons using the neural differentiation protocol (**Figure D.3**).

At present, several samples with variable neuronal complexity are awaiting immunostaining analysis to quantify the neuronal complexity. The next phase of this work is to apply this platform to human neurons that model RTT and Fragile X Syndrome (FXS).

D.3 Significance of Research

This work has established a high-throughput platform that incorporates an innovative approach to gene edit and measures neuronal complexity using automated methods without requiring manual tracing of neurons, which is labor-intensive. These methods will shed light on the molecular mechanisms underpinning lower neurite complexity in RTT patients, and potentially identify molecular targets for therapeutic interventions. Because the approach can easily be utilized with different genome editors targeting any portion of the genome, this platform can be applied to a broad spectrum of experiments aimed at understanding the genetic and molecular mechanisms of many other neurological disorders, including autism and schizophrenia.

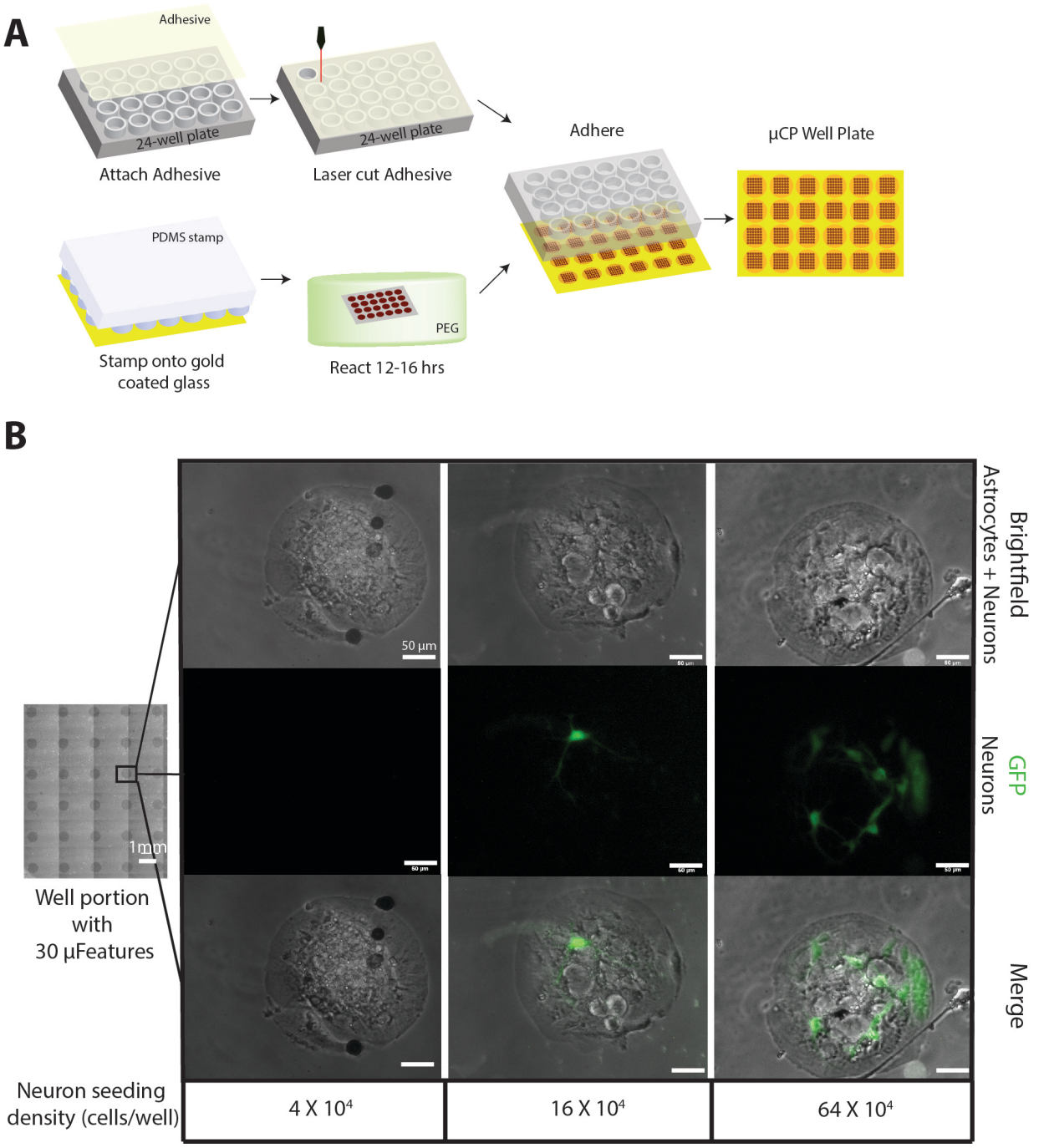


Figure D.1: Caption next page

Figure D.1: **Astrocyte-neuronal co-cultures on patterned 24-well platform.** A) Schematic showing the generation of microscale well platform. Gold-coated glass is stamped with initiator solution using PDMS mold and then reacted in PEG solution. Glass is then combined with a standard bottomless 24-well plate that has a double side adhesive attached. B) Representative images of mouse astrocyte-H9 GFP neuron co-cultures (BF) on patterned well plates at different neuron (GFP channel) seeding densities. Each island is 100 μm in radius. Images are taken 3 days after seeding the H9-GFP neurons.

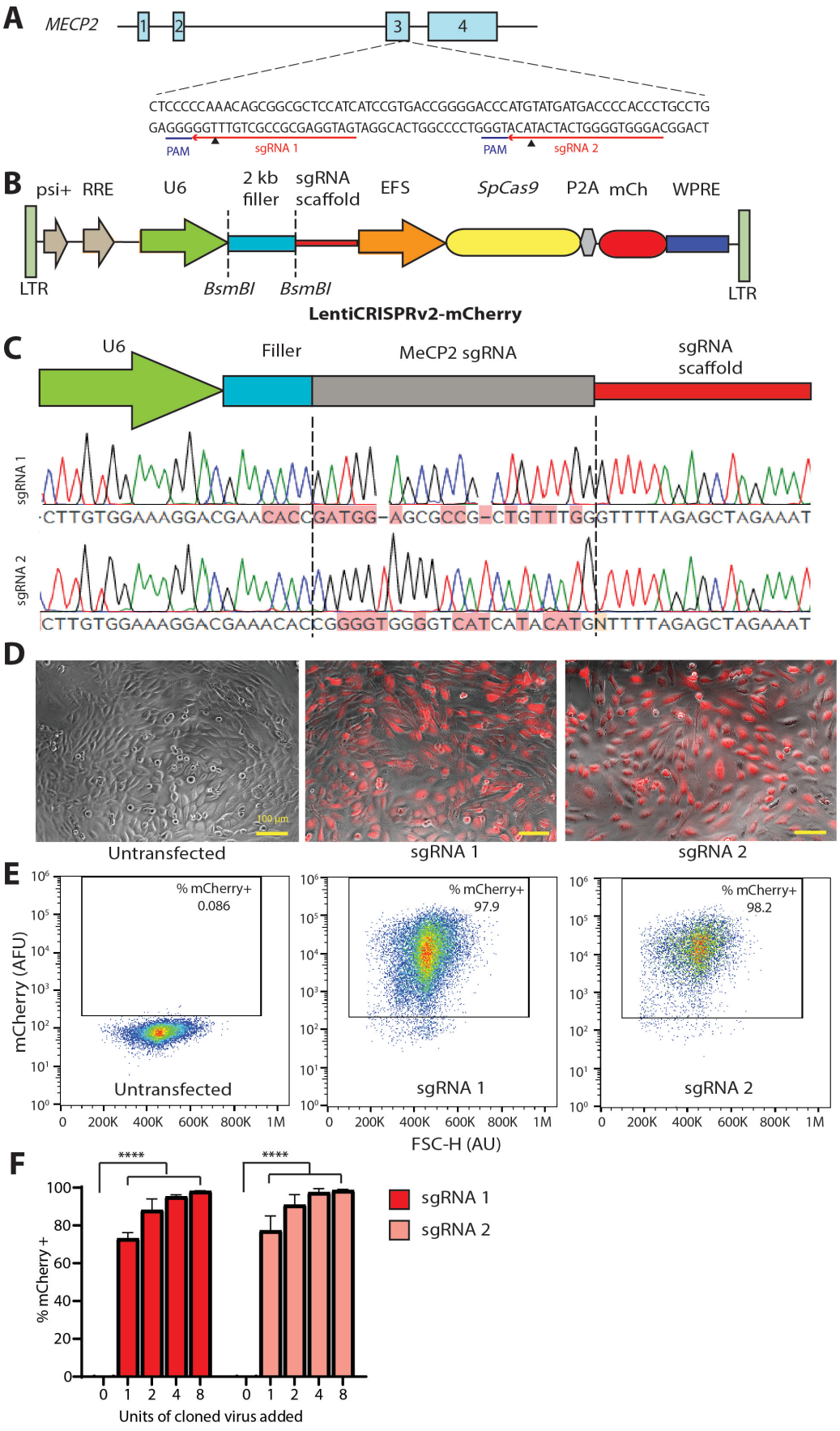

A
MECP2
1
2
3
4
CTCCCCCAAACAGCGGCGCTCCATCATCCGTGACCGGGGACCCATGTATGATGACCCCACCCTGCCTG
GAGGGGGTTTGTCGCCGCGAGGTAGTAGGCACTGGCCCCTGGGTACATACTACTGGGGTGGGACGGACT
PAM
sgRNA 1
PAM
sgRNA 2
B
psi+
RRE
U6
2 kb filler
sgRNA scaffold
EFS
SpCas9
P2A
mCh
WPRE
LTR
BsmBI
BsmBI
LTR
LentiCRISPRv2-mCherry
C
U6
Filler
MeCP2 sgRNA
sgRNA scaffold
sgRNA 1
CTTGTGGAAAGGACGAACACCGATGG-AGCGCCG-CTGTTTGGGTTTTAGAGCTAGAAAT
sgRNA 2
CTTGTGGAAAGGACGAAACACCGGGGTGGGGTCATCATACATGNTTTTAGAGCTAGAAAT
D
100 µm
Untransfected
sgRNA 1
sgRNA 2
E
% mCherry+
0.086
% mCherry+
97.9
% mCherry+
98.2
mCherry (AFU)
Untransfected
sgRNA 1
sgRNA 2
FSC-H (AU)
F
****
****
% mCherry +
sgRNA 1
sgRNA 2
Units of cloned virus added

Figure D.2

Figure D.2: **Production and testing of a genome editor to generate a human RTT model.** This example illustrates the production and testing of a lentivirus with a CRISPR-Cas9 sgRNA targeting exon 3 of *MECP2*. A) Schematic of the human *MECP2* locus with the two sgRNAs target sequences, protoadjacent motif (PAM) sequences, and cut-sites (black arrow) labeled. B) Schematic of the LentiCRISPRv2-mCherry plasmid consisting of SpCas9, sgRNA, and mCherry expression constructs. C) Sanger sequencing results of the cloned plasmid to verify the presence of the sgRNA insert. D) Representative images of untransfected 3T3 cells and cells transduced with 8 units of sgRNA1 and sgRNA2 virus per well (1 unit of virus = 1.5 μL of freshly prepared virus). Brightfield shows the 3T3 cells and the red channel shows the mCherry+ 3T3 cells. E) Flow cytometry analysis of untransfected 3T3 cells and cells transduced with 8 units of sgRNA1 and sgRNA2 virus per well. Percentage of mCherry+ cells are labeled. F) Plot of the percentage of mCherry+ cells versus the units of sgRNA1 and sgRNA2 virus added per well of 3T3s (n=2). *p*-values generated by two-way ANOVA using the Dunnett's test for multiple comparisons to untransfected cells; ns = $p \geqslant 0.05$, * for $p < 0.05$, ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$.

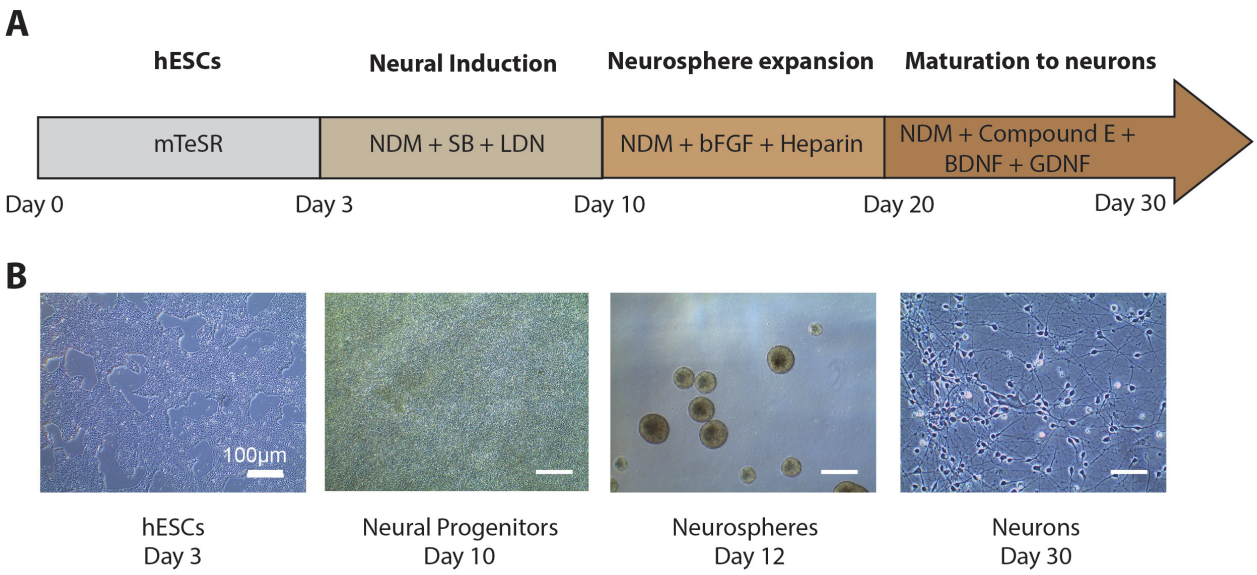


Figure D.3: **Neuronal differentiation of human embryonic stem cells using dual SMAD inhibition.** A) Schematic of the protocol for neural differentiation using dual SMAD inhibition. The protocol includes 4 steps: 1) Growing human embryonic stem cells (hESCs) to 50% confluency (days 0-3). 2) Neural induction (days 3-10) to neural progenitors, by using Neural Differentiation Medium (NDM = DMEM/F12, Neurobasal, B27, N2) and blocking SMAD signaling using TGF-β inhibitor SB431542 and BMP inhibitor LDN193189. 3) Neurosphere formation by dissociating cells using dispase, and further neurosphere maintenance and expansion (days 10-20) in low attachment flasks by supplementing NDM with bFGF growth factor and heparin to stabilize the bFGF. 4) Neural maturation (days 20-30), by dissociating neurospheres with accutase and re-plating them on Matrigel-coated plates in the presence of brain-derived neurotrophic factor (BDNF), glia-derived neurotrophic factor (GDNF), and compound E (for blockade of Notch signaling). B) Brightfield representative images of H9 hESCs, neural progenitors, neurospheres, and neurons on days 3, 10, 12, 30 of neural differentiation. The plating of dissociated neurospheres in this protocol produced the most consistent attachment of mature neurons on our platform.

BIBLIOGRAPHY

1. De Los Angeles, A. *et al.* Hallmarks of pluripotency. en. *Nature* **525,** 469–478. doi:10.1038/nature15515 (Sept. 2015).

2. Avior, Y., Sagi, I. & Benvenisty, N. Pluripotent stem cells in disease modelling and drug discovery. en. *Nature Reviews Molecular Cell Biology* **17,** 170–182. doi:10.1038/nrm.2015.27 (Mar. 2016).

3. Trounson, A. & DeWitt, N. D. Pluripotent stem cells progressing to the clinic. en. *Nature Reviews Molecular Cell Biology* **17,** 194–200. doi:10.1038/nrm.2016.10 (Mar. 2016).

4. Deinsberger, J., Reisinger, D. & Weber, B. Global trends in clinical trials involving pluripotent stem cells: a systematic multi-database analysis. en. *npj Regenerative Medicine* **5,** 15. doi:10.1038/s41536-020-00100-4 (Dec. 2020).

5. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. en. *Nature* **292,** 154–156. doi:10.1038/292154a0 (July 1981).

6. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. en. *Cell* **126,** 663–676. doi:10.1016/j.cell.2006.07.024 (Aug. 2006).

7. Thomson, J. A. Embryonic Stem Cell Lines Derived from Human Blastocysts. en. *Science* **282,** 1145–1147. doi:10.1126/science.282.5391.1145 (Nov. 1998).

8. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. en. *Nature Reviews Molecular Cell Biology* **17,** 183–193. doi:10.1038/nrm.2016.8 (Mar. 2016).

9. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. en. **318,** 5 (2007).

10. Zhao, X.-y. *et al.* iPS cells produce viable mice through tetraploid complementation. en. *Nature* **461,** 86–90. doi:10.1038/nature08267 (Sept. 2009).

11. Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. en. *Nature* **448,** 318–324. doi:10.1038/nature05944 (July 2007).

12. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. en. *Nature* **448,** 313–317. doi:10.1038/nature05934 (July 2007).

13. Howden, S. E. *et al.* Genetic correction and analysis of induced pluripotent stem cells from a patient with gyrate atrophy. eng. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 6537–6542. doi:10.1073/pnas.1103388108 (Apr. 2011).

14. Tucker, B. A., Mullins, R. F. & Stone, E. M. Stem cells for investigation and treatment of inherited retinal disease. en. *Human Molecular Genetics,* ddu124. doi:10.1093/hmg/ddu124 (Mar. 2014).

15. Lustremant, C. *et al.* Human induced pluripotent stem cells as a tool to model a form of Leber congenital amaurosis. eng. *Cellular Reprogramming* **15,** 233–246. doi:10.1089/cell.2012.0076 (June 2013).

16. Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. en. *Nature* **435,** 646–651. doi:10.1038/nature03556 (June 2005).

17. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. en. *Nature Biotechnology* **29,** 143–148. doi:10.1038/nbt.1755 (Feb. 2011).

18. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. en. *Science* **339,** 823–826. doi:10.1126/science.1232033 (Feb. 2013).

19. Narsinh, K. H. *et al.* Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells. en. *Journal of Clinical Investigation* **121,** 1217–1221. doi:10.1172/JCI44635 (Mar. 2011).

20. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. en. *Genes & Development* **24,** 2239–2263. doi:10.1101/gad.1963910 (Oct. 2010).

21. Buganim, Y. *et al.* The Developmental Potential of iPSCs Is Greatly Influenced by Reprogramming Factor Selection. *Cell Stem Cell* **15,** 295–309. doi:10.1016/j.stem.2014.07.003 (Sept. 2014).

22. Ohi, Y. *et al.* Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. en. *Nature Cell Biology* **13,** 541–549. doi:10.1038/ncb2239 (May 2011).

23. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. en. *Nature* **471,** 68–73. doi:10.1038/nature09798 (Mar. 2011).

24. Zhao, T., Zhang, Z.-N., Rong, Z. & Xu, Y. Immunogenicity of induced pluripotent stem cells. en. *Nature* **474,** 212–215. doi:10.1038/nature10135 (June 2011).

25. Beers, J. *et al.* Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions. en. *Nature Protocols* **7,** 2029–2040. doi:10.1038/nprot.2012.130 (Nov. 2012).

26. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. en. *Nature Methods* **8,** 424–429. doi:10.1038/nmeth.1593 (May 2011).

27. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. en. *Nature* **462,** 595–601. doi:10.1038/nature08592 (Dec. 2009).

28. Carey, B. W. *et al.* Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell stem cell* **9,** 588–598 (2011).

29. Hanna, J. *et al.* Metastable Pluripotent States in NOD-Mouse-Derived ESCs. *Cell Stem Cell* **16,** 566–568. doi:10.1016/j.stem.2015.04.007 (May 2015).

30. Chen, G., Guo, Y., Li, C., Li, S. & Wan, X. Small Molecules that Promote Self-Renewal of Stem Cells and Somatic Cell Reprogramming. en. *Stem Cell Reviews and Reports* **16,** 511–523. doi:10.1007/s12015-020-09965-w (June 2020).

31. Lin, T. & Wu, S. Reprogramming with Small Molecules instead of Exogenous Transcription Factors. en. *Stem Cells International* **2015,** 1–11. doi:10.1155/2015/794632 (2015).

32. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced Pluripotent Stem Cells Generated Without Viral Integration. *Science* **322,** 945–949. doi:10.1126/science.1162494 (2008).

33. Okita, K. *et al.* An Efficient Nonviral Method to Generate Integration-Free Human-Induced Pluripotent Stem Cells from Cord Blood and Peripheral Blood Cells. en. *STEM CELLS* **31,** 458–466. doi:10.1002/stem.1293 (Mar. 2013).

34. Woltjen, K. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. en. **458,** 6 (2009).

35. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Ef cient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. en. **85,** 15.

36. Kim, D. *et al.* Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. en, 9 (2010).

37. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells using synthetic modified mRNA. en, 23 (2013).

38. Yu, J. *et al.* Human Induced Pluripotent Stem Cells Free of Vector and Transgene Sequences. en. *Science* **324,** 797–801. doi:10.1126/science.1172482 (May 2009).

39. Masuda, S. *et al.* Chemically induced pluripotent stem cells (CiPSCs): a transgene-free approach. en, 2.

40. Yoshioka, N. *et al.* Efficient Generation of Human iPS Cells by a Synthetic Self-Replicative RNA. en, 21 (2014).

41. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. en. *Cell* **131,** 861–872. doi:10.1016/j.cell.2007.11.019 (Nov. 2007).

42. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. en. *Nature* **460,** 49–52. doi:10.1038/nature08180 (July 2009).

43. Guo, S. *et al.* Nonstochastic Reprogramming from a Privileged Somatic Cell State. en. *Cell* **156,** 649–662. doi:10.1016/j.cell.2014.01.020 (Feb. 2014).

44. Wakao, S. *et al.* Multilineage-differentiating stress-enduring (Muse) cells are a primary source of induced pluripotent stem cells in human fibroblasts. en. *Proceedings of the National Academy of Sciences* **108,** 9875–9880. doi:10.1073/pnas.1100816108 (June 2011).

45. Di Stefano, B. *et al.* C/EBP poises B cells for rapid reprogramming into induced pluripotent stem cells. en. *Nature* **506,** 235–239. doi:10.1038/nature12885 (Feb. 2014).

46. Rais, Y. *et al.* Corrigendum: Deterministic direct reprogramming of somatic cells to pluripotency. en, 1.

47. Shakiba, N. *et al.* Cell competition during reprogramming gives rise to dominant clones. en. *Science* **364,** eaan0925. doi:10.1126/science.aan0925 (Apr. 2019).

48. Yunusova, A. M., Fishman, V. S., Vasiliev, G. V. & Battulin, N. R. Deterministic versus stochastic model of reprogramming: new evidence from cellular barcoding technique. en. *Open Biology* **7,** 160311. doi:10.1098/rsob.160311 (Apr. 2017).

49. Buganim, Y. *et al.* Single-Cell Expression Analyses during Cellular Reprogramming Reveal an Early Stochastic and a Late Hierarchic Phase. en. *Cell* **150,** 1209–1222. doi:10.1016/j.cell.2012.08.023 (Sept. 2012).

50. Huangfu, D. *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. en. *Nature Biotechnology* **26,** 1269–1275. doi:10.1038/nbt.1502 (Nov. 2008).

51. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. en. *Nature Biotechnology* **26,** 101–106. doi:10.1038/nbt1374 (Jan. 2008).

52. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. en. *Nature Reviews Genetics* **14,** 427–439. doi:10.1038/nrg3473 (June 2013).

53. Polo, J. M. *et al.* A Molecular Roadmap of Reprogramming Somatic Cells into iPS Cells. en. *Cell* **151,** 1617–1632. doi:10.1016/j.cell.2012.11.039 (Dec. 2012).

54. Mattout, A., Biran, A. & Meshorer, E. Global epigenetic changes during somatic cell reprogramming to iPS cells. en. *Journal of Molecular Cell Biology* **3,** 341–350. doi:10.1093/jmcb/mjr028 (Dec. 2011).

55. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. en. *Nature* **502,** 462–471. doi:10.1038/nature12749 (Oct. 2013).

56. Maherali, N. *et al.* Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. en. *Cell Stem Cell* **1,** 55–70. doi:10.1016/j.stem.2007.05.014 (June 2007).

57. Koche, R. P. *et al.* Reprogramming Factor Expression Initiates Widespread Targeted Chromatin Remodeling. en. *Cell Stem Cell* **8,** 96–105. doi:10.1016/j.stem.2010.12.001 (Jan. 2011).

58. Tran, K. A. *et al. Defining reprogramming checkpoints from single-cell analysis of induced pluripotency* en. preprint (Developmental Biology, Jan. 2019). doi:10.1101/531244.

59. Meshorer, E. *et al.* Hyperdynamic Plasticity of Chromatin Proteins in Pluripotent Embryonic Stem Cells. en. *Developmental Cell* **10,** 105–116. doi:10.1016/j.devcel.2005.10.017 (Jan. 2006).

60. Stephens, A. D. *et al.* Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins. en. *Molecular Biology of the Cell* **29** (ed Misteli, T.) 220–233. doi:10.1091/mbc.E17-06-0410 (Jan. 2018).

61. Stephens, A. D., Banigan, E. J. & Marko, J. F. Chromatin's physical properties shape the nucleus and its functions. en. *Current Opinion in Cell Biology* **58,** 76–84. doi:10.1016/j.ceb.2019.02.006 (June 2019).

62. Orkin, S. H. & Hochedlinger, K. Chromatin Connections to Pluripotency and Cellular Reprogramming. en. *Cell* **145,** 835–850. doi:10.1016/j.cell.2011.05.019 (June 2011).

63. Cordie, T. *et al.* Nanofibrous Electrospun Polymers for Reprogramming Human Cells. en. *Cellular and Molecular Bioengineering* **7,** 379–393. doi:10.1007/s12195-014-0341-z (Sept. 2014).

64. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. en. *Science* **324,** 1029–1033. doi:10.1126/science.1160809 (May 2009).

65. Nishimura, K., Fukuda, A. & Hisatake, K. Mechanisms of the Metabolic Shift during Somatic Cell Reprogramming. en. *International Journal of Molecular Sciences* **20,** 2254. doi:10.3390/ijms20092254 (May 2019).

66. Mathieu, J. & Ruohola-Baker, H. Metabolic remodeling during the loss and acquisition of pluripotency. en. *Development* **144,** 541–551. doi:10.1242/dev.128389 (Feb. 2017).

67. Rigaud, V. O. C., Hoy, R., Mohsin, S. & Khan, M. Stem Cell Metabolism: Powering Cell-Based Therapeutics. en. *Cells* **9,** 2490. doi:10.3390/cells9112490 (Nov. 2020).

68. Kida, Y. S. *et al.* ERRs Mediate a Metabolic Switch Required for Somatic Cell Reprogramming to Pluripotency. en. *Cell Stem Cell* **16,** 547–555. doi:10.1016/j.stem.2015.03.001 (May 2015).

69. Ishida, T., Nakao, S., Ueyama, T., Harada, Y. & Kawamura, T. Metabolic remodeling during somatic cell reprogramming to induced pluripotent stem cells: involvement of hypoxia-inducible factor 1. en. *Inflammation and Regeneration* **40,** 8. doi:10.1186/s41232-020-00117-8 (Dec. 2020).

70. Hawkins, K. E. *et al.* NRF2 Orchestrates the Metabolic Shift during Induced Pluripotent Stem Cell Reprogramming. en. *Cell Reports* **14,** 1883–1891. doi:10.1016/j.celrep.2016.02.003 (Mar. 2016).

71. Ma, T. *et al.* Atg5-independent autophagy regulates mitochondrial clearance and is essential for iPSC reprogramming. en. *Nature Cell Biology* **17,** 1379–1387. doi:10.1038/ncb3256 (Nov. 2015).

72. Prigione, A. *et al.* Human Induced Pluripotent Stem Cells Harbor Homoplasmic and Heteroplasmic Mitochondrial DNA Mutations While Maintaining Human Embryonic Stem Cell–like Metabolic Reprogramming. en, 12.

73. Panopoulos, A. D. *et al.* The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. en. *Cell Research* **22,** 168–177. doi:10.1038/cr.2011.177 (Jan. 2012).

74. Zhu, S. *et al.* Reprogramming of Human Primary Somatic Cells by OCT4 and Chemical Compounds. en. *Cell Stem Cell* **7,** 651–655. doi:10.1016/j.stem.2010.11.015 (Dec. 2010).

75. Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T. & Yamanaka, S. Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. en. *Cell Stem Cell* **5,** 237–241. doi:10.1016/j.stem.2009.08.001 (Sept. 2009).

76. Musunuru, K. Genome editing of human pluripotent stem cells to generate human cellular disease models. en. *Disease Models and Mechanisms* **6,** 896–904. doi:10.1242/dmm.012054 (July 2013).

77. Simara, P., Motl, J. A. & Kaufman, D. S. Pluripotent stem cells and gene therapy. en. *Translational Research* **161,** 284–292. doi:10.1016/j.trsl.2013.01.001 (Apr. 2013).

78. Sarkar, A. *et al.* Efficient Generation of CA3 Neurons from Human Pluripotent Stem Cells Enables Modeling of Hippocampal Connectivity In Vitro. en. *Cell Stem Cell* **22,** 684–697.e9. doi:10.1016/j.stem.2018.04.009 (May 2018).

79. Holmgren, G., Sartipy, P., Andersson, C. X., Lindahl, A. & Synnergren, J. Expression Profiling of Human Pluripotent Stem Cell-Derived Cardiomyocytes Exposed to Doxorubicin—Integration and Visualization of Multi-Omics Data. en. *Toxicological Sciences* **163,** 182–195. doi:10.1093/toxsci/kfy012 (May 2018).

80. Czerniecki, S. M. *et al.* High-Throughput Screening Enhances Kidney Organoid Differentiation from Human Pluripotent Stem Cells and Enables Automated Multidimensional Phenotyping. en. *Cell Stem Cell* **22,** 929–940.e4. doi:10.1016/j.stem.2018.04.022 (June 2018).

81. Ebert, A. D., Liang, P. & Wu, J. C. Induced Pluripotent Stem Cells as a Disease Modeling and Drug Screening Platform: en. *Journal of Cardiovascular Pharmacology* **60,** 408–416. doi:10.1097/FJC.0b013e318247f642 (Oct. 2012).

82. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. en. *Science* **337,** 816–821. doi:10.1126/science.1225829 (Aug. 2012).

83. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife Sciences* **2.** doi:10.7554/eLife.00471 (2013).

84. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (New York, N.Y.)* **339,** 819–823. doi:10.1126/science.1231143 (Feb. 2013).

85. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. eng. *Nature Biotechnology* **31,** 230–232. doi:10.1038/nbt.2507 (Mar. 2013).

86. Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. en. *Nature Biotechnology.* doi:10.1038/nbt.2501 (2013).

87. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. en. *Nature Reviews Microbiology* **18,** 67–83. doi:10.1038/s41579-019-0299-x (Feb. 2020).

88. Doudna, J. A. The promise and challenge of therapeutic genome editing. en. *Nature* **578,** 229–236. doi:10.1038/s41586-020-1978-5 (Feb. 2020).

89. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. en. *Nature Biotechnology* **38,** 824–844. doi:10.1038/s41587-020-0561-9 (July 2020).

90. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. en. *Nature Reviews Molecular Cell Biology* **20,** 490–507. doi:10.1038/s41580-019-0131-5 (Aug. 2019).

91. Smith, J. *et al.* A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. en. *Nucleic Acids Research* **34,** e149–e149. doi:10.1093/nar/gkl720 (Dec. 2006).

92. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. eng. *Nature* **507,** 62–67. doi:10.1038/nature13011 (Mar. 2014).

93. O'Connell, M. R. *et al.* Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516,** 263–266. doi:10.1038/nature13769 (Dec. 2014).

94. Rothkamm, K., Krüger, I., Thompson, L. H. & Löbrich, M. Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. en. *Molecular and Cellular Biology* **23,** 5706–5715. doi:10.1128/MCB.23.16.5706-5715.2003 (Aug. 2003).

95. Stark, J. M., Pierce, A. J., Oh, J., Pastink, A. & Jasin, M. Genetic Steps of Mammalian Homologous Repair with Distinct Mutagenic Consequences. en. *Molecular and Cellular Biology* **24,** 9305–9316. doi:10.1128/MCB.24.21.9305-9316.2004 (Nov. 2004).

96. Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. en. *DNA Repair* **7,** 1765–1771. doi:10.1016/j.dnarep.2008.06.018 (Oct. 2008).

97. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. en. *Nature Reviews Molecular Cell Biology* **18,** 495–506. doi:10.1038/nrm.2017.48 (Aug. 2017).

98. Mjelle, R. *et al.* Cell cycle regulation of human DNA repair and chromatin remodeling genes. en. *DNA Repair* **30,** 53–67. doi:10.1016/j.dnarep.2015.03.007 (June 2015).

99. Liang, F., Han, M., Romanienko, P. J. & Jasin, M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. en. *Proceedings of the National Academy of Sciences* **95,** 5172–5177. doi:10.1073/pnas.95.9.5172 (Apr. 1998).

100. Ira, G. *et al.* DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. en. *Nature* **431,** 1011–1017. doi:10.1038/nature02964 (Oct. 2004).

101. Ernst, M. P. *et al.* Ready for Repair? Gene Editing Enters the Clinic for the Treatment of Human Disease. en. *Molecular Therapy - Methods & Clinical Development* **18,** 532–557. doi:10.1016/j.omtm.2020.06.022 (Sept. 2020).

102. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. en. *Nature* **563,** 646–651. doi:10.1038/s41586-018-0686-x (Nov. 2018).

103. Bae, S., Kweon, J., Kim, H. S. & Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. en. *Nature Methods* **11,** 705–706. doi:10.1038/nmeth.3015 (July 2014).

104. Ata, H. *et al.* Robust activation of microhomology-mediated end joining for precision gene editing applications. en. *PLOS Genetics* **14** (ed Burgess, S. M.) e1007652. doi:10.1371/journal.pgen.1007652 (Sept. 2018).

105. Iyer, S. *et al.* Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. en. *Nature* **568,** 561–565. doi:10.1038/s41586-019-1076-8 (Apr. 2019).

106. Grajcarek, J. *et al.* Genome-wide microhomologies enable precise template-free editing of biologically relevant deletion mutations. en. *Nature Communications* **10,** 4856. doi:10.1038/s41467-019-12829-8 (Dec. 2019).

107. Truong, L. N. *et al.* Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. en. *Proceedings of the National Academy of Sciences* **110,** 7720–7725. doi:10.1073/pnas.1213431110 (May 2013).

108. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. eng. *Nature Methods* **11,** 783–784. doi:10.1038/nmeth.3047 (Aug. 2014).

109. Smurnyy, Y. *et al.* DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. en. *Nature Chemical Biology* **10,** 623–625. doi:10.1038/nchembio.1550 (Aug. 2014).

110. Hockemeyer, D. *et al.* A Drug-Inducible System for Direct Reprogramming of Human Somatic Cells to Pluripotency. en. *Cell Stem Cell* **3,** 346–353. doi:10.1016/j.stem.2008.08.014 (Sept. 2008).

111. Zou, J. *et al.* Gene Targeting of a Disease-Related Gene in Human Induced Pluripotent Stem and Embryonic Stem Cells. en. *Cell Stem Cell* **5,** 97–110. doi:10.1016/j.stem.2009.05.023 (July 2009).

112. Hotta, A. & Yamanaka, S. From Genomics to Gene Therapy: Induced Pluripotent Stem Cells Meet Genome Editing. en. *Annual Review of Genetics* **49,** 47–70. doi:10.1146/annurev-genet-112414-054926 (Nov. 2015).

113. Kim, D. & Kim, J.-S. DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. en. *Genome Research* **28,** 1894–1900. doi:10.1101/gr.236620.118 (Dec. 2018).

114. Tsai, S. Q. *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. en. *Nature Methods* **14,** 607–614. doi:10.1038/nmeth.4278 (June 2017).

115. Lazzarotto, C. R. *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR—Cas9 genome-wide activity. en. *Nature Biotechnology.* doi:10.1038/s41587-020-0555-7 (June 2020).

116. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. en. *Nature Biotechnology* **33,** 187–197. doi:10.1038/nbt.3117 (Feb. 2015).

117. Wienert, B. *et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq* en. preprint (Molecular Biology, Nov. 2018). doi:10.1101/469635.

118. Lombardo, A. *et al.* Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. en. *Nature Biotechnology* **25,** 1298–1306. doi:10.1038/nbt1353 (Nov. 2007).

119. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. en. *Nature Biotechnology* **32,** 677–683. doi:10.1038/nbt.2916 (July 2014).

120. Veres, A. *et al.* Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell* **15,** 27–30. doi:10.1016/j.stem.2014.04.020 (July 2014).

121. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. en. *Nature Reviews Genetics* **19,** 770–788. doi:10.1038/s41576-018-0059-1 (Dec. 2018).

122. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. en. *Nature* **576,** 149–157. doi:10.1038/s41586-019-1711-4 (Dec. 2019).

123. Cebrian-Serrano, A. & Davies, B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. en. *Mammalian Genome* **28,** 247–261. doi:10.1007/s00335-017-9697-4 (Aug. 2017).

124. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. en. *Nature* **523,** 481–485. doi:10.1038/nature14592 (July 2015).

125. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. en. *Nature* **556,** 57–63. doi:10.1038/nature26155 (Apr. 2018).

126. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. en. *Nature Biotechnology* **33,** 538–542. doi:10.1038/nbt.3190 (May 2015).

127. Chu, V. T. *et al.* Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. en. *Nature Biotechnology* **33,** 543–548. doi:10.1038/nbt.3198 (May 2015).

128. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. en. *eLife* **3,** e04766. doi:10.7554/eLife.04766 (Dec. 2014).

129. Uusi-Mäkelä, M. I. E. *et al.* Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). en. *PLOS ONE* **13** (ed Riley, B. B.) e0196238. doi:10.1371/journal.pone.0196238 (Apr. 2018).

130. Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. en. *Nucleic Acids Research* **44,** 6482–6492. doi:10.1093/nar/gkw524 (July 2016).

131. Chakrabarti, A. M. *et al.* Target-Specific Precision of CRISPR-Mediated Genome Editing. en. *Molecular Cell* **73,** 699–713.e6. doi:10.1016/j.molcel.2018.11.031 (Feb. 2019).

132. Chen, G. *et al.* A biodegradable nanocapsule delivers a Cas9 ribonucleoprotein complex for in vivo genome editing. en. *Nature Nanotechnology* **14,** 974–980. doi:10.1038/s41565-019-0539-2 (Oct. 2019).

133. Liu, B. *et al.* Inhibition of histone deacetylase 1 (HDAC1) and HDAC2 enhances CRISPR/Cas9 genome editing. en. *Nucleic Acids Research* **48,** 517–532. doi:10.1093/nar/gkz1136 (Jan. 2020).

134. Janssen, J. M., Chen, X., Liu, J. & Gonçalves, M. A. The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-Editing Events. en. *Molecular Therapy - Nucleic Acids* **16,** 141–154. doi:10.1016/j.omtn.2019.02.009 (June 2019).

135. Fujita, T., Yuno, M. & Fujii, H. Allele-specific locus binding and genome editing by CRISPR at the p16INK4a locus. en. *Scientific Reports* **6,** 30485. doi:10.1038/srep30485 (Aug. 2016).

136. Daer, R. M., Cutts, J. P., Brafman, D. A. & Haynes, K. A. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. en. *ACS Synthetic Biology* **6,** 428–438. doi:10.1021/acssynbio.5b00299 (Mar. 2017).

137. Kallimasioti-Pazi, E. M. *et al.* Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. en. *PLOS Biology* **16** (ed Misteli, T.) e2005595. doi:10.1371/journal.pbio.2005595 (Dec. 2018).

138. Horlbeck, M. A. *et al.* Nucleosomes impede Cas9 access to DNA in vivo and in vitro. en. *eLife* **5,** e12677. doi:10.7554/eLife.12677 (Mar. 2016).

139. Chen, Y. *et al.* Using local chromatin structure to improve CRISPR/Cas9 efficiency in zebrafish. en. *PLOS ONE* **12** (ed Gong, Z.) e0182528. doi:10.1371/journal.pone.0182528 (Aug. 2017).

140. Isaac, R. S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. en. *eLife* **5,** e13450. doi:10.7554/eLife.13450 (Apr. 2016).

141. Yarrington, R. M., Verma, S., Schwartz, S., Trautman, J. K. & Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. en. *Proceedings of the National Academy of Sciences* **115,** 9351–9358. doi:10.1073/pnas.1810062115 (Sept. 2018).

142. Clarke, R. *et al.* Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. en. *Molecular Cell* **71,** 42–55.e8. doi:10.1016/j.molcel.2018.06.005 (July 2018).

143. Kalousi, A. & Soutoglou, E. Nuclear compartmentalization of DNA repair. en. *Current Opinion in Genetics & Development* **37,** 148–157. doi:10.1016/j.gde.2016.05.013 (Apr. 2016).

144. Clouaire, T. & Legube, G. DNA double strand break repair pathway choice: a chromatin based decision? en. *Nucleus* **6,** 107–113. doi:10.1080/19491034.2015.1010946 (Mar. 2015).

145. Jeggo, P. A. & Downs, J. A. Roles of chromatin remodellers in DNA double strand break repair. en. *Experimental Cell Research* **329,** 69–77. doi:10.1016/j.yexcr.2014.09.023 (Nov. 2014).

146. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. en. *Nature Reviews Molecular Cell Biology* **20,** 698–714. doi:10.1038/s41580-019-0152-0 (Nov. 2019).

147. Aymard, F. *et al.* Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. en. *Nature Structural & Molecular Biology* **21,** 366–374. doi:10.1038/nsmb.2796 (Apr. 2014).

148. Clouaire, T. *et al.* Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. en. *Molecular Cell* **72,** 250–262.e6. doi:10.1016/j.molcel.2018.08.020 (Oct. 2018).

149. Schep, R. *et al.* Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. en. *Molecular Cell* **81,** 2216–2230.e10. doi:10.1016/j.molcel.2021.03.032 (May 2021).

150. Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin: Second in review series on chromatin dynamics. en. *EMBO reports* **3,** 224–229. doi:10.1093/embo-reports/kvf053 (Mar. 2002).

151. Grunstein, M. Histone acetylation in chromatin structure and transcription. en. *Nature* **389,** 349–352. doi:10.1038/38664 (Sept. 1997).

152. Shin, H. R. *et al.* Small-molecule inhibitors of histone deacetylase improve CRISPR-based adenine base editing. en. *Nucleic Acids Research* **49,** 2390–2399. doi:10.1093/nar/gkab052 (Feb. 2021).

153. Park, H., Shin, J., Choi, H., Cho, B. & Kim, J. Valproic Acid Significantly Improves CRISPR/Cas9-Mediated Gene Editing. en. *Cells* **9,** 1447. doi:10.3390/cells9061447 (June 2020).

154. Zhang, J.-P. *et al.* HDAC inhibitors improve CRISPR-mediated HDR editing efficiency in iPSCs. en. *Science China Life Sciences*. doi:10.1007/s11427-020-1855-4 (Jan. 2021).

155. Singh, R., Kuscu, C., Quinlan, A., Qi, Y. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. en. *Nucleic Acids Research* **43,** e118–e118. doi:10.1093/nar/gkv575 (Oct. 2015).

156. Nishizawa, M. *et al.* Epigenetic Variation between Human Induced Pluripotent Stem Cell Lines Is an Indicator of Differentiation Capacity. en. *Cell Stem Cell* **19,** 341–354. doi:10.1016/j.stem.2016.06.019 (Sept. 2016).

157. Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. en. *Nature Biotechnology* **28,** 521–526. doi:10.1038/nbt.1632 (May 2010).

158. Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M. & Nolan, G. P. A Continuous Molecular Roadmap to iPSC Reprogramming through Progression Analysis of Single-Cell Mass Cytometry. en. *Cell Stem Cell* **16,** 323–337. doi:10.1016/j.stem.2015.01.015 (Mar. 2015).

159. Joutsijoki, H., Haponen, M., Rasku, J., Aalto-Setälä, K. & Juhola, M. Machine Learning Approach to Automated Quality Identification of Human Induced Pluripotent Stem Cell Colony Images. en. *Computational and Mathematical Methods in Medicine* **2016,** 1–15. doi:10.1155/2016/3091039 (2016).

160. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. en. *Nature Biotechnology* **27,** 1033–1037. doi:10.1038/nbt.1580 (Nov. 2009).

161. Wakao, S. *et al.* Morphologic and Gene Expression Criteria for Identifying Human Induced Pluripotent Stem Cells. en. *PLoS ONE* **7** (ed Pera, M.) e48677. doi:10.1371/journal.pone.0048677 (Dec. 2012).

162. Pfannkuche, K., Fatima, A., Gupta, M. K., Dieterich, R. & Hescheler, J. Initial Colony Morphology-Based Selection for iPS Cells Derived from Adult Fibroblasts Is Substantially Improved by Temporary UTF1-Based Selection. en. *PLoS ONE* **5** (ed Callaerts, P.) e9580. doi:10.1371/journal.pone.0009580 (Mar. 2010).

163. Tokunaga, K. *et al.* Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. en. *Scientific Reports* **4,** 6996. doi:10.1038/srep06996 (May 2015).

164. Kato, R. *et al.* Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. en. *Scientific Reports* **6,** 34009. doi:10.1038/srep34009 (Sept. 2016).

165. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. en. *Nature Genetics* **38,** 431–440. doi:10.1038/ng1760 (Apr. 2006).

166. Silva, M. *et al.* Generating iPSCs: Translating Cell Reprogramming Science into Scalable and Robust Biomanufacturing Strategies. en. *Cell Stem Cell* **16,** 13–17. doi:10.1016/j.stem.2014.12.013 (Jan. 2015).

167. Schopperle, W. M. & DeWolf, W. C. The TRA-1-60 and TRA-1-81 Human Pluripotent Stem Cell Markers Are Expressed on Podocalyxin in Embryonal Carcinoma: THE TRA-1-60 AND TRA-1-81 HUMAN PLURIPOTENT STEM CELL MARKERS ARE EXPRESSED ON PODOCALYXIN IN EMBRYONAL CARCINOMA. en. *STEM CELLS* **25,** 723–730. doi:10.1634/stemcells.2005-0597 (Mar. 2007).

168. Nelakanti, R. V., Kooreman, N. G. & Wu, J. C. Teratoma Formation: A Tool for Monitoring Pluripotency in Stem Cell Research. en. *Current Protocols in Stem Cell Biology* **32.** doi:10.1002/9780470151808.sc04a08s32 (Feb. 2015).

169. Sheridan, S. D., Surampudi, V. & Rao, R. R. Analysis of Embryoid Bodies Derived from Human Induced Pluripotent Stem Cells as a Means to Assess Pluripotency. en. *Stem Cells International* **2012,** 1–9. doi:10.1155/2012/738910 (2012).

170. Müller, F.-J. *et al.* A bioinformatic assay for pluripotency in human cells. en. *Nature Methods* **8,** 315–317. doi:10.1038/nmeth.1580 (Apr. 2011).

171. Tsankov, A. M. *et al.* A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. en. *Nature Biotechnology* **33,** 1182–1192. doi:10.1038/nbt.3387 (Nov. 2015).

172. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. en. *Nature Reviews Molecular Cell Biology* **14,** 357–368. doi:10.1038/nrm3584 (June 2013).

173. Paull, D. *et al.* Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. en. *Nature Methods* **12,** 885–892. doi:10.1038/nmeth.3507 (Sept. 2015).

174. Hansson, J. *et al.* Highly Coordinated Proteome Dynamics during Reprogramming of Somatic Cells to Pluripotency. en. *Cell Reports* **2,** 1579–1592. doi:10.1016/j.celrep.2012.10.014 (Dec. 2012).

175. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. en. *Nature* **454,** 49–55. doi:10.1038/nature07056 (July 2008).

176. Toh, K. C., Ramdas, N. M. & Shivashankar, G. V. Actin cytoskeleton differentially alters the dynamics of lamin A, HP1 and H2B core histone proteins to remodel chromatin condensation state in living cells. en. *Integrative Biology* **7,** 1309–1317. doi:10.1039/C5IB00027K (2015).

177. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. en. *Nature* **453,** 948–951. doi:10.1038/nature06947 (June 2008).

178. Kind, J. *et al.* Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. en. *Cell* **163,** 134–147. doi:10.1016/j.cell.2015.08.040 (Sept. 2015).

179. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. en. *Molecular Cell* **38,** 603–613. doi:10.1016/j.molcel.2010.03.016 (May 2010).

180. Sakurai, K. *et al.* Kinome-wide Functional Analysis Highlights the Role of Cytoskeletal Remodeling in Somatic Cell Reprogramming. en. *Cell Stem Cell* **14,** 523–534. doi:10.1016/j.stem.2014.03.001 (Apr. 2014).

181. Li, R. *et al.* A Mesenchymal-to-Epithelial Transition Initiates and Is Required for the Nuclear Reprogramming of Mouse Fibroblasts. en. *Cell Stem Cell* **7,** 51–63. doi:10.1016/j.stem.2010.04.014 (July 2010).

182. Liao, B. *et al.* MicroRNA Cluster 302–367 Enhances Somatic Cell Reprogramming by Accelerating a Mesenchymal-to-Epithelial Transition. en. *Journal of Biological Chemistry* **286,** 17359–17364. doi:10.1074/jbc.C111.235960 (May 2011).

183. Samavarchi-Tehrani, P. *et al.* Functional Genomics Reveals a BMP-Driven Mesenchymal-to-Epithelial Transition in the Initiation of Somatic Cell Reprogramming. en. *Cell Stem Cell* **7,** 64–77. doi:10.1016/j.stem.2010.04.015 (July 2010).

184. Gingold, J. A. *et al.* A Genome-wide RNAi Screen Identifies Opposing Functions of Snai1 and Snai2 on the Nanog Dependency in Reprogramming. en. *Molecular Cell* **56,** 140–152. doi:10.1016/j.molcel.2014.08.014 (Oct. 2014).

185. Zhang, X.-B. Cellular Reprogramming of Human Peripheral Blood Cells. en. *Genomics, Proteomics & Bioinformatics* **11,** 264–274. doi:10.1016/j.gpb.2013.09.001 (Oct. 2013).

186. Chou, B.-K. *et al.* Efficient human iPS cell derivation by a non-integrating plasmid from blood cells with unique epigenetic and gene expression signatures. en. *Cell Research* **21,** 518–529. doi:10.1038/cr.2011.12 (Mar. 2011).

187. Kim, D. H. *et al.* Single-Cell Transcriptome Analysis Reveals Dynamic Changes in lncRNA Expression during Reprogramming. en. *Cell Stem Cell* **16,** 88–101. doi:10.1016/j.stem.2014.11.005 (Jan. 2015).

188. Lujan, E. *et al.* Early reprogramming regulators identified by prospective isolation and mass cytometry. en. *Nature* **521,** 352–356. doi:10.1038/nature14274 (May 2015).

189. Nguyen, Q. H. *et al.* Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. en. *Genome Research* **28,** 1053–1066. doi:10.1101/gr.223925.117 (July 2018).

190. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. en. *Cell* **176,** 928–943.e22. doi:10.1016/j.cell.2019.01.006 (Feb. 2019).

191. Harkness, T. *et al.* High-content imaging with micropatterned multiwell plates reveals influence of cell geometry and cytoskeleton on chromatin dynamics. en. *Biotechnology Journal* **10,** 1555–1567. doi:10.1002/biot.201400756 (Oct. 2015).

192. Howden, S. E. *et al.* Simultaneous Reprogramming and Gene Correction of Patient Fibroblasts. en. *Stem Cell Reports* **5,** 1109–1118. doi:10.1016/j.stemcr.2015.10.009 (Dec. 2015).

193. Carlson-Stevermer, J. *et al.* High-Content Analysis of CRISPR-Cas9 Gene-Edited Human Embryonic Stem Cells. en. *Stem Cell Reports* **6,** 109–120. doi:10.1016/j.stemcr.2015.11.014 (Jan. 2016).

194. Sun, N. *et al.* Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. en. *Proceedings of the National Academy of Sciences* **106,** 15720–15725. doi:10.1073/pnas.0908450106 (Sept. 2009).

195. Quintanilla, R. H., Asprer, J. S. T., Vaz, C., Tanavde, V. & Lakshmipathy, U. CD44 Is a Negative Cell Surface Marker for Pluripotent Stem Cell Identification during

Human Fibroblast Reprogramming. en. *PLoS ONE* **9** (ed Sampaolesi, M.) e85419. doi:10.1371/journal.pone.0085419 (Jan. 2014).

196. Silva, J. *et al.* Nanog Is the Gateway to the Pluripotent Ground State. en. *Cell* **138,** 722–737. doi:10.1016/j.cell.2009.07.039 (Aug. 2009).

197. Jain, N., Iyer, K. V., Kumar, A. & Shivashankar, G. V. Cell geometric constraints induce modular gene-expression patterns via redistribution of HDAC3 regulated by actomyosin contractility. en. *Proceedings of the National Academy of Sciences* **110,** 11349–11354. doi:10.1073/pnas.1300801110 (July 2013).

198. Carpenter, A. E. *et al.* [No title found]. en. *Genome Biology* **7,** R100. doi:10.1186/gb-2006-7-10-r100 (2006).

199. Wold, S., Esbensen, K. & Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **2.** Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists, 37–52. doi:https://doi.org/10.1016/0169-7439(87)80084-9 (1987).

200. Van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9,** 2579–2605 (2008).

201. Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. en. *Proceedings of the National Academy of Sciences* **108,** 8257–8262. doi:10.1073/pnas.1017017108 (May 2011).

202. Tanabe, K., Nakamura, M., Narita, M., Takahashi, K. & Yamanaka, S. Maturation, not initiation, is the major roadblock during reprogramming toward pluripotency from human fibroblasts. en. *Proceedings of the National Academy of Sciences* **110,** 12172–12179. doi:10.1073/pnas.1310291110 (July 2013).

203. Singh, A. M. & Dalton, S. The Cell Cycle and Myc Intersect with Mechanisms that Regulate Pluripotency and Reprogramming. en. *Cell Stem Cell* **5,** 141–149. doi:10.1016/j.stem.2009.07.003 (Aug. 2009).

204. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. en. *Nature* **564,** 219–224. doi:10.1038/s41586-018-0744-4 (Dec. 2018).

205. Zhao, T. *et al.* Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. en. *Cell Stem Cell* **23,** 31–45.e7. doi:10.1016/j.stem.2018.05.025 (July 2018).

206. Liu, E., Gordonov, S., Treiser, M. D. & Moghe, P. V. Parsing the early cytoskeletal and nuclear organizational cues that demarcate stem cell lineages. en. *Cell Cycle* **9,** 2108–2117. doi:10.4161/cc.9.11.11864 (June 2010).

207. Bauwens, C. L. *et al.* Control of Human Embryonic Stem Cell Colony and Aggregate Size Heterogeneity Influences Differentiation Trajectories. en. *Stem Cells* **26,** 2300–2310. doi:10.1634/stemcells.2008-0183 (Sept. 2008).

208. Peerani, R. *et al.* Niche-mediated control of human embryonic stem cell self-renewal and differentiation. en. *The EMBO Journal* **26,** 4744–4755. doi:10.1038/sj.emboj.7601896 (Nov. 2007).

209. Caiazzo, M. *et al.* Defined three-dimensional microenvironments boost induction of pluripotency. en. *Nature Materials* **15,** 344–352. doi:10.1038/nmat4536 (Mar. 2016).

210. Roberts, B. *et al.* Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. en. *Molecular Biology of the Cell* **28** (ed Drubin, D. G.) 2854–2874. doi:10.1091/mbc.e17-03-0209 (Oct. 2017).

211. Haupt, A. *et al.* Endogenous Protein Tagging in Human Induced Pluripotent Stem Cells Using CRISPR/Cas9. en. *Journal of Visualized Experiments,* 58130. doi:10.3791/58130 (Aug. 2018).

212. Walsh, A. J. *et al.* Optical Metabolic Imaging Identifies Glycolytic Levels, Subtypes, and Early-Treatment Response in Breast Cancer. en. *Cancer Research* **73,** 6164–6174. doi:10.1158/0008-5472.CAN-13-0527 (Oct. 2013).

213. Dong, B. *et al.* Superresolution intrinsic fluorescence imaging of chromatin utilizing native, unmodified nucleic acids for contrast. en. *Proceedings of the National Academy of Sciences* **113,** 9716–9721. doi:10.1073/pnas.1602202113 (Aug. 2016).

214. Bickmore, W. A. & van Steensel, B. Genome Architecture: Domain Organization of Interphase Chromosomes. en. *Cell* **152,** 1270–1284. doi:10.1016/j.cell.2013.02.001 (Mar. 2013).

215. Treiser, M. D. *et al.* Cytoskeleton-based forecasting of stem cell lineage fates. en. *Proceedings of the National Academy of Sciences* **107,** 610–615. doi:10.1073/pnas.0909597107 (Jan. 2010).

216. Kim, J. J. *et al.* Optical High Content Nanoscopy of Epigenetic Marks Decodes Phenotypic Divergence in Stem Cells. en. *Scientific Reports* **7,** 39406. doi:10.1038/srep39406 (Apr. 2017).

217. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. en. *Nature Biotechnology* **29,** 928–933. doi:10.1038/nbt.1977 (Oct. 2011).

218. Schepers, K. *et al.* Dissecting T cell lineage relationships by cellular barcoding. en. *Journal of Experimental Medicine* **205,** 2309–2318. doi:10.1084/jem.20072462 (Sept. 2008).

219. Wong, E. T. Reproducible doxycycline-inducible transgene expression at specific loci generated by Cre-recombinase mediated cassette exchange. en. *Nucleic Acids Research* **33,** e147–e147. doi:10.1093/nar/gni145 (Sept. 2005).

220. Muthusamy, T., Mukherjee, O., Menon, R., Prakash Bangalore, M. & Panicker, M. M. A Method to Identify and Isolate Pluripotent Human Stem Cells and Mouse Epiblast Stem Cells Using Lipid Body-Associated Retinyl Ester Fluorescence. en. *Stem Cell Reports* **3,** 169–184. doi:10.1016/j.stemcr.2014.05.004 (July 2014).

221. Hirata, N. *et al.* A Chemical Probe that Labels Human Pluripotent Stem Cells. en. *Cell Reports* **6,** 1165–1174. doi:10.1016/j.celrep.2014.02.006 (Mar. 2014).

222. Luni, C. *et al.* High-efficiency cellular reprogramming with microfluidics. en. *Nature Methods* **13,** 446–452. doi:10.1038/nmeth.3832 (May 2016).

223. Singh, A. *et al.* Adhesion strength–based, label-free isolation of human pluripotent stem cells. en. *Nature Methods* **10,** 438–444. doi:10.1038/nmeth.2437 (May 2013).

224. Sha, J., Lippmann, E. S., McNulty, J., Ma, Y. & Ashton, R. S. Sequential Nucleophilic Substitutions Permit Orthogonal Click Functionalization of Multicomponent PEG Brushes. en. *Biomacromolecules* **14,** 3294–3303. doi:10.1021/bm400900r (Sept. 2013).

225. Wojcik, K. & Dobrucki, J. W. Interaction of a DNA intercalator DRAQ5, and a minor groove binder SYTO17, with chromatin in live cells-Influence on chromatin organization and histone-DNA interactions. en. *Cytometry Part A* **73A,** 555–562. doi:10.1002/cyto.a.20573 (June 2008).

226. Martin, R. M., Leonhardt, H. & Cardoso, M. C. DNA labeling in living cells. en. *Cytometry Part A* **67A,** 45–52. doi:10.1002/cyto.a.20172 (Sept. 2005).

227. Andrews, P. W. *et al.* Harmonizing standards for producing clinical-grade therapies from pluripotent stem cells. en. *Nature Biotechnology* **32,** 724–726. doi:10.1038/nbt.2973 (Aug. 2014).

228. Inoue, H., Nagata, N., Kurokawa, H. & Yamanaka, S. iPS cells: a game changer for future medicine. en. *The EMBO Journal* **33,** 409–417. doi:10.1002/embj.201387098 (Mar. 2014).

229. Yaffe, M. P., Noggle, S. A. & Solomon, S. L. Raising the standards of stem cell line quality. en. *Nature Cell Biology* **18,** 236–237. doi:10.1038/ncb3313 (Mar. 2016).

230. Kreiner, T. & Irion, S. Whole-Genome Analysis, Stem Cell Research, and the Future of Biobanks. en. *Cell Stem Cell* **12,** 513–516. doi:10.1016/j.stem.2013.04.024 (May 2013).

231. McNulty, J. D. *et al.* High-precision robotic microcontact printing (R-CP) utilizing a vision guided selectively compliant articulated robotic arm. en. *Lab on a Chip* **14,** 1923. doi:10.1039/c3lc51137e (2014).

232. *The AGT cytogenetics laboratory manual* Fourth edition. en (eds Arsham, M. S., Barch, M. J., Lawce, H. J. & Technologists, A. o. G.) (Wiley, Blackwell, Hoboken, New Jersey, 2017).

233. International Standing Committee on Human Cytogenomic Nomenclature., Shaffer, L. G., Schmid, M. & McGowan-Jordan, J. *ISCN : an international system for human cytogenomic nomenclature (2016)* English (Karger, Basel; Paris, 2016).

234. Riccardi, C. & Nicoletti, I. Analysis of apoptosis by propidium iodide staining and flow cytometry. en. *Nature Protocols* **1,** 1458–1461. doi:10.1038/nprot.2006.238 (Aug. 2006).

235. Mandai, M. *et al.* Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration. en. *New England Journal of Medicine* **376,** 1038–1046. doi:10.1056/NEJMoa1608368 (Mar. 2017).

236. Takagi, S. *et al.* Evaluation of Transplanted Autologous Induced Pluripotent Stem Cell-Derived Retinal Pigment Epithelium in Exudative Age-Related Macular Degeneration. en. *Ophthalmology Retina* **3,** 850–859. doi:10.1016/j.oret.2019.04.021 (Oct. 2019).

237. Schweitzer, J. S. *et al.* Personalized iPSC-Derived Dopamine Progenitor Cells for Parkinson's Disease. en. *New England Journal of Medicine* **382,** 1926–1932. doi:10.1056/NEJMoa1915872 (May 2020).

238. Yamanaka, S. Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. en. *Cell Stem Cell* **27,** 523–531. doi:10.1016/j.stem.2020.09.014 (Oct. 2020).

239. Madrid, M., Sumen, C., Aivio, S. & Saklayen, N. Autologous Induced Pluripotent Stem Cell–Based Cell Therapies: Promise, Progress, and Challenges. en. *Current Protocols* **1.** doi:10.1002/cpz1.88 (Mar. 2021).

240. Silva, M. *et al.* Generating iPSCs: Translating Cell Reprogramming Science into Scalable and Robust Biomanufacturing Strategies. en. *Cell Stem Cell* **16,** 13–17. doi:10.1016/j.stem.2014.12.013 (Jan. 2015).

241. Huang, C.-Y. *et al.* Human iPSC banking: barriers and opportunities. en. *Journal of Biomedical Science* **26,** 87. doi:10.1186/s12929-019-0578-x (Dec. 2019).

242. Jo, H.-Y. *et al.* Development of genetic quality tests for good manufacturing practice-compliant induced pluripotent stem cells and their derivatives. en. *Scientific Reports* **10,** 3939. doi:10.1038/s41598-020-60466-9 (Dec. 2020).

243. Sullivan, S. *et al.* Quality control guidelines for clinical-grade human induced pluripotent stem cell lines. en. *Regenerative Medicine* **13,** 859–866. doi:10.2217/rme-2018-0095 (Oct. 2018).

244. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. en. *Nature Methods* **15,** 917–920. doi:10.1038/s41592-018-0111-2 (Nov. 2018).

245. Kandel, M. E. *et al.* Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments. en. *Nature Communications* **11,** 6256. doi:10.1038/s41467-020-20062-x (Dec. 2020).

246. Nguyen, T. C. *et al.* Virtual organelle self-coding for fluorescence imaging via adversarial learning. en. *Journal of Biomedical Optics* **25.** doi:10.1117/1.JBO.25.9.096009 (Sept. 2020).

247. Wieslander, H., Gupta, A., Bergman, E., Hallström, E. & Harrison, P. J. *Learning to see colours: generating biologically relevant fluorescent labels from bright-field images* en. preprint (Bioinformatics, Jan. 2021). doi:10.1101/2021.01.18.427121.

248. Christiansen, E. M. *et al.* In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. en. *Cell* **173,** 792–803.e19. doi:10.1016/j.cell.2018.03.040 (Apr. 2018).

249. Molugu, K. *et al.* Tracking and Predicting Human Somatic Cell Reprogramming Using Nuclear Characteristics. en. *Biophysical Journal* **118,** 2086–2102. doi:10.1016/j.bpj.2019.10.014 (May 2020).

250. Skala, M. C. *et al.* In vivo multiphoton fluorescence lifetime imaging of protein-bound and free nicotinamide adenine dinucleotide in normal and precancerous epithelia. en. *Journal of Biomedical Optics* **12,** 024014. doi:10.1117/1.2717503 (2007).

251. Huang, S., Heikal, A. A. & Webb, W. W. Two-Photon Fluorescence Spectroscopy and Microscopy of NAD(P)H and Flavoprotein. en. *Biophysical Journal* **82,** 2811–2825. doi:10.1016/S0006-3495(02)75621-X (May 2002).

252. Chance, B., Schoener, B., Oshino, R., Itshak, F. & Nakase, Y. Oxidation-reduction ratio studies of mitochondria in freeze-trapped samples. NADH and flavoprotein fluorescence signals. en. *Journal of Biological Chemistry* **254,** 4764–4771. doi:10.1016/S0021-9258(17)30079-0 (June 1979).

253. Georgakoudi, I. & Quinn, K. P. Optical Imaging Using Endogenous Contrast to Assess Metabolic State. en. *Annual Review of Biomedical Engineering* **14,** 351–367. doi:10.1146/annurev-bioeng-071811-150108 (Aug. 2012).

254. Nakashima, N., Yoshihara, K., Tanaka, F. & Yagi, K. Picosecond fluorescence lifetime of the coenzyme of D-amino acid oxidase. en. *Journal of Biological Chemistry* **255,** 5261–5263. doi:10.1016/S0021-9258(19)70779-0 (June 1980).

255. Lakowicz, J. R. Fluorescence lifetime imaging of free and.protein-bound NADH Proc. Nati. Acad. Sci. USA Vol. 89, pp. 1271-1275, February 1992 Biochemistry. en, 5.

256. Datta, R., Heaster, T. M., Sharick, J. T., Gillette, A. A. & Skala, M. C. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. en. *Journal of Biomedical Optics* **25,** 1. doi:10.1117/1.JBO.25.7.071203 (May 2020).

257. Pouli, D. *et al.* Label-free, High-Resolution Optical Metabolic Imaging of Human Cervical Precancers Reveals Potential for Intraepithelial Neoplasia Diagnosis. en. *Cell Reports Medicine* **1,** 100017. doi:10.1016/j.xcrm.2020.100017 (May 2020).

258. Fink, J. *et al.* Comparative study and improvement of current cell micro-patterning techniques. en. *Lab Chip* **7,** 672–680. doi:10.1039/B618545B (2007).

259. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology* **32,** 381–386. doi:10.1038/nbt.2859 (Apr. 2014).

260. Liazoghli, D., Roth, A. D., Thostrup, P. & Colman, D. R. Substrate Micropatterning as a New in Vitro Cell Culture System to Study Myelination. en. *ACS Chemical Neuroscience* **3,** 90–95. doi:10.1021/cn2000734 (Feb. 2012).

261. Csucs, G., Michel, R., Lussi, J. W., Textor, M. & Danuser, G. Microcontact printing of novel co-polymers in combination with proteins for cell-biological applications. en. *Biomaterials* **24,** 1713–1720. doi:10.1016/S0142-9612(02)00568-9 (May 2003).

262. Dong, H. Y., Wilkes, S. & Yang, H. CD71 is Selectively and Ubiquitously Expressed at High Levels in Erythroid Precursors of All Maturation Stages: A Comparative Immunochemical Study With Glycophorin A and Hemoglobin A. en. *Am J Surg Pathol* **35,** 10 (2011).

263. Zhang, J., Nuebel, E., Daley, G. Q., Koehler, C. M. & Teitell, M. A. Metabolic Regulation in Pluripotent Stem Cells during Reprogramming and Self-Renewal. en. *Cell Stem Cell* **11,** 589–595. doi:10.1016/j.stem.2012.10.005 (Nov. 2012).

264. Zhang, J. *et al.* Metabolism in Pluripotent Stem Cells and Early Mammalian Development. en. *Cell Metabolism* **27,** 332–338. doi:10.1016/j.cmet.2018.01.008 (Feb. 2018).

265. Ryall, J. G., Cliff, T., Dalton, S. & Sartorelli, V. Metabolic Reprogramming of Stem Cell Epigenetics. en. *Cell Stem Cell* **17,** 651–662. doi:10.1016/j.stem.2015.11.012 (Dec. 2015).

266. Sun, H. *et al.* Metabolic switch and epithelial–mesenchymal transition cooperate to regulate pluripotency. en. *The EMBO Journal* **39.** doi:10.15252/embj.2019102961 (Apr. 2020).

267. Xu, X. *et al.* Mitochondrial Regulation in Pluripotent Stem Cells. en. *Cell Metabolism* **18,** 325–332. doi:10.1016/j.cmet.2013.06.005 (Sept. 2013).

268. Kolenc, O. I. & Quinn, K. P. Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD(P)H and FAD. en. *Antioxidants & Redox Signaling* **30,** 875–889. doi:10.1089/ars.2017.7451 (Feb. 2019).

269. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. en. *Nature Methods* **16,** 1226–1232. doi:10.1038/s41592-019-0582-9 (Dec. 2019).

270. Sullivan, W. J. *et al.* Extracellular Matrix Remodeling Regulates Glucose Metabolism through TXNIP Destabilization. en. *Cell* **175,** 117–132.e21. doi:10.1016/j.cell.2018.08.017 (Sept. 2018).

271. Hu, H. *et al.* Phosphoinositide 3-Kinase Regulates Glycolysis through Mobilization of Aldolase from the Actin Cytoskeleton. en. *Cell* **164,** 433–446. doi:10.1016/j.cell.2015.12.042 (Jan. 2016).

272. Bays, J. L., Campbell, H. K., Heidema, C., Sebbagh, M. & DeMali, K. A. Linking E-cadherin mechanotransduction to cell metabolism through force-mediated activation of AMPK. en. *Nature Cell Biology* **19,** 724–731. doi:10.1038/ncb3537 (June 2017).

273. Park, J. S. *et al.* Mechanical regulation of glycolysis via cytoskeleton architecture. en. *Nature* **578,** 621–626. doi:10.1038/s41586-020-1998-1 (Feb. 2020).

274. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. en. *Journal of Open Source Software* **3,** 861. doi:10.21105/joss.00861 (Sept. 2018).

275. Amancio, D. R. *et al.* A Systematic Comparison of Supervised Classifiers. en. *PLoS ONE* **9** (ed Shen, H.-B.) e94137. doi:10.1371/journal.pone.0094137 (Apr. 2014).

276. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. eng. *Nature Methods* **14,** 979–982. doi:10.1038/nmeth.4402 (Oct. 2017).

277. Moran, P. A. P. Notes on Continuous Stochastic Phenomena. en, 8 (2021).

278. Prieto, J. *et al.* Early ERK1/2 activation promotes DRP1-dependent mitochondrial fission necessary for cell reprogramming. en. *Nature Communications* **7,** 11124. doi:10.1038/ncomms11124 (Sept. 2016).

279. Folmes, C. D. *et al.* Somatic Oxidative Bioenergetics Transitions into Pluripotency-Dependent Glycolysis to Facilitate Nuclear Reprogramming. en. *Cell Metabolism* **14,** 264–271. doi:10.1016/j.cmet.2011.06.011 (Aug. 2011).

280. Cacchiarelli, D. *et al.* Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. eng. *Cell* **162,** 412–424. doi:10.1016/j.cell.2015.06.016 (July 2015).

281. Reid, M. A., Dai, Z. & Locasale, J. W. The impact of cellular metabolism on chromatin dynamics and epigenetics. en. *Nature Cell Biology* **19,** 1298–1306. doi:10.1038/ncb3629 (Nov. 2017).

282. Mueller, F., Mayhew, S. G. & Massey, V. Effect of temperature on the absorption spctra of free and protein-bound flavines. en. *Biochemistry* **12,** 4654–4662. doi:10.1021/bi00747a017 (Nov. 1973).

283. Sato, K., Nishina, Y., Shiga, K. & Tanaka, F. Hydrogen-bonding dynamics of free flavins in benzene and FAD in electron-transferring flavoprotein upon excitation. en. *Journal of Photochemistry and Photobiology B: Biology* **70,** 67–73. doi:10.1016/S1011-1344(03)00056-3 (May 2003).

284. Schneckenburger, H., Wagner, M., Weber, P., Strauss, W. S. & Sailer, R. Autofluorescence Lifetime Imaging of Cultivated Cells Using a UV Picosecond Laser Diode. en. *Journal of Fluorescence* **14,** 649–654. doi:10.1023/B:JOFL.0000039351.09916.cc (Sept. 2004).

285. Zbinden, A. *et al.* Fluorescence lifetime metabolic mapping of hypoxia-induced damage in pancreatic pseudo-islets. en. *Journal of Biophotonics* **13.** doi:10.1002/jbio.202000375 (Dec. 2020).

286. D'Arcangelo, E. & McGuigan, A. P. Micropatterning strategies to engineer controlled cell and tissue architecture in vitro. en. *BioTechniques* **58.** doi:10.2144/000114245 (Jan. 2015).

287. Mack, A. A., Kroboth, S., Rajesh, D. & Wang, W. B. Generation of Induced Pluripotent Stem Cells from CD34+ Cells across Blood Drawn from Multiple Donors with Non-Integrating Episomal Vectors. en. *PLoS ONE* **6** (ed Borlongan, C. V.) e27956. doi:10.1371/journal.pone.0027956 (Nov. 2011).

288. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. eng. *Nature Methods* **8,** 409–412. doi:10.1038/nmeth.1591 (May 2011).

289. Nanbo, A., Sugden, A. & Sugden, B. The coupling of synthesis and partitioning of EBV's plasmid replicon is revealed in live cells. en. *The EMBO Journal* **26,** 4252–4262. doi:10.1038/sj.emboj.7601853 (Oct. 2007).

290. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. en. *Nature Biotechnology* **33,** 58–63. doi:10.1038/nbt.3070 (Jan. 2015).

291. Suhr, S. T. *et al.* Mitochondrial Rejuvenation After Induced Pluripotency. en. *PLoS ONE* **5** (ed Klymkowsky, M.) e14095. doi:10.1371/journal.pone.0014095 (Nov. 2010).

292. Qian, T. *et al. Label-free imaging for quality control of cardiomyocyte differentiation* en. preprint (Bioengineering, June 2021). doi:10.1101/2021.06.01.446615.

293. Rodimova, S. A. *et al.* Metabolic activity and intracellular pH in induced pluripotent stem cells differentiating in dermal and epidermal directions. en. *Methods and Applications in Fluorescence* **7,** 044002. doi:10.1088/2050-6120/ab3b3d (Sept. 2019).

294. Meleshina, A. V. *et al.* Probing metabolic states of differentiating stem cells using two-photon FLIM. en. *Scientific Reports* **6,** 21853. doi:10.1038/srep21853 (Apr. 2016).

295. Hirano, K. & Namihira, M. FAD influx enhances neuronal differentiation of human neural stem cells by facilitating nuclear localization of LSD1. en. *FEBS Open Bio* **7,** 1932–1942. doi:10.1002/2211-5463.12331 (Dec. 2017).

296. Jiménez-Torres, J. A., Beebe, D. J. & Sung, K. E. en. in *The Tumor Microenvironment* (eds Ursini-Siegel, J. & Beauchemin, N.) 59–69 (Springer New York, New York, NY, 2016). doi:10.1007/978-1-4939-3801-8_5.

297. Salick, M. R. *et al.* Micropattern width dependent sarcomere development in human ESC-derived cardiomyocytes. en. *Biomaterials* **35,** 4454–4464. doi:10.1016/j.biomaterials.2014.02.001 (May 2014).

298. Holmes, G., Donkin, A. & Witten, I. *WEKA: a machine learning workbench* en. in *Proceedings of ANZIIS '94 - Australian New Zealnd Intelligent Information Systems Conference* (IEEE, Brisbane, Qld., Australia, 1994), 357–361. doi:10.1109/ANZIIS.1994.396988.

299. Shi, Y., Inoue, H., Wu, J. C. & Yamanaka, S. Induced pluripotent stem cell technology: a decade of progress. en. *Nature Reviews Drug Discovery* **16,** 115–130. doi:10.1038/nrd.2016.245 (Feb. 2017).

300. De Masi, C., Spitalieri, P., Murdocca, M., Novelli, G. & Sangiuolo, F. Application of CRISPR/Cas9 to human-induced pluripotent stem cells: from gene editing to drug discovery. en. *Human Genomics* **14,** 25. doi:10.1186/s40246-020-00276-2 (Dec. 2020).

301. Sinha, D. *et al.* Human iPSC Modeling Elucidates Mutation-Specific Responses to Gene Therapy in a Genotypically Diverse Dominant Maculopathy. en. *SSRN Electronic Journal.* doi:10.2139/ssrn.3377258 (2019).

302. Jacob, A. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. en. *Cell Stem Cell* **21,** 472–488.e10. doi:10.1016/j.stem.2017.08.014 (Oct. 2017).

303. Readhead, B. *et al.* Expression-based drug screening of neural progenitor cells from individuals with schizophrenia. en. *Nature Communications* **9,** 4412. doi:10.1038/s41467-018-06515-4 (Dec. 2018).

304. Novosadova, E. V. *et al.* The Use of Human Induced Pluripotent Stem Cells for Testing Neuroprotective Activity of Pharmacological Compounds. en. *Biochemistry (Moscow)* **84,** 1296–1305. doi:10.1134/S0006297919110075 (Nov. 2019).

305. Ruillier, V. *et al.* Rescuing compounds for Lesch-Nyhan disease identified using stem cell–based phenotypic screening. en. *JCI Insight* **5,** e132094. doi:10.1172/jci.insight.132094 (Feb. 2020).

306. Smith, C. *et al.* Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. en. *Cell Stem Cell* **15,** 12–13. doi:10.1016/j.stem.2014.06.011 (July 2014).

307. Foltz, L., Howden, S., Thomson, J. & Clegg, D. Functional Assessment of Patient-Derived Retinal Pigment Epithelial Cells Edited by CRISPR/Cas9. en. *International Journal of Molecular Sciences* **19,** 4127. doi:10.3390/ijms19124127 (Dec. 2018).

308. Schuster, S., Saravanakumar, S., Schöls, L. & Hauser, S. Generation of a homozygous CRISPR/Cas9-mediated knockout human iPSC line for the STUB1 locus. en. *Stem Cell Research* **34,** 101378. doi:10.1016/j.scr.2018.101378 (Jan. 2019).

309. Carlson-Stevermer, J. *et al.* Design of efficacious somatic cell genome editing strategies for recessive and polygenic diseases. en. *Nature Communications* **11,** 6277. doi:10.1038/s41467-020-20065-8 (Dec. 2020).

310. Sun, J. *et al.* CRISPR/Cas9 editing of APP C-terminus attenuates -cleavage and promotes -cleavage. en. *Nature Communications* **10,** 53. doi:10.1038/s41467-018-07971-8 (Dec. 2019).

311. Wang, S. *et al.* Rescue of premature aging defects in Cockayne syndrome stem cells by CRISPR/Cas9-mediated gene correction. en. *Protein & Cell* **11,** 1–22. doi:10.1007/s13238-019-0623-2 (Jan. 2020).

312. Jacków, J. *et al.* CRISPR/Cas9-based targeted genome editing for correction of recessive dystrophic epidermolysis bullosa using iPS cells. en. *Proceedings of the National Academy of Sciences* **116,** 26846–26852. doi:10.1073/pnas.1907081116 (Dec. 2019).

313. Li, H. L. *et al.* Precise Correction of the Dystrophin Gene in Duchenne Muscular Dystrophy Patient Induced Pluripotent Stem Cells by TALEN and CRISPR-Cas9. en. *Stem Cell Reports* **4,** 143–154. doi:10.1016/j.stemcr.2014.10.013 (Jan. 2015).

314. Xie, F. *et al.* Seamless gene correction of -thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and *piggyBac*. en. *Genome Research* **24,** 1526–1533. doi:10.1101/gr.173427.114 (Sept. 2014).

315. Liu, G.-H. *et al.* Targeted Gene Correction of Laminopathy-Associated LMNA Mutations in Patient-Specific iPSCs. en. *Cell Stem Cell* **8,** 688–694. doi:10.1016/j.stem.2011.04.019 (June 2011).

316. Zou, J., Mali, P., Huang, X., Dowey, S. N. & Cheng, L. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* **118 17,** 4599–608 (2011).

317. Hotta, A. & Yamanaka, S. From Genomics to Gene Therapy: Induced Pluripotent Stem Cells Meet Genome Editing. en. *Annual Review of Genetics* **49,** 47–70. doi:10.1146/annurev-genet-112414-054926 (Nov. 2015).

318. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. en. *Nature Biotechnology* **32,** 279–284. doi:10.1038/nbt.2808 (Mar. 2014).

319. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. en. *Cell* **154,** 1380–1389. doi:10.1016/j.cell.2013.08.021 (Sept. 2013).

320. Ryan, D. E. *et al.* Improving CRISPR–Cas specificity with chemical modifications in single-guide RNAs. en. *Nucleic Acids Research* **46,** 792–803. doi:10.1093/nar/gkx1199 (Jan. 2018).

321. Amrani, N. *et al.* NmeCas9 is an intrinsically high-fidelity genome-editing platform. en. *Genome Biology* **19,** 214. doi:10.1186/s13059-018-1591-1 (Dec. 2018).

322. Tóth, K. F. *et al.* Trichostatin A-induced histone acetylation causes decondensation of interphase chromatin. en. *Journal of Cell Science* **117,** 4277–4287. doi:10.1242/jcs.01293 (Aug. 2004).

323. Damodaran, K. *et al.* Compressive force induces reversible chromatin condensation and cell geometry–dependent transcriptional response. en. *Molecular Biology of the Cell* **29** (ed Weaver, V. M.) 3039–3051. doi:10.1091/mbc.E18-04-0256 (Dec. 2018).

324. Kurgan, G. *et al.* CRISPAltRations: A validated cloud-based approach for interrogation of double-strand break repair mediated by CRISPR genome editing. en. *Molecular*

Therapy - Methods & Clinical Development **21,** 478–491. doi:10.1016/j.omtm.2021.03.024 (June 2021).

325. Langley, S. A., Karpen, G. H. & Langley, C. H. Nucleosomes Shape DNA Polymorphism and Divergence. en. *PLoS Genetics* **10** (ed Pritchard, J. K.) e1004457. doi:10.1371/journal.pgen.1004457 (July 2014).

326. Liu, H., Zhai, J., Luo, K. & Liu, L. Chromatin structure is distinct between coding and non-coding single nucleotide polymorphisms. en. *BMC Molecular Biology* **15,** 22. doi:10.1186/1471-2199-15-22 (2014).

327. Holoch, D. & Moazed, D. RNA-mediated epigenetic regulation of gene expression. en. *Nature Reviews Genetics* **16,** 71–84. doi:10.1038/nrg3863 (Feb. 2015).

328. Pulecio, J., Verma, N., Mejía-Ramírez, E., Huangfu, D. & Raya, A. CRISPR/Cas9-Based Engineering of the Epigenome. en. *Cell Stem Cell* **21,** 431–447. doi:10.1016/j.stem.2017.09.006 (Oct. 2017).

329. Gelles, J. D. & Chipuk, J. E. Robust high-throughput kinetic analysis of apoptosis with real-time high-content live-cell imaging. en. *Cell Death & Disease* **7,** e2493–e2493. doi:10.1038/cddis.2016.332 (Dec. 2016).

330. Sharma, A., Singh, K. & Almasan, A. Histone H2AX phosphorylation: a marker for DNA damage. eng. *Methods in Molecular Biology (Clifton, N.J.)* **920,** 613–626. doi:10.1007/978-1-61779-998-3_40 (2012).

331. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. en. *Science* **368,** 290–296. doi:10.1126/science.aba8853 (Apr. 2020).

332. Nam, A. S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. en. *Nature* **571,** 355–360. doi:10.1038/s41586-019-1367-0 (July 2019).

333. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. en. *Nature Genetics* **53,** 403–411. doi:10.1038/s41588-021-00790-6 (Mar. 2021).

334. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. en. *Nature Methods* **14,** 417–419. doi:10.1038/nmeth.4197 (Apr. 2017).

335. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. en. *F1000Research* **4,** 1521. doi:10.12688/f1000research.7563.1 (Dec. 2015).

336. Jakobs, S. High resolution imaging of live mitochondria. en. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1763,** 561–575. doi:10.1016/j.bbamcr.2006.04.004 (May 2006).

337. Lukinavičius, G. *et al.* Fluorogenic probes for live-cell imaging of the cytoskeleton. en. *Nature Methods* **11,** 731–733. doi:10.1038/nmeth.2972 (July 2014).

338. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. en. *Nature Biotechnology* **37,** 547–554. doi:10.1038/s41587-019-0071-9 (May 2019).

339. Downing, T. L. *et al.* Biophysical regulation of epigenetic state and cell reprogramming. en. *Nature Materials* **12,** 1154–1162. doi:10.1038/nmat3777 (Dec. 2013).

340. Tung, J. C. *et al.* Tumor mechanics and metabolic dysfunction. en. *Free Radical Biology and Medicine* **79,** 269–280. doi:10.1016/j.freeradbiomed.2014.11.020 (Feb. 2015).

341. Glaser, A., McColl, B. & Vadolas, J. GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-mediated Genome Editing. en. *Molecular Therapy - Nucleic Acids* **5,** e334. doi:10.1038/mtna.2016.48 (2016).

342. Finley, A. & Copeland, R. A. Small Molecule Control of Chromatin Remodeling. en. *Chemistry & Biology* **21,** 1196–1210. doi:10.1016/j.chembiol.2014.07.024 (Sept. 2014).

343. Gnyszka, A., Jastrzębski, Z. & Flis, S. DNA Methyltransferase Inhibitors and Their Emerging Role in Epigenetic Therapy of Cancer. en. *ANTICANCER RESEARCH,* 8 (2013).

344. Almassalha, L. M. *et al.* Label-free imaging of the native, living cellular nanoarchitecture using partial-wave spectroscopic microscopy. en. *Proceedings of the National Academy of Sciences* **113,** E6372–E6381. doi:10.1073/pnas.1608198113 (Oct. 2016).

345. Tidball, A. M. *et al.* Rapid Generation of Human Genetic Loss-of-Function iPSC Lines by Simultaneous Reprogramming and Gene Editing. English. *Stem Cell Reports* **0.** doi:10.1016/j.stemcr.2017.07.003 (Aug. 2017).