CHEMICAL SCREEN FOR EPIGENETIC BARRIERS TO SINGLE ALLELE ACTIVATION OF *OCT4*

Kathryn Margaret Headley

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of PhD in the Curriculum of Genetics and Molecular Biology in the School of Medicine.

Chapel Hill
2019

Approved by:

Nate Hathaway

Ian Davis

Stephen Frye

Paul Wade

Greg Wang

ABSTRACT

Kathryn Margaret Headley: Chemical screen for barriers to single-allele activation of *Oct4*. (Under the direction of Nate Hathaway)

Induced pluripotent stem cells (iPSCs) have applications in many research fields. However, current methods to produce iPSCs have been inefficient and largely carcinogenic. The current study uses the chromatin *in vivo* assay (CiA) mouse platform explore barriers to small molecule facilitated activation of a single allele of *Oct4*. The C*iA:Oct4* allele contains an engineered EGFP reporter that replaces one allele of the *Oct4* gene combined with an upstream Gal4 array in the promoter at this engineered locus. A high throughput small molecule screen was performed with and without recruitment of a transcriptional activator, VP16. From this screen, we identified that Azacytidine and multiple different HDAC inhibitors facilitated C*iA:Oct4* gene activation. Furthermore, an HDAC inhibitor, Mocetinostat, was found to increase reprogramming efficiency during transcription factor reprogramming approximately 20 fold. Our results identified more recently discovered HDAC inhibitors which could be generally useful for future reprogramming studies. Furthermore, our results highlight the differences between testing small molecules using single allele activation versus whole cell reprogramming

To my best friend, confidant, and everyday hero,
Thank you. I couldn't have done this without you.
I am the luckiest girl in the world to have you as a sister.

ACKNOWLEDGEMENTS

I would like to thank my principal investigator, Dr. Nate Hathaway who was an amazing mentor and teacher. Through his mentorship, I learned the value of positivity and adaptability in the face of scientific challenges. I would also like to thank my committee members, Dr. Ian Davis, Dr. Stephen Frye, Dr. Paul Wade, and Dr. Greg Wang who provided thoughtful and discerning feedback which helped guide my project to success. Finally, I would also like to thank my family whose unwavering resilience in their own struggles taught me how to face personal and professional struggles with courage, kindness, and love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

53BP1	p53-Binding Protein 1
ATM	Ataxia-telangiectasia mutated
ATP	Adenosine triphosphate
Aza	Azacytidine
BrDU	Bromodeoxyuridine / 5-bromo-2'-deoxyuridine
Cas	CRISPR-associated proteins
CiA	Chromatin *in-vivo* Assay
ciPSCs	Chemically induced pluripotent stem cells
CO2	Carbon dioxide
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNAs
CtIP	C-terminal binding protein 1
ctrl	Control
DBD	DNA Binding Domain
dCas9	Catalytically dead Cas9
DDR	DNA damage response
DMEM	Dulbecco`s Modified Eagle Media -
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNA-PKcs	DNA-dependent protein kinase, catalytic subunit
DNMT	DNA methyltransferase

DNMTi	DNA methyltransferase inhibitor
DSB	Double strand break
EGFP	Enhanced Green Fluorescent Protein
EmGFP	Emerald Green Fluorescent Protein
EpiG	Epigenetic-targeted compound library
ESC	Embryonic Stem Cells
EXO1	Exonuclease I
FBS	Fetal bovine serum
FSC	Forward scatter
GFP	Green fluorescent protein
H2A	Histone 2A variant
H2AK13	Histone 2A lysine 13
H2AK15	Histone 2A lysine 15
H2AK15ub	Histone 2A lysine 15 ubiquitinated
H2B	Histone 2B
H2B-mC	Histone 2B mCherry
H3K4	Histone 3 Lysine 4
H4K20me2	Histone 4 Lysine 20 dimethylated
HATs	Histone acetyltransferases
HDAC	Histone Deactylase
HDACi	Histone Deactylase inhibitor
HDR	Homology directed repair

HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
hESC	human Embryonic Stem Cell
HMTase	Histone lysine methyltransferase
HMTs	Histone lysine methyltransferase
iPSC	Induced pluripotent stem cell
IVF	*In vitro* fertilization
KD	Knockdown
Kdm4b	Lysine demethylase 4B
KEAP1	Kelch-like ECH-associated protein 1
Klf4	Kruppel-like factor 4
KO	Knockout
KRAB	Krueppel-associated box
MAPK	Mitogen-activated protein kinase
MDC1	Mediator of DNA damage checkpoint 1
MEF	Mouse Embryonic Fibroblast
Moc	Mocetinostat
MRN	MRE11-RAD50-NBS1
NaB	Sodium butyrate
NEAA	Non-essential amino acids
NHEJ	Non homologous End Joingin
Oct4	Octamer-binding transcription factor 4
OSKM factors	Oct4, Sox2, Klf, c-Myc

PAM	Protospacer adjacent motif
PBS	phosphate-buffered saline
PCR	Polymerase chain reaction (PCR)
PEI	Polyethylenimine
Penn-Strep	Penicillin-Streptomycin
PI	Propidium Iodide
PI3-kinase	Phosphoinositide 3-kinase
PK	Protein Kinase
POU5f1	POU Class 5 Homeobox 1
PRC2	Polycomb repressive complex 2
PTM	Post Translational Modification
RA	Retinoic Acid
RIF1	Replication timing regulatory factor 1
RNA	Ribonucleic acid
RNF168	Ring finger protein 168
RNF8	Ring finger protein 8
ROCK	Rho-associated protein kinase
RPA	Replication protein A
SAHA	Suberoylanilide hydroxamic acid
SCNT	Somatic cell nuclear transfer
sgRNA	single guide RNA
Sox2	(sex determining region Y)-box 2

SSC	Side scatter
TET	Ten-Eleven Translocation (TET) family protein
TGF	Transforming growth factor
TGFBeta	Transforming growth factor beta
tracrRNA	trans-activating crRNA
tRNA	trans-activating crRNA
TSA	Trichostatin A
TTD	Tandem Tudor Domain
VPA	Valproic Acid
WRN	Werner syndrome helicase

CHAPTER 1: INTRODUCTION TO OVERCOMING iPSC GENERATION BARRIERS

Stem cells: Background

Stem cells, or undifferentiated cells, have the potential to become any cell type. This makes stem cells desirable for numerous therapeutic applications in research fields ranging from cancer, heart damage repair, repair of tissues weakened by autoimmune disease, and many more. Stem cells can differentiate into specialized cell types by obeying cellular developmental commands. Stem cells are categorized by two types of cells: embryonic stem cells and adult stem cells. Embryonic stem cells (ESCs), derived from the inner cell mass of a blastocyst, are the most versatile stem cells with potential to become all cell types in a mature organism. Adult stem cells are more limited in differentiation choices, acting as a reservoir of precursor cells, which can differentiate into specialized cells (often in response to initiations of tissue regeneration and repair pathways). Differentiation, once thought of as a unidirectional pathway, has now become a multidirectional pathway as scientists have developed methods to reprogram differentiated cells. However, current reprogramming methods do not produce therapeutic quality reprogrammed cells. This is likely due to imperfect cell conversions, defects in epigenetic profiles, mutations due to adenovirus insertion models, defects in transcriptome, and overexpression of factors which promote carcinogenesis. Optimizations of reprogramming methods will require an in-depth understanding of reprogramming pathways, factors which manipulate them, and how manipulations of these pathways work to affect cellular identity.

Embryonic stem cells vs. adult stem cells

Embryonic stem cells have the capability to become almost any cell type as well as the ability to proliferate indefinitely *in vitro*. Stem cells were first studied when scientists disaggregated cells of embryos prior to *in vitro* fertilization (IVF) implantation in the 1960s (Edwards, 2001). The disaggregated cells, now known to be embryonic stem cells, were noticed as cells with distinct morphologies and proliferative potential (Edwards, 2001). The 1980s-1990s was a time of stem cell discovery in which researchers started to characterize ESCs archetypal stem cell characteristics including: a self-renewing capability, consistent karyotyping after several passages *in vitro*, active telomerase, and potential for differentiation into specialized cells, potential for teratoma formation, potential to form colonies of clonal cells, chimera formation potential following injection into a blastocyst, and the expression of pluripotent markers such as *Oct4* and *Sox2* (Bongso and Richards, 2004; Evans and Kaufman, 1981; Thomson et al., 1995). ESCs have potential for use in therapeutic applications through tissue engineering and regenerative medicine. However, ethical considerations limit the use of embryonic stem cells in research and also have made it difficult for clinical developments. The question of whether an embryo, destroyed for ESC harvesting, is a person has become an ethical research barrier. This ethical barrier makes it difficult to obtain donors for embryos, and results in increased ethical oversight of research and decreased potential for collaborations as legal and ethical standards differ between research institutions (Bongso and Richards, 2004).

Adult stem cells, an alternative to ESCs, are multipotent stem cells present throughout life in the body primarily in a nondividing state. Adult stem cells are activated in response to cellular need for a particular cell type or in response to tissue damage. Adult stem cells are found primarily in the skin, blood, blood vessels, bone marrow, skeletal muscles, and the liver, but are

present in low levels which make isolation of adult stem cells from these tissues difficult. Therapeutic limitations include: difficult culturing *in vitro* due to a slow rate of division and passaging can result in chromosome shortening in length (Bongso and Richards, 2004). Current stem cell treatments and clinical trials largely refer to adult stem cell treatments, namely, hematopoietic stem-cell transplants (Gratwohl et al., 2015). However, adult stem cell treatments are currently branching out into clinical trials such as heart disease (Patel et al., 2016), stroke (Steinberg et al., 2016), sickle cell anemia (Bernaudin et al., 2007), spinal cord injury (Lima et al., 2010), multiple sclerosis (Burt et al., 2015), diabetes (Cai et al., 2016; Gaipov et al., 2016), retinal/optic nerve disease (Weiss et al., 2015), and lupus (Burt, 2006).

Cellular reprogramming

Cell reprogramming methods have now become available which allow researchers to manipulate cell fate; most notably, the generation of stem cells derived from somatic cells. Stem cells reprogrammed from somatic tissue have many applications in therapeutic development and circumvent issues of tissue rejection as well as ethical considerations posed by stem cells. The first reprogrammed cells were created by the introduction of a tadpole intestinal epithelia nuclei into a somatic cells from an enucleated frog egg, creating a cloned tadpole of donor tadpole by Sir John Gurdon (Omole and Fakoya, 2018). This method of removing nuclei and replacing with a donor somatic nucleus was later named somatic cell nuclear transfer (SCNT) and was the first to show that the somatic cell nucleus contained all the genetic information required for the generation of an entire organism. Notably, the SCNT method was later mirrored in a method by Sir Ian Wilmut to create the first cloned mammal, “Dolly” the sheep (Omole and Fakoya, 2018).

The next major development in reprogramming was the generation of induced pluripotent stem cells (iPSCs) in 2006 by Yamanaka et al. In this method, an adenovirus was used to exogenously express four factors (Oct4, Sox2, Klf4, c-Myc (OSKM factors)) to Mouse Embryonic Fibroblast (MEF) cells to reprogram the cells into to an ESC-like state, called iPSCs (Omole and Fakoya, 2018). Gurdon and Yamanaka received the Nobel prize in 2012 for their advances to cellular reprogramming technologies (Omole and Fakoya, 2018). Excitingly, iPSC generated through transcription factor reprogramming, with the OSKM factors, had similar pluripotent differentiation potential as ESCs and contained many of the same cellular characteristics of ESCs; including the ability to self-renew and maintain cellular identity in culture. Importantly, iPSCs had further applications. Somatic cells could be harvested from a patient and reprogramed *in vitro* into a desired tissue, and either returned to the patient or tested. Testing can involve either a drug specific screen, searching for a beneficial response with the patient's tissue, or cells can be genetically altered *in vitro* for further testing then differentiated into a specialized tissue and transplanted back to a patient.

Both the SCNT method and transcription factor reprogramming methods can produce reprogrammed pluripotent cells. However, none of these methods can produce cells which sufficiently resemble stem cells for use in therapeutics. Namely, epigenetic and transcriptional defects have been found in both SCNT and iPSCs compared to ESCs (Krupalnik and Hanna, 2014). Also, SCNT cells have a low efficiency for successful progeny and stem cell generation, likely due to incomplete reprogramming (Campbell et al., 2007). Additionally, SCNT requires donation of human eggs, which has ethical considerations and can be a hindrance for therapeutic development. Benefits for SCNT cells include a method for generation which does not utilize transcription of factors which promote carcinogenesis (Krupalnik and Hanna, 2014).

Nevertheless, iPSCs likely have greater therapeutic potential, because somatic cells can easily and ethically be harvested in a patient and can consistently generate iPSCs from harvested cells.

Key Regulatory Pathways in iPSC Generation

Cell reprogramming is primarily regulated by the control of gene expression through transcriptional control of gene networks such as epithelial-mesenchymal transition pathways, nuclear receptor pathways metabolism pathways, and cellular stress pathways. Regulation of the transcriptome is governed by multiple chromatin regulatory mechanisms including DNA accessibility, in which genes related to pluripotency are highly regulated in adult cells. Cell reprogramming involves promoting, establishing, and maintaining the reprogrammed transcriptome and erasing previous somatic cell transcriptional and epigenetic memories (Nashun et al., 2015).

DNA, epigenetic writers, readers, and erasers:

DNA accessibility in eukaryotic cells is a critical component of regulation of gene expression. The DNA double helix is tightly wrapped and wound around histone octamers which make up nucleosomes, which are further packaged into chromatin fibers which make up a chromosome. Chromatin exists in generally two states: heterochromatin (tightly compacted, associated with transcriptional repression) or euchromatin (open chromatin, available for transcription). Chromatin compaction and accessibility of DNA is generally influenced by chemical modifications on DNA and histone proteins. Chemical modifications are regulated by epigenetic factors such as writers, readers, and erasers.

Writers, such as DNA methyltransferases, histone lysine methyltransferases, and histone acetyltransferases, can "write" or introduce chemical modifications to DNA or histone tails which can influence chromatin accessibility. Epigenetic "reader" enzymes can identify and respond to chemical modification profiles. Erasers, such as histone deacetylase (HDAC)

complexes, remove or “erase” post translational modifications which likewise affect chromatin compaction. Diverse groups of enzymes that work to change the epigenetic profile of loci lead to expression profiles. Not surprisingly, epigenetic writers, readers, and erasers have a major influence on cell fate. Perturbing these cellular pathways has been shown to be helpful in cellular reprogramming. Central epigenetic modifications are decided at key points in development and are typically stable and inheritable from parent cell to daughter cell, this ‘epigenetic memory’ is influential in maintenance of developmental state.

Epigenetic writers:

Chemical modifications are diverse and governed by several enzymes and direct transcriptional activation or repression (Biswas and Rao, 2018). The most well studied of the modifications are methylation and acetylation modifications. In order to reprogram a somatic cell to an embryonic like state, methylation and acetylation patterns will need to be rewritten to promote a pluripotent resembling epigenetic profile. Methylation modifications can occur on both DNA and histones while acetylation modifications occur solely on histones (Biswas and Rao, 2018). DNA methyltransferases (DNMTs) are enzymes that transfer a methyl group from S-adenosyl-L-methionine to the 5’ position of the nitrogenous base, cytosine, in DNA (Biswas and Rao, 2018). There are five types of DNMTs: DNMT1, DNMT2, DNMT3A, DNMT3B, and DNMT3L. DNMT1 is the most prevalent in adult cells and maintains methylation levels in the cell (Biswas and Rao, 2018). DNMT1 is prevalent at the replication fork where it catalyzes methylation from parental to daughter strand. DNMT1 is important in maintaining genome wide methylation changes throughout various stages during development. DNMT2 methylates the 38th cytosine residue to in the anticodon loop of tRNAs providing stability to the tRNA (Biswas and

Rao, 2018). DNMT3A and DNMT3B are DNA methyltransferases involved with genome wide methylation of DNA, they are important for the generation of a methylation pattern during gametogenesis and early embryogenesis (Biswas and Rao, 2018). DNMT3L is a catalytically inactive protein, which is an important regulation for *de novo* methylation, however has been known to associate with DNMT3A and DNMT3B to increase catalytic activity (Biswas and Rao, 2018; Gowher et al., 2005).

Histone modifications also represent a major epigenetic gene regulator. Histone methylations occur on lysine residues and can be methylated up to three times. Specific modifications are associated with transcriptional activation (i.e. H3K4me(2,3), H3K9me1, H3K27me1, H3K36me3, and HK20me1) and repression (i.e., H3K9me(2,3), H3K27me(2,3), and H4K20me3). Histone lysine methyltransferase (HMTs) are the writer enzymes which catalyze methyl group transfer from adenosyl-methionine to lysine residues. The polycomb repressive complex 2 (PRC2) with histone methyltransferase activity is another notable writer complex involved in producing the transcriptional repressive mark H2K27me(2,3). Histone acetylation on lysine residues results in a loss of positive charge which can lead to a decreased association with DNA, often associated with transcriptional activation (Biswas and Rao, 2018).

Epigenetic readers:

Cellular interpretation of chemical modifications is performed by epigenetic reader proteins. Notably, many epigenetic readers are also chromatin modifiers, which initiate cellular responses based on their interpretation of epigenetic modifications. Reading capabilities are primarily based on domains which recognize particular modifications. Histone methylation reader domains are exampled by chromodomains, tudor domains, malignant brain tumor, plant

homeodomain domains. Histone acetylation readers also play a role in the recognition of posttranslational marks, the most studied of which are bromodomains (Biswas and Rao, 2018; Park et al., 2016). These domains recognize specific histone modifications and initiate cellular responses. Mbd3 (a DNA methylation reader protein) depletion during transcription factor reprogramming has been associated with increasing reprogramming efficiency. However, iPSC generated in Mbd3 depleted cells were insufficient quality for therapeutic use (Brumbaugh and Hochedlinger, 2013; Chin et al., 2009; Ji et al., 2016; Rais et al., 2013; Zhou et al., 2008; Zviran and Hanna, 2014).

Epigenetic erasers:

The epigenetic eraser proteins, histone demethylases and histone deacetylase play important roles in gene regulation (Biswas and Rao, 2018; D'Oto et al., 2016). Histone deacetylases (HDAC) catalyze the removal of the ε-amino acetyl group from lysine residues. This returns a positive charge to the lysine residues enhancing the DNA (negatively charged) to histone interaction, allowing for more tightly wrapped DNA, which increases transcriptional silencing (Biswas and Rao, 2018). Notably, chemical inhibition of HDACs has been shown to facilitate reprogramming (Huangfu et al., 2008). On the other hand, expression of demethylases has been associated with both increases and decreases in reprogramming efficiencies. For example, overexpression of Kdm4b demethylase in combination with multiple methyltransferase deficiencies promoted reprogramming (Huang et al., 2011; Ji et al., 2016; Pang et al., 2011). Conversely, JMJD3 a histone H3K27 demethylase expression has been shown to be associated with reprogramming repression (Ji et al., 2016; Yoo et al., 2011).

It is also important to mention that microRNAs play a critical role in gene expression and is a notable epigenetic barrier in reprogramming. Importantly, microRNAs most notably miR26b, are involved with targeting mRNA regions associated with DNA methylation (Ji et al., 2016; Maherali et al., 2007). Also, the miR-290-295 clusters have been shown to be important regulators of the ESC-specific cell cycle. (Ji et al., 2016; Judson et al., 2009).

Transcriptional regulation:

It is clear that transcription factor expression plays a powerful role in reprogramming. Yamanka and others have demonstrated that ectopic expression of only four transcription factors allows for the reprogramming of cells from a differentiated state to a pluripotent state. In these experiments, it was found that within two to three weeks modulation of expression of hundreds of genes occurs, as well as chromatin remodeling (Carey et al., 2009). An important transcription factor family involved with epigenetic profile modifications is the TET family. They catalyze the conversion of methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmc). Expression of Tet1 and Tet2 has been shown to increase cell reprogramming efficiency (Costa et al., 2013; Doege et al., 2012; Ji et al., 2016). TET1 and TET2 have been shown to interact with Nanog, Sox2, and Oct4 (Costa et al., 2013; Gao et al., 2013; Ji et al., 2016; Zhu et al., 2014) Interestingly, deletion of Tet1 has also been shown to slightly enhance reprogramming efficiency. Tet2 inactivation reduced reprogramming 70%, Tet3 had an insignificant effect (Hu et al., 2014; Ji et al., 2016) Knock-down of tet1-3 in ES cells had no effect on pluripotency. However, in MEF cells, triple deficiencies (such as tet2 deletion from tet1/tet3 double knock-out or tet3 deficiency from tet1/tet2 double knockout, or tet1/tet2/tet3 triple knockout) inhibits reprogramming. The triple

knockout MEFs inability to be reprogrammed is likely due to a failure of the mesenchymal to epithelial transition (Hu et al., 2014; Ji et al., 2016).

Cell morphology pathways are also important when considering reprogramming methods. The TGF- β family proteins are important regulators of morphogenesis and cell fate, and have been shown to be a successful target for increasing reprogramming efficiency (Mullen and Wrana, 2017). TGF- β family proteins interact with transmembrane serine/threonine kinase receptors which signal several gene pathways. Namely, TGF-β receptor activation can regulate Smad signaling pathways which regulate several transcriptional profiles often in conjunction with interactions with epigenetic modifiers such as histone acetyl transferases, p300, and HDACs (Mullen and Wrana, 2017).

Other important regulatory networks affecting reprogramming efficiency are metabolic response pathways. A cyclin-dependent kinase inhibitor, p21, involved with cell cycle progression has been shown to decrease reprogramming efficiency (Ji et al., 2016; Pasque et al., 2012). Similarly, cell depletion of p53, a tumor suppressor protein involved with DNA damage response (and targets p21) has been shown to increase reprogramming efficiency (Ji et al., 2016).

Protein kinases (PKs) represent a common target for regulating cell fate. PK regulate several processes involved in cell identity including cell cycle, transcription, and metabolic switching. PKs transfer phosphate groups, usually from adenosine triphosphate (ATP) or guanosine triosephosphate (GTP) to protein substrates which regulate signaling pathways. Notably, protein kinases regulate several steps in the Wnt signaling pathway that have been shown to be important in pluripotency maintenance (Hime and Abud, 2013; Stadtfeld and Hochedlinger, 2010). The mitogen-activated protein kinase (MAPK) pathway is another important protein kinase pathway involved with pluripotency. MAPK are involved with

signaling responses to cellular stimuli, promoting cell proliferation, differentiation , migration, and apoptosis (Neganova et al., 2017).

Chemical Approaches for Enhancing iPSC Generation

Current iPSC generation techniques have several therapeutic flaws that need to be overcome in order to move iPSCs to clinical settings. First, the original methods to produce iPSCs were very inefficient with only ~0.067% of MEF cells successfully reprogrammed into iPSCs using the Yamanaka method (Takahashi et al., 2007). Additionally, generated iPSCs have been demonstrated to be largely carcinogenic. This is, in part, likely due to the current method of iPSC generation which uses a virus gene insertion model of pluripotency factors to promote cell plasticity (Lin and Wu, 2015a). Small-molecule-only reprogramming methods for iPSCs have been seen as a safer alternative to the original methods. Small molecules often have reversible effects, dose response, do not result in genetic chances, and can be combined with other small molecules to regulate multiple pathways at the same time. Furthermore, small molecules are simple to store, often cheap to produce, and easy to test in a research setting. Here we outline current small molecules that have been successful in iPSC generation efficiency, organized by factors which influence cellular identity including; chromatin profile, mesenchymal-epithelial transition, nuclear receptor pathways metabolism pathways, and cellular stress pathways (Lin and Wu, 2015a; Nie et al., 2012).

Epigenetic profiles such as DNA methylation and histone post translational modifications are key features in reprogramming cellular identity. The first molecules shown to increase reprogramming efficiency were well characterized HDACis and DNMTis and were used to facilitate SCNT (Kishigami et al., 2006) and iPSC generation studies (Huangfu et al., 2008). From these studies, HDACi valproic acid (VPA) was shown to increase reprogramming efficiency ~100-fold which has since been shown to be a leading molecule in increasing iPSC generation efficiency. Other epigenetically relevant small molecules identified by these screens

and further screens which increase iPSC generation efficiency include: HDACi SAHA (10-fold), TSA (10-fold). HDACi sodium butyrate (NaB) (100-fold), (Mali et al., 2010; Zhang et al., 2014), DMNT inhibitor decitabine (~3 fold), and a dual DMNT and HDAC inhibitor RSC133 (~3 fold) (Lee et al., 2012; Lin and Wu, 2015a). Similarly, small molecules inhibiting histone methylation pathways have also been shown to increase iPSC efficiency. Histone methyltransferase inhibitors, 3-Deazaneplanocin A (DZNep) hydrochloride and H3K4 demethylation inhibitor, Parnate increase efficiency 65-fold, and ~3-fold respectively (Lin and Wu, 2015a).

Morphological structures including cell polarity, cell boundaries, and cell-cell interactions are key factors which differ between somatic cells and iPSCs and must be reprogrammed for complete iPSC transformation (Gill et al., 2003; Lin et al., 2009a). Inhibiting an epithelial to mesenchymal transition regulating protein, TGF-β, with small molecule, A83-01 (Zhu et al., 2010), has been shown to increase iPSC generation 7-fold during transcription factor reprogramming. Similarly, small molecule, SB431542, inhibits three factors involved in regulation of morphological transitions (ALK4, ALK5, and ALK7) and increases iPSC generation 200-fold (Lin et al., 2009a; Lin and Wu, 2015a). Likewise, Thiazovivin, an inhibitor of Rho-associated protein kinase (ROCK), which is known to be involved in cell shape and movement regulation, increases iPSC generation efficiency 200-fold (Lin et al., 2009b; Lin and Wu, 2015a).

Cellular communication of extracellular influences is often initiated by receptor binding and plays an important role in the regulation of proliferation, differentiation, cell motility, and survival (all of which affect cellular identity). Notably, the MAPK/ERK kinase pathway has been shown to be an effective target in chemically facilitated reprogramming studies. Other inhibitors in this pathway, Compound B10 and PD0325901, increase iPSC generation efficiency

3-fold and ~200 fold respectively (Li and Rana, 2012; Lin et al., 2009b). In another pathway involved with interpretation and response to extracellular stimuli, Retinoic acid receptor agonists, AM580, TTNPB (Brewer et al., 1993; Ross-Innes et al., 2010), have also been shown to increase iPSC efficiency ~200-fold and ~15 fold respectively (Hou et al., 2013; Lin and Wu, 2015a; Wang et al., 2011; Yang et al., 2015).

Inhibiting proteins in metabolism pathways has also been successful in increasing iPSC generation efficiency. mTOR inhibitor, Rapamycin, increases iPSC generation efficiency 4.8 fold (Chen et al., 2011). And IP3K inhibitor, Compound B8, increases iPSC generation efficiency 3-fold (Zhu et al., 2010). Lastly, inhibitors of cellular stress response proteins in somatic cells can increase iPSC generation efficiency. Specially, a Hypoxia-inducible factor pathway activator, Quercetin, increases reprogramming efficiency 3-fold (Lin and Wu, 2015b; Zhu et al., 2010).

Combinatorial approaches to iPSC generation

In order to fully reprogram adult cells, all the above factors will likely need to be modified to obtain therapeutic quality iPSCs. In a step towards generating iPSCs from only small molecules, researchers have started testing combinations of inhibitors and testing for reprogramming. Many steps towards generating chemically induced pluripotent stem cells (ciPSCs) have focused on replacing factors from transcription factor reprogramming methods with small molecules. Markedly, BIX01294 (G9a HMTase inhibitor) and RG108 (a DNMT inhibitor) have been shown to replace Oct4 in transcription factor reprogramming methods lacking exogenous *Oct4* expression (Lin and Wu, 2015a; Shi et al., 2008b, 2008a). Similarly, TGFβ inhibitors, 616452 and SB431542, and L-channel calcium agonist, BayK8644 (in

combination with BIX01294), have been shown to replace Sox2 in transcription factor reprogramming methods (Huangfu et al., 2008). And Kenpaullone (Lyssiotis et al., 2009) has been shown to be able to replace c-Myc and Klf in transcription factor reprogramming methods respectively. The first combination cocktail showing potential for ciPSCs generation named VC6TF (VPA, CHIR99021, 616452, tranylcypromine, and Forskolin). However, ciPSCs generated from this method were incompletely reprogramed as evidenced by undetectable *Oct4* and *NANOG* expression (Hou et al., 2013).

In summary, small molecule inhibitors regulating chromatin profile, mesenchymal-epithelial transition, signaling to extracellular responses, and cellular stress have been shown to increase iPSC generation efficiency. Likewise, small molecules have also been identified which can replace transcription factors during transcription factor reprogramming. Likely, future methods to produce ciPSCs will be a combination of small molecules which work to increase pluripotency factors associated with transcription factor reprogramming while also promoting pluripotent chromatin remodeling, signaling, morphological, and cellular stress pathways.

REFERENCES

Bernaudin, F., Socie, G., Kuentz, M., Chevret, S., Duval, M., Bertrand, Y., Vannier, J.-P., Yakouben, K., Thuret, I., Bordigoni, P., Fischer, A., Lutz, P., Stephan, J.-L., Dhedin, N., Plouvier, E., Margueritte, G., Bories, D., Verlhac, S., Esperou, H., Coic, L., Vernant, J.-P., Gluckman, E., 2007. Long-term results of related myeloablative stem-cell transplantation to cure sickle cell disease. Blood 110, 2749–2756. https://doi.org/10.1182/blood-2007-03-079665

Biswas, S., Rao, C.M., 2018. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. Eur. J. Pharmacol. 837, 8-24. https://doi.org/10.1016/j.ejphar.2018.08.021

Bongso, A., Richards, M., 2004. History and perspective of stem cell research. 18, 827-842 https://doi.org/10.1016/j.bpobgyn.2004.09.002

Brewer, G.J., Torricelli, J.R., Evege, E.K., Price, P.J., 1993. Optimized survival of hippocampal neurons in B27-supplemented neurobasal, a new serum-free medium combination. J. Neurosci. Res. 35, 567–576. https://doi.org/10.1002/jnr.490350513

Brumbaugh, J., Hochedlinger, K., 2013. Removing reprogramming roadblocks: Mbd3 depletion allows deterministic iPSC generation. Cell Stem Cell 13, 379–81. https://doi.org/10.1016/j.stem.2013.09.012

Burt, R.K., 2006. Nonmyeloablative Hematopoietic Stem Cell Transplantation for Systemic Lupus Erythematosus. JAMA 295, 527. https://doi.org/10.1001/jama.295.5.527

Burt, R.K., Balabanov, R., Han, X., Sharrack, B., Morgan, A., Quigley, K., Yaung, K., Helenowski, I.B., Jovanovic, B., Spahovic, D., Arnautovic, I., Lee, D.C., Benefield, B.C., Futterer, S., Oliveira, M.C., Burman, J., 2015. Association of Nonmyeloablative Hematopoietic Stem Cell Transplantation With Neurological Disability in Patients With Relapsing-Remitting Multiple Sclerosis. JAMA 313, 275. https://doi.org/10.1001/jama.2014.17986

Cai, J., Wu, Z., Xu, X., Liao, L., Chen, J., Huang, L., Wu, W., Luo, F., Wu, C., Pugliese, A., Pileggi, A., Ricordi, C., Tan, J., 2016. Umbilical Cord Mesenchymal Stromal Cell With Autologous Bone Marrow Cell Transplantation in Established Type 1 Diabetes: A Pilot Randomized Controlled Open-Label Clinical Study to Assess Safety and Impact on Insulin Secretion. Diabetes Care 39, 149–157. https://doi.org/10.2337/dc15-0171

Campbell, K.H.S., Fisher, P., Chen, W.C., Choi, I., Kelly, R.D.W., Lee, J.-H., Xhu, J., 2007. Somatic cell nuclear transfer: Past, present and future perspectives. Theriogenology 68, S214–S231. https://doi.org/10.1016/j.theriogenology.2007.05.059

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., Jaenisch, R., 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. 106, 157–162. https://doi.org/10.1073/pnas.0811426106

Chen, T., Shen, L., Yu, J., Wan, H., Guo, A., Chen, J., Long, Y., Zhao, J., Pei, G., 2011. Rapamycin and other longevity-promoting compounds enhance the generation of mouse induced pluripotent stem cells. Aging Cell 10, 908–911. https://doi.org/10.1111/j.1474-9726.2011.00722.x

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiuwu, O., Richter, L., Zhang, J., Khvorostov, I., Ott, V., Grunstein, M., Lavon, N., Benvenisty, N., Croce, C.M., Clark, A.T., Baxter, T., Pyle, A.D., Teitell, M.A., Pelegrini, M., Plath, K., Lowry, W.E., 2009. Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. Cell Stem Cell 5, 111–123. https://doi.org/10.1016/j.stem.2009.06.008

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P. V, Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., Das, S., Levasseur, D.N., Li, Z., Xu, M., Reik, W., Silva, J.C.R., Wang, J., 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature 495, 370–4. https://doi.org/10.1038/nature11925

D'Oto, A., Tian, Q.-W., Davidoff, A.M., Yang, J., 2016. Histone demethylases and their roles in cancer epigenetics. J. Med. Oncol. Ther. 1, 34–40.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., Levine, R.L., Nik, S., Chen, E.I., Abeliovich, A., 2012. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature 488, 652–5. https://doi.org/10.1038/nature11333

Edwards, R.G., 2001. IVF and the history of stem cells. Nature. 413, 349-351. https://doi.org/10.1038/35096649

Evans, M.J., Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–6. https://doi.org/10.1038/292154a0

Gaipov, A., Turebekov, Z., Serebrennikova, D., Kozina, L., Askarov, M., Tuganbekova, S., 2016. SP381AUTOLOGOUS BONE MARROW-DERIVED STEM CELLS TRANSPLANTATION IN TYPE I DIABETES MELLITUS. Nephrol. Dial. Transplant. 31, i217–i217. https://doi.org/10.1093/ndt/gfw169.03

Gao, Y., Chen, J.J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., Yao, C., Liu, X., Lu, Z., Xu, Z., Kang, L., Chen, J.J., Wang, H., Cai, T., Gao, S., 2013. Replacement of Oct4 by Tet1 during iPSC Induction Reveals an Important Role of DNA

Methylation and Hydroxymethylation in Reprogramming. Cell Stem Cell 12, 453–469. https://doi.org/10.1016/j.stem.2013.02.005

Gill, S.S., Patel, N.K., Hotton, G.R., O'Sullivan, K., McCarter, R., Bunnage, M., Brooks, D.J., Svendsen, C.N., Heywood, P., 2003. Direct brain infusion of glial cell line–derived neurotrophic factor in Parkinson disease. Nat. Med. 9, 589–595. https://doi.org/10.1038/nm850

Gowher, H., Liebert, K., Hermann, A., Xu, G., Jeltsch, A., 2005. Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. J. Biol. Chem. 280, 13341–13348. https://doi.org/10.1074/jbc.M413412200

Gratwohl, A., Pasquini, M.C., Aljurf, M., Atsuta, Y., Baldomero, H., Foeken, L., Gratwohl, M., Bouzas, L.F., Confer, D., Frauendorfer, K., Gluckman, E., Greinix, H., Horowitz, M., Iida, M., Lipton, J., Madrigal, A., Mohty, M., Noel, L., Novitzky, N., Nunez, J., Oudshoorn, M., Passweg, J., van Rood, J., Szer, J., Blume, K., Appelbaum, F.R., Kodera, Y., Niederwieser, D., 2015. One million haemopoietic stem-cell transplants: a retrospective observational study. Lancet Haematol. 2, e91–e100. https://doi.org/10.1016/S2352-3026(15)00028-9

Hime, G.R., Abud, H.E., 2013. The Stem Cell State, in: Advances in Experimental Medicine and Biology. pp. 1–4. https://doi.org/10.1007/978-94-007-6621-1_1

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., Ge, J., Xu, J., Zhang, Q., Zhao, Y., Deng, H., 2013. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science (80-.). 341, 651–654. https://doi.org/10.1126/science.1239278

Hu, X., Zhang, L., Mao, S.-Q., Li, Z., Chen, J., Zhang, R.-R., Wu, H.-P., Gao, J., Guo, F., Liu, W., Xu, G.-F.G.-L., Dai, H.-Q., Shi, Y.G., Li, X., Hu, B., Tang, F., Pei, D., Xu, G.-F.G.-L., 2014. Tet and TDG Mediate DNA Demethylation Essential for Mesenchymal-to-Epithelial Transition in Somatic Cell Reprogramming. Cell Stem Cell 14, 512–522. https://doi.org/10.1016/j.stem.2014.01.001

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., Hui, L., 2011. Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature 475, 386–389. https://doi.org/10.1038/nature10116

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., Melton, D. a, 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat. Biotechnol. 26, 795–7.

Ji, P., Manupipatpong, S., Xie, N., Li, Y., 2016. Induced Pluripotent Stem Cells: Generation

Strategy and Epigenetic Mystery behind Reprogramming. Stem Cells Int. 2016, 8415010.

https://doi.org/10.1155/2016/8415010

Judson, R.L., Babiarz, J.E., Venere, M., Blelloch, R., 2009. Embryonic stem cell-specific microRNAs promote induced pluripotency. Nat. Biotechnol. 27, 459–61. https://doi.org/10.1038/nbt.1535

Kishigami, S., Mizutani, E., Ohta, H., Hikichi, T., Thuan, N. Van, Wakayama, S., Bui, H.-T., Wakayama, T., 2006. Significant improvement of mouse cloning technique by treatment with trichostatin A after somatic nuclear transfer. Biochem. Biophys. Res. Commun. 340, 183–189. https://doi.org/10.1016/J.BBRC.2005.11.164

Krupalnik, V., Hanna, J.H., 2014. The quest for the perfect reprogrammed cell. Nature 511, 160–162. https://doi.org/10.1038/nature13515

Lee, J., Xia, Y., Son, M.-Y., Jin, G., Seol, B., Kim, M.-J., Son, M.J., Do, M., Lee, M., Kim, D., Lee, K., Cho, Y.S., 2012. A Novel Small Molecule Facilitates the Reprogramming of Human Somatic Cells into a Pluripotent State and Supports the Maintenance of an Undifferentiated State of Human Pluripotent Stem Cells. Angew. Chemie Int. Ed. 51, 12509–12513. https://doi.org/10.1002/anie.201206691

Li, Z., Rana, T.M., 2012. A kinase inhibitor screen identifies small-molecule enhancers of reprogramming and iPS cell generation. Nat. Commun. 3, 1085. https://doi.org/10.1038/ncomms2059

Lima, C., Escada, P., Pratas-Vital, J., Branco, C., Arcangeli, C.A., Lazzeri, G., Santana Maia, C.A., Capucho, C., Hasse-Ferreira, A., Peduzzi, J.D., 2010. Olfactory Mucosal Autografts and Rehabilitation for Chronic Traumatic Spinal Cord Injury. Neurorehabil. Neural Repair 24, 10–22. https://doi.org/10.1177/1545968309347685

Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., Ding, S., 2009a. A chemical platform for improved induction of human iPSCs. Nat. Methods 6, 805–808. https://doi.org/10.1038/nmeth.1393

Lin, T., Wu, S., 2015a. Reprogramming with small molecules instead of exogenous transcription factors. Stem Cells Int. 2015, 1–11. https://doi.org/10.1155/2015/794632

Lyssiotis, C.A., Foreman, R.K., Staerk, J., Garcia, M., Mathur, D., Markoulaki, S., Hanna, J., Lairson, L.L., Charette, B.D., Bouchez, L.C., Bollong, M., Kunick, C., Brinker, A., Cho, C.Y., Schultz, P.G., Jaenisch, R., 2009. Reprogramming of murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. Proc. Natl. Acad. Sci. 106, 8912–8917. https://doi.org/10.1073/pnas.0903860106

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K., Hochedlinger, K., 2007. Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. Cell Stem Cell 1, 55–70. https://doi.org/10.1016/j.stem.2007.05.014

Mali, P., Chou, B.K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R.A., Ohm, J.E., Yu, W., Baylin, S.B., Yusa, K., Bradley, A., Meyers, D.J., Mukherjee, C., Cole, P.A., Cheng, L., 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells 28, 713–720. https://doi.org/10.1002/stem.402

Mullen, A.C., Wrana, J.L., 2017. TGF-β Family Signaling in Embryonic and Somatic Stem-Cell Renewal and Differentiation. Cold Spring Harb. Perspect. Biol. 9. https://doi.org/10.1101/cshperspect.a022186

Nashun, B., Hill, P.W.S., Hajkova, P., 2015. Reprogramming of cell fate: epigenetic memory and the erasure of memories past. EMBO J. 34, 1296–308. https://doi.org/10.15252/embj.201490649

Neganova, I., Chichagova, V., Armstrong, L., Lako, M., 2017. A critical role for p38MAPK signalling pathway during reprogramming of human fibroblasts to iPSCs. Sci. Rep. 7, 41693. https://doi.org/10.1038/srep41693

Nie, B., Wang, H., Laurent, T., Ding, S., 2012. Cellular reprogramming: A small molecule perspective. Curr. Opin. Cell Biol. https://doi.org/10.1016/j.ceb.2012.08.010

Omole, A.E., Fakoya, A.O.J., 2018. Ten years of progress and promise of induced pluripotent stem cells: historical origins, characteristics, mechanisms, limitations, and potential applications. PeerJ 6, e4370. https://doi.org/10.7717/peerj.4370

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Südhof, T.C., Wernig, M., 2011. Induction of human neuronal cells by defined transcription factors. Nature 476, 220–223. https://doi.org/10.1038/nature10202

Park, M., Keung, A.J., Khalil, A.S., 2016. The epigenome: the next substrate for engineering. Genome Biol. 17, 183. https://doi.org/10.1186/s13059-016-1046-5

Pasque, V., Radzisheuskaya, A., Gillich, A., Halley-Stott, R.P., Panamarova, M., Zernicka-Goetz, M., Surani, M.A., Silva, J.C.R., 2012. Histone variant macroH2A marks embryonic differentiation in vivo and acts as an epigenetic barrier to induced pluripotency. J. Cell Sci. 125, 6094–6104. https://doi.org/10.1242/jcs.113019

Patel, A.N., Henry, T.D., Quyyumi, A.A., Schaer, G.L., Anderson, R.D., Toma, C., East, C., Remmers, A.E., Goodrich, J., Desai, A.S., Recker, D., DeMaria, A., 2016. Ixmyelocel-T for patients with ischaemic heart failure: a prospective randomised double-blind trial. Lancet 387, 2412–2421. https://doi.org/10.1016/S0140-6736(16)30137-4

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., Maza, I., Mor, N., Baran, D., Weinberger, L., Jaitin, D.A., Lara-Astiaso, D., Blecher-Gonen, R., Shipony, Z., Mukamel, Z., Hagai, T., Gilad, S., Amann-Zalcenstein, D., Tanay, A., Amit, I., Novershtern, N., Hanna, J.H., 2013. Deterministic direct reprogramming of somatic cells to pluripotency. Nature 502, 65–70. https://doi.org/10.1038/nature12587

Ross-Innes, C.S., Stark, R., Holmes, K.A., Schmidt, D., Spyrou, C., Russell, R., Massie, C.E., Vowler, S.L., Eldridge, M., Carroll, J.S., 2010. Cooperative interaction between retinoic acid receptor- and estrogen receptor in breast cancer. Genes Dev. 24, 171–182. https://doi.org/10.1101/gad.552910

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Schöler, H.R., Ding, S., 2008a. Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds. Cell Stem Cell 3, 568–574. https://doi.org/10.1016/j.stem.2008.10.004

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Schöler, H.R., Ding, S., 2008b. A Combined Chemical and Genetic Approach for the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell. https://doi.org/10.1016/j.stem.2008.05.011

Stadtfeld, M., Hochedlinger, K., 2010. Induced pluripotency: history, mechanisms, and applications. Genes Dev. 24, 2239–2263. https://doi.org/10.1101/gad.1963910

Steinberg, G.K., Kondziolka, D., Wechsler, L.R., Lunsford, L.D., Coburn, M.L., Billigen, J.B., Kim, A.S., Johnson, J.N., Bates, D., King, B., Case, C., McGrogan, M., Yankee, E.W., Schwartz, N.E., 2016. Clinical Outcomes of Transplanted Modified Bone Marrow–Derived Mesenchymal Stem Cells in Stroke. Stroke 47, 1817–1824. https://doi.org/10.1161/STROKEAHA.116.012995

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., Yamanaka, S., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131, 861–872. https://doi.org/10.1016/j.cell.2007.11.019

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., Hearn, J.P., 1995. Isolation of a primate embryonic stem cell line. Proc. Natl. Acad. Sci. 92, 7844–7848. https://doi.org/10.1073/pnas.92.17.7844

Wang, Q., Xu, X., Li, J., Liu, J., Gu, H., Zhang, R., Chen, J., Kuang, Y., Fei, J., Jiang, C., Wang, P., Pei, D., Ding, S., Xie, X., 2011. Lithium, an anti-psychotic drug, greatly enhances the generation of induced pluripotent stem cells. Cell Res. 21, 1424–1435. https://doi.org/10.1038/cr.2011.108

Weiss, J.N., Levy, S., Malkin, A., 2015. Stem Cell Ophthalmology Treatment Study (SCOTS) for retinal and optic nerve diseases: a preliminary report. Neural Regen. Res. 10, 982–8. https://doi.org/10.4103/1673-5374.158365

Yang, J., Wang, W., Ooi, J., Campos, L.S., Lu, L., Liu, P., 2015. Signalling through retinoic acid receptors is required for reprogramming of both mouse embryonic fibroblast cells and epiblast stem cells to induced pluripotent stem cells. Stem Cells 33, 1390–1404. https://doi.org/10.1002/stem.1926

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., Crabtree, G.R., 2011. MicroRNA-mediated conversion of human fibroblasts to neurons. Nature 476, 228–31. https://doi.org/10.1038/nature10323

Zhang, Z., Xiang, D., Wu, W.-S., 2014. Sodium Butyrate Facilitates Reprogramming by Derepressing OCT4 Transactivity at the Promoter of Embryonic Stem Cell–Specific miR-302/367 Cluster. Cell. Reprogram. 16, 130–139. https://doi.org/10.1089/cell.2013.0070

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., Melton, D.A., 2008. In vivo reprogramming of adult pancreatic exocrine cells to β-cells. Nature 455, 627–632. https://doi.org/10.1038/nature07314

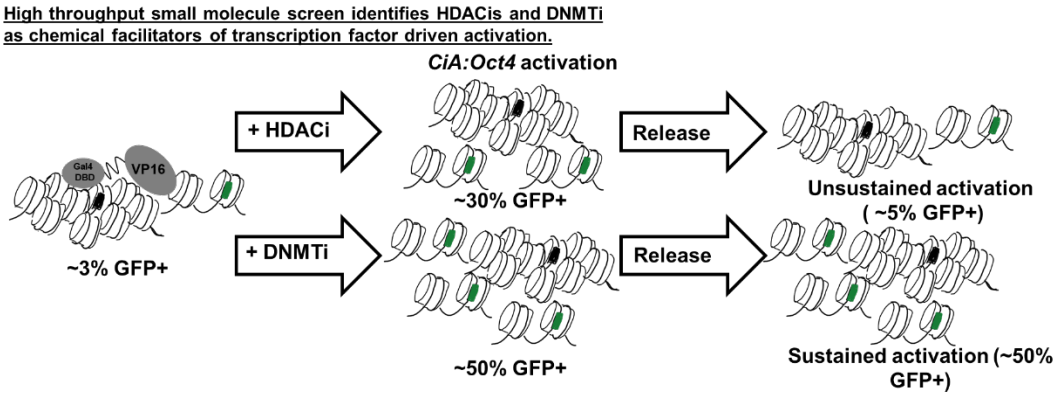
Zhu, G., Li, Y., Zhu, F., Wang, T., Jin, W., Mu, W., Lin, W., Tan, W., Li, W., Street, R.C., Peng, S., Zhang, J., Feng, Y., Warren, S.T., Sun, Q., Jin, P., Chen, D., 2014. Coordination of engineered factors with TET1/2 promotes early-stage epigenetic modification during somatic cell reprogramming. Stem cell reports 2, 253–61. https://doi.org/10.1016/j.stemcr.2014.01.012

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., Ding, S., 2010. Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell Stem Cell 7, 651–655. https://doi.org/10.1016/j.stem.2010.11.015

Zviran, A., Hanna, J.H., 2014. Lucky iPSCs. Genome Biol. 15, 109. https://doi.org/10.1186/gb4167

CHAPTER 2: CHEMICAL SCREEN FOR EPIGENETIC BARRIERS TO SINGLE ALLELE ACTIVATION OF *OCT4*[1]

Graphical Abstract



Here we employ a novel cellular platform capable of resolving dynamics of epigenetic transformations that occur when a single *Oct4* allele is activated by a directed transcription factor in fibroblast cells. We identified compounds that break down chromatin-based barriers to *Oct4* activation through blocking either histone deacetylation or DNA methylation pathways. We also provide a new way to lineage trace changes in gene activity in individual cells over time and through cell division.

---

[1]This chapter previously appeared as an article in Stem Cell Research. The original citation is as follows: Headley K. "Chemical screen for epigenetic barriers to single allele activation of *Oct4*," Stem Cell Research, 38 (July 2019): 101470.

Abstract

Here we utilized the chromatin *in vivo* assay (CiA) mouse platform to directly examine the epigenetic barriers impeding the activation of the *CiA:Oct4* allele in mouse embryonic fibroblasts (MEF)s when stimulated with a transcription factor. The *CiA:Oct4* allele contains an engineered EGFP reporter replacing one copy of the *Oct4* gene, with an upstream Gal4 array in the promoter that allows recruitment of chromatin modifying machinery. We stimulated gene activation of the *CiA:Oct4* allele by binding a transcriptional activator to the Gal4 array. As with cellular reprogramming, this process is inefficient with only a small percentage of the cells re-activating *CiA:Oct4* after weeks. Epigenetic barriers to gene activation potentially come from heavy DNA methylation, histone deacetylation, chromatin compaction, and other posttranslational marks (PTM) at the differentiated *CiA:Oct4* allele in MEFs. Using this platform, we performed a high-throughput chemical screen for compounds that increased the efficiency of activation. We found that Azacytidine and newer generation histone deacetylase (HDAC) inhibitors were the most efficient at facilitating directed transcriptional activation of this allele. We found one hit form our screen, Mocetinostat, improved iPSC generation under transcription factor reprogramming conditions. These results separate individual allele activation from whole cell reprogramming and give new insights that will advance tissue engineering.

Introduction

Regenerative medicine aims to replace damaged tissues with healthy engineered tissues (Tian et al., 2012; Walia et al., 2012). Many current regenerative medicine techniques use human derived stem cells (hESCs) from a donor to regenerate damaged tissues upon stem cell injection or to regenerate tissues *in vitro* which can be transplanted into the patient (Bongso and Richards, 2004; Mao and Mooney, 2015; Olson et al., 2011). Induced pluripotent stem cell (iPSC) therapies are a promising alternative within the regenerative medicine field allowing for individual treatments using iPSCs derived from a patient's own somatic cells (Kastenberg and Odorico, 2008; Mao and Mooney, 2015). The iPSC method avoids any potential ethical ramifications and has the advantage of treating patients with their own tissues. Furthermore, iPSCs specific tests can be done *in vitro* to personalize treatments (Bongso and Richards, 2004; Li and Li, 2014). Yet, a major barrier to application of iPSCs in clinical practice is that current iPSCs generated using the transcription factor induced reprogramming methods are inefficient and sometimes carcinogenic (Li et al., 2011; Medvedev et al., 2010; Takahashi and Yamanaka, 2006). Recent regenerative medicine research has found methods to efficiently generate safer iPSCs (Attwood and Edel, 2019; Cyranoski, 2018; Feng et al., 2009; Li and Li, 2014; Sanal, 2014; Sharma, 2016). Some of these techniques include small molecule facilitation of induced reprogramming which have resulted in more efficient cellular reprogramming (Feng et al., 2009; Ichida et al., 2009; Li et al., 2011; Nie et al., 2012; Shi et al., 2008; Yuan et al., 2011; Zhu et al., 2010).

Previous studies have identified small molecules capable of increasing the efficiency of iPSC generation with transcription factor driven reprogramming methods. There has also been success in using small molecules to replace some transcription factors. However, finding an

efficient small molecule cocktail that can alone efficiently activate reprogramming has been challenging (Li et al., 2009; Nie et al., 2012; Shi et al., 2008; Yuan et al., 2011; Zhou and Ding, 2010; Zhu et al., 2010). Klf4, c-Myc, Oct4, and Sox2 are typically employed in reprogramming, these transcription factors irreversibly affect hundreds of genes. We wanted to examine epigenetic barriers to activation of a key pluripotency factor, *Oct4*. In this study, we performed a screen to identify small molecules that facilitate single allele activation in combination with a single transcriptional activator docked at the chromatin *in vivo* assay at *Oct4* (*CiA:Oct4*) allele. For this study, we chose to utilize a simian virus 40 large T antigen (SVT) infected cell line to immortalize our cells. This method made cells easier to array for a high throughput screen without having to worry about cell density or senescence. Notably, SVT immortalized cells have effectively been used by multiple groups to in regenerative medicine models (Kellermann et al., 1990, 1987; Poliard et al., 1995).

Oct4 expression is highly correlated with iPSC generation and is a key phenotypic indicator of successful iPSC generation (Hathaway et al., 2012; Ichida et al., 2009; Lin and Wu, 2015; Radzisheuskaya and Silva, 2014; Shi and Jin, 2010; Shimozaki et al., 2003; Zeineddine et al., 2014). The Oct4 protein, encoded by the *POU5f1* (POU domain, class 5, transcription factor locus and belonging to the POU (Pit, Oct, Unc)) family, is described as a master pluripotency factor (Zeineddine et al., 2014). *Oct4* expression acts as a gatekeeper, driving molecular signaling cascades which maintain pluripotency in stem cells. *Oct4* is rapidly repressed as cells differentiate during mammalian development (Radzisheuskaya and Silva, 2014; Zeineddine et al., 2014). Hence, *Oct4* is a highly regulated genetic locus. The *Oct4* locus contains a distal enhancer, proximal enhancer, and proximal promoter which are regulated tightly throughout development (Kellner and Kikyo, 2010). Many different factors bind and regulate this locus. Notably, Dnmt3a

and Dnmt3b methylate DNA at all three regulatory regions around the *Oct4* locus and promote silencing of the gene. Additionally, Oct4 can form complexes with Nanog and HDAC2 resulting in silencing of the *Oct4* locus (Liang et al., 2008). High DNA methylation and low histone acetylation are present in somatic cells where *Oct4* has been completely silenced (Kellner and Kikyo, 2010). Fittingly, Azacytidine (DNA methyl transferase inhibitor (DNMTi)), Suberoylanilide Hydroxamic Acid (SAHA) (histone deacetylase inhibitor (HDACi)), and Valproic Acid (VPA) (HDACi) were among the first identified epigenetically relevant small molecules capable of increasing *Oct4* activation during transcription factor induced reprogramming (Feng et al., 2009; Huangfu et al., 2008). Other more recently discovered small molecules, such as Oct4-activating compound 1 (Li et al., 2012), BIX-01294 (Shi et al., 2008), RG108 (Shi et al., 2008), Sodium butyrate (Mali et al., 2010), AM580 (Wang et al., 2011), Tranylcypromine (Li et al., 2009), and DZNep (Hou et al., 2013) increase iPSC generation (Huangfu et al., 2008; Ichida et al., 2009; Nie et al., 2012) and also activate *Oct4* expression during transcription factor induced reprogramming methods. Among these identified small molecules, VPA was considered to be an effective *Oct4* activator under transcription factor induced reprogramming methods, providing a substantial increase in iPSC colony production (Feng et al., 2009).

We have developed a screening strategy using the CiA system in mouse embryonic fibroblast (MEF) cells. The CiA platform is a murine cell line with one *Oct4* allele replaced with an enhanced green fluorescent protein (EGFP) preceded by a Gal4 binding domain to which chromatin modifying machinery can be recruited through direct protein fusions to GAL4 or chemically induced proximity. The other *Oct4* allele in *CiA:Oct4* cells is wild type. From *CiA:Oct4* mice we generated MEF cell lines. We tested access to transcriptional machinery by recruiting a VP16 transcriptional activator to the *CiA:Oct4* locus as a GAL4 fusion protein, and

observed a small amount of *CiA:Oct4* activation (~3% at the timepoint screened) as measured by GFP expression. We then performed a screen with a library of 959 small molecules to identify compounds that enhanced the ability of the tethered transcription factor to activate the C*iA:Oct4* locus. We validated the top small molecule activators from this screen with dose response analysis and compared it to previously described iPSC enhancers VPA, SAHA, and TSA. We found that small molecules identified by our screen outperformed VPA, SAHA, and TSA in single allele *Oct4* gene activation with VP16 recruitment. We then performed single-cell analysis of chosen successful *Oct4* activators for 60 hours following small molecule addition from small molecules DNMTi: Azacytidine and HDACis: Mocetinostat and Entinostat. From this experiment, we found that on a single-cell level, cells spontaneously turn on *CiA:Oct4* resulting in GFP expression that is passed on to daughter cells. Finally, we tested Mocetinostat with traditional four factor reprogramming and found this compound increased iPSC generation efficiency.

Results

Small molecule screen for facilitators of *CiA:Oct4* activation

To identify small molecules targeting epigenetic pathways which restrict efficient activation of the *CiA:Oct4* locus, a high-throughput small molecule screen was performed (Fig. 1). We used an in-house curated set of small molecules with an epigenetic-targeted compound library (EpiG library), which contained a set of 959 small molecules. Some molecules are well characterized with known targets, others are derivatives from molecules that contain scaffolds similar to epigenetic inhibitors. This screen was performed with recruitment of the transcriptional activator VP16 or with a Gal4-DNA binding protein alone as a control.

Cells were infected with a Gal4-VP16 lentivirus and selected with puromycin. Compounds were added at 10μM to cells on Day 0 and gene activation was measured by high-throughput flow cytometry after four days of compound treatment. (Fig. 2.1A). As a counter screen we used a lentiviral infection of a Gal4 protein alone without any transcriptional activation component (Fig. 2.S1A). Flow cytometry readings for both screens were gated as indicated (Fig. 2.S1B). Compounds were considered "hits" when greater than five percent of cells activated GFP. Compounds with high background fluorescence in the Gal4 counter screen were removed. The top 23 small molecule activators were rescreened for validation with a sequential dose curve treatment with concentrations ranging from 10μM to 0.3μM (Fig. 2.S2A). Flow cytometry gating was performed as indicated (Fig. 2.S2B). Based on dose response data, five small molecules were chosen for further analysis for activation of the *CiA:Oct4* locus including: Mocetinostat, Droxinostat, Entinostat, Tacedinaline, and Azacytidine. Azacytidine is a known potent DNMTi previously identified for increasing *Oct4* activation during transcription factor reprogramming conditions. Intriguingly, Mocetinostat Tacedineline and Entinostat all target HDAC -1, -2, and -3

(Supplemental Table 2). The identification of HDAC inhibitors and DNMTi *Oct4* activators reinforced the importance of histone acetylation and DNA methylation on maintenance of chromatin state at the *Oct4* locus. It is important to note that although this study exclusively monitors *Oct4* expression, the four small molecules detailed in this study have widespread transcriptional perturbations which have been extensively documented in literature (Bijangi-Vishehsaraei et al., 2010; Cai et al., 2015; Delcuve et al., 2013; Fournel et al., 2008; Haberland et al., 2009; Lauffer et al., 2013; Liu et al., 2016; Loprevite et al., 2005; LoRusso et al., 1996; McCourt et al., 2012; Moradei et al., 2007; Píškala et al., 1981; Rosato et al., 2003; Saito et al., 1999; Wood et al., 2010; Yu et al., 2015). It is also possible that the facilitation of *Oct4* activation examined results from indirect effects of these inhibitors.

Validation of lead molecules

Hit compounds were validated and optimal compound concentration for gene activation was examined by a second round of dose response test on *CiA:Oct4* MEF cells (Fig. 2.2). To track the amount of gene activation and gain knowledge of cell transduction rates, *CiA:Oct4* MEF cells were infected with a lentiviral construct containing a Histone H2B monoCherry (H2B-mCh) tracer with a self-cleaving P2A peptide separating either a Gal4-VP16 or a Gal4-only control (Fig. 2.1A, gated as shown in Fig. 2.S3). GFP and mCh were visualized using flow cytometry four days later. Cells were fluorescence gated and the mCh positive cells were evaluated for GFP level as indicated (Fig. 2.S3). Since only mCh cells are considered in this analysis the activation rates are higher as cells with lower transduction expression are excluded. For comparison, the Gal4-H2B-mCh-VP16 infected cells showed an average activation of 12% with a standard deviation of 2.8. Mocetinostat demonstrated 29% *CiA:Oct4* activation at 0.625μM. Tacedinaline demonstrated 20% *CiA:Oct4*

activation at 10μM. Entinostat demonstrated 32% *CiA:Oct4* activation at 0.312 μM. Azacytidine demonstrated the most effective activation at 5μM (57%), but 2.5μM treated cells had better cell morphology by microscope analysis and still had 45% activation. Droxinostat did not demonstrate significant activation following rescreening and was removed from further study.

Interestingly, four of the five small molecules (Mocetinostat, Tacedinaline, Entinostat, and Azacytidine) identified by this screen were more effective than VPA, SAHA, and TSA in single allele *CiA:Oct4* activation. In this assay, the activation in the presence of VPA treatment was not significant. This could be due to moderate cell death we observed in the presence of VPA (data not shown). Likewise, SAHA demonstrated no significant activation while TSA allowed for mild increased activation at a dose of 0.08μM (16% *CiA:Oct4* activation). It should be noted that the time frame of our analysis was much shorter than the time frame of whole cell reprogramming, and the barriers of single allele activation may be different than network activation by transcription factor cocktails.

Temporal analysis of chemical facilitated *CiA:Oct4* activation

To understand the dynamics of small molecule facilitated gene activation by a directed transcription factor in a population of cells, gene activation was monitored by time-lapse microscopy and flow cytometry over 70 hours following small molecule treatment as indicated (Fig. 2.3A). Digital analysis of images was used to identify total cell population in a frame of view and then to count GFP positive cells (Supplemental text and Fig. 2.S4). This approach of monitoring gene activation in live cells allowed us to identify key transformation points in allele activation. We found that Entinostat and Mocetinostat accelerated transcription factor driven gene

activation, with activation peaks detected by 30 hours (Fig. 2.3B). Azacytidine showed slower gene activation from hours 0-30 (Fig. 2.3B), while rapid gene activation from hours 30-60 and peak activation at hour 70. These results suggested that HDAC inhibition results the facilitation of early *CiA:Oct4* activation; however, at later time points some effects are lost. Comparably, Azacytidine resulted in slow and constant triggered activation in conjunction with tethered transcriptional machinery. To further understand the durability of small molecule effects on activation of the *Oct4* locus by transcription activator docking, cells were treated with compound for four days then released for four days by washout of small molecule (Fig. 2.4A). We found cells with higher transcriptional activator driven expression from HDACi treatment rapidly lost gene activation after four days of HDACi washout. Comparatively, cells treated with Azacytidine and directed transcriptional activator maintained higher levels gene activation even after four days of small molecule release (Fig. 2.4B).

Single cell analysis of chemical facilitated *CiA:Oct4* activation

To study *CiA:Oct4* activation response on a single-cell level to transcriptional activator tethering in conjunction with HDACi and DNMTi treatment, cells were tracked through the H2B-mCh tracer and single-cell nuclear GFP intensity was quantified at each time point. We found that *CiA:Oct4* nuclear GFP average mean intensity increased at different rates in individual cells tracked. However, there was a clear difference in the stimulated activation between control cells and small molecule treated cells. Untreated control cells had gradual expression changes in general while small molecule treated samples demonstrated spontaneous rapid allele activation. A common theme throughout both control and small molecule treated cells was that daughter cells tended to maintain parental expression patterns after cell divisions. Namely, cells that were GFP

negative tended to stay GFP negative and cells that were GFP positive tended to have progeny that were also GFP positive (Fig. 2.5). These findings are consistent with the model where *Oct4* expression is driven by the expression of the *Oct4* alleles passed down from parental cells (Wolff et al., 2018). As a control, the expression of GFP compared to nuclear mCh expression was also tracked (Fig. 2.S5). In conclusion, these results lead us to believe treatment with Mocetinostat and/or Azacytidine are the most effective compounds among those tested to facilitate *Oct4* activation by transcriptional activators.

Small molecule effects on cell cycle and viability

To understand the effect of small molecules on cell cycle, we used a standard propidium iodide staining assay to measure total DNA content per cell. To understand and effects on cell viability, we performed an alamarBlue assay which measures metabolically active live cells. We treated cells for five days with small molecule as indicated (Fig. 2.S6A). On the fourth day all cell wells were split to ensure logarithmic growth at our assay point. On the fifth day, both cell viability and cell cycle analyses were conducted. It was found that cell viability was not changed in Mocetinostat, Entinostat, and Tacedinealine at optimal treatment concentrations from our dose response analysis, while Azacytidine and VPA standard treatment resulted in measurable cell

cytotoxicity (Fig S6B). We did not determine any large perturbations to the cell cycle upon propidium iodide staining (Fig. 2.S6C, Gated in Fig. 2.S6D).

Mocetinostat increases *CiA:Oct4* activation during transcription factor reprogramming.

As a final test to see if molecules identified by our single allele activation method could help advance cell reprogramming techniques, we compared Mocetinostat identified here with Azacytidine and generated iPSC by 4-factor reprogramming. We infected *CiA:Oct4* MEFs with a polycistronic vector containing Oct4, Sox2, Klf4, and cMyc separated by self-cleaving peptides with a tetracycline inducible promoter system (Carey et al., 2009). (Fig. 2.6A). We found that Mocetinostat increased activation of Oct4-GFP, a phenotypic indicator of cell reprogramming to 22% GFP+ (Fig. 2.6B, Fig. 2.S7D). The control (Doxycycline treated cells without small molecule addition) demonstrated a lower level of GFP expression. Notably, the addition of small molecules to all polycistronic vector infected cells resulted in small activation potentially from the small molecules facilitate overcoming doxycycline control of the four-factor cassette (Fig. 2.S7C). We confirmed successful iPSC colony generation through alkaline phosphatase staining and morphological changes which resembled iPSC colonies (Fig. 2.S7B, Fig. 2.S8).

Discussion

We determined from this small molecule screen and follow-up studies that four compounds (Mocetinostat, Tacedinaline, Entinostat, and Azacytidine) demonstrated robust and reproducible single *CiA:Oct4* allele activation when used in conjunction with transcriptional activator recruitment. In ideal conditions, Azacytidine demonstrated a ~60% *CiA:Oct4* activation, which is the highest change in Oct4-GFP expression recorded in a population of cells due to a single transcription factor and small molecule combination acting on *Oct4*. Interestingly, of the top five small molecule activators from the original screen, four were HDAC inhibitors and the top hit is a previously described *Oct4* activator and DNA methylation inhibitor, Azacytidine (Huangfu et al., 2008). This reinforces previous findings that DNA methylation and histone acetylation play major roles in determining *Oct4* expression levels. But also adds new classes of HDAC inhibitors that should be further examined in iPSC generation work. Notably, Mocetinostat, Tacedinaline, Entinostat, and Azacytidine outperformed TSA, SAHA, and VPA suggesting that single allele activation may not have the same requirements as whole cell network transcription factor reprogramming conditions.

We were able to further reveal gene activation dynamics through our small molecule treatment and release study (Fig. 2.4B). We found that HDACi resulted in rapid gene activation which was rapidly lost upon small molecule release. Comparatively, DNMTi resulted in slower gene activation which was maintained even after the small molecule was removed from the system. We believe HDAC inhibition resulted in rapid reversible gene activation while DNMTi resulted in slow and more static gene activation. Previous studies have supported the idea that loss of histone acetylation results in reversible epigenetic memory, while DNA methylation accumulation results in irreversible epigenetic memory (Bintu et al., 2016). Our study

demonstrates that the other side of the model is true as well; it supports a model through which histone acetylation accumulation results in rapid and reversible gene activation, while DNA de-methylation results in irreversible gene activation. Finally, we demonstrated that one small molecule identified by this screen, Mocetinostat, lead to a 22% of *CiA:Oct4* activation at an early timepoint in iPSC generation. Our work indicates that Mocetinostat could be a strong candidate for future small molecule facilitated iPSC generation studies.

Conclusion

In conclusion, we identified the following small molecules: Azacytidine, Mocetinostat, Tacedinaline, and Entinostat which stimulated high single allele *Oct4* activation when combined with the directed recruitment of transcriptional machinery. Our results provide for a robust epigenetic screen for endogenous single allele *Oct4* activation chemical enhancers combining a directed transcription factor and small molecule. Additionally, we demonstrated dynamics of *Oct4* single allele activation through treatment using HDACi or DNMTi pathways. We found that HDAC inhibition seemed to result in primary peak activation occurring by 30 hours while DNMT inhibition resulted in gradual activation with peak activation by hour 60. Interestingly, DNMT inhibition resulted in activation that was sustained even after four days release of small molecules, while HDAC inhibition resulted in activation that was almost completely lost after four days. This demonstrated models of epigenetic memory where histone acetylation levels are more dynamic than DNA methylation levels and can result in corresponding more dynamic activation with histone acetylation accumulation compared to slower DNA methylation loss. We further found exploration of *CiA:Oct4* MEFs expression on a single-cell level revealed that *Oct4* activation was spontaneous throughout the experiment and active *CiA:Oct4* expression state can be stably passed through cellular generations. Finally, we found that the small molecule Mocetinostat identified in this study was successful in increasing iPSC generation.

Methods

Generation of *CiA:Oct4* SVT-MEFs

CiA:Oct4 MEF cell lines immortalized by infection of simian virus 40 large T antigen, were obtained and cultured as previously described (Hathaway et al., 2012). Briefly, cells were cultured at 37ºC 5% CO_2 conditions. Base media was either FluoroBrite DMEM Media (ThermoFisher, A1896701) for imaging, or DMEM (Corning, MT10013CV) for standard cell culture. Media was supplemented with 10% FBS (Gibco, Lot:1972526), 10mM HEPES pH 7.5, 10mM NEAA, 0.1% 1000X 2-betamercaptoethanol (Gibco, 21985023), 1% 100X Penn-Strep (Corning, 30-002-CI). Additionally, L-Glutamine (Corning, 25005CI) at 4mM was added to FluoroBrite media.

Description of plasmids:

nLV-EF-1a-Gal4-VP16-PGK-Puro (N114, Addgene, Plasmid #44014) and nLV-EF-1a-Gal4-Stop-PGK-Puro (N113, Addgene, Plasmid #44176) were previously described.

nLV-EFn-1a-Gal4-VP16-P2A-H2B-mCh-PGK-Puro (K114mC) was developed by a PCR stitching Gal4-VP16-P2A P2A-H2B-mCh and in fusion cloning the product into a NotI linearized nLV-Dual Promoter EF-1a-MCS-PGK-Puro (N103) using In-fusion HD cloning kit (Clontech). Plasmid and plasmid map are available on Addgene: TetO-FUW-OKSIM (Addgene, Plasmid #20321) and FUW-M2rtTA (Addgene, Plasmid # 20342).

Lentiviral infection of *CiA:Oct4* SVT-MEFs

15 million 293T lentiX cells (Clontech) were co-transfected with gene delivery vector (N114, K114mC, or N113) and packaging vectors pspax2 (Addgene, Plasmid #12260) and pMD2.G (Addgene, Plasmid # 12259) with PEI (Polysciences Inc, 24765) and cultured for 48 hours to produce lentivirus. Lentivirus was pelleted via ultracentrifugation with a Beckman SW32Ti rotor a ~72,000 xg and resuspended in 150uL PBS. 60,000 *CiA:Oct4* MEFs were infected with 30uL of concentrated lentivirus. Puromycin selection of MEF cells was performed at a concentration of 2.5 ug/ml.

Small molecule screen

EpiG set of three 384-well compound plates was used in assay, compounds were screened at 10μM. *CiA:Oct4* MEFs were cultured in standard conditions then infected with lentivirus (N114, N113) and treated with small molecules for four days at 10μM. Screens were performed in three separate screens. Cells were analyzed by Flow Cytometry on the iQue Screener Plus. Analysis gating was performed using FlowJo as indicated (Fig. 2.S1B).

Dose-response of small molecule treatment

CiA:Oct4 MEFs were cultured in standard conditions then infected with lentivirus (N114, N113, K114mC) and treated with small molecules for four days in a dose dependent manner and then released from small molecule treatment for four days. The small molecule treatment on the *CiA:Oct4* MEFs were dosed as follows: Droxinostat (10μM, 5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM**, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, 0.005μM). Mocetinostat (1.25μM, 0.625μM, 0.3125μM, 0.1256μM. 0.08μM, 0.04μM**, 0.20μM, 0.01μM, 0.005μM,

0.002μM, 0.001μM, 0.0006μM), Tacedinaline (10μM, 5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM**, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, 0.005μM) Entinostat (2.5μM, 1.25μM, 0.625μM, 0.3125μM, 0.1256μM, 0.08μM**, 0.04μM, 0.02μM, 0.01μM, 0.005μM, 0.002μM, 0.001μM), Azacytidine (10μM, 5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM**, 0.156μM, 0.078μM, 0.039μM, 0.019μM,0.010μM, 0.005μM), TSA (0.16μM, 0.08μM, 0.04μM, 0.02μM, 0.01μM, 0.005μM**, 0.0025μM, 0.0013μM, 0.0006μM, 0.0003μM, 0.0002μM, 0.0002μM) VPA (5000μM, 2500μM, 1250μM, 625μM, 312.5μM, 156.25μM**, 78.12μM, 39.06μM, 19.5μM, 9.7μM, 4.8μM, 2.4μM) SAHA (10μM, 5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM**, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, 0.005μM). (n≥3 except at indicated ** where n=2) Cells were imaged by the IN Cell Analyzer 2200 on Day 4 and Day 8 following lentiviral infection. Cells were analyzed by Flow Cytometry on the iQue Screener Plus. Analysis gating was performed using FlowJo as indicated (Fig. 2.S3).

Small molecule time-lapse imaging

CiA:Oct4 MEFs were cultured in standard conditions then infected with lentivirus (N114, N113, K114mC) and treated with small molecules for four days in a dose dependent manner. The small molecule treatment dosage was follows: Droxinostat (10μM, 5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM, 0.156μM, 0.078μM, 0.039μM, 0.019μM, and 0.010μM) Mocetinostat (5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, and 0.005μM), Tacedinaline (5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, and 0.005μM), Entinostat (5μM, 2.5μM. 1.25μM, 0.625μM, 0.312μM, 0.156μM, 0.078μM, 0.039μM, 0.019μM, 0.010μM, and 0.005μM). *CiA:Oct4* SVT-

MEFs were imaged every two hours after 24 hours (for 14 hours) and after 48 hours (for 14 hours) by the GE IN Cell Analyzer, as well as once every 24 hours.

Single-cell analysis

CiA:Oct4 MEFs were cultured in standard conditions and treated with small molecules for four days in a dose dependent manner. The small molecule treatment dosage was follows: 2.5µM for Azacytidine, 630nM Entinostat and 80nM Mocetinostat. *CiA:Oct4* SVT-MEFs were imaged every 35 minutes from hours 0 to 60 by the GE IN Cell Analyzer. Scale bar in videos is 50um. (Supplemental Videos) Cells were segmented, tracked and annotated in a semi-automatic way as described previously (Borland et al., 2018) using a set of scripts developed in Fiji (Schindelin et al., 2012). GFP (Oct4) and H2B-mCherry signals were calculated as a mean value of pixels within defined nuclear regions. Family trees were rendered using EteToolkit library (Huerta-Cepas et al., 2016) in Python 4.5.4 Anaconda (Anaconda, 2016). Cell death rate was calculated as a ratio of tracks ending in cell death to all possible track endings, namely: end of the experiment, cell leaving a field of view, mitosis or cell death.

Microscope image acquisition

IN Cell Analyzer 2200: Chip type front illuminated sCMOS, Chip size 2560x2160 pixels. Pixel size 6.5µm. Readout speeds 95MHz, 286Mhz, Readout modes Rolling shutter, global shutter. Camera interface Camera-link. Bit depth 15 bit. Quantum efficiency ~60% dynamic range 1:15,000. Read noise 1.5 e at 33 fps 2e at 100fps. Magnification (20X objective) IN Cell Analyzer 2200 software for acquisition and IN Cell Developer for image processing. Pictures of cells were

taken at 37 degrees Celsius in FluoroBrite media. Images were taken with the FITC 525, Brightfield, and Cy3 filters. Images were taken in 2-D imaging setting.

Cell viability/proliferation

CiA:Oct4 MEFs were cultured in standard conditions and treated with small molecules for five days in a dose dependent manner as indicated (Fig. 2.S6A). High, Medium, and Low treatment conditions are as follows:. Mocetinostat (High = 1.6μM, Medium = 0.16μM, Low = 0.05μM, n=8), Tacedinaline (High = 30μM, Medium = 10μM, Low = 3.3μM, n=8), Entinostat (High = 12.5μM, Medium = 1.25μM, Low =0.42μM, n=8), DMSO (n=28), Azacytidine (High = 25μM, Medium = 2.5μM, Low = 0.8μM, n=8), VPA (High =6000μM, Medium = 2000μM, Low = 667μM, n=8)). Cells were split on day four to 10,000 cells/ml. alamarBlue reagent (Cat # DAL1025) was added on Day 5 to 10% of well volume with standard conditions and incubated for 16 hours before visualization on the GloMax Discover Serial Number 9700000261 and Software Version 3.0.0.

Cell cycle analysis

CiA:Oct4 MEFs were cultured in standard conditions and treated with small molecules for five days in a dose dependent manner as indicated in (Fig. 2.S6A). Treatment conditions are as follows: Mocetinostat (0.16μM, n=7), Tacedinaline (10μM, n=4), Entinostat (0.42μM, n=3), DMSO (n=10), Azacytidine (2.5μM, n=2), VPA (2000μM, n=3)). To stain for cell cycle phases, a BrDU/Propidium Iodide assay was performed using the EZ-BrDU Kit (TNB-6600) and the kit's provided protocol. Gating of cells was performed as indicated (Fig. 2.S6D).

Induction of pluripotent stem cells with small molecule treatment

CiA:Oct4 SVT-MEF cells were infected with TetO-FUW-OKSIM and FUW-M2rtTA on Day -15 as indicated in Fig. 2.7A. On Day 0, cells were treated with either DMSO (Control), 2.5μM Azacytidine, or 156nM Mocetinostat accompanied with (Fig. 2.7B) or without Doxycycline (Fig. 2.S7C). Flow cytometry was performed on Day 4. Cells cultured for longer than four days were treated with small molecules alternating on and off every 2-3 days. Gating strategy is demonstrated in Supplementary Fig. 2.7C. Imaging of cells for Fig. 2.7A was performed on Day 4. (Fig. 2.S7A). Alkaline phosphatase staining was performed with Reprocell Alkaline Phosphate Staining Kit (Cat # NC0088922). Alkaline phosphatase staining was performed at various times ranging from 20-60 days after infection as indicated in figures.

Acknowledgements

We would like to thank the members of the Hathaway lab and the Center for Integrative Chemical Biology and Drug Discovery for helpful advice and technical assistance, especially S. Frye, K. Pearce, L. James, and B. Hardy. The authors thank B. Price for data visualization with Python Programming. We thank the Flow Cytometry Core University of North Carolina funded by P30 CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center. We thank P. Vignaux for a careful editing of the paper. We also thank S. Coquery and J. Dow for technical assistance with flow cytometry. This work was in part supported by grants from the U.S. National Institutes of Health, Grant R01GM118653 (to N.A.H.) and Grant DP2-HD091800-01 (to J.E.P.). K.M.H was supported National Science Foundation Grant No. DGE-1650116 and by a training grant from NIGMS under award T32 GM119999. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Anaconda, 2016. Anaconda Software Distribution. Comput. Softw.

Attwood, S.W., Edel, M.J., 2019. iPS-Cell Technology and the Problem of Genetic Instability-Can It Ever Be Safe for Clinical Use? J. Clin. Med. 8, 288. https://doi.org/10.3390/jcm8030288

Bijangi-Vishehsaraei, K., Huang, S., Safa, A.R., Saadatzadeh, M.R., Murphy, M.P., 2010. 4-(4-Chloro-2-methylphenoxy)-N-hydroxybutanamide (CMH) targets mRNA of the c-FLIP variants and induces apoptosis in MCF-7 human breast cancer cells. Mol. Cell. Biochem. 342, 133–142. https://doi.org/10.1007/s11010-010-0477-7

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., Elowitz, M.B., 2016. Dynamics of epigenetic regulation at the single-cell level. Science. 351, 720–724. https://doi.org/10.1126/science.aab2956

Bongso, A., Richards, M., 2004. History and perspective of stem cell research. Best Pract. Res. Clin. Obstet. Gynaecol. 18, 827–842. https://doi.org/10.1016/j.bpobgyn.2004.09.002

Borland, D., Yi, H., Grant, G.D., Kedziora, K.M., Chao, H.X., Haggerty, R.A., Kumar, J., Wolff, S.C., Cook, J.G., Purvis, J.E., 2018. The Cell Cycle Browser: An Interactive Tool for Visualizing, Simulating, and Perturbing Cell-Cycle Progression. Cell Syst. 7, 180–184.e4. https://doi.org/10.1016/j.cels.2018.06.004

Cai, J., Zhang, Q., Lin, K., Hu, L., Zheng, Y., 2015. The Effect of MGCD0103 on CYP450 Isoforms Activity of Rats by Cocktail Method. Biomed Res. Int. 2015, 517295. https://doi.org/10.1155/2015/517295

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., Jaenisch, R., 2009. Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. 106, 157–162. https://doi.org/10.1073/pnas.0811426106

Cyranoski, D., 2018. 'Reprogrammed' stem cells to be tested in people with Parkinson's disease. Nature. https://doi.org/10.1038/d41586-018-07407-9

Delcuve, G.P., Khan, D.H., Davie, J.R., 2013. Roles of histone deacetylases in epigenetic regulation: Emerging paradigms from studies with inhibitors, in: Epigenetics and Pathology: Exploring Connections between Genetic Mechanisms and Disease Expression. Apple Academic Press, pp. 143–171. https://doi.org/10.1201/b16304

Feng, B., Ng, J.H., Heng, J.C.D., Ng, H.H., 2009. Molecules that Promote or Enhance

Reprogramming of Somatic Cells to Induced Pluripotent Stem Cells. Cell Stem Cell. 4, 301-302. https://doi.org/10.1016/j.stem.2009.03.005

Fournel, M., Bonfils, C., Hou, Y., Yan, P.T., Trachy-Bourget, M.-C., Kalita, A., Liu, J., Lu, A.-H., Zhou, N.Z., Robert, M.-F., Gillespie, J., Wang, J.J., Ste-Croix, H., Rahil, J., Lefebvre, S., Moradei, O., Delorme, D., Macleod, A.R., Besterman, J.M., Li, Z., 2008. MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo. Mol. Cancer Ther. 7, 759–68. https://doi.org/10.1158/1535-7163.MCT-07-2026

Haberland, M., Montgomery, R.L., Olson, E.N., 2009. The many roles of histone deacetylases in development and physiology: Implications for disease and therapy. Nat. Rev. Genet. 1, 32-42. https://doi.org/10.1038/nrg2485

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., Crabtree, G.R., 2012. Dynamics and memory of heterochromatin in living cells. Cell 149, 1447–1460.

Hou, P., Li, Y., Zhang, X., Liu, C., Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., Ge, J., Xu, J., Zhang, Q., Zhao, Y., Deng, H., 2013. Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science. 341, 651–654. https://doi.org/10.1126/science.1239278

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., Melton, D. a, 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat. Biotechnol. 26, 795–7.

Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33, 1635–1638. https://doi.org/10.1093/molbev/msw046

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., Huangfu, D., Akutsu, H., Liu, D.R., Rubin, L.L., Eggan, K., 2009. A Small-Molecule Inhibitor of Tgf-β Signaling Replaces Sox2 in Reprogramming by Inducing Nanog. Cell Stem Cell 5, 491–503.

Kastenberg, Z.J., Odorico, J.S., 2008. Alternative sources of pluripotency: science, ethics, and stem cells. Transplant. Rev. 22, 215–222. https://doi.org/10.1016/j.trre.2008.04.002

Kellermann, O., Buc-Caron, M.H., Gaillard, J., 1987. Immortalization of precursors of endodermal, neuroectodermal and mesodermal lineages, following the introduction of the simian virus (SV40) early region into F9 cells. Differentiation 35, 197–205. https://doi.org/10.1111/j.1432-0436.1987.tb00169.x

Kellermann, O., Buc-Caron, M.H., Marie, P.J., Lamblin, D., Jacob, F., 1990. An immortalized osteogenic cell line derived from mouse teratocarcinoma is able to mineralize in vivo and in vitro. J. Cell Biol. 110, 123–132. https://doi.org/10.1083/jcb.110.1.123

Kellner, S., Kikyo, N., 2010. Transcriptional regulation of the Oct4 gene, a master gene for pluripotency. Histol. Histopathol. 25, 405–412.

Lauffer, B.E.L., Mintzer, R., Fong, R., Mukund, S., Tam, C., Zilberleyb, I., Flicke, B., Ritscher, A., Fedorowicz, G., Vallero, R., Ortwine, D.F., Gunzner, J., Modrusan, Z., Neumann, L., Koth, C.M., Lupardus, P.J., Kaminker, J.S., Heise, C.E., Steiner, P., 2013. Histone Deacetylase (HDAC) Inhibitor Kinetic Rate Constants Correlate with Cellular Histone Acetylation but Not Transcription and Cell Viability. J. Biol. Chem. 288, 26926–26943. https://doi.org/10.1074/jbc.M113.490706

Li, S., Li, Q., 2014. A promising approach to iPSC-based cell therapy for diabetic wound treatment: Direct lineage reprogramming. Mol. Cell. Endocrinol. 393, 8–15. https://doi.org/10.1016/j.mce.2014.05.025

Li, W., Tian, E., Chen, Z.-X., Sun, G., Ye, P., Yang, S., Lu, D., Xie, J., Ho, T.-V., Tsark, W.M., Wang, C., Horne, D.A., Riggs, A.D., Yip, M.L.R., Shi, Y., 2012. Identification of Oct4-activating compounds that enhance reprogramming efficiency. Proc. Natl. Acad. Sci. 109, 20853–20858. https://doi.org/10.1073/pnas.1219181110

Li, W., Zhou, H., Abujarour, R., Zhu, S., Joo, J.Y., Lin, T., Hao, E., Schöler, H.R., Hayek, A., Ding, S., 2009. Generation of human-induced pluripotent stem cells in the absence of exogenous Sox2. Stem Cells 27, 2992–3000. https://doi.org/10.1002/stem.240

Li, Y., Zhang, Q., Yin, X., Yang, W., Du, Y., Hou, P., Ge, J., Liu, C., Zhang, W., Zhang, X., Wu, Y., Li, H., Liu, K., Wu, C., Song, Z., Zhao, Y., Shi, Y., Deng, H., 2011. Generation of iPSCs from mouse fibroblasts with a single gene, Oct4, and small molecules. Cell Res. 21, 196–204. https://doi.org/10.1038/cr.2010.142

Liang, J., Wan, M., Zhang, Y., Gu, P., Xin, H., Jung, S.Y., Qin, J., Wong, J., Cooney, A.J., Liu, D., Songyang, Z., 2008. Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. Nat. Cell Biol. 10, 731–739. https://doi.org/10.1038/ncb1736

Lin, T., Wu, S., 2015. Reprogramming with small molecules instead of exogenous transcription factors. Stem Cells Int. 2015, 1–11. https://doi.org/10.1155/2015/794632

Liu, J., Li, G., Wang, X., Wang, L., Zhao, R., Wang, J., Kong, Y., Ding, J., Li, J., Zhang, L., 2016. Droxinostat, a Histone Deacetylase Inhibitor, Induces Apoptosis in Hepatocellular Carcinoma Cell Lines via Activation of the Mitochondrial Pathway and Downregulation of

FLIP. Transl. Oncol. 9, 70–78. https://doi.org/10.1016/j.tranon.2016.01.004

Loprevite, M., Tiseo, M., Grossi, F., Scolaro, T., Semino, C., Pandolfi, A., Favoni, R., Ardizzoni, A., 2005. In vitro study of CI-994, a histone deacetylase inhibitor, in non-small cell lung cancer cell lines. Oncol. Res. 15, 39–48.

LoRusso, P.M., Demchik, L., Foster, B., Knight, J., Bissery, M.C., Polin, L.M., Leopold, W.R., Corbett, T.H., 1996. Preclinical antitumor activity of CI-994. Invest. New Drugs 14, 349–56.

Mali, P., Chou, B.K., Yen, J., Ye, Z., Zou, J., Dowey, S., Brodsky, R.A., Ohm, J.E., Yu, W., Baylin, S.B., Yusa, K., Bradley, A., Meyers, D.J., Mukherjee, C., Cole, P.A., Cheng, L., 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells 28, 713–720. https://doi.org/10.1002/stem.402

Mao, A.S., Mooney, D.J., 2015. Regenerative medicine: Current therapies and future directions. Proc. Natl. Acad. Sci. 112, 14452–14459. https://doi.org/10.1073/pnas.1508520112

McCourt, C., Maxwell, P., Mazzucchelli, R., Montironi, R., Scarpelli, M., Salto-Tellez, M., O'Sullivan, J.M., Longley, D.B., Waugh, D.J.J., 2012. Elevation of c-FLIP in castrate-resistant prostate cancer antagonizes therapeutic response to androgen receptor-targeted therapy. Clin. Cancer Res. 18, 3822–33. https://doi.org/10.1158/1078-0432.CCR-11-3277

Medvedev, S.P., Shevchenko, A.I., Zakian, S.M., 2010. Induced Pluripotent Stem Cells: Problems and Advantages when Applying them in Regenerative Medicine. Acta Naturae 2, 18–28.

Mitchell, R.R., Szabo, E., Benoit, Y.D., Case, D.T., Mechael, R., Alamilla, J., Lee, J.H., Fiebig-Comyn, A., Gillespie, D.C., Bhatia, M., 2014. Activation of neural cell fate programs toward direct conversion of adult human fibroblasts into tri-potent neural progenitors using OCT-4. Stem Cells Dev. 23, 1937–46.

Moradei, O.M., Mallais, T.C., Frechette, S., Paquin, I., Tessier, P.E., Leit, S.M., Fournel, M., Bonfils, C., Trachy-Bourget, M.-C., Liu, J., Yan, T.P., Lu, A.-H., Rahil, J., Wang, J., Lefebvre, S., Li, Z., Vaisburg, A.F., Besterman, J.M., 2007. Novel aminophenyl benzamide-type histone deacetylase inhibitors with enhanced potency and selectivity. J. Med. Chem. 50, 5543–6. https://doi.org/10.1021/jm701079h

Nie, B., Wang, H., Laurent, T., Ding, S., 2012. Cellular reprogramming: A small molecule perspective. Curr. Opin. Cell Biol. 6, 784-792. https://doi.org/10.1016/j.ceb.2012.08.010

Olson, J.L., Atala, A., Yoo, J.J., 2011. Tissue Engineering: Current Strategies and Future

Directions. Chonnam Med. J. 47, 1. https://doi.org/10.4068/cmj.2011.47.1.1

Píškala, A., Hanna, N.B., Buděsínský, M., Cihák, A., Veselý, J., 1981. Synthesis and biological activity of 2'-deoxy-6-methyl-5-azacytidine and its alpha-D-anomer. Nucleic Acids Symp. Ser. 83–6.

Poliard, A., Nifuji, A., Loric, S., Lamblin, D., Launay, J.M., Kellermann, O., Kellermann, O., 1995. Immortalization of committed precursor cells from mouse teratocarcinoma using an adenovirus-SV40 recombinant plasmid. Methods Cell Sci. 17, 103–109. https://doi.org/10.1007/BF00986658

Radzisheuskaya, A., Silva, J.C.R., 2014. Do all roads lead to Oct4? The emerging concepts of induced pluripotency. Trends Cell Biol. 5, 275-284.

Rosato, R.R., Almenara, J.A., Grant, S., 2003. The histone deacetylase inhibitor MS-275 promotes differentiation or apoptosis in human leukemia cells through a process regulated by generation of reactive oxygen species and induction of p21CIP1/WAF1 1. Cancer Res. 63, 3637–45.

Saito, A., Yamashita, T., Mariko, Y., Nosaka, Y., Tsuchiya, K., Ando, T., Suzuki, T., Tsuruo, T., Nakanishi, O., 1999. A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. Proc. Natl. Acad. Sci. U. S. A. 96, 4592–7.

Sanal, M.G., 2014. A highly efficient method for generation of therapeutic quality human pluripotent stem cells by using naive induced pluripotent stem cells nucleus for nuclear transfer. SAGE open Med. 2, 2050312114550375. https://doi.org/10.1177/2050312114550375

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: An open-source platform for biological-image analysis. Nat. Methods. https://doi.org/10.1038/nmeth.2019

Sharma, R., 2016. iPS Cells—The Triumphs and Tribulations. Dent. J. 4, 19. https://doi.org/10.3390/dj4020019

Shi, G., Jin, Y., 2010. Role of Oct4 in maintaining and regaining stem cell pluripotency. Stem Cell Res. Ther. 1, 39.

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Schöler, H.R., Ding, S., 2008. Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds. Cell Stem Cell 3, 568–574. https://doi.org/10.1016/j.stem.2008.10.004

Shimozaki, K., Nakashima, K., Niwa, H., Taga, T., 2003. Involvement of Oct3/4 in the enhancement of neuronal differentiation of ES cells in neurogenesis-inducing cultures. Development 130, 2505–12.

Takahashi, K., Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Tian, C., Ambroz, R.J., Sun, L., Wang, Y., Ma, K., Chen, Q., Zhu, B., Zheng, J.C., 2012. Direct conversion of dermal fibroblasts into neural progenitor cells by a novel cocktail of defined factors. Curr. Mol. Med. 12, 126–37.

Walia, B., Satija, N., Tripathi, R.P., Gangenahalli, G.U., 2012. Induced Pluripotent Stem Cells: Fundamentals and Applications of the Reprogramming Process and its Ramifications on Regenerative Medicine. Stem Cell Rev. Reports. 8, 100-115. https://doi.org/10.1007/s12015-011-9279-x

Wang, W., Yang, J., Liu, H., Lu, D., Chen, X., Zenonos, Z., Campos, L.S., Rad, R., Guo, G., Zhang, S., Bradley, A., Liu, P., 2011. Rapid and efficient reprogramming of somatic cells to induced pluripotent stem cells by retinoic acid receptor gamma and liver receptor homolog 1. Proc. Natl. Acad. Sci. 108, 18283–18288. https://doi.org/10.1073/pnas.1100893108

Wolff, S.C., Kedziora, K.M., Dumitru, R., Dungee, C.D., Zikry, T.M., Beltran, A.S., Haggerty, R.A., Cheng, J., Redick, M.A., Purvis, J.E., 2018. Inheritance of OCT4 predetermines fate choice in human embryonic stem cells. Mol. Syst. Biol. 14, e8140. https://doi.org/10.15252/msb.20178140

Wood, T.E., Dalili, S., Simpson, C.D., Sukhai, M.A., Hurren, R., Anyiwe, K., Mao, X., Suarez Saiz, F., Gronda, M., Eberhard, Y., MacLean, N., Ketela, T., Reed, J.C., Moffat, J., Minden, M.D., Batey, R.A., Schimmer, A.D., 2010. Selective inhibition of histone deacetylases sensitizes malignant cells to death receptor ligands. Mol. Cancer Ther. 9, 246–56. https://doi.org/10.1158/1535-7163.MCT-09-0495

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., Qi, L.S., 2015. Small molecules enhance crispr genome editing in pluripotent stem cells. Cell Stem Cell 16, 142–147. https://doi.org/10.1016/j.stem.2015.01.003

Yuan, X., Wan, H., Zhao, X., Zhu, S., Zhou, Q., Ding, S., 2011. Brief report: Combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. Stem Cells 29, 549–553. https://doi.org/10.1002/stem.594

Zeineddine, D., Hammoud, A.A., Mortada, M., Boeuf, H., 2014. The Oct4 protein: more than a magic stemness marker. Am J Stem Cells 3, 74–82.

Zhou, H., Ding, S., 2010. Evolution of induced pluripotent stem cell technology. Curr. Opin. Hematol. 8, 288. https://doi.org/10.1097/MOH.0b013e328339f2ee

Zhu, S., Li, W., Zhou, H., Wei, W., Ambasudhan, R., Lin, T., Kim, J., Zhang, K., Ding, S., 2010. Reprogramming of human primary somatic cells by OCT4 and chemical compounds. Cell Stem Cell 7, 651–655. https://doi.org/10.

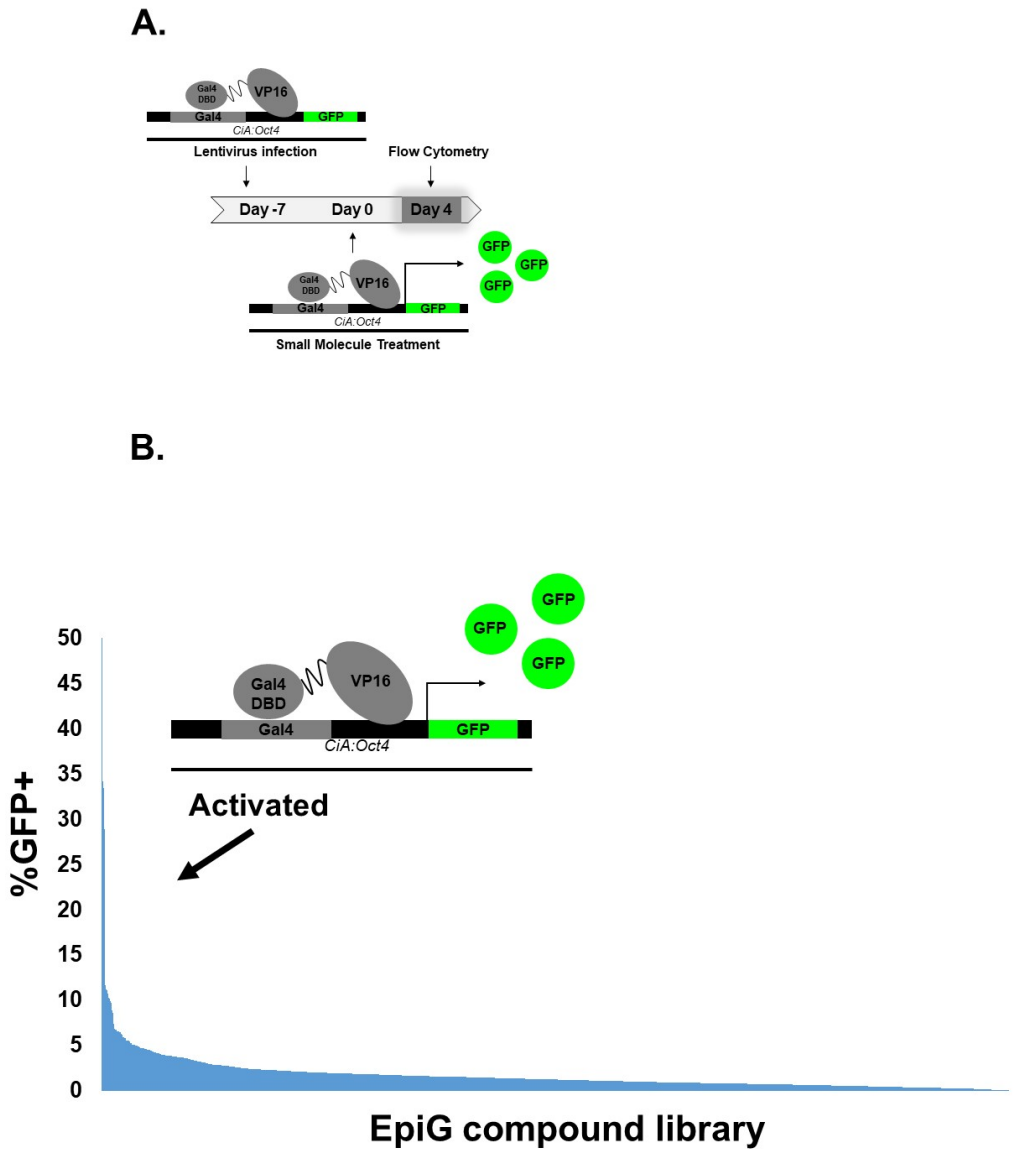
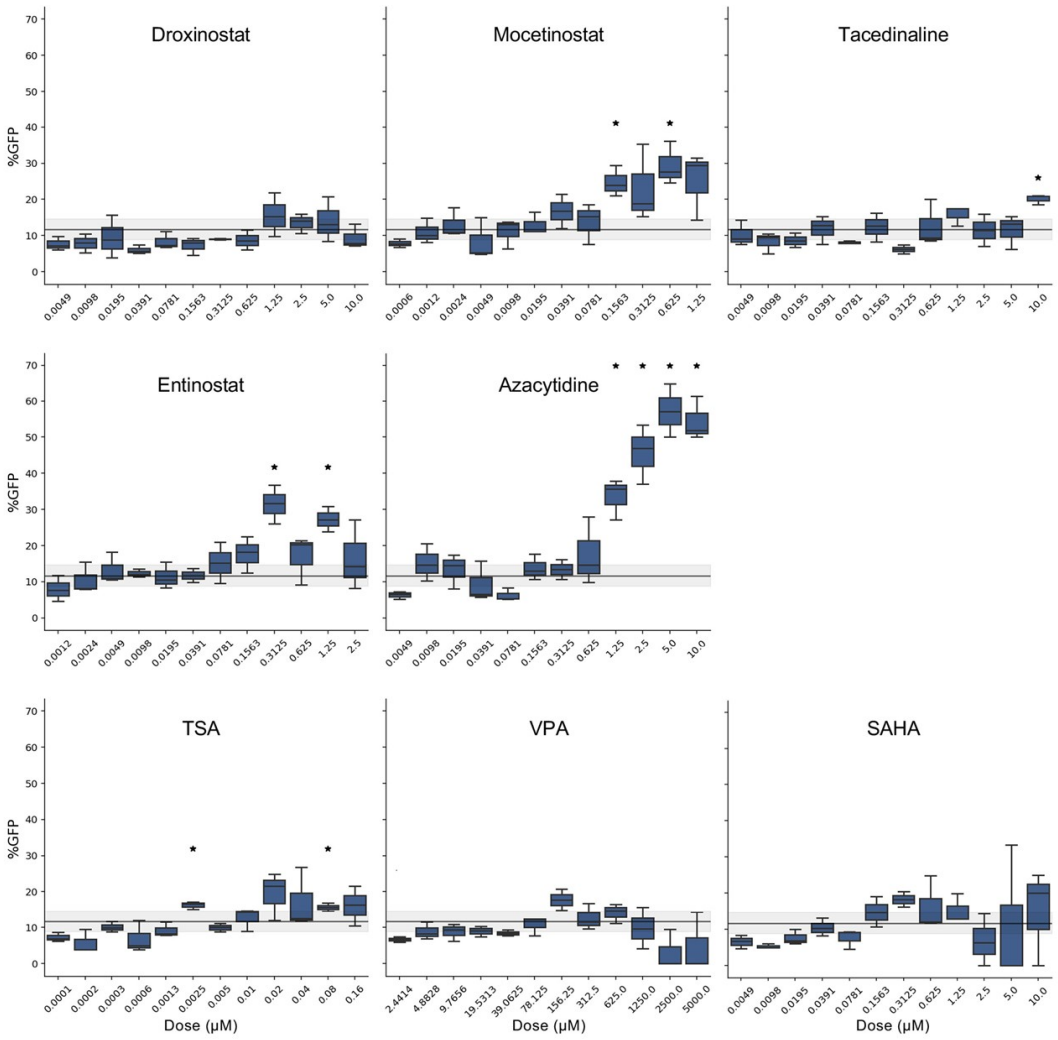


Figure 2.1: Small molecule high throughput screen reveals chemical facilitators of *CiA:Oct4* activation. **(A)** Schematic representation of experimental timeline. Addition of lentivirus occurred at Day -7. Selection for proper transduction was added on Day -5. Small molecules were added to media on Day 0 and flow cytometry analysis was performed on Day 4. **(B)** ~960 small molecules were screened, results represented as %GFP activation after four days of small molecule treatment ordered from highest (left) to lowest (right) percentage of GFP positive cells.

Figure 2.2: Dose response of five selected top hit compounds to triage compounds worth further analysis. **(A)** Schematic representation of procedural timeline. Lentivirus infection occurred on Day -7. Small molecule was added to cells on Day 0. Flow cytometry analysis was performed on Day 4. **(B)** Small molecule treatment dose response demonstrates best dosage for small molecule treatment and comparison to well characterized HDACis: TSA, SAHA, and VPA. The control average is shown as gray line and 95% confidence interval is shown as a gray shadow around this line. ($p \leq 0.05$*) Error bars represent standard deviation. The control average activation is represented with a gray line and the surrounding gray shadow represents the 95% confidence interval of the control average.



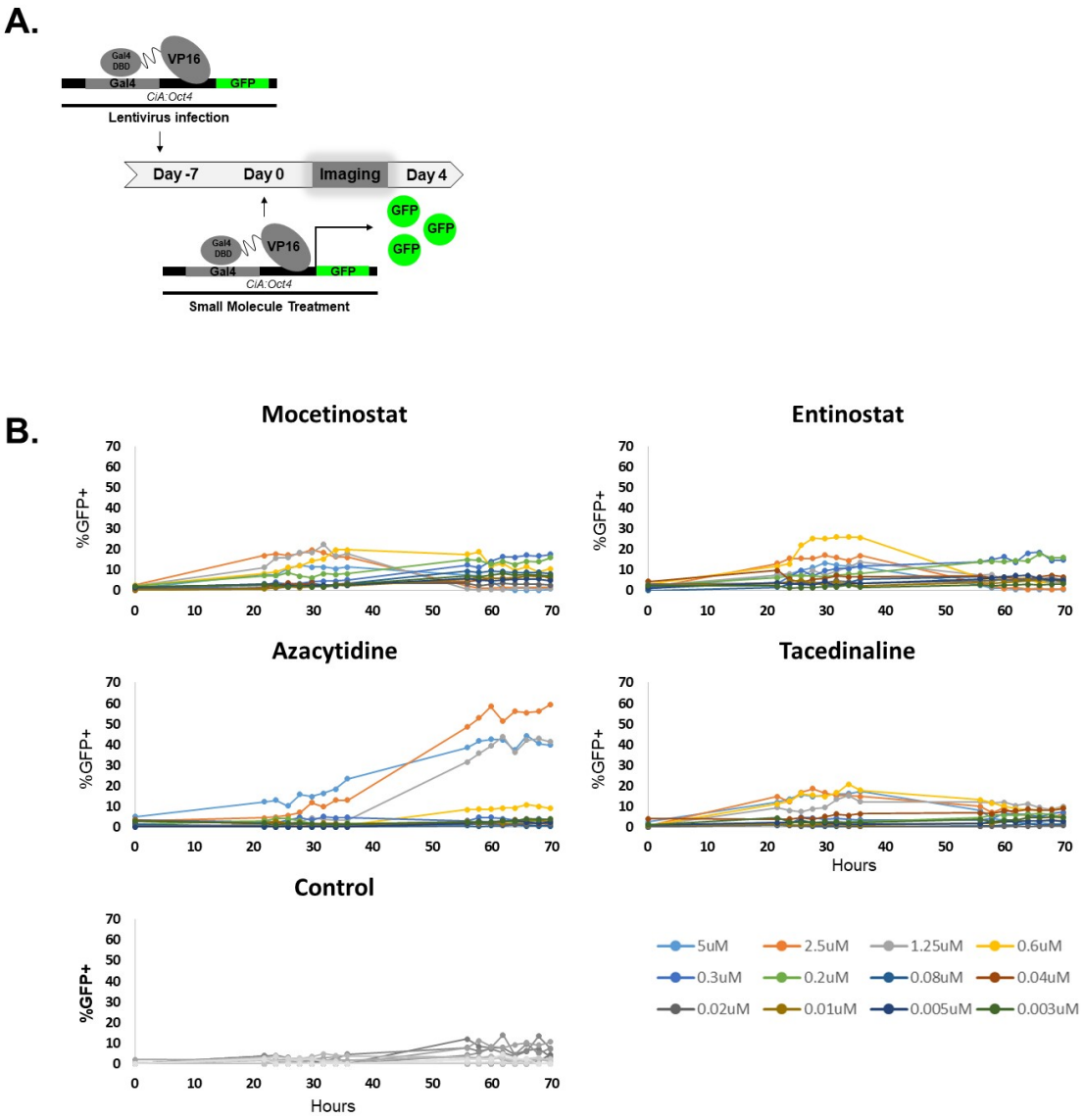


Figure 2.3: Live cell imaging of *CiA:Oct4* during recruitment transcriptional activator and treatment of indicated small molecule. **(A)** Schematic representation of procedural timeline. Cells were infected with Lentivirus on Day -7. Small Molecules were added to cells at the indicated doses on Day 0 and imaged at the indicated times until Day 4 **(B)** Time-lapse imaging reveals dynamics of HDACi facilitated *CiA:Oct4* activation vs. DNTMi facilitated *CiA:Oct4* activation. High content time-lapse imaging data was collected at the indicated times from hours 0 to 70. Analysis was performed using GE Cell Developer to count GFP+ nuclei and mC+ nuclei over time. % GFP+ cells were calculated by dividing GFP counts by mC counts over time (See Supplementary Figure 2.4 for image example).

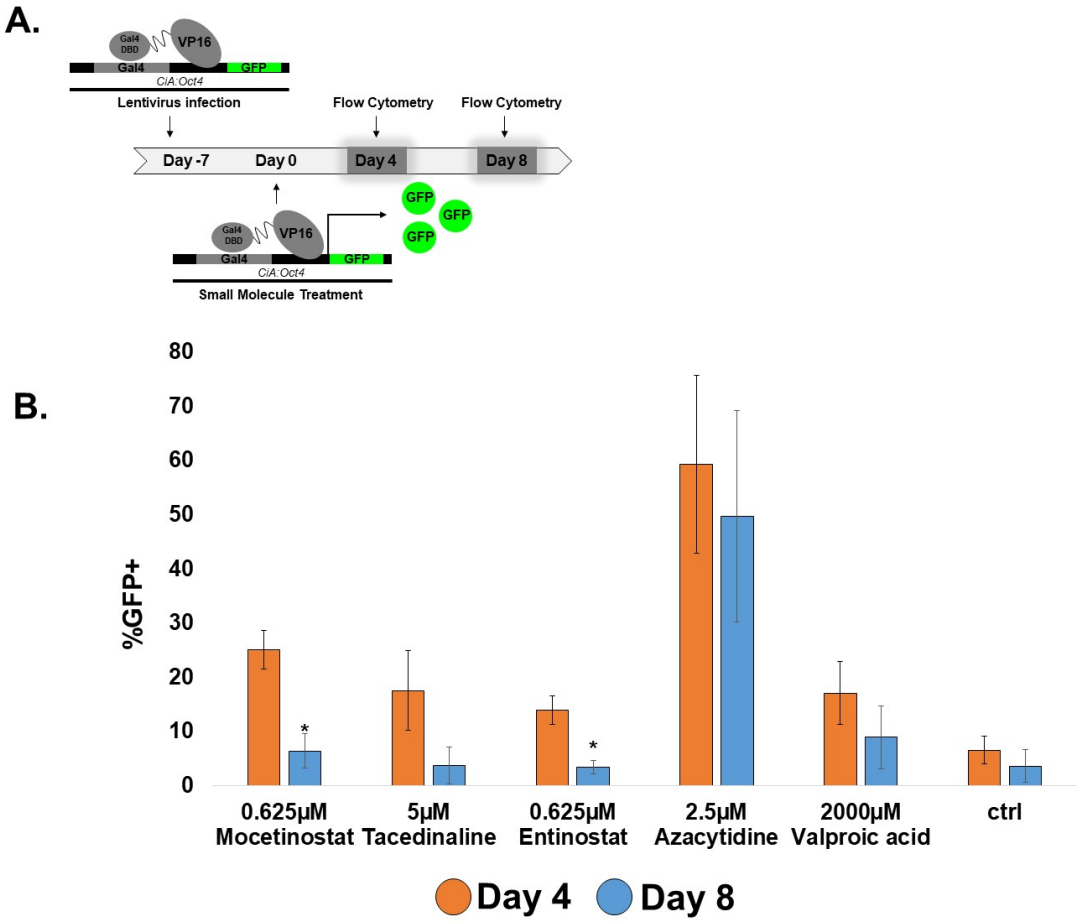


Figure 2.4: Flow cytometry analysis of memory of small molecule facilitation of *CiA:Oct4* activation before and after a 4 day washout. **(A)** Schematic representation of procedural timeline**.** Cell infection occurred on Day -7. Cells were treated with small molecules on Day 0 through Day 4. Flow cytometry was performed on the cells on Day 4. Small molecule treatment was released on Day 4. Flow cytometry was performed 4 days after release, on Day 8. **(B)** DNMTi results in long-term gene activation while HDACi has short-term gene activation demonstrated by small molecule release. Day 4 in orange shows percent GFP positive cells identified by flow cytometry of cells treated with the indicated small molecule. Day 8 in blue shows percent GFP cells identified by flow cytometry of cells treated with the indicated small molecule than released from small molecule treatment for four days. ($p \leq 0.05$*). Error bars represent standard deviation.

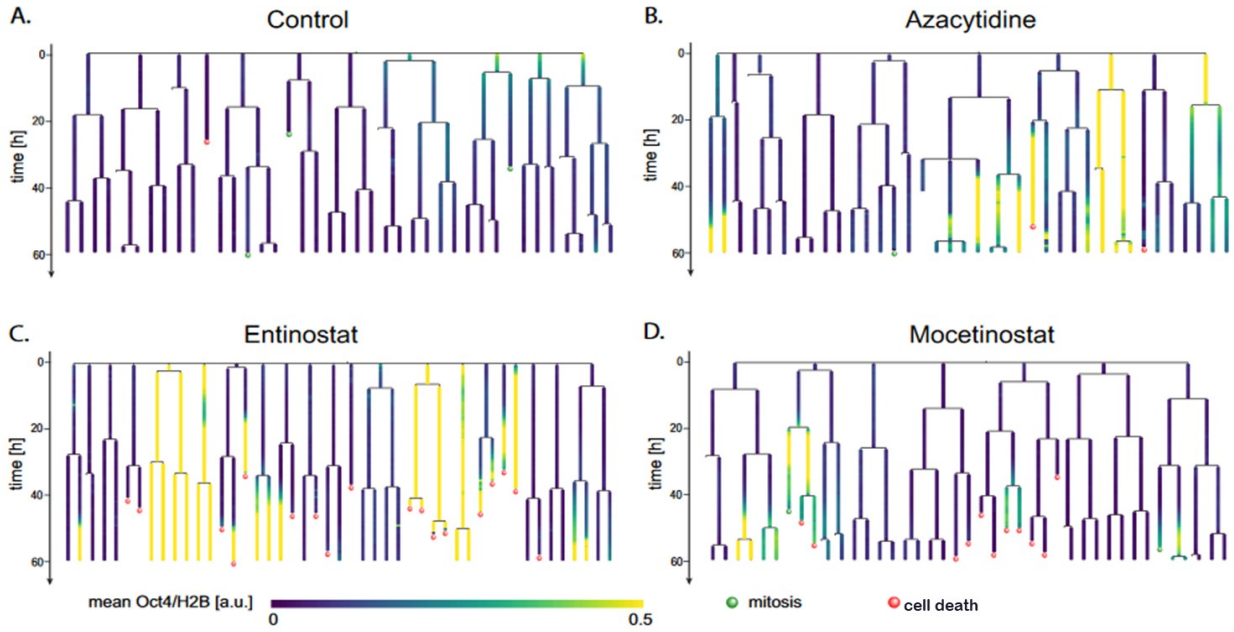


Figure 2.5: Single cell traces from a time-lapse imaging experiment showing GFP(Oct4)/H2B-mCherry ratio of cell families growing in different conditions: A. control; B. Azacytidine; C. Entinostat; D. Mocetinostat. Green dots indicate mitosis of cells which offspring was not tracked. Red dots indicate cell death. Cell death rate (see materials and methods): control - 2.5% (1/40); Azacytidine – 5% (2/38); Entinostat – 38% (18/47); Mocetinostat – 27% (11/41). Total duration of the experiments 60 hours.

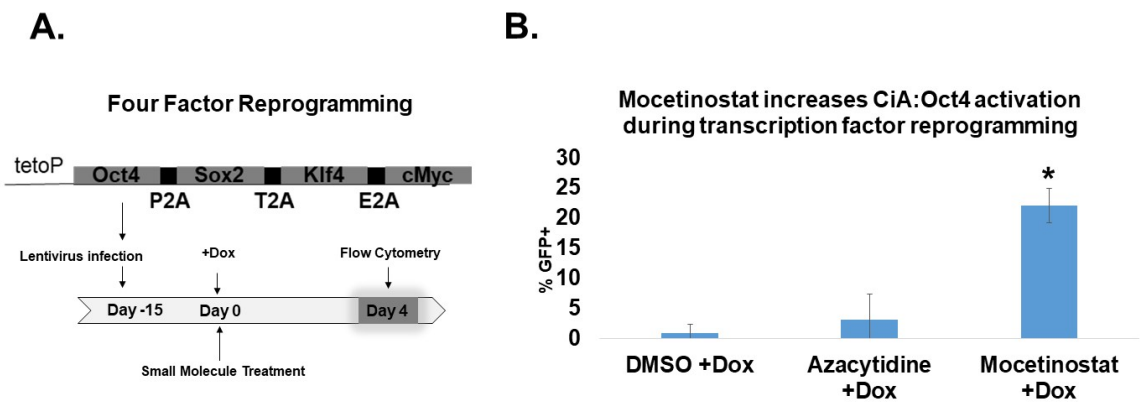
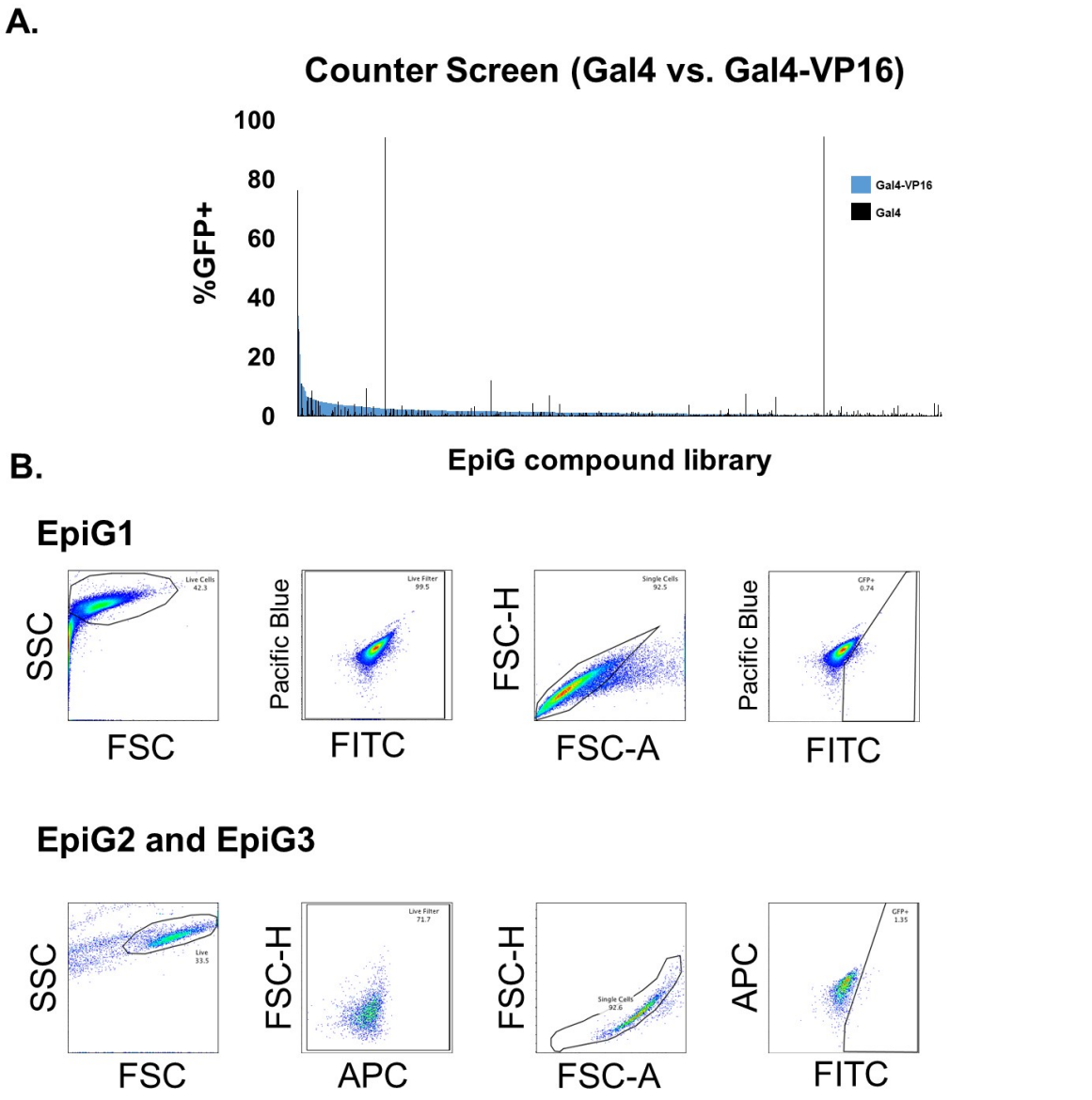
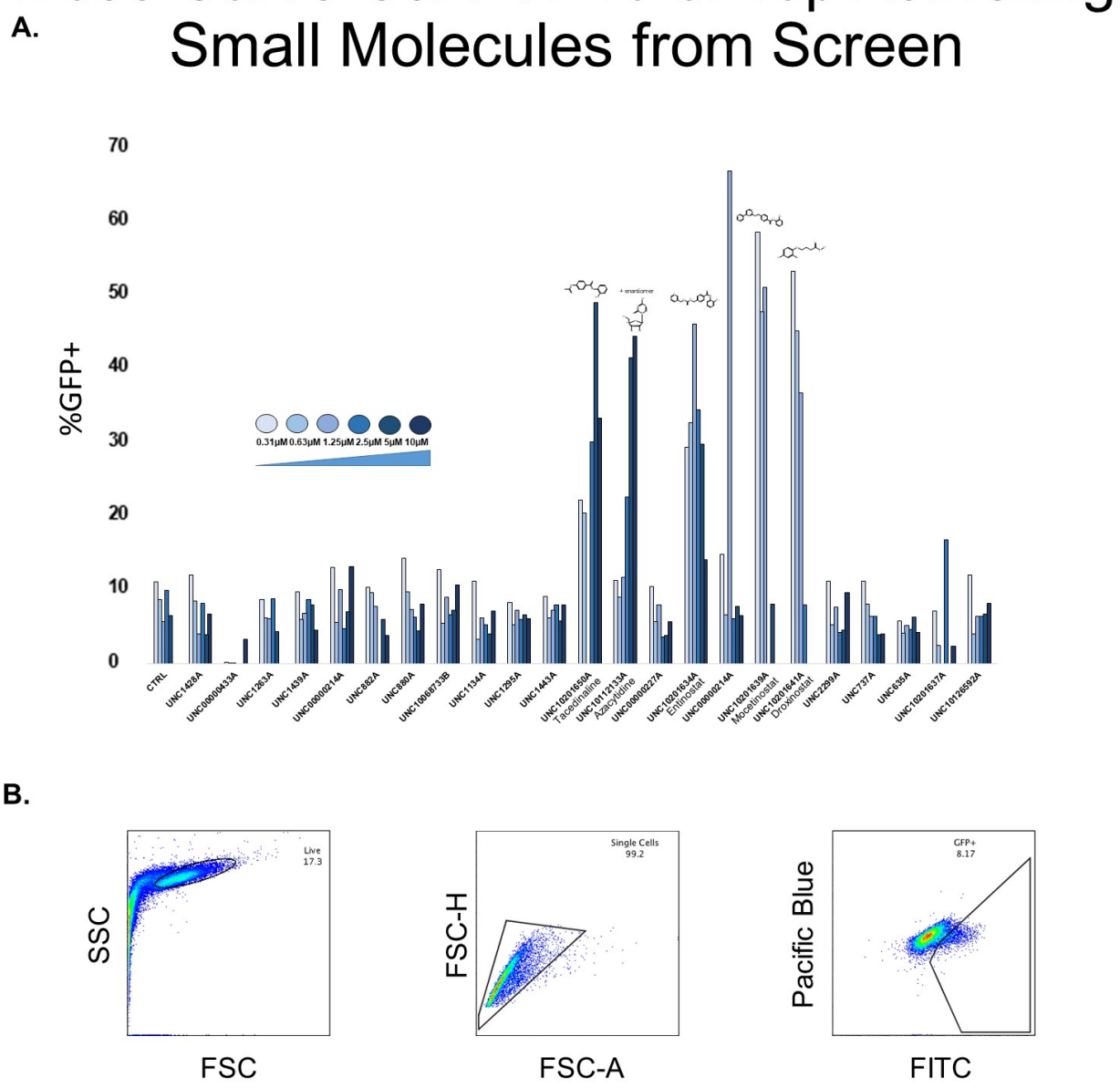


Figure 2.6: Mocetinostat treatment increases transcription factor reprogramming. (A) Schematic representation of timeline. Cells were infected on Day -15, Cells were treated with small molecules on Day 0 and flow cytometry was performed on Day 4 (B) Mocetinostat treated cells demonstrated increased Oct4 activation during transcription factor reprogramming with polycistronic vector for Oct4, Sox2, Klf4, and cMyc. ($p \leq 0.05$*) Error bars represent standard deviation and p-values are representative comparison to DMSO (+Dox).



Supplemental Figure 2.1) Gal4 only counter-screen. **(A)** Screening results demonstrate low background of GFP expression in top hits. GFP expression of Gal4 infected cells treated with small molecules does not correlate with GFP expression of Gal4-VP16 infected cells treated with the same molecule. In top small molecules, GFP background is typically low. **(B)** Cell gating workflow example.



Supplemental Figure 2.2) Rescreen of the top compounds that facilitated the highest percentage of GFP positive cells. **(A)** This work identified five small molecule facilitators of *CiA:Oct4* activation. Small molecules: UNC10201641A - Droxinostat, UNC10201639A – Mocetinostat, UNC10201634A – Entinostat, UNC10112133A – Azacytidine, and UNC10201650A – Tacedinaline all demonstrated dose dependent increases in GFP activation. These identified small molecules have potential as *CiA:Oct4* activators. **(B)** Flow cytometry gating and example rescreen of top *CiA:Oct4* activating small molecules. Pacific Blue acts as a measure for cell autofluorescence.

Figure 2 Gating

A.

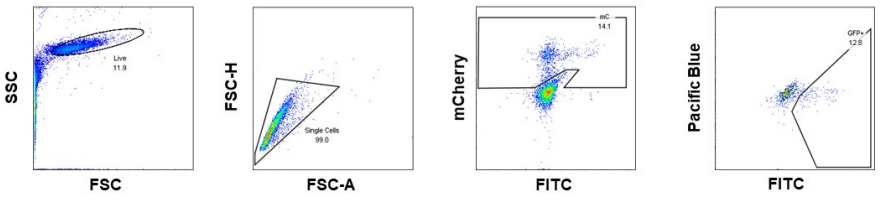
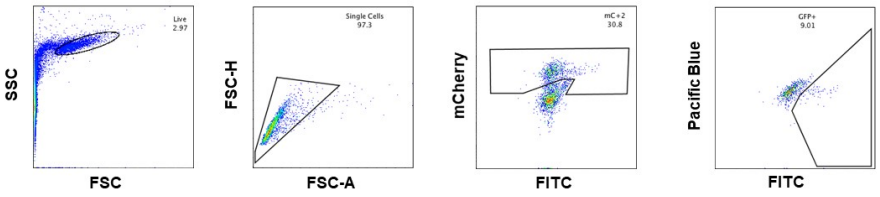


Figure 4 Gating

Day 4

B.



Day 8

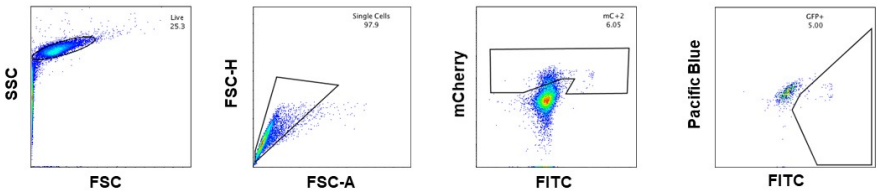
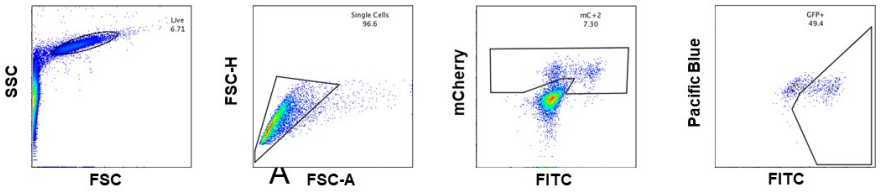
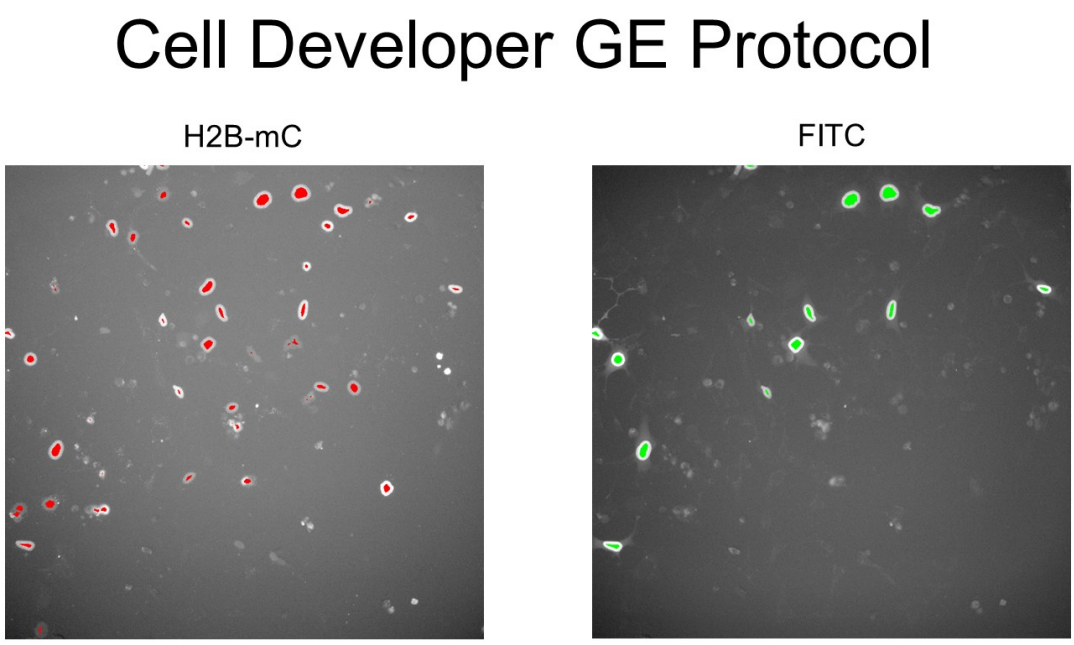


Figure 5 Gating

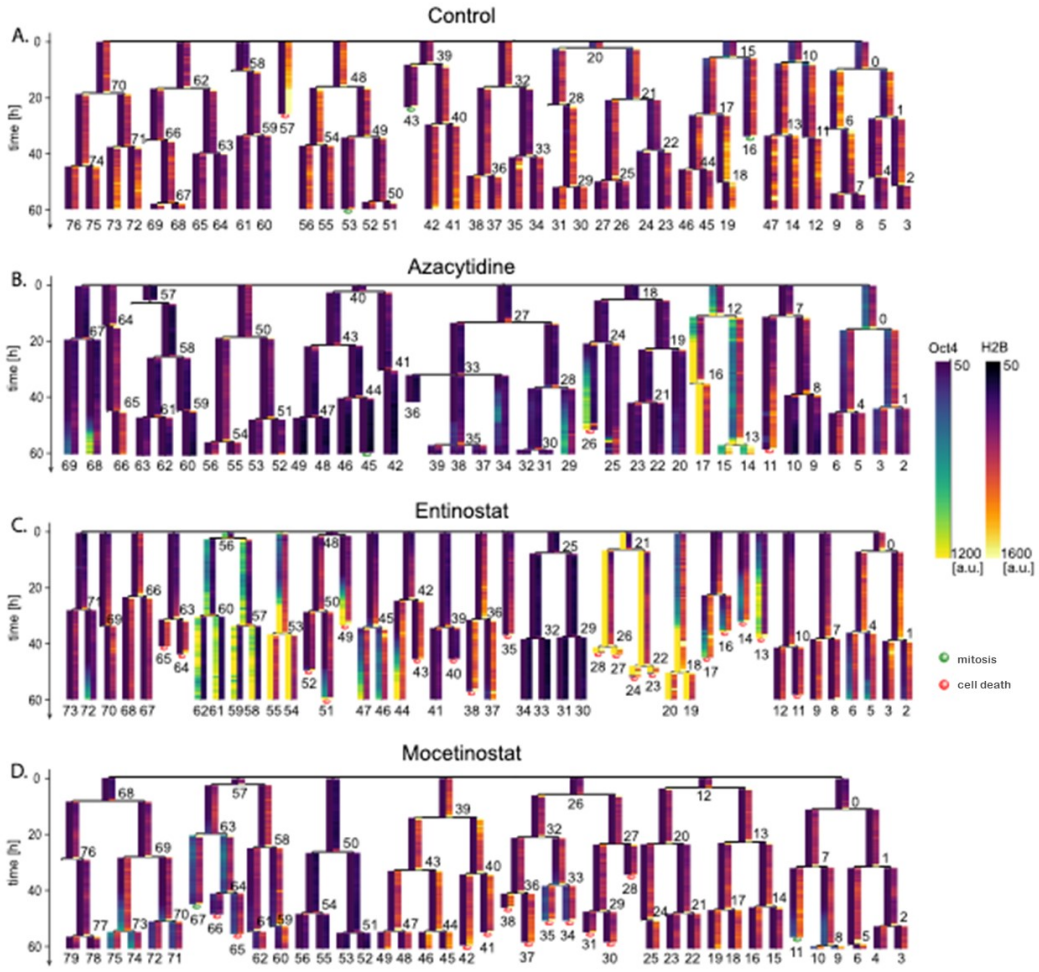
C.



Supplemental Figure 2.3) General gating strategy for multiple figures with an example sample for each. **(A)** Flow cytometry gating for dose response of Gal4-VP16-P2A-H2B-mC (Figure 2 gating). Pacific Blue measures autofluorescence. **(B)** Flow cytometry gating for small molecule treatment and release (Figure 4 gating). ($p \leq 0.05$*)**.** Autofluorescence of cells is measured by Pacific Blue. **(C)** Flow cytometry gating for sequential HDACi treatment for two days then DNMTi treatment for two days or vice-versa. (Figure 5 gating). Pacific Blue is autofluorescence of cells.

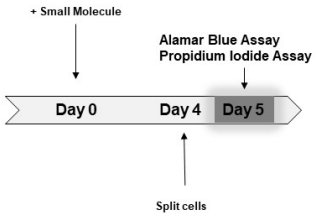


Supplementary Figure 2.4) Example cell counting from GE Cell Developer**.** Cells were assigned GFP+ or mCh+ given the assigned GE cell developer protocol. The images here provide an example of program calculated picked cells. Red representing H2B-mCh. GFP is represented as green and is a marker for *CiA:Oct4* expression.

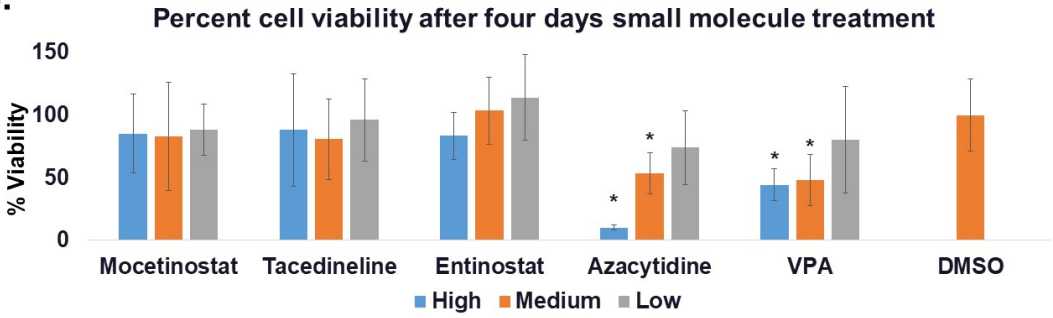


Supplementary Figure 2.5) This supplementary figure accompanies main Figure 6. Single cell traces showing separately GFP (Oct4) and H2B-mCherry for each cell growing in different conditions: **(A)** control; **(B)** Azacytidine; **(C)** Entinostat; **(D)** Mocetinostat. H2B signal (50-1600 [a.u.]) is visualized using 'Inferno' look-up table (Matplotlib) and shown at the right side of cell branches. Oct 4 signal (50-1200 [a.u.]) is visualized using 'Viridis' look-up table (Matplotlib) and shown to the left of H2B trace. Numbers of full cell cycles (beginning and ending in mitosis within time constraints of the movies) of cells shown in A-D. Mean duration of cell cycles correspond to unique identifiers of cells as shown in supplementary movies (1-4). Red dots indicate tracks ending in cell death, green dots indicate tracks ending in mitosis without offspring being tracked.

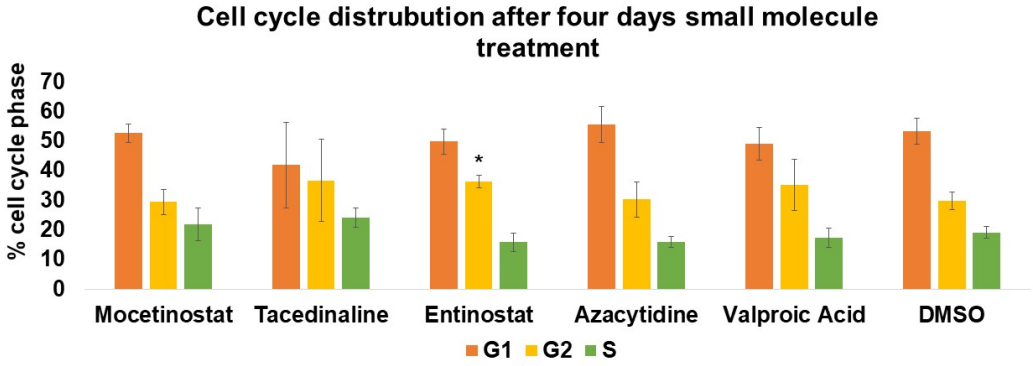
A.



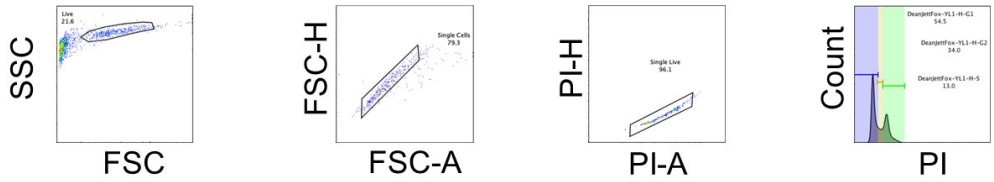
B.



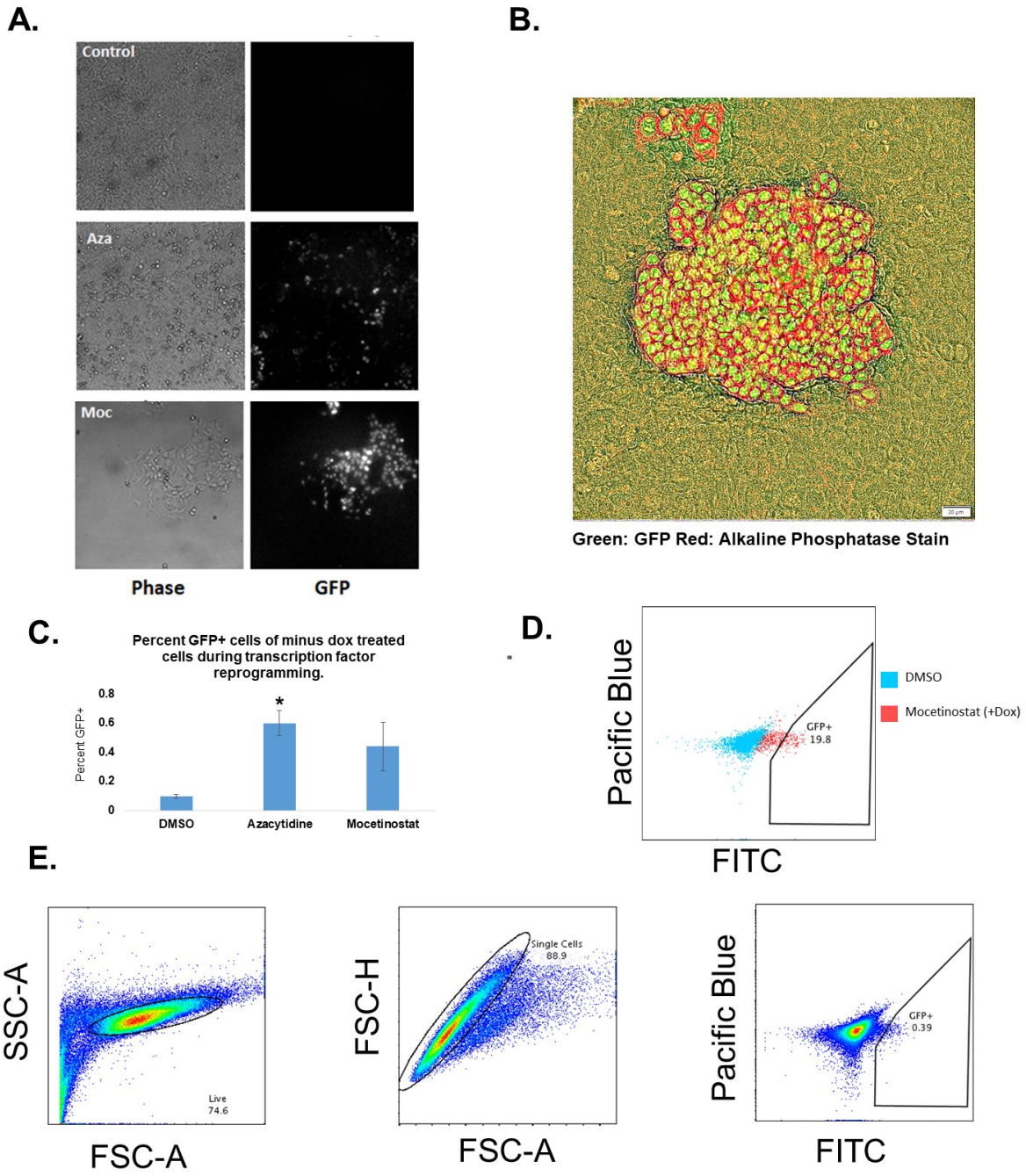
C.



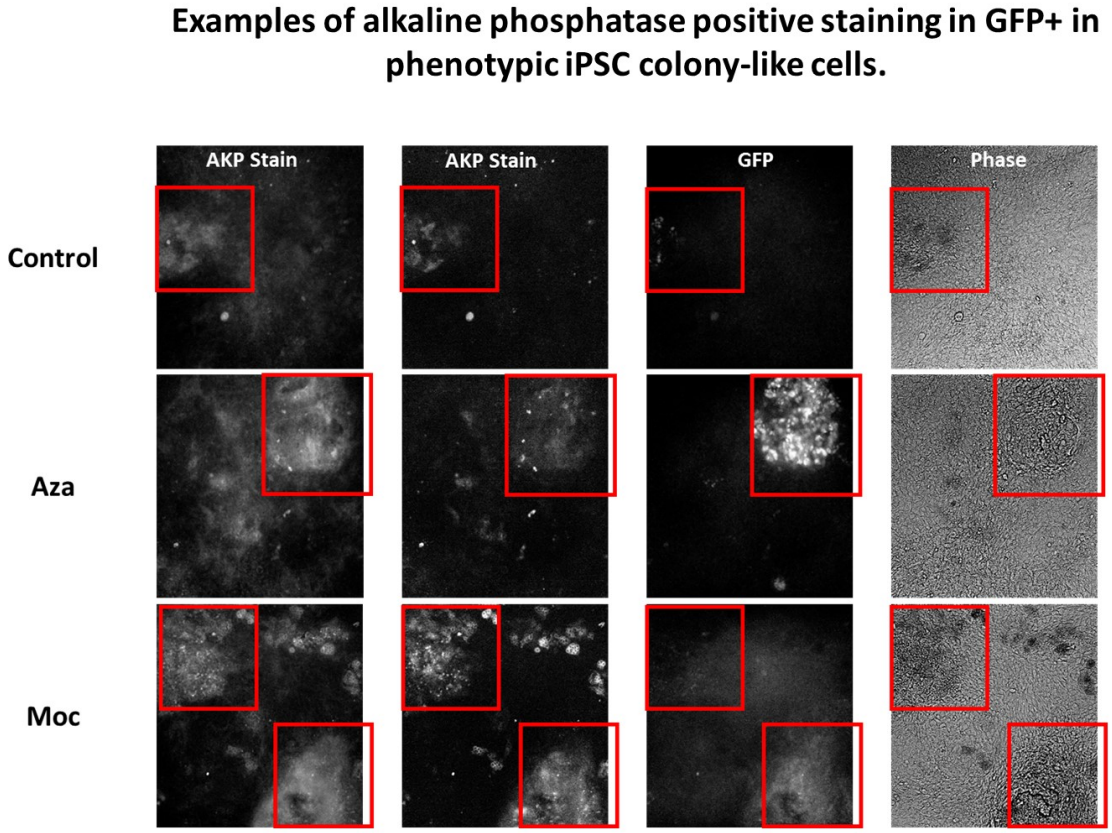
D.



Supplemental Figure 2.6) Azacytidine and VPA decreased cell viability while none of the tested compounds significantly affected cell cycle progression. A) Schematic of small molecule treatment prior to viability/proliferation and cell cycle assays. B) Percent viable cells calculated from alamarBlue staining on SVT-MEF CiAOct4 +/- cells after four days small molecule treatment (upper panel) (*p<0.05, n=8-16) C) Cell cycle analysis on four day small molecule treated cells following BrDU/propidium iodide assay and flow cytometry. D) Gating strategy for propidium iodide assay in Fig. S6C.



Supplementary Figure 2.7) A) Mocetinostat and Azacytidine treatment results in visually apparent GFP+ cells. Example iPSC images 18 days post infection. B) iPSC colonies are distinct colonies which grow with a round cell shape phenotype. Example iPSC colony image 56 days post infection. Overlay of mCherry (alkaline phosphatase stain), FITC (GFP), and Brightfield. C) Percent GFP+ cells of –dox treated cells during transcription factor reprogramming. D) Mocetinostat treated cells have increased GFP activation compared to DMSO cells during transcription factor reprogramming. Example flow cytometry gating comparing Mocetinostat (+Dox) treated cells to DMSO treated cells during transcription factor reprogramming. E) General gating strategy for Mocetinostat and Azacytidine treatment during transcription factor reprogramming method using polycistronic vector containing Oct4, Sox2, Klf4, and cMyc (Figure 7). Autofluorescence of cells is measured by Pacific Blue.



Supplementary Figure 2.8) Examples of alkaline phosphatase staining in comparison to GFP+ cells with iPSC colony morphology. These images are taken at 20 days post infection (Control and Azacytidine) and 30 days post infection (Mocetinostat).

Compound	Percent Activation	Known Targets	Reference
Azacytidine	29% GFP+	Methyltransferase inhibitor	Pikala et al.
Tacedinaline	10% GFP+	HDAC 1, 2, and 3 inhibitor	Moradei et al. Loprevite et al. LoRusso et al. Graziano et al.
Entinostat	10% GFP+	HDAC 1 2 and 3 inhibitor	Lauffer et al. Rosato et al. Saito et al. Zhang et al.
Droxinostat	10% GFP+	HDAC 3, 6 and 8 inhibitor	Liu et al. Wood et al. MMcCourt et al. Bijangi-Vishehsaraei et al. Wood et al.
Mocetinostat	5% GFP+	HDAC Inhibitor against HDAC 1, 2, 3, and 11	Fournel et al. Cai et al.

Supplemental Table 2.1) Summary of top small molecules and their known targets.

Table References

Bijangi-Vishehsaraei, K., Huang, S., Safa, A.R., Saadatzadeh, M.R., Murphy, M.P., 2010. 4-(4-Chloro-2-methylphenoxy)-N-hydroxybutanamide (CMH) targets mRNA of the c-FLIP variants and induces apoptosis in MCF-7 human breast cancer cells. Mol. Cell. Biochem. 342, 133–142. https://doi.org/10.1007/s11010-010-0477-7

Cai, J., Zhang, Q., Lin, K., Hu, L., Zheng, Y., 2015. The Effect of MGCD0103 on CYP450 Isoforms Activity of Rats by Cocktail Method. Biomed Res. Int. 2015, 517295. https://doi.org/10.1155/2015/517295

Fournel, M., Bonfils, C., Hou, Y., Yan, P.T., Trachy-Bourget, M.-C., Kalita, A., Liu, J., Lu, A.-H., Zhou, N.Z., Robert, M.-F., Gillespie, J., Wang, J.J., Ste-Croix, H., Rahil, J., Lefebvre, S., Moradei, O., Delorme, D., Macleod, A.R., Besterman, J.M., Li, Z., 2008. MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo. Mol. Cancer Ther. 7, 759–68. https://doi.org/10.1158/1535-7163.MCT-07-2026

Graziano, M.J., Galati, A.J., Walsh, K.M. Immunotoxicity of the anticancer drug CI-994 in rats: effects on lymphoid tissue. Arch. Toxicol. 73, 168–74.

Lauffer, B.E.L., Mintzer, R., Fong, R., Mukund, S., Tam, C., Zilberleyb, I., Flicke, B., Ritscher, A., Fedorowicz, G., Vallero, R., Ortwine, D.F., Gunzner, J., Modrusan, Z., Neumann, L., Koth, C.M., Lupardus, P.J., Kaminker, J.S., Heise, C.E., Steiner, P., 2013. Histone Deacetylase (HDAC) Inhibitor Kinetic Rate Constants Correlate with Cellular Histone Acetylation but Not Transcription and Cell Viability. J. Biol. Chem. 288, 26926–26943. https://doi.org/10.1074/jbc.M113.490706

Liu, J., Li, G., Wang, X., Wang, L., Zhao, R., Wang, J., Kong, Y., Ding, J., Li, J., Zhang, L., 2016. Droxinostat, a Histone Deacetylase Inhibitor, Induces Apoptosis in Hepatocellular Carcinoma Cell Lines via Activation of the Mitochondrial Pathway and Downregulation of FLIP. Transl. Oncol. 9, 70–78. https://doi.org/10.1016/j.tranon.2016.01.004

Loprevite, M., Tiseo, M., Grossi, F., Scolaro, T., Semino, C., Pandolfi, A., Favoni, R., Ardizzoni, A., 2005. In vitro study of CI-994, a histone deacetylase inhibitor, in non-small cell lung cancer cell lines. Oncol. Res. 15, 39–48.

LoRusso, P.M., Demchik, L., Foster, B., Knight, J., Bissery, M.C., Polin, L.M., Leopold, W.R., Corbett, T.H., 1996. Preclinical antitumor activity of CI-994. Invest. New Drugs 14, 349–56.

McCourt, C., Maxwell, P., Mazzucchelli, R., Montironi, R., Scarpelli, M., Salto-Tellez, M., O'Sullivan, J.M., Longley, D.B., Waugh, D.J.J., 2012. Elevation of c-FLIP in castrate-resistant prostate cancer antagonizes therapeutic response to androgen receptor-targeted therapy. Clin. Cancer Res. 18, 3822–33. https://doi.org/10.1158/1078-0432.CCR-11-3277

Moradei, O.M., Mallais, T.C., Frechette, S., Paquin, I., Tessier, P.E., Leit, S.M., Fournel, M., Bonfils, C., Trachy-Bourget, M.-C., Liu, J., Yan, T.P., Lu, A.-H., Rahil, J., Wang, J., Lefebvre, S., Li, Z., Vaisburg, A.F., Besterman, J.M., 2007. Novel aminophenyl benzamide-type histone deacetylase inhibitors with enhanced potency and selectivity. J. Med. Chem. 50, 5543–6. https://doi.org/10.1021/jm701079h

Píškala, A., Hanna, N.B., Budešínský, M., Cihák, A., Veselý, J., 1981. Synthesis and biological activity of 2'-deoxy-6-methyl-5-azacytidine and its alpha-D-anomer. Nucleic Acids Symp. Ser. 83–6.

Rosato, R.R., Almenara, J.A., Grant, S., 2003. The histone deacetylase inhibitor MS-275 promotes differentiation or apoptosis in human leukemia cells through a process regulated by generation of reactive oxygen species and induction of p21CIP1/WAF1 1. Cancer Res. 63, 3637–45.

Saito, A., Yamashita, T., Mariko, Y., Nosaka, Y., Tsuchiya, K., Ando, T., Suzuki, T., Tsuruo, T., Nakanishi, O., 1999. A synthetic inhibitor of histone deacetylase, MS-27-275, with marked in vivo antitumor activity against human tumors. Proc. Natl. Acad. Sci. U. S. A. 96, 4592–7.

Wood, T.E., Dalili, S., Simpson, C.D., Sukhai, M.A., Hurren, R., Anyiwe, K., Mao, X., Suarez Saiz, F., Gronda, M., Eberhard, Y., MacLean, N., Ketela, T., Reed, J.C., Moffat, J., Minden, M.D., Batey, R.A., Schimmer, A.D., 2010. Selective inhibition of histone deacetylases sensitizes malignant cells to death receptor ligands. Mol. Cancer Ther. 9, 246–56. https://doi.org/10.1158/1535-7163.MCT-09-0495

CHAPTER 3: INTRODUCTION TO OVERCOMING CRISPR-CAS9 GENOME EDITING BARRIERS

History of CRISPR-Cas9 genome editing

Introduction

Precise and efficient genome editing has been a goal for researchers for many decades. Manipulations to DNA sequence through deletions, insertions, or base changes allow researchers to directly understand the role that DNA sequence plays in cellular functions. CRISPR-Cas9 genome editing is now a widely accepted powerful genome editing tool allowing for specific changes to the genome. CRISPR-Cas9 genome editing applications have revolutionized molecular biology, allowing for greater understanding of; molecular networks, epigenetic mechanisms, imaging, genetic therapeutics, agriculture and more (Adli, 2018).

CRISPR-Cas9 genome editing history

Clustered regularly interspaced short palindromic repeats (CRISPR) sequences were first discovered in 1987, by Ishino et al. after sequencing a 1.7kbp *E.Coli* DNA fragment of an isozyme of alkaline phosphatase (iap) during the study of phosphate metabolism (Ishino et al., 2018; Makino et al., 2016). The sequence was notable at the time because of it was difficult to sequence (potentially due to hairpin structures) and had noteworthy repeated patterns of DNA sequence. These DNA sequences included palindromic sequences ~30bps in length and

separated by ~36bps nonpalindromic “spacer” sequences. At the time of the discovery of these sequences, it was unclear of the function of CRISPR elements due to lack of genomic information (Ishino et al., 2018).

Advancements in genomics allowed for further understanding of this unique sequence function. In the early 1990s, during a classification study of archaea DNA fragments, it was found that various strands of archaea contained similar structured palindromic repeats and spacer sequences to those previously identified in a bacterial system by Ishino et al (Mojica et al., 1995, 1993). Further, using bioinformatics methods it was found that over 80 microbe sequences identified spacer regions of CRISPR sequences were identical complements to that of viruses and conjugative plasmids (Lander, 2016; Mojica et al., 2005). This suggested a functional system important to both bacterial and archaea systems. Several hypotheses, during this time, suggested that spacer elements serve as a genetic memory of previous invaders such as viral infections and phage transfers and form of cell immunity response against phage infection of foreign DNA elements (Bolotin et al., 2005; Lander, 2016; Pourcel et al., 2005).

In 2006, during a study of bacterial cultures for dairy fermentation, a clearer relationship between CRISPR elements and cell immunity was discovered. It was found that specific bacterial strains had particular resistances to viral infections and that their resistance to infection correlated to the sequence of the CRISPR spacer region. More specifically, the strains with spacer regions homologous to the viral DNA of the invader were more resistant to infection (Horvath et al., 2008). Further tests demonstrated that bacteria could add spacers to CRISPR sequences following a phage infection, homologous in sequence to the phage invader, and providing future resistance to the bacteria (Horvath et al., 2008). In another aspect of this study, the functional roles of Cas7 and Cas9 were elucidated. It was found that Cas7 was required in

order for the bacteria culture to gain resistance to the phage but was not required for continued resistance. In contrast, Cas9 was always required for phage resistance. From this study it was suggested that Cas9 was the endonuclease which cut invading nucleic acids and provided resistance against infection (Bolotin et al., 2005; Lander, 2016; Makarova et al., 2006). This activity of Cas9 was confirmed by Deveau et al. who demonstrated that Cas9 cut double stranded DNA at consistent and specific sequences next to the PAM sequence guided by crRNA sequence (Deveau et al., 2008; Horvath et al., 2008; Lander, 2016).

Programming CRISPR-Cas9

In the early 2000s, systems were created which could program the CRISPR system. CRISPR arrays were inserted into bacterial genomes which could create an immunity system against any virus which they called, "A flu shot for bacteria" (Makarova et al., 2006). In a step towards genome editing, an *in vitro* system programmed CRISPR components to cut purified DNA sequences with crRNAs and tracrRNAs targeted for specific loci. Additionally, it was found that tracrRNAs and crRNAs could be fused to create a single-guide RNA which could be used with the same specific cut efficiency in their place (Jinek et al., 2012; Lander, 2016). In 2012, the CRISPR-Cas9 system became a genome editing platform using a Cas9, a tracrRNA or sgRNA, and a CRISPR array where it was shown to successfully mutate, delete, or insert target genes with high precision and efficiency (Cong et al., 2013; Hwang et al., 2013; Mali et al., 2013b). This system has since become the most successful genome editing technique to date and has become permissive in many research fields and clinical applications.

Optimization of CRISPR-Cas9 Genome Editing

There are six types of CRISPR systems. The type II system has gained the most prevalence in genome editing research. In the type II system, the tracrRNA, Cas9 protein, and pre-crRNAs are expressed from the CRISPR locus and assembled into a mature crRNA:tracrRNA:Cas9 complex following post-transcriptional processing and work together to identify and cleave target sequences. In this process, the complex scans the genome looking for a protospacer adjacent motif (PAM) sequence and then targets and cleaves a sequence complementary to its guide RNA (crRNA:tracrRNA), 10-12 nucleotides away (Lino et al., 2018).

The first genome editing using the type II system was demonstrated by Doudna and Charpentier. In their study, targeted double strand breaks (DSBs) were successfully made in DNA of the bacteria, streptococcus pyogenes (Jinek et al., 2012; Lino et al., 2018). It was identified that Cas9 could be directed by a smaller sgRNA instead of a larger crRNA:tracrRNA. This system became the most prevalent tool currently used in CRISPR-Cas9 genome editing (Lino et al., 2018). This further evolved to become a platform an insertion/mutation editing platform through the introduction of a targeting vector (Cho et al., 2013). In this genome editing adaptation, the targeting vector provides a template for DNA which can undergo homology directed repair (Cong et al., 2013; Mali et al., 2013a).

CRISPR-Cas9 genome editing system modifications

Several modifications to the CRISPR-Cas9 system has allowed for unique genetic and epigenetic alterations. A catalytically-dead Cas9 (dCas9) was engineered which can bind and

target a DNA sequence. On its own, dCas9 targeting can interfere with transcription, an approach known as CRISPRi (Larson et al., 2013). Conversely, dCas9 can also be fused with a transcriptional activator (such as VP64) to activate target loci, a technique also known as CRISPRa (Mali et al., 2013a). Similarly, epigenetic readers, writers, and erasers such as HATs, HDACs, DNMTs, KRAB, and more can also be fused to the Cas9 to change epigenetic profiles at target loci (Bintu et al., 2016; Hilton et al., 2015).

Current avenues for CRISPR-Cas9 genome editing optimization center on sgRNA selection, Cas9 optimization, and promoting specific DNA damage response pathways. Chosen sgRNAs are not always selective for their directed loci during CRISPR-Cas9 genome editing. sgRNAs can bind base-mismatched sequences which can lead to unspecific binding and cleavage, and cause unwanted genome edits (Doench et al., 2014; Fu et al., 2013; Hsu et al., 2013; Moreno-Mateos et al., 2015; Wang et al., 2014; Xu et al., 2015). Bioinformatics tools have become available to help design sgRNAs with minimal potential for binding off-target sequences (Brazelton et al., 2015).

In an optimization to the CRISPR-Cas9 system, Cas9 protein can be modified to allow for more sgRNA selection options. Cas9 from various species of bacteria identify different PAM sequences. Cas9 proteins which can target different PAM sequences provides more genomic sequence options for PAM sequences, and likewise, more options to for Cas9 driven targeting with sgRNAs (Anders et al., 2016; Hirano et al., 2016; Kleinstiver et al., 2015; Lino et al., 2018).

Additionally, the CRISPR-Cas9 system has been optimized to more efficiently target the correct loci. To reduce off-target cleavage, a CRISPR-Cas9 system was developed which facilitates specificity of double stand cleavage. In this system, an engineered Cas9 produces ssDNA nicks. Two Cas9s both producing a nick on both strands at the same location results in a

double strand break required for genome editing. The requirement of two Cas9 systems at the same location for genome editing to occur increases the likelihood genome editing occurs only at the targeted location and increases genome editing specificity (Cho et al., 2014; Lino et al., 2018; Mali et al., 2013a; Ran et al., 2013). In a comparable system, a dCas9 was fused to a FokI nuclease dimer. The requirement of two dCas9-FokI dimers to bind at the same location to enact a double strand break, likewise, increases specificity. Also, studies have been successful in increasing specific cleavage by creating Cas9 mutants with customizable DNA binding affinities (Kleinstiver et al., 2016; Slaymaker et al., 2016).

Another system with significant therapeutic applications utilizes a transfer RNA adenosine deaminase fused to a dCas9. This system performs adenine/thymine to guanine/cytosine base conversion at targeted loci with no induction of double strand breakage with a high efficiency (~50%). This system has applications in therapeutics to edit small cancer-causing mutations with decreased potential for genome instability and off-target mutations posed by double strand breakage (Gaudelli et al., 2017).

HDR vs. NHEJ

During CRISPR-Cas9 genome editing, to precisely perform an insertion mutation, the cell must undergo homology directed repair (HDR). The HDR pathway occurs very infrequently occurring 0.5-20% of the time compared to 20-60% non-homologous end joining (NHEJ) (Kleinstiver et al., 2015). To overcome this limitation, cell cycle synchronization into late S and G2 phase has been successful increasing HDR efficiency (Lin et al., 2014) Additionally, researchers have identified small molecule inhibitors that have been successful at promoting

HDR over NHEJ. Such small molecules include: Scr7, L755507, Brefeldin A, and CRISPY mix (Robert et al., 2015; Srivastava et al., 2012; Tomkinson et al., 2013; Vartak and Raghavan, 2015; Yu et al., 2015). Of these compounds a notable compound used to increase HDR is Scr7. Scr7 is known to increase efficiency ~19 fold, however, has notable toxicity concerns (Chu et al., 2015; Srivastava et al., 2012; Vartak and Raghavan, 2015).

DNA Double-Strand Break Repair and 53BP1

DNA damage response (DDR) is pivotal to genome integrity maintenance and genetic stability in cells. Improper DDR has serious implications including defects in transcription and cell cycle progression and is commonly associated with diseases such as immune deficiency, neurological degeneration, premature aging, and cancer (Giglia-Mari et al., 2011). DDR is a largely based on recognition and response to damaged chromatin. Chromatin damage occurs frequently and can be due to several sources such as environmental, radiation exposure, chemical exposure, or reactive oxygen species (ROS). An estimated 10,000-100,000 lesions can occur in a cell per day and thusly DDR is highly regulated to avoid cellular catastrophe (Giglia-Mari et al., 2011). DDR results in signals which lead to DNA repair or cell cycle arrest. DNA repair on single strand breaks occurs either through "cut and patch" mechanisms on single strand break repair pathways (such as small lesion repair) or through more complicated mechanisms for DSB repair (such as HDR or NHEJ). In small lesion repair on a single DNA strand, nucleotide lesions are recognized by DNA glycosylases, replaced, and sealed using base excision repair-specific DNA polymerase β and XRCC1/Ligase III. Nucleotide excision repair (NER) responds to helical destabilizing lesions on single strands (Giglia-Mari et al., 2011). In NER, the mutated short single-stranded DNA segment is cut away and the homologous DNA strand acts as a template for synthesis of a new strand and similarly ligated into place (Giglia-Mari et al., 2011).

The decision to repair DSB through HDR or NHEJ is primarily based on cell phase, however is also influenced by other factors. NHEJ pathways primarily occur in post-mitotic and cycling cells in G1 during which times cells must rapidly seal DNA to proceed with pivotal cellular processes (Giglia-Mari et al., 2011). In the NHEJ mechanism, DSBs are bound by the Ku70/Ku80 hetero-dimer. Ku70/Ku80 binding activates the PI3-kinase DNA-PK which further

recruits Artemis nuclease and the MRE11/Rad50/NBs1 (MRN) complex for DNA end-processing and XRCC4/LigaseIV complex for end ligation. NHEJ is error prone due to a fast repair mechanism and often results in mutation of a few nucleotides.

Homology directed repair pathway

During the S and G2 phase of the cell cycle, a homologous sister chromatid template is available for reference to repair for the broken DNA. In the HDR pathway, MRN is recruited to the DSB and holds the broken DNA strands together (de Jager et al., 2001; Giglia-Mari et al., 2011). MRN recruits the CtIP nuclease which as a complex to initiate end resection concomitantly with exonuclease I (EXO1) (Giglia-Mari et al., 2011; Limbo et al., 2007; Sartori et al., 2007; Takeda et al., 2007). RPA binds the single stranded DNA region and then is exchanged with a RAD51 nucleo-protein filament. RAD51 replacement allows for strand invasion to the homologous sister DNA. A "D-loop" is formed by two separated DNA strands and an invaded third stand of DNA between them. In the D loop, the invading DNA strangle strand portion is swapped for the DNA found in the homologous sister DNA. This results in a repaired DSB with sequencing deriving from the homologous sister DNA. Although cell stage is the primary determinant of whether DSB repair will proceed with NHEJ or HDR, other factors exist which influence pathway choice including telomeric factors, ubiquitylation, SUMOylation, histone modifications, and DNA damage checkpoint proteins.

CRISPR-Cas9 genome editing and DSB repair

CRISPR-Cas9 genome editing has been efficient in creating DNA mutations through initiating DSBs at precise loci using Cas9 and a sgRNA and initiating of cellular DNA repair

mechanisms to produce desired mutations. DNA which has received CRISPR-Cas9 induced DSBs will proceed through the NHEJ or HDR pathway. Notably, scientists seeking to identify nonsense mutations often use methods directed toward promoting the NHEJ pathway which produce small changes to the genome often resulting in small insertion/deletion mutations. Conversely, scientists seeking to insert a specific sequence into the genome will seek cells which have gone through the HDR pathway during CRISPR-Cas9 genome editing where a provided template acts as the template for repair instead of a homologous sister chromatid DNA during the HDR pathway.

53BP1 and DDR

A checkpoint protein in DDR pathway is the tumor suppressor p53-binding protein 1 (53BP1). 53BP1 has no enzymatic activity on its own, however, acts as an epigenetic reader of histone methylation and ubiquitination marks induced by chromatin breaks and interacts with partner proteins to signal DDR response pathways (Botuyan et al., 2018; Li and Zou, 2005; Lu and Wang, 2013). 53BP1 has an important regulatory function in DNA damage response where it binds as an oligomer to H4K20me(1,2) and H2AK15ub. 53BP1 binding blocks the HDR pathway from occurring by antagonizing long-range DNA end resection.

DSBs are initially recognized by the MRE11-RAD50-NBS1 (MRN) complex and recruits auto-phosphorylated, active ataxia-telangiectasia mutated (ATM) monomers. The ATM phosphorylates histone 2A variant X (H2A.X) at Ser129 (γH2A.X). Mediator of DNA damage checkpoint protein 1 (MDC1) upon recognition of γH2A.X, phosphorylates MDC1. MDC1 phosphorylation results in a positive feedback loop recruiting MRN and active ATM which further recruits more MDC1. Further MDC1 phosphorylation events on chromatin results in

recruitment of E3 ubiquitin ligase RING finger 8 (RNF8), RNF163, and UBC13 signaling a cascade. This cascade results in ubiquitination of the damaged chromatin at H2AK13 and H2AK15 and these ubiquitin modifications further recruit more RNF168. 53BP1 is recruited to the sites of chromatin by recognizing and binding to the H2AK15 site through its ubiquitylation-dependent recruitment motif and to H4K20me(1,2) through the Tandem Tudor Domain (TTD) (Panier and Boulton, 2014).

The accumulation of 53BP1 on chromatin results in an inability to proceed with DNA end resection. (Bothmer et al., 2010; Daley and Sung, 2014). The mechanism of action of how 53BP1 accumulation blocks end resections is not fully understood. However, it has been suggested that it may be the result of cooperation with another protein, RIF1 binding to 53BP1, concurrently accumulating on damaged chromatin (Di Virgilio et al., 2013). Yeast studies have suggested that RIF1 may oligomerize for form a Rif1-Rif2-Rap1 complex which may interact with 53BP1 and stabilize the chromatic state inhibiting access to nucleases and chromatin modifying enzymes, thus inhibiting end resection mechanisms (Bouwman et al., 2010; Bunting et al., 2010). In another pathway which likely inhibits end resection, 53BP1 is thought to inhibit BRCA1 association with the MRN-CtIP complex formation on DNA, an important first step in homology directed repair. Further, it also been suggested that BRCA1 may displace NHEJ factors including, 53BP1, acting as an antagonist for the NHEJ pathway (Bothmer et al., 2010; Bouwman et al., 2010; Bunting et al., 2010; Daley and Sung, 2014; Escribano-Díaz et al., 2013; Feng et al., 2013).

The many CRISPR-Cas9 genome editing advances; starting from understanding an acquired immunity system in bacteria and archaea, to programming the system, to developing methods to use the system to genome edit, exemplify the successful implementation of the

research process. Importantly, however, there is still a clear need for optimization of the CRISPR-Cas9 system to increase efficacy and efficiency. Notably, identifying methods to proficiently direct DNA repair mechanisms to increase the occurrence of HDR over NHEJ, will likely have a profound effect on our ability to rapidly produce precise genome edits.

REFERENCES

Adli, M., 2018. The CRISPR tool kit for genome editing and beyond. Nat. Commun. 9, 1911. https://doi.org/10.1038/s41467-018-04252-2

Anders, C., Bargsten, K., Jinek, M., 2016. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. Mol. Cell 61, 895–902. https://doi.org/10.1016/j.molcel.2016.02.020

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., Elowitz, M.B., 2016. Dynamics of epigenetic regulation at the single-cell level. Science (80-.). 351, 720–724. https://doi.org/10.1126/science.aab2956

Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D., Dusko, S., Correspondence, E., Ehrlich, S.D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. 8, 151. https://doi.org/10.1099/mic.0.28048-0

Bothmer, A., Robbiani, D.F., Feldhahn, N., Gazumyan, A., Nussenzweig, A., Nussenzweig, M.C., 2010. 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. J. Exp. Med. 207, 855–65. https://doi.org/10.1084/jem.20100244

Botuyan, M.V., Cui, G., Drané, P., Oliveira, C., Detappe, A., Brault, M.E., Parnandi, N., Chaubey, S., Thompson, J.R., Bragantini, B., Zhao, D., Chapman, J.R., Chowdhury, D., Mer, G., 2018. Mechanism of 53BP1 activity regulation by RNA-binding TIRR and a designer protein. Nat. Struct. Mol. Biol. 25, 591–600. https://doi.org/10.1038/s41594-018-0083-z

Bouwman, P., Aly, A., Escandell, J.M., Pieterse, M., Bartkova, J., van der Gulden, H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., Haffty, B.G., Tommiska, J., Blomqvist, C., Drapkin, R., Adams, D.J., Nevanlinna, H., Bartek, J., Tarsounas, M., Ganesan, S., Jonkers, J., 2010. 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat. Struct. Mol. Biol. 17, 688–695. https://doi.org/10.1038/nsmb.1831

Brazelton, V.A., Zarecor, S., Wright, D.A., Wang, Y., Liu, J., Chen, K., Yang, B., Lawrence-Dill, C.J., 2015. A quick guide to CRISPR sgRNA design tools. GM Crops Food 6, 266–276. https://doi.org/10.1080/21645698.2015.1137690

Bunting, S.F., Callén, E., Wong, N., Chen, H.-T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C.-X., Finkel, T., Nussenzweig, M., Stark, J.M., Nussenzweig, A., 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. Cell 141, 243–254.

https://doi.org/10.1016/j.cell.2010.03.012

Cho, S.W., Kim, S., Kim, J.M., Kim, J.-S., 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. 31, 230–232. https://doi.org/10.1038/nbt.2507

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., Kim, J.-S., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132–141. https://doi.org/10.1101/gr.162339.113

Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., Kühn, R., 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat. Biotechnol. 33, 543–548. https://doi.org/10.1038/nbt.3198

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). 339, 819–823. https://doi.org/10.1126/science.1231143

Daley, J.M., Sung, P., 2014. 53BP1, BRCA1, and the choice between recombination and end joining at DNA double-strand breaks. Mol. Cell. Biol. 34, 1380–8. https://doi.org/10.1128/MCB.01639-13

de Jager, M., van Noort, J., van Gent, D.C., Dekker, C., Kanaar, R., Wyman, C., 2001. Human Rad50/Mre11 is a flexible complex that can tether DNA ends. Mol. Cell 8, 1129–35.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S., 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. 190, 1390–400. https://doi.org/10.1128/JB.01412-07

Di Virgilio, M., Callen, E., Yamane, A., Zhang, W., Jankovic, M., Gitlin, A.D., Feldhahn, N., Resch, W., Oliveira, T.Y., Chait, B.T., Nussenzweig, A., Casellas, R., Robbiani, D.F., Nussenzweig, M.C., 2013. Rif1 Prevents Resection of DNA Breaks and Promotes Immunoglobulin Class Switching. Science. 339, 711–715. https://doi.org/10.1126/science.1230624

Doench, J.G., Hartenian, E., Graham, D.B., Tothova, Z., Hegde, M., Smith, I., Sullender, M., Ebert, B.L., Xavier, R.J., Root, D.E., 2014. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. Nat. Biotechnol. 32, 1262–1267. https://doi.org/10.1038/nbt.3026

Escribano-Díaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T.F., Tkáč, J., Cook,

M.A., Rosebrock, A.P., Munro, M., Canny, M.D., Xu, D., Durocher, D., 2013. A Cell Cycle-Dependent Regulatory Circuit Composed of 53BP1-RIF1 and BRCA1-CtIP Controls DNA Repair Pathway Choice. Mol. Cell 49, 872–883. https://doi.org/10.1016/j.molcel.2013.01.001

Feng, L., Fong, K.-W., Wang, J., Wang, W., Chen, J., 2013. RIF1 Counteracts BRCA1-mediated End Resection during DNA Repair. J. Biol. Chem. 288, 11135–11143. https://doi.org/10.1074/jbc.M113.457440

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826. https://doi.org/10.1038/nbt.2623

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I., Liu, D.R., 2017. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature 551, 464–471. https://doi.org/10.1038/nature24644

Giglia-Mari, G., Zotter, A., Vermeulen, W., 2011. DNA damage response. Cold Spring Harb. Perspect. Biol. 3, a000745. https://doi.org/10.1101/cshperspect.a000745

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. 33, 510–517. https://doi.org/10.1038/nbt.3199

Hirano, S., Nishimasu, H., Ishitani, R., Nureki, O., 2016. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. Mol. Cell 61, 886–894. https://doi.org/10.1016/j.molcel.2016.02.018

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J. Bacteriol. 190, 1401–12. https://doi.org/10.1128/JB.01415-07

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., Zhang, F., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 31, 827–832. https://doi.org/10.1038/nbt.2647

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., Joung, J.K., 2013. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. 31, 227–229. https://doi.org/10.1038/nbt.2501

Ishino, Y., Krupovic, M., Forterre, P., 2018. History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. J. Bacteriol. 200, 7. https://doi.org/10.1128/jb.00580-17

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science (80-.). https://doi.org/10.1126/science.1225829

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., Joung, J.K., 2016. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490–495. https://doi.org/10.1038/nature16526

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V. V, Zheng, Z., Joung, J.K., 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. 33, 1293–1298. https://doi.org/10.1038/nbt.3404

Lander, E.S., 2016. The Heroes of CRISPR. Cell 164, 18–28. https://doi.org/10.1016/j.cell.2015.12.041

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., Qi, L.S., 2013. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. 8, 2180–2196. https://doi.org/10.1038/nprot.2013.132

Li, L., Zou, L., 2005. Sensing, signaling, and responding to DNA damage: organization of the checkpoint pathways in mammalian cells. J. Cell. Biochem. 94, 298–306. https://doi.org/10.1002/jcb.20355

Limbo, O., Chahwan, C., Yamada, Y., de Bruin, R.A.M., Wittenberg, C., Russell, P., 2007. Ctp1 Is a Cell-Cycle-Regulated Protein that Functions with Mre11 Complex to Control Double-Strand Break Repair by Homologous Recombination. Mol. Cell 28, 134–146. https://doi.org/10.1016/j.molcel.2007.09.009

Lin, S., Staahl, B.T., Alla, R.K., Doudna, J.A., 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766. https://doi.org/10.7554/eLife.04766

Lino, C.A., Harper, J.C., Carney, J.P., Timlin, J.A., 2018. Delivering CRISPR: a review of the challenges and approaches. Drug Deliv. 25, 1234–1257. https://doi.org/10.1080/10717544.2018.1474964

Lu, R., Wang, G.G., 2013. Tudor: a versatile family of histone methylation 'readers.' Trends Biochem. Sci. 38, 546–555. https://doi.org/10.1016/j.tibs.2013.08.002

Makarova, K.S., Grishin, N. V, Shabalina, S.A., Wolf, Y.I., Koonin, E. V, 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct 1, 7. https://doi.org/10.1186/1745-6150-1-7

Makino, K., Nakata, A., Ishino, Y., Amemura, M., Shinagawa, H., 2016. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. 169, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., Church, G.M., 2013a. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838. https://doi.org/10.1038/nbt.2675

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013b. RNA-guided human genome engineering via Cas9. Science (80-.). 339. https://doi.org/10.1126/science.1232033

Mojica, F.J., Ferrer, C., Juez, G., Rodríguez-Valera, F., 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. Mol. Microbiol. 17, 85–93.

Mojica, F.J., Juez, G., Rodríguez-Valera, F., 1993. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. Mol. Microbiol. 9, 613–21.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174–82. https://doi.org/10.1007/s00239-004-0046-3

Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.-D., Fernandez, J.P., Mis, E.K., Khokha, M.K., Giraldez, A.J., 2015. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. Nat. Methods 12, 982–988. https://doi.org/10.1038/nmeth.3543

Panier, S., Boulton, S.J., 2014. Double-strand break repair: 53BP1 comes into focus. Nat. Rev. Mol. Cell Biol. 15, 7–18. https://doi.org/10.1038/nrm3719

Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653–663. https://doi.org/10.1099/mic.0.27437-0

Ran, F.A., Hsu, P.D., Lin, C.-Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., Zhang, F., 2013. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell 154, 1380–1389. https://doi.org/10.1016/j.cell.2013.08.021

Robert, F., Barbeau, M., Éthier, S., Dostie, J., Pelletier, J., 2015. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. 7, 93. https://doi.org/10.1186/s13073-015-0215-6

Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. Nature 450, 509–514. https://doi.org/10.1038/nature06337

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., Zhang, F., 2016. Rationally engineered Cas9 nucleases with improved specificity. Science (80-.). 351, 84–88. https://doi.org/10.1126/science.aad5227

Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B., Raghavan, S.C., 2012. An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. Cell 151, 1474–1487. https://doi.org/10.1016/j.cell.2012.11.054

Takeda, S., Nakamura, K., Taniguchi, Y., Paull, T.T., 2007. Ctp1/CtIP and the MRN Complex Collaborate in the Initial Steps of Homologous Recombination. Mol. Cell 28, 351–352. https://doi.org/10.1016/j.molcel.2007.10.016

Tomkinson, A.E., Howes, T.R.L., Wiest, N.E., 2013. DNA ligases as therapeutic targets. Transl. Cancer Res. 2, 1219.

Vartak, S. V., Raghavan, S.C., 2015. Inhibition of nonhomologous end joining to increase the specificity of CRISPR/Cas9 genome editing. FEBS J. 282, 4289–4294. https://doi.org/10.1111/febs.13416

Wang, T., Wei, J.J., Sabatini, D.M., Lander, E.S., 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science. 343, 80–84. https://doi.org/10.1126/science.1246981

Xu, H., Xiao, T., Chen, C.-H., Li, W., Meyer, C.A., Wu, Q., Wu, D., Cong, L., Zhang, F., Liu, J.S., Brown, M., Liu, X.S., 2015. Sequence determinants of improved CRISPR sgRNA design. Genome Res. 25, 1147–1157. https://doi.org/10.1101/gr.191452.115

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., Qi, L.S.S., La Russa, M., Xie, M., Ding, S., Qi, L.S.S., 2015. Small molecules enhance crispr

genome editing in pluripotent stem cells. Cell Stem Cell 16, 142–147. https://doi.org/10.1016/j.stem.2015.01.003

CHAPTER 4: EXAMINING 53BP1 SMALL MOLECULE INHIBITORS EFFECT ON CRISPR-CAS9 GENOME EDITING.

Abstract

In order to achieve specific genomic insertions using CRISPR-Cas9 editing, the DDR to double strand breaks must proceed through the homology-dependent repair pathway (HDR). Small molecule inhibitors have been successful in promoting HDR over NHEJ, increasing the efficiency of precise CRISPR-Cas9 editing. Here we present a 10 compound test using small molecule inhibitors with relevance or history of diminishing the NHEJ or promoting the HDR pathway during CRISPR-Cas9 genome editing. In our study, we assayed ten compounds in a CRISPR-Cas9 system that targets Rosa-26 with a promoterless EmGFP. Thus, there should be a correlation with knock-in and GFP positive cells as only cells that are knocked in at the correct location downstream of the Rosa-26 promoter should express EmGFP . We also evaluated two 53BP1 inhibitors, UNC2170 and MFP6008, which previously had not been tested in CRISPR-Cas9 genome editing studies. We found combinations of UNC2170 or MFP6008 with MLN4924 we are able to increase GFP expression efficiency four-fold. While potentially exciting, we still need to validate these results by polymerase chain reaction (PCR) or sequencing and in other systems to examine the effects of small molecule inhibition on favoring HDR pathways.

Introduction

CRISPR-Cas9 genome editing is becoming a staple of genetic research with the ability to edit any desired target gene. CRISPR-Cas9 genome editing is the programmed response to double strand breaks initiated by Cas9 undergoing DNA damage repair (DDR). DNA damage repair can either proceed through the homology directed repair (HDR) or non-homologous end joining (NHEJ) pathway. To perform precise genome edits, such as inserting a specific gene in a particular locus, the cell will need to undergo the HDR pathway, which involves repairing the double strand break by copying template DNA strand (containing desired genome edits) into the location of the double strand break. Unfortunately, the HDR pathway is occurs less frequently (0.5-20%) compared to NHEJ (20-60%). This low HDR occurrence rate makes it difficult and time-costly to procure a precise genome edit, with many clones having to be evaluated to obtain a correct edit. Several studies have focused on identifying small molecules which increase the occurrence of HDR over NHEJ. Examples of such molecules are Scr7, L755507, Brefeldin A, and mix containing multiple compounds (NU7026, Trichostatin A, MLN4924, and NSC 15520) (Robert et al., 2015; Srivastava et al., 2012; Tomkinson et al., 2013; Vartak and Raghavan, 2015; Yu et al., 2015). Additionally, a 53BP1 peptide inhibitor has been shown to increase precise genome editing approximately two-fold (Canny et al., 2018).

To identify and compare small molecules which can increase the efficiency of precise CRISPR-Cas9 genome editing, we performed a ten-compound assay with notable HDR or NHEJ interfering small molecules from literature testing for GFP expression following programed EmGFP gene insertion. Among the ten small molecules tested were 53BP1 inhibitors, MFP6008 and UNC2170, previously uncharacterized in regulation of HDR over NHEJ (Gupta et al., 2014). As it was shown previously that promoting or repressing multiple key steps in the HDR and

NHEJ can further increase genome editing efficiency (Riesenberg and Maricic, 2018), we performed additional experiments with dual compound treatments (with either MFP6008 or UNC2170).

The eight compounds tested in our screen were UNC2170, UNC2892, MFP6008, (+)-JQ1, NU7441, NSC19630, MLN4924, KI696, Resveratrol, and Trichostatin A. The key steps which they regulate are described in Supplemental Figure 4.3. UNC2170 and MFP6008 inhibit 53BP1, a negative regulator of end-resection. NU7441 inhibits DNA-PKcs binding within the NHEJ pathway, which inhibits further progression in the NHEJ pathway (Riesenberg and Maricic, 2018). Trichostatin A and Resveratrol enhance ATM, promoting homologous recombination (Riesenberg and Maricic, 2018). MLN4924 enhances CtIP which promotes end resection, which likewise promotes homologous recombination. NSC19630 likely blocks the association of RPA to p53 and Rad9, which promotes availability of RPA for 3' coating of ssRNA, thereby promoting homologous recombination (Riesenberg and Maricic, 2018). KI696, a KEAP1 inhibitor results in a decrease in PALB2 degradation, which allows for increases in PALB2-BRACA2 formation and RAD51 foci formation promoting homologous recombination (Orthwein et al., 2015). Similarly, we also tested (+)-JQ1, a bromodomain inhibitor which inhibits DNA-PKs downregulating the NHEJ pathway (Aggarwal et al., 2011; Li et al., 2018; Samadder et al., 2016).

In this study, we were able to increase GFP expression by roughly four-fold using a combination of a 53BP1 small molecule inhibitor (either UNC2170 or MFP6008) inhibiting negative regulators of end resection and a small molecule promoting end resection (MLN4924). Our current assay only has measured for GFP expression by manner of a promoterless GFP template with homology to the Rosa26 gene locus right behind the promoter. Thus, the assay

should correlate with insertion of this construct into the promoter region of this gene. However, we still need to evaluate and measure HDR vs. NHEJ repair by other assays in the future such as PCR and sequencing or similar method to evaluate if genome editing was successful in these assays. We also have to date only tested this in one cell line with one targeting construct, future studies should test the applicability of this approach with other genes and in other cell types.

Results:

Canny et al. demonstrated that inhibition of 53BP1 using a 53BP1 peptide inhibitor (pcDNA3-Flag::UbvG08 I44A, deltaGG) favored HDR facilitating CRISPR-Cas9 genome-editing efficiency (Canny et al., 2018). This peptide was a ubiquitin variant peptide which bound to the a 53BP1 Tudor domain, blocking the site from further chromatin binding, and limiting 53BP1 accumulation at DNA damage sites. Transfection of this peptide during CRISPR-Cas9 genome editing facilitated the HDR pathway (Canny et al., 2018). In this experiment, a dysfunctional mutant peptide (pcDNA3-Flag::UbvG08 P69L, L70V, I44A, deltaGG) was also constructed; it served as a negative control for peptide transfection (Canny et al., 2018). Canny et al. demonstrated that the 53BP1 inhibitor peptide transfection resulted in an approximately two-fold increase in successful HDR. To understand the effects of small molecule 53BP1 inhibition, compared to 53BP1 peptide inhibition (and dysfunctional mutant 53BP1 peptide inhibition), we tested the effects of 53BP1 peptide transfection and 53BP1 small molecule inhibition during programmed gene insertion of EmGFP. Cells were transfected and treated with small molecules as indicated in Figure 4.1A. Additionally, this experiment served as a validation of our technique demonstrating around a two-fold increase in GFP expression efficiency three days after transfection in line with previous reports (Fig. 4S2) (Canny et al., 2018). We found that UNC2170 and MFP6008 facilitate GFP expression after CRISPR-Cas9 system treatment approximately 50% (Fig. 4.1). In the future, we will cross validate this with sequencing. Additionally, small molecules 53BP1 inhibition tended to increase GFP expression efficiency after CRISPER-Cas9 transfection with donor template plasmid in all conditions, even in the 53BP1 peptide inhibitor treated condition (Fig. 4.1B).

Small molecules which increase precise CRISPR-Cas9 genome editing, by promoting homology directed repair pathways, have become more prevalent in the last five years as genome editing has become more popular. Using a mix of small molecules involved with promoting or silencing key DNA repair pathways, we tested to see the effects on CRISPR-Cas9 programmed gene insertion. We treated MEF cells with small molecules UNC2170, UNC2892, MFP6008, (+)-JQ1, NU7441, NSC19630, MLN4924, KI696, Resveratrol, and Trichostatin A during transfection with sgRNA/Cas9 plasmid and targeting plasmid with an EmGFP following a timeline as indicated in Figure 4.2A. Expression of GFP after transfection should correlate EmGFP insertion into the Rosa-26 locus, but this needs to be formally verified by PCR and sequencing. It was observed that all compounds tested generally increased GFP expression (Fig. 4.2Bi). Individual treatments of MLN4924, KI696, and MFP6008 demonstrated a significant ($p<0.05$*) increase in GFP expression following reagent transfection.

Given that small molecule 53BP1 inhibition had not been tested in combination with the other DNA-repair regulating small molecules in previous literature, and that previous studies have been successful in increasing the occurrence of HDR over NHEJ through inhibition of multiple steps in HDR and NHEJ, we wanted to determine the effect of 53BP1 inhibition in combination with other inhibitors which promote homologous recombination. We found that dual treatments with 53BPI small molecule inhibition and any other small molecule (except NU7441, which had toxicity concerns) in the screen generally increased GFP expression following CRISPR-Cas9 system treatment in an additive manner (Fig. 4.2ii.-iii.). When treated in combination with either 53BP1 inhibitor (either UNC2170 or MFP6008), all dual combination treatments but two (UNC2170 combination with NU7441 and MFP6008 combination Resveratrol) resulted in a significant ($p<0.05$*) increase in GFP expression following

transfection. Likewise, when comparing the efficiencies of either MFP6008 solo treatment to dual treatment of MFP6008 and other small molecule inhibitors, it was found that MFP6008 increased significantly ($p<0.05$*) in combination treatment with KI696, Trichostatin A, and (+)-JQ1. Similarly, when comparing efficiencies of individual UNC2170 treatment to combination with other small molecules tested, it was found that MLN4924, (+)-JQ1, KI696, Trichostatin A and NSC19630 significantly ($p<0.05$) increased GFP expression after transfection with template and CRISPER/Cas9 constructs. We also found it noteworthy that the two leading dual treatments with MFP6008 and UNC2170 were combinations with MLN4924 and (+)-JQ1. We found a four-fold increase in homologous recombination efficiency through treatment with a 53BP1 inhibitor (either MFP6008 or UNC2170). This mirrors the efficiency others have seen with similar compounds with CRISPR-Cas9 genome editing, however, uses only two small molecules (either UNC2170 and MLN4924 or MFP6008 and MLN4924).

Discussion:

From our study, we found that small molecule inhibition of 53BP1 (either UNC2170 or MFP6008) increases GFP expression following CRISPR-Cas9 transfection by approximately 50%. We are actively testing the results on the DNA sequence to assay for precise gene insertion following transfection. Furthermore, UNC2170 or MFP6008 supplemented by MLN4924 increases GFP expression after CRISPR-Cas9 transfection with donor EmGFP targeting the Rosa-26 locus by four-fold. Interestingly, both 53BP1 inhibitors and MLN4924 have activity to promote end resection: 53BP1 inhibition results in inhibition of 53BP1 accumulation on chromatin, associated with blocking end resection an important step in HDR. Likewise, MLN4924 enhances CtIP activity to promote end resection. Given the response of these compounds in this assay and that end resection is early on in the HDR pathway, end resection is potentially an important determinant of the mechanism driving the HDR pathway.

Several other of the dual combination treatments resulted in increased homologous recombination efficiencies. Namely, treatment with (+)-JQ1 and a 53BP1 small molecule inhibitor (either MFP6008 or UNC2170) resulted in three-fold GFP activation and Trichostatin A, NSC19630, and KI696 resulted in ~2.5-fold increase in GFP expression after CRISPR-Cas9 transfection and compound treatment. Given the success of others work, testing the additive effects of promoting and negatively regulating multiple steps in the HDR and NHEJ pathways, a logical next step of this study would be to combine effective small molecules together in a combinatorial study to test for further increases to homologous recombination efficiency during CRISPR-Cas9 genome editing. Additionally, we currently have not confirmed successful insertion of the EmGFP into the Rosa-26 locus from this method. PCR followed by sequencing, southern blot, and additional molecular techniques will be used in the future to test successful

precise genome editing, which will further experimentally test our method and results. Additionally, we will use assays that are capable of quantitatively testing resolution of DSB by HDR or NHEJ following CRISPR-Cas9 gene editing. This remains an exciting area for future study as small molecule facilitation of precise genome editing has countless applications both in research laboratories and in emerging CRISPR-Cas9 based therapeutics.

Methods

Wild type MEF cell lines (lines lacking the CiA transgene) were transfected using the protocol provided by the Primary Cell 4D-Nucleofector™ Kit L using the 100µL Nucleocuvette™ vessels. Cell medium was either FluoroBrite DMEM Media (ThermoFisher, A1896701) for imaging, or DMEM (Corning, MT10013CV) for standard cell culture. All media was supplemented with 15% FBS (Gibco, Lot:1972526), 10 mM HEPES pH 7.5, 10 mM NEAA, 0.1% 1000 × 2-betamercaptoethanol (Gibco, 21,985,023), 1% 100× Penn-Strep (Corning, 30-002-CI). To FluoroBrite media, l-Glutamine (Corning, 25005CI) at 4 mM was added.

Prior to each experiment, cell medium and compounds were preincubated in a humidified 37°C/5% CO_2 incubator for 30 minutes. To initiate transfection of cells, 1 million *CiA:Oct4-/-* cells were resuspended in 82µL nucleofector solution, 18µL supplement solution, and 5µg DNA and added to a 100µL Nucleocuvette™. Cells were transfected using program code EH-100, incubated at room temperature for 10 minutes, and resuspended with pre-37°C/5% CO_2 incubator-equilibrated medium and split equally into a 24 well plate with pre-equilibrated medium and compounds. 24 hours after transfection, media was replaced by compound supplemented medium and puromycin at 2.5µg/ml. After 24 hours puromycin selection, media was replaced with compound supplemented medium for one additional day, and then with normal media every two days. Cells were analyzed by Flow Cytometry on the iQue Screener Plus on the days indicated in Fig. 4.1A and Fig. 4.2A. Analysis gating was performed using FlowJo as indicated (Fig. 4.S1A-B).

DNA for peptide small molecule comparison assay was 1.6µg KH501 (adapted from pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene, Plasmid #62988) with sgRNA (CACCGTCGTGATCTGCAACTCCAGT)), 1.6µg pR26-SA-EmGFPex, and 1.6µg either

pcDNA3-Flag::UbvG08 I44A, deltaGG (Addgene, Plasmid #74939) or pcDNA3-Flag::UbvG08 P69L, L70V, I44A, deltaGG (Addgene, Plasmid #74940) DNA for small molecule screen was made up of 2.5µg KH501 and 2.5µg pR26-SA-EmGFPex.

Cells were treated at different compound concentrations primarily based on literature evaluations: UNC2170 (Perfetti et al., 2015): 100µM, UNC2892 (Perfetti et al., 2015): 100µM, MFP6008*: 100µM, (+)-JQ1 (Li et al., 2018; Wang et al., 2015): 0.2µM, NU7441 (Riesenberg and Maricic, 2018) 20µM, NSC19630 (Aggarwal et al., 2011; Brosh and Jr, 2013): 1µM, MLN4924 (Riesenberg and Maricic, 2018): 0.5µM, KI696 (Orthwein et al., 2015): 1µM, Resveratrol (Riesenberg and Maricic, 2018): 1µM, Trichostatin A (Riesenberg and Maricic, 2018): 0.01µM. (*=identified in a fragment-based screen by Collaborator Dr. Stephen Hedley at Monash University)

REFERENCES

Aggarwal, M., Sommers, J.A., Shoemaker, R.H., Brosh, R.M., 2011. Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. Proc. Natl. Acad. Sci. 108, 1525–1530. https://doi.org/10.1073/pnas.1006423108

Brosh, R.M., Jr, 2013. DNA helicases involved in DNA repair and their roles in cancer. Nat. Rev. Cancer 13, 542–58. https://doi.org/10.1038/nrc3560

Canny, M.D., Moatti, N., Wan, L.C.K., Fradet-Turcotte, A., Krasner, D., Mateos-Gomez, P.A., Zimmermann, M., Orthwein, A., Juang, Y.-C., Zhang, W., Noordermeer, S.M., Seclen, E., Wilson, M.D., Vorobyov, A., Munro, M., Ernst, A., Ng, T.F., Cho, T., Cannon, P.M., Sidhu, S.S., Sicheri, F., Durocher, D., 2018. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR–Cas9 genome-editing efficiency. Nat. Biotechnol. 36, 95–102. https://doi.org/10.1038/nbt.4021

Gupta, A., Hunt, C.R., Chakraborty, S., Pandita, R.K., Yordy, J., Ramnarain, D.B., Horikoshi, N., Pandita, T.K., 2014. Role of 53BP1 in the regulation of DNA double-strand break repair pathway choice. Radiat. Res. 181, 1–8. https://doi.org/10.1667/RR13572.1

Li, X., Baek, G., Ramanand, S.G., Sharp, A., Gao, Y., Yuan, W., Welti, J., Rodrigues, D.N., Dolling, D., Figueiredo, I., Sumanasuriya, S., Crespo, M., Aslam, A., Li, R., Yin, Y., Mukherjee, B., Kanchwala, M., Hughes, A.M., Halsey, W.S., Chiang, C.-M., Xing, C., Raj, G. V, Burma, S., de Bono, J., Mani, R.S., 2018. BRD4 Promotes DNA Repair and Mediates the Formation of TMPRSS2-ERG Gene Rearrangements in Prostate Cancer. Cell Rep. 22, 796–808. https://doi.org/10.1016/j.celrep.2017.12.078

Orthwein, A., Noordermeer, S.M., Wilson, M.D., Landry, S., Enchev, R.I., Sherker, A., Munro, M., Pinder, J., Salsman, J., Dellaire, G., Xia, B., Peter, M., Durocher, D., 2015. A mechanism for the suppression of homologous recombination in G1 cells. Nature 528, 422–426. https://doi.org/10.1038/nature16142

Perfetti, M.T., Baughman, B.M., Dickson, B.M., Mu, Y., Cui, G., Mader, P., Dong, A., Norris, J.L., Rothbart, S.B., Strahl, B.D., Brown, P.J., Janzen, W.P., Arrowsmith, C.H., Mer, G., McBride, K.M., James, L.I., Frye, S. V., 2015. Identification of a Fragment-like Small Molecule Ligand for the Methyl-lysine Binding Protein, 53BP1. ACS Chem. Biol. 10, 1072–1081. https://doi.org/10.1021/cb500956g

Riesenberg, S., Maricic, T., 2018. Targeting repair pathways with small molecules increases precise genome editing in pluripotent stem cells. Nat. Commun. 9, 2164. https://doi.org/10.1038/s41467-018-04609-7

Robert, F., Barbeau, M., Éthier, S., Dostie, J., Pelletier, J., 2015. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. Genome Med. 7, 93. https://doi.org/10.1186/s13073-015-0215-6

Samadder, P., Aithal, R., Belan, O., Krejci, L., 2016. Cancer TARGETases: DSB repair as a pharmacological target. Pharmacol. Ther. 161, 111–131. https://doi.org/10.1016/J.PHARMTHERA.2016.02.007

Srivastava, M., Nambiar, M., Sharma, S., Karki, S.S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R.K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B., Raghavan, S.C., 2012. An Inhibitor of Nonhomologous End-Joining Abrogates Double-Strand Break Repair and Impedes Cancer Progression. Cell 151, 1474–1487. https://doi.org/10.1016/j.cell.2012.11.054

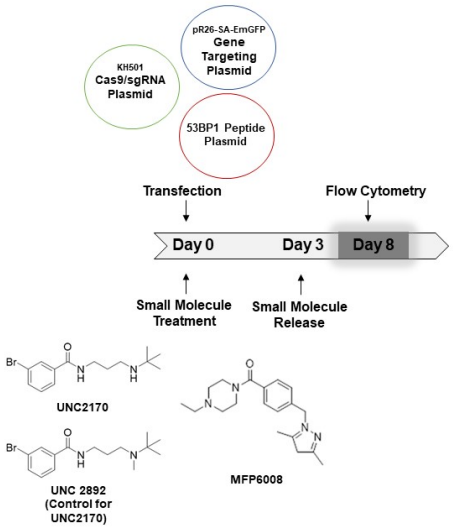
Tomkinson, A.E., Howes, T.R.L., Wiest, N.E., 2013. DNA ligases as therapeutic targets. Transl. Cancer Res. 2.

Vartak, S. V., Raghavan, S.C., 2015. Inhibition of nonhomologous end joining to increase the specificity of CRISPR-Cas9 genome editing. FEBS J. 282, 4289–4294. https://doi.org/10.1111/febs.13416

Wang, L., Xie, L., Ramachandran, S., Lee, Y., Yan, Z., Zhou, L., Krajewski, K., Liu, F., Zhu, C., Chen, D.J., Strahl, B.D., Jin, J., Dokholyan, N. V, Chen, X., 2015. Non-canonical Bromodomain within DNA-PKcs Promotes DNA Damage Response and Radioresistance through Recognizing an IR-Induced Acetyl-Lysine on H2AX. Chem. Biol. 22, 849–61. https://doi.org/10.1016/j.chembiol.2015.05.014

Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S., Qi, L.S.S., La Russa, M., Xie, M., Ding, S., Qi, L.S.S., 2015. Small molecules enhance crispr genome editing in pluripotent stem cells. Cell Stem Cell 16, 142–147. https://doi.org/10.1016/j.stem.2015.01.003

A.



B.

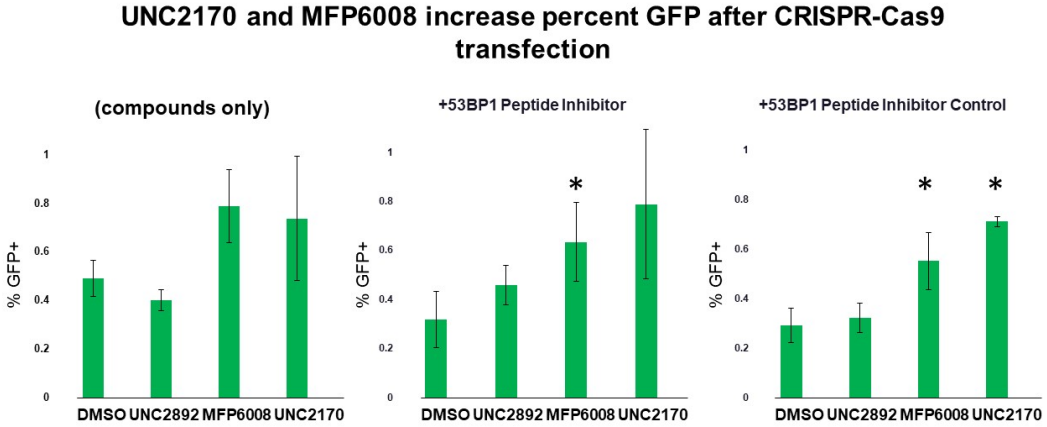


Figure 4.1: Small molecules, UNC2170 and MFP6008, facilitate GFP activation after CRISPR-Cas9 transfection. **(A)** Schematic representation of experimental timeline. Transfection of sgRNA/Cas9, targeting plasmid, and 53BP1 peptide (or control peptide) on Day 0. Small molecule were removed from media on Day 3 and flow cytometry analysis was performed on Day 8. **(B)** Percent GFP+ cells from flow cytometry analysis on Day 8 after transfection with 53BP1 peptide inhibitor, and treatment with small molecule 53BP1 inhibitors, UNC2170 and MFP6008. (Heteroscedastic, two tailed distribution t-test $p<0.05$*, n=3)

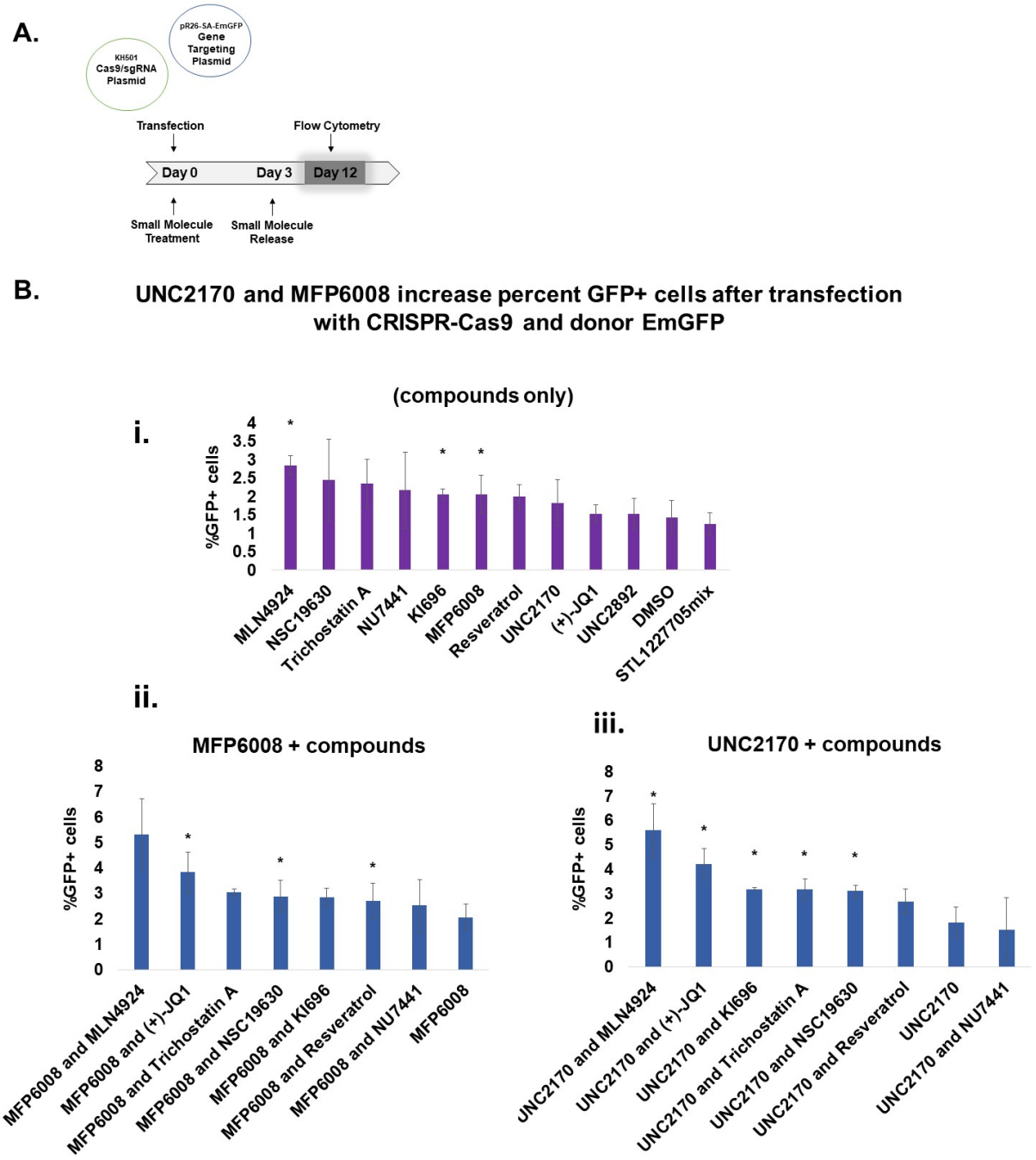


Figure 4.2: Small molecules, UNC2170 and MFP6008, increase percent GFP+ cells after CRISPR-Cas9 transfection. **(A)** Schematic representation of experimental timeline. Transfection of sgRNA/Cas9, targeting plasmid, and 53BP1 peptide (or control peptide) on Day 0. Small molecules were removed from media on Day 3 and flow cytometry analysis was performed on Day 8. B) (*$p<0.05$, n=3) **(B)** Percent GFP+ cells twelve days after transfection. Cells were treated with compounds (i), compounds + MFP6008 (ii), or compounds + UNC2170 (iii). (Heteroscedastic, two tailed distribution t-test for (i) compared to DMSO, (ii) compared to MFP6008, (iii) compared to UNC2170, $p<0.05$*, n=3 or 6.)

Figure 4.1 Gating

A.

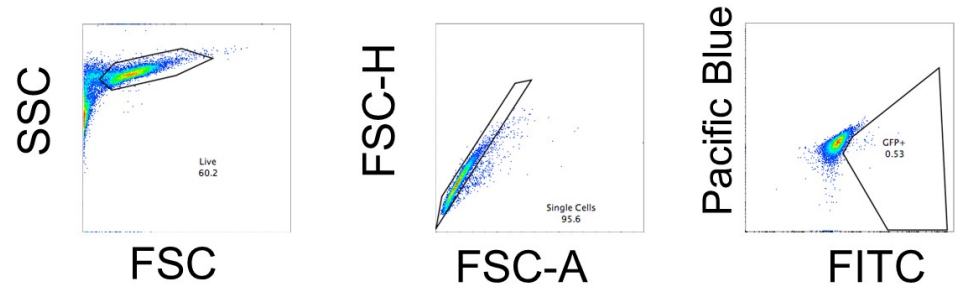
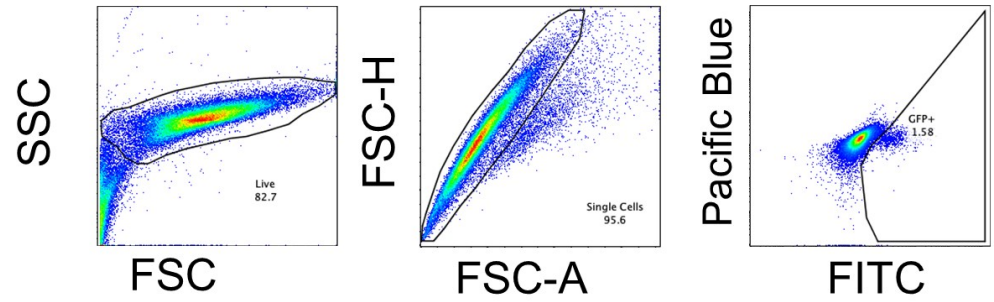
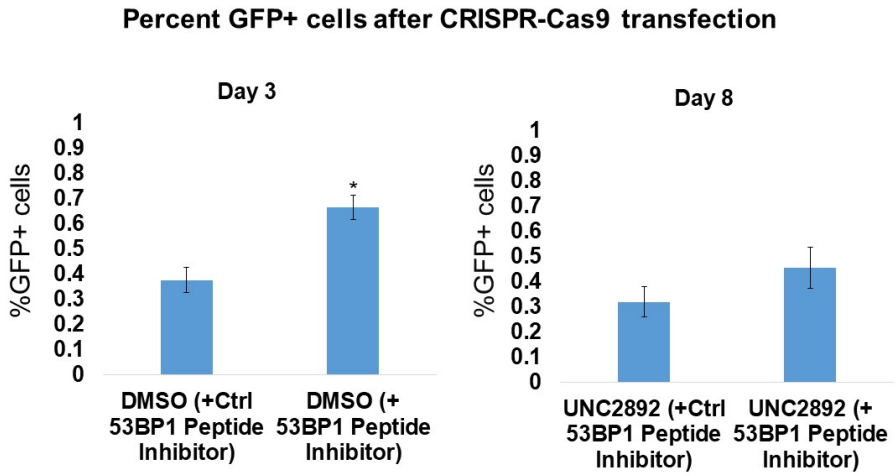


Figure 4.2 Gating

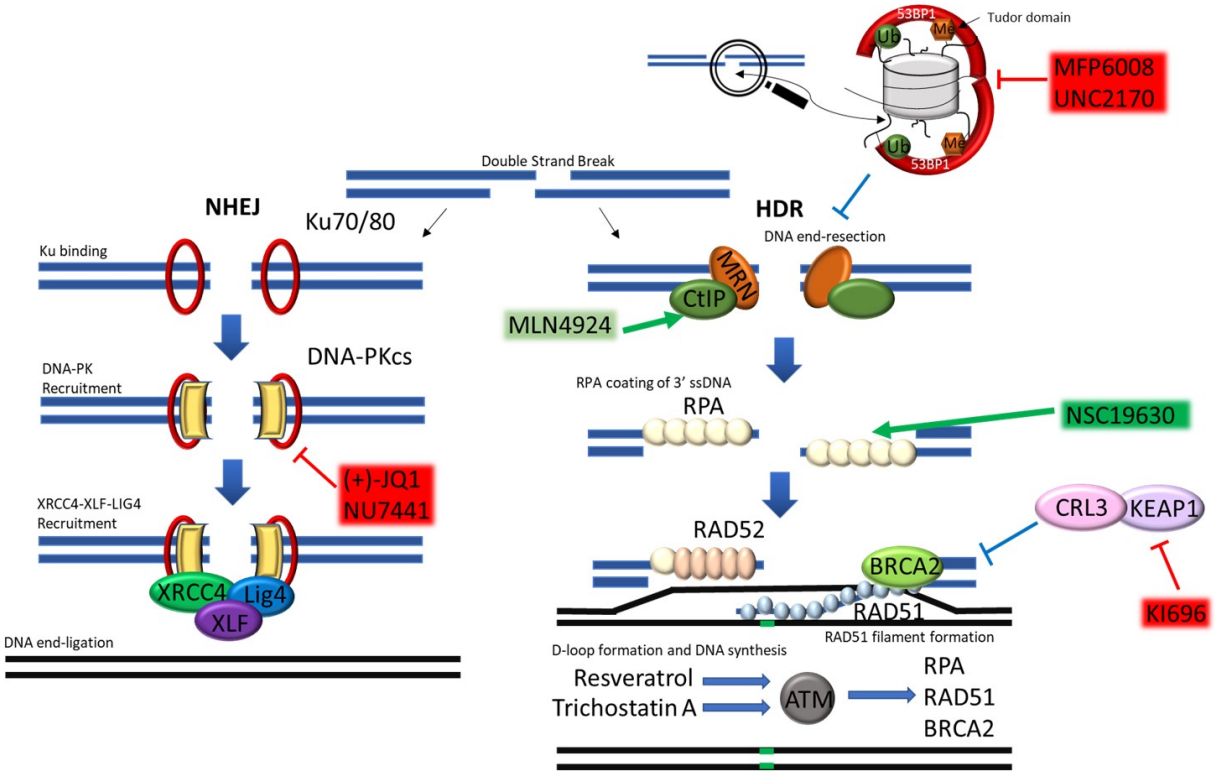
B.



Supplemental Figure 4.1: General gating strategy with an example sample for each. **(A)** Flow cytometry gating for Figure 4.1B and Supplementary Figure 4.S2. **(B)** Flow cytometry gating for Figure 4.2B.



Supplemental Figure 4.2: 53BP1 peptide inhibitor effect on GFP expression after CRISPR-Cas9 transfection. Flow cytometry analysis at Day 3 and Day 8 comparing increases in percent GFP with 53BP1 Peptide inhibitor or 53BP1 peptide inhibitor negative control. (Heteroscedastic, two tailed distribution t-test. $p<0.05$*, n=3)



Supplemental Figure 4.3: Small molecules: MLN4924, NSC19630, Trichostatin A, NU7441, KI696, MFP6008, Resveratrol, UNC2170, and (+)-JQ1 anticipated mode of action for either promoting (green) or negatively regulating (red) key steps in either the NHEJ or HDR pathway. MFP6008 and UNC2170 inhibit the DNA damage response protein, 53BP1 which binds to H4K20me2 modifications induced by double-strand DNA breaks, blocking end-resection, and inhibiting BRCA1 recruitment to DSB sites. (+)-JQ1 and NU7441 decrease the efficacy of DNA-PKcs. NSC19630 is a WRN helicase inhibitor which likely blocks the association of RPA to p53 and RAD9. KI696 is a KEAP1 inhibitor. Freeing up PALB2, by inhibiting KEAP1-CRL3, allows for PALB2-BRCA2-BRCA1 complex recruitment promoting RAD51 filament formation in the HDR pathway. Trichostain A and Resveratrol enhance ATM which in turn enhance factors which facilitate the HDR pathway.

CHAPTER 5: DISCUSSION

Conclusions and future directions

Understanding the barriers to *Oct4* activation in somatic cells is critical for efficient and safe methods of iPSC generation. We conducted a high throughput small molecule screen (~960 small molecules with epigenetic relevance) identifying small molecules which could facilitate *Oct4* single allele activation in *CiA:Oct4* mouse embryonic fibroblast (MEF) cells. From this screen, four small molecules were identified which facilitate *Oct4* activation as measured by GFP florescence expressed from the reporter allele up to ~50% with simultaneous recruitment of the transcription factor, VP16. All of the lead small molecules identified in our screen: Mocetinostat, Tacedinaline, Entinostat and Azacytidine outperformed VPA, TSA, and SAHA in single allele activation of *Oct4*. VPA, TSA, and SAHA are all effective in increasing iPSC generation during transcription factor reprogramming. These findings highlighted the difference between small molecule direct single allele activation and network cellular reprogramming of *Oct4* with the OSKM factors, which affect hundreds of genes.

To understand the dynamics of small molecule inhibition on *Oct4* activation, high content time-lapse imaging was performed. With this analysis, we were able to directly visualize and measure HDACi and DNMTi induced whole population activation. With this technique we demonstrated that HDACi induced activation results in quicker activation peaking (~30 hours) compared to DNMTi induced activation (~60 hours). Further dynamics were elucidated during a HDACi and DNMTi induced activation and small molecule release study. From this study, we found that HDACi induced activation was lost after four days, while DNMTi induced activation was retained after four days. This discovery is consistent with models of ‘epigenetic memory’

where DNA-methylation results in irreversible epigenetic memory, while loss of histone acetylation results in reversible epigenetic memory (Bintu et al., 2016). Finally, application of our lead HDACi, Mocetinostat, to reprogramming MEF cells to iPSCs during transcription factor reprogramming demonstrated increased iPSC generation efficiency.

Identification of three HDACi and a DNTMi which outperform VPA, TSA, and SAHA on single allele activation of Oct4.

Before this study, it was clear that epigenetic factors played major roles in successful reprogramming. Other laboratories had tested well known HDACis and DNMTis and were able to increase reprogramming efficiencies (Dokmanovic et al., 2007; Huangfu et al., 2008). Nevertheless, previous studies have looked to identify small molecules that increased iPSC generation under transcription factor reprogramming conditions, which affect hundreds of genes simultaneously in response to high ectopic transcription factor expression. As future iPSC generation methods will likely not be using a method which transfects the oncogenic associated OSKM factors, identification of small molecules that facilitate activation of key pluripotent factors, such as *Oct4,* outside of the transcription factor reprogramming systems are desirable for these techniques to advance tissue engineering. From our high throughput small molecule screen and subsequent dose response study we identified four small molecules that activate single allele *Oct4* with simultaneous recruitment of VP16. HDACi (Tacedinaline, Entinostat, Mocetinostat) and a DNMTi (Azacytidine) activated *Oct4* with efficiencies up to 50% as measured by GFP expression from our reporter allele. Furthermore, it was identified that these newly identified molecules outperformed VPA, TSA, and SAHA which are known to increase *Oct4* activation during transcription factor reprogramming. Future studies could further test small molecules to

understand single allele activation of other pluripotency related genes including Nanog, Sox2, Klf, and c-Myc. Additionally, as this study did not specifically test *Oct4* transcription levels in comparison to other loci, it is possible that the effect quantified in this screen is the result of a global gene response to DNMTi and HDACi. Further tests should quantify the specificity of the expression response of DNMTi and HDACi on several loci in comparison to *Oct4*.

Additionally, an important caveat to these identified top activating small molecules from this screen is that this method utilizes recruitment of transcription factor, VP16, which could prime the *CiA:Oct4* locus for transcriptional activation. Artificially increased acetylation that may result from VP16 recruitment could bias the screen towards identifying histone deacetylase complexes as activators. Conspicuously, notable small molecules in the set such as various histone methyltransferase inhibitors, which we expected to show *CiA:Oct4* activation, did not demonstrate apparent activation with our screen. Notably, small molecules which activate this locus which were not identified by our screen could have been overlooked due to high-background or conducing high throughput screening as a single chosen concentration in our assay. Additionally, a subset of compounds demonstrated noticeable toxicity by cell number and cell morphology at the 10 μM dose we used for our screening assay which could have influenced *CiA:Oct4* activation levels. Some epigenetic regulatory pathways could indeed play a role in reprogramming cell identity, but could have been missed in our high throughput screening approach due to the right concentration not being tested or the possibility that a single screening timepoint did not capture all molecules that influenced this pathway.

Dynamics of HDACi and DNMTi induced single allele activation of Oct4

Epigenetic memory is dynamic, where modifications result in transcriptional effects that fluctuate over time (Bintu et al., 2016). Understanding epigenetic memory and the interplay with epigenetic small molecule regulators is becoming more and more important as many of these small molecules are being used in either the clinics or tested in clinical trials. By using high content time-lapse microscopy, we were able to visualize and calculate population changes in single allele *Oct4* expression over time in response to HDACi or DNMTi treatments. We found that activation peaked earlier in HDACis (~30 hours) compared to DNMTi (~60-70 hours) following small molecule treatment and transcription factor recruitment. This result is suggestive of a model of gene dynamics where accumulation of acetylation results in fast gene activation while gene activation from methylation loss is slower activation. Notably, to test this model in more depth would require testing additional loci and cell lines. Another avenue that has not been explored is testing the degree to which HDACi or DNMTi enhanced expression of *Oct4*, which could be calculated by intensity of the GFP in the pictures. Additionally, further studies can be performed to understand the dynamics of the small molecules treated concomitantly or sequentially to reveal the interplay between these pathways and single allele Oct4 activation.

Dynamics of HDACi and DNMTi induced single allele activation of Oct4 and subsequent small molecule release.

We also sought to understand the reversibility of these small molecules during single allele Oct4 activation. Using a four day treat and release method and analysis with flow cytometry, we were able to identify differences in loss of activation between DNTMi and HDACi. Interestingly, DNMTi induced activation was sustained after release of compound treatment, while HDACi induced activation was lost. This small molecule effect mirrors the idea

that acetylation is a 'reversible' epigenetic memory, while DNA methylation is a 'irreversible' epigenetic memory (Bintu et al., 2016). Importantly, to understand if this is a specific effect on Oct4, or is a larger more generalized effect on the genome, qPCR would need to be performed looking at changes in gene expression upon HDACi and DNMTi treatment and release at several different loci.

Single-cell tracking of HDACi and DNMTi induced *CiA:Oct4* activation.

From our study, we tracked single cell *CiA:Oct4* gene expression during a 60-hour video, obtained from high content time-lapse imaging. We did not identify any notable differences between DNMTi induced activation and HDACi induced activation, and noted that activation tended to be spontaneous and heritable. However, our sample size for this study was fairly small with ~12-20 starting cells and subsequent progeny. A clearer understanding of DNMTi and HDACi induced gene activation and potential differences between the two processes could be elucidated by a larger sample size.

Mocetinostat's effect on reprogramming efficiency

Understanding that Mocetinostat treatment increases single allele activation of *Oct4* in our experimental conditions, we sought also to identify Mocetinostat effect on iPSC generation. Excitingly, we found that Mocetinostat treatment increased iPSC generation efficiency ~20 fold after four days of small molecule and doxycycline treatment. This demonstrates that our screen identifying small molecules which facilitate single allele activation of Oct4 identified, a small molecule previously uncharacterized in reprogramming studies. Subsequent studies can further test the other two small molecules, Tacedinaline and Entinostat in iPSC generation. In another

avenue of study, Mocetinostat could be combined with current cocktails with facilitate iPSC generation. Furthermore, if subsequent testing reveals a specific interaction with Mocetinostat to *Oct4* activation, Mocetinostat could be tested to see if it can replace *Oct4* in current iPSC generation transcription factor reprogramming methods. Since Moceinostat works to increase reprogramming efficiency from MEFs to iPSCs, it would be interesting to test its effect in other cell-fate conversion methods such as direct fibroblast neuron transcription factor driven transdifferentiation.

REFERENCES

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., Elowitz, M.B., 2016. Dynamics of epigenetic regulation at the single-cell level. Science (80-.). 351, 720–724. https://doi.org/10.1126/science.aab2956

Dokmanovic, M., Clarke, C., Marks, P.A., 2007. Histone deacetylase inhibitors: overview and perspectives. Mol. Cancer Res. 5, 981–989.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., Melton, D.A., 2008. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat. Biotechnol. 26, 795–7. https://doi.org/10.1038/nbt1418