Validating Guide RNA Hits For Improving CRISPRa-Reprogramming

Master's thesis
University of Turku
Department of Life Technologies
Molecular Systems Biology
01. 2021
Sign. _____
Harri Heikkinen

ABSTRACT

UNIVERSITY OF TURKU
Department of Life Technologies
HARRI HEIKKINEN: Validating guide RNA hits for improving CRISPRa-reprogramming
Master's thesis, 60 p,
Biochemistry
01. 2021

Somatic cells can be reprogrammed into induced pluripotent stem cells (iPSCs) by ectopically overexpressing the transcription factors OCT4, SOX2, KLF4 and c-Myc. This technique has since its invention been a topic of intense research and hold great promise for use in biomedicine. However, the reprogramming process itself remains inefficient and slow. Recently CRISPR/Cas9-mediated gene activation (CRISPRa) has been employed for the reprogramming of the human somatic cells into iPSCs. CRISPRa has proven advantageous in many aspects, such as the ability to activate many genes simultaneously and endogenously. In this method, precisely targetable guide RNAs mediate an enzymatically inactive mutant of Cas9, nuclease-dead Cas9 (dCas9)-initiated activation of the target sequence through a transactivator protein. However, CRISPRa still needs to be optimized as it is not very efficient. For optimizing the CRISPRa method, we need to target additional endogenous genomic sequences which are known to be able to increase reprogramming efficiency. Prior CRISPRa screen, carried out by the Otonkoski research group, identified enrichment of genes *KLF2, KLF5* and *KLF17*, which are expected to increase reprogramming efficiency. In addition, Otonkoski and his group did a single cell RNA sequencing at different time points during reprogramming, which identified two of the expressed candidate genes: epithelial cell adhesion molecule (*EpCAM*) and *ZNF486.*

The aim of this thesis was to validate the enriched guides of *KLF2, KLF5* and *KLF17* for their endogenous gene activation efficiency, and to create and validate functional guides for the endogenous activation of *EpCAM* and *ZNF486*. In addition, our task was to define the effect the expression of these guides have in the pluripotent reprogramming process.

The results of this thesis describe endogenous gene activation of targeted sequences using synthesized guides, with a correlation between efficiency and the distance to the gene promoter site. Additionally, while the reprogramming of human foreskin fibroblast cells using CRISPRa was successful, the incorporation of guides for *EpCAM* in the cell populations did not increase the efficiency of the reprogramming process. In fact, the presence of EpCAM seemed to inhibit the development of iPSCs, at least in fibroblasts.

Keywords:
Pluripotent stem cells, Pluripotent reprogramming, CRISPR-Cas, iPSCs, CRISPRa, EpCAM, KLF

Table of Contents

Abbreviations

AP	Alkaline phosphatase
bp	Base pairs
Cas	CRISPR associated protein
cDNA	Complementary DNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activation
crRNA	CRISPR RNA
DD	Dihydrofolate reductase-derived destabilisation domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EB	Embryoid body
EBNA	Eppstein-Barr virus nuclear antigen
E. coli	*Escherichia coli*
EDTA	Ethylenediamine-tetraacetic acid
EEA	EGA-enriched Alu element motif
EMT	Epithelial to mesenchymal transition
EpCAM	Epithelial Cell Adhesion Molecule
ESC	Embryonic stem cell
E8	Essential 8 medium
et al.	And others ("*et alteri*")
Fw	Forward
GFP	Green fluorescent protein
gRNA	Guide RNA
HEK293	Human embryonic kidney cells 293
HFF	Human foreskin fibroblast
H3K27ac	Histone H3 lysine 27 acetylation mark
H3K27me3	Histone H3 lysine 27 trimethylation mark
H3K9me3	Histone H3 lysine 9 trimethylation mark
H3K4	Histone H3 lysin 4 methylation mark
ICC	Immunocytochemistry
iPSC	Induced pluripotent stem cell
KLF	Krüppel-like factor

LCL	Lymphoblastoid cell line
M	Molar (mol/l)
MET	Mesenchymal to epithelial transition
miRNA	Micro RNA
miR	Micro RNA
mRNA	Messenger RNA
OCT4	Octamer binding protein 4
OriP	Origin of plasmid replication
PAM	Protospacer adjacent motif
PCR	Polymerase chain reaction
PFA	Paraformaldehyde pH
Phusion	*Pyrococcus*-like enzyme fused with processivity-enhancing domain
Rv	reverse
RNA	Ribonucleic acid
ROCK	Rho-associated protein kinase
Rpm	Revolutions per minute
RT	Room temperature
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
scRNA-seq	Single cell RNA sequencing
sgRNA	Single guide RNA
SOX2	Sex-determining region Y-box containing Transcription factor 2
TAE	Tris-acetate-EDTA-buffer
TBE	Tris-borate-EDTA-buffer
TMP	Trimethoprim
tracrRNA	Trans-activating RNA
Tris	Tris-(hydroxymethyl)- aminomethane
VP	Viral protein
ZFP	Zinc-finger protein

1 Introduction

1.1 Pluripotent stem cells

Stem cells are classified as cells that can both differentiate into other types of cells and divide to produce more stem cells (Romito and Cobellis, 2016). After zygote formation the first cells forming the morula are totipotent stem cells, going onward to form the embryonic and placental tissues (Condic, 2014). The cells forming the embryo further differentiate into pluripotent stem cells, leaving behind the ability to form placental tissues (Condic, 2014). Pluripotency is a transient state *in vivo*, and the pluripotent cells soon differentiate into different adult tissues, losing their pluripotency (Condic, 2014). However, pluripotent stem cells can be derived from the early embryo and maintained indefinitely *in vitro* (Evans & Kaufman, 1981; Thomson et al., 1998). Formerly human pluripotent stem cells used in biomedical research have been obtained from embryonic tissues, destroying the embryo in the process (de Wert and Mummery, 2003). For this reason, the use of embryonic stem cells was limited and controversial. This problem has fortunately been alleviated by the development of method for obtaining induced pluripotent stem cells (iPSCs) from somatic cells by pluripotent reprogramming (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007).

1.2 Pluripotent reprogramming

Pluripotent reprogramming is the process of turning somatic cells into induced pluripotent stem cells by introducing a specific set of reprogramming factors into the cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). The first set of transcription factors for inducing pluripotency, developed by Takahashi and Yamanaka in 2006, consisted of octamer-binding transcription factor 4 (Oct4), sex determining region Y-box 2 (Sox2), krüppel-like factor 4 (Klf4) and c-Myc (Takahashi and Yamanaka, 2006). The combination of these factors is also known as OSKM, deriver from the genes' names' first letters. Later, Thomson et al. developed another set of factors which used LIN28 and NANOG instead of KLF4 and C-MYC, named Thomson factors (Yu et al., 2007). These generated iPSCs are indistinguishable from embryonic stem cells and likewise are pluripotent (Takahashi and Yamanaka, 2006) and the experiments have since sparked intense research in the field. Generation of iPSCs has many applications within disease modeling, regenerative medicine and immunotherapy, while overcoming the major ethic

drawbacks prevalent in embryonic stem cell research (Kiskinis and Eggan, 2010). For example, iPSCs are unique to the host, as they can be derived from patients own somatic cells, bypassing immune rejection (Kiskinis and Eggan, 2010). Generating iPSCs also does not require an embryo to work with, as they can be made from many types of somatic cells (Kiskinis and Eggan, 2010). To this date, a vast amount of different methods of inducing pluripotency have been described, including: different transcription factors, small molecules and chemicals, and mRNAs (Anokye-Danso et al., 2012; Jere Weltner, 2018a). However, the current methods for inducing pluripotency are still far from perfect (Weltner, 2018a). For example, iPSCs may contain some epigenetic marks and possible unwanted residues from the reprogramming process that can affect their differentiation and use in disease modelling (Chin et al., 2012; Halevy and Urbach, 2014). Furthermore, the inefficiency of this process still remains a major hindrance in pluripotent reprogramming, with most methods only inducing pluripotency in only a fraction of cells (Malik and Rao, 2013).

1.3 Reprogramming phases

Cellular reprogramming occurs in a multistep, albeit dynamic, process which is typically thought to happen in two waves: A stochastic first wave acting to suppress somatic genes, and a more deterministic second wave that activate and maintain the expression of pluripotency-related genes (Polo et al., 2012). Following a binding of core reprogramming factors to distal enhancer sequences, and later to promoter regions, C-Myc begins the first transcriptional wave (Soufi et al., 2012). The first three days of the first transcriptional wave facilitate a rapid proliferation of the cells, which also seems to support the self-renewal and maintenance of the iPSCs (Hansson et al., 2015; Polo et al., 2012). This also increases energy demand, and the cell metabolism changes from oxidative phosphorylation to glycolysis (Folmes et al., 2011; Nishimura et al., 2019). In addition, factors regulating chromatin organization and RNA processing are upregulated in this early process (Buganim et al., 2013; Hansson et al., 2015)

After this, the second transcriptional wave occurs with Klf4 working with Oct4 and Sox2 to promote the pluripotent gene expression (Chronis et al., 2017; Polo et al., 2012). This then leads to a more deterministic and hierarchical cell changes and establishes the core pluripotency circuitry, forming mature iPSCs (Polo et al., 2012).

1.4 Mesenchymal-to-epithelial transition (MET) in reprogramming

Epithelial-to-mesenchymal transition (EMT) is a developmental process initiated earliest in the gastrulation stage (R. Li et al., 2010). In EMT, changes in cell contact interactions lead to loss of epithelial features in the cell and development of mesenchymal features (R. Li et al., 2010). This is mediated through the change in expression of about 4000 genes, with the most prevalent changes occurring in downregulation of E-cadherin (which maintains epithelial cohesion) and upregulation of Snail, in turn repressing epithelial regulators (R. Li et al., 2010; Zavadil et al., 2001). EMT is an important process for determining cell fate (R. Li et al., 2010), and it is a part of the process in which early pluripotent cells develop into mesenchymal ones. As such, the reversion of mesenchymal fibroblasts into pluripotent cells have been thought to require the reverse transition, mesenchymal-to-epithelial transition (MET). This has been noted to be the case (Samavarchi-Tehrani et al., 2010).

Samavarchi-Tehrani et al., (2010) classify the reprogramming to fall into three different phases: initiation, maturation and stabilization. They note that the initiation phase involves a strong induction of MET, with upregulation of epithelial junctional components and epithelial like colonies. One of the important factors facilitating the transition from the initiation MET process to the maturation phase is the TGF-b-related bone morphogenetic protein (BMP) signaling pathway (Samavarchi-Tehrani et al., 2010). BMP pathway works in part by inducing miR(microRNA)-205 and miR-200 with the OSKM factors, which promotes MET (Samavarchi-Tehrani et al., 2010). When this process was initiated removal of BMP did not halt the reprogramming process (Samavarchi-Tehrani et al., 2010). This indicates that BMP is mostly relevant in early MET initiation.

Other important components in MET are the TGF-b factors. TGF-b1 is a known EMT inducer and has been noted to block reprogramming, as have TGFb2 and TGFb3 (R. Li et al., 2010). TGF-b1 has been noted to inhibit the upregulation of epithelial markers, and to repress the downregulation of MET-activating Snail (Peinado et al., 2003). The OSKM factors seem to reduce the expression of TGF-β factors, with Sox2 and Oct4 suppressing TGF-β receptor 3, require for activation of Snail, while Oct4 and Klf4 suppressed TGF-β3 (R. Li et al., 2010). c-Myc suppressed TGF-β1, providing complete inhibition of the TGF-β factors (R. Li et al., 2010). In addition, incompletely reprogrammed cells

contained high levels of remaining TGF-β1 and Snail, with low levels of E-cadherin (R. Li et al., 2010). Usage of exogenous TGF-β inhibitors also increased reprogramming efficiency, further suggesting the direct role of TGF-β factors in EMT and reprogramming repression (R. Li et al., 2010). This way, the OSKM factors used in pluripotent reprogramming facilitate EMT by repressing pro-MET factors such as TGF-β1 and Snail.

After MET has initiated the reprogramming, removal of the OSKM factors in 5 days along the initiation reversed the reprogramming process, meaning that the reprogramming is still unstable at this phase (Samavarchi-Tehrani et al., 2010). The maturation phase then involved high levels of Nanog and Sall4. When Nanog and Sall4 reached their peak levels at day 9, many colonies were able to keep their initiated changes despite discontinuation of ectopic OSKM expression (Samavarchi-Tehrani et al., 2010). During the reprogramming event, many genes act as barriers to limit the reversal of somatic cells into pluripotency (Polo et al., 2012). Many of these genes seem to be active during the transition stage between the first and the second wave, and this is thus considered an important rate-limiting step of pluripotent reprogramming (Stadelmann, 2019; Warren, 2019).

1.5 Epigenetics and barriers of reprogramming

The purpose of these rate-limiting barriers is to prevent formation of tumors and to keep the cell proliferation in check.

The first reprogramming barrier is caused primarily by the p53 pathway via p21 mediated cell cycle arrest (Kawamura et al., 2009). The tumor-suppressing p53 pathway is activated by the introduction of reprogramming factors in the cell and inhibits the reprogramming of the cells by apoptosis, cell cycle arrest and senescence (Kawamura et al., 2009).

Temporarily downregulating the p53 pathway can increase reprogramming efficiency (Kawamura et al., 2009; Zhao et al., 2008). However, as the pathway also protects the cell from genome instability, completely turning off the pathway may result in declined genomic quality of the iPSCs (Hong et al., 2009; Zhao et al., 2008).

The second barrier in reprogramming is the inaccessibility of the core pluripotency genes, as they are packed tightly in inaccessible heterochromatin, caused by repressive DNA

and histone modifications, like H3K9me3 or H3K27me3 (Chen et al., 2013; Delgado-Olguín and Recillas-Targa, 2011). This chromatic structure needs to be remodeled for the reprogramming factors to reach their targets (Delgado-Olguín and Recillas-Targa, 2011). This can be achieved by removing repressive modifications or adding chromatin opening marks like acetyl groups, and by inhibiting histone deacetylases or DNA methyltransferases (Delgado-Olguín and Recillas-Targa, 2011; Huangfu et al., 2008).
In addition, the incomplete transgenic reprogramming often result in a cell-type specific memory of the somatic state, which is maintained through differentiation (Lister et al., 2011; Van Den Hurk et al., 2016).

1.6 Reprogramming process in other cell types

Despite the intense research going into reprogramming fibroblasts, the aspects of the reprogramming process cannot be considered universal. There is still a need to determine corresponding pathways and dynamics in other cell types. Nefzger et al., (2017) have characterized mouse neutrophils and keratinocytes in addition to fibroblasts. These experiments show that there are many cell-type specific differences in their pluripotent reprogramming, especially in the initiation stage involving cell identity loss (Nefzger et al., 2017).
Furthermore, neither neutrophils or keratinocytes seemed to go through the primitive streak-like developmental process which have been described in fibroblasts (Nefzger et al., 2017). In addition, despite MET being an important process in fibroblast reprogramming, keratinocytes do not require this, as they never pass through the EMT differentiation process (Nefzger et al., 2017). The forced expression of OSKM factors resulted in mostly cell-type specific changes (Nefzger et al., 2017).
However these effects also tend to mask a conserved transcription network common to all three cell types, with up to half of gene required for pluripotency being highly correlated (Nefzger et al., 2017). In addition, all cell types seemed to go through the two transcriptional waves described in literature, with the first one coinciding with cell identity loss and the second one with pluripotency network reactivation (Nefzger et al., 2017). Also Egr1 downregulation, central for reprogramming success, appears to be common to all three cell types (Nefzger et al., 2017; Worringer et al., 2014).

1.7 Vectors

There are many different methods of delivering the reprogramming factors into the cells, each with their own advantages and drawbacks (Hu, 2014). The methodology for iPSC generation can be divided into integrative, that include retroviruses and DNA transposons, and non-integrative, including RNA, Sendai viruses, proteins, and episomal plasmids (Hu, 2014).

1.7.1 Integrative vectors

One of the first vectors to be used in pluripotent reprogramming were retroviruses (Takahashi and Yamanaka, 2006). These retroviruses have most of their viral genes removed, leaving only the ones necessary for gene expression and delivery (Hu, 2014). They are most often derived from Moloney Mouse Leukemia Virus γ-retroviral vectors, or lentiviruses from HIV (Hu, 2014). The γ-retroviral vectors need a proliferating cell to infect, but the lentiviruses can infect non-dividing cells (Kitamura et al., 2003). However, retroviral vectors have the problem of uncontrolled genomic integration that might result in insertional mutagenesis (Fusaki et al., 2009).

Adeno-associated virus has also been considered a viable vector for reprogramming, as it is non-toxic and capable of long-term transduction also in non-dividing cells (Hu, 2014). However, there seems to be a tendency to for AAV to integrate only on cells going through reprogramming (J. Weltner et al., 2012). More commonly used Sendai viruses also offer a more persistent expression of transgenes due to its episomal replication, with no need for repeated transductions (Fusaki et al., 2009).

Integrative transposons can be used for reprogramming, consisting of a vector donor plasmid containing the transgenic cargo and terminal repeat sequences, and a helper plasmid with the transposase (Wang et al., 2008a). The most typical transposon vectors capable of transposing to cells for pluripotent reprogramming are the PiggyBac transposon, and the Sleeping Beauty transposon (Wang et al., 2008b). Integrative transposons provide a more stable expression of genes, have a larger cargo capacity, and can be used to control the temporal expression of transgenes and secondary reprogramming of differentiated cells (Woltjen et al., 2009). They also can be removed from the genome using transposase re-expression (Yusa et al., 2009).

1.7.2 Non-integrative vectors

Second way to induce pluripotent reprogramming is with episomal plasmid vectors (Weltner, 2018a). Because of the transient expression of non-replicating plasmid vectors, they needed serial plasmid transfections, and were in efficient (Jia et al., 2010). Later replicative episomal plasmid vectors containing oriP and EBNA1 sequences were introduced, which removed the need for serial transfections and gave a more persistent factor expression (Junying et al., 2009). The oriP and EBNA1 are the origin of replication and a protein sequence supporting the function of the oriP and the stability of the plasmid, respectively, originating from the Epstein-Barr virus (Yates et al., 2000). These sequences facilitate a plasmid replication once per cell cycle, but because of imperfect plasmid retention, they over time are removed from the genome, resulting in non-integrative reprogramming method (Hu, 2014; Okita et al., 2011).

Some other used viral vectors are adenovirus or Sendai virus -based episomal viruses (Hu, 2014). Adenoviruses have a large cargo capacity, but are typically inefficient, and not commonly in use anymore (Hu, 2014).

Reprogramming can be initiated with modified mRNA targeting the reprogramming factors (Warren et al., 2010). Transfection of mRNA causes typically an immune response, inhibiting growth, and for this reason the mRNA has to be modified by inhibiting the immune factors, particularly RIG-I, PKR or TRL7/8 (Seth et al., 2006). In addition, alphavirus based episomal RNA vectors can be used. These have the advantage of being non-integrative, but can be toxic to the cell (Weltner, 2018a).

Recombinant proteins have been used in reprogramming, and rely on a cell penetrating peptide to fuse with the cell and release the cargo (Hu, 2014). This is typically achieved with a protein transduction domain, naturally occurring in HIV and herpes simplex virus (Hu, 2014). These methods are currently somewhat inefficient (Weltner, 2018a).

1.8 CRISPR-Cas

CRISPR (clustered regularly interspaced short palindromic repeats) is a family of bacterial DNA sequences originally acting as a bacterial immune system against foreign DNA (Kurata et al., 2018). The CRISPR DNA, along with a CRISPR-associated protein (Cas) forms the system known as CRISPR-Cas, a method commonly used in genetic

engineering. The CRISPR-Cas -system consists of a leader sequence followed by short repeats separated by unique spacers (Louwen et al., 2014). The CRISPR-Cas system is activated from viral stress and saves the invading DNA particles into the spacers (Louwen et al., 2014). These spacer sequences are then transcribed into CRISPR-RNA (crRNA) (Kurata et al., 2018). The crRNA joins with a cleaving Cas-endonuclease and targets and degrades subsequent invading DNA sequences (Louwen et al., 2014).

This system is now commonly used in genetic engineering, with the most used CRISPR subtype relying on a protein Cas9 (Kurata et al., 2018). The Cas9 system consists of the Cas9 endonuclease, nuclease targeting CRISPR-RNAs (crRNAs), and a trans-activating CRISPR RNA (tracrRNA) (Kurata et al., 2018). In many systems however, the CRISPR RNAs are combined into a single-guide RNA (gRNA), which can be easily manipulated to target and cut wanted sequences (Kurata et al., 2018). Multiple gRNAs can also be concatenated into a single vector to allow for simultaneous targeting of multiple genes and is typically achieved by Golden Gate cloning (Weltner, 2018a).

1.9 CRISPR gene activation system

In addition to cutting the DNA, CRISPR-Cas can also be used in targeted gene activation (CRISPRa). CRISPRa utilizes catalytically inactivate Cas-protein (dCas9) (Gilbert et al., 2013). An additional transactivation domain is attached to this dCas9 protein, which functions by calling relevant transcription factors, such as the mediator complex and Polymerase II, to the gRNA-Cas complex binding site to initiate gene transcription (Balboa et al., 2015). Common trans-activation domains used for activation include NF-kappa B p65 subunit or multiple repeats of HSV1 VP16 peptides, such as in the dCas9VP192 -sequence developed by Balboa et al. (2015), which utilizes 12 VP16 repeats (Weltner, 2018a). The VP192-system can be fused with a doxycycline activated TetON-promoter for controlling its activity in the cell (Sokka J, 2019; Weltner, 2018a).

1.1 CRISPRa-reprogramming

By activating various reprogramming factors in the cell, CRISPRa can be used in pluripotent reprogramming. This was demonstrated by Liu et al. in 2018, by using a CRISPRa SunTag-activator to reprogram mouse fibroblasts into pluripotency via modulating Oct4 and Sox2 (Liu *et al*., 2018). Additionally, pluripotent reprogramming in

human cells was achieved by Weltner et al. via a dCas9-VP192 activator. This was achieved solely by targeting the core OSKM pluripotency factors, with the addition of Lin28a, but this method remained inefficient. Incorporating an EEA motif targeting to CRISPRa-reprogramming has since made the method nearly as effective as earlier established methods (Weltner et al., 2018b). The EEA motif is a highly conserved sequence in the human genome typically enriched in the promoter areas of genes involved in early embryonic development (Töhönen et al., 2015). The enhanced reprogramming is thought to be mediated in part by activation of NANOG and REX1 (Warren, 2019; Weltner et al., 2018b).

The CRISPRa -mediated reprogramming has the advantage of being able to induce the expression of reprogramming factors within the cell, rather than with forced exogenous factors, to effectively initiate the reprogramming process (Balboa et al., 2015). CRISPRa can also potentially help bypass the epigenetic barrier in reprogramming by opening areas in the genome. In addition, the guide RNAs for the CRISPRa reprogramming system are easy to design and allows the targeting and studying of complex reprogramming factor systems and their interactions in the cell. Alternative reprogramming factors can be screened for reprogramming potential and individually validated to optimize the pluripotent reprogramming outcome.

1.10 Guide screening

Screens were previously carried out by our lab to find candidate genes to enhance the reprogramming outcome. This was done by transfecting into cells a library of guide RNAs targeting about 40 transcription factors interacting with the pluripotency network. These cells were then carried through the reprogramming and the most efficient colonies were selected. The selected colonies were then carried through the selection process through 3 generations, with varying amounts of transfection material. Guides that promote more efficient reprogramming are expected to enrich in the reprogrammed iPSC colonies, and can be then screened for further evaluation. However, guides can enrich in the colonies for other reasons than enhancing the pluripotency, like by chance. To rule out passenger guides that are randomly included in the reprogramming colonies, hits from the screen must be validated. The thesis work will be focused on validating the hits from a previously performed CRISPRa screen for guide RNAs that can enhance reprogramming. The guide

RNAs will first be assessed for their efficiency in activating their target genes, followed by assessing their effect on pluripotent reprogramming efficiency.

1.11 Single cell RNA sequencing

Additionally, our lab carried out a single-cell RNA analysis to determine expression patterns of different genes during pluripotent reprogramming. Two emerging genes of interest were the genes *Epcam,* activating in the later stages of reprogramming, and *Znf486.*

Single cell RNA sequencing is a method for examining the gene expression at the single cell resolution. When analyzing cells in bulk, the varying transcriptomic levels even out. Single cell sequencing reveals heterogeneity in seemingly homogenous cell populations, with the possibility of identifying sub-populations and characterizing relations within cell states.

To achieve this, cells are suspended into a suspension which disrupts the cell adhesions. The cells are then separated with microdroplet separation, where each droplet contains the reagents necessary for the sequencing the included singular cell. The cell material is lysed, and RNA is reverse transcribed into cDNA, with unique short primers containing molecular identifier to track the individual mRNA molecules. The cDNA is amplified an additional time to produce cDNA libraries, which can be pooled and sequenced with next generation sequencing and analyzed with bioinformatic approaches. This method can then be utilized to investigate emergence of heterogenous iPSC subpopulations during reprogramming.

1.12 iPSC characterization

To successfully research pluripotent reprogramming, one needs to know if the cells in question can reach pluripotency, and whether they actually do so (Martí et al., 2013). This is also important for the objective comparison between different pluripotent cell lines (Martí et al., 2013). Some of the commonly used methods include immunocytochemistry and identification of three germ layers in embryoid bodies (Martí et al., 2013). Additionally alkaline phosphatase (AP) -staining can be used to validate the amount of pluripotent colonies generated by different samples (Singh et al., 2012).

Immunocytochemistry typically aims to detect common pluripotency markers, such as OCT4, NANOG, SOX2, TRA-1-60 or TRA-1-81, using secondary antibody binding (Nethercott et al., 2011). Immunocytochemistry allows for the detection and localization of these markers in cell populations (Nethercott et al., 2011).

Embryoid bodies are suspended aggregates of stem cells (Lin et al., 2020). In the absence of FGF2 these aggregates spontaneously differentiate into all three germ layers, and are used in order to both create differentiated cells for research, and to confirm the pluripotency of cells (Lin et al., 2020).

AP-staining takes advantage of the capability for the common enzyme, alkaline phosphatase, to change conformation of a colorimetric agent from soluble to precipitated state, allowing its visual detection (Martí et al., 2013). Alkaline phosphatase is present in many human cells types, and is particularly expressed at high levels in pluripotent stem cells (Singh et al., 2012). While it is not a definitive method for defining pluripotent cells, it can be used to differentiate iPSCs from non-AP containing differentiated cells (Singh et al., 2012).

1.13 Pluripotency factors

The original set of pluripotency factors developed by Takahashi and Yamanaka, combining Oct4, Sox2, Klf4 and Myc, are routinely used in pluripotent reprogramming. There have been many different combinations of pluripotency factors able to induce pluripotency, but Oct4 and Sox2 typically act as central factors in typical reprogramming factor sets (Weltner, 2018a).

1.13.1 OCT4

One of the most important pieces of the core pluripotency factors is the octamer-binding transcription factor 4 (Oct4) (Radzisheuskaya and Silva, 2014). Oct4 alone can induce pluripotency in mouse neural stem cells, but over expression of Oct4 causes mesodermal differentiation and its under-expression causes trophectodermal or endodermal differentiation (Kim et al., 2008). Oct4 is expressed in all pluripotent cells, and activates in the zygote during embryonic genome activation, facilitating proper embryonic development (M. Li and Belmonte, 2018). Oct4 acts by binding with Sox2, forming a heterodimer, and maintains pluripotency by promoting self-renewal (M. Li and Belmonte, 2018). Oct4 has also been demonstrated to facilitate mesenchymal to epithelial transition

(MET), and downregulate MET-counteracting factors (Radzisheuskaya and Silva, 2014). It is also shown to bind to heterochromatin containing regions and open them, activating genes and thus acting as an epigenetic modificator (Gaspar-Maia et al., 2011). It is not known to have a role in somatic cells, and its expression is silenced with cell lineage commitment (Jaenisch and Young, 2008). Due to its unique binding to Sox2, Oct4 has previously thought to be the only reprogramming factor which cannot be replaced by other members of its family during pluripotent reprogramming (Radzisheuskaya and Silva, 2014). Recently however, reprogramming has been achieved just with SOX2, KLF4 and C-MYC, while substituting OCT4 with endoderm lineage specifiers (Velychko et al., 2019).

1.13.2 SOX2

The sex-determining region Y-box 2 (Sox2), binding together with Oct4, is another core reprogramming factor in the pluripotency network (Boyer et al., 2005). Like Oct4, it is expressed in the early embryo and is required for its proper development by playing a part in establishing the inner blastocyst cell mass, as well regulating the development of the trophectodermal and extraembryonic endodermal cell lineages (M. Li and Belmonte, 2018; Sarkar and Hochedlinger, 2013). In addition however, Sox2 also is expressed in later embryonic development stages, and facilitates tissue regeneration and adult tissue homeostasis (Sarkar and Hochedlinger, 2013). Together with Oct4, Sox2 is thought to inhibit differentiation, and Sox2 is also involved in promoting MET (Radzisheuskaya and Silva, 2014). Sox2 can be replaced by other Sox-family members, such as Sox1 and Sox3 to in cellular reprogramming (Sarkar and Hochedlinger, 2013).

1.13.3 KLF4

Klf4, or Krüppel-like factor 4, is yet another important transcription factor which regulates cell proliferation, apoptosis, differentiation, embryonic development and tissue homeostasis maintenance (Ghaleb and Yang, 2017). It belongs to the KLF family of zinc finger transcription factors, which bind to DNA and control its expression (Ghaleb and Yang, 2017). Klf4 is essential for the proper development of mice, and knockdown of *Klf4* results in prenatal death (Bialkowska et al., 2017).

Klf4 can be either a transcriptional activator or repressor, and acts as an oncogene with both tumor suppressing and promoting qualities (Ghaleb and Yang, 2017). Klf4 is also thought to have a role in EMT, either as a suppressor or activator (Ghaleb and Yang,

2017), and mediates the interaction of Oct4 and Sox2 (Bialkowska et al., 2017; Radzisheuskaya and Silva, 2014). The removal of KLF4 during reprogramming can halt the process, which can then be continued by re-introducing KLF4 to the cell (Nishimura et al., 2014). KLF4 is furthermore not required for pluripotent reprogramming, but increases its effectiveness and speed (Jaenisch and Young, 2008). Klf4 can be replaced by Klf2 or Klf5, and simultaneous knockdown of all these genes in ESCs results in differentiation (Bourillot and Savatier, 2010).

1.13.4 c-Myc

The last core factor belonging to the Yamanaka factor cocktail is Myc, also known as c-Myc. *Myc* is an oncogene, well known for its role in proliferation of many cells (Hanahan and Weinberg, 2011). It joins with MAX to form a heterodimer and binds DNA sequences and amplifies gene expression on a wide range (Dang, 2012). While Myc does not itself bind to heterochromatin, it is observed to play a part in epigenetic modification by effecting the expression of other genes during tumorigenesis (Poli et al., 2018).

The *Myc* transcription network is responsible for influencing many cell growth pathways, and the *Myc* network is distinct from the typical core pluripotency circuit (Kim et al., 2008). Myc acts in these pathways either as an activator or repressor depending on the context (Kim et al., 2008). C-MYC increases both speed and efficiency of pluripotent reprogramming, but is not mandatory for iPSC generation and can be either omitted or replaced by other members of the family (Nakagawa et al., 2010). These includes also L-MYC and N-MYC. In some cases, L-MYC has proven even more effective in iPSC generation, and also provides the advantage of being less tumorigenic (Nakagawa et al., 2010).

1.13.5 LIN28A

LIN28A and LIN28B are RNA binding transcription factors expressed in pluripotent stem cells (M. Li and Belmonte, 2018). Specifically, LIN28A binds to let-7 pre-miRNAs blocking their maturation (M. Li and Belmonte, 2018), thus inhibiting their cell cycle promoting effects on MYC and Cyclin D1 (Lee et al., 2016; Ruiz et al., 2011). They are used in reprogramming, and enhances its effectiveness (Zhang et al., 2016). LIN28A and LIN28B are thought to be miRNA regulators with roles in development, cellular metabolism, cell cycle and oncogenesis (Zhang et al., 2016).

1.13.6 EpCAM

The epithelial cell-adhesion molecule (Epcam) is a transmembrane glycoprotein present in many epithelial tissues, particularly epithelial cancers (L. Huang et al., 2018). It mediates cell-to-cell adhesion as well as cell signaling, migration, proliferation and differentiation (L. Huang et al., 2018). Epcam has not been noted to be expressed in other than epithelial tissues, and its expression is elevated in many advanced carcinomas (Patriarca et al., 2012). As such, it can be used as a diagnostic marker. Epcam silencing in cancer has been observed to significantly reduce the proliferation and migration of epithelial cancer cells (Münz et al., 2004). The role of Epcam in cell proliferation can me partly mediated by its ability to rapidly upregulate the oncogene c-Myc (Münz et al., 2004). In addition, Epcam and its associated protein Cldn7 were noted to be upregulated in mouse embryonic fibroblast reprogramming, and that overexpression and inhibition of Epcam or Cldn7 resulted in respective enhancement or impairment of the reprogramming process (H. P. Huang et al., 2011). This could be caused due to its p53-p21 pathway repression, and upregulation of Oct4 (H. P. Huang et al., 2011). Additionally, EpCAM and EpEX has been noted to both increase OSKM reprogramming efficiency, but also replace multiple OSKM factors (Kuan et al., 2017). Using EpCAM and EpEX, either Klf4 or Oct4 alone were capable of reprogramming fibroblasts into iPSCs (Kuan et al., 2017).

1.13.7 ZNF486 and zinc-finger proteins

A zinc finger is a structural protein motif containing one or more stabilizing zinc ions (Han et al., 2020). Zinc fingers typically bind to DNA sequences, but also RNA, proteins and lipids (Cassandri et al., 2017). This attachment then guides cellular organization, development, adhesion, protein folding and, importantly, gene expression (Cassandri et al., 2017). Zinc finger proteins are developed into many distinct families with different tasks, with two particularly important families in this case being the Krüppel-like factor (KLF) family, and Krüppel-associated box (KRAB) containing zinc finger proteins (KZFPs). The KZFP family of zinc finger proteins are known to control transposable elements in early embryonic development of humans (Pontis et al., 2019). These transposable elements can then affect gene expression via KLF stimulation (Pontis et al., 2019). ZNF486 is one of these zinc finger proteins (Pontis et al., 2019), and while there is not much data of ZNF486 functions, it is thought to be activated during pluripotency reversal and thus potential gene for affecting pluripotent reprogramming, based on earlier screens by our lab.

1.13.8 Krüppel-like factors 2, 5 and 17

Like KLF4, transcription factors KLF2, KLF5 and KLF17 belong to the KLF family of zinc finger proteins (Pollak et al., 2018). Particularly KLF2 and KLF5 hold similar functions as they can be used to replace KLF4 in reprogramming (Bourillot and Savatier, 2010). KLF2 is thought to be involved in many processes in human, including lung development, epithelial integrity, T-cell differentiation and adipogenesis (Jha and Das, 2017). It is also implicated to mediate the development of embryonic blood cells and blood vessels (Jha and Das, 2017). Likewise, KLF5 is involved in cell proliferation and differentiation, and thought to regulate skeletal muscle regeneration, myogenic differentiation and mitochondria function among many others (Pollak et al., 2018). KLF17 is present in primate inner cell mass alongside NANOG (Blakeley et al., 2015). Primed cells lose KLF17 expression (Guo et al., 2016). It acts in the epithelial-mesenchymal transtition and has an anti-metastatic function in breast cancer (Gumireddy et al., 2010). KLF17 is known to interact with p53, where p53 is thought to promote KLF17 functions, while KLF17 enhances p53 transcription in a positive feedback-loop (Zhou et al., 2016).

2 Aims of the Thesis

The main aim of this work is to derive a new more efficient way of producing human iPSC using CRISPRa. This is done by validating the genes identified from our CRISPRa reprogramming guide screen (*KLF2, KLF5* and *KLF17*), and single cell RNA sequencing (*EpCAM* and *ZNF486*). The specific aims of the work are as follows:

1. Evaluation of selected guide RNAs in activation of their respective target endogenous genes in HEK293 cells.

For this aim, target genes will be selected by single cell RNA sequencing, and the enrichment of guide RNAs targeting the specific gene resulting from guide screens. Guide RNAs selected by single cell RNA sequencing targeting *EpCAM* and *ZNF486* will be tested guide by guide for endogenous gene activation efficiency. The guide RNAs selected from screens targeting *KLF2, KLF5* and *KLF17* will be tested guide by guide for gene activation efficiency and the results compared to the level of enrichment per guide in the reprogramming guide screens.

2. Functional testing of selected gRNAs

For this aim, guides selected in the aim 1. will be tested for their effect in improving reprogramming efficiency. This will be done using guides targeting endogenous OSKML (*OCT4, SOX2, KLF4, MYC* and *LIN28*) as basal reprogramming factors in addition to guides targeting genes *EpCAM, ZNF486, KLF2, KLF5* and *KLF17* each. Another set of guides targeting the same genes will also be selected based on their enrichment in the reprogramming screen. These two sets will be compared for their effect on reprogramming efficiency and for the kinetics of the reprogramming process, measured by total number of alkaline positive iPSC colonies.

3 Materials and Methods

3.1 Primers and guide assembly

3.1.1 RT-PCR

RNA was extracted by first lysing the cells in 350 µl LBP lysis buffer. The lysate was then purified using NucleoSpin RNA plus purification kit (Macherey-Nagel), according to manufacturer's instructions. RNA concentrations were measured with Simplinano spectrophotometer. One microgram of RNA was denatured in 65 °C for 1 min and synthesized into cDNA by reverse transcription. RT master mix was firstly created (dNTP 0.3 mM, Oligo dT primer 5 µM, Random primers 0.1 µg, Ribolock RNAse inhibitor 20U, M-MLV reverse transcriptase 100U in M-MLV Reverse transcriptase buffer, total volume 8.7 µl). Molecular biology grade water (Lonza) was used in all PCR reactions. To each 1 µg of denatured sample RNA, water was added to a total of 11.3 µl. 8.7 µl of RT master mix was added on samples for a total volume of 20 µl, which were then incubated at 37 °C for 90 min. Enzyme was denatured at 90 °C for 5 min.

3.1.2 qPCR

Gene expression was quantified in 20 µl qPCR reactions. A master mix containing 4 µl 5x HOT FIREPol EvaGreen qPCR Mix Plus (ROX) (Solis Biodyne) reagent and 10 µl water per reaction. 1 µl of sample DNA was added per reaction. Qiagen QIAgility pipetting robot was used to pipet 5 µl 2 mM forward and reverse primers (Table 1.) diluted in water to 15 µl of sample mix, making two technical replicates for each reaction. The reaction cycle was run in Qiagen Rotor-Gene Q qPCR machine for at 95 °C for 15 min, after which temperatures 95 °C for 25 s, 61 °C for 25 s, and 72 °C for 25 s was cycled for 35 cycles.

Table 1. qPCR primers for ZNF486 and EpCAM

qPCR primers	Forward	Reverse	Annealing T (C)	Length (bp)
ZNF486	5'GCTTTTTAAGGAAGTTCTCAACCC	5'GAACTTGTTACAGGCTTGCCA	59	95
EpCAM	5'GCGAGTGAGAACCTACTGGA	5'AACGCGTTGTGATCTCCTTCT	59	110

After qPCR, the reactions' melting curve was analyzed by increasing the temperature from 75 °C by 1 °C every 5 seconds. The results were analyzed using the ddCt method.

3.1.3 gRNA design

gRNAs were designed using Benchling sequence design tools (Benchling [Biology Software]. (2019). Retrieved from https://benchling.com). The upstream region of the genes between -50 and -500 bp from the transcription start site was targeted, and 5 gRNAs for each gene were made (Table 2.). gRNAs expression cassettes were then assembled using PCR (dNTP 0.2 mM, 1 aggc Forward Primer 0.5 µM, 1 aggc Reverse Primer 0.5 µM, Phusion polymerase 0.02U, U6 tailed promoter 5 ng, tracR tailed template 5 ng and gRNA oligos 0.02 µM in HF buffer according to manufacturer's instructions, total volume 100 µl).

Table 2. gRNA oligo sequences for ZNF486 and EpCAM

gRNA oligo sequences	**ZNF486**	**EpCAM**
Guide 1	GTGGAAAGGACGAAACACCG ATTATCCAATCAGAGACGCT GTTTTAGAGCTAGAAATAG	GTGGAAAGGACGAAACACCG TCACTCCCCCAACTCCCGGG GTTTTAGAGCTAGAAATAG
Guide 2	GTGGAAAGGACGAAACACCG ACTCAGTCCCTGAGTGACAG GTTTTAGAGCTAGAAATAG	GTGGAAAGGACGAAACACCG CAGGTGCTTCGTCTCCATGG GTTTTAGAGCTAGAAATAG
Guide 3	GTGGAAAGGACGAAACACCG CCAGGCCCACCCTAGAGCAA GTTTTAGAGCTAGAAATAG	GTGGAAAGGACGAAACACCG GGGAACTGGATAGAGGAACG GTTTTAGAGCTAGAAATAG
Guide 4	GTGGAAAGGACGAAACACCG TAGAAAATCAACTTTTACTT GTTTTAGAGCTAGAAATAG	GTGGAAAGGACGAAACACCG ATTGCCAGGTAAAAGCTCAA GTTTTAGAGCTAGAAATAG
Guide 5	GTGGAAAGGACGAAACACCG ATATATGTGTAAATCTCGAG GTTTTAGAGCTAGAAATAG	GTGGAAAGGACGAAACACCG AGGTGTTCCCAGGCTGAGCG GTTTTAGAGCTAGAAATAG

The reaction was run on 98 °C for 3 min after which temperatures 98 °C for 10 sec, 52 °C for 30 sec and 72 °C for 12 sec were cycled 35 times. Reaction was ended with 72 °C for 8 min. Successful assembly of cassettes was confirmed with electrophoresis in 1.5 % agarose gel in TBE. Products were purified using NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel). (Balboa et al., 2015)

3.1.4 HEK293 Transfection for guide validation

To test the gene activation levels of guide RNAs, HEK293 cells were seeded on 24-well plates on MEF media 100 000 cells per well. The next day the cells were transfected with 400 ng of CRISPR-Cas activator plasmid DNA, 200 ng of constructed guide RNA, 10 µl

water and 3.6 µl Fugene -transfection reagent for a total of 25 µl reaction. The cell RNA was then purified using NucleoSpin RNA purification kit (Macherey-Nagel).

3.2 Plasmid preparation

3.2.1 Golden gate assembly for EpCAM and ZNF486

To prepare for the Golden gate-ligation, guides were assembled using linker primers (dNTP 0.2 mM, Phusion polymerase 0.02U, U6 tailed promoter 5 ng, tracR tailed template 5 ng and gRNA oligos 0.02 µM in HF buffer according to manufacturer's instructions, total volume 100 µl. 5 ng of each linker primer was used, depicted in Tables 3 and 4.).

Table 3. Guide cassette assembly PCR primers for EpCAM and ZNF486

EpCAM plasmid	**Fw. Primer**	**Rev. Primer**
EpCAM guide 1	1 aggc fw.	1 aggc rev
EpCAM guide 2	2 aggc fw.	5 aggc rev.
ZNF486 plasmid	**Fw. Primer**	**Rev. Primer**
ZNF486 guide 2	1 aggc fw.	3 aggc rev
ZNF486 guide 3	4 aggc fw.	5 aggc rev.
Combination plasmid	**Fw. Primer**	**Rev. Primer**
EpCAM guide 1	1 aggc fw.	1 aggc rev
EpCAM guide 2	2 aggc fw.	2 aggc rev.
ZNF486 guide 2	3 aggc fw.	3 aggc rev
ZNF486 guide 3	4 aggc fw.	5 aggc rev.

Table 4. Guide cassette assembly PCR primers for KLF2, KLF5 and KLF17.

QPCR guide	**Fw. Primer**	**Rev. Primer**	**Screen guide**	**Fw. Primer**	**Rev. Primer**
KLF2 guide 1	1 aggc fw.	1 aggc rev.	KLF2 guide 1	1 aggc fw.	1 aggc rev.
KLF2 guide 2	2 aggc fw.	2 aggc rev.	KLF2 guide 5	2 aggc fw.	2 aggc rev.
KLF5 guide 1	3 aggc fw.	3 aggc rev.	KLF5 guide 1	3 aggc fw.	3 aggc rev.
KLF5 guide 5	4 aggc fw.	4 aggc rev.	KLF5 guide 5	4 aggc fw.	4 aggc rev.
KLF17 guide 3	5 aggc fw.	5 aggc continuation rev.	KLF17 guide 3	5 aggc fw.	5 aggc continuation rev.
KLF17 guide 6	6 aggc fw.	5 aggc rev.	KLF17 guide 1	6 aggc fw.	5 aggc rev.

Golden gate-ligation was then used to assemble five plasmids. Three of these plasmids were for EpCAM and ZNF486: two guide sequences were selected to a plasmid for ZNF486 activation, two guides for EpCAM activation plasmid and additionally all four sequences were inserted to a singular EpCAM & ZNF486 -plasmid. Two clones were made for each plasmid with *E. coli* transfection.

Additionally, Golden gate -ligation was used to assemble two more plasmids for KLF2, KLF5 and KLF17. First plasmid contained guides selected by RT-qPCR results and the second contained guides selected by earlier screens. The cassettes were assembled using a PCR (T4 DNA ligase 5 U, Esp3l Restriction enzyme 10 U, DTT 10 mM, GG-dest plasmid 150 ng and guide cassettes 50 ng each in T4 ligase buffer according to manufacturer's instructions, total volume 20 µl). Reaction was run at 37 °C for 2 min, 16 °C for 5 min and 80 °C for 20 min, 50 cycles.

3.2.2 Plasmid cloning using e.coli

10 µl of GG-dest plasmid DNA was transfected to 66 µl of competent DH5-α *E. coli* cells (New England Biolabs, NEB5) by incubating the cells on ice for 30 min, and heat shocking them using a heat block at 42 °C for 1 min, returning the cells on ice for 1 min. The solution was spread on pre-made ampicillin-LB coated petri dishes, and left to incubate in 37 °C for 16 hours overnight.

The next day, two colonies were picked from each plate using a pipette tip, and submerged to 5 ml of LB media supplemented with 5 µl of ampicillin.

3.2.3 Plasmid restriction and ligation

In order to transfect the guides into fibroblast cells, the guide insert was isolated from the GG-dest backbone and ligated into a GG-EBNA backbone (GG-EBNA plasmid solution 1 µl, insert sequence solution 1 µl, T4 DNA ligase 5U in T4 DNA ligase buffer according to manufacturer's instructions, total volume 10 µl). To confirm band sizes and to ensure compatible ends for joining inserts and plasmid backbones, the plasmid DNA was digested with restriction enzymes NotI and EcoR1 (Plasmid DNA 800 ng, NotI 5U and EcoRl 5U in FastDigest Green buffer according to manufacturer's instructions, total volume 10 µl). For GG-dest and GG-EBNA digestion, 800 and 500 ng of plasmid was used, respectively. The reaction was incubated in 37 °C for 30 min. For band size confirmation, the digested DNA was run on 1 % TAE gel, 100 V.

3.2.4 Plasmid isolation using miniprep and midiprep

The plasmids were isolated using GeneJET Plasmid Miniprep Kit (Thermo Scientific, K0503), and stored in -20 C, and later using NucleoBond Xtra Midi EF (Macherey Nagel, 740420.50). The DNA concentrations were measured using SimpliNano spectrophotometer (Table 5.).

Table 5. Plasmid concentrations

for EpCAM and ZNF486	**C (ng/ul)**	**Cassettes for KLF2, KLF5 and KLF17**	**C (ng/ul)**
EpCAM clone 1	153.7	qPCR clone 1	206.0
EpCAM clone 2	204.7	Screen clone 1	150.9
ZNF486 clone 1	259.0	Screen clone 2	170.0
ZNF486 clone 2	167.8	Screen clone 3	185.4
Combination clone 1	210.8	Screen clone 4	174.1
Combination clone 2	209.0	Screen clone 5	187.5

3.2.5 Glycerol stocks

The prepared *E. coli* cultures containing transformed plasmids were preserved as 2 ml Eppendorf's stocks, with 1:4 glycerol (400 ul) and 3:4 bacterial culture in LB (1200 ul). The stock tubes were stored in -80 C.

3.2.6 Sanger sequencing

5 µl of 80-100 ng/ul template plasmid DNA was mixed with 5 µl of 100 µM primer solution. Different sample was prepared for each primer. Samples were sent to Eurofins Genomics for sequencing (LightRun Tube Service). To sequence EpCAM and ZNF486, T7 forward and SP6 reverse primers were used. For KLF2, KLF5 and KLF17, primers were made from the sequences of KLF5 g5 in forward, KLF5 g5 in reverse, KLF17 g3 in reverse, T7 forward primer and a Bbsl reverse sequence to substitute for a possibly faulty SP6 primer (Table 6.).

Table 6. Sanger sequencing primers

Primer	Sequence
KLF5 g5 forward	GCCCGACCGCACCCCTCTC
KLF5 g5 reverse	GAGAGGGGTGCGGTCGGGC
KLF 17 g3 reverse	ATTCCACCCCTCAGTCTGA
Bbsl reverse	ATGGCTGATTATGATCTAGAGTC
T7 forward	TAATACGACTCACTATAGGG
SP6 reverse	ATTTAGGTGACACTATAG

3.3 Cell culturing

3.3.1 Cell cultures

Human foreskin fibroblasts (HFFs) and human embryonic kidney 293 -cells (HEK293s) were grown in Mouse embryonic fibroblast (MEF) -media. The cells were cultured on BioLite cell culture dishes (Thermo Scientific™, 130182) or Costar 6-, 12-, 24-well cell culture plates (Corning, 3335-7). Cells were passaged when confluency reached 80 %. For this, the cells were submerged in TrypLE™ (Gibco™, 12604021) for 3 min at 37 C, and the plated were gently tapped to dislodge the cells. Cell media 2 times the volume of TrypLE™ was added to neutralise the trypsinisation, and the suspension was centrifuged for 4 min at 200 g. The supernatant was removed, and cells were resuspended in MEF-medium on a cell culture dish at a ratio of 1:20 to 1:40.

LCL cells were cultured in 150 cm2 cell culture flasks (Corning, CLS430825-50EA), using LCL medium (RPMI 1640 Medium (Gibco™, 11875093) supplemented with GlutaMAX™, 15% FBS and 1 µM Sodium Pyruvate (Lonza, BE13-115E). The cells were passaged by first centrifuging them in a falcon tube at 200 g for 5 min, and then resuspended in 1ml PBS. The cell density was counted using Countess Automated Cell Counter (Invitrogen™), with 10 µl of cell suspension and 10 µl of 0.4% Trypan Blue Stain (Invitrogen™, T10282) on the counting slide. A cell density of 0.5 – 1.5 x 106 cells/ml was added on flasks for optimal cell growth.

3.3.2 Cryopreservation and thawing of cells

Cells were thawed by hand and centrifuged at 200 g for 5 min. The cells were then washed by resuspension to PBS and the centrifugation was repeated. The pellet was then suspended to the used cell media and incubated on culture dishes in 37 C.

To freeze the cells, they were first detached submerging them in TrypLE™ for 3 min at 37 °C. The plate was tapped lightly to dislodge the cells and centrifuged in a falcon tube for 4 min at 200 g. The pellet was suspended in 0.5 ml cell medium and 0.5 ml freezing medium (80 % fetal bovine serum (FBS) and 20 % dimethyl sulfoxide, (DMSO, Sigma, D8418-100ML), for a total DMSO concentration of 10 %). The cell suspension was then frozen at -80 °C in cryogenic vial for one week and transferred to -150 °C for long term preservation.

3.4 Transfection and reprogramming

3.4.1 LCL transfection and reprogramming

LCL cells were transfected with electroporation using Invitrogen Neon -transfection system with 1650 V, 10 ms with 3 x pulse, and seeded on matrigel on 6-well plates in 2 ml LCL media per well with 0.25 mM Sodium butyrate (NaB) (Table 7.). On day 3, the transfected cells were passaged to 2 ml new LCL media with 0.25 mM NaB, by pelleting cells with 200 g centrifugation for 5 min, and counting 500 000 cells for every 1 ml of media. For the counting, Countess automatic cell counter was used with 10 µl of cell suspension and 10 µl trypan blue for each counting plate. From day 6 onwards, cell media was changed to every 2 days with the same method, except using hES (human Embryonic Stem cell)-media with 0.25 mM NaB. From day 10, all non-adherent cells were removed from the wells, and 2 ml E8 media without NaB was added for each well. The E8 media was then changed every 2 days until AP-staining.

Table 7. Conditions for LCL transfections

EEA Plasmid	C (µg)	-EEA Plasmid	C (µg)
GG-EBNA-EPCAM	1.5	GG-EBNA-EPCAM	1.5
GG-EBNA-O3S2K2M2L1-PP	1.125	GG-EBNA-O3S2K2M2L1-PP	1.5
GG-EBNA-mir302/367-7g	1.125	GG-EBNA-mir302/367-7g	1.5
PCXLE- dCas9-VP192-T2A-EGFP-shP53	1.125	PCXLE- dCas9-VP192-T2A-EGFP-shP53	1.5
GG-EBNA-EEA-5g-PGK-Puro	1.125		

3.4.2 HFF transfection and reprogramming

HFF cells were transfected with electroporation using Invitrogen Neon -transfection system with 1650 V, 10 ms with 3 x pulse, and seeded on matrigel on 10 ml petri dishes 10 ml MEF-media per well with 0.25 mM Sodium butyrate (NaB) (Table 8.). Cell media was changed every other day. From day 4 onwards, the cell media was changed to 1:1 ratio of MEF to hES media with 0.25 mM NaB. On day 16, E8 flex-media was changed to the cells. Cells were then stained with the Alkaline Phosphatase staining method on day 19.

Table 8. Conditions for HFF transfections

EpCAM plasmids	C (µg)	EEA plasmids	C (µg)	TdT Plasmids	C (µg)
GG-EBNA-EPCAM	1.5	GG-EBNA-EEA-5g-PGK-Puro	1.5	GG-EBNA-TdT-g1	1.5
GG-EBNA-O3S2K2M2L1-PP	1.125	GG-EBNA-O3S2K2M2L1-PP	1.5	GG-EBNA-O3S2K2M2L1-PP	1.5
GG-EBNA-mir302/367-7g	1.125	GG-EBNA-mir302/367-7g	1.5	GG-EBNA-mir302/367-7g	1.5
PCXLE- dCas9-VP192-T2A-EGFP-shP53	1.125	PCXLE- dCas9-VP192-T2A-EGFP-shP53	1.5	PCXLE- dCas9-VP192-T2A-EGFP-shP53	1.5

3.5 iPSC characterization

3.5.1 Alkaline Phosphatase Staining

The reprogrammed LCL-colonies at day 21 were washed with PBS once and fixed in 4 % paraformaldehyde (Fisher Chemical, 10131580) for 15 min RT in 1x PBS. The cells were then washed with PBS twice times and strained with NBT/BCIP solution (Tris-HCl pH 9.5 100mM, NaCl 100mM, $MgCl_2$ 50mM, NBT/BCIP stock (Sigma-Aldrich) 20 µl in water) for about 10 minutes each or when colonies were clearly visible. The solution was then replaced with PBS and the wells were scanned and counted.

3.5.2 Immunocytochemistry

Visualization of expressed genes was done with immunocytochemical staining. The cells were fixed in 4 % paraformaldehyde for 15 min, permeabilized with 0.5 % triton-X in PBS for 10 min and blocked with Ultra V Block for 10 min. Cells were then incubated with primary antibody o/n in +6 °C (Table 9.).

Table 9. Primary antibodies for immunocytochemical staining:

Primary Antibody	Manufacturer	Host Species	Dilution	Catalog
NANOG	Cell signaling technologies	Rabbit	1:250	4903S
OCT4	Santa Cruz Biotechnology	Rabbit	1:500	SC-9081
TRA-1-60	Invitrogen	Mouse	1:200	MA1-023
TRA-1-81	Invitrogen	Mouse	1:200	MA1-024

After the primary antibodies were aspirated, the cells were washed with PBS. Cells were then incubated with the secondary antibody in 0.1 % tween in 1:500 ratio for 30 min RT (Table 10.). After this, secondary antibodies were aspirated, and the stained cells were immersed in DAPI and imaged with fluorescence microscope.

Table 10. Secondary antibodies for immunocytochemical staining:

Secondary Antibody/DNA dye	Manufacturer	Host Species	Dilution	Catalog
Hoechst	Thermo Fisher scientific	DNA dye	1:1000	62249
Anti-Rabbit IgG (H+L) Alexa Fluor 594	Invitrogen	Donkey	1:500	A21207
Anti-Mouse IgG (H+L) Alexa Fluor 594	Invitrogen	Donkey	1:500	A21203
Anti-Mouse IgG (H+L) Alexa Fluor 488	Invitrogen	Donkey	1:500	A21202
Anti-Rabbit IgG (H+L) Alexa Fluor 488	Invitrogen	Donkey	1:500	A21206

3.5.3 Embryoid body formation

The iPSC colonies were detached in EDTA for 3 min, and transferred to non-adherent dishes (Corning, 3471) in hES-medium (KnockOut DMEM; Gibco), supplemented with 20 % KnockOut serum replacement (Gibco), 1 % GlutaMAX supplement (Gibco), 0.1 mM beta-mercaptoethanol (Gibco) and 1 % non-essential amino acids (Gibco), without

bFGF. Cell media was changed every other day. For this, cells were carefully transferred to 10 ml falcon tubes, and pelleted on 200 rpm for 4 min. After aspirating media, cells were suspended in 1 ml hES without bFGF and transferred back to non-adherent dishes in 4 ml media. On day 11, some embryoid bodies were plated on gelatin-coated 24-well dishes. On day 4 from plating, embryoid bodies were fixed in 4 % PFA for 30 min RT, submerged in PBS, and moved to +6 °C. On day 10, fixed embryoid bodies were incubated with primary antibodies in +6 °C o/n (Table 11.).

Table 11. Primary antibodies for immunocytochemical staining of embryoid bodies:

Primary antibody	**Manufacturer**	**Host species**	**Dilution**	**Catalog**
Sox17	R&D Systems	Goat	1:500	AF1924
α-smooth muscle actin	Sigma	Mouse	1:500	A2547
β-tubulin III	Abcam	Rabbit	1:500	Ab18207

Afterwards, cells were gently washed with PBS, incubated with secondary antibodies (Table 12.), and imaged for SOX17, α-SMA (α- smooth muscle actin) and TUBB3 (β-tubulin III).

Table 12. Secondary antibodies for immunocytochemical staining of embryoid bodies:

Secondary Antibody/DNA dye	**Manufacturer**	**Host Species**	**Dilution**	**Catalog**
Hoechst	Thermo Fisher scientific	DNA dye	1:1000	62249
Anti-Mouse IgG (H+L) Alexa Fluor 488	Invitrogen	Donkey	1:500	A21202
Anti-Rabbit IgG (H+L) Alexa Fluor 488	Invitrogen	Donkey	1:500	A21206
Anti-Goat IgG (H+L) Alexa Fluor 488	Invitrogen	Donkey	1:500	A11055

3.6 Statistical analysis

To analyse RT-qPCR data, comparative delta Ct method was used. This method normalises the Ct-values to a chosen housekeeping gene, in this case Cyclophilin G (ΔCtT). Additionally, it normalized them to a calibrator standard (Golden control VI, ΔCtC). The housekeeping gene is subtracted from the resulting activation values. Then the normalized golden control value is subtracted from normalized activation value.
In the case of EpCAM & ZNF486, the guides were also normalized to NT control to give fold change. In the case of KLF2, KLF5 and KLF17, the guides were normalized to their respective pools in addition to Cyclophilin G and Golden control.
In inductions, AP-positive colonies were counted by hand. Clear, moderately sized colonies were counted as positive, while ambiguous dots were left out (Fig 14.).
Error bars in both experiments represent the standard error mean (SEM). Single factor ANOVA was used to evaluate the significance level.

3.7 Cell Imaging

Cells were monitored with Tissue Culture microscope (Leica Miscrosystems, DM IL LED). Images were taken with an attached Leica EC3 digital camera with Leica LAS EZ software. Fluorescent images were taken with Evos FL cell imaging system (Life Technologies). Colour detection was done via installed DAPI, GFP and Texas red light cubes. Images were processed with Adobe photoshop to optimize brightness and contrast and cropping.

4 Results

The CRISPRa reprogramming method can be optimized by screening for guide RNAs that improve the reprogramming outcome. Guides that promote more efficient reprogramming are expected to enrich in the reprogrammed iPSC colonies. However, to rule out passenger guides that are randomly included in the reprogramming colonies, hits from the screen must be validated. The thesis work will be focused on validating the hits from a previously performed CRISPRa screen for guide RNAs that can enhance reprogramming. The guide RNAs will first be assessed for their efficiency in activating their target genes, followed by assessing their effect on pluripotent reprogramming efficiency.

4.1 HFF-cell line characterization

HFF cell lines induced with OSKML and CRISPR-activator were characterized for their reprogramming capability by the expression of OCT4 and NANOG using immunocytochemical staining. This was done to create a cell line known to be capable of pluripotency reversal using CRISPRa and endogenous target expression. On day 14, all cell lines showed positive for pluripotency markers OCT4 and NANOG, and cell surface markers TRA-1-60 and TRA-1-81 (Fig 1. a and b) The cells were also developed into suspended embryoid bodies and allowed to differentiate. The cells were then stained for markers SOX17, α-smooth muscle actin and β-tubulin III using same immunocytochemical staining methods, and tested positive for all three (Fig 1. c).

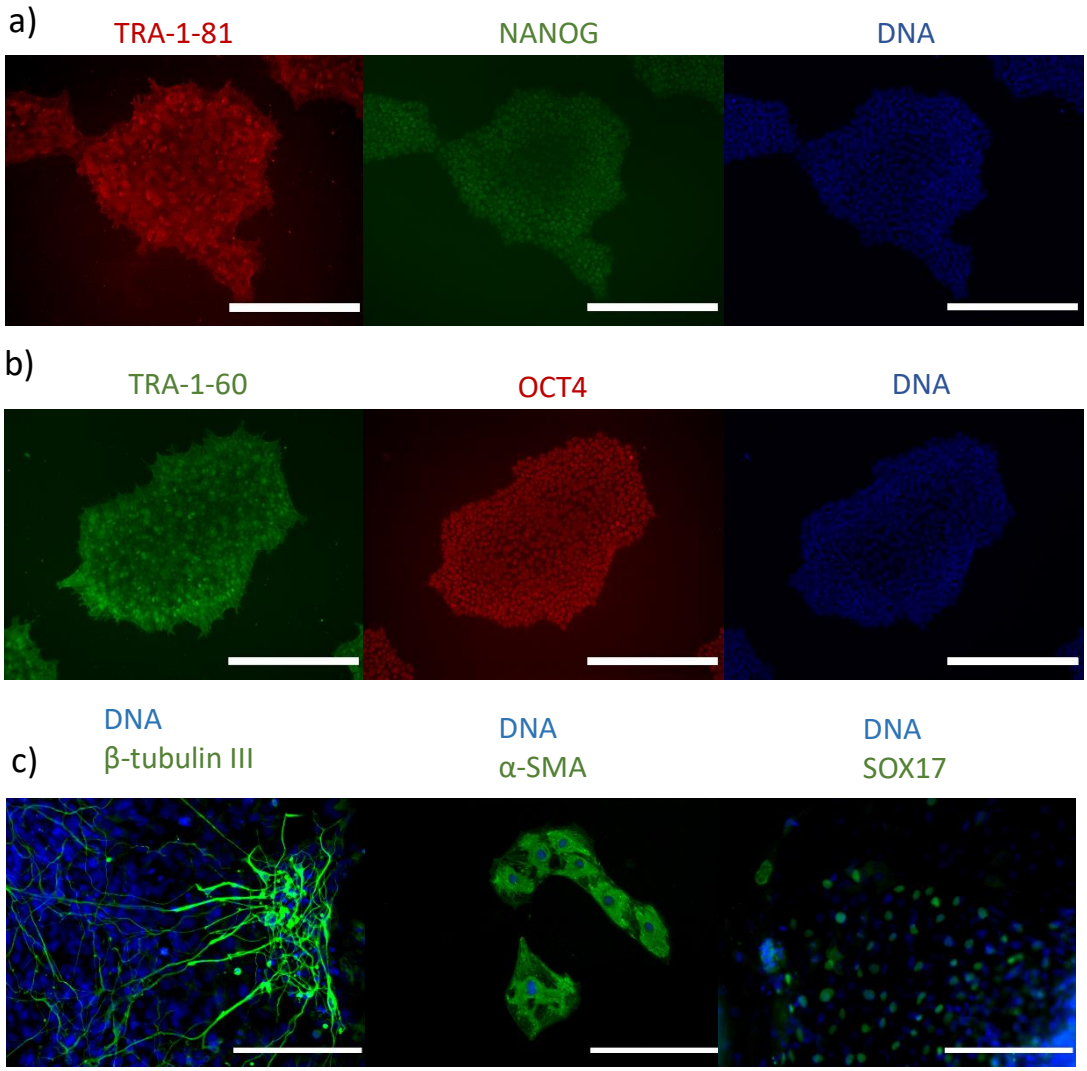


Figure 1. HFF cell line characterization. Expression of pluripotency markers in reprogrammed HFF iPSC colonies. Co-staining of NANOG, TRA-1-81 and DNA (a), as well as OCT4, TRA-1-60 and DNA (b). Picture c) depicts expression of germ layer-speicific markers in embryoid bodies. Ectodermal (β-tubulin III), mesodermal (α-smooth muscle actin) and endodermal (SOX17). Scale bars = 400 µM

4.2 Guide assembly and HEK293 transfection

Five plasmids targeting *EpCAM* and 5 plasmids targeting *ZNF486* were designed in Benchling. As for *KLF2, KLF5* and *KLF17*, guide oligos were already made in previous experiments and were extracted and donated by a colleague for use. Five guides were used for *KLF2* and six guides for *KLF5* and *KLF17* each.

The plasmids were first assembled into cassettes using PCR. Successful guide assembly was confirmed with electrophoresis (Fig 2.). Assembled guides were then transfected into HEK293 cells, alongside a dCas9 activator-encoding plasmid. A non-transfected control (NT) acted as a reference, to which a dCas9 plasmid was transfected without any guides.

A sample was made for each guide plasmid, and pool samples were made for each gene, containing all five guides for the given gene. These pools were used as a common parameter to normalize the five genes to. Three biological replicates were used. Three days later the cells were lysed, and the resulting lysate RNA was reverse transcribed into cDNA using reverse transcriptase reaction.

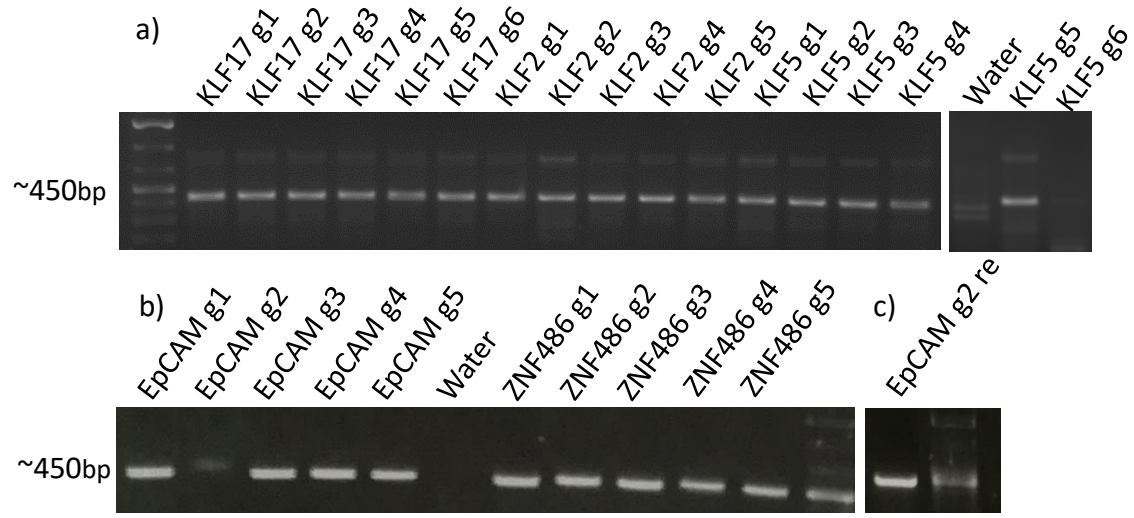


Figure 2. Guide assembly test. Electrophoresis was used to confirm successful guide assembly for KLF2, KLF5, KLF17 (a) and for ZNF486 and EpCAM (b). In c), EpCAM guide 2 was reassembled as it did not show correctly in first PCR.

After HEK293 transfections and RNA extraction, the gene activation levels of *ZNF486* and *EpCAM* for single guides and pools was tested with PCR (Fig 3. a and b). This was to see whether noticeable gene activation could be observed and see if performing RT-qPCR can be considered viable. For KLF2, KLF5 and KLF17, expression levels of pools were tested using the same method (Fig 3. c). Sometimes residual DNA left over from the extraction can interfere with qPCR results. To see if this residual DNA is present, cDNA conversion was repeated using a control sample without reverse transcriptase enzyme (Fig 3. d) No significant amounts of DNA was detected.

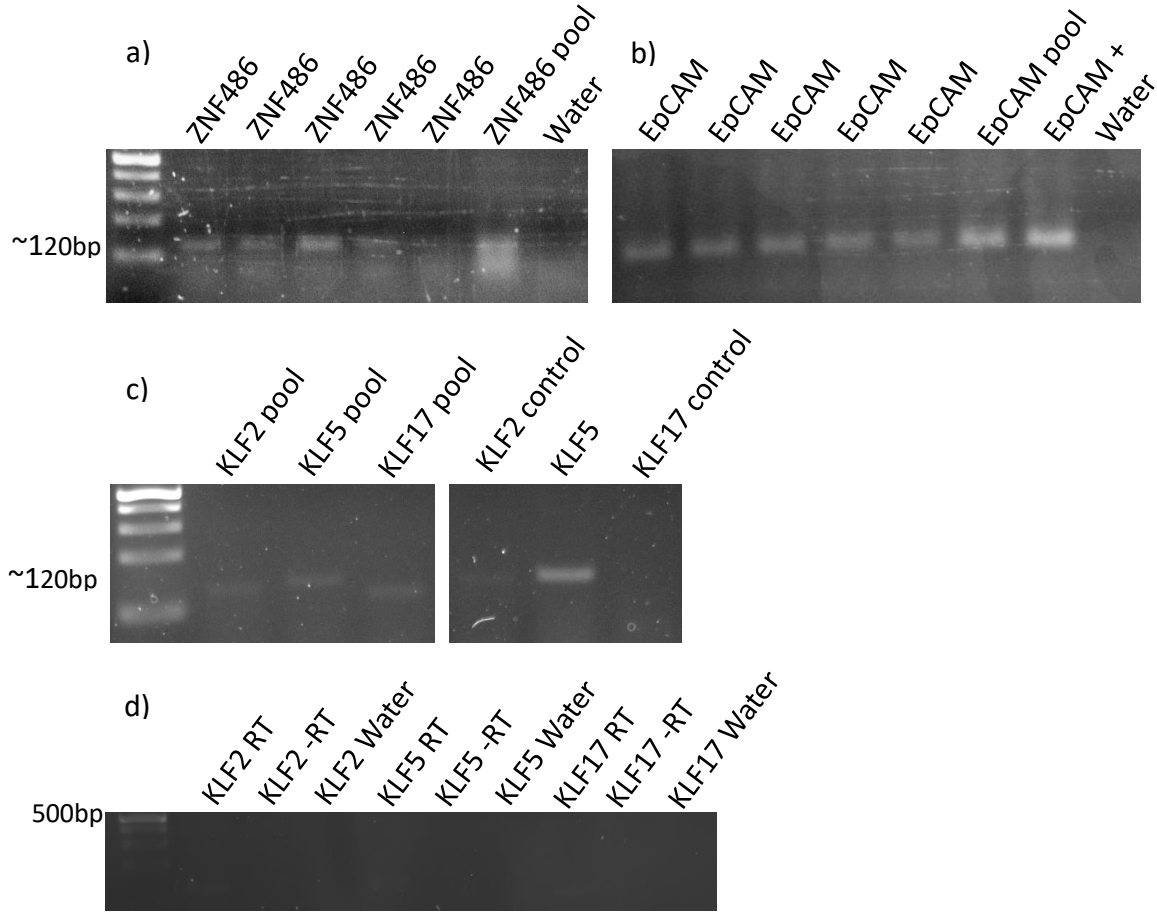


Figure 3. Single guide activation PCR, pool activation PCR and DNA interference PCR. PCR was used to confirm single guide gene expression and pool gene expression to see if the expression levels are sufficient for RT-qPCR. PCR for ZNF486 (a) and EpCAM (b). PCR for pool expression levels for KLF2, KLF5, KLF17 (c). DNA interference PCR for KLF2, KLF5 and KLF17 (d). RT depicts sample with reverse transcriptase, -RT depicts sample without.

4.3 Guide RT-qPCR analysis

To validate the CRISPRa gene expression upregulation, the sample cDNA was amplified using RT-qPCR. Samples were compared to the NT control. The expression was also normalized to a housekeeping gene (in this case Cyclophilin G), to see if the expression was different from a cells baseline expression. In addition, samples were normalized to a Golden Control -DNA sample, which served as a reference point across all replicates. In addition to the three biological replicates, two technical replicated were made by the RT-qPCR process. DNA obtained from iPSC cells was used to test the qPCR primers' attachment to their targets, as iPSC DNA is expected to contain small amounts of the primers' target sequences (Fig 4. a and b).

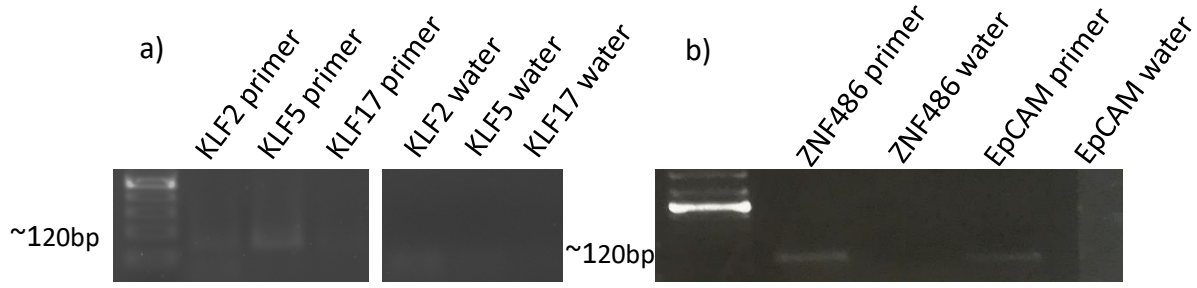


Figure 4. Primer attachment test. PCR to validate the qPCR primers' attachment to their target sequences in D0 iPSC DNA. (a) for KLF2, KLF5, KLF17 and (b) for ZNF486 and EpCAM.

4.4 EpCAM guides resulted in increased gene activation

The results show the highest expression in EpCAM guide 1, which then diminishes gradually towards guide 5 (Fig 5. a). This also suggests a correlation between the guide's distance from the promoter and its effectiveness. Guides were designed from the closest (guide 1), to farthest (guide 5) from the promoter. EpCAM guides 1 and 2 thus showed the highest efficiency and were selected for plasmid assembly and cloning for the following reprogramming. In the case of ZNF486, there were no clear differences in the guide-induced expression levels compared to the NT-control (Fig 5. b). ZNF486 guides 2 and 3 were selected nevertheless for plasmid assembly and cloning, as they seemed most promising in terms of gene activation.

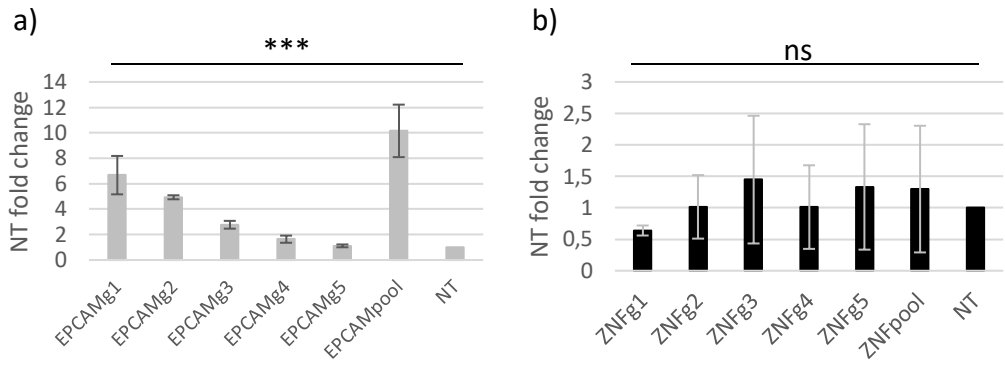
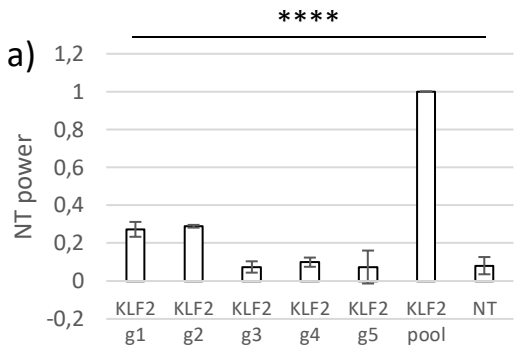
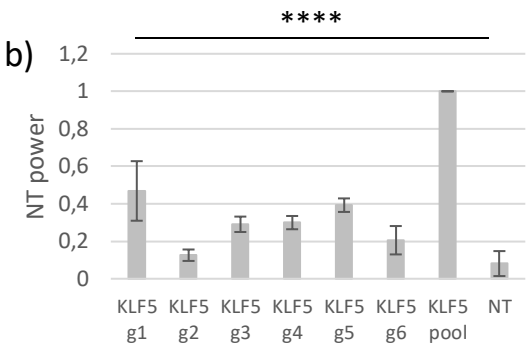


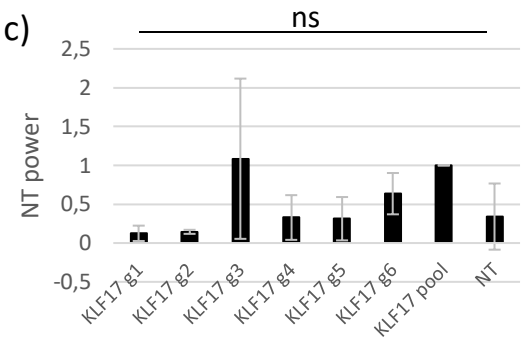
Figure 5. Gene activation levels of EpCAM and ZNF486 guides in HEK293 cells. a) Gene expression levels of EpCAM targeting guides and a pool containing all 5 guides compared to non-transfected control. (Single factor ANOVA *p = 0,000213, significant mean difference indicated by ´***´). b) Gene expression levels of ZNF486 targeting guides and a pool compared to a non-transfected control. (Single factor ANOVA *p = 0,987, non-significant mean difference indicated by ´ns´). Both genes were normalized first to Cyclophilin G housekeeping gene and the to Golden control DNA. Genes were then normalized again to NT for fold change. Error bars depict standard error mean of all three replicates.

4.5 Screen-selected candidate guides showed greatest activation in KLFs

In the case of Krüppel-like factors, KLF2 showed greatest activation in guides 1 and 2 (Fig 6. a), KLF5 in guides 1 and 5 (Fig 6. b), and KLF17 in guides 3 and 6 (Fig 6. c), although the variance between replicates of KLF17 was significant. These guides were then compared to guides enriched in guide screening experiments (Fig 6. d). The purpose of this was to give insight on how accurately these guide screens would reflect the gene activation potential of the enriched genes, and whether this method could be used to select guides for activation in subsequent experiments. Comparing the two sets of guides, we can see that 4 out of 6 guides in the RT-qPCR analysis correspond to their screen equivalents. Thus, guides enriched in screens seem to typically enrich along with their gene activation potential. This should theoretically also relate to the pluripotency inducing potential of their respective genes. Like with EpCAM, the guide closest to the promoter site (guide 1), was the most selected single guide of all the guides both in screens and RT-qPCR. However, in RT-qPCR, KLF17 guide 1 seemed to have the least activation.







d)

Gene	RT-qPCR selected guide	Screen selected guide
KLF2	Guides 1 and 2	Guides 1 and 5
KLF5	Guides 1 and 5	Guides 1 and 5
KLF17	Guides 3 and 6	Guides 1 and 3

Figure 6. Gene activation levels of KLF2, KLF5 and KLF17 guides in HEK293 cells. a) Gene expression levels of KLF2 targeting guides and a pool containing all 5 guides compared to non-transfected control. (Single factor ANOVA *p = 1,6e-12, extremely significant difference indicated by ´****´) b) Gene expression levels of KLF5 targeting guides and a pool compared to non-transfected control. (Single factor ANOVA *p = 1,13e-19, extremely significant difference indicated by ´****´) c) Gene expression levels of KLF17 targeting guides and a pool compared to non-transfected control. All genes were normalized first to Cyclophilin G housekeeping gene and the Golden control DNA. (Single factor ANOVA *p = 0,294, non-significant mean difference indicated by ´ns´) d) Comparison of guides enriched in screens and guides with the highest activation in RT-qPCR for KLF2, KLF5 and KLF17. Guides were normalized to their pools. Error bars depict standard error mean of all three replicates.

4.6 ZNF486 & EpCAM cloning

The selected guides were assembled into three plasmids using Golden gate -ligation, with one containing EpCAM g1 and g2, second ZNF486 g2 and g3, and third containing all four guides (EpCAM g1, g2, ZNF486 g2, g3). The plasmids were first cloned in *E. coli*, and guide inserts were then inserted into EBNA backbone for another cloning and subsequent electroporation. In each step, correct band sized were confirmed with electrophoresis from EcoRI- and SpeI- digested minipreps (Fig. 7). Final midiprep preparations from plasmids were confirmed with sanger sequencing and made into glycerol stocks. However, it was decided that due to the gene activation of ZNF486 guides being negligible, work was continued only with the EpCAM plasmid.

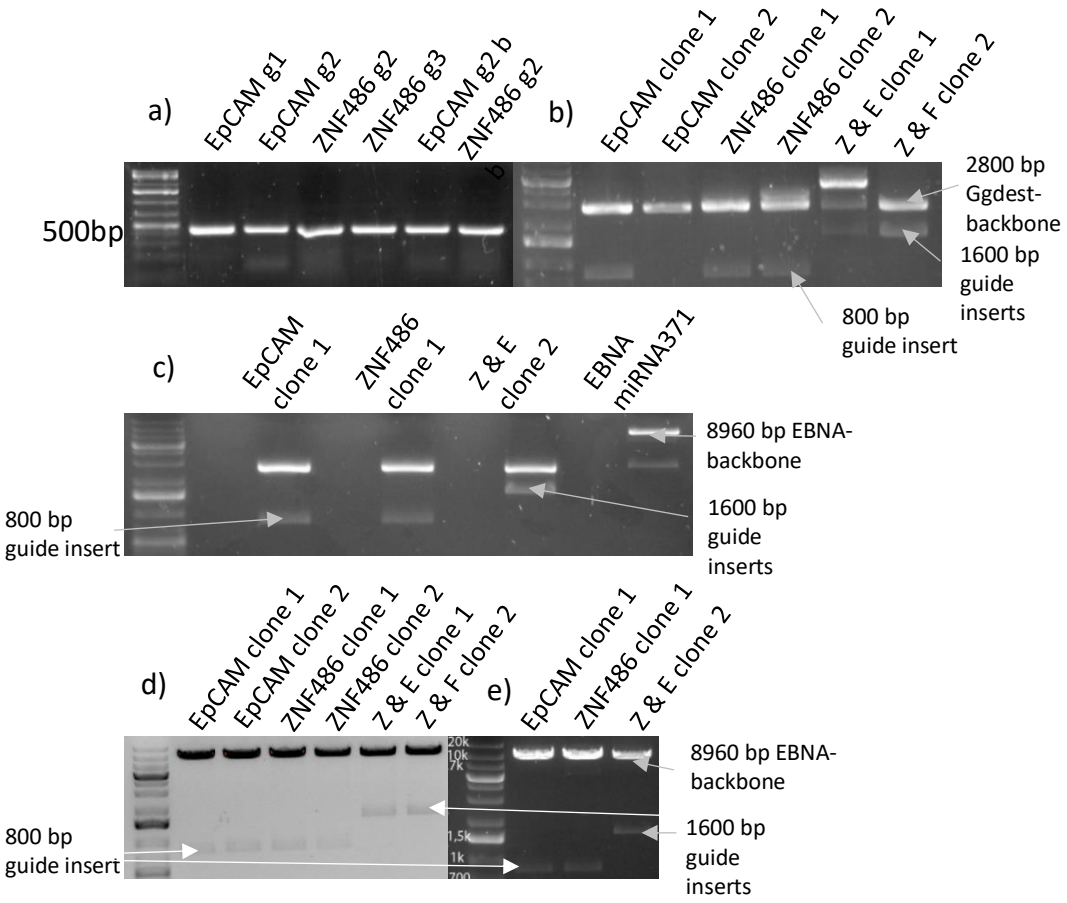


Figure 7. Digestion of GGdest- and EBNA plasmid clones for EpCAM and ZNF486 in *E. coli*. a) Guide assembly PCR of individual guides for Golden Gate -ligation. b) Ggdest-backbone miniprep digestion samples, two clones per plasmid. c) EBNA-backbone digestion samples. d) EBNA backbone miniprep digestion. e) Final EBNA-backbone MIDIprep digestion. In photos, EpCAM clones contain EpCAM guides 1 and 2, ZNF486 clones contain ZNF486 guides 2 and 3, and Z & E clones contain all four guides.

4.7 KLF Cloning

In the case of Krüppel-like factors, two distinct plasmids were made. The first plasmid contained all 6 guides selected by the RT-qPCR analysis. The second plasmid contained all 6 guides selected based on earlier screens. Following the same methodology as with ZNF486 and EpCAM, guides were first inserted into GG-dest backbone followed by EBNA backbone. During all steps, correct band sized were confirmed by digestion and electrophoresis (Fig 8.). Results of guide assembly PCR also show additional dim bands. Given their small size, these are thought to be interlocked primer sequences. Final midiprep preparations from plasmids were confirmed with sanger sequencing and made into glycerol stocks.

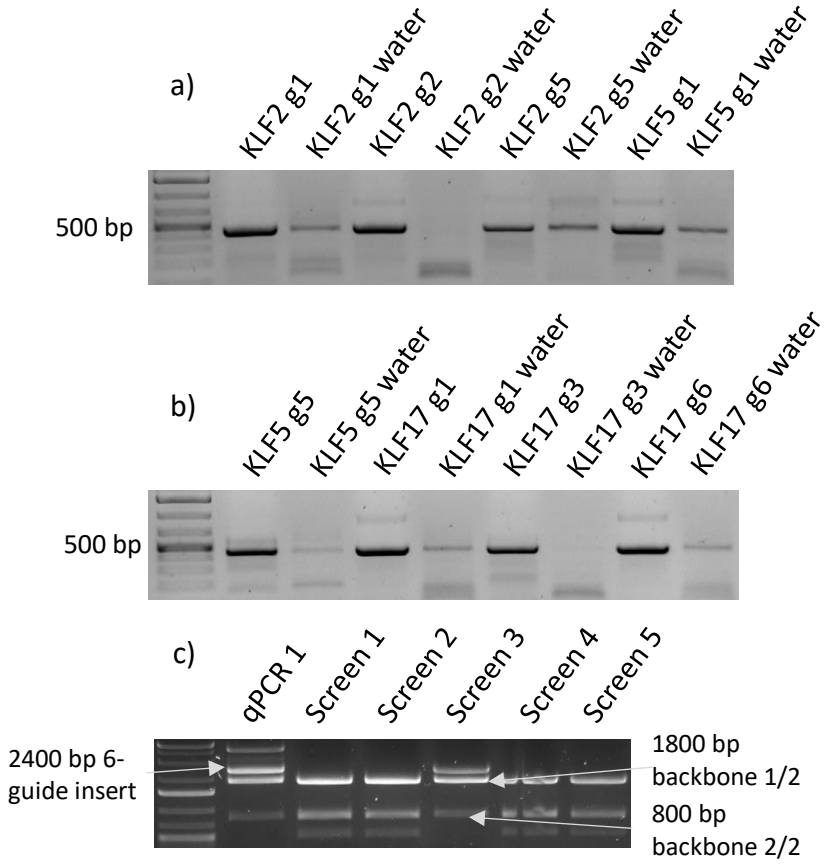


Figure 8. Digestion of GGdest plasmid clones for KLF2, KLF5 and KLF17 in *E. coli.* a) & b) Guide assembly PCR of individual guides for Golden gate -ligation. c) Ggdest-backbone miniprep digestion. Plasmids contain six total guides, two of KLF2, two of KLF5 and two of KLF17.

4.8 EpCAM expression resulted in negligible pluripotent colonies in LCL cells

To determine the effects of EpCAM in reprogramming efficiency, LCL cells were electroporated with dCas9 CRISPR-activator (pCXLE-dCas9VP192-T2A-EGFP-shP53 plasmid), endogenous OSKML activating guides, EpCAM guide plasmid and an EEA plasmid. A sample without EEA was used as a control. Three replicates were made. The cells were compared to CRISPRa + miRNA302 guide plasmid inductions. Transgene reprogramming was used as a positive control. Plasmids used in each electroporation are summarized in Table 13. LCL cells electroporated with plasmids containing the dCas9-activator and guide plasmids were evaluated by their GFP expression three days later (Fig 9. a). Cells treated with EpCAM seemed to grow faster than controls, but their morphology changed towards iPSC morphology slower. EpCAM samples still had attached cell clusters at day 21. Cells not treated with EEA also had a more differentiated and spread out morphology compared to EEA treated cells. All EpCAM samples either

with EEA or without showed a very limited amount of colony formation, compared to transgene and miRNA302 samples (Fig 9. b, c and d). CRISPRa in the absence of EEA unexpectedly showed also no colony formations whatsoever. This could indicate a mistake in inductions. Whether this would also affect EpCAM sample colony formation is to be speculated.

Electroporation	Plasmids
CRISPRa	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902)
Transgene	- pCXLE-hOCT3/4-shp53-F (Addgene #27077) - pCXLE-hSK (Addgene #27078) - pCXLE-hUL (Addgene #27080)
CRISPRa + EEA	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-EEA-5guides-PGK-Puro (Addgene #102898)
EpCAM	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-EPCAM
EpCAM + EEA	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-EPCAM - GG-EBNA-EEA-5guides-PGK-Puro (Addgene #102898)

Table 13. Electroporation conditions for LCL cells. Plasmids transfected for each electroporation referred in Figure 9.

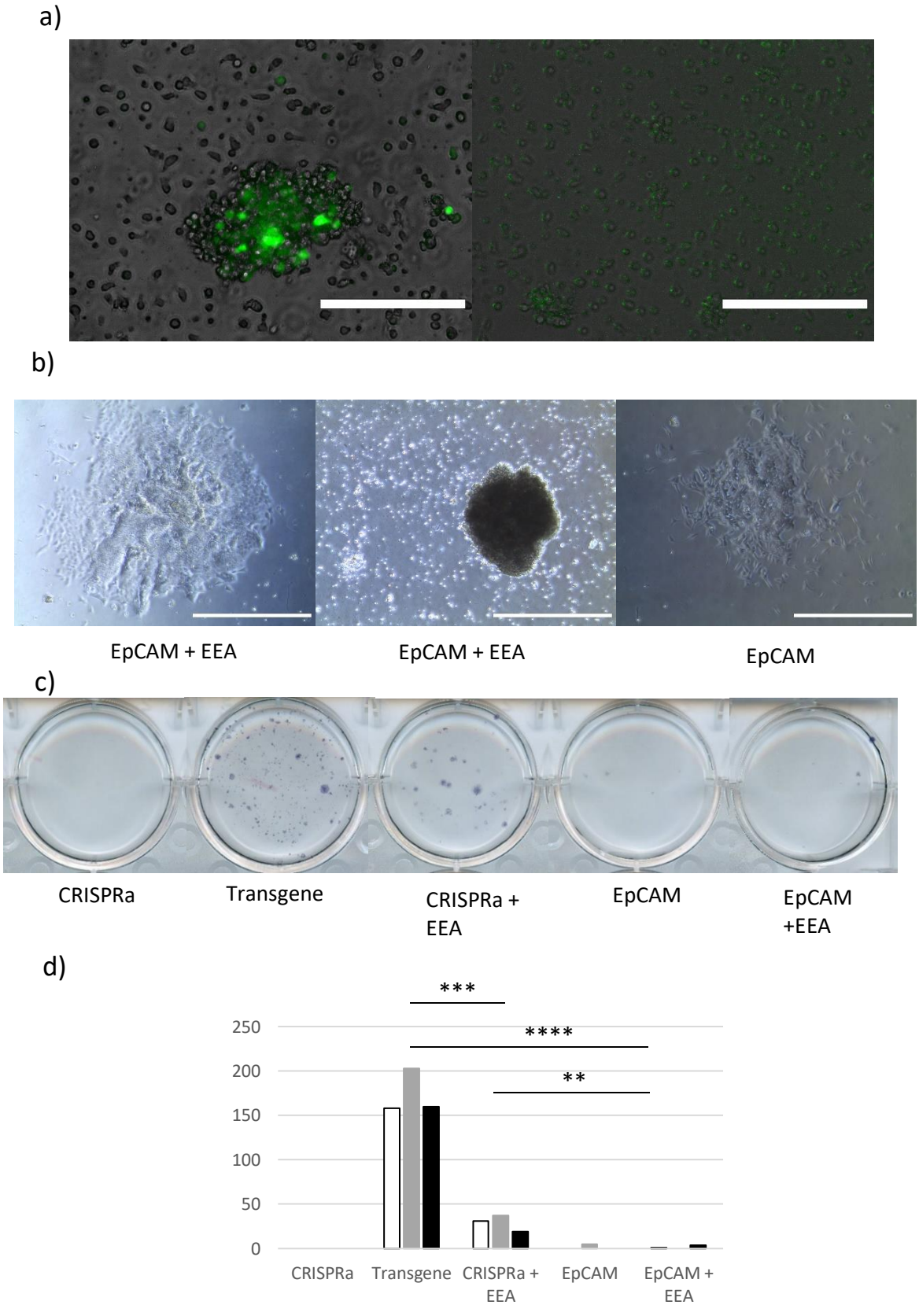


Figure 9. Comparison of reprogramming efficiency in LCL cells between samples reprogrammed using EpCAM without EEA, EpCAM with EEA, miRNA, just CRISPRa, and using transgenes. a) GFP fluorescence of LCL cells at day 3. Scale bars = 400 µM b) Difference in cell morphology between EpCAM with and without EEA at day 21. Scale bars =

400 µM c) Alkaline phosphatase-stained colonies with 1.0 x 10^6 electroporated cells. d) Number of reprogrammed colonies for each method. (Single factor ANOVA, between EpCAM, EpCAM + EEA and Transgene *p = 1,03e-05, between Transgene and CRISPRa + EEA *p = 0,000753, between EpCAM, EpCAM + EEA, CRISPR + EEA *p = 0,0015).

4.9 EpCAM expression did not cause an increase of pluripotent colonies in HFF cells

Additionally, the effect of EpCAM along with a dCas9 activator plasmid was tested on HFF cells. Plasmids used in each electroporation are summarized in Table 14. From early on, EEA treated cells seemed to form pluripotent colonies faster than EpCAM or TdT samples. Otherwise, their morphology did not seem to differ noticeably. From the results can be seen that the control EEA samples effectively created iPSCs, with approximately 100 colonies on each plate, except for one plate which appears to have failed (Fig 10.). Negative TdT control likewise brought up just several individual colonies, as expected. However, on EpCAM-induced samples only one or two colonies can be seen, indicating that endogenous expression of EpCAM on HFF, and possibly on LCL cells, does not increase the formation of iPSCs. Figure 11. depicts which colonies were counted for the graph.

Electroporation	Plasmids
EpCAM	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-EPCAM
EEA	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-EEA-5guides-PGK-Puro (Addgene #102898)
TdT	- pCXLE-dCas9VP192-T2A-EGFP-shP53 (Addgene #69535) - GG-EBNA-MIR302g7-PGK-Puro - GG-EBNA-O3S2K2M2K2L1-PP (Addgene #102902) - GG-EBNA-TdT

Table 14. Elektroporation conditions for HFF cells. Plasmids tranfected for each elektroporation referred in Figure 10.

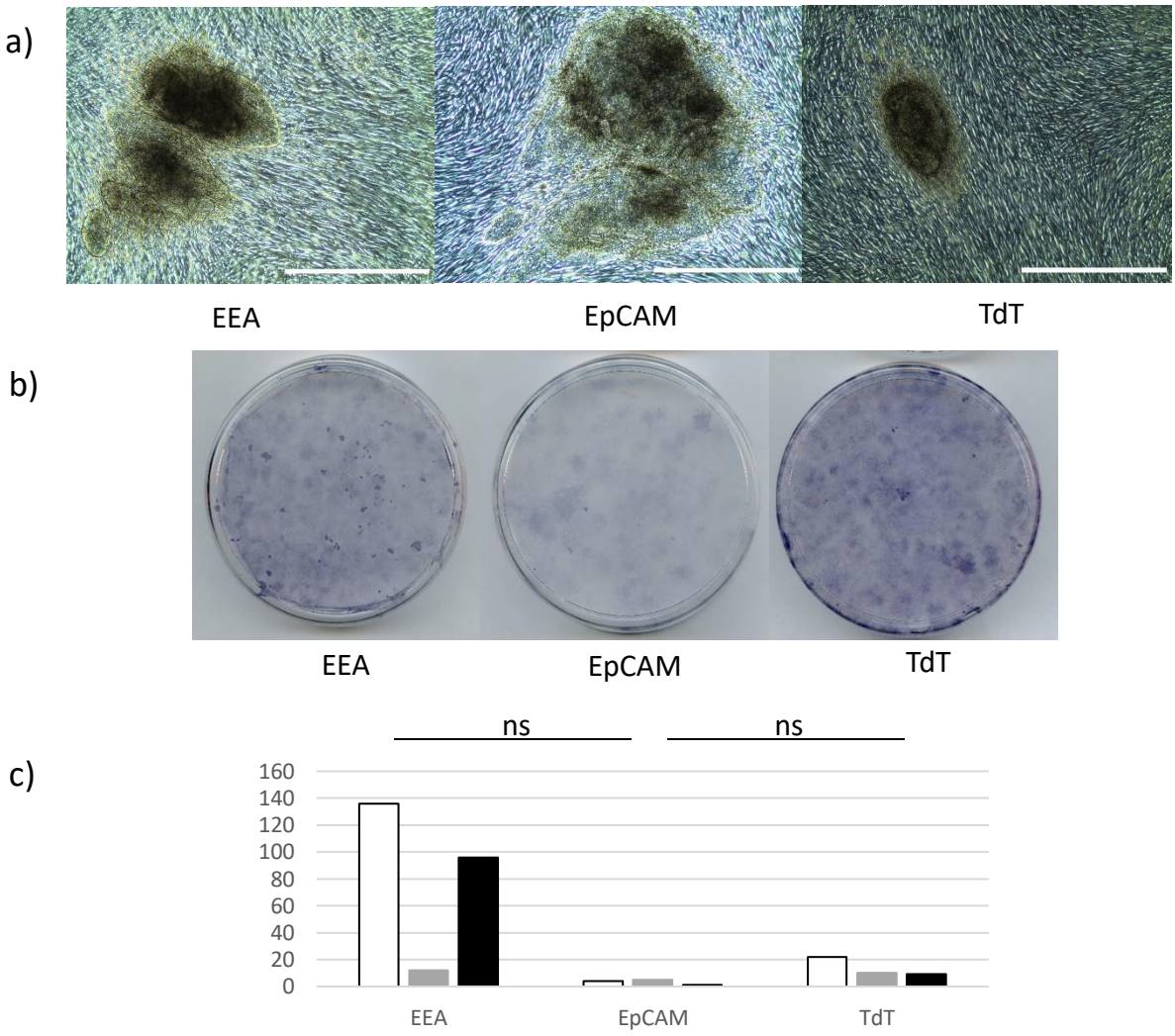


Figure 10. Comparison of reprogramming efficiency in HFF cells between samples reprogrammed using EEA, EpCAM or TdT. a) Difference in cell morphology between samples at day 21. Scale bars = 400 µM b) Alkaline phosphatase stained colonies with 1.0 x 10^6 electroporated cells. c) Number of reprogrammed colonies for each method. (Single factor ANOVA between EEA and EpCAM *p = 0,0998, non-significant mean difference indicated by 'ns'. Between EpCAM and TdT, *p = 0,0762, non-significant mean difference indicated by 'ns')

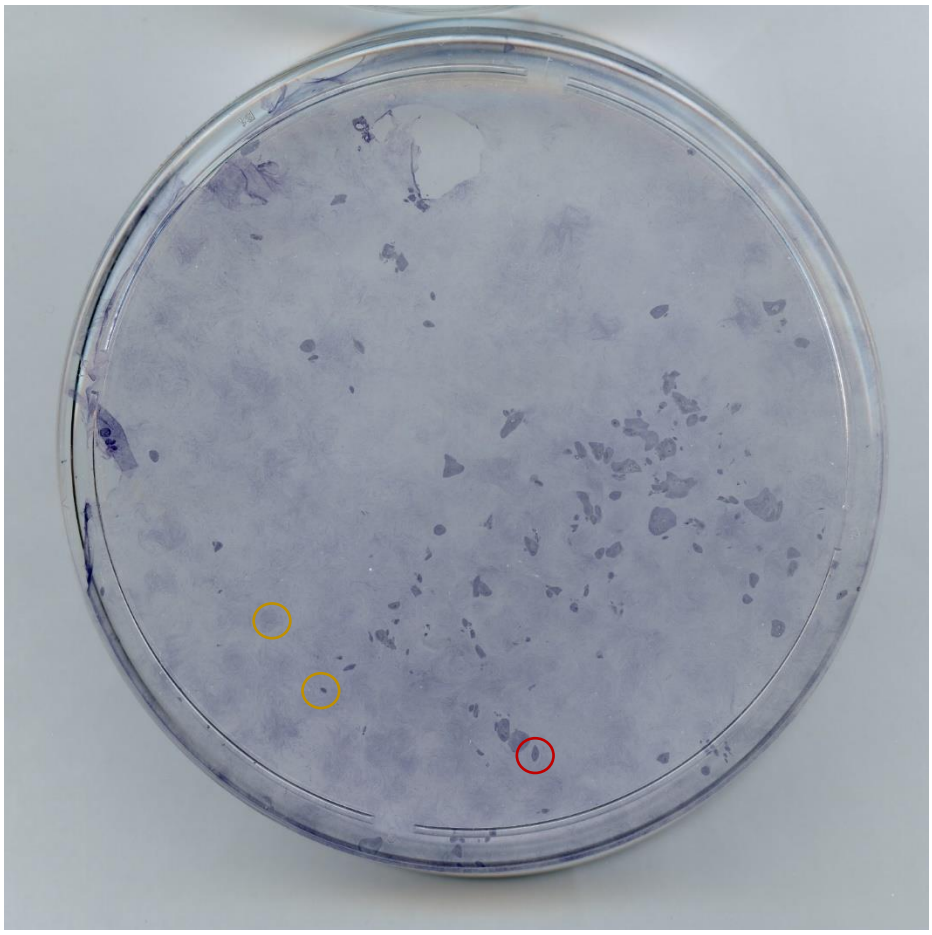


Fig 11. Zoom in of EEA-treated HFF cells. Medium sized colonies resembling iPSC morphology were counter as a reprogrammed colony, example circled in red. Very small specks were not counted as they can be ambiguous, nor were spots of background color. Examples circled in yellow.

5 Discussion

5.1 Selected guides increase endogenous expression

By validating the gene activation levels of screened guide RNAs we aimed to find the two most effective guides for pluripotency factor activation. Additionally, for KLF2, KLF5 and KLF17 we aimed to confirm that the enriched guides were also mostly the ones that best increased gene expression, and likely thus has an effect on reprogramming outcome. This needs to be confirmed on later studies.

All 5 target genes were successfully transfected into HEK293 cells, which caused endogenous activation of the targeted genes. The expression levels of the obtained sample DNA was then validated using RT-qPCR. All genes, with the exception of ZNF486, showed clear differences in activation levels compared to the non-transfected control, with a trend for sites closer to the promoter site to show more activation than the ones further away. This makes sense, as a closer promoter site likely allows for a more efficient contact for the transcription machinery to initiate transcription. However, the interaction between CRISPRa and factors aiding in transcription is not well established, and can affect the gene expression in unexpected ways. The differing gene activation could be affected by blocking of the transcription factors based on the position of the guide sequences. Pool samples, as expected, showed the highest activation. However, using too many endogenous guides at the same time can have diminishing effects on activation, and as such these pool samples of five to six guides are likely thus not optimal for pluripotency induction. Best effects are expected to be observed by pooling together 2 of the best activator guides for each given gene.

5.2 Enrichment reflects guides' gene activation potential

In the case of KLF2, KLF5 and KLF17, four of the six guides which caused highest gene activation in RT-qPCR experiment, also were the ones enriched in the earlier guide screening. These guides were the KLF2 guide 1, KLF5 guides 1 and 5, and KLF17 guide 3. On RT-qPCR, KLF2 guide 2 however shower higher activation in contrast to the enriched guide 5, as was the case with KLF17 guide 1, which was replaced with guide 6 in screens. As such, this indicates that screening guides in bulk by enrichment levels reflects their gene activation levels by most part, and could be used for finding candidate guides for pluripotency enhancement. All enriched guides however caused activation of the target genes, although enriched KLF2 guide 5 and KLF17 guide 1 showed small or

no difference compared to the non-transfected control. For what reason these particular genes were enriched compared to the better ones in RT-qPCR remains to be answered. As the enrichment in screens ultimately reflects the genes' impact on the pluripotency outcome, this might be part due to arbitrary differences with the guides' effects in regards to the final pluripotent state. There also tends to be a limit to how much gene activation is useful, and smaller endogenous activation could be enough to get the full potential. This could be experimented by testing different guides separately in pluripotency inductions, however it needs to be considered whether this is practical. Additionally, as the guides were screened for their effects in bulk, some non-acting guides might affect the efficacy of other guides, enhancing the reprogramming process indirectly. There could be potential in researching the guide combinations, which enriched a particular colony, together. This could give insight to potential co-acting genes in the pluripotency network. Additionally, to see how these guide sets derived from different activation experiments differ in pluripotency enhancement would be interesting to see in the future. As the difference is only between two guides out of six, and accounting for the possible negligible effects between guides, the end result would likely be not much different. Whether the three screened KLF guides enhance the pluripotency in general in addition to the typical OSKML factors is also an important aspect to see.

5.3 EpCAM did not enhance pluripotency in LCL and HFF cells using CRISPRa

Introducing EpCAM to either LCL or HFF cells during the reprogramming process, EpCAM did not seem to enhance the reprogramming. On the contrary, it seemed to inhibit pluripotency generation somewhat. This difference was not however statistically significant in LCL cells, but compared to transgene and CRISPRa + EEA inductions in HFF cells a significant difference can be seen. Amongst the LCL controls, the failed CRISPRa transfection caused some suspicion if the transfections failed. However, other inductions showed expected results, and the following HFF transfections showing similar lack of EpCAM induced iPSCs, it is possible that the transfections worked but just were not able to induce pluripotency.

The cells were successfully cultivated with the typical CRISPRa reprogramming protocol by integrating the DDdCas9VP192 activator along with the OSKML transcription factor -targeting guides, using NaB. EpCAM, negative TdT control and positive EEA control

grew normally with no great differences, except that EEA-induced cells created pluripotent colonies faster. We hypothesized that as EEA induction has been observed to greatly enhance the pluripotency, and TdT should not cause pluripotency enhancement, cells induced with EpCAM alone should reside within EEA and TdT controls (Weltner et al., 2018b). However, EpCAM induced HFF cells showed even less reprogrammed colonies than TdT. This is unexpected, as EpCAM has been shown earlier to enhance pluripotency in fibroblasts (H. P. Huang et al., 2011; Kuan et al., 2017). As the experiment showed similar results in two different occasions with distinct cell types, it raises the question whether this particular method of CRISPR activation or the inducible activators themselves counter this enhancement of reprogramming.

5.4 EpCAM in reprogramming

EpCAM has been observed to increase c-Myc expression in both epithelial cells and fibroblasts and increase cancer cell proliferation (Münz et al., 2004). As c-Myc is a common pluripotency factor, EpCAM was expected to increase reprogramming efficiency (Münz et al., 2004). This has also been confirmed with previous experiments involving EpCAM in reprogramming (H. P. Huang et al., 2011). EpCAM has also been observed to increase OCT4 expression, inhibit the p53 pathway and increase reprogramming efficiency in fibroblasts (H. P. Huang et al., 2011; Kuan et al., 2017). Why EpCAM seemed to inhibit pluripotent reprogramming in this case is not known. The role of EpCAM in cell development is complex, and is known to be relevant in proliferation and differentiation of cells, also during gastrulation (L. Huang et al., 2018). Additionally, in adult cells, it is exclusively expressed in epithelial cells, and thus its effects might differ in the reprogramming of different cell types (Keller et al., 2019).

One obvious explanation is that both of these induction experiments simply failed. According to electrophoresis experiments and sanger sequencing, upregulation of EpCAM using guides was present, and incorporation of EpCAM-targeting guides into inducible plasmids was successful. In this case, the fault would likely reside in unsuccessful electroporation. However, the cells managed to survive in same quantities across samples and the development of all different cell samples was roughly similar. There were some differences, however.

In LCL cells, the morphology of EpCAM only induced cells started to resemble differentiated cells earlier than in EpCAM + EEA cells. The proliferation was also faster.

In HFF cells, EpCAM only induced cells showed normal development and morphology resembling pluripotency. The cells formed somewhat larger colonies compared to EEA and TdT, although not by a large margin.

In LCL cells, it is possible that EpCAM induced cells developed faster, but were kept for too long, resulting in faster differentiation compared to EpCAM + EEA cells. According to previous studies, the development of EpCAM induced stem cells did not differ dramatically, and were able to efficiently induce pluripotency in roughly the same timeframe. However, these studies have been done using fibroblasts (H. P. Huang et al., 2011). At current state, pluripotency induction using EpCAM in LCL cells has not been researched, and the role of EpCAM in their developmental pathway might be different.

In HFF cells, in which EpCAM reprogramming has been successful, the reason might be that the conditions for EpCAM cells was unsuitable because of some differences in the methodology involving CRISPR-activation. In addition, because the conditions required for different cell stages in reprogramming are different, cells developing at different speeds could fall into unsuitable cell media when following standard protocol. This could cause the cells to die faster.

In both cases, these possible factors affecting pluripotent reprogramming outcome using EpCAM could be tested by several means. Firstly, by changing the cell media conditions for samples types individually based on their cell morphology, rather than sharing the timeframe within all samples types, and observing whether there are notable differences in their development. And secondly, by experimenting how induction with EpCAM and CRISPRa alone might affect the cell morphology, via for example ICC-staining for relevant markers.

In addition, as there is not much research on EpCAM reprogramming using LCL cells or other blood cells, it would be useful to experiment whether these cells can be reprogrammed into iPSCs using EpCAM by other means. As the gene activation in this case is done by non-integrative plasmid electroporation, the transient EpCAM expression might drastically affect the outcome compared to a more stable expression. EpCAM, like other genes, can have differing effects depending on the pluripotent state. For this reason, an inducible EpCAM activation could be one notable method to test.

5.5 Future prospects

Guide screening by enrichment appears to be applicable to find candidate guides which effectively upregulate their target genes. Theoretically this would also apply to their effect in the reprogramming outcome. In the future it could be valuable to test enriched guide sets and their co-acting effects. It would also be useful to determine if there are other causes which affect the guide enrichment. Additionally, the candidate genes resulting from screens were not limited to tested KLFs, and the other genes also need to be evaluated for their gene activation and effects in reprogramming.

Of course, guides targeting KLF2, KLF5 and KLF17 evaluated here need to be tested in the future for their reprogramming enhancing effects too. KLF2 and KLF5 could also be used to test if they can replace endogenous KLF4 to achieve pluripotency.

Further experiments are needed to evaluate whether EpCAM upregulation is applicable in a CRISPR-activation system. Likely the CRISPRa method has some slightly distinct mechanics compared to conventional methods which need to be addressed to fully utilize genes studied here.
In addition, the cells generated via EpCAM reprogramming would need to be characterized in the future. AP-staining is not sufficient for determining the cell's pluripotency, and as such the cells should be stained for their pluripotency markers via ICC-staining, and via development to embryoid bodies.

6 Acknowledgements

I want to thank my primary supervisors Jere Weltner and Ras Trokovic for recruiting me into this fascinating project as a part of your team. Thank you Jere for all the instructions on CRISPR and human cell reprogramming, help with the thesis and having the patience to explain things to me over and over again. Your passion and knowledge in the field is really something to strive for. And thank you Ras for all the help in understanding the experiments and keeping things logical, and of course for bringing a happy and welcoming atmosphere, making it a joy to get to the lab every morning.

I want to also thank Timo Otonkoski for providing me with this opportunity to work in his team. You lead an excellent and determined team and this has been a wonderful experience in a friendly and encouraging environment. It is no wonder your team has achieved so much.

Thank you Joonas Sokka, Eeva-Mari Jouhilahti and Solja Eurola for always being there to openly help me with any problem encountered. Joonas, I appreciate your will to go out of your way to help others and your assistance with the cell maintenance. Eeva-Mari, thank you for teaching all the various assays, helping understand the results and keeping them logical, and also being frank with things. Thank you Solja for the help with the PCR and all other unending questions I had.

Additionally, thank you Hossam, Christian, Sami and Hazem for always being there to help with the endless conundrum called qPCR, along with your kindness and positive attitude.

Thank you Anni for introducing me to stem cell maintenance and helping with all the orders, and thank you both Anni and Heli for keeping the lab running and for getting reagents ready even on a short notice. Thank you also to Jarkko, Väinö, Tom, Matthew, Ro, Laura, Vikash, and any others I might have missed, for you help, advice and maintaining a good atmosphere in the lab.

Finally, I would like to thank my family for for their love and support throughout my studies. You’re the best.

7 References

Anokye-Danso, F., Trivedi, C., & Juhr, D. (2012) Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency. *Bone **23***:1–7. DOI: 10.1016/j.stem.2011.03.001.Highly. [Epub ahead of print]

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., & Otonkoski, T. (2015) Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. *Stem Cell Reports **5***:448–459. DOI: 10.1016/j.stemcr.2015.08.001. [Epub ahead of print]

Bialkowska, A. B., Yang, V. W., & Mallipattu, S. K. (2017) Krüppel-like factors in mammalian stem cells and development. *Development (Cambridge) **144***:737–754. DOI: 10.1242/dev.145441. [Epub ahead of print]

Blakeley, P., Fogarty, N. M. E., Valle, I., Wamaitha, S. E., Hu, T. X., Elder, K., Snell, P., Christie, L., Robson, P., Niakan, K. K., Blakeley, P., Fogarty, N. M. E., Valle, I., Wamaitha, S. E., Hu, T. X., Elder, K., Snell, P., Christie, L., Robson, P., & Niakan, K. K. (2015) Erratum to Defining the three cell lineages of the human blastocyst by single-cell RNA-seq (Development, (2015) 142, 3151-3165). *Development (Cambridge) **142***:3613. DOI: 10.1242/dev.131235. [Epub ahead of print]

Bourillot, P. Y., & Savatier, P. (2010) Krüppel-like transcription factors and control of pluripotency. *BMC Biology **8***:8–10. DOI: 10.1186/1741-7007-8-125. [Epub ahead of print]

Boyer, L., Lee, T., & Cole, M. (2005) Core Transcriptional Regulatory Circuitry in Human Embryonic. *Cell **2005***:144–147. DOI: 10.1016/j.cell.2005.08.020.Core. [Epub ahead of print]

Buganim, Y., Faddah, D., & Jaenisch, R. (2013) Mechanisms and models of somatic cell reprogramming Yosef. *Bone **23***:1–7. DOI: 10.1038/nrg3473.Mechanisms. [Epub ahead of print]

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., & Raschellà, G. (2017) Zinc-finger proteins in health and disease. *Cell Death Discovery **3***:. DOI: 10.1038/cddiscovery.2017.71. [Epub ahead of print]

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., Guo, L., Zhu, J., Zhao, X., Peng, T., Zhang, Y., Chen, S., Li, X., Li, D., Wang, T., & Pei, D. (2013) H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genetics **45***:34–42. DOI: 10.1038/ng.2491. [Epub ahead of print]

Chin, M., Mason, M., & Xie, W. (2012) Induced Pluripotent Stem Cells and Embryonic Stem Cells Are Distinguished by Gene Expression Signatures. *Bone **23***:1–7. DOI: 10.1016/j.stem.2009.06.008.Induced. [Epub ahead of print]

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., Plath, K., Comprehensive, J., Program, B., & Angeles, L. (2017) *Cooperative binding of transcription factors orchestrates reprogramming **168***:442–459. DOI: 10.1016/j.cell.2016.12.016.Cooperative. [Epub ahead of print]

Condic, M. L. (2014) Totipotency: What it is and what it is not. In *Stem Cells and Development* (Vol. 23, Issue 8, pp. 796–812) Mary Ann Liebert Inc. DOI: 10.1089/scd.2013.0364. [Epub ahead of print]

Dang, C. (2012) MYC on the path to cancer. *Cell* ***76***:211–220. DOI: 10.1016/j.cell.2012.03.003.MYC. [Epub ahead of print]

Delgado-Olguín, P., & Recillas-Targa, F. (2011) Chromatin structure of pluripotent stem cells and induced pluripotent stem cells. *Briefings in Functional Genomics* ***10***:37–49. DOI: 10.1093/bfgp/elq038. [Epub ahead of print]

Evans, M. & Kaufman, M. (1981). Establisment in Culture of Pluripotential Cells from Mouse Embryos. *Nature,* ***292*****:** 154-156.

Folmes, C. D. L., Nelson, T. J., Martinez-Fernandez, A., Arrell, D. K., Lindor, J. Z., Dzeja, P. P., Ikeda, Y., Perez-Terzic, C., & Terzic, A. (2011) Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metabolism* ***14***:264–271. DOI: 10.1016/j.cmet.2011.06.011. [Epub ahead of print]

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., & Hasegawa, M. (2009) Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy Series B: Physical and Biological Sciences* ***85***:348–362. DOI: 10.2183/pjab.85.348. [Epub ahead of print]

Gaspar-Maia, A., Alajem, A., Meshorer, E., & Ramalho-Santos, M. (2011) Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology* ***12***:36–47. DOI: 10.1038/nrm3036. [Epub ahead of print]

Ghaleb, A. M., & Yang, V. W. (2017) Krüppel-like factor 4 (KLF4): What we currently know Amr. *Encyclopedia of Signaling Molecules* 2776–2776. DOI: 10.1007/978-3-319-67199-4_105444. [Epub ahead of print]

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Gloria, A., Torres, S. E., Stern-ginossar, N., Brandman, O., Whitehead, H., Doudna, J. A., Lim, W. A., & Jonathan, S. (2013) *CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes* ***154***:442–451. DOI: 10.1016/j.cell.2013.06.044.CRISPR-Mediated. [Epub ahead of print]

Gumireddy, K., Li, A., Gimotty, P. a, Klein-szanto, A. J., Louise, C., Katsaros, D., Coukos, G., Zhang, L., & Huang, Q. (2010) *Transition and Metastasis in Breast Cancer* ***11***:1297–1304. DOI: 10.1038/ncb1974.KLF17. [Epub ahead of print]

Guo, G., Von Meyenn, F., Santos, F., Chen, Y., Reik, W., Bertone, P., Smith, A., & Nichols, J. (2016) Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* ***6***:437–446. DOI: 10.1016/j.stemcr.2016.02.005. [Epub ahead of print]

Halevy, T., & Urbach, A. (2014) Comparing ESC and iPSC—Based Models for Human Genetic Disorders. *Journal of Clinical Medicine* ***3***:1146–1162. DOI: 10.3390/jcm3041146. [Epub ahead of print]

Han, G., Lu, C., Guo, J., Qiao, Z., Sui, N., Qiu, N., & Wang, B. (2020) C2H2 Zinc Finger Proteins: Master Regulators of Abiotic Stress Responses in Plants. *Frontiers in Plant Science* ***11***:1–13. DOI: 10.3389/fpls.2020.00115. [Epub ahead of print]

Hanahan, D., & Weinberg, R. A. (2011) Hallmarks of cancer: The next generation. *Cell* ***144***:646–674. DOI: 10.1016/j.cell.2011.02.013. [Epub ahead of print]

Hansson et al. (2015) Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Physiology & Behavior* ***176***:139–148. DOI: 10.1016/j.celrep.2012.10.014.Highly. [Epub ahead of print]

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., & Yamanaka, S. (2009) Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* ***460***:1132–1135. DOI: 10.1038/nature08235. [Epub ahead of print]

Hu, K. (2014) Vectorology and factor delivery in induced pluripotent stem cell reprogramming. In *Stem Cells and Development* (Vol. 23, Issue 12, pp. 1301–1315) Mary Ann Liebert Inc. DOI: 10.1089/scd.2013.0621. [Epub ahead of print]

Huang, H. P., Chen, P. H., Yu, C. Y., Chuang, C. Y., Stone, L., Hsiao, W. C., Li, C. L., Tsai, S. C., Chen, K. Y., Chen, H. F., Ho, H. N., & Kuo, H. C. (2011) Epithelial cell adhesion molecule (EpCAM) complex proteins promote transcription factor-mediated pluripotency reprogramming. *Journal of Biological Chemistry* ***286***:33520–33532. DOI: 10.1074/jbc.M111.256164. [Epub ahead of print]

Huang, L., Yang, Y., Yang, F., Liu, S., Zhu, Z., Lei, Z., & Guo, J. (2018) Functions of EpCAM in physiological processes and diseases (Review). *International Journal of Molecular Medicine* ***42***:1771–1785. DOI: 10.3892/ijmm.2018.3764. [Epub ahead of print]

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. *Nature Biotechnology* ***26***:795–797. DOI: 10.1038/nbt1418. [Epub ahead of print]

Jaenisch, R., & Young, R. (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Canadian Television Today* ***132***:66–101. DOI: 10.2307/j.ctv6cfqtv.7. [Epub ahead of print]

Jha, P., & Das, H. (2017) KLF2 in regulation of NF-κB-mediated immune cell function and inflammation. *International Journal of Molecular Sciences* ***18***:. DOI: 10.3390/ijms18112383. [Epub ahead of print]

Jia, F., Wilson, K., & Sun, N. (2010) A nonviral minicircle vector for deriving Human iPSC cells. *Early Human Development* ***83***:1–11. DOI: 10.1016/j.earlhumdev.2006.05.022. [Epub ahead of print]

Junying, Y., Kejin, H., Kim, S. O., Shulan, T., Stewart, R., Slukvin, I. I., & Thomson, J. A. (2009) Human induced pluripotent stem cells free of vector and transgene sequences. *Science* ***324***:797–801. DOI: 10.1126/science.1172482. [Epub ahead of print]

Kawamura, T., Suzuki, J., Wang, Y. V., Menendez, S., Morera, L. B., Raya, A., Wahl, G. M., & Belmonte, J. C. I. (2009) Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* ***460***:1140–1144. DOI: 10.1038/nature08311. [Epub ahead of print]

Keller, L., Werner, S., & Pantel, K. (2019) Biology and clinical relevance of EpCAM. *Cell Stress* ***3***:165–180. DOI: 10.15698/cst2019.06.188. [Epub ahead of print]

Kim, J., Chu, J., Shen, X., Wang, J., & Orkin, S. H. (2008) *An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells Acetylation Is Indispensable for p53 Activation* ***005***:2008.

Kiskinis, E., & Eggan, K. (2010) Progress toward the clinical application of patient-specific pluripotent stem cells. *Journal of Clinical Investigation* ***120***:51–59. DOI: 10.1172/JCI40553. [Epub ahead of print]

Kitamura, T., Koshino, Y., Shibata, F., Oki, T., Nakajima, H., Nosaka, T., & Kumagai, H. (2003) Retrovirus-mediated gene transfer and expression cloning: Powerful tools in functional genomics. In *Experimental Hematology* (Vol. 31).

Kuan, I. I., Liang, K. H., Wang, Y. P., Kuo, T. W., Meir, Y. J. J., Wu, S. C. Y., Yang, S. C., Lu, J., & Wu, H. C. (2017) EpEX/EpCAM and Oct4 or Klf4 alone are sufficient to generate induced pluripotent stem cells through STAT3 and HIF2α. *Scientific Reports* ***7***:1–14. DOI: 10.1038/srep41852. [Epub ahead of print]

Kurata, M., Wolf, N. K., Lahr, W. S., Weg, M. T., Kluesner, M. G., Lee, S., Hui, K., Shiraiwa, M., Webber, B. R., & Moriarity, B. S. (2018) Highly multiplexed genome engineering using CRISPR/Cas9 gRNA arrays. *PLoS ONE* ***13***:. DOI: 10.1371/journal.pone.0198714. [Epub ahead of print]

Lee, H., Han, S., Kwon, C. S., & Lee, D. (2016) Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein and Cell* ***7***:100–113. DOI: 10.1007/s13238-015-0212-y. [Epub ahead of print]

Li, M., & Belmonte, J. (2018) Deconstructing the pluripotency gene regulatory network. *Physiology & Behavior* ***176***:139–148. DOI: 10.1038/s41556-018-0067-6.Deconstructing. [Epub ahead of print]

Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., Qin, B., Xu, J., Li, W., Yang, J., Gan, Y., Qin, D., Feng, S., Song, H., Yang, D., … Pei, D. (2010) A mesenchymal-to-Epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* ***7***:51–63. DOI: 10.1016/j.stem.2010.04.014. [Epub ahead of print]

Lin, Y., Chen, G., Engineering, G., & Facility, C. (2020) Embryoid body formation from human pluripotent stem cells in chemically defined E8 media. *Nonwovens* 56–116. DOI: 10.1201/9781315365022-10. [Epub ahead of print]

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., Ogmalley, R., Castanon, R., Klugman, S., Downes, M., Yu, R., Stewart, R., Ren, B., Thomson, J. A., Evans, R. M., & Ecker, J. R. (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* ***471***:68–73. DOI: 10.1038/nature09798. [Epub ahead of print]

Malik, N., & Rao, M. S. (2013) A review of the methods for human iPSC derivation. *Methods in Molecular Biology* ***997***:23–33. DOI: 10.1007/978-1-62703-348-0_3. [Epub ahead of print]

Martí, M., Mulero, L., Pardo, C., Morera, C., Carrió, M., Laricchia-Robbio, L., Esteban, C. R., & Belmonte, J. C. I. (2013) Characterization of pluripotent stem cells. *Nature Protocols* ***8***:223–253. DOI: 10.1038/nprot.2012.154. [Epub ahead of print]

Münz, M., Kieu, C., Mack, B., Schmitt, B., Zeidler, R., & Gires, O. (2004) The carcinoma-associated antigen EpCAM upregulates c-myc and induces cell proliferation. *Oncogene* ***23***:5748–5758. DOI: 10.1038/sj.onc.1207610. [Epub ahead of print]

Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T., & Yamanaka, S. (2010) Promotion of direct reprogramming by transformation-deficient Myc. *Proceedings*

of the National Academy of Sciences of the United States of America ***107***:14152–14157. DOI: 10.1073/pnas.1009374107. [Epub ahead of print]

Nefzger, C. M., Rossello, F. J., Chen, J., Liu, X., Knaupp, A. S., Firas, J., Paynter, J. M., Pflueger, J., Buckberry, S., Lim, S. M., Williams, B., Alaei, S., Faye-Chauhan, K., Petretto, E., Nilsson, S. K., Lister, R., Ramialison, M., Powell, D. R., Rackham, O. J. L., & Polo, J. M. (2017) Cell Type of Origin Dictates the Route to Pluripotency. *Cell Reports* ***21***:2649–2660. DOI: 10.1016/j.celrep.2017.11.029. [Epub ahead of print]

Nethercott, H. E., Brick, D. J., & Schwartz, P. H. (2011) Immunocytochemical analysis of human pluripotent stem cells. *Methods in Molecular Biology* ***767***:201–220. DOI: 10.1007/978-1-61779-201-4_15. [Epub ahead of print]

Nishimura, K., Fukuda, A., & Hisatake, K. (2019) Mechanisms of the metabolic shift during somatic cell reprogramming. *International Journal of Molecular Sciences* ***20***:. DOI: 10.3390/ijms20092254. [Epub ahead of print]

Nishimura, K., Kato, T., Chen, C., Oinam, L., Shiomitsu, E., Ayakawa, D., Ohtaka, M., Fukuda, A., Nakanishi, M., & Hisatake, K. (2014) Manipulation of KLF4 expression generates iPSCs paused at successive stages of reprogramming. *Stem Cell Reports* ***3***:915–929. DOI: 10.1016/j.stemcr.2014.08.014. [Epub ahead of print]

Okita, K., Matsumura, Y., Sato, Y., Okada, A., Morizane, A., Okamoto, S., Hong, H., Nakagawa, M., Tanabe, K., Tezuka, K.-I., Shibata, T., Kunisada, T., Takahashi, M., Takahashi, J., Saji, H., & Yamanaka, S. (2011) *A more efficient method to generate integration-free human iPS cells*.

Patriarca, C., Macchi, R. M., Marschner, A. K., & Mellstedt, H. (2012) Epithelial cell adhesion molecule expression (CD326) in cancer: A short review. *Cancer Treatment Reviews* ***38***:68–75. DOI: 10.1016/j.ctrv.2011.04.002. [Epub ahead of print]

Peinado, H., Quintanilla, M., & Cano, A. (2003) Transforming growth factor β-1 induces Snail transcription factor in epithelial cell lines. Mechanisms for epithelial mesenchymal transitions. *Journal of Biological Chemistry* ***278***:21113–21123. DOI: 10.1074/jbc.M211304200. [Epub ahead of print]

Poli, V., Fagnocchi, L., Fasciani, A., Cherubini, A., Mazzoleni, S., Ferrillo, S., Miluzio, A., Gaudioso, G., Vaira, V., Turdo, A., Giaggianesi, M., Chinnici, A., Lipari, E., Bicciato, S., Bosari, S., Todaro, M., & Zippo, A. (2018) MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. *Nature Communications* ***9***:. DOI: 10.1038/s41467-018-03264-2. [Epub ahead of print]

Pollak, N. M., Hoffman, M., Goldberg, I. J., & Drosatos, K. (2018) Krüppel-Like Factors: Crippling and Uncrippling Metabolic Pathways. *JACC: Basic to Translational Science* ***3***:132–156. DOI: 10.1016/j.jacbts.2017.09.001. [Epub ahead of print]

Polo, J. M., Anderssen, E., Walsh, R. M., Schwarz, B. A., Nefzger, C. M., Lim, S. M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., Bar-nur, O., Cheloufi, S., Stadtfeld, M., Figueroa, E., Robinton, D., Natesan, S., Melnick, A., & Zhu, J. (2012) *A molecular roadmap of cellular reprogramming into iPS cells Jose* ***151***:1617–1632. DOI: 10.1016/j.cell.2012.11.039.A. [Epub ahead of print]

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T. W., Jaenisch, R., & Trono, D. (2019) Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. *Cell Stem Cell* ***24***:724-735.e5. DOI: 10.1016/j.stem.2019.03.012. [Epub ahead of print]

Radzisheuskaya, A., & Silva, J. C. R. (2014) Do all roads lead to Oct4? The emerging concepts of induced pluripotency. *Trends in Cell Biology* ***24***:275–284. DOI: 10.1016/j.tcb.2013.11.010. [Epub ahead of print]

Romito, A., & Cobellis, G. (2016) Pluripotent stem cells: Current understanding and future directions. In *Stem Cells International* (Vol. 2016) Hindawi Publishing Corporation. DOI: 10.1155/2016/9451492. [Epub ahead of print]

Ruiz, S., Panopoulos, A., & Herrerías, A. (2011) A high proliferation rate is required for cell reprogramming and maintenance of human embryonic stem cell identity. *Bone* ***23***:1–7. DOI: 10.1016/j.cub.2010.11.049.A. [Epub ahead of print]

Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H. K., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A., & Wrana, J. L. (2010) Functional genomics reveals a BMP-Driven mesenchymal-to-Epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* ***7***:64–77. DOI: 10.1016/j.stem.2010.04.015. [Epub ahead of print]

Sarkar, A., & Hochedlinger, K. (2013) The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. *Bone* ***23***:1–7. DOI: 10.1016/j.stem.2012.12.007.The. [Epub ahead of print]

Seth, R. B., Sun, L., & Chen, Z. J. (2006) Antiviral innate immunity pathways. *Cell Research* ***16***:141–147. DOI: 10.1038/sj.cr.7310019. [Epub ahead of print]

Singh, U., Quintanilla, R. H., Grecian, S., Gee, K. R., Rao, M. S., & Lakshmipathy, U. (2012) Novel Live Alkaline Phosphatase Substrate for Identification of Pluripotent Stem Cells. *Stem Cell Reviews and Reports* ***8***:1021–1029. DOI: 10.1007/s12015-012-9359-6. [Epub ahead of print]

Soufi, A., Donahue, G., & Zaret, K. (2012) Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome Abdenour. *Bone* ***23***:1–7. DOI: 10.1016/j.cell.2012.09.045.Facilitators. [Epub ahead of print]

Stadelmann, C. (2019) *Improving Human Pluripotent Reprogramming by Targeted Activation of the miR-302/367 Cluster Using CRISPRa*.

Takahashi, K., & Yamanaka, S. (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* ***126***:663–676. DOI: 10.1016/j.cell.2006.07.024. [Epub ahead of print]

Thomson, J., Itskovitz-Eldor, J., Shapiro, S., Waknitz, M., Swiegiel, J., Marshall, V. and Jones, J. (1998) Embryonic stem cell lines derived from human blastocysts. *Science,* ***282*****:** 1145-7.

Töhönen, V., Katayama, S., Vesterlund, L., Jouhilahti, E. M., Sheikhi, M., Madissoon, E., Filippini-Cattaneo, G., Jaconi, M., Johnsson, A., Bürglin, T. R., Linnarsson, S., Hovatta, O., & Kere, J. (2015) Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nature Communications* ***6***:1–9. DOI: 10.1038/ncomms9207. [Epub ahead of print]

Van Den Hurk, M., Kenis, G., Bardy, C., Van Den Hove, D. L., Gage, F. H., Steinbusch, H. W., & Rutten, B. P. (2016) Transcriptional and epigenetic mechanisms of cellular reprogramming to induced pluripotency. In *Epigenomics* (Vol. 8, Issue 8, pp. 1131–1149) Future Medicine Ltd. DOI: 10.2217/epi-2016-0032. [Epub ahead of print]

Velychko, S., Adachi, K., Kim, K. P., Hou, Y., MacCarthy, C. M., Wu, G., & Schöler, H. R. (2019) Excluding Oct4 from Yamanaka Cocktail Unleashes the Developmental Potential of iPSCs. *Cell Stem Cell **25***:737-753.e4. DOI: 10.1016/j.stem.2019.10.002. [Epub ahead of print]

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., & Liu, P. (2008a) Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America **105***:9290–9295. DOI: 10.1073/pnas.0801017105. [Epub ahead of print]

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., & Liu, P. (2008b) *Chromosomal transposition of PiggyBac in mouse embryonic stem cells*. www.pnas.org/cgi/content/full/

Warren, L., Manos, P. D., Ahfeldt, T., Loh, Y. H., Li, H., Lau, F., Ebina, W., Mandal, P. K., Smith, Z. D., Meissner, A., Daley, G. Q., Brack, A. S., Collins, J. J., Cowan, C., Schlaeger, T. M., & Rossi, D. J. (2010) Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell **7***:618–630. DOI: 10.1016/j.stem.2010.08.012. [Epub ahead of print]

Weltner, J., Anisimov, A., Alitalo, K., Otonkoski, T., & Trokovic, R. (2012) Induced Pluripotent Stem Cell Clones Reprogrammed via Recombinant Adeno-Associated Virus-Mediated Transduction Contain Integrated Vector Sequences. *Journal of Virology **86***:4463–4467. DOI: 10.1128/jvi.06302-11. [Epub ahead of print]

Weltner, Jere (2018a) *Novel Approaches for Pluripotent Reprogramming* (Issue 53).

Weltner, Jere, Balboa, D., Katayama, S., Bespalov, M., Krjutškov, K., Jouhilahti, E. M., Trokovic, R., Kere, J., & Otonkoski, T. (2018b) Human pluripotent reprogramming with CRISPR activators. *Nature Communications **9***:. DOI: 10.1038/s41467-018-05067-x. [Epub ahead of print]

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H. K., & Nagy, A. (2009) PiggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature **458***:766–770. DOI: 10.1038/nature07863. [Epub ahead of print]

Worringer, K. A., Rand, T. A., Hayashi, Y., Sami, S., Takahashi, K., Tanabe, K., Narita, M., Srivastava, D., & Yamanaka, S. (2014) The let-7/LIN-41 pathway regulates reprogramming to human induced pluripotent stem cells by controlling expression of prodifferentiation genes. *Cell Stem Cell **14***:40–52. DOI: 10.1016/j.stem.2013.11.001. [Epub ahead of print]

Yates, J. L., Camiolo, S. M., & Bashaw, J. M. (2000) The Minimal Replicator of Epstein-Barr VirusoriP. *Journal of Virology **74***:4512–4522. DOI: 10.1128/jvi.74.10.4512-4522.2000. [Epub ahead of print]

Yu, J., Vodyanik, M. & Smuga-Otto, K., Vodyanik M., Antosiewicz-Bourget, J., Frane, J., Tian, J., Nie, J., Jonsdottir, G., Ruotti, V., Stewart, R., Slukvin, I. & Thomson,

J. (2007) Induced pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* ***318*****:** 1917-20.

Yusa, K., Rad, R., Takeda, J., & Bradley, A. (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nature Methods* ***6***:363–369. DOI: 10.1038/nmeth.1323. [Epub ahead of print]

Zavadil, J., Bitzer, M., Liang, D., Yang, Y. C., Massimi, A., Kneitz, S., Piek, E., & Böttinger, E. P. (2001) Genetic programs of epithelial cell plasticity directed by transforming growth factor-β. *Proceedings of the National Academy of Sciences of the United States of America* ***98***:6686–6691. DOI: 10.1073/pnas.111614398. [Epub ahead of print]

Zhang, J., Ratanasirintrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S. B., Yu, C., Ross, C. A., Cacchiarelli, D., Xia, Q., Seligson, M., Shinoda, G., Xie, W., Cahan, P., Wang, L., Ng, S. C., Tintara, S., Trapnell, C., Onder, T., Loh, Y. H., … Daley, G. Q. (2016) LIN28 Regulates Stem Cell Metabolism and Conversion to Primed Pluripotency. *Cell Stem Cell* ***19***:66–80. DOI: 10.1016/j.stem.2016.05.009. [Epub ahead of print]

Zhao, Y., Yin, X., Qin, H., Zhu, F., Liu, H., Yang, W., Zhang, Q., Xiang, C., Hou, P., Song, Z., Liu, Y., Yong, J., Zhang, P., Cai, J., Liu, M., Li, H., Li, Y., Qu, X., Cui, K., … Deng, H. (2008) Two Supporting Factors Greatly Improve the Efficiency of Human iPSC Generation. In *Cell Stem Cell* (Vol. 3, Issue 5, pp. 475–479). DOI: 10.1016/j.stem.2008.10.002. [Epub ahead of print]

Zhou, S., Tang, X., & Tang, F. (2016) Krüppel-like factor 17, a novel tumor suppressor: its low expression is involved in cancer metastasis. *Tumor Biology* ***37***:1505–1513. DOI: 10.1007/s13277-015-4588-3. [Epub ahead of print]