Understanding the Role of ZNF765 and/or L1PA Retroelements in Defining Embryonic Stem Cell Phenotypes
Abstract

Many regulatory mechanisms during early development remain largely unexplored in humans. These mechanisms vary throughout species and by better understanding what these mechanisms are, we can gain a better grasp on how malfunctions in those mechanisms result in early developmental diseases in humans. To begin to delve into this issue, it is important to learn which genetic components serve a role in regulating those mechanisms. I believe that ZNF765 and/or the retroelements that are regulated by ZNF765 are such components that play an important developmental role. I want to show that they help define naïve or primed embryonic stem phenotypes. I plan to generate datasets of the transcriptomes of prime and naïve cells in the presence and absence of ZNF765 to see if naïve-associated genes, prime-associated genes, or retroelements are up-regulated or down-regulated when the expression of ZNF765 is altered. To accomplish this, I have shown that I can create conditions where ZNF765 expression is greatly reduced through CRISPRi in prime stem cells. I have also verified that I can activate naïve-associated genes through a naïve conversion using feeder-free media. These findings will allow me or other Haussler lab members to verify the same conditions in naïve cells to generate transcriptome datasets that will let us determine whether ZNF765 and or the retroelements regulated by it play an early developmental regulatory role in prime and naïve cells.
Does the transcriptome of the naïve state lacking ZNF765 make the cell’s phenotypes "prime-like"?

GRAPHICAL ABSTRACT
Acknowledgements

I would first and foremost like to thank Dr. David Haussler for granting me the opportunity to work and learn in his lab. I also want to thank Sofie Salama for encouraging me to pursue this project and for guiding me throughout this journey. I would also want to give particular thanks to both Dr. Jason Fernandes and Kristof Tigyi for their incredible guidance and for sharing their wisdom with me when much patience was required. Moreover, I would like to thank the Howard Hughes Medical Institute and the Institute for Biology of Stem Cells (IBSC) at UC Santa Cruz who have provided me with the resources to work on this project, particularly Bari Holm Nazario from IBSC who introduced me to the Haussler lab. Furthermore, I want to thank the Research Mentoring Internship (RMI) and the RMI director Zia Isola for encouraging me to pursue research in my field. Lastly, I would like to thank my parents for without whom none of this would have been possible and for sacrificing so much in order for me to follow my dreams.
## Contents

1 Introduction  

2 Background  
   2.1 Retroelements  
   2.2 KRAB-Zinc Finger Proteins  
   2.3 Embryonic Stem Cell States  
   2.4 CRISPRi in Gen1Cs  
   2.5 KZNF regulation of LINE1  

3 ZNF765 knock-down in Gen1C prime cells  
   3.1 ZNF765g1 and ZNF765g2 Gen1Cs Dox induction  
   3.2 RNA Isolation and Purification  
   3.3 qPCR of ZNF765g1 and ZNF765g2 KD and WT conditions  
   3.4 Results  

4 Determining whether RSeT Feeder-Free Medium successfully converts ZNF765g2  
   Gen1Cs to a naïve state  
   4.1 Naïve conversion of ZNF765g2 Gen1C  
   4.2 qPCR of RSeT™ ZNF765g2 WT Gen1C  
   4.3 Results  

5 Future Work: Determining whether ZNF765 and/or retroelements make naïve  
   ZNF765 KD condition less naïve  
   5.1 qPCR ZNF765g2 KD and WT conditions in naïve cells  
   5.2 Poly(A)-selection  
   5.3 Library prep  
   5.4 Results  

6 Conclusion  

A RNA Isolation and Purification
List of Figures

1  Heatmap of differentially expressed transposable elements between prime and naïve embryonic cells .................. 3
2  Full retrotransposition cycle of L1 ........................................... 7
3  Formation of heterochromatin through Kap-1 recruitment .......... 8
4  Evolutionary arms race between KZNFs and retroelements ........... 9
5  Generating Gen1Cs to contain gRNA plasmid ........................... 12
6  ChIP-seq data of ZNF649, ZNF93, and ZNF765 on L1PA5 .......... 13
7  Dox-induction leads to repression of targeted KZNF gene .......... 16
8  ZNF765g1 and ZNF765g2 Target Sequences ............................ 17
9  GFP and mCherry fluorescence ............................................. 19
10 ZNF765 qPCR knock-downs .............................................. 20
11 Knock-downs using different ZNF765 primers ......................... 21
12 Naïve cell morphology ..................................................... 23
13 qPCR data of ZNF765g2 Gen1C RSeT conversion ....................... 26
14 Flowchart of transcriptome dataset generation ........................ 27
15 Bioanalyzer traces for prime 765g2 WT Library Prep samples ...... 30
16 Standard curve from ZNF765 prime libraries ........................... 31
17 Expected results from prime and naïve WT and KD dataset .......... 32

List of Tables

1  Expression Values of KZNFs Between Prime and Naïve States ....... 4
Expression Values of ZNF765, ZNF649, and ZNF93 Between Prime and Naïve States
1 Introduction

Although humans share many aspects of biology with other species, many key differences between humans and other mammals have been observed in embryonic development that may explain the evolution of complex human-specific traits. One of these differences is observed in early embryonic stem cell phenotypes in mice and human. An early mammalian embryo is unique because it can exist in a totipotent ground state that can give rise to extraembryonic cells, the ground state is also called the naïve state (Nichols and Smith, 2009). Early embryonic stem cells in both mice and human can be modeled in cell culture dishes using pluripotent stem cells. Pluripotent stem cells can be cultured in two states (primed and naïve) when derived from mice, whereas, until recently, pluripotent human stem cells could only be cultured in a single state (primed), suggesting key biological differences between mouse and human naïve states (Tesar et al., 2007; Nichols and Smith, 2009). Mouse embryonic stem cells (mESCs) are taken from the inner cell mass of pre-implantation embryos. These cells are defined by the naïve state where they are able to self-renew and give rise to the germ layer as well as extraembryonic components (Evans and Kaufman, 1981; Nichols and Smith, 2009). However, mouse epiblast stem cells (mEpiSCs) are derived from a post-implantation stage and are classified as pluripotent stem cells which can only produce the germ layer (Nichols and Smith, 2009).

In comparison, human embryonic stem cells (hESCs) and human induced pluripotent stem cells (iPSCs) resemble the primed state despite being derived from pre-implantation embryos; although, hESCs and iPSCs recent advances have identified new culture conditions that that allows hESCs to enter a naïve state(Theunissen et al., 2016). By being able to study human cells in a naïve state, we can begin to understand the mechanisms and components that play a role in early human development that cannot be modeled in mouse systems. These studies could further help us understand genetic diseases that arise when those mechanisms and components fail to function properly.

Transposable elements (TEs) are genetic elements that are able to duplicate themselves and reintegrate back into the genome. One way to characterize naïve hESCs is through their unique TE expression patterns. Figure 1 shows that there are 10,000 TEs that are dif-
ferentially expressed between prime hESCs and naïve hESCs that are derived in different naïve media growth factors indicating that TEs have a regulatory role in early embryonic cell development. Retroelements are a type of transposable element (TE) which make up at least 45% of the human genome (Ayarpadikannan et al., 2014). Retroelements are capable of mobilizing and inserting themselves at random locations of the genome. When transcriptionally active, one known method in which intrinsic human retroelements can duplicate and reintegrate themselves into the genome is by using reverse transcriptase and integrase through the proteins encoded by LINE retroelements (Ayarpadikannan et al., 2014). The random integrations of retroelements can result in mutations that interfere with important genetic sequences. Such disruptions have been linked to genetic diseases (Hancks Kazazian, 2016). To combat the potential harm of retroelements to the host, KZNFs regulate their expression by repressing them. KZNFs target specific sequences of retroelements. Once bound, KZNFs recruit KAP-1 which assembles a repressor complex and leads to the formation of heterochromatin (Ecco et al., 2017). Heterochromatin formation tightly packages the DNA, making it less accessible for transcription factors which results in that segment of DNA being minimally transcribed. This repression puts selective pressure on the retroelements to mutate in order to escape the KZNFs. KZNFs must then develop a way to keep up with these mutational drifts, resulting in an evolutionary arms race between retroelements and KZNFs (Jacobs et al., 2014).

An evolutionary battle between retroelements and KZNFs could contribute to the regulation of development in cells. The evolutionary battle has resulted in the repurposing of endogenous retroelements such as the regulatory roles that they play that affects cell fate and development. For instance, when a particular LINE-1 family that is found to be highly expressed in mice zygotes is inhibited, it caused the zygote to arrest at extremely early stages of the zygote division (Sciamanna et al., 2011). This means that that LINE-1 family has a significant developmental role in ensuring that the mouse zygote properly divides. Furthermore, the regulation of retroelements through KZNFs is important early in development because mutations that result from new retroelement insertions can incorporate into the germline and be propagated. Analyzed data that shows the expression of KZNFs between primed and naïve H9 human embryonic stem cells (hESCs) which can
This heatmap consists of the expression of 10,000 transposable elements where red represents overexpression and blue represents underexpression of the genetic elements while yellow signifies no change in expression. Between prime samples and naïve samples grown in different media conditions, 12i/L/DOX+RI, 12/L/DOX+RI, 5i/L/A, and 4i/L/A, there is a significant group of transposable elements that is more highly expressed in the naïve condition than in the prime condition and vise versa.

be observed in Table 1 suggests that KZNFs are differentially expressed between both stem cell states.

I have decided to focus on the role that ZNF765 might play in maintaining stem cell states. This particular KZNF is of interest based on ChIP-seq data we have analyzed (Imbeault et al., 2017; Theunissen et al., 2016). The data shows KAP-1 recruitment by ZNF765 on LINE-1 primate specific 5 (L1PA5) retroelements in a naïve state only. This could indicate that the differential binding between ZNF765 and L1PA5 results in phenotypic differences between both stem cell states. My goal was to develop a platform where I can use CRISPR-based technology to observe the significance of ZNF765 and L1PA5 in determining stem cell states. I wish to employ this platform to create datasets of tran-
### Table 1: Expression Values of KZNFs Between Prime and Naïve States

This table shows the top 5 differentially expressed KZNFs in prime vs naïve H9s based on the highest log2 fold change value between both states.

<table>
<thead>
<tr>
<th>KZNF</th>
<th>Mean Naïve Transcript Expression Value</th>
<th>Mean Prime Transcript Expression Value</th>
<th>Log2FoldChange</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF208</td>
<td>2349</td>
<td>1</td>
<td>-9.93</td>
</tr>
<tr>
<td>ZNF729</td>
<td>8285</td>
<td>11</td>
<td>-9.37</td>
</tr>
<tr>
<td>ZNF650</td>
<td>1046</td>
<td>1</td>
<td>-9.26</td>
</tr>
<tr>
<td>ZNF728</td>
<td>958</td>
<td>1</td>
<td>-9.15</td>
</tr>
<tr>
<td>ZNF257</td>
<td>22977</td>
<td>3</td>
<td>-9.08</td>
</tr>
</tbody>
</table>

scriptomes between prime and naïve stem cell states by using CRISPRi to knock-down the ZNF765 gene in a Gen1C iPSC cell line in both prime and naïve states. I also wish to set these different ZNF765 expression conditions in the prime and naïve cells because by analyzing the transcriptomes of prime and naïve cells when ZNF765 expression is being altered, I will be able to observe which prime-associated and naïve-associated genes as well as retrotransposable elements are being affected by the alteration. Since ZNF765 is expressed nearly twice as much in the naïve state, it is expected that the less ZNF765 that is present in the naïve condition causes those cells to more closely resemble the prime condition phenotypes. The phenotypic differences however could potentially be more directly affected by the retrotransposable elements that are regulated by ZNF765 since retroelements have been shown to have regulatory developmental roles in mice. For this reason, observing which retroelements are differentially expressed between the prime and naïve state in the presence and absence of ZNF765 would be significant. Based on the comparisons, I will determine whether differential expression of ZNF765 in prime and naïve cells plays a role in the phenotypic differences or of retroelement regulation between both states and whether this plays a significant role in early development. Furthermore, the methods developed in this study can also be applied to other KZNFs that are of interest, such as KZNFs with higher differential expression values or other KZNFs that bind to L1PAs to observe how KZNFs in general might serve a developmental regu-
latory role.
2 Background

2.1 Retroelements

Transposable Elements (TEs), or “jumping genes”, are segments of DNA that can mobilize throughout the genome (Pray, 2008). TEs are classified under two main categories, class I and class II. Class I elements are able to mobilize through an RNA transcript intermediate, hence their name, retrotransposable elements. However, for class II TEs, the element does not require an RNA intermediate to transpose. Focusing on class I TEs, TEs have either autonomous or non-autonomous properties where non-autonomous TEs are dependent on the genetic compositions of autonomous TEs (Wessler and Susan, 2006). Autonomous retroelements contain the necessary machinery, such as reverse transcriptase and integrase, that is required to complete a full retrotransposition cycle; whereas, non-autonomous retroelements take advantage of the machinery found in autonomous retroelements to complete a retrotransposition cycle. Retroelements can further be classified into Long Terminal Repeat (LTR) retrotransposons and non-LTR retrotransposons. The abundance and activity of TEs varies throughout species and these properties can be linked to divergence of species that have high similarities within their genomes. For instance, human and chimpanzee genomes share 96% of genetic sequences and the variation between both genomes is largely caused by the difference in retroelements and retroelement expression (Wessler and Susan, 2006; Waterson et al., 2005). This indicates that retroelements could be responsible for differences in phenotypes between closely related species.

Long Interspersed Nuclear Elements-1 (L1s) are active, autonomous, non-LTR retrotransposons that are present in the human genome. (Kazazian and Moran, 1998). More than 45% of the human genome is composed of TEs and 17% of them are L1 retroelements (Cordaux and Batzer, 2009). There are over 500,000 copies of L1 in the human genome due to their continued activity. Figure 2 shows a full retrotransposition cycle of an active L1 retroelement in which it is transcribed and reintegrated through proteins encoded by the L1 element itself. Insertions of active L1 elements have been identified as causal agents in Hemophilia A as well as other genetic diseases such as cancer. The mutations caused by
L1 in Hemophilia A mostly occur in the germline or in early development (Kazazian and Moran, 1998). Additionally, LINEs and other retroelements have high GC content, which is prone to DNA methylation and therefore has the potential to silence genes that are in close proximity to the retroelements (Reyad et al., 2016). Consequently, the regulation of retroelements is crucial during development.

Figure 2: Full retrotransposition cycle of L1

This highlights the steps required and encoded by L1 to produce a full retrotransposition cycle when transcriptionally active. After transcription, the retroelement goes through an RNA intermediate, which then uses reverse transcriptase and integrase to reintroduce itself at random throughout the genome.

2.2 KRAB-Zinc Finger Proteins

Krüppel-associated box domain (KRAB) zinc finger proteins (KZNFs) make up the largest family of human transcription regulators and are responsible for repressing many families of retroelements (Imbeault et al., 2017). KZNFs contain an N-terminal KRAB domain followed by zinc finger domains. The KRAB domain is responsible for recruiting KRAB-associated protein-1 (Kap-1) while the zinc finger domains are able to recognize and bind to retroelements (Bellefroid et al., 1991; Huntley et al., 2006). Zinc finger domains bind specific nucleic acid sequences which can correspond to specific retroelement recognition motifs, (DNA patterns that serve a biological function), and once the KZNF is bound, the KRAB domain is able to recruit Kap-1. (Wolf and Goff, 2009). Kap-1 recruits other transcription factors which promote the formation of heterochromatin
through changes of histone modification at the bound sites (Iyengar and Farnham, 2011). Furthermore, the formation of heterochromatin results in the repression of gene transcription by tightly packing the DNA so transcription factors are less able to access and transcribe the condensed sequences (Sripathy et al., 2006). Figure 3 shows how KZNFs repress the transcription of retroelements by recognizing a retrotroelement and recruiting Kap-1 and its associated factors that modify both histones and DNA to generate heterochromatin.

![Diagram of KZNF, Kap-1, and DNA modifications](image)

**Figure 3: Formation of heterochromatin through Kap-1 recruitment**

*When a KZNF recognizes a retroelement, it binds to it, recruits Kap 1, and promotes the formation of heterochromatin by packing the DNA through histone modifications. The blue circles represent DNA methylation and the red circles represent histone methylation (H3K9me3) which are associated with heterochromatin formation. Adapted from Ecco et al., 2017.*

Although KZNFs are able to repress the transcription and mobilization of retroelements, some retroelements remain active within the human genome which means that active retroelements have escaped the binding of particular KZNFs. Retroelements are able to mutate the recognition sites bound by KZNFs either gradually through point mutations or all at once through large deletions (Jacobs et al., 2014; Fernandes et al., 2018). In order to prevent retroelements from wreaking havoc throughout the genome, KZNFs must find a way to regulate the propagation of retroelements. Figure 4 shows how a retroelement that is initially bound by a KZNF mutates its recognition site and escapes the binding of the KZNF. It then shows how a new KZNF must then repress the mutated
retroelement to keep up with its uncontrolled spread. This results in a constant battle between repression and escape and this phenomena is described as an evolutionary arms race (Jacobs et al., 2014).

The escape of retroelements from KZNF can be observed as mutations within the retroelement occur. KZNFs then experience selective pressure to control the mutational drifts and the battle continues (Provided by Dr. Jason Fernandes).

The regulation of retroelements is particularly significant in embryonic stem cells because retroelements that mobilize and propagate at this stage are passed on to the germline and consequently can become fixed in the genome. As previously mentioned, this uncontrolled propagation can result in genetic diseases, some of which occur during early development. This indicates that if propagation of retroelements occurs in embryonic stem cells, the retroelements could potentially cause developmental diseases. Therefore, KZNFs must play an important role in early development.

Figure 4: Evolutionary arms race between KZNFs and retroelements

The escape of retroelements from KZNF can be observed as mutations within the retroelement occur. KZNFs then experience selective pressure to control the mutational drifts and the battle continues (Provided by Dr. Jason Fernandes).
2.3 Embryonic Stem Cell States

Embryonic stem cells are undifferentiated cells and are found in two different states, a naïve state which goes through a transitional phase to result in a prime state. This progression occurs before lineage commitment of somatic cells, or fully differentiated cells (Nichols and Smith, 2009). Prime cells resemble post-implantation epiblasts, or the inner cell mass of an embryo, that are pluripotent and are able to produce all three of the germ layers in early development of an embryo (Nichols and Smith, 2009). The germ layers are made up of the ectoderm, mesoderm, and endoderm which produce the embryo. However, naïve embryonic stem cells are able to form extraembryonic and embryonic lineages and are found in a preimplantation state of the epiblast (Nichols and Smith, 2009). This means that naïve cells are totipotent and are able to create placenta as well as the three germ layers.

These naïve cells, or extraembryonic cells, allow the embryo to survive in the post-implantation uterine environment. Human-induced pluripotent stem cells (iPSCs) are derived from adult tissue and can be reverted back to a prime-like state by the transient introduction of reprogramming transcription factors and selection for embryonic stem cell-like properties (Takahashi and Yamanka, 2006). Using RSeT™ Feeder-Free Medium (Stemcell Technologies, Cat 05970), prime iPSCs and embryonic stem cells (ESCs) can be converted to a naïve state, which makes iPSCs or ESCs ideal candidates for studying different mechanisms and properties between prime and naïve embryonic stem cells.

2.4 CRISPRi in Gen1Cs

Clustered regularly interspaced short palindromic repeats interference (CRISPRi) can be used to study loss of function of targeted genes of interest in iPSCs with specificity (Mandegar et al. 2016). This mechanism differs from CRISPR nuclease (CRISPRn) because it contains a catalytically dead CRISPR associated protein (dCas9). CRISPRn can result in in-frame insertion/deletions (INDELs) that can have negative consequences within the genome. Due to the Cas9 nuclease activity being deactivated, INDELs can be avoided in CRISPRi (Qi et al., 2013; Shi et al., 2015). In the presence of a guide RNA, dCas9 sta-
bly localizes to a specific site of the genome and can inhibit transcription initiation near its guide RNA-specified binding site. The dCas9 was also fused with a KRAB domain to create a more efficient transcriptional interference since KRAB is a well-known transcriptional repressor (Gilbert et al., 2014). The dCas9-KRAB is guided to the targeted gene through a guide RNA (gRNA) which then represses that gene creating a knock-down of the gene (Mali et al., 2013; Mandegar et al., 2016).

Gen1Cs are a stable cell line which contain inducible CRISPRi properties by having a dCas9-KRAB, mCherry, (a fluorescence gene used for detection of dCas9-KRAB), Tetacycline operator (TetO) promoter cassette (Mandegar et al., 2016). When Doxycycline (Dox) is added to the Gen1Cs, the TetO promoter becomes active and the dCas9-KRAB and mCherry are expressed which gives the CRISPRi system reversible properties. The Gen1Cs were also transduced with a construct containing an mU6 promoter, (bound by RNA pol III and used for shorter RNA products), a KZNF guide RNA sequence, a green fluorescence protein sequence (GFP), and zeocin resistance. Figure 5 shows how the Gen1Cs containing the dCas9-KRAB cassette were introduced to the gRNA plasmids. The gRNA plasmid genes are constantly expressed and when dCas9-KRAB is present (+Dox), they form a complex that binds to the promoter of the targeted KZNF gene and represses it. Through this process, loss of function of ZNF765 can be studied in iPSCs.

2.5 KZNF regulation of LINE1

Previous studies have focused on the KZNFs that bind to primate specific LINE1 retroelements (L1PA) because they are the only autonomous, active retrotransposons in humans. Retroelements spread in bursts which give rise to newer retroelements. One such example are the LIPA6 retroelements that are shared by primates. When L1PA6s that are found in all primates combated KZNF escape in the evolutionary arms race, newer versions of the retroelements emerged since shared retroelements within a species mutate and evolve within those species or within one specific species, such as L1PA5 which is only found in more closely related, younger primate families (Jacobs et al., 2014). Chromatin Immunoprecipitation Sequencing (ChIP-seq) is used to detect protein binding to
Figure 5: Generating Gen1Cs to contain gRNA plasmid

This figure shows how the Gen1Cs that contain the Dox-inducible, dCas9-KRAB, mCherry cassette were transduced with the mU6 promoter, gRNA targeting KZNF, GFP, and zeocin resistance plasmid so that the plasmid could randomly integrate into the genome.

specific genomic locations for transcription factors, such as KZNF or Kap-1. Determining which KZNFs may regulate L1PA was done by analyzing ChIP-seq data that was previously collected (Imbeault et al., 2017, Lambert et al., 2016).

The ChIP-seq data revealed binding of ZNF93, ZNF649, and ZNF765 to motifs near
the transcription start site on a consensus L1PA5 sequence. It was also done by observing ChIP-seq of Kap-1 binding to L1PA5 genetic locations that line up with the binding of the three KZNFs previously stated (Jacobs et al., 2014). The binding of the ZNF649, ZNF93, ZNF765, and Kap-1 through ChIP-seq can be observed in Figure 6. Since Kap-1 recruitment for ZNF765 binding locations is lost in a prime state, it could potentially mean that ZNF765 is differentially expressed and/or differentially active between the prime and naïve states. The data on Table 2 shows that ZNF765 is indeed being differentially expressed between H9, (a human embryonic stem cell line), prime states and H9 naïve states where it is more highly expressed in the naïve state based on its log2foldchange. The -1.77 log2foldchange value for ZNF765 means that the expression of ZNF765 is nearly doubled in naïve cells. Despite also seeing differential expression for ZNF649 and ZNF93, ZNF765 remains of interest since ZNF649 and ZNF93 do not show differential recruitment of Kap-1 to L1PA5 and also have higher expression levels than ZNF765 in the prime state. The differential expression levels with the ChIP-seq data could indicate that ZNF765 and/or the retroelements being regulated by ZNF765 are responsible for determining embryonic stem cell phenotypes. For this reason, ZNF765 was chosen to be studied.

Figure 6: ChIP-seq data of ZNF649, ZNF93, and ZNF765 on L1PA5

This figure displays analysis of ChIP-seq data that shows binding locations on the L1PA5 retroelement sequence. It shows ZNF649 and ZNF93 binding and Kap-1 recruitment in both naïve and primed phenotypes of cells. It also shows a loss of Kap-1 recruitment to the ZNF765 binding site in primed phenotypes of cells. Source: Raney et al. (2014) Bioinformatics. 30(7):1003-5.
ZNF765, ZNF649, and ZNF93 Differential Expression between Prime and Naïve H9s

<table>
<thead>
<tr>
<th>KZNF</th>
<th>Mean Naïve Transcript Expression Value</th>
<th>Mean Prime Transcript Expression Value</th>
<th>Log2FoldChange</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZNF765</td>
<td>2612</td>
<td>763</td>
<td>-1.77</td>
</tr>
<tr>
<td>ZNF649</td>
<td>6027</td>
<td>1701.263</td>
<td>-1.8197697</td>
</tr>
<tr>
<td>ZNF93</td>
<td>2605.019</td>
<td>1248.598</td>
<td>-1.0552363</td>
</tr>
</tbody>
</table>

Table 2: Expression Values of ZNF765, ZNF649, and ZNF93 Between Prime and Naïve States

This table shows the expression of ZNF765, ZNF649, and ZNF93 in prime vs naïve H9s based on the log2 fold change value between both states. From this data, it is apparent that there is higher expression of all three KZNFs in naïve cells.
3 ZNF765 knock-down in Gen1C prime cells

Gen1C is a cell line that can be used to study KZNFs and retroelements in a human iPSC context due to its CRISPRi properties that allow KZNF genes of interest to be knocked-down. Through a knock-down, the transcriptome can be studied in the presence and absence of the targeted KZNF, such as retroelement mRNA that is produced when a retroelement becomes transcriptionally active in the absence of a KZNF that represses the retroelement. To determine whether ZNF765 and/or retroelements play a role in defining embryonic stem cell phenotypes, I had to establish a platform in which I could perform a successful ZNF765 knock-down in prime cells. In order for a successful knock-down to occur, the dCas9-KRAB complex produced when the Dox antibiotic is added, (knock-down or KD condition), must bind to the scaffold of the gRNA, which has to bind with specificity to the targeted gene. The mechanism in which Dox-induction represses or knocks-down the targeted ZNF765 gene in Gen1Cs can be observed in Figure 7 where the Tet Response Element (TRE) promoter, which is typically inhibited and repressed by a tetracycline repressor, becomes transcriptionally active so KRAB-dCas9 and mCherry are produced. The KRAB-dCas9 is then able to bind to the KZNF gRNA which is constantly being expressed along with GFP and bound to the targeted KZNF gene. The protein complex formed then represses or knocks-down the targeted KZNF gene. Furthermore, when Dox is added, the cells fluoresce green and red due to the GFP and mCherry production as opposed to wild-type (WT) Gen1Cs which only fluoresce green since only GFP is produced.

The David Haussler wet lab has in-house stocks of Gen1Cs expressing gRNAs that were designed to target various KZNF genes. For ZNF765, there are guide 1 (g1) and guide 2 (g2) stocks. The guides are significant in the CRISPRi system because they have to be unique to the ZNF765 gene to avoid off-target effects that inadvertently knock-down other genes. Furthermore, the plasmid containing the guide RNA integrates itself at random in the genome when transduced, which could potentially interfere with a successful knock-down or other necessary cellular activity by disrupting essential genes. Figure 8 demonstrates the sequences that ZNF765g1 and ZNF765g2 target. These sequences are
Figure 7: Dox-induction leads to repression of targeted KZNF gene

This figure shows how when Dox is added, it binds to the inhibitor of the Tet Response Element (TRE) promoter so that KRAB-dCas9 and mCherry (red fluorescence) can be produced. The gRNA and GFP (green fluorescence) are always transcriptionally active, so after Dox-induction, the KRAB-dCas9 is able to repress or knock-down the targeted gene by forming a repressor complex with the KZNF guide that is bound to the targeted gene. Zeocin is added to select for cells that produce the gRNA.

found upstream and near the ZNF765 gene transcription site on chromosome 19. The ZNF765g1 target sequence, (highlighted in blue) lies on the template strand while the ZNF765g2 target sequence, (highlighted in purple), lies on the non-template strand. To establish a ZNF765 knock-down platform in prime cells, I first had to verify that the in-house stocks of ZNF765g1 and ZNF765g2 resulted in an 85-95% change in expression levels of ZNF765.

3.1 ZNF765g1 and ZNF765g2 Gen1Cs Dox induction

The ZNF765g1 and ZNF765g2 in-house stocks were thawed and passaged, (transferred and expanded to a new plates), a minimum of two times before Dox induction. The cells were grown and maintained with mTeSR\textsuperscript{TM}1 medium (Stemcell Technologies, Cat 85850) and stored in an incubator maintained at 37° C with 5% CO2 to grow. I added Zeocin (50µg/mL) to the mTeSR\textsuperscript{TM}1 medium to select against cells that lacked the gRNA plasmid, which contains Zeocin resistance. I plated ZNF765g1 and ZNF765g2 Gen1Cs
Figure 8: ZNF765g1 and ZNF765g2 Target Sequences

UCSC genome browser image showing the first exon (dark blue rectangle) of ZNF765 and part of the first intron (dark blue arrows). Both ZNF765g1 and ZNF765g2 lie near the transcription site of the ZNF765 gene on chromosome 19. ZNF765g1 (5’ AACAGAGCGAAACTCACCCG 3’) is highlighted in light blue and ZNF765g2 (5’ GAGTGACTATCCCACCGCCG 3’) is highlighted in purple and respectively lie on the template and non-template strands.

Each on two 6cm petri dishes to set up a knock-down (KD) and a wild-type (WT) condition for each guide. I then performed Dox-induction by adding Dox (1μg/mL) to one of the two plates for ZNF765g1 and ZNF765g2 to produce dCas9-KRAB and mCherry. To ensure that dCas9-KRAB and gRNA were being produced, I used a live cell imaging fluorescence scope to detect GFP and mCherry production using GFP and TexasRed/TRITC filters, respectively.

3.2 RNA Isolation and Purification

After 48-36 hrs of Dox induction, I extracted and isolated the RNA for each KD and WT condition of the ZNF765g1 and ZNF765g2 prime cells using the Direct-zol RNA kit TRIzol reagent from Zymo Research (Zymo Research, Cat R2052) which lyses cells without degrading RNA. I then performed RNA purification as per the adapted RNA Clean and Concentrator-5 protocol (Zymo Research, Cat R1013). Protocol details for RNA isolation and purification can be found on Appendix A.

3.3 qPCR of ZNF765g1 and ZNF765g2 KD and WT conditions

Once the RNA was isolated and purified, I converted each of the samples to cDNA using the iScript cDNA synthesis Kit (BioRad, Cat 1708890), protocol found on Appendix B, in order to stabilize the RNA to perform quantitative polymerase chain reaction (qPCR) on a QuantStudio 6 Flex Real-Time PCR system. To quantify the amount of ZNF765
present in each sample, I used a QuantiTect SYBR Green PCR Kit (Qiagen, Cat 204141) to detect extension of a ZNF765 template through SYBR Green which fluoresces when a nucleotide is added to a ZNF765 DNA strand. Adapted protocol for SYBR Green PCR kit can be found in Appendix C.

3.4 Results

Through the live cell imaging obtained after Dox induction, I expected to see green fluorescence both in the KD and WT conditions for ZNF765g1 and ZNF765g2 since the gRNA should constantly be produced by the cells. Furthermore, I expected to see mCherry production in only the KD samples since mCherry is indicative of dCas9-KRAB production. Figure 9 depicts the results obtained from the live cell imaging which shows that mCherry and therefore dCas9-KRAB were being expressed in the KD samples for ZNF765g1 and ZNF765g2 while the GFP and gRNA were being produced in all samples. However, the intensity of the mCherry is much lower in the ZNF765g1 samples than the intensity observed in the ZNF765g2 samples. A partial reason for observing the low intensity is due to the overlaying of the images when initially collected. For this reason, I stopped overlaying the pictures when imaging for the ZNF765g2 samples. However, the difference in intensities is overwhelming and cannot solely be explained by the image overlay. The ZNF765g1 KD condition fluorescence also indicates that the sample is producing very low amounts of mCherry and therefore dCas9.

In order to properly confirm that a successful knock-down has been performed, the expression levels of ZNF765 were quantified through qPCR. A housekeeper gene, a gene that is active in all tissues, B2M was used to compare the ZNF765 expression levels in the qPCR results. When ZNF765 is knocked-down, I expected the expression level of ZNF765 of the KD condition compared to the WT condition to be lower since the knock-down should have less ZNF765. From the qPCR results shown in Figure 10, I determined that ZNF765g1 does not perform an adequate knock-down since the ZNF765 expression levels compared to -Dox or WT were the same for the KD condition. The low red fluorescence intensity seen in Figure 9 for the 765g1 KD condition sample correlates with the
unsuccessful knock-down observed from the qPCR results. However, for ZNF765g2, the ZNF765 expression levels compared to WT show a successful knock-down with an 85% decrease of ZNF765g2 in the KD condition compared to the ZNF765g2 WT condition.

Figure 9: GFP and mCherry fluorescence
This image, pictured at a 10X magnification, shows how GFP is produced in all of the samples for ZNF765g1 and ZNF765g2 at similar intensities. It also shows how the KD conditions displays TexasRed/TRITC fluorescence which is red. The 765g1 samples show much lower red fluorescence compared to 765g2 samples which is partially due to the channels being overlayed.
Overall, I performed qPCR on the ZNF765g2 Gen1Cs a total of three times, where I was able to detect inconsistencies between the ZNF765g2 knock-downs. The first time I performed qPCR was to confirm that the knock-down was successful. The second time that the knock-down was performed was for the samples that I gathered to generate the RNAseq datasets and the third time was for the samples that I wanted to use for biological replicated to show that the results are reproducible. Each time showed high variation in the results for the knock-down. I initially believed that this was caused by improper design of the 765 primer that I was using for the qPCR since I found that it maps to the two flanking KZNF genes of ZNF765. For this reason I decided to repeat the experiment using three different primers. The results of that qPCR can be seen on Figure 11 which shows the knock-down results of the three ZNF765 Gen1C samples using three different ZNF765 primers. The second primer was the original primer that was used and the first and third one was a primer that does not map to any other genes in the human genome. Each primer showed between a 60-98% knock-down between all the samples. However, the results were inconclusive and the experiment will have to be repeated because I used a ZNF649 Gen1C sample as a control which failed. The ZNF649 sample has been confirmed to have a 98% knock-down through RNAseq. When using it as a control, I got about a 54% knock-down which means that the results are not credible. Since the control failed, I am unable to determine whether the inconsistencies are technical or biological and have to
repeat the experiment. Although, the varying results that I have seen between repetitions of the qPCR using the same primers and samples indicates that the variation is caused by technical errors. Given that it is a biological error, I will select clones that show strong knock-downs and grow a new line from those clones to use for my experiments.

**Figure 11: Knock-downs using different ZNF765 primers**

*This figure shows the knock-downs of three different ZNF765g2 Gen1Cs. They show between a 60-98% knock-down throughout all the samples. In addition, it shows the knock-down of a control sample, ZNF649, which resulted in about a 54% knock-down.*

4 Determining whether RSeT Feeder-Free Medium successfully converts ZNF765g2 Gen1Cs to a naïve state

Mouse Embryonic Fibroblast (MEF) feeder cells are often used in cell culture to grow human iPSCs because they secrete growth factors that allow stem cells to grow while maintaining their pluripotency. Furthermore, most published protocols include MEFs
in the growth conditions used to convert prime iPSCs to a naïve state (Theunissen et al., 2016). This method of embryonic stem cell conversion is not ideal because of its long culturing time (about 6 months) and because of the MEF contamination (mouse RNA) that results when isolating RNA for experiments. However, the RSeT™ Feeder-Free Medium (Stemcell Technologies, Cat 05970), successfully converts stem cells from a prime state to a naïve state without the use of feeders within a week (Hunter et al., STEMcell Technologies, Inc., n.d.).

In order to compare the transcriptomes of prime and naïve cells when ZNF765 is differentially expressed, I converted the prime ZNF765g2 Gen1Cs to a naïve state in order to be able to establish a knock-down of ZNF765 in naïve cells that are derived from the prime cell line that I am using for the ZNF765 knock-downs. This will allow me to compare the transcriptomes of prime and naïve cells in the presence and absence of ZNF765. This means that in addition to establishing a ZNF765 knock-down platform in prime 765g2 Gen1Cs, I also have to establish a knock-down of ZNF765 in naïve ZNF765g2 Gen1Cs. To achieve this, I first had to confirm that a successful naïve conversion was performed using the feeder-free RSeT media. After verifying that the naïve conversion was successful, I will have to ensure that a ZNF765g2 knock-down is also successful in the converted cells.

4.1 Naïve conversion of ZNF765g2 Gen1C

I initiated conversion of the WT condition of ZNF765g2 prime Gen1Cs to a WT condition of ZNF765g2 naïve Gen1Cs by adding RSeT™ Feeder Free Medium (StemCell Technologies, Cat 05975) to the prime Gen1Cs. In order to select for cells expressing the gRNA, 50µg/mL of Zeocin were added to the RSeT™ Medium. Successful conversion to a naïve state can be initially observed within one to two days after adding RSeT™ through a brightfield microscope since naïve cells show a “dome-like” growth pattern while prime cells display a flatter morphology. Figure 12 shows how the cells that were fed with RSeT™ Medium exhibit a “dome-like” growth compared to the flat behavior found in prime cells. They also grow in smaller colonies or in an elongated fashion instead of the
round-like growth exhibited by the Gen1Cs grown in mTeSR. Before harvesting the cells with the RSeT™ Medium, they were passaged five times. Figure 12 also shows the morphology of the cells in RSeT before harvesting after the fifth passage. They show less heterogeneity in “dome-like” behavior but continue to exhibit an elongated-like growth pattern. After I completed the appropriate amount of passages and isolated the RNA, I had to confirm that the naïve conversion was successful by performing qPCR on the naïve cells to observe whether genes that are expressed in a naïve state only were being expressed in the cells that were grown in RSeT™ Medium.

![Figure 12: Naïve cell morphology](image)

These images show the “dome-like” or “halo-like” naïve morphology that is being exhibited by ZNF765g2 Gen1Cs two days after adding RSeT™ Medium and shows how the morphology compares to that of cells grown in mTeSR where flatter colonies grow. The RSeT colonies also appear to grow in an elongated form compared to the mTeSR round growth. These snapshot also depict the morphology of the 765g2 Gen1Cs that were harvested after being passaged five times in RSeT feeder-free media. The cells exhibit less “dome-like” behavior but still had a bright hue or halo around the edges, indicating that they are growing upward. They also continued to grow in an elongated fashion that differs from cells grown in mTeSR.

### 4.2 qPCR of RSeT™ ZNF765g2 WT Gen1C

In order to verify that the RSeT™ ZNF765g2 WT condition Gen1Cs were successfully converted to a naïve state, I isolated and purified RNA from ZNF765g2 Gen1C cultures after 5 passages in RSeT as per the protocol described in Appendix A using a Direct-Zol RNA kit and an RNA Clean and Concentrator-5 kit. I then converted the purified RNA to cDNA using the iScript cDNA synthesis kit and ran the cDNA in the QuantStudio 6 Flex
Real-Time PCR system to perform qPCR, protocols can be found in Appendices B and C. Differences between prime and naïve cells can be observed from differential gene expression. Based on the genes that when expressed are associated with naïve pluripotency and were used by StemCell Technologies to determine that their media can convert prime cells to naïve cells, I chose to look at the following genes: DPPA5, DPPA3, DNMT3L, KLF2, and KLF17 (Hunter et al., StemCell Technologies, n.d.).

4.3 Results

To determine whether the ZNF765g2 Gen1Cs were properly converted to a naïve state, higher naïve-associated gene expression levels were expected in the RSeT sample compared to the mTeSR sample. Furthermore, lower expressions of NANOG, SOX11, and ZIC2, (prime-associated genes that have shown lower or no change in expression levels in naïve MEF-grown H9s compared to prime H9s), were expected to be observed in the qPCR data. In qPCR, the $C_T$ value is a number that shows how many cycles were required for a cDNA sample to cross a threshold. Since the less cDNA molecules that are present in a sample require more cycles to cross the $C_T$ threshold, less cDNA molecules correlate to a higher $C_T$ value. So when qPCR is performed properly, the higher the the $C_T$ value is, the less that the targeted gene is being expressed in the samples.

In qPCR, because each cDNA molecule is doubled when a cycle passes, to get the amount of cDNA molecules that are present after the sample crosses the $C_T$ threshold, the exponential formula $2^n$ can be used where $n$ is the $C_T$ value. This is known as the Log2FoldChange. To get the Log2FoldChange values of the ZNF765g2 RSeT samples compared to the mTeSR sample, I used the $\Delta\Delta C_T$ method where I take my sampled $C_T$s and subtract them from the $C_T$ of an internal control housekeeper gene, or gene that is always being expressed at the same level in all cell tissues, to get a $\Delta C_T$ value. I then am able to get the $\Delta\Delta C_T$ by subtracting the experimental $\Delta C_T$ from the control $\Delta C_T$. Afterwards, I use this number in the Log2FoldChange formula to compare how much more a gene is being expressed in the experimental sample (RSeT/ naïve conversion) relative to the control sample (mTeSR/prime).
Figure 13 shows the data that was acquired from the naïve conversion experiment when using RSeT feeder-free media. From that data I was able to observe more expression of the naïve-associated genes KLF2, DPPA5, and DNMT3L in the ZNF765g2 RSeT (naïve conversion) sample compared to the mTeSR sample since the Log2FoldChange was a minimum of 8.77. In addition, NANOG expression remained unchanged between the two samples as expected. However, the prime-associated genes SOX11 and ZIC2 were also more highly expressed in the RSeT sample. The SOX11 Log2FoldChange data (74.8) is very unreliable because of how highly it indicates that it is being expressed in the RSeT sample. Nonetheless, this could indicate that the cells were converted to an intermediate step of a naïve/prime-like state or that half the cells that were harvested from one plate were in a prime state while the other half were in a naïve state.

While the results show some indication of naïve- and prime-associated gene expression levels between samples, I will have to conduct the experiment again because the results have high error in the computation. The higher error is due to the large $C_T$ values that were measured. The closer that a $C_T$ value is to the amount of cycles that are run, the higher that the error in $C_T$ measured is since it relies on the detection of very few samples. Many of the mTeSR samples had an undefined $C_T$ meaning that not enough cycles were run to detect how many cDNA molecules of that sample were present. This means that the cDNA molecules added to the reaction mix were either very low or not present all. To calculate for those samples I set their $C_T$ value equal to the amount of reactions that I ran on the experiment (40) to simulate for the low presence of the targeted genes. For this reason, I will repeat the experiment to measure the samples in a more visible range by adding more cDNA of each sample. This will shift the $C_T$ to lower values, ideally between 15-25 cycles. This range will give me much more accurate results where the error decreases. Furthermore, I will repeat the experiment with a parental Gen1C line that contains no guide RNA to verify that the results of the genes that were more highly expressed in RSeT media are reproducible in another Gen1C line, making the results more reliable. If repetitions of the experiment show that the RSeT media does not successfully convert Gen1C to a naïve state, I will have to use MEFs to culture a naïve ZNF765g2 Gen1C line.
This figure displays the Log2FoldChange of the ZNF765g2 Gen1C RSeT samples (naïve conversion) compared to the mTeSR (prime) samples. This data shows that from the naïve-associated genes (KLF2, KLF17, DPPA3, DPPA5, and DNMT3L), all except KLF17 and DPPA3 express the naïve-associated genes by a minimum of 400 times more in the RSeT sample compared to the mTeSR sample. However, the data also shows that the prime-associated genes, SOX11 and ZIC2, are also more highly expressed in the RSeT sample compared to the mTeSR sample.

### 5 Future Work: Determining whether ZNF765 and/or retroelements make naïve ZNF765 KD condition less naïve

In order to determine whether ZNF765 and/or retroelements play a role in defining embryonic stem cell phenotypes, I have to compare the transcriptomes of both the WT and KD conditions of prime and naïve ZNF765g2 Gen1Cs. To do this, I also have to establish the knock-down in the ZNF765g2 Gen1C naïve cells. Then to compare the transcriptomes, I must perform RNAseq after a series of steps that are required to select for mRNA. Figure 14 shows the steps that must be followed to perform RNAseq. The steps in the figure require that once I have isolated the RNA for the WT and KD conditions
in the prime and naïve cells, I must poly(A)-select the isolated RNA and Library prep
the poly(A)-selected RNA. In order to poly(A)-select, I will use the NEXTflex™ Poly(A)
Beads (Bioo Scientific, Cat 512979). The beads are designed to isolate pure, intact mRNA
through magnetic based separation.

Once the mRNA is isolated, I will Library Prep the mRNA using the NEXTflex™
Rapid Directional qRNA-Seq™ Kit - Set A (Bioo Scientific, Cat 5130-02D). This kit pre-
pares conventional directional, strand-specific RNA libraries that can be sequenced by
Illumina. It also contains the Molecular Indexing™ feature which is used for labeling.
The directionality allows the Library prepped mRNA to be sequenced through the cDNA
strand. Furthermore, 96 distinct molecular labels, which provide a unique molecular
identification in the form of an 8 nucleotide barcode tag, are incorporated into the se-
quencing adapters (oligonucleotide that can be ligated to RNA molecules). The Molecu-
lar Indexing™ feature described in the NEXTflex™ Rapid Directional qRNA-Seq™ Kit
- Set A allow for absolute, digital measurements of gene expression levels. Once I have
Library Prepped the samples, I can send them to be RNAseq-ed.

After I have samples sequenced at UC Berkeley, I will then enlist the help of a bioin-
formatics specialist, Dr. Sol Katzman, to generate DEseq datasets, (datasets of the tran-

---

Figure 14: Flowchart of transcriptome dataset generation

This figure demonstrates the steps that are required to be able to analyze the transcriptomes of the ZNF765g2
prime and naïve Gen1Cs for the WT and KD conditions. After the conditions are set in the prime and naïve
states and the RNA is isolated, poly(A)-selection must be performed. Then the poly(A)-selected RNA can
be library prepped and RNAseq-ed to generate the transcriptome datasets that will be used for comparison.

27
scriptome which show expression levels of specific genes), of each of the conditions in each of the samples. These datasets will be used to compare the transcriptomes between the WT and KD conditions in the prime and naïve cells.

5.1 qPCR ZNF765g2 KD and WT conditions in naïve cells

After confirming that the naïve conversion was successful, I plan to set up the KD and WT conditions for the naïve ZNF765g2 Gen1Cs by adding Dox (1 µg/mL) and image the cells using the live cell imaging microscope. The guidelines for the ‘RNA isolation and purification’, ‘cDNA synthesis’, and ‘qPCR of ZNF765 Gen1C KD’ will be followed which can be found on Appendices A, B, and C to confirm that the ZNF765 knock-down was successful in the naïve ZNF765g2 Gen1Cs.

5.2 Poly(A)-selection

After performing the RNA isolation, I will use the NEXTflex™ Poly(A) Beads to isolate the mRNA through Poly(A)-selection for the WT and KD conditions of prime and naïve ZNF765g2 Gen1Cs. Since mRNA is biologically modified to contain a Poly(A)-tail, (a series of multiple Adenosine nucleotides added to the 3’ end of mRNA, to protect and stabilize the mRNA), it can be used to bind the mRNA to the magnetic beads found in the kit which contain oligo(T)-nucleotides, (a series of multiple Tyrosine nucleotides which bind to poly-Adenosine tail). The protocol for Poly(A)-selection can be found in the NEXTflex Poly(A) Beads handbook (Cat NOVA-512980). This step has already been performed on the ZNF765g2 WT and KD prime Gen1Cs.

5.3 Library prep

Once the mRNA is isolated, I will be able to use the NEXTflex™ Rapid Directional qRNA-Seq™ Kit - Set A to Library Prep each of the WT and KD conditions of prime and naïve ZNF765g2 Gen1Cs. The protocol for the Library Prep can be found on the NEXTflex Rapid Directional qRNA-Seq Kit (Cat NOVA-5130). This step has been performed on the ZNF765g2 WT and KD Gen1C samples. I will be sending the samples to QB3 Vincent J.
Coates Genomics Sequencing Laboratory at UC Berkeley where they will be RNAseq-ed with Illumina sequencers.

5.4 Results

In order to send the library preps to be sequenced by an Illumina sequencer, such as the HiSeq or MiSeq machines, the samples must pass some quality control (QC) tests. The first QC is performed by running the library prepped samples through a Bioanalyzer. A Bioanalyzer uses a microfluidics capillary-based electrophoresis method to detect DNA sample sizes. The DNA generates a ladder through capillaries based on their sizes which is driven electrophoretically. Since RNAseq data that is generated from Illumina sequencing requires that the sample RNA be fragmented, the Bioanalyzer shows that the fragmented sizes of the samples are within a required range. According to the Rita San-Bento Illumina Technical Support, an ideal trace has a flatline between the lower and upper markers which can be observed in Figure 15 (Rita San-Bento Illumina Technical Support, 2012). I was able to observe this trace mark on my prime ZNF765g2 WT and KD library prepped samples which can also be observed in Figure 15. The trace mark for those samples lie within 240-1800bp. Since the Bioanalyzer traces were satisfactory, I was able to continue with the Library Prep QC tests by running a KAPA quantification protocol.

For the KAPA quantification protocol, I had to run varying dilutions of each of my Library Prepped samples and had to compare them to dilutions of a set of standards through qPCR. The standard curve that was generated from the dilutions, observed in Figure 16, is composed of the concentration of the samples in pM and the $C_q$ values, in this case called the $C_q$, allowed me to calculate what the working concentration is for each of my samples in ng/µL. This calculation was generated by observing where each of the sample dilutions lie on the standard curve that was generated from the standard dilutions while factoring in the average sample size in bp of the samples. This is significant because in order to get proper results from the RNAseq, the samples must be sent at an equal concentration of 5nM. Since all the samples will be mixed in a pool, if samples
Figure 15: Bioanalyzer traces for prime 765g2 WT Library Prep samples

This data shows how an ideal Bioanalyzer Trace has a flatline between the lower and upper markers. From the Bioanalyzer results gathered from the prime 765g2 WT and KD Library Preps, an satisfactory bioanalyzer trace is detected for both samples. The fragments lie between 240 and 1800 bp.
were at different concentrations, then we would not be able to compare differential gene expression between samples.

Standard curve generated from prime ZNF765g2 WT and KD libraries

![Graph showing the standard curve with the equation y = -3.3545x + 12.961 and R² = 0.9985.](image)

**Figure 16: Standard curve from ZNF765 prime libraries**

This standard curve which from the linear regression model has a slope of -3.3545, y-intercept of 12.961, and R² of 0.9985, shows the average Cq based on the concentration of the standards. The values gathered from the linear regression model allowed me to calculate the working concentration of each library prep sample.

All of these experiments must also be performed on the naïve ZNF765g2 WT and KD library preps that will be generated. Once the QCs are met, the samples are able to be sent out to be RNAseqed at the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. The results will be sent to Dr. Sol Katzman, a bioinformatics specialist, who will generate the DEseq datasets of the prime and naïve ZNF765g2 KD and WT samples. From these datasets I will focus on naïve-associated genes, prime-associated genes, and retroelements that are either going to be up-regulated or down-regulated between the samples. I could potentially see retroelements or prime-associated genes whose expressions go up when ZNF765 is knocked-down in the prime datasets. I am also interested to see which retroelements and prime-associated genes are up-regulated when ZNF765 is knocked-down in the naïve samples. The dataset that I am most interested in is that of the naïve KD condition sample because if I see many up-regulated prime-associated genes in the KD naïve sample, I can determine that the phenotype of the naïve KD is "prime"-like, indicating that ZNF765 plays a significant role in early developmental cell regulation.
Figure 17 shows an image of how I expect to see the naïve KD sample phenotypes to make the cell resemble the prime sample phenotypes. Furthermore, if retroelements are also up-regulated in the naïve KD sample, I can surmise that the retroelements regulated by ZNF765 at this stage of an embryonic cell could potentially also have important regulatory roles during early development.

6 Conclusion

This project aimed to further our understanding of early human development. By better studying early human development, we can begin to delve deeper into what makes humans different from other species. Furthermore, by understanding the mechanisms that play a role in human development, we can start to explore how malfunctions of those mechanisms lead to genetic/developmental diseases and disorders.

Through the experiments conducted for this project, I sought to determine the role that ZNF765 and/or the retroelements that are regulated by that particular KZNF have in defining embryonic stem cell phenotypes. In large part, this consisted of testing two CRISPRi platforms to determine whether I could set up conditions where ZNF765 can be
successfully knocked-down in prime and naïve cells. To accomplish this, I used CRISPRi to repress the transcription of ZNF765 to set the KD condition in prime and naïve Gen1C iPSCs. I confirmed that the platform was successful in repressing ZNF765 in the prime cells by performing qPCR since the knock-downs showed at least a 50% repression but have to determine if inconsistencies between these knock-downs are caused by biological or technical errors. I was able to show that the RSeT feeder-free medium was successful in converting the prime ZNF765g2 Gen1Cs into a more naïve state by finding that naïve-associated genes, KLF2, DPPA5, and DNMT3L are more highly expressed in the RSeT grown ZNF765g2 Gen1Cs. However, I need to determine if prime-associated genes SOX11, and ZIC2 expressions are also more highly expressed in the RSeT grown samples. By being able to grow naïve ZNF765g2 Gen1Cs, I will be able to run the same knock-down qPCR experiments that were performed on the prime cells to see if I can perform a successful knock-down in the naïve-converted cells. Once I have established this, I will have to compare the transcriptomes of the different WT and KD conditions in prime and naïve Gen1Cs. To accomplish this, I will collect the total RNA of the KD and WT conditions of prime and naïve ZNF765g2 Gen1Cs and poly-(A) select to isolate the mRNA. I will be able to sequence the transcriptomes of the prime samples by library prepping the mRNA of each of the conditions on the isolated mRNA and send it to QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley to perform RNAseq through Illumina sequencing. When I complete this, I will be able to compare the transcriptomes of prime and naïve cells when ZNF765 is absent, and will look to see how the absence of ZNF765 affects retroelement expression and naïve-associated gene expressions. This comparison will allow me to determine how much of an impact ZNF765 has in defining embryonic stem cell phenotypes and therefore determine if ZNF765, or the retroelements regulated by ZNF765, has an early developmental role.
References


A RNA Isolation and Purification

**Purpose:** This experiment is conducted to isolate and purify the total RNA from a stem cell culture sample grown in a 6cm petri dish.

**Overview:** This kit allows for the preparation of and purification of RNA directly from the sample. It is able to do so cleanly by using Zymo-Spin™ IICR Columns which perform easy separations through a membrane.

**Materials:**
- cultured stem cells on 6cm dish
- DPBS, no calcium
- cell scraper
- 15mL conical tube
- centrifuge and microcentrifuge
- 1.5mL eppendorf tube
- Direct-zol™ RNA MiniPrep (Cat R2052)
- RNA Clean & Concentrator™-5 (R1015)

**Methods:**
Step 1 (Isolate total RNA in TRIzol):
(a) Take an 80%-85% confluent 6cm plate that is growing the desired cell culture
(b) Wash the plate two times with DPBS
(c) Add 2mL DPBS and scrape cells off with cell scraper. Make sure to get as many cells as possible.
(d) Collect the scraped cells with 3mL DPBS and place inside a 15mL conical tube.
(e) Spin down cells by placing the conical tube in a centrifuge machine. Spin at 300g for 3-5 min.
(f) Aspirate the media. Be careful to not disturb the cell pellet.

(g) Resuspend cell pellet with 1mL DPBS and transfer resuspended mixture to a 1.5mL eppendorf tube.

(e) Spin down cells by placing eppendorf tube in a centrifuge machine.

(f) Carefully aspirate media using a P-1000 micropipette

(g) Add an equal volume of TRIzol and 95%-100% ethanol. Resuspend the cell pellet first using the TRIzol and then add the ethanol. (Ex. Add 600µL TRIzol and 600µ ethanol)

At this point, the total RNA has been harvested and can be stored in -80°C or it can continue to be worked on.

Step 2 part I (Sample Preparation):

* Perform all steps at room temperature and centrifugation at 10,000-16,000 x g for 30 seconds, unless specified.

(a) Transfer 600µL of the mixture into a Zymo-Spin™ IICR Column in a Collection Tube and centrifuge. Repeat to collect all harvested RNA. Transfer the column into a new collection tube and discard the flowthrough.

(b) Add 400µL Direct-zol™ RNA PreWash to the column and centrifuge. Discard the flow-through and repeat this step.

(c) Add 700µL RNA Wash Buffer to the column and centrifuge for 2 minutes to ensure complete removal of the wash buffer. Transfer the column carefully into an RNase-free tube. Spin down again. Transfer the remaining liquid to the RNase-free tube.

(d) To elute RNA, add 50µL of DNase/RNase-Free Water directly to the column matrix, let sit for 5 min, and centrifuge.

RNA can continue to be used or it can be stored in -80°C

Step 2 part II (RNA purification):

Sources: Direct-zol™ RNA MiniPrep (Cat R2052) *All centrifugation steps should be performed at 10,000 – 16,000 x g. RNA species 17 nt will be recovered.

(a) Add 2 volumes RNA Binding Buffer to each sample1 and mix. (Ex. mix 100 µL buffer and 50 µL sample)

(b) Add an equal volume of ethanol (95-100%) and mix. (Ex. add 150 µL ethanol)

(c) Transfer the sample to the Zymo-Spin™ IC Column in a Collection Tube and cen-
trifuge for 30 seconds. Discard the flow-through.

(d) Add 400 μL RNA Prep Buffer to the column and centrifuge for 30 seconds. Discard the flow-through.

(e) Add 700 μL RNA Wash Buffer to the column and centrifuge for 30 seconds. Discard the flow-through.

(f) Add 400 μL RNA Wash Buffer to the column and centrifuge for 2 minutes. Repeat to ensure complete removal of the wash buffer. Transfer the column carefully into an RNase-free tube.

(g) Add 15 μL DNase/RNase-Free Water directly to the column matrix, let sit for 5 min to ensure surface has absorbed the water, and centrifuge for 30 seconds.

(h) DNaseI treatment:

1. For each sample to be treated, prepare DNase I reaction mix in an RNase-free tube (not provided). Mix well by gentle inversion:

<table>
<thead>
<tr>
<th>DNAseI treatment:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA sample (10 μg) volume adjusted with water or TE buffer</td>
<td>40 μL</td>
</tr>
<tr>
<td>DNase I</td>
<td>5 μL</td>
</tr>
<tr>
<td>DNA Digestion Buffer</td>
<td>5 μL</td>
</tr>
<tr>
<td>50 μL total</td>
<td></td>
</tr>
</tbody>
</table>

2. Incubate at room temperature (20-30°C) for 15 minutes.

(i) Prepare adjusted RNA Binding Buffer. Mix an equal volume of buffer and ethanol (95-100%). (Ex. mix 50 μL buffer and 50 μL ethanol)

(j) Add 2 volumes of the adjusted buffer to the sample and mix. (Ex. mix 100 μL adjusted buffer and 50 μL sample)

(k) Transfer the mixture to the Zymo-Spin™ Column and centrifuge for 30 seconds.

(l) Add 400 μl RNA Prep Buffer to the column and centrifuge for 30 seconds. Discard the flow-through.

(m) Add 700 μl RNA Wash Buffer to the column and centrifuge for 30 seconds. Discard the flow-through.

(n) Add 400 μl RNA Wash Buffer to the column and centrifuge for 2 minutes to ensure
complete removal of the wash buffer. Transfer the column carefully into an RNase-free tube (not provided).

(o) Add 15 µL DNase/RNase-Free Water directly to the column matrix and centrifuge for 30 seconds.

(p) Nanodrop and store a 1:10 dilution.

The eluted RNA can be used immediately or stored at -70°C.

**Source:** Direct-zol™ RNA MiniPrep (Cat R2052) and RNA Clean & Concentrator™-5 (R1015)
B  iScript cDNA Synthesis

Purpose: The goal is to generate cDNA from RNA using reverse transcriptase and to amplify the cDNA using PCR.

Overview: cDNA is generated from RNA to stabilize the easily degradable RNA (so it can be worked with causing until cDNA has been generated). The PCR then amplifies the cDNA to have higher yields of DNA. The cDNA can then be used for further analysis such as qPCR.

Materials:

- iScript cDNA Synthesis Kit (Cat 1708890)
- PCR tubes
- thermocycler

Methods:

Generate the following Reaction Mix

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume per rxn, µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5x iScript Reaction Mix</td>
<td>4</td>
</tr>
<tr>
<td>iScript Reverse Transcriptase</td>
<td>1</td>
</tr>
<tr>
<td>Nuclease-Free Water</td>
<td>14</td>
</tr>
<tr>
<td>RNA template</td>
<td>1</td>
</tr>
<tr>
<td>RNA template</td>
<td>1</td>
</tr>
</tbody>
</table>

Use the following PCR settings on thermocycler

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume per rxn, µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>priming</td>
<td>5 min at 25°C</td>
</tr>
<tr>
<td>Reverse Transcription</td>
<td>20 min at 46°C</td>
</tr>
<tr>
<td>RT inactivation</td>
<td>1 in at 95°C</td>
</tr>
<tr>
<td>Optional</td>
<td>Hold at 4°C</td>
</tr>
</tbody>
</table>

Source: iScript cDNA Synthesis Kit (Cat 1708890)
C qPCR using QuantiTect SYBR Green using mTeSR and RSeT ZNF765g2 Gen1Cs as examples

**Purpose:** The goal for this experiments is to quantify the expression of a target gene using qPCR and SYBR green.

**Overview:** SYBR is an interlarding dye that allows samples $C_T$ values to be calculated through fluorescence everytime a nucleotide is added to the sample template during amplification.

**Materials:**

- 2x QuantiTect SYBR Green PCR Master Mix (Cat 204141)
- Fast-96 well plate
- qPCR machine
- primers
- copper plate holder

**Methods:** *Make sure to vortex every reagent and reaction mix

(a) First set up an excel sheet where the corresponding targeted genes are aligned with the targeted primers.

The excel figure is an example of what this looks like for the mTeSR and RSeT 765g2 Gen1C samples.
(b) After the excel sheet is set up, make a 1:10 dilution of the primers that will be used. In this case the primers are as follows: B2M, GAPDH, KLF2, KLF4, DPPA3, DPPA5, DNMT3L, NANOG, SOX11, and ZIC2.

(c) Make master mixes for each individual primer using the master mix table. Make sure to calculate for error by calculating for one extra sample. For instance, since I use the B2M primer a total of 4 times, I calculate for 5 reactions mixes.
Master Mix Table

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume per rxn, µL</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>2x QuantiTect SYBR Green PCR Master Mix</td>
<td>12.5 µL</td>
<td>1x</td>
</tr>
<tr>
<td>Forward and Reverse Primers</td>
<td>1</td>
<td>0.6µM</td>
</tr>
<tr>
<td>Nuclease-Free Water</td>
<td>10.5</td>
<td>NA</td>
</tr>
<tr>
<td>template</td>
<td>1</td>
<td>less than 500ng/rxn</td>
</tr>
<tr>
<td><strong>total</strong></td>
<td><strong>25µL</strong></td>
<td></td>
</tr>
</tbody>
</table>

(d) Add 24µL of each master mix to it’s corresponding well based on the primer. Use an automated micropipettter for this. Make sure you do this on a Fast-96 well plate that is on a copper hold on ice.

(e) Add 1µL of each sample to it’s corresponding template well

(f) Add a seal to the Fast-96 well plate

(g) Vortex the plate

(h) Use the qPCR set up Table for setting up the machine

qPCR set up Table

<table>
<thead>
<tr>
<th>Step</th>
<th>Time</th>
<th>Temperature, °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR Initial Activation Step</td>
<td>15min</td>
<td>95</td>
</tr>
<tr>
<td>Denaturation</td>
<td>15 s</td>
<td>94</td>
</tr>
<tr>
<td>Annealing</td>
<td>30 s</td>
<td>50-60</td>
</tr>
<tr>
<td>Extension (obtain data acquisition at this step)</td>
<td>30 s</td>
<td>72</td>
</tr>
<tr>
<td>Number of Cycles</td>
<td>35-40</td>
<td></td>
</tr>
</tbody>
</table>

(i) Analyze data using the ΔΔCT method on Excel

Source: Qiagen QuantiTect SYBR Green PCR Handbook (Cat 204141)