Chapter 04: Bioinformatic analysis to identify cell cycle targets of pluripotency factors
4.1 Introduction

Mouse embryonic stem cells (mES) stem cells have a highly modified cell cycle that allows them to proliferate rapidly, with doubling times ranging from 8 – 10 hours (Solter et al., 1971; Power and Tam, 1993), by contrast to the typical doubling time of somatic cells (>24 hours). This rapid proliferation is brought about by modifying the regulators of the cell cycle, including cell cycle checkpoints, altering expression patterns of key cell cycle genes and tuning the metabolic state to sustain this rapid proliferation (White and Dalton, 2005; Shyh-Chang et al., 2013).

The rapid proliferation seen in mES is tightly interlinked to the pluripotent state. As seen in embryo development, when pluripotent cells differentiate to multipotent cells, it is accompanied by a lengthening of the cell cycle resulting in reduced proliferation rates (Lange and Calegari, 2010). The lengthening of cell cycle is primarily due to the increase in the length of G1 by the activation of retinoblastoma (RB) and Cdk inhibitors (Cdkns) (White et al., 2005; Lange and Calegari, 2010). Eventually the multipotent progenitors generated during sequential developmental stages differentiate to specialized cell types that do not divide i.e. irreversibly exit from the cell cycle, or they form committed progenitor cells (adult stem cells) that can reversibly exit the cell cycle (Lange and Calegari, 2010; Cheung and Rando, 2013).

In mES cells, Oct-3/4, Sox2 and Nanog form a core transcriptional network that maintains the pluripotent state (Okamoto et al., 1990; Rosner et al., 1990; Scholer et al., 1990; Pesce et al., 1998; Avilion et al., 2003; Chambers et al., 2003; Mitsui et al., 2003; Hart et al., 2004; Boyer et al., 2005; Ivanova et al., 2006; Masui et al., 2007). These transcription factors also regulate a large number of cell cycle genes and in turn are regulated by them. Changes in the cell cycle length are probably best exemplified in reprogramming of somatic cells to an induced pluripotent state which involves a switch from a slow cycling state of a somatic cell to a rapid cycling state of iPS (Takahashi and Yamanaka, 2006; Polo et al., 2012). Further, somatic cells that proliferate faster have a greater propensity to reprogram, suggesting that the rapid proliferation seen in embryonic stem cells may be a key characteristic of pluripotency (Ruiz et al., 2011; Guo et al., 2014). However, the mechanisms that link pluripotency to cell cycle kinetics are not well defined.
Reprogramming of somatic cells to iPS and mouse embryonic stem cells are useful model systems to dissect out the mechanisms that link pluripotency with the rapid cell cycle seen in embryonic stem cells. As outlined in the introduction, a transcription factor network maintains pluripotency. Some cell cycle targets of the pluripotency network (PPN) are identified, but explicit links to others are yet to be made. One method to identify regulatory links between pluripotency and cell cycle is through ChIP-seq data mining. A large number of published ChIP-seq data sets are available in public repositories for many of the transcription factors that maintain the pluripotent state in mES. These data sets are a valuable resource to mine potential links between pluripotency factors and cell cycle genes. Using published ChIP-seq data sets for studies in mES as well as for reprogramming MEFs to iPS, it should be possible to identify sites of enrichment for these transcription factors on the genomic loci encoding cell cycle genes. These enrichment sites might serve as regulatory regions through which the pluripotency factors could regulate the function of these cell cycle genes. A bioinformatics analysis of published ChIP-seq data sets was therefore undertaken to map out these regions, as a good starting point to identify genes that may be regulated by pluripotency factors for further experimental studies.

4.2 Results

4.2.1 Compilation of ChIP-seq data sets from public repositories

To create a list of data sets to mine for potential links between pluripotency factors and cell cycle genes, data sets from NCBI Gene Expression Omnibus (GEO) were collected using keywords related to cell cycle, pluripotency, embryonic stem cells and reprogramming (Edgar, 2002; Barrett et al., 2013). Only data sets that contained ChIP-seq data were considered, since ChIP-chip or other array based methods do not include non-promoter regions such as 3‘ untranslated regions (UTR) or intronic regions which could contain potential enrichment sites (Raha et al., 2010; Marinov, 2018). From this list, only those data sets that contained enrichment data for either the core pluripotency factors Oct-3/4, Sox2/Nanog or Klf4 were considered.

Once the list was compiled (Table 4.1), the ChIP-seq data sets were processed using Model-based Analysis of ChIP-Seq (MACS) and enrichment peak calls were made using a
setting of q-value <0.005 (Q-values are weighted p-values factoring in false discovery rates of detection) (Zhang et al., 2008). The processed data was then converted to file formats compatible with the UCSC genome browser using the public server of usegalaxy.org (Afgan et al., 2016). The compatible file formats were then uploaded onto the UCSC server (Kent et al., 2002; Kent et al., 2005; Raney et al., 2014; Afgan et al., 2016; Casper et al., 2018). Once the data was uploaded, selected cell cycle genes were visualised on the UCSC browser and loci examined for sites of enrichment (Figure 4.1).

Use keywords to find ChIP-seq data sets on NCBI GEO related to cell cycle, pluripotency, embryonic stem cells and reprogramming

Select all ChIP-seq data sets that contain Oct-3/4, Sox2, Nanog or Klf4 enrichment files

Process ChIP-seq data sets with Model-based Analysis (MACS) of ChIP-Seq

Enrichment peak calls made using a setting of q-value <0.005

Convert files to UCSC genome browser compatible format using usegalaxy.org

Visualise genes on UCSC genome browser and identify sites of enrichment on cell cycle genes

**Figure 4.1 Flowchart for methodology used for bioinformatic analysis**

Flowchart outlining the methodology used to determine occupancy by either Oct-3/4, Sox2, Nanog and Klf4 at the loci of cell cycle genes as a means of identifying potential early targets of the Yamanaka factors during reprogramming
<table>
<thead>
<tr>
<th>GEO access number</th>
<th>Transcription factor</th>
<th>Publication title</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE52373</td>
<td>Klf4</td>
<td>C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells</td>
<td>(Di Stefano et al., 2014)</td>
</tr>
<tr>
<td>GSE90893</td>
<td>Oct-3/4, Klf4, Sox2, cMyc</td>
<td>Cooperative Binding of Transcription Factors Orchestrates Reprogramming</td>
<td>(Chronis et al., 2017)</td>
</tr>
<tr>
<td>GSE48666</td>
<td>Nanog</td>
<td>CpG island-mediated global gene regulatory modes in mouse embryonic stem cells</td>
<td>(Beck et al., 2014)</td>
</tr>
<tr>
<td>GSE28455</td>
<td>Sox2</td>
<td>Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells</td>
<td>(Adachi et al., 2013)</td>
</tr>
<tr>
<td>GSE11724</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells</td>
<td>(Marson et al., 2008)</td>
</tr>
<tr>
<td>GSE43231</td>
<td>Oct-3/4, Nanog</td>
<td>Distinct and combinatorial functions of Jmjd2b/Kdm4b and Jmjd2c/Kdm4c in mouse embryonic stem cell identity</td>
<td>(Das et al., 2014)</td>
</tr>
<tr>
<td>GSE36570</td>
<td>Oct-3/4, Klf4, Sox2, cMyc</td>
<td>Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome</td>
<td>(Soufi et al., 2012)</td>
</tr>
<tr>
<td>GSE56312</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Ground State Conditions Induce Rapid Reorganization of Core Pluripotency Factor Binding before Global Epigenetic Reprogramming</td>
<td>(Galonska et al., 2015)</td>
</tr>
<tr>
<td>GSE</td>
<td>Treatment</td>
<td>Description</td>
<td>Reference</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>GSE67520</td>
<td>Oct-3/4</td>
<td>Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming</td>
<td>(Chen et al., 2016)</td>
</tr>
<tr>
<td>GSE11431</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Integration of external signalling pathways with the core transcriptional network in embryonic stem cells</td>
<td>(Chen et al., 2008)</td>
</tr>
<tr>
<td>GSE49848</td>
<td>Klf4</td>
<td>Klf4 and Klf5 differentially inhibit mesoderm and endoderm differentiation in embryonic stem cells</td>
<td>(Aksoy et al., 2014)</td>
</tr>
<tr>
<td>GSE44288</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Master transcription factors and mediator establish super-enhancers at key cell identity genes</td>
<td>(Whyte et al., 2013)</td>
</tr>
<tr>
<td>GSE74636</td>
<td>Oct-3/4, Sox2</td>
<td>ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network.</td>
<td>(Matsuda et al., 2017)</td>
</tr>
<tr>
<td>GSE43275</td>
<td>Oct-3/4, Sox2</td>
<td>Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm</td>
<td>(Aksoy et al., 2013)</td>
</tr>
<tr>
<td>GSE56098</td>
<td>Oct-3/4</td>
<td>Reorganization of enhancer patterns in transition from naive to primed pluripotency</td>
<td>(Buecker et al., 2014)</td>
</tr>
<tr>
<td>GSE35496</td>
<td>Sox2</td>
<td>SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state</td>
<td>(Lodato et al., 2013)</td>
</tr>
<tr>
<td>GSE61475</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Transcription factor binding dynamics during human ES cell differentiation</td>
<td>(Tsankov et al., 2015)</td>
</tr>
<tr>
<td>GSE46130</td>
<td>Oct-3/4, Sox2, Nanog</td>
<td>Transcriptional and epigenetic dynamics during specification of human embryonic stem cells</td>
<td>(Gifford et al., 2013)</td>
</tr>
<tr>
<td>GSE92846</td>
<td>Oct-3/4, Klf4, Sox2</td>
<td>Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells.</td>
<td>(Liu et al., 2017b)</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>GSE22934</td>
<td>Oct-3/4</td>
<td>Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network</td>
<td>(Ang et al., 2011)</td>
</tr>
</tbody>
</table>
4.2.2 Identification of potential enrichment sites of Oct-3/4, Sox2, Nanog, Klf4 on the loci of cell cycle genes

Enrichment tracks from the data sets GSE43231, GSE44288, GSE22934, GSE35496, GSE43275 and GSE11724 are shown in Figures 4.2, 4.5, 4.7, 4.9, 4.11, 4.13, 4.17, 4.19, 4.21, 4.23, 4.25, 4.27, 4.29, 4.31, 4.33, 4.35, 4.37, 4.39, 4.41, and 4.43 (Marson et al., 2008; Ang et al., 2011; Aksoy et al., 2013; Lodato et al., 2013; Whyte et al., 2013; Das et al., 2014). Data from GSE90893 is displayed in the following Figures 4.3, 4.4, 4.6, 4.8, 4.10, 4.12, 4.14, 4.15, 4.16, 4.18, 4.20, 4.22, 4.24, 4.26, 4.28, 4.30, 4.32, 4.34, 4.36, 4.38, 4.40, 4.42, and 4.44 (Chronis et al., 2017), with the enrichment tracks from top to bottom representing

- Enrichment for Oct-3/4 in mES
- Enrichment for Oct-3/4 in pre-IPS clone 1
- Enrichment for Oct-3/4 in pre-IPS clone 2
- Enrichment for Oct-3/4 48 hours post doxycycline induction in MEFs
- Enrichment for Klf4 in mES
- Enrichment for Klf4 in pre-IPS clone 1
- Enrichment for Klf4 in pre-IPS clone 2
- Enrichment for Klf4 in MEFs
- Enrichment for Sox2 in mES
- Enrichment for Sox2 in pre-IPS clone 1
- Enrichment for Sox2 in pre-IPS clone 2
- Enrichment for Sox2 48 hours post doxycycline induction in MEFs

The MEFs used in the study for data set GSE90893 were derived from a mouse that contained a heterozygous R26-M2rtTA allele and a single doxycycline inducible polycistronic cassette that coded for Oct-3/4, Klf4, Sox2, c-Myc (Sridharan et al., 2009). The pre-IPS clones (partially reprogrammed cells) used for the data set GSE90893 were created using retroviral expression of Oct4, Sox2, Klf4, and cMyc in Nanog-GFP reporter MEFs (Mikkelsen et al., 2008; Chronis et al., 2017). The two clones represent the “late intermediate” stage of reprogramming in which the cells have gained ES cell like morphology, repressed MEF gene expression but have not activated the expression of endogenous core pluripotency factors (Mikkelsen et al., 2008). An enrichment peak in pre-IPS was only considered if it was evident in tracks derived from both the clones.
<table>
<thead>
<tr>
<th>Cell cycle gene</th>
<th>Enrichment in mES</th>
<th></th>
<th></th>
<th>Enrichment during early stages of reprogramming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyclin A</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin B</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin D2</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cyclin D3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cyclin E</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cdk1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Cdk2</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdk4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdk6</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>p21 (Cdkn1a)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p27 (Cdkn1b)</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cell cycle gene</td>
<td>Enrichment in mES but absent in pre iPS</td>
<td>Enrichment in pre iPS but absent in mES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oct-3/4</td>
<td>Klf4</td>
<td>Sox2</td>
<td>Oct-3/4</td>
</tr>
<tr>
<td>Cyclin A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclin B</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cyclin D1</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cyclin D2</td>
<td>Yes</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cyclin D3</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cyclin E</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cdk1</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cdk2</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cdk4</td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Cdk6</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>p21 (Cdkn1a)</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p27 (Cdkn1b)</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: There could be multiple enrichment sites on a gene loci displaying different occupancies, hence a site may be present in both categories.
4.2.2.1 **Cyclin A:**
Cyclin A regulates the S-phase of the cell cycle along with Cdk2. In mES, it is constitutively expressed unlike in somatic cells where expression is from late G1 to early M-phase (Stead et al., 2002; Fujii-Yamamoto et al., 2005). Enrichment was seen for Oct-3/4 at the Cyclin A transcription start site (TSS) in mES which corresponds to the Cell Cycle Regulatory Element (CCRE), a repressor element that regulates Cyclin A expression (Huet et al., 1996) (Figure 4.2). A Klf4 enrichment site was also seen upstream of the TSS in mES. Enrichment was seen for Oct-3/4, Sox2 and Klf4 at the TSS during early reprogramming suggesting a direct role of these factors in regulating expression of Cyclin A during the early stages of reprogramming (Figure 4.3).

4.2.2.2 **Cyclin B**
Cyclin B is important in regulating the M-phase of the cell cycle along with Cdk1 and is the only Cyclin in mES to display an oscillatory behaviour in mES (Stead et al., 2002; Fujii-Yamamoto et al., 2005). Enrichment was seen for Klf4 at the 3’ untranslated region (UTR) in mES and pre-iPS which could be a novel transcriptional site of regulation (Figure 4.4). The 3’ UTR of the Cyclin B mRNA is crucial for its translation in MEFs, whether the 3’ UTR region in the gene is involved in its transcriptional regulation is not known (Malureanu et al., 2010). Klf4 also showed enrichment at the TSS in pre-iPS which was absent in mES suggesting that this site may play a potential role in successful reprogramming (Figure 4.4).
Figure 4.2 Enrichment peaks at the Cyclin A locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.3 Enrichment peaks at the Cyclin A locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.4 Enrichment peaks at the Cyclin B locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.3 Cyclin D1

Cyclin D1 regulates G1 with Cdk4 or Cdk6. It is expressed at low levels in mES during the highly reduced G1 (Faast et al., 2004). Overlapping enrichment sites were seen for Oct-3/4 and Klf4 at regions upstream of the TSS in mES (Figure 4.5 & Figure 4.6). Oct-3/4 and Sox2 regulate the expression of Cyclin D1 via miR-302 in human embryonic stem cells (huES), whether they also directly regulate transcription is not known and this site could be a novel potential regulatory site (Card et al., 2008). Overlapping enrichment was seen for Oct-3/4, Sox2 and Klf4 at the intronic region between exon 2 and 3 during early reprogramming suggesting that these factors may play a role in regulating the expression of Cyclin D1 during initiation of reprogramming (Figure 4.6). The same site showed enrichment of Klf4 in pre-iPS but was absent in mES (Figure 4.6). The intronic enrichment site is a potential novel regulatory region. Overlapping enrichment was also seen for Oct-3/4 and Klf4 at the TSS in pre-iPS that was absent in mES (Figure 4.6). These results suggest these sites may be important regulatory regions for successful reprogramming of pre-iPS to mES.

4.2.2.4 Cyclin D2

Cyclin D2 regulates G1 with Cdk4 or Cdk6 and is not expressed in mES (Faast et al., 2004). Overlapping enrichment was seen for Oct-3/4, Sox2, and Nanog at the 3' UTR mES a potentially novel site of transcriptional regulation (Figure 4.7). The 3' UTR region of the Cyclin D2 mRNA is crucial for its translation during corticogenesis, whether the 3' UTR region is also important for transcription is not known (Tsunekawa et al., 2012). As Cyclin D2 is not expressed in mES, this site could be potential site of gene repression. Due to the overlap with the gene 9330179D12Rik at regions near the TSS, it was difficult to determine whether the enrichment peaks were for Cyclin D2 or the gene 9330179D12Rik (Figure 4.8).
Figure 4.5 Enrichment peaks at the Cyclin D1 locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.6 Enrichment peaks at the Cyclin D1 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.7 Enrichment peaks at the Cyclin D2 locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.8 Enrichment peaks at the Cyclin D2 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.5 Cyclin D3

Cyclin D3 regulates G1 with Cdk4 or Cdk6. It is expressed at low levels in mES during highly reduced G1 (Faast et al., 2004). Cyclin D3 gene codes for three transcripts that all code for the same protein (https://www.ncbi.nlm.nih.gov/gene/12445). Overlapping enrichment sites were seen for Oct-3/4, Sox2 and Nanog upstream of the TSS, two intronic regions between exon 1 and exon 2 of the longer transcript (marked A & B in the figure) in mES (Figure 4.9). Overlapping enrichment was also seen for Oct-3/4, Sox2 and Klf4 at intronic region A during early reprogramming (Figure 4.10). Overlapping enrichment was seen for Oct-3/4 and Klf4 at intronic region A in pre-iPS (Figure 4.10). Multiple enrichment sites were seen for Klf4 along the locus of the shorter transcript in pre-iPS and iPS (Figure 4.10). All of these sites are potential novel regulatory regions in mES highlighting an as yet undefined role in regulating the mES cell cycle and its link to pluripotency.

4.2.2.6 Cyclin E

Cyclin E regulates the G1/S phase of the cell cycle along with Cdk2. In mES, it is constitutively expressed unlike in somatic cells where expression is from late G1 to S-phase (Stead et al., 2002; Fujii-Yamamoto et al., 2005). Overlapping enrichment was seen for Oct-3/4, Nanog and Klf4 at regions upstream of TSS in mES (Figure 4.11). Enrichment was seen for Oct-3/4 at the TSS during early reprogramming suggesting a potential role in regulating expression at this stage (Figure 4.12). Enrichment was seen for Klf4 at the same site which was present in MEFs and pre-iPS but absent in mES, suggesting a potential role in regulating successful reprogramming (Figure 4.12). Both these sites are potentially novel regulatory regions.
Figure 4.9 Enrichment peaks at the Cyclin D3 locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.10 Enrichment peaks at the Cyclin D3 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.11 Enrichment peaks at the Cyclin E locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.12 Enrichment peaks at the Cycin E locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.7 Cdk1

Cdk1 regulates the M-phase along with Cyclin B and exhibits maximal activity during G2/M in mES (Faast et al., 2004). Overlapping enrichment was seen for Oct-3/4, Sox2, Nanog at a region ~ 7kb upstream of the TSS in mES which could be a novel site for potential regulation (Figure 4.13). Oct-3/4 indirectly regulates Cdk1 activity via Cd25c and partners with Cdk1 to represses transcription of Cdx2 a master regulator of the Trophectoderm lineage in mES (Niwa et al., 2005; Li et al., 2012b; Zhao et al., 2014). Cdk1 also phosphorylates Nanog in vitro and knockout of Cdk1 results in loss of pluripotency in mES (Zhang et al., 2011; Brumbaugh et al., 2014; Neganova et al., 2014; Huskey et al., 2015). Whether Oct-3/4, Sox2 and Nanog regulate expression of Cdk1 at this novel site is not known. Also there was no corresponding enrichment for Oct-3/4 and Sox2 at this site in pre-iPS suggesting that occupancy by Oct-3/4 & Sox2 may lead to successful reprogramming (Figure 4.14). Enrichment was seen for Oct-3/4, Klf4 and Sox2 at the TSS during early reprogramming suggesting a potential role during early reprogramming (Figure 4.14). Enrichment was also seen for Klf4 at the TSS in mES, pre-iPS and MEFs highlighting a crucial role of Klf4 in regulating Cdk1 expression in general (Figure 4.14).

4.2.2.8 Cdk2

Cdk2 regulates G1/S-phase with Cyclin E and S/G2 phase with Cyclin A2. In mES, it exhibits high activity throughout the cell cycle and is important for maintaining the rapid cell cycle seen (Stead et al., 2002). Enrichment was seen for Oct-3/4 at the TSS during early reprogramming suggesting that Oct-3/4 may play a role in regulating the expression of Cdk2 during initiation of reprogramming (Figure 4.15). Enrichment was also seen for Klf4 at the TSS in mES and pre-iPS (Figure 4.15). During reprogramming, Cdk2 phosphorylates Sox2 at S39 and S253 which while not important to maintain the pluripotent state, is essential for reprogramming MEFs (Ouyang et al., 2015). Enrichment was seen for Klf4 at the intronic region between exon 1 and 2 in MEFs and pre-iPS which was absent in mES suggesting a potential role in regulating successful reprogramming (Figure 4.15). This intronic region could be a novel regulatory region.
Figure 4.13 Enrichment peaks at the Cdk1 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.14 Enrichment peaks at the Cdk1 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.15 Enrichment peaks at the Cdk2 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.9 Cdk4
Cdk4 regulates G1 with Cyclin D1/D2/D3. It is expressed at low levels in mES during the highly reduced G1 (Faast et al., 2004). Overlapping enrichment was seen for Oct-3/4, Sox2 & Klf4 at the TSS during early reprogramming which suggests a potential role in early reprogramming (Figure 4.16). Enrichment was seen for Klf4 at the TSS in pre-iPS which was absent in mES suggesting a potential role in regulating successful reprogramming (Figure 4.16)

4.2.2.10 Cdk6
Cdk6 regulates G1 with Cyclin D1/D2/D3. It exhibits low levels of activity in mES during the highly reduced G1 (Faast et al., 2004). Cdk6 displayed many overlapping enrichment sites for Oct-3/4, Sox2, Nanog and Klf4 along the gene body in mES highlighting an as yet undefined role in regulating the mES cell cycle and its link to pluripotency (Figure 4.17 & Figure 4.18). Many of these sites corresponded to overlapping sites for Oct-3/4, Sox2 & Klf4 during early reprogramming, as well either in pre-iPS or mES or both (Figure 4.17 & Figure 4.18). These results suggest that Cdk6 could be extremely important in also regulating the reprogramming process.
Figure 4.16 Enrichment peaks at the Cdk4 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.17 Enrichment peaks at the Cdk6 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.18 Enrichment peaks at the Cdk6 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.11 p21 (Cdkn1a)

p21 is an inhibitor of Cyclin/Cdk activity and is not expressed in mES cells (Stead et al., 2002). The p21 gene expresses from two transcripts that encodes for the same protein (https://www.ncbi.nlm.nih.gov/gene/12575). Enrichment of Oct-3/4 was seen at the TSS which has been highlighted in an earlier study in mES (Lee et al., 2010) (Figure 4.19). This validates the advantage of using a bioinformatic approach to discover novel potential regulatory regions. There was overlapping enrichment of Klf4 and Sox2 at the TSS during early reprogramming that persisted to the pre-IPS stage but disappeared in mES suggesting that this site could be crucial for reprogramming and maintaining the pluripotent cell cycle (Figure 4.20). There were many other sites along the gene locus that showed alternate occupancy between pre-IPS and mES (Figure 4.20). As p21 is a potent inhibitor of Cdk4 and is a known repressor of reprogramming, these sites could be of vital importance for successful reprogramming (Ruiz et al., 2011).

Note: The experimentally validated binding site for Oct-3/4 at the TSS of p21 is absent in the data set GSE90893 (Chronis et al., 2017). This could be due to the different antibodies used by both studies for ChIP. As the site is experimentally validated, it was considered present for all analysis done.

4.2.2.12 p27 (Cdkn1b)

p27 is an inhibitor of Cyclin/Cdk activity and is not expressed in mES cells (Stead et al., 2002). Enrichment was seen for Nanog ~4kb upstream of the TSS in mES which has experimentally been shown for Nanog enrichment in mES (Munst et al., 2016) (Figure 4.21). The same site is occupied by Sox2 during early reprogramming, suggesting a potential repressive role of Sox2 on p27 expression (Figure 4.22). Overlapping enrichment was seen for Oct-3/4, Klf4 & Sox2 at the TSS during early reprogramming (Figure 4.22). Enrichment was seen for Klf4 at the 3' UTR in pre-IPS which is absent in mES, which could be a novel regulatory region that regulates successful reprogramming to iPS (Figure 4.22).
Figure 4.19 Enrichment peaks at the p21 (Cdkn1a) locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.20 Enrichment peaks at the p21 (Cdkn1a) locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.21 Enrichment peaks at the p27 (Cdkn1b) locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.22 Enrichment peaks at the p27 (Cdkn1b) locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.2.2.13 Control genes

Twelve genes were randomly selected from a list of ten thousand mouse genes in the MGI database (http://www.informatics.jax.org/) using the random list generator https://www.random.org/lists/. The genes were Akp3, Apoa4, Car1, Cygb, Defb7, Hp, Il17b, Lhx1, Lpo, Myo7a, Saa2, & Sept14. These genes were analysed for enrichment sites for Oct-3/4, Klf4, Sox2 and Nanog in mES from the published data sets used. They were also analysed for enrichment sites for Oct-3/4, Sox2 and Klf4 during the early stages of reprogramming. None of the genes showed any significant enrichment for any of the factors probed for (Figure 4.23 to Figure 4.46).
**Figure 4.23 Enrichment peaks at the Akp3 locus in mES**

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.24 Enrichment peaks at the Akp3 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.25 Enrichment peaks at the Apoa4 locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.26 Enrichment peaks at the Apoa4 locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.27 Enrichment peaks at the Car1 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.28 Enrichment peaks at the Car1 locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.29 Enrichment peaks at the Cygb locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.30 Enrichment peaks at the Cygb locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.31 Enrichment peaks at the Defb7 locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.32 Enrichment peaks at the Defb7 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.33 Enrichment peaks at the Hp locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.34 Enrichment peaks at the Hp locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
**Figure 4.35 Enrichment peaks at the II17b locus in mES**

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.36 Enrichment peaks at the Il17b locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.37 Enrichment peaks at the Lhx1 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.38 Enrichment peaks at the Lhx1 locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.39 Enrichment peaks at the Lpo locus in mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.40 Enrichment peaks at the Lpo locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.41 Enrichment peaks at the Myo7a locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.42 Enrichment peaks at the Myo7a locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.43 Enrichment peaks at the Saa2 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.44 Enrichment peaks at the Saa2 locus in MEFs, early reprogramming, pre iPS and mES

The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.45 Enrichment peaks at the Sept14 locus in mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
Figure 4.46 Enrichment peaks at the Sept14 locus in MEFs, early reprogramming, pre iPS and mES
The black arrow indicates the regions of transcription factor enrichment. The direction for gene coding is shown in light green. The scale bar on the top of the image denotes distance on the genome, while the numbers below indicate the genomic region being viewed. Genes are indicated in blue, with the blue boxes denoting exons and the blue line denoting introns. The large blue boxes denote the translated part of the exon while the small blue boxes the untranslated region. The consensus coding sequence is indicated in dark green, with the dark green boxes denoting translated region from the exons and the dark green line denoting intronic regions. The track names are on the left of the image.
4.3 Discussion

mES cells display a rapid proliferation rate which is a key characteristic of pluripotent stem cells. This rapid proliferation is brought about by modifying the regulators of the cell cycle, including cell cycle checkpoints, altering expression patterns of key cell cycle genes and altering the metabolic state to sustain this rapid proliferation (White and Dalton, 2005; Shyh-Chang et al., 2013). To identify potential regulatory regions through which the pluripotency factors could regulate the function of cell cycle regulators, a bioinformatic data mining of published ChIP-seq data sets for core pluripotency factors, Oct-3/4, Sox2, Nanog and Klf4 was carried out.

4.3.1 Strong interdependence between pluripotency factors and cell cycle genes

12 key cell cycle genes were analysed for enrichment peaks on their gene loci and with the exception of Cdk4, all the cell cycle genes contained at least one enrichment peak for the core pluripotency factors across their gene body with many cell cycle genes containing multiple enrichment peaks (Table 4.2). These results further emphasised that there is a strong interdependence between the core pluripotency factors and the rapid cell cycle seen in mES.

4.3.2 Interdependence of Cyclin A and Cyclin E with pluripotency factors

Cyclin A knockouts mES do not proliferate unlike Cyclin E knockout mES that can (Kalaszczyńska et al., 2009; Liu et al., 2017a). Oct-3/4 and Klf4 both have enrichment sites located adjacent to each other near the Cyclin A TSS suggesting they could help maintain cell proliferation by positively regulating the Cyclin A expression.

Cyclin E has been found to be important for stabilising Oct-3/4, Sox2 and Nanog proteins but is not as crucial for mES proliferation (Liu et al., 2017a). Oct-3/4, Sox2 and Nanog have an overlapping enrichment site near the Cyclin E TSS. This site could be part of a potential feedback loop mechanism regulating the protein levels of these core pluripotency factors and hence maintain the pluripotent state.
4.3.3 Cyclin D3 and Cdk6 could be crucial for maintaining pluripotency or determining cell fate

Cyclin D3 and Cdk6 contain multiple enrichment peaks for Oct-3/4, Sox2, Nanog and Klf4 across its gene loci. Cyclin D3 is expressed at low levels and Cdk6 is the predominant Cdk partner for Cyclin D in mES (Faast et al., 2004). In huES cells, all the three D-type cyclins are expressed and their expression levels change depending on the cell fate they differentiate into (Pauklin and Vallier, 2013). During differentiation towards endoderm, all the D-type cyclins are downregulated, whereas during differentiation to the mesodermal lineage, there is an upregulation of Cyclin D2 accompanied by minor increase in Cyclin D1 & Cyclin D3 (Pauklin and Vallier, 2013). During differentiation towards neuro-ectodermal lineage, there is an upregulation of all the D-type cyclins (Pauklin and Vallier, 2013). Further, Cyclin D1–3 directly regulate many lineage determinant genes (Pauklin et al., 2016). It could be possible that the multiple overlapping sites of enrichment seen in Cyclin D3 may be potential regulatory regions that actually control cell fate. As Cdk6 partners with Cyclin D3 in mES, this could also apply to the multiple sites of enrichment seen in Cdk6.

4.3.4 Yamanaka factors target cell cycle genes during initial phase of reprogramming

With the exception of Cyclin B1 and Cyclin D2, all the other cell cycle genes analysed contained at least one binding site for Oct-3/4, Sox2 or Klf4 with many cell cycle genes displaying multiple enrichment sites (Table 4.2). During the initial stages of reprogramming, there is a switch in the cell cycle from a slow cycling state to a faster cycling state (Polo et al., 2012). It is possible that these genes are targeted by the Yamanaka factors to facilitate the switch to a faster cycling state.

4.3.5 Enrichment sites that are present or absent in mES could facilitate complete conversion to iPS

Many enrichment sites were discovered that were occupied by the transcription factors in pre-iPS but were absent in mES or vice versa (Table 4.3). These sites could be potential regulatory regions that the transcription factors need to occupy or vacate for completion of reprogramming. As a slow cycling state is considered one of the key barriers to successful reprogramming, these sites are of potential interest in improving
reprogramming efficiency (Ruiz et al., 2011; Guo et al., 2014). A good example is Cyclin D2 which is not expressed in mES and displayed enrichment for Oct-3/4 and Sox2 at the 3’ UTR which was absent in pre-iPS, suggesting that these sites could be potential sites of repression. Pre iPS cells express Cyclin D2, and it could be that in order to completely reprogram pre-iPS, Oct-3/4 and Sox2 have to specifically repress Cyclin D2 at these sites (Mikkelsen et al., 2008).

4.3.6 Pluripotency factors regulate cell cycle genes at non promoter regions

Many of the enrichment sites were located at TSS of many genes, but there were also a large number enrichment sites across the introns and UTRs. These could be of importance as it is emerging that many non TSS regions are important sites of regulation as seen by data published from the ENCODE (Encyclopaedia of DNA Elements) project. Most of the sites described here in this bioinformatic analysis are novel, and present a great opportunity to study a potential link between the pluripotency network and cell cycle regulation in mES.

In summary, genome-wide transcription factor binding data from public repositories was mined to identify potential sites of regulation by the core pluripotency factors Oct-3/4, Sox2, Nanog and Klf4. Through analysis of the loci of cell cycle genes, many transcription factor binding sites were identified that could represent potential regulatory regions. Some of the sites identified have already been experimentally validated (Lee et al., 2010; Munst et al., 2016) suggesting biological significance of the data mining analysis. The cell cycle genes showing significant transcription factor enrichment peaks were selected for further analysis in gene expression changes during the early stages of reprogramming.