The Expression and Regulation of Genes Correlating with human Embryonic Stem Cell (hESC) Pluripotency and Self-Renewal

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy (PhD) in the Faculty of Life Sciences

2014

Goabaone Gaobotse
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of figures</td>
<td>7</td>
</tr>
<tr>
<td>List of Tables</td>
<td>11</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>13</td>
</tr>
<tr>
<td>Abstract</td>
<td>15</td>
</tr>
<tr>
<td>Declaration</td>
<td>16</td>
</tr>
<tr>
<td>Copyright statement</td>
<td>16</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>17</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>18</td>
</tr>
<tr>
<td>1.1 Human Embryonic Stem Cells</td>
<td>18</td>
</tr>
<tr>
<td>1.1.1 History of human embryonic and mouse embryonic stem cells</td>
<td>18</td>
</tr>
<tr>
<td>1.1.2 Derivation of human embryonic stem cells</td>
<td>19</td>
</tr>
<tr>
<td>1.1.3 Characterization and culturing</td>
<td>21</td>
</tr>
<tr>
<td>1.1.4 Requirements for hESC culturing</td>
<td>24</td>
</tr>
<tr>
<td>1.1.5 Induced pluripotent stem cells</td>
<td>25</td>
</tr>
<tr>
<td>1.2 Regulation of hESC pluripotency and self-renewal</td>
<td>28</td>
</tr>
<tr>
<td>1.2.1 Intrinsic and extrinsic factors</td>
<td>28</td>
</tr>
<tr>
<td>1.2.2 Transcription factor interactions</td>
<td>32</td>
</tr>
<tr>
<td>1.2.3 Signalling and molecular pathways</td>
<td>37</td>
</tr>
<tr>
<td>1.2.4 Epigenetic Regulation</td>
<td>40</td>
</tr>
<tr>
<td>1.3 Characteristics, functions and interactions of YY1 and YY1AP1</td>
<td>42</td>
</tr>
<tr>
<td>1.4 DNA Microarrays</td>
<td>46</td>
</tr>
<tr>
<td>1.4.1 Design and concept</td>
<td>46</td>
</tr>
<tr>
<td>1.4.2 The application of microarrays on ESCs</td>
<td>48</td>
</tr>
<tr>
<td>1.5 Differentiation of hESCs</td>
<td>49</td>
</tr>
<tr>
<td>1.6 Summary</td>
<td>50</td>
</tr>
<tr>
<td>1.7 Hypotheses, Aims and Objectives</td>
<td>51</td>
</tr>
</tbody>
</table>
Chapter 2: Materials and methods ................................................................. 53

2.1 Culture of mouse embryonic fibroblasts (MEFs) ........................................... 53
  2.1.1 Derivation of MEFs .................................................................................. 53
  2.1.2 MEF culture and passaging ................................................................. 55
  2.1.3 Mitomycin C inactivation and freezing of MEFs .................................... 55
  2.1.4 Thawing of MEFs .................................................................................. 56

2.2 Culture of human embryonic stem cells (hESCs) .......................................... 56
  2.2.1 MEFs feeder-based system .................................................................... 56
  2.2.2 Feeder-free system ................................................................................ 57
  2.2.3 Freezing and thawing of hESCs ............................................................. 57

2.3 Differentiation of hESCs .................................................................................. 58
  2.3.1 Activin-A and FGF-2 withdrawal protocols ........................................ 58
  2.3.2 EB differentiation protocol .................................................................... 58
  2.3.3 Chondrogenic protocol .......................................................................... 59

2.4 293FT, MCF-7 and iPS cell culture ................................................................. 61

2.5 Gene identification .......................................................................................... 61
  2.5.1 Microarray data preparation .................................................................. 61
  2.5.2 Bioinformatics analysis of microarray data to mine potential pluripotency genes ................................................................. 63
  2.5.3 Functional analysis to identify potential stem cell maintenance regulatory genes ................................................................. 64

2.6 Polymerase Chain Reaction (PCR) ............................................................... 65
  2.6.1 Primer design ........................................................................................ 65
  2.6.2 RNA Extraction ...................................................................................... 67
  2.6.3 DNase Treatment of RNA ...................................................................... 67
  2.6.4 cDNA synthesis ...................................................................................... 68
  2.6.5 Gel preparation ....................................................................................... 69
  2.6.6 Gel electrophoresis ................................................................................ 69

2.7 Quantitative Real-Time PCR (qRT-PCR) .................................................... 70
  2.7.1 Primer design ......................................................................................... 70
  2.7.2 qRT-PCR ................................................................................................ 72
  2.7.3 Data analysis .......................................................................................... 75
2.8 Immunofluorescence staining ................................................................................. 75
  2.8.1 Single gene staining ..................................................................................... 75
  2.8.2 Co-localization staining ............................................................................. 76
  2.8.3 Microscopy .................................................................................................. 76

2.9 Plasmid vector cloning ......................................................................................... 77
  2.9.1 shRNA designation, annealing and phosphorylation ..................................... 77
  2.9.2 Vector cloning and restriction digest ............................................................. 79
  2.9.3 Gel purification and vector dephosphorylation ............................................. 81
  2.9.4 Ligation and transformation ....................................................................... 83
  2.9.5 DNA sequencing of ligated vector ............................................................... 83

2.10 Transfection and knockdown ............................................................................ 84
  2.10.1 Calcium phosphatase transfection of 293FTs ............................................ 84
  2.10.2 Lipofectamine transfection of MCF-7s .................................................... 84
  2.10.3 Nucleofection of hESCs .......................................................................... 85
  2.10.4 DOX-induced knockdown of MCF-7s, hESCs and iPSCs ............................ 85
  2.10.5 EB formation from knocked down cells .................................................... 86

2.11 Flow cytometry ................................................................................................. 88
  2.11.1 Cell sorting .............................................................................................. 88

2.12 Western Blotting ............................................................................................... 88
  2.12.1 Harvesting of cells ................................................................................... 88
  2.12.2 BCA protein quantification ...................................................................... 89
  2.12.3 Preparation of SDS gels and western blotting .......................................... 91

2.13 Immunoprecipitation (IP) ................................................................................ 96
  2.13.1 Collection of samples .............................................................................. 96
  2.13.2 Running and analysis .............................................................................. 96

2.14 Chromatin Immuno-Precipitation Sequencing (ChIP-Seq) ......................... 97
  2.14.1 Cross-linking ............................................................................................ 97
  2.14.2 Pre-clearing .............................................................................................. 98
  2.14.3 Immunoprecipitation, washes and DNA Purification ............................... 98
  2.14.4 ChIP solutions ....................................................................................... 100
  2.14.5 DNA sequencing ..................................................................................... 100
  2.14.6 Results analysis ...................................................................................... 101
Chapter 3: Mining and identification of potential pluripotency associated genes from Affymetrix microarray data ................................................................. 102

3.1 Selection of candidate genes from Affymetrix microarray data .................. 103
   3.1.1 Assessment of microarray replicates’ consistency and reliability .......... 103
   3.1.2 Selection of genes targeted by pluripotency transcription factors .......... 109
   3.1.3 Comparison of gene expression trends with published data .............. 110

3.2 Identification of genes with best fit to pluripotency ............................... 114
   3.2.1 Selection of target genes interacting with Oct4 ............................ 114
   3.2.2 Functional clustering with DAVID annotation database ................. 116
   3.2.3 Functional evaluation of gene clusters for assembly of a final gene list ................................................................................................................. 118

3.3 Discussion ......................................................................................... 120

Chapter 4: Assessing the expression of selected target genes in hESCs .......... 124

4.1 Validation of target gene expression in hESCs ...................................... 125
   4.1.1 hESC culturing and maintenance ................................................. 125
   4.1.2 Gene expression analysis by qualitative PCR .............................. 128

4.2 Evaluation of the effects of hESC differentiation on the expression of target genes ................................................................................................. 130
   4.2.1 Activin-A withdrawal induced differentiation ............................. 130
   4.2.2 Differentiation through the withdrawal of FGF-2 ....................... 132
   4.2.3 Activin-A + FGF-2 withdrawal differentiation ......................... 135
   4.2.4 Embryoid body differentiation ................................................ 140
   4.2.5 Gene expression relative to Oct4 ............................................... 142
   4.2.6 Microscopic evaluation of differentiated hESCs ....................... 144

4.3 Expression of YY1AP1 and YY1 in hESCs ........................................ 146
   4.3.1 YY1AP1 is consistently down-regulated during differentiation ....... 147
   4.3.2 Correlations between YY1AP1 and YY1 expression during
differentiation .......................................................................................... 148
   4.3.3 YY1 expression during human embryo development ................... 150
   4.3.4 Intracellular localisation of YY1AP1 with Oct4 and YY1 in hESCs andiPSCs ................................................................................................. 152

4.4 Discussion ....................................................................................... 155
Chapter 5: YY1AP1 short-hairpin knockdown in hESCs ................................................. 161
5.1 Cloning and evaluation of a pLVTHM-YY1AP1 inducible vector system ................................................................. 162
  5.1.1 Construct shRNA design and vector preparation ........................................... 162
  5.1.2 pLVTHM-YY1AP1 plasmid transfection into 293FT and MCF-7 cell ............ 164
5.2 Generation of a GFP positive hESC population ................................................ 168
  5.2.1 Nucleofection of hESCs with pLVTHM-YY1AP1 plasmid .................. 168
  5.2.2 Enrichment of a pure GFP positive cell population .......................... 170
5.3 DOX induced knockdown of YY1AP1 in pluripotent hESCs ............................ 173
5.4 Discussion ............................................................................................................. 181
Chapter 6: Characterization of YY1AP1 knocked down cells ................................ 185
6.1 Analysis of gene expression in YY1AP1 knocked down cells ............................ 186
  6.1.1 Evaluation of cell cycle, apoptotic and differentiation marker expression in knocked down cells ................................................................. 186
  6.1.2 Protein level expression patterns in YY1AP1 knocked down and normal hUES1 cells ..................................................................................... 197
  6.1.3 Gene protein interaction analysis by immunoprecipitation ....................... 200
  6.1.4 Differentiation of YY1AP1 knocked down cells and analysis of gene expression after differentiation ............................................................ 202
6.2 ChIP-Seq Analysis of gene Interactions ............................................................ 215
6.3 Discussion ............................................................................................................. 218
Chapter 7: General Discussion ............................................................................... 222
References ............................................................................................................... 233
Total word count: 63 236
LIST OF FIGURES

Chapter 1
Figure 1.1. Derivation and growth of hESCs ............................................................... 20
Figure 1.2. Induction of iPSCs .................................................................................. 26
Figure 1.3. Oct4 expression in hESCs ........................................................................ 30
Figure 1.4. Transcription factor regulation of pluripotency ....................................... 36
Figure 1.5. Pluripotency signalling pathways ........................................................... 39
Figure 1.6. Proposed models of YY1 transcriptional activation and repression .......... 43
Figure 1.7. Principle of microarrays ......................................................................... 47

Chapter 2
Figure 2.1. Growth factor withdrawal differentiation of hESCs .......................... 60
Figure 2.2. Derivation of the TE, ICM and hESCs microarray data generation ....... 62
Figure 2.3. PCR amplification and dissociation plots ............................................. 73
Figure 2.4. pLVTHM plasmid vector ....................................................................... 87
Figure 2.5 BSA standard curve ................................................................................ 90

Chapter 3
Figure 3.1. Gene expression distribution in hUES3 and hUES7 stem cells .......... 104
Figure 3.2. Gene expression distribution in hUES3 and TE cells ........................ 105
Figure 3.3. Total number of pluripotency transcription factors bound to ten 10% cohorts of genes expressed in hUES3, hUES7 and MAN1 stem cells .......... 111
Figure 3.4. Oct4 promoter on chromosome 6 ......................................................... 115
Figure 3.5. DAVID functional annotation output ............................................... 117
Figure 3.6. DAVID functional annotation output for genes with Oct4 binding sites .... 118

Chapter 4
Figure 4.1. Morphological and immunofluorescence evaluation of hUES7 p29 pp12 cells .................................................................................................................. 126
Figure 4.2 qRT-PCR analysis of gene transcript expression of markers in pluripotent stem cells ........................................................................................................ 127
Figure 4.3 Agarose gel electrophoresis for target gene transcripts in hUES7 cells .... 129
Figure 4.4. Q-RT-PCR for control and target gene transcripts in hUES7 and MAN7 cells during a 10-day Activin-A withdrawal differentiation protocol..................................................131
Figure 4.5. q-RT-PCR for target gene transcript in hUES7 and MAN7 cells during a 10-day FGF-2 withdrawal differentiation .................................................................133
Figure 4.6. Comparison of target gene transcript expression in Activin-A and FGF-2 withdrawal during 10-Day protocols for hUES7 and MAN-7 cells...............................134
Figure 4.7. The expression of germ layer marker transcripts in hUES7 cells at Day 10 of two separate Activin-A and FGF-2 removal differentiation protocols .....................136
Figure 4.8. The expression of germ layer marker transcripts in hUES7 cells at Days 0 and 7 of a combined Activin-A plus FGF-2 exclusion differentiation protocol ..........139
Figure 4.9. Target gene transcript expression in HUES7 and MAN7 stem cells during a 10 day EB protocol ........................................................................................................141
Figure 4.10. Target gene transcript expression relative to Oct4 ..................................143
Figure 4.11. Phase-contrast images of hUES7 and MAN7 undergoing differentiation .145
Figure 4.12. Combined target gene expression in hUES7 and MAN7 cells during Activin-A, FGF-2 and EB differentiation protocols .........................................................147
Figure 4.13. The expression of YY1AP1 and YY1 transcripts during the withdrawal of Activin-A, Activin-A plus FGF-2 and during a Chondrogenic differentiation protocol in hESCs ........................................................................................................149
Figure 4.14. The expression of YY1 transcript during human embryo development ....151
Figure 4.15. YY1AP1, YY1 and Oct4 co-localization staining in hESCs and iPSCs .........153
Figure 4.16. YY1AP1 and YY1 expression in hUES7 hESCs and ZK2012L iPSCs .........154

Chapter 5
Figure 5.1. Gel electrophoresis of undigested and digested pLVTHM plasmid vectors .................................................................................................................................163
Figure 5.2. Positive DNA sequencing of a) ligated YY1AP1 shRNA and b) ligated β2M shRNA .................................................................................................................................165
Figure 5.3 RT-PCR showing YY1AP1 expression in 293-FT and MCF-7 cells ............165
Figure 5.4. Transfection of YY1AP1 plasmid into 293-FT and MCF-7 cells ...............166
Figure 5.5. The expression of YY1AP1 and YY1 in YY1AP1 knocked down MCF-7s ......167
Figure 5.6. Nucleofection of YY1AP1 shRNA into hESCs and iPSCs .......................169
Figure 5.7. Multiple transfection of hESCs .................................................................171
Figure 5.8. GFP-based sorting of hESCs ....................................................................172
Figure 5.9. Dox titration of hUES1 cells ................................................................. 174
Figure 5.10. Goat anti YY1AP1 immunofluorescence staining and transcript levels of YY1AP1 and YY1 in GFP positive YY1AP1 knocked down hUES1s .................................................. 175
Figure 5.11. GFP expression in control hUES1 cells transfected with β2M shRNA and empty vector during YY1AP1 knock down ................................................................. 177
Figure 5.12. Morphology of untreated and Dox treated hUES1 cells during YY1AP1 knock down .................................................................................................................. 178
Figure 5.13. YY1AP1 and YY1 transcripts expression in control hUES1 cells during YY1AP1 knock down .......................................................................................................... 179
Figure 5.14. YY1AP1 and YY1 protein expressions in YY1AP1 knock down hUES1 cells 180

Chapter 6
Figure. 6.1. YY1AP1 immunostaining in YY1AP1 knocked down hUES1 cells .......... 187
Figure. 6.2. qRT-PCR for analysis of gene transcript expression in YY1AP1 knocked down hUES1 cells ........................................................................................................... 187
Figure 6.3. qRT-PCR for endodermal, mesodermal and ectodermal markers in YY1AP1 knocked down hUES1 cells .......................................................................................... 188
Figure 6.4. qRT-PCR for analysis of gene transcript expression in B2M knocked down hUES1 cells .................................................................................................................. 189
Figure. 6.5. qRT-PCR for endodermal, mesodermal and ectodermal markers in B2M knocked down hUES1 cells .................................................................................................. 190
Figure. 6.6. qRT-PCR for analysis of gene transcript expression in hUES1 cells transfected with a control pLVTHM empty vector Figure 3.6. DAVID functional annotation output for genes with Oct4 binding sites .................................................. 191
Figure. 6.7. qRT-PCR for endodermal, mesodermal and ectodermal markers in knocked hUES1 transfected with a control pLVTHM empty vector .................................................... 192
Figure. 6.8. qRT-PCR for analysis of gene transcript expression in Doxycycline-treated hUES1 cells .................................................................................................................. 193
Figure. 6.9. qRT-PCR for endodermal, mesodermal and ectodermal markers in hUES1 cells treated with Doxycycline ............................................................................................. 194
Figure. 6.10. qRT-PCR for analysis of gene transcript expression in untreated hUES1 cells ................................................................................................................................. 195
Figure. 6.11. qRT-PCR for endodermal, mesodermal and ectodermal markers in untreated hUES1 control cells ............................................................................................... 196
Figure. 6.12. YY1AP1 and YY1 proteins are expressed in MCF-7s ......................... 197
Figure 6.13. Western blotting for protein expression during YY1AP1 knockdown in hUES1 cells ................................................................. 198
Figure 6.14. Protein expression during YY1AP1 knockdown in hUES1 cells .......... 199
Figure 6.15. Western blotting on immunoprecipitated YY1AP1 and YY1 proteins in 293-FT and hUES1 cells ............................................................................................................ 201
Figure 6.16. Knock down and control EBs formed during 10 days of differentiation of YY1AP1 deficient hUES1 cells .................................................................................... 203
Figure 6.17. qRT-PCR for gene transcript expression analysis in EBs from YY1AP1 knocked down hUES1 cells ..................................................................................... 204
Figure 6.18. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from YY1AP1 knocked down hUES1 cells ................................................................. 205
Figure 6.19. qRT-PCR showing transcript expression of differentiation markers in EBs made from YY1AP1 knocked down hUES1 cells .................................................................... 206
Figure 6.20. qRT-PCR for gene transcript expression analysis in EBs from β2M knocked down hUES1 cells ............................................................................................................. 207
Figure 6.21. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from β2M knocked down hUES1 cells .......................................................................... 208
Figure 6.22. qRT-PCR for gene transcript expression in EBs from hUES1 cells transfected with a control pLVTHM empty vector ................................................................................. 209
Figure 6.23. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from hUES1 cells transfected with a control pLVTHM empty vector .................. 210
Figure 6.24. qRT-PCR for gene transcript expression in EBs from hUES1 cells treated with 1µg/ml of Doxycycline ........................................................................................................ 211
Figure 6.25. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from hUES1 cells treated with 1µg/ml of Doxycycline ................................................................. 212
Figure 6.26. qRT-PCR for gene transcript expression in EBs from control untreated hUES1 ........................................................................................................................................ 213
Figure 6.27qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from control untreated hUES1 cells .................................................................................. 214
Figure 6.28. Agarose gel electrophoresis for fragment size determination of sheared chromatin ................................................................................................................................... 216
Figure 6.29. qRT-PCR for gene transcript expression analysis in YY1 immunoprecipitated chromatin DNA and mock DNA from hUES1 cells ................................................. 216

Chapter 7

Figure 7.1. A proposed model of the potential role of YY1AP1 during the self-renewal of stem cells ........................................................................................................................................ 227
LIST OF TABLES

Chapter 2

Table 2.1. Different media and constituents used in the culture of different cell types 54
Table 2.2. Qualitative PCR primers and their annealing temperatures and cycle numbers .................................................................................................................................................. 66
Table 2.3. Digestion reagents and their volumes used for DNase treatment of RNA .....68
Table 2.4 qRT-PCR primers and their amplicon sizes in base pairs .................................. 71
Table 2.5.qRT-PCR master mix reagents and their volumes per single 10 μl and 25 μl reactions ............................................................................................................................................................................. 74
Table 2.6. General qRT-PCR thermal profile .................................................................... 74
Table 2.7. Primary antibodies used for immunofluorescence staining ..........................76
Table 2.8. Secondary antibodies used for immunofluorescence staining ......................77
Table 2.9.shRNA annealing reagents .............................................................................. 78
Table 2.10.shRNA phosphorylation reagents ................................................................ 79
Table 2.11. Vector restriction digest reagents ................................................................. 81
Table 2.12. Vector phosphorylation ................................................................................ 82
Table 2.13.shRNA and plasmid vector ligation components ........................................... 83
Table 2.14. H1 forward primer ....................................................................................... 83
Table 2.15 Polyacrylamide gel electrophoresis separation gels .................................... 91
Table 2.16. Polyacrylamide gel electrophoresis stacking gel components .................... 92
Table 2.17. Primary antibodies used for western blotting .............................................. 94
Table 2.18. Secondary antibodies used for western blotting .......................................... 94
Table 2.19. Western blotting reagents, buffers and their components ............................ 95
Table 2.20. ChIP-Seq buffers and their components ..................................................... 100
Chapter 3

Table 3.1. Mean expression score values of genes expressed in hUES3, hUES7, MAN1 and TE cells.................................................................................................................................107
Table 3.2. Number of genes expressed in hUES3, hUES7, MAN1 and TE replicates .....108
Table 3.3. Total numbers of pluripotency transcription factors targeting 10% cohorts of t-stats ranked genes expressed in MAN1 cells.........................................................................................113
Table 3.4. Final target gene list ..................................................................................................119
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>EB</td>
<td>Embryoid body</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EF1a</td>
<td>Elongation factor-1a</td>
</tr>
<tr>
<td>FACs</td>
<td>Fluorescence activated cell sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal bovine serum</td>
</tr>
<tr>
<td>FGF2</td>
<td>Fibroblast growth factor 2</td>
</tr>
<tr>
<td>G1</td>
<td>Gap1</td>
</tr>
<tr>
<td>G2</td>
<td>Gap2</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>hESC</td>
<td>Human embryonic stem cell</td>
</tr>
<tr>
<td>ICM</td>
<td>Inner cell mass</td>
</tr>
<tr>
<td>iPSC</td>
<td>Induced pluripotent stem cell</td>
</tr>
<tr>
<td>KD</td>
<td>Knockdown</td>
</tr>
<tr>
<td>M-phase</td>
<td>Mitotic-phase</td>
</tr>
<tr>
<td>MAD2L2</td>
<td>Mitotic arrest deficient-like 2</td>
</tr>
<tr>
<td>MCF-7</td>
<td>Michigan cancer foundation-7</td>
</tr>
<tr>
<td>MEF</td>
<td>Mouse embryonic fibroblast</td>
</tr>
<tr>
<td>mESC</td>
<td>Mouse embryonic stem cell</td>
</tr>
<tr>
<td>Oct4</td>
<td>Octamer-binding transcription factor 4</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PFA</td>
<td>Paraformaldehyde</td>
</tr>
<tr>
<td>ROCK</td>
<td>Rho-associated protein kinase</td>
</tr>
<tr>
<td>S-phase</td>
<td>Synthesis-phase</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>shRNA</td>
<td>Short hairpin RNA</td>
</tr>
<tr>
<td>Sox2</td>
<td>Sex determining region Y-box 2</td>
</tr>
<tr>
<td>SSEA</td>
<td>Stage-specific embryonic antigen</td>
</tr>
<tr>
<td>t-stat</td>
<td>Test-statistic</td>
</tr>
<tr>
<td>TBS-T</td>
<td>Tris buffered saline-tween</td>
</tr>
<tr>
<td>TE</td>
<td>Trophectoderm</td>
</tr>
</tbody>
</table>
Tet0  
Tetracycline operator

Tra-  
Trafalgar-

YY1  
Yin-yang 1

YY1AP1  
Yin-yang 1 associated protein 1
ABSTRACT

The Expression and Regulation of Genes Correlating with human Embryonic Stem Cell Pluripotency and Self-Renewal

A thesis submitted for the Degree of PhD to the Faculty of Life Sciences, University of Manchester

Goabaone Gaobotse

December 2014

Stem cell pluripotency and self-renewal are two important attributes of human embryonic stem cells which have led to enhanced interest in stem cell research. Understanding the mechanisms that underlie the regulation and maintenance of these properties is imperative to the clinical application of stem cells. Pluripotency and self-renewal are regulated by different genes, transcription factors and other co-factors such as FoxD3 and Klf4. Oct4, Nanog and Sox2 are central to the stem cell regulatory circuitry. They form interactions with co-factors to promote cell proliferation and inhibit differentiation by negatively regulating differentiation markers. However, there are other novel pluripotency associated factors yet to be studied. In this study, bioinformatics and functional analyses were employed to identify a potential pluripotency gene called YY1AP1 from our lab’s pre-existing microarray data. YY1AP1, a transcription regulatory gene, showed consistent down-regulation with induced cell differentiation. It was further investigated. First, its co-localization with Oct4 in both hESCs and iPSCs was confirmed by immunofluorescence staining. Knockdown experiments were then performed on this gene to investigate effects of knocking it down on gene expression in hESCs. Knocked-down cells were characterized for markers of pluripotency and differentiation at the transcript level. Results showed a down-regulation of pluripotency genes with no specific promotion of any of the germ layer markers. Gene expression at the protein level in knocked down cells was then assessed for YY1AP1, and its binding partner YY1, and pluripotency markers. Results showed that proteins of YY1AP1, YY1, Oct4, Nanog and CTCF were down regulated while the tumour suppressor gene protein, p53, was up-regulated in YY1AP1 deficient stem cells. Protein to protein interaction studies showed that YY1AP1, YY1, Nanog and CTCF proteins directly interacted with each other. Differentiation of YY1AP1 deficient cells into EBs led to an almost complete shutdown of all gene expression, an indication that the cells did not form ‘real’ EBs. Differentiation of YY1AP1 ablated cells did not support any lineage promotion either. These results suggest a potentially new role for YY1AP1 in proliferation and self-renewal of stem cells through its possible direct binding to CTCF or its indirect binding to CTCF in complex with YY1.
DECLARATION

No part of this thesis has been submitted in support of an application for another degree or qualification to the University of Manchester or any other university or other institute of learning.

COPYRIGHT STATEMENT

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property

and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on Presentation of Theses.
ACKNOWLEDGEMENTS

First I would like to thank my supervisor, Prof. Sue Kimber, for her great supervision over the course of this study. I appreciate her constant encouragement throughout the project. I would also like to thank my advisor Prof. Ray Boot-Handford who was always there to offer me advice.

I thank Nicola Bates who was a constant source of technical and moral support. Her help was invaluable. I also thank Dr. Aaron Webber who helped with the bioinformatics analysis of microarray data. I thank Dr. Aixin Cheng for his constant offer of scientific advice and his kind donation of the plasmid vector. I also thank Dr. Andy Hayes, Dr. Ian Donaldson and Ms. Stacey Holden of the DNA Sequencing Unit for their immense help during my Chip-Seq experiments. I am especially grateful to Dr. Donaldson for his help with the Chip-Seq data analysis. I also pass my gratitude to Dr. Daniela Palacios for her kind donation of the ChiP protocol. I thank Mr. Michael Jackson for his help and training with flow cytometry. I also appreciate the assistance of Mr. Stephen Fawkes with sonication during ChiP experiments.

I would like to thank my friends and colleagues Dr. Tao Wang, Dr. Puwapong Nimkingratana and Sundas Chohan who were always there for me professionally and personally. I especially thank Sundas for making my stay here wonderful with her constant words of encouragement and jokes.

Finally, I pass my gratitude to BIUST for funding this work and making this research possible. To my parents and siblings; thank you for your constant support and for bearing my absence from home all these years. I hope this thesis will make you all proud. I dedicate this work to my nephew Letso.
CHAPTER 1

INTRODUCTION

The discovery of stem cells has brought great excitement to the medical research fraternity. Stem cells can potentially be utilized in different areas of biology such as novel drug discovery, organ development and transplantation, spinal cord injury cures and many other areas. Although discovered over three decades ago, there is still a need for further research on stem cells before their clinical and therapeutic potential can be fully realized. Stem cells have two crucial properties; pluripotency and self-renewal. The mechanisms of pluripotency and self-renewal maintenance in stem cells need to be fully understood so that scientists can devise efficient ways of differentiating them into desired cell lineages for clinical applications. In this chapter, published studies and literature on the history of stem cells will be reviewed. The important attributes of stem cells and the different ways that these cells are cultured in the laboratory will also be looked at. Genes, transcription factors and signalling pathways that maintain stem cell properties will also be discussed. Finally, the hypotheses and objectives of this study, which informed its experimental design, will also be succinctly stated. The aims formulated to try and satisfy these hypotheses will be elucidated.

1.1 Human embryonic stem cells

1.1.1 History of human and mouse embryonic stem cells

Stem cells are cells in multicellular organisms that have two important properties. One property of stem cells is that they can give rise to derivatives of all the three germ layers; endoderm, mesoderm and ectoderm. These germ layers can then differentiate into almost all the different cell types in the adult body. This property is called pluripotency. Another property of stem cells is that they can mitotically divide under correct laboratory conditions to produce more of themselves through a phenomenon called self-renewal. Stem cells are generally divided into two classes; adult and embryonic stem cells (ESCs). Even though it has been discovered that adult stem cells can differentiate into a greater range of cell types than originally thought, their
tendency to only differentiate into a small number of cells has made them less appealing to researchers (Gepstein, 2002). During the early stages of stem cell research, it was thought that ESCs bore a close resemblance to the epiblast-derived ectodermal cells (Zwaka and Thomson, 2005). The name embryonic stem (ES) cell was coined by Evans and Kaufman after isolating the first ESCs from mice blastocysts in the early 1980s (Evans and Kaufman 1981; Martin 1981). From cultures of these isolated ESCs, they realized that the cells were able to grow either in vitro or develop, in vivo, into tumour cells when injected into mice.

After the successful isolation of ESCs, scientists began to investigate their ability to differentiate into somatic tissues. In their study, Wobus et al. (1984) discovered that delayed mouse blastocysts of strain 129/ter Sv gave rise to the ESC line ESC-BLC1. Their in vivo differentiation studies showed multi-tissue tumours with all the three germ layers (endoderm, mesoderm and ectoderm), resulting from the inoculation of embryonal stem cells into syngeneic mice. In their in vitro studies, stem cells differentiated into neural-like structures that resembled endoderms.

1.1.2 Derivation of human embryonic stem cells

With the in vitro potential of embryo-derived mouse stem cells established, scientists were then tasked with trying to isolate human embryonic stem cells (hESCs). This involved the use of embryos donated by couples undergoing In vitro fertilization (IVF). The embryos were cultured in the laboratory to produce blastocysts whose individual ICMs were separated into independent stem cells (Bongso et al., 1994; Figure 1.1).
Figure 1.1 Derivation and growth of hESCs. The blastocyst is derived from the morula. The embryo comprises of the inner cell mass (ICM) and the trophectoderm (TE) which are both isolated from a 5-8 day old whole blastocyst. Stem cells are derived from the ICM by immunosurgery and then grown on MEF coated plates. MEFs provide the nutrition required to keep the hESCs pluripotent and self-renewing. After their expansion on MEFs, hESCs are differentiated by growth in suspension into embryoid bodies (EBs). EBs can differentiate into various specialised cells of the body such as smooth muscle, skeletal muscle and hematopoietic cells.
The differentiation of hESCs is induced by the introduction of stringent conditions such as the use of specific growth factors and active substratum (Pouton and Haynes, 2007). Three types of cells with the capacity to differentiate into multiple cell lineages have been derived from human tissues. These are hESCs, human embryonal carcinoma (EC) cells and human embryonic germ (EG) cells (Gepstein, 2002). ECCs are malignant tumour stem cells comprising of somatic and extraembryonic (Andrews et al., 2005). On the other hand, EGCs are derived from the late embryo primordial cells and develop into male and female gametes. ECCs were isolated first among the three cell types from germ cell tumours. However, EC cell cultures are rarely used in clinical practice because they have a lower differentiation capacity when compared to hESCs and can also give rise to tumours since they are derived from tumours (Gepstein, 2002).

1.1.3 Characterization and culturing

Stem cells have special characteristics that make them unique from other cell types. These attributes, such as their ability to generate progenitors of differentiated cell types and their possession and expression of particular transcription factors and other specific markers (Hoffman and Carpenter, 2005), are used to characterize them.

Human ESCs can be distinguished by their possession of surface markers such as glycoproteins like SSEA-4, TRA-1-60 and TRA-1-81 (Andrews et al., 1984). They also possess other surface antigens such as CD9, CD117 and CD135 (Kaufman et al., 2001; Hoffman and Carpenter, 2005). On the contrary, mouse embryonic stem cells (mESCs) on the other hand express embryonic antigens such as SSEA1, which are only specific to them (Ginis et al., 2004). Transcription factors used to characterize hESCs are responsible for the maintenance of their pluripotency and self-renewal. This includes Oct4 and Sox2 which play major roles in the regulation of stem cell pluripotency and self-renewal (Yuan, 1995). Nanog, a homeobox transcription factor, has also been identified as a hESC regulatory transcription factor. On the other hand, the pluripotency of mESCs is mainly regulated by Leukaemia Inhibitory Factor (LIF), which is not required for stem cell regulation (Ginis et al., 2004).
In their study working as part of the International Stem Cell Initiative Consortium, Adewumi et al. (2007) characterized 59 hESC lines in 17 different laboratories across the world. They discovered that all these cell lines showed similar expression patterns for a number of hESC markers such as SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, Nanog and Oct4. Even though they found certain gene-dependent differences in the expression of some lineage markers, Adewumi and colleagues deduced that while the hESC lines they studied were derived in different laboratories from different genetic histories by different methodologies, they still showed similar expression patterns for specific marker antigens and genes.

The culturing of hESCs should be carried out under specific conditions, characterized by the presence of well-defined matrices and media to allow the maintenance of the morphologic, phenotypic and karyotypic attributes of the cells (Korin and Levenberg, 2007). Generally, hESCs are differentiated into 3-dimensional cell aggregates called embryoid bodies (EBs) which imitate some of the early embryonic developmental events (Korin and Levenberg, 2007). EBs are capable of mimicking events of early embryonic development due to their ability to promote the interaction of cells with other cells as well as the interaction of the cells with their environment (Dang et al., 2004). Wnt signalling pathway is important in during embryoid body formation by helping in the self-organisation of the EBs. According to ten Berge et al. (2008) the activation of the Wnt signalling pathway leads to the formation of the primitive streak during EB formation. Wnt5a and Wnt11 have been reported to be important in the regulation of hESCs differentiation into EBs (Hwang et al., 2009). It is reported that the regulation of EB formation by Wnts is size significant. When Wnt5a is highly, smaller EBs are formed while on the contrary, high expression levels of Wnt11 leads to the formation of big EBs (Hwang et al., 2009). This is found to be highly significant during lineage specification of hESCs.

Even though hESCs can now be cultured in feeder-free environments using Dulbecco’s modified eagles medium (DMEM) supplemented with serum, feeder-free culture systems are not completely reliable hence the continued use of feeder-based systems. Mitomycin C inactivated mouse embryonic fibroblasts (MEFs) are commonly used to grow stem cells because they contain soluble factors that are crucial for the continuous differentiation of the cells (Korin and Levenberg, 2007; Caramasa et al.,
However, xenogeneic contaminants may be transferred from the MEFs into the cultured cells (Mallon et al., 2006). Human ESC culture systems require the use of serum to keep the cells expanding while maintaining their undifferentiated state. However serum, such as Foetal Bovine Serum (FBS), may contain compounds that may be hazardous to the growing hESCs (Korin and Levenberg, 2007). Batch to batch variability of serum can negatively impact the growth and pluripotency of cells. KnockOut Serum Replacement (KOSR), which commercially available serum, is now commonly being used because it reduces variations in the culture conditions and lowers the differentiation of the cells during proliferation.

There are different ways of culturing hESCs in the laboratory. These include maintaining the cells as large clusters of colonies, before slicing them into pieces and transferring them into MEF embedded culture dishes (Costa et al., 2008). Human ESCs can also be passaged using enzymes for short periods of time (not more than 25 passages) to potentially avoid the development of chromosomal abnormalities over time while non-enzymatic mechanical passaging can be used to maintain cell cultures for longer periods of time (Costa et al., 2008). Irrespective of the method used to passage hESC colonies, it is crucial that the cells are kept in clusters (Korin and Levenberg, 2007). More recently, Rho-associated protein kinase (ROCK) inhibitor has been used in the promotion stem cell adhesion and proliferation in culture. ROCK, an inhibitor of p160-Rho-associated coiled-coil kinase, has been shown to have anti-apoptotic attributes by enhancing cell survival after thawing through the ROCK dependent signalling pathway (Li et al., 2008). In their study of the effect of ROCK inhibitor on the survival of hESCs, Watanabe et al. (2007) showed that culturing hESCs in the absence led to up to 27% increase in cloning efficiency without any effect on the pluripotency of the cells. In another related study, hESCs were thawed with and without ROCK inhibitor at different plating densities (Li et al., 2009). It was found out that after cryopreservation, more cell survival was observed in cells treated with ROCK inhibitor than in untreated control cells. In fact, when attached cells were washed with fresh media, more than 10 times of cells treated with ROCK inhibitor remained attached when compared to the controls (Li et al., 2009).
1.1.4 Requirements for hESC culturing

The transfer of possible contaminants from feeder cells into hESCs needs to be eliminated because it may complicate the transplantation of differentiated cells into a recipient patient. This concern has led to a need for the development of feeder-free culture methods to improve the clinical application of hESC derived cultures (Mallon et al., 2006). Researchers have and continue to develop optimized protocols for culturing hESCs under feeder free conditions. The feeder free culturing system of hESCs first was developed 2001 (Xu et al., 2001). Here, hESCs were cultured on matrigel or laminin in MEF conditioned medium. The cells were proliferated successfully over 130 passages and found to express hESC pluripotency markers such as Oct4, SSEA1 and Tra-1-60. Additionally, the cells formed teratomas in mice and gave rise to derivatives of the three germ layers (Xu et al., 2001). These results signalled that hESCs cultured under feeder free conditions maintained stem cell pluripotency and morphology.

A serum-free medium called TeSR1 has been developed for feeder-free culturing of hESCs over long periods of time. In making this medium, Ludwig et al. (2006) used animal derived proteins, such as bovine serum albumin (BSA), combined with cloned zebrafish basic fibroblast growth factor (zbFGF) as pre-added components of the medium. Stempro is another commercial feeder-free medium commonly used in some stem cell laboratories. Akopian et al. (2010), of the International Stem Cell Initiative Consortium, carried out a study to compare TeSR1 and Stempro media. They used 10 ESC lines to assess eight methods of cell culture using medium containing fibroblast growth factor 2 (FGF-2) as a replacement for knockout serum. MEFs were used as positive controls. They discovered that cells were maintained only on MEFs, TeSR1 and Stempro. Cells in the other types of media could not be maintained because of apoptosis, poor attachment and differentiation. Akopian and colleagues predicted that the maintenance of cells cultured in TeSR1 and Stempro was likely due to the advanced production, quality control and sophisticated mixture of growth factors in these media. Baxter et al. (2009) have also developed a serum and feeder-free culture system which can offer long-term (at least 10 passages) support for the self-renewal of hESCs while keeping their differentiation levels minimal. This culture system utilises components like FGF-2, activin-A, NT4, B27 and N2 and each component has a specific
role to play. FGF-2 and activin A, for example, inhibit hESCs differentiation (Dvorak et al., 2005; Vallier et al., 2005) while NT4 acts as a survival factor (Pyle et al., 2006) and combines with FGF-2 to stimulate cell growth. Recently, matrigel has been commonly used as an alternative substrate for the growth of hESCs.

Laminins are important components of the hESCs culture system. In their study, Vuoristo et al. (2013) investigated the potential of a human carcinoma cell line, called JAR, to support hESC proliferation based on the ability of this cell line to produce large quantities of laminins. Interestingly, cells from FES29 and H9 hESC lines as well as HEL11.4, an iPSC cell line, were able to proliferate and survive on matrix developed from JAR cells in conditioned medium. These cells were also successfully differentiated into neurons and hepatocytes (Vuoristo et al., 2013). These findings presented yet another effective non-feeder dependent hESC culture system which could in the clinical application of hESCs.

1.1.5 Induced pluripotent stem cells

Induced pluripotent stem cells (iPSCs) are stem cells made from somatic cells by inducing the expression of pluripotency-associated transcription factors in these somatic cells, in order to make them pluripotent (Figure 1.2). iPSCs are very important in stem cell research as they could eventually eliminate dependence on embryo-derived human embryonic stem cells. These cells have been reported to exhibit characteristics that are similar to natural embryo-derived stem cells. The general concept and first generation of iPSC generation were first described and undertaken by Takahashi and Yamanaka (2006) who showed that pluripotent cells could be induced from mouse embryonic and adult fibroblasts by the introduction of Oct4, Sox2, c-Myc and Klf4. In this study, they hypothesized that genes involved in the functional maintenance of embryonic stem cells might also be able to induce an embryonic state in adult cells. They selected 24 genes as candidate pluripotency inducers in somatic cells and transduced them into MEFs. They found that some combinations of these factors led to the expression of ESC markers in MEFs. More importantly, they found that the combination of Oct4, Sox2, c-Myc and KLF4 induced pluripotency in the MEFs.
The second generation of iPSCs was undertaken by Yamanaka and colleagues by introducing four pluripotency transcription factors (Oct4, Sox2 and Klf4 and c-Myc) into adult human dermal fibroblasts (HDFs) (Takahashi et al., 2007). These HDF derived iPSCs were reported to be morphologically similar to hESCs. Moreover, they proliferated and expressed similar pluripotency genes and surface markers like embryo-derived hESCs. These iPSCs were also reported to be capable of differentiating into the three primary germ layers in-vitro. They also formed teratomas, which are tumours that possess derivative tissues of more than one primary germ layer (Takahashi et al., 2007). Teratomas are used to test cells for pluripotency. iPSCs have also been generated by using lentiviruses to carry Oct4, Nanog, Sox2 and Lin28 into human fibroblasts and transform them into pluripotent cells (Yu et al., 2007). These iPSCs were derived from human somatic cells and were shown to express hESC markers such as SSEA3, SSEA4, Tra-1-60 and TRA-1-81. The cells had correct number of chromosomes which visually appeared normal (i.e. normal karyotype). The cells could also differentiate into derivative cells of the three primary germ layers (Yu et al., 2007).

Figure 1.2.Induction of iPSCs. Reprogramming factors Oct4, Sox2, c-Myc and Klf4 are introduced into cultured somatic cells derived from adult organisms. These somatic cells then develop into pluripotent iPSCs with the ability to differentiate into cells of the three primary germ layers. *Alternatively, Oct4, Nanog, Sox2 and Lin28 have also been combined to generate iPSCs. Usual sources of somatic cells during reprogramming include blood cells, hepatocytes, melanocytes, neural cells and fibroblasts such as mouse embryonic and dermal cells. Because of genetic differences between patients or cell donors, there is cell to cell variability of the somatic cells. This can affect the viability of reprogramming depending on the reprogramming potential of the each batch of cells. The cell type chosen is also crucial for optimum reprogramming.
One of the greatest challenges in disease modelling is the difficulty to replicate and study *in vitro*, the *in vivo* events of disease formation. This inability to develop *in vitro* disease models is worsened by diseases that occur in organs of the body that are very difficult to access (Unternaehrer and Daley, 2011). iPSCs can be a good source of pluripotent cells that could be used in modelling of diseases. Mature somatic cells from patients with genetically defined diseases can be used to generate iPSCs. Because the resultant iPSCs show exact genetic changes in the donor, they can be characterized to understand specific phenotypic changes that occur during disease progression in the patient. These disease-specific iPSCs could then be used during drug screenings to develop substances that can stop or reverse the phenotypic changes observed during disease progression.

Since their discovery, the ability of human somatic cell derived iPSCs to differentiate into other types of body cells has been well documented. iPSCs have been reported to have the ability to differentiate into chondrogenic progenitor cells (Guzzo et al. 2013). In this study, Guzzo and colleagues cultured human iPSC-derived mesenchymal like progenitor cells in the presence of BMP-2. This led to the development of progenitor cells of different types of cartilages like articular, transient and fibro cartilages. This study is an indicator of the massive potential of iPSCs in the clinical treatment of articular joint illnesses. iPSCs have also been successfully differentiated into neural cells. In their study to compare the neural differentiation of hESCs and iPSCs, Hu B-Y et al. (2010) reported that the same transcription factors involved in the differentiation of hESC to neural progenitor cells are also involved in the differentiation of iPSCs toward the same lineage over a similar time frame. These results might be a mark of the future application of iPSCs in clinical and research studies of neural disorders. iPSCs also offer immunological advantages.

IPSCs are heavily utilized in the study of diseases of the central nervous system. In their *in vitro* study, Mariani et al. (2012) used iPSCs to investigate cortical formation in humans. IPSCs were cultured in suspension in the presence of specific factors to generate 3D aggregates that comprised of radial glia, intermediate progenitors and cortical neurons. Gene expression analysis of these aggregates showed that they expressed transcription factors that regulate dorsal patterning and growth. Gene expression patterns in these aggregates also greatly correlated with that
of early cerebral cortical wall in humans (Mariani et al., 2012). These findings show the potential of iPSCs in studying brain development and brain diseases. There are however challenges to the clinical application of iPSCs in the modeling brain disorders. These include differences in iPSC generation methods and genetic variations between individuals and the genetic instability of neurons (Marchetto et al., 2011). Different methods of generating iPSCs can lead to the generation of genetically variable iPSC cultures which can impact the neuronal differentiation capacity of the cells. This can also be an attribute of the genetic differences between individual somatic cell donors.

1.2 Regulation of hESC pluripotency and self-renewal

1.2.1 Intrinsic and extrinsic factors

The full potential of human embryonic stem cells in clinical research and therapy can only be achieved through a thorough understanding of the signalling and transcription factors that govern their pluripotency and self-renewal. Boyer et al. (2005) revealed that the human embryonic stem cell transcriptional pluripotency circuitry, although being principally regulated by Oct4, Nanog and Sox2, involved many other genes which are direct targets of these three core pluripotency transcription factors. Many of these genes are activators of important homeodomain proteins.

Stem cell pluripotency regulation is centred on three transcription factors; Oct4, Nanog and Sox2. According to Kashyap et al. (2009), these three factors work in combination with polycomb repressive complexes (PRCs) and miRNAs to regulate and maintain stem cell properties. They also work with other regulatory complexes to control cellular pluripotency and expression. They increase or decrease each other’s levels of expression to directly regulate many other genes (Chambers and Smith, 2004). Oct4 acts by binding to the ATGCAAT octamer sequence or by alternatively binding to the A/T rich sites of its target genes (Chambers and Smith, 2004). It is highly expressed in the inner cell mass (ICM) but down regulated in the trophectoderm (TE), which contains a mass of differentiating/differentiated cells. Differentiating cells are those just beginning to change lineage while the differentiated cells are those that have already changed lineage and expresses markers of differentiation. Oct4 gets down regulated with differentiation and upregulated during cellular reprogramming.
(Figure 1.3). However, it requires the presence of other factors to function effectively. For example, in the absence of the Leukaemia Inhibitory Factor (LIF), Oct4 alone is incapable of preventing the differentiation of mouse ESCs (Pan and Thomson, 2007). Therefore, the presence of other pluripotency transcription factors and different signalling pathways is important for Oct4 to function optimally. The expression of Oct4 is quickly reduced when the trophoblast lineage is formed (Babaie et al., 2007).

In their study of transcriptional regulation in hESCs, Boyer et al. (2005) used genome-scale-location analysis to identify Oct4, Nanog and Sox2 target genes. Their results showed that Oct4, Nanog and Sox2 co-occupy the majority of their target genes. Most of these target genes encode different transcription factors that are involved in many important developmental events. These results emphasize the cooperation between Oct4, Nanog and Sox2 during the regulation of hESC self-renewal and pluripotency. Identifying target genes of these core pluripotency factors can help in the understanding stem cell self-renewal and pluripotency circuitries.

Transcription factors also play important roles during the differentiation of hESCs into specialized cell lineages. Transcription factors are master regulators of stem cell fate. They determine whether cells should remain in an undifferentiated pluripotent state or differentiate into a particular lineage. They are responsible for the stimulation and inhibition of certain gene expression programmes (Niwa et al., 2000). Oct4 has been reported to play a central role in initiating the formation of cells that are important during mouse embryonic stem cell differentiation (Nichols et al., 1998; Niwa et al, 2000).

In their study to determine conditions necessary for the regulation of developmental potency of ESCs by Oct4, Niwa et al. (2000) found that precise levels of Oct4 determine three specific hESC lineages. Their findings showed that a less than 2-fold elevation in Oct4 expression results in differentiation of mESCs into the primitive endoderm and mesoderm. On the other hand, inhibition of the expression of Oct4 leads to a decline in stem cell pluripotency and the de-differentiation of the cells towards the trophectoderm. Interestingly, a 50% reduction in the expression level of Oct4 is enough to induce the differentiation of stem cells into trophectoderm cells while a 50% overexpression of Oct4 results in the cells differentiating towards the
endodermal and mesodermal lineages (Niwa et al., 2000). These findings signal the principal role of Oct4 in the regulation of pluripotency for lineage commitment control. Figure 1.3 below illustrates the dynamics of stem cell maintenance by Oct4 during reprogramming and differentiation.

Figure 1.3. Oct4 expression in hESCs. When cells undergo differentiation, Oct4 is down regulated and replaced with markers of differentiation leading to the development of different specialised cells. Oct4 levels get restored when cells undergo reprogramming back into pluripotent cells. Each specialised cell expresses markers that are specific to it. Oct4 down regulation during differentiation and up regulation during reprogramming usually occurs at the same time with other pluripotency regulatory transcription factors such as Oct4 and Nanog.
In their investigation of molecular events during hESC differentiation, Babaie et al. (2007) used RNA-interference (RNAi) to suppress the expression of Oct4 in a hESC line and create microarray transcriptional profiles. They then used these profiles to identify genes whose activity and regulation was dependant on Oct4 in human cells. They found changes in the expression of over 1000 genes, including those under the direct regulation of Oct4, such as Sox2, REX1 and GATA6. Among these are genes involved in Wnt, BMP and FGF signalling pathways. Other genes are involved in apoptosis and chromatin remodelling/modification, which are all part of the regulatory circuitry involved in gain or loss of pluripotency. Chromatin remodelling processes such as DNA methylation regulate the conformation and assembly of chromatin to determine whether pluripotency genes are activated or repressed. Therefore, it is these dynamic changes in chromatin structure that regulate the maintenance of pluripotency in stem cells.

Chambers et al. (2003) used expression cloning to identify and isolate a factor that determines self-renewal in mouse ESCs. They found that Nanog mRNA was present in both undifferentiated pluripotent mouse ESCs and hESCs but absent in differentiated cells. They also found that the activity of Nanog occurred in parallel with the activation of STAT3 by cytokine for the initiation of stem cell self-renewal. These findings might indicate the central role that Nanog plays in ESC identity. During embryo development, Nanog regulates the fate of pluripotent ICM cells and helps to preserve the pluripotency of the epiblast, therefore avoiding primitive endodermal differentiation (Chambers et al., 2003; Pan and Thomson, 2007). The over-expression of Nanog in mESCs enables the self-renewal of mESCs without the mandatory LIF and BMP signalling pathways. Over-expressing Nanog also enable hESCs to grow in the absence of feeder layers (Chambers and Smith, 2004; Pan and Thomson, 2007). According to Pan and Thomson (2007), the ability of high levels of Nanog to maintain stem cells in an undifferentiated state in the absence of LIF might be an indicator that it is a direct effector of the LIF-STAT3 signalling pathway during pluripotency maintenance. However, as Chambers and Smith (2004) elaborate, the presence of the BMP and LIF signalling pathways, as well as Oct4, can enhance the activity of Nanog and result in efficient and maximum self-renewal of the cells.
The roles of the core pluripotency and self-renewal regulatory transcription factors in ESCs have been shown to be mutual. Therefore, these genes, through their direct interaction with each other, work synergistically to influence the maintenance of stemness. In their investigation of the protein interaction network of pluripotency regulators in mESCs, Wang et al. (2006) used affinity purification to isolate Nanog protein and identify and purify genes that directly interact with it. Through the use of shRNAs, the Nanog interacting proteins were inhibited in mESCs. It was found out that the inhibition of these genes led to a decline in cell pluripotency. The isolated Nanog protein interaction network was made up of well-known regulators of stem cell pluripotency, including Oct4. It was also found that the decline in any one member of this protein network had a detrimental effect on the overall pluripotency of the cells (Wang et al., 2006a). These findings signal the interdependence of pluripotency transcription factors on one another during the maintenance of stem cell properties. A similar analysis on mESCs has also been performed (van den Berg et al., 2010). This time, Oct4 protein was enriched and its interacting partners were purified. Once again, a pool of pluripotency regulatory proteins was observed including new proteins that had never been previously associated with stem cell pluripotency regulation (van den Berg et al., 2010).

LIF is an important component of the mouse embryonic stem cell maintenance system. However, LIF is not needed for the maintenance of pluripotency and self-renewal in hESCs. A member of the cytokine family called Interleukin 6 (IL6); LIF has been shown by Smith et al. (1988) and Williams et al. (1988) that it can be a direct replacement of differentiation inhibitory activity (DIA). DIA is a soluble anti-differentiation factor produced mainly by different sources. Williams et al. (1988) showed that in the absence if DIA mESCs started to differentiate while the introduction of LIF keeps the mESCs in a pluripotent state.

1.2.2 Transcription factor interactions

Oct4, Nanog and Sox2 form the core of the human embryonic stem cell pluripotency and self-renewal regulatory network (Figure 1.4). However, for effective and sustainable maintenance of pluripotency and self-renewal, these three genes work
together with other genes and transcription factors. In their study, Chia et al. (2010) used genome wide RNA interference studies to identify genes that are involved in the regulation of the pluripotency and self-renewal of hESCs. They discovered that the hESC regulatory network comprised of a complex made up of transcriptional regulators and chromatin remodelling factors. Altogether, these findings show that there is a wealth of other unidentified genes and transcription factors involved in stem cell maintenance.

One of the most common methods of identifying potential pluripotency regulatory transcription factors and genes is through the identification of protein binding partners of the core pluripotency transcription factors. In their study of the protein interaction network of hESC pluripotency factors, van den Berget et al. 2010 used an affinity purification protocol to purify proteins that interact with the pluripotency core factor Oct4. They purified well known Oct4 binding partners such as Sall4, Dax1 and Essrb resulting in an Oct4 interactome of over 160 proteins comprising of transcription factors, chromatin modifiers and remodelling factors. Their findings once again demonstrated that the hESC regulatory machinery extends beyond the core regulators of Oct4, Nanog and Sox2 and that it is a complex network involving different transcription factors and genes.

Oct4 and Nanog have been identified as the principle regulators of pluripotency and self-renewal in hESCs. However, the role played by Sox2 in the hESCs pluripotency circuitry has and is still being elucidated. It has been found that knocking down Sox2 in pluripotent hESCs leads to loss of pluripotency and changes in stem cell morphology (Fong et al., 2008). In this study, Fong et al. (2008) found out that RNAi induced down-regulation of Sox2 also led to the differentiation of the cells towards trophectoderm and the down-regulation of other pluripotency-associated factors, including Oct4 and Nanog. Interestingly, knocking down Oct4 and Nanog in hESCs in separate experiments led to a similar change in morphology observed during Sox2 knock down (Fong et al. 2008). These results signal the close association between Oct4, Nanog and Sox2 in pluripotency regulation and how the interaction between them is crucial for proper maintenance of the stem cell state. The regulatory network of ESCs involves coordination of a complex network of different transcription factors and genes, as well as Oct4, Nanog and Sox2 during pluripotency and differentiation.
fate decisions in the cells. These transcription factors and genes form reciprocal combinatorial complexes and pathways which determine whether cells should remain in a stem cell state or differentiate. The stem cell fate decisions depend on the equilibrium between self-renewal, proliferation and differentiation and the balance among these different cell states (Kashyap et al., 2009). According to Kashyap and colleagues, Oct4, Nanog and Sox2 control cell differentiation by epigenetically modulating the expression of pro-differentiation genes. Loss of Oct4 leads to loss of pluripotency while its overexpression surprisingly induces ESC differentiation. This means that pluripotency maintenance in the cells require precise levels of Oct4. Nanog on the other hand plays an important role in self-renewal rather than pluripotency maintenance in the cells. In combination with Oct4, Sox2 drives the expression of specific pluripotency target genes in stem cells which include Nanog.

Despite their synergistic regulation and maintenance of hESC pluripotency and self-renewal, Oct4, Nanog and Sox2 have been shown to influence different cell fate decisions. It has been reported that in the presence of BMP4, Oct4 induces the differentiation of hESCs towards mesendoderm while low levels of Oct4 pushes the cells towards the ectodermal lineage without BMP4 and the extraembryonic lineage in the presence of BMP4 (Wang et al., 2012). On the other hand, Nanog represses the differentiation of hESCs towards the ectodermal lineage while Sox2 represses mesendodermal differentiation of the cells (Wang et al., 2012).

The absence of Oct4 promotes stem cell differentiation and loss of the undifferentiated state and self-renewal capacity. In their study, Hay et al. (2004) used siRNAs to knockdown Oct4 in mouse and human embryonic stem cells. Knocking down Oct4 in mESCs led to an up regulation of the trophoblast marker CDX2 and multiple endodermal markers including Gata6. Gata6 was also found to be upregulated in Oct4 deficient hESCs (Hay et al., 2004). This shows the important role that Oct4 plays in the maintenance of pluripotency in both mouse and human embryonic stem cells.

RNAi induced knockdown of Oct4 and Beta-2-microglobulin in hESCs and human embryonal carcinoma cells (hECs) led to the down regulation of pluripotency markers such as SSEA3 and Tra-1-60, a signal that differentiation was induced (Matin et al., 2004). The knockdown of these two genes also resulted in the up regulation of trophectoderm markers hCG and Gcm1 in both cell lines. These findings elucidate the
similar role that Oct4 play in both mouse and human ESCs to maintain them in their undifferentiated states.

Among the three core regulators of stem cell pluripotency, Nanog was the last to be discovered (Chamber et al., 2003; Mitsui et al., 2003). However, its role in the pluripotency circuitry is being thoroughly investigated. siRNA induced down regulation of the Nanog transcript and protein in hESCs led to the up regulation of endodermal markers such as Gata4 and laminin B1 as well as the up regulation of trophectoderm markers such CDX2 and hCG (Hyslop et al., 2005). These results suggest that Nanog prevents the differentiation of hESCs by inhibiting the expression of endoderm and trophectoderm markers which, up on the elimination of Nanog, are activated.

Even though Oct4, Nanog and Sox2 are central to pluripotency maintenance in stem cells, they also cooperate with other transcription factors in this regulation. PRDM14 is one such factor. Through genomic RNAi studies, PRDM14 has been shown to directly regulate Oct4 (Chia et al., 2010). In fact it was found that PRDM14 co-localises with Oct4, Nanog and Sox2 in hESCs, signalling its involvement in the stem cell pluripotency regulation circuitry. Along with Oct4, Sox2 and Klf4, PRDM14 was also able to promote the reprogramming of human fibroblasts (Chia et al., 2010).

Nuclear factor related to kappa-B-binding protein (NFRKB) is another transcription that has been shown to interact with the core pluripotency circuitry to regulate stem cell pluripotency. A component of the INO80 chromatin remodelling complex in hESCs, NKRKB knockdown leads to the down regulation of Oct4 and its introduction during the reprogramming of human fibroblasts with Oct4, Nanog, Sox2 and Klf4 led to a two-fold enhancement of reprogramming while its knockdown led to inefficiencies in reprogramming (Chia et al., 2010).

TRIM6 has been identified as a novel member of the pluripotency maintenance circuitry in mouse embryonic stem cells (Sato et al., 2012). In this study, Sato and colleagues found that TRIM6 directly interacts with the pro-oncogene c-Myc and negatively regulate its transcriptional activity. The repression of c-Myc by TRIM6 leads to the maintenance of the pluripotency of mESCs. In fact, it was reported in the same study that the knockdown of TRIM6 led to elevations in c-Myc expression and the down regulation of Nanog expression leading to the differentiation of the cells. This therefore means that in mESCs, in the absence of TRIM6, c-Myc induces
differentiation by repressing the expression of Nanog, which is required for pluripotency maintenance in the cells. The negative regulation of c-Myc by TRIM6 prohibits differentiation through the inhibition of c-Myc transcriptional activity. This study showed how, through its repression of c-Myc expression, TRIM6 regulates pluripotency of mESCs in cooperation with core pluripotency factors like Nanog.

Figure 1.4. Transcription factor regulation of pluripotency. Transcription factors have been identified as master regulators of stem cell pluripotency. Oct4, Nanog and Sox2 are central to the embryonic stem cell pluripotency maintenance and their down regulation causes pluripotency collapse and induction of cell differentiation. These transcription factors work with other cofactors and modifiers. Oct4 itself was the first pluripotency regulator to be reported (Nichols et al. 1998). The ablation of Oct4 leads to the formation of a dominant negative form of Oct4 which induces cell differentiation by interrupting the activity of the wild-type Oct4. As mentioned earlier, only an optimum amount of Oct4 can maintain cell self-renewal which is necessary for pluripotency maintenance (Niwa et al., 2000). This therefore means that there must be a regulatory network in stem cells that keep Oct4 levels optimal so that pluripotency can be maintained. This network includes Nanog and Sox2. Nanog has been reported to have two transactivators and is thus believed to activate the Oct4 promoter during Oct4 expression regulation in stem cells (Pan and Pei, 2003). Sox2 and Oct4 have been shown to be co-regulators of gene expression. Sox2 can regulate the expression of other transcription factors that affect Oct4 expression, and therefore contribute to keeping Oct4 levels optimum in the stem cells, so that pluripotency is maintained (Li et al., 2007). Another member of the pluripotency regulatory network is FoxD3. This transcription factor can activate the Oct4 promoter in a sequence-specific fashion (Pan et al., 2006). As explained by Niwa and colleagues (2010), because Nanog, Sox2 and FoxD3 all activate the expression of Oct4, the over expression of Oct4 causes it to repress itself since it would then be difficult to keep its level of expression optimum. Indeed Oct4, Nanog and Sox2 have been shown to co-occupy a significant proportion of their target genes (Boyer et al., 2005). Green line- promotion, red line-repression. TFs-Transcription Factors.
1.2.3 Signalling and molecular pathways

There is marked contrast between signalling and molecular pathways that regulate pluripotency in mouse and human embryonic stem cells (mESCs). Signalling pathways play an important role in maintaining stem cell properties such as self-renewal and differentiation. The JAK/STAT pathway, for example maintains mESC pluripotent state in the presence of LIF while LIF does not induce pluripotency of hESCs. Notch signalling is not active in pluripotent hESCs but is enriched in differentiated hESC colonies. The P13/AKT pathway maintains mESC pluripotency and its inhibition in hESCs leads to the down regulation of Oct4, Nanog and Sox2. The role of the NF-kB signalling pathway in ESCs is not well understood but this pathway is believed to be enriched in undifferentiated hESCs and its inhibition leads to cell differentiation and death. The Wnt signalling pathway maintains self-renewal in ESCs in association with the TGF-Beta pathway which is also involved in cell fate determination.

The LIF/STAT3 pathway has been found to be crucial in the regulation and maintenance of pluripotency in mESCs. In their study of mESC pluripotency and self-renewal signalling networks, Matsuda et al. (1999) investigated self-renewal regulation in mESCs by Glycoprotein 130 (gp130), a transmembrane protein which forms a complex with LIF and the LIF receptor. They found out that gp130 is required for the activation of the STAT3 signalling pathway. The activation of this pathway is sufficient to maintain the mESCs in a pluripotent non-differentiating state (Matsuda et al., 1999).

The regulation of mESC pluripotency and self-renewal by the LIF-STAT3 pathway is in collaboration with other factors. In investigating effectors of the LIF-STAT3 pathway, Cartwright et al. (2005) reported that the transcription factor Myc was a target of this pathway. It was found out that Myc was highly prevalent in pluripotent cells but upon LIF withdrawal, its level was significantly reduced and the use of a Myc antagonist was found to induce cell differentiation (Cartwright et al., 2005). These findings might be a signal that Myc is capable of maintaining stem cell pluripotency and self-renewal even in the absence of LIF.

The pluripotency and self-renewal of hESCs is regulated by a mechanism comprising of different signalling and molecular pathways. In their study, Dreesen and Brivanlou (2007) analysed major signalling pathways involved in the regulation of
hESCs pluri-potency and self-renewal. They identified the Wnt, TGF-β, NF-κB, MAPK/ERK and the P13K/AKT signalling pathways (Figure 1.5) as the ones at the forefront of regulating and maintaining stem cell properties.

During their study investigating changes in gene expression during stem cell differentiation, Armstrong et al. (2006) found that elements of the MAPK/ERK and the NF-κB signalling pathways were heavily down-regulated in hESCs following the differentiation of the cells into embryoid bodies. This observation was indicative of the roles that these pathways play in maintaining stem cells pluripotent. In fact, the authors reported an over 4-fold loss of some of the MAPK/ERK signalling pathway components such as RASAL2 and SOS1 during stem cell differentiation. Additionally, Armstrong and colleagues reported that inhibiting the expression of components of the MAPK/ERK pathway in hESCs resulted in loss of pluripotency of the cells and an increase in cell differentiation and death. This suggests that the MAPK/ERK signalling pathway is active in pluripotent stem cells but lost when cells differentiate, again signalling the importance of these two signalling pathways in human embryonic stem cell maintenance. Evidence that MAPK/ERK signalling is lost during differentiation was described by Armstrong et al. (2006) who showed by western blotting and flow cytometry that components of this pathway such as MAP4K1, RAF and NRAS were down regulated upon differentiation. They reported that the application of U0126, a specific inhibitor of MAPK/ERK signalling, led to cell differentiation and ultimate cell death, which was evident of the importance of this pathway in the pluripotency and viability of hESCs. Therefore, the MAPK/ERK is active in undifferentiated cells but lost during differentiation.

The P13K/AKT signalling pathway is another pathway that is important in stem cell maintenance. Armstrong et al. (2006) reported an up-regulation of the P13K/AKT signalling pathway components in pluripotent stem cells. According to the authors, the inhibition of this signalling pathway led to a decline in stem cell pluripotency-associated markers Oct4, Nanog and Sox2 indicating that the cells had lost pluripotency. The P13K/AKT signalling pathway Activin-A induced differentiation towards the endodermal lineage.

The TGF-β pathway is yet another signalling pathway that is important in the maintenance of hESC pluripotency. The activin/nodal branch of the TGF-β pathway is
said to be mainly responsible for this function. James et al. (2005) reported that TGF-β/activin/nodal part of the TGF-β pathway is activated by SMAD2/3 in pluripotent stem cells and helps to maintain the expression of pluripotency markers in the cells (Figure 1.5). Furthermore, the authors found out that SMAD2/3 is also required for the activation of the Wnt signalling pathway, which is also a pluripotency maintenance signalling pathway.

Figure 1.5. Pluripotency signalling pathways. Schematic showing some of the signalling pathways involved in the maintenance of stem cell pluripotency and self-renewal. Activation of the TGFβ/Activin/Nodal, FGF and IGF pathways lead to the activation of Smad, Ras/Raf and PI3K proteins in the cytoplasm. This leads to the activation of their respective pathways in both the cytoplasm and nucleus leading to the up-regulation of Oct4, Nanog and Sox2.
1.2.4 Epigenetic Regulation

The pluripotency of stem cells is important for their in vitro application in research and potential in vivo clinical application. ESCs can be induced to differentiate to specific lineages. It is important to understand the regulation of stem cell transition from pluripotency to lineage specialization. Epigenetic regulation is a result of many factors such as chromatin modifications and Polycomb group (PcG) repressive complexes. PcG repressive complexes are a group of proteins involved in the epigenetic silencing of chromatin (Christopherson and Helin, 2010). Epigenetic modifications can help us understand how a balance is struck between pluripotency and lineage specification through differentiation. Epigenetic regulation refers to the regulation of heritable gene expression alterations that occur without any changes in the DNA sequence (Christopherson and Helin, 2010) and it includes chromatin remodelling, histone modification and DNA modifications such as methylations. Epigenetic regulation is very important in the regulation of stem cell pluripotency. Changes in gene expression mediate the expression of important regulatory genes in ESCs. This can be achieved by the promotion of pluripotency transcription factors such as Oct4 and Nanog and the transient inhibition of genes that drive cell differentiation. Epigenetic regulation of gene expression depends on the state and properties of chromatin within the cells. Epigenetic stability is tightly controlled in hESCs to maintain pluripotency but this changes during cell differentiation. The derivation and culture of the ESCs can alter the epigenetic structure of the cells which can affect cellular gene expression (Xie et al., 2013). Therefore; epigenetic modifications determine cell fate decisions by activating or silencing genes to regulate self-renewal, pluripotency and differentiation of cells.

Chromatin comprises of DNA, histones and proteins (Hu et al., 2012). Chromatin modification is the primary way through which epigenetic regulation occurs. It has been shown that in hESCs, there is a loose association between chromatin proteins and the DNA-histone complex in comparison to terminally differentiated cells where such association in more compact (Meshorer et al., 2006). This loose association allows easier remodelling of the chromatin during differentiation. Most chromatin re-modellers are ATP-dependent and function by regulating the interaction between DNA and histones and thereby contributing to the
looser association of chromatin components in hESCs. Some of the well-known chromatin remodelling complexes are NuRD, SWI/SNF and chromodomain helicase DNA binding protein (CHD) (Hu et al., 2006).

During their analysis of epigenetic regulation during embryonic development, Xie et al. (2013) induced the differentiation of hESCs into different lineages including the neural, mesendodermal and mesenchymal stem cells. They then investigated chromatin modifications and DNA methylations of each formed lineage. It was found that early embryonic development promoters interacted with H3K27me3 upon silencing while promoters regulating late embryonic development displayed DNA methylation upon silencing (Xie et al., 2013). These findings suggest that specific epigenetic regulation mechanisms are involved in regulating specific stages during the differentiation of hESC.

PcG repressive complexes have been shown to regulate both mouse and human ESCs (Lee et al., 2006; Boyer et al., 2006a). PcG proteins are a group of transcriptional repressor proteins with the ability to induce chromatin remodelling and activate gene silencing for the maintenance of the identity of the cells (Christopherson and Helin, 2010). These proteins are involved in early embryonic development. PcG repressive complexes are not required for ESC pluripotency maintenance and are divided into two protein complexes; Polycomb Repressive Complex 1 (PRC1) and Polycomb Repressive Complex 2 (PRC2). These two complex groups are differentiated according to function and composition (Christopherson and Helin, 2006). In their study of development in hESCs, Lee et al. (2006) mapped SUZ12, a member of PRC2 to the whole hESC genome. They found that this protein was distributed in over 200 genes which are involved in regulating important human development processes. Moreover, it was realised that genes that are targeted by the PRC2 protein SUZ12 were activated during the differentiation of hESCs. (Lee et al. 2006). Interestingly, Lee and colleagues found that the pluripotency transcription factors; Oct4, Nanog and Sox2 were also bound to the majority of the PRC2 target genes. These findings might signal how pluripotency factors repress the expression of PRC2 target genes in hESCs to maintain the pluripotency.

Polycomb proteins have been shown to repress a large number of developmental genes in mESCs (Boyer et al., 2006a) in order to block the
differentiation of the cells. In their study of polycomb regulation of development in mESCs, using gene-wide analysis, Boyer et al. (2006a) found that both PRC1 and PRC2 occupied a vast number of developmental genes. It was found out that the PcG target genes were however activated upon the initiation of the differentiation of the cells (Boyer et al. 2006a). Therefore, Polycomb protein complexes play an important role in repressing developmental pathways in ESCs in order to keep them pluripotent. Understanding molecular events that regulate stem cell state maintenance and differentiation such as epigenetic regulation will greatly enhance progress towards their clinical application.

1.3 Characteristics, functions and interactions of YY1 and YY1AP1

YY1 is a zinc-finger transcription factor and a member of the GLI-Kruppel gene family. A polycomb group protein, YY1 can repress or activate transcription depending on the types of interactions it forms with its target genes (Shi et al., 1991). Shi and colleagues first described this protein as binding to overlapping regions in the adeno-associated virus p5 promoter region where it induces transcriptional repression. Different models for the mechanisms (Figure 1.6) of YY1’s transcriptional repression and activation have been described. It has been reported that YY1 can compete with transcription activating factors at binding sites of target gene promoters to interfere with the activation of transcription. It is this negative regulation of transcriptional activators by YY1 at target gene promoters that induces transcriptional repression through the ultimate displacement of the activators (Shi et al., 1997). Alternatively, YY1 can block activation of transcription directly, by binding distally to a transcriptional activator at the promoter region of a target gene (Guo et al., 1997). In this mechanism, transcription is repressed through the inhibition of the upstream promoter by YY1. Thomas and Seto (1999) described another mechanism of transcriptional repression by YY1. In this mechanism, YY1 recruits co-repressors to the target gene promoter where it forms a complex with them to induce repression, or chromatin-remodelling in order to induce repression. Likewise, different models have been proposed on how YY1 induces transcriptional activation. YY1 can activate transcription through direct interaction with other transcription factors that enhances its
transcriptional activation (Nguyen et al., 2004). Alternatively, as Thomas and Seto (1999) propose, YY1 induced transcription can be induced indirectly through the recruitment of transcription co-activators to the target gene promoter by YY1.

Figure 1.6. Proposed models of YY1 transcriptional activation and repression a) YY1 (in black & white) can activate transcription by binding directly to a transcription factor at the promoter region b) YY1 can repress transcription and gene expression by binding distally to a transcriptional activator at the gene promoter region in order to block the initiation of transcription.
YY1 has been described to have many cellular functions. It has been reported to have roles in cell division, differentiation and survival. In their study, Sui et al. (2004) knocked down YY1 from lymphoid cells and found out that the down-regulation of YY1 led to a decline in the proliferation rate of cells. YY1 has also been reported to be highly abundant in prostate cancer tissues (Seligson et al., 2005), which are highly proliferative cells. In differentiation, contrasting roles of YY1 have been reported. For example, while it has been reported to be a differentiation repressor (Xu et al., 2004), other studies have reported it as an inducer of cell differentiation (Donohoe et al., 1999; Affar et al. 2006; He et al., 2007; Liu et al., 2007). YY1 has also been reported to have anti-apoptotic functions. The knockdown of YY1 in lymphoid cells led to an increase in apoptosis due to an increase in p53 brought by the decline in the ubiquitination of p53 (Sui et al., 2004). Sui et al. (2004) also suggest that YY1 directly interacts with p53 only in complex with Hdm2. Hdm2 is the human equivalent of the mouse gene mdm2. It has been reported that YY1 expression induces the formation of the p53-Hdm2 association leading to the ubiquitination of p53 by Hdm2 (Gronroos et al., 2004). These findings suggest that YY1 plays an anti-apoptotic role by regulating the degradation of p53. Moreover, its regulation of p53 is mediated by its interaction with Hdm2.

YY1 regulation of the tumour suppressor gene p53 has been reported to involve the inhibition of p53 from interacting with its co-factors. In their study, Gronroos et al. (2004) found that YY1 represses p53 by inhibiting its interaction with its co-activator gene p300. This does not only block the accumulation of p53, but also that of its interacting genes. Gronroos and colleagues found that silencing the expression of YY1 led to the accumulation of p53 and its interacting genes. These findings demonstrate the role that YY1 play in regulating the transcriptional activity of p53.

YY1 has been reported to interact at the protein level with Oct4 only in complex with CTCF in mESCs (Donohoe et al. 2007; Donohoe et al., 2009). YY1 has also been reported to induce the differentiation of hESCs towards the mesodermal lineage (Morikawa et al., 2003; Gregoire at al., 2013). Transcription factor binding site analysis of genes upstream of genes expressed in hESC derived cardiac progenitor cells revealed the presence of YY1 in these cardiac progenitor cells (Gregoire et al., 2013).
was found that, in concert with GATA4, YY1 was able to induce activation of the transcription of Nkx2.5, a marker of cardiac lineage commitment. In fact, Gregoire and colleagues reported that inducing loss of function in cardiac cells led to a decline in YY1 while the overexpression of YY1 in hESCs promoted the differentiation of the cells into cardiac progenitor cells. These findings, indicating the important role of YY1 in promoting cardiogenesis through its transcriptional regulation of Nkx2.5, could be important in clinical applications involving cardiogenesis and cardiac rehabilitation.

In mouse embryonic stem cells (mESCs), YY1 has been implicated in the chromatin remodelling complex (Vella et al., 2012). It is said to bind to chromatin near the transcription start site of its target genes. Its down regulation resulted in the loss of its target genes and an increase in the accumulation of non-coding RNAs (Vella et al., 2011). This signals the potential role of YY1 in regulating chromatin remodelling in mESCs as well as the formation of small RNAs.

YY1 plays many roles in developmental processes. It has been identified as an interacting factor of Smads, a protein group involved in TGF-β and BMP transduction (Kurisaki et al., 2003). In their study, Kurisaki and colleagues reported that YY1 represses the transcription of TGF-β and BMP. It also interacts with Smad proteins 1-4 and represses the transcription of these proteins. siRNA induced silencing of YY1 resulted in the transcriptional activation of TGF-β and BMP and promoted cell differentiation (Kurisaki et al., 2003). These findings signals that YY1 regulates TGF-β and BMP induced cell differentiation by repressing the transcription of Smad proteins.

YY1AP1 is a YY1 co-transcriptional activator gene. YY1AP1, also known as YY1AP, is located on chromosome 1q21.3 and is ubiquitously expressed in some human and cancer cells. YY1AP1 co-localizes with YY1 in cell nuclei and also has two YY1 binding sites (Wang et al., 2001; Wang et al., 2004). The functions of YY1 as a transcriptional co-activator have yet to be fully understood.

YY1AP1 is a YY1 transcriptional regulatory binding partner with a molecular weight of 90kDa molecular weight. It forms a protein comprised of 750 amino acids (Ohtomo et al., 2007). YY1AP1 has been found to play some role in cell cycle regulation. Affymetrix arrays were undertaken to study cell cycle exit after human T cells were deprived of IL-2. YY1AP1 was found to be up-regulated along with cell cycle regulatory genes (Chechlinska et al., 2009). Screening of the human fetal liver cDNA
library has revealed that YY1AP1 directly interacts with the mitotic arrest deficient-like 2 (MAD2L2), a cell-cycle regulatory gene (Li et al., 2007). These genes were also found to co-localize in the nucleus of HeLa cells. Furthermore, the overexpression of YY1AP1 inhibited the proliferation of cells through the induction of cell cycle arrest (Li et al., 2007). Taken together, these studies signal the potential role of YY1AP1 as a novel cell cycle regulatory gene.

1.4 DNA Microarrays

1.4.1 Design and concept

DNA microarrays are used for the simultaneous evaluation and analysis of many different genes. This makes them very efficient because doing gene-by-gene analysis is laborious, for example, by RT-PCR or q-RT-PCR. Microarrays are basically made of probes bound to a solid substrate such as glass. The basic concept of DNA microarrays (Figure 1.7) involves the printing of PCR products onto glass slides to produce high intensity cDNA. First, RNA is extracted from both the reference and experimental samples. The RNA molecules are then reverse transcribed and the resulting cDNA molecules are fluorescently labelled, usually with Cys dyes (Cy3 and Cy5). The cDNA molecules of both samples are then hybridized to probes on DNA chips. After hybridization, the slides are washed and scanned with a laser microscope. Hybridization between the samples and probes will create a fluorescence intensity indicative of how abundant the probe or gene is in both the reference and experimental samples. As is the case with other scientific experiments, microarrays need to be standardized to compensate for any expression errors. These errors might be caused by differences in equipment and protocols used by different laboratories to carry out their microarray experiments (Leung and Cavalieri, 2003). Evaluation of microarray data before analysis is also important as it eliminates poor quality spots from the arrays.
Figure 1.7. Principle of microarrays. RNA is extracted from both samples and reverse transcribed into cDNA. Both cDNA samples are labelled with fluorescent dyes and hybridized overnight on a microarray chip and observed under a fluorescent microscope. The intensity of emitted fluorescence from each array is converted into an expression value which is representative of the level of expression of the gene of interest.
1.4.2 The application of microarrays on ESCs

Microarrays have been used to identify expression patterns which allow cells to acquire different lineages through the differentiation of embryonic stem cells (ESCs) (Chang et al., 2006). Therefore, microarrays allow the understanding and monitoring of expression signals that define stem cells. DNA microarrays are a powerful and universal tool that takes advantage of the accurately specific complementary base-pairing to give researchers the ability to manage RNA and DNA in a way that has never been experienced before for any other biological reagent (Brown and Botstein, 1999). It is on that principle that microarrays are used to analyse RNA and DNA samples harvested from both human and mouse embryonic stem cells cultured under laboratory conditions. Because of their ability to analyse many samples at the same time, microarrays have offered an alternative way of conducting global transcript analysis on stem cells for application in tissue engineering for therapeutic application and in the discovery of novel drugs and pharmaceuticals (Fernandes et al. 2009).

During their induction of iPSCs from mouse and adult fibroblasts, Takahashi et al. (2007) used DNA microarrays to determine and compare global gene expression of their generated iPSCs with gene expression profiles of hESCs and MEFs. They found that there was a close correlation between gene expression in the iPSCs and hESCs with very little correlation between iPSCs and MEFs gene expression profiles. Moreover, these microarrays allowed them to identify those genes that were upregulated in iPSCs and hESCs. In this way, microarrays allow comparison of gene expression patterns between separate samples.

During the investigation of lineage specification of BMP4 in hESCs, DNA microarrays were used to determine markers expressed in cells that had been differentiated through the over-expression of BMP4 in comparison to the untreated hESCs (Xu et al., 2002). In this study, over 43 000 cDNA probes were analysed with only 19 of them showing expression in all the samples. Therefore, DNA microarrays can be used in cell differentiation studies to identify genes that are upregulated and those down regulation when hESCs undergo lineage transition.
DNA microarrays are also critical in studying the genome wide expression patterns of genes to determine their interacting partners in hESCs. DNA microarrays were used in the genome-wide analysis of Oct4, Nanog and CTCF binding sites in hESCs and mESCs (Kunarso et al., 2010). CTCF (CCCTC-binding factor) is a transcriptional repressor zinc finger protein involved in different cellular processes (Chaumeil and Skok, 2012). Kunarso and colleagues found that the binding profiles of Oct4 and Nanog shared only 5% homology between hESCs and mESCs while CTCF displayed a reasonably conserved binding sites profile across the two species. This might signal lack of conservation of binding profiles of Oct4 and Nanog in hESCs and mESCs.

Yamazoe and Iwata (2005) investigated the suitability of high-throughput microarray methods in screening feeder cells responsible for inducing both human and mouse embryonic stem cell differentiation into different lineages. They prepared cell binding microarrays for PA6 cells, human umbilical vein endothelial cells and COS-1 cells. They then cultured these cells for 8 days and applied mouse embryonic stem cells to the arrays. They found out, through immunocytochemistry, that PA6 cells were able to induce neural differentiation of the stem cells, a demonstration of the power of high throughput microarray analysis in screening many cells responsible for inducing hESC differentiation. Microarrays allow scientists to study certain groups of cells involved in the maintenance of stemness. For example, they can be used to study stem cell signalling pathway regulatory genes, stem cell pluripotency and self-renewal transcription factors among others. All these would be otherwise near impossible to undertake manually on a gene by gene basis. In this way, microarrays offer a great platform for the study of the biology of stem cells and the mechanisms that are involved in regulating and maintaining that biology.

1.5 Differentiation of hESCs

The ability of stem cells to give rise to potentially all types of adult body cells has been proven in vitro by many scholars. One of the most common methods of hESC pluripotency confirmation is the testing of their ability to give rise to EBs, which comprise of the three primary germ layers; endoderm, mesoderm and ectoderm (see 1.1.3).
Soon after the first isolation of human embryonic stem cells in 1998, scientists started doing studies to verify their pluripotency. One such study was performed by Itskovitz-Eldor and colleagues in 2000. In this study, the authors cultured aggregates of hESCs in suspension into EBs and, through qRT-PCR, demonstrated that different tissue lineage markers were expressed in those EBs. For example, they reported the expression of alpha-fetoprotein (an endodermal marker), zeta-globin (a mesodermal marker) and neuro-filaments (an ectodermal marker). These findings showed that the EBs comprised of all the three primary germ layers, a strong suggestion that the stem cells were truly pluripotent. In mice, stem cell pluripotency is commonly assessed by the injection of pluripotent stem cells into a 2-8 cell stage mouse embryo blastocyst and observing the formation of foetus and adult multi-celled organisms containing the introduced stem cell derivatives in the resultant chimeras. For obvious ethical considerations, this method of pluripotency verification cannot be applied to hESCs and human beings.

1.6 Summary

The pluripotency and self-renewal capabilities of stem cells have brought a lot of scientific interest to these cells. Studies have revealed that Oct4, Nanog and Sox2 are central to stem cell self-renewal and pluripotency through their interactions with other genes and transcription factors. Studies continue to be done on these factors to increase knowledge on their regulatory mechanisms. Because there are other factors that work with Oct4 and Nanog to maintain stem cell pluripotency, it is important that these factors are also investigated to improve our understanding of the mechanisms of stem cell pluripotency and self-renewal. Microarray experiments have revealed the expression of thousands of genes involved in the regulation of different ESC lines. Regenerative medicine will greatly benefit from the exploitation of the unique properties of embryonic stem cells. It is therefore the intention of this study to identify and investigate the maintenance of human embryonic stem cell properties. The success of this study would not only improve our understanding of stem cell regulation, but would also open up novel areas of future research on other potential pluripotency target genes identified.
1.7 Hypothesis, Aims and Objectives

Stem cells can be a potentially inexhaustible source of cells for differentiation into specific tissue lineages. They may be able to be applied in areas of medicine such as cell and organ transplantation, subsequently eliminating problems such of organ rejection because of 100% genetic compatibility between organ and recipient. This is because patients would get organs developed from iPSCs that had been derived from their own somatic cells.

Hypothesis

The hypothesis of this study is that Oct4 and Nanog in isolation are unable to maintain the pluripotency of embryonic stem cells. There are other known factors involved in this regulation and there are likely to be as yet unidentified pluripotency-associated genes that function and are expressed in a similar pattern to Oct4 and Nanog and that also contribute to the maintenance of the pluripotency and self-renewal of human embryonic stem cells.

Aims

In order to test this hypothesis, the following aims were set;

1. To identify novel pluripotency-associated gene(s) from pre-existing Affymetrix microarray data in our laboratory
2. To investigate the role(s) of one such key gene in regulating pluripotency

Objectives

To achieve the aims of this study, the following objectives were set;

1. To mine Affymetrix microarray data, using specialized bioinformatics softwares, for genes whose expression in hESCs correlate with Oct4 and Nanog expression based on t-test values.
2 To perform an extensive literature search and functional analysis in order to identify genes with best fit to pluripotency

3 To verify the expression of the target genes in stem cells through immunofluorescence staining and qualitative RT-PCR and qRT-PCR experiments.

4 To perform differentiation protocols in order to monitor how such target transcript expression changes on differentiation of hESCs

5 To perform knockdown experiments on target gene(s) and investigate how this affects hESC pluripotency

6 To characterize KD cells and differentiate them through EB formation in order to determine the role of such an identified gene in hESC and their differentiated derivatives
CHAPTER 2

MATERIALS AND METHODS

2.1 Culture of mouse embryonic fibroblasts (MEFs)

2.1.1 Derivation of MEFs

Using sterile instruments and solutions, MEFs were derived from a 13.5 day old pregnant mouse from the Biological Services Unit of the University of Manchester. Mice were killed by a schedule 1 method. The abdomen of the mouse was sterilized with 70% ethanol. Uterine horns were exposed by cutting through the abdominal skin and peritoneum. They were placed in a petri dish containing Phosphate Buffered Saline (PBS) (PAA Laboratories; w/o Ca$^{2+}$ and Mg$^{2+}$). Foetuses were then removed from the embryonic sac and the placenta and membranes were removed and discarded. The foetuses were decapitated and the remaining carcasses were washed three times with PBS (PAA; w/o Ca$^{2+}$ and Mg$^{2+}$). The carcasses were then placed on a clean petri dish and minced finely with a clean scalpel blade. Two ml of trypsin: EDTA (0.25% trypsin w/v, Gibco; 5mM EDTA, Sigma) was added to the petri dish. The dish was incubated for 15 minutes at 37$^\circ$C. After incubation, 5 ml of MEF medium (table 2.1, medium 1) was added and the mixture was transferred into a 15ml centrifuge tube and mixed vigorously using a pipette. The mixture was then transferred into a T75 flask and a further 15ml of MEF medium was added. The flask was incubated overnight at 37$^\circ$C. After overnight incubation, media in the flask was replaced to remove any floating cellular debris. MEFs were allowed to reach at least 90% confluency before passaging.
<table>
<thead>
<tr>
<th>Medium</th>
<th>Application</th>
<th>Constituents</th>
<th>Final Concentration</th>
<th>Supplying Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MEF medium</td>
<td>MEFs, 293FTs, MCF7s</td>
<td>DMEM Medium</td>
<td>88.5%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-Glutamine 2mM (L-Glu)</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Fetal Bovine Serum (FBS)</td>
<td>10%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Penicillin/Streptomycin (Pen/Strep)</td>
<td>0.5%</td>
<td>Gibco</td>
</tr>
<tr>
<td>2. hES medium</td>
<td>hESCs cultured on MEFs</td>
<td>DMEM F12</td>
<td>76.3%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-Essential Amino Acids</td>
<td>1%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-Glutamine 2mM</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Insulin, transferrin, selenium (ITS)</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pen/Strep</td>
<td>0.5%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knock-Out Serum Replacement (KOSR)</td>
<td>20%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-Mercaptoethanol 0.09 mM</td>
<td>367µl</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lipid supplement x100</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td>3. *Feeder-Free Base medium (FF medium)</td>
<td>hESCs, cultured free iPSCs feeder-free</td>
<td>Advanced DMEM F12</td>
<td>90%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L-Glutamine 2mM</td>
<td>1%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSA (0.5g BSA + 50ml PBS)</td>
<td>0.10%</td>
<td>Sigma/PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>β-Mercaptoethanol 0.09 mM</td>
<td>0.2%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-Essential Amino Acids</td>
<td>1%</td>
<td>PAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lipid Supplement x100</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N2 Supplement x100</td>
<td>1%</td>
<td>Gibco</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B27 Supplement x50</td>
<td>2%</td>
<td>Gibco</td>
</tr>
<tr>
<td>4. mTESr1</td>
<td>hESCs, cultured free iPSCs feeder-free</td>
<td>mTESR1™ medium</td>
<td>1:1</td>
<td>Stem Cell Tech.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mTESR1™ supplement</td>
<td></td>
<td>Stem Cell Tech.</td>
</tr>
</tbody>
</table>

Table 2.1 Different media and constituents used in the culture of different cell types
*Before use, 20ng/ml of FGF2 (Autogen Bioclear), 10ng/ml of ActivinA (R&D Systems) and 2ng/ml of NT4 (Preprotech) were added to each FF Base medium tube.

2.1.2 MEF culture and passaging

After reaching at least 90% confluency, media was discarded and MEFs were washed once with PBS (PAA). Meanwhile, 10cm cell culture dishes were coated with gelatin (Sigma, St Louis, USA) for at-least 30 minutes at room temperature. To detach the MEFs from the flask, 5ml of TrypLE™ Express (Invitrogen, Paisley, USA) was added and the flask was incubated for about 2 minutes in a 37°C, 5% CO₂ incubator. In order to neutralize the TrypLE™ Express, 10ml of MEF media (table 2.1, medium 1) was added to the flask. The cell suspension was then transferred into a 15ml tube and pipetted gently up and down to make a single cell suspension. The suspension was centrifuged for 5 minutes at 2000rpm (Heraeus, Thermo Scientific, UK). After centrifugation, the supernatant was discarded and the pellet was resuspended in 9ml of MEF medium. The MEFs were then plated at a ratio of 1:3 in 10cm gelatin-coated culture dishes each containing 7ml of MEF media. The dishes were incubated at 37°C, 5% CO₂ until the cells attained confluency.

2.1.3 Mitomycin C inactivation and freezing of MEFs

Mitomycin C (Sigma, St Louis, USA) was mixed with PBS (PPA Laboratories Inc.) at a final concentration of 10μg/ml. Media was removed from MEF culture dishes. MEFs were then treated with 5ml of pre-warmed Mitomycin C: PBS mixture and incubated for 3 hours at 37°C, 5% CO₂. After incubation, the mixture was removed and the MEFs were washed with PBS. The MEFS were dissociated with the application of TrypLE™ Express as described in Section 2.1.2. The cells were counted with a haemocytometer. They were then centrifuged (Sigma) for 5 minutes and re-suspended in an appropriate volume of ice-cold freezing solution (90% FBS, 10% DMSO-Dimethyl Sulfoxide, both from Sigma). One ml volumes of the cell suspension were transferred into 1.6ml cryopreservation vials (Fischer Scientific UK Ltd). The vials were stored overnight at -78°C before being transferred to a liquid nitrogen tank for long-term storage.
2.1.4 Thawing of MEFs

Prior to the thawing of MEFs, fresh cell culture plates were coated with 0.1% gelatin (20 ml of 1% Gelatin, 180 ml of sterile H₂O) and incubated for at least 1 hour at 37°C. MEF vials were removed from liquid nitrogen tank and thawed in a 37°C water bath until only a sliver of ice could be seen. The vials were cleaned with 70% ethanol before opening to remove any contamination source from the water bath. MEFs in the vials were transferred into 50 ml tubes containing 10 ml of cold MEF media (table 2.1). The tubes were then centrifuged for 2 minutes at 800xg at room temperature. The supernatant was discarded and the pellet re-suspended in 5 ml MEF media (table 2.1, medium 1). Cells were counted with a haemocytometer. About 3 x 10⁵ cells were plated per 35 mm dish (or 1 x 10⁵ cells per Organ Culture Dish, OCD) by re-suspending them in 2 ml of MEF media for each well of the gelatin pre-coated cell culture plates. The plates were carefully placed into a 37°C, 5% CO₂ cell culture incubator and left undisturbed for at least 24 hours.

2.2 Culture of human embryonic stem cells (hESCs)

2.2.1 MEFs feeder-based system

First, six-well plates were coated with 0.1% gelatin solution (1% gelatin and 9 parts of sterile water) and incubated for 30 minutes in a 37°C, 5% CO₂ incubator. Gelatin enhances the adhesion of MEFs to culture plates. Mitomycin C inactivated MEFs from liquid nitrogen storage were thawed centrifuged (2 minutes at 8000xg) and plated onto the gelatin coated plates (as described in section 2.1.4.). The plates were incubated at 37°C, 5% CO₂. MEF media was replaced daily. After the MEFs had reached 80-90% confluency, MEF media was discarded and vials of hUES1, hUES3, hUES7 (All from Harvard; Cowan et al, 2004) or MAN7 (North West Embryonic Stem Cell Centre, University of Manchester) from liquid nitrogen were thawed and plated onto the confluent MEFs in 3 ml hES media per well of a six-well culture plate (table 2.1, medium 2). Cells were fed 48 hours after plating and daily thereafter. They were passaged at least 3 times before being transferred into a feeder-free culture system.
2.2.2 Feeder-free system

Fibronectin solution (1 part fibronectin, Millipore, Billerica, MA, USA: 19 parts PBS w/o Ca\textsuperscript{2+} and Mg\textsuperscript{2+}, PAA Laboratories Inc.) was applied to 6-well plates and the plates were incubated overnight at 4°C. Following incubation, hUES1, hUES3, hUES7 or MAN7 cells were removed from MEFs as follows. First, media in the plates was discarded. The cells were then washed 2 times with PBS (w/o Ca\textsuperscript{2+} and Mg\textsuperscript{2+}, PAA Laboratories Inc.). After washing, 0.5 ml of trypLE\textsuperscript{TM} Express was applied to each well to detach the cells from the plates. The plates were incubated for 2 minutes at 37°C, 5% CO\textsubscript{2}. After incubation, 2ml of hES medium was added to each well to inactivate trypLE\textsuperscript{TM} Express. Cells were collected into 15ml tube which was then centrifuged (Sigma 3-16K centrifuge, SciQuip, Shropshire, UK) for 3 minutes at 700xg. The supernatant was discarded and the tube was gently tapped to break up the cell pellet. The cells were then resuspended in hESC feeder-free media (table 2.1, medium 3). Fibronectin solution was then removed from the plates that had been incubated overnight. Cells were plated onto the fibronectin coated wells in 3ml hESC feeder-free media per well. They were fed with pre-warmed feeder-free media 48 hours after plating and every 24 hours thereafter. The hESCs were passaged at a 1:3 ratio using TrypLE\textsuperscript{TM} Express (as described above in this section) until they reached at least 80% confluency. Alternatively, cells were also cultured on Matrigel (Becton Dickenson). Matrigel is a protein complex derived from mouse sarcoma cells and it is commonly used as a substrate on which hESCs are cultured. Matrigel aliquots were thawed for 2 hours at 4°C to avoid gel formation. This is because at 37°C, Matrigel proteins aggregate together and form a layer of film. After thawing, Matrigel was diluted in 24ml of DMEM-F12 medium (Gibco). One ml of the Matrigel mixture was added to each well of a 6-well plate and the plate was incubated for at least 1 hour at room temperature before cells were cultured and passaged as described above.

2.2.3 Freezing and thawing of hESCs

A vial of hESCs was collected from liquid nitrogen storage tank and thawed as described in Section 2.1.4. However, instead of using MEF media, hES media was prepared and used as described in section 2.2.2. After thawing, hESCs were
resuspended in hES media, centrifuged for 3 minutes at 700xg and then resuspended in 3ml of hES media per well for plating on MEF feeder layers (Section 2.2.1). Plates were incubated in a 37°C, 5% CO₂ incubator. Media was unchanged the day after plating but changed daily thereafter. Cells were cultured until confluent enough for transfer into a feeder-free system.

2.3 Differentiation of hESCs

2.3.1 Activin-A and FGF-2 withdrawal protocols

During the Activin-A (R&D Systems, Abingdon, UK) and FGF-2 (Autogen Bioclear, Caine, UK) differentiation protocols, hUES7, MAN7 and iPSCs (ZK2012L) were cultured and maintained in a feeder-free system (Section 2.2.2) for a period of 10 days. For the Activin-A removal protocol, Activin-A was excluded from hESC FF media while FGF-2 was excluded during the FGF-2 removal protocol (Figure 2.1). In a third protocol, both Activin-A and FGF-2 were simultaneously excluded from hESC FF media but unlike the other 2 protocols, cells were grown for 7 days only. During the Activin-A and FGF-2 removal protocols, RNA was collected at Days 0, 5 and 10 respectively while during the Activin-A plus FGF-2 exclusion protocol, RNA was collected at Days 0, 1, 3, 5, 6 and 7. RNA was treated and reverse-transcribed into cDNA (Section 2.6.3 – 2.6.5) which was used during qRT-PCR (Sections 2.7.3 – 2.7.4).

2.3.2 EB differentiation protocol

HUES7 and MAN7 cells that were used during the embryoid body (EB) differentiation protocol were first cultured on MEF feeders (section 2.2.1) before being transferred into a feeder free system (section 2.2.2). HESCs with at least 80% confluency were harvested as described in section 2.2.2. and cultured in suspension in non-adhesive 60mm culture dishes containing 15ml MEF medium (table 2.1, medium 1). The EBs were cultured for a period of 10 days in a 37°C, 5% CO₂ tissue culture incubator. Media was changed every 4 days by transferring the EBs in MEF medium in-to 15ml tube. The tube was left undisturbed for 15 minutes to all the EBs to settle at the bottom of the tube. As much of the media as possible was carefully discarded without disturbing the settled EBs. Fresh media was added and the EBs were transferred into a fresh non-
adhesive 60mm culture dish and put back into the incubator. RNA was extracted (as in section 2.6.3) from the EBs at Days 0, 5 and 10 of the protocol. The RNA was treated (sections 2.6.4) and reverse transcribed into cDNA (Section 2.6.5) which was used for qRT-PCR (sections 2.7.3 – 2.7.4).

2.3.3 Chondrogenic protocol

The protocol for the directed differentiation of hESCs can be found in Oldershaw et al (2010) published by our lab. In brief, this protocol involves differentiation of hESCs into chondrocytes (cartilage cells) by using a sequence of mesoderm and chondrocyte inducing growth factors. In this protocol, RNA was collected at Days 0, 4, 8 and 14, treated, reverse transcribed (Section 2.6.3-2.6.5) and used during qRT-PCR (Sections 2.7.3 –2.7.4).
Figure 2.1. Growth factor withdrawal differentiation of hESCs. Schematic showing the differentiation of stem cells by growth factor withdrawal. hESCs are cultured in feeder-free stem cell media for 10 days with the omission of either Activin-A or FGF-2 from the media. RNA is extracted at Days 0, 5 and 10 of both the Activin-A removal and FGF-2 removal differentiation protocols. This RNA is then converted to cDNA which is then used to analyse gene expression through protocols such as q-RT-PCR.
2.4 293FT, MCF-7 and iPS cell culture

293FT cells (Invitrogen, USA) and MCF-7 cells were grown and passaged in 6-well plates in MEF medium (table 2.1, medium 1) in exactly the same way that MEFs were cultured and passaged (section 2.1.2). Media was changed every day. IPSCs were cultured in feeder-free medium (table 2.1, medium 3) and passaged in exactly the same way that pluripotent hESCs were cultured and split under-feeder free conditions (see section 2.2.2).

2.5 Gene identification

2.5.1 Microarray data preparation

Microarray data (Affymetrix, Santa Clara, California, United States) has been previously prepared in the lab of Prof. Sue Kimber, Faculty of Life Sciences, University of Manchester. The aim of generating this microarray data was to compare gene expression in the inner cell mass (ICM) of the human blastocyst stage embryo to gene expression in hESCs derived from such ICMs. The isolated trophectoderm (TE), the first differentiated cell type formed at the blastocyst stage, was used as a relevant differentiated cell control (see figure 2.2). Stem cells that were used during the generation of the microarray data were extracted from human embryos donated by IVF patients from St Mary’s Hospital in Manchester. The use of embryos for derivation in Manchester is licensed by the Human Fertility and Embryology Authority (HFEA research licence R0176) and the Local Ethics Committee. The use of embryos for research is licensed under HFEA licence R0026. HUES3 and hUES7 stem cells used were derived at Harvard University by Cowan et al. (2004) from blastocyst stage embryos produced by IVF procedure, with approval from the Harvard institutional review board. Stem cells used during the microarray experiments were cultured in Stempro media (Invitrogen, UK). Acquired Affymetrix microarray data was normalized using the gene expression analysis and normalisation MAS 5.0 method.
Figure 2.2. Derivation of the TE, ICM and hESCs microarray data generation. Schematic showing how microarray data was generated in our lab. TE and ICM cell samples were derived from blastocysts donated by patients undergoing IVF. Some ICM cells were cultured into hESCs. RNA was collected from ICM, TE and hES cells. These RNA samples were then converted to cDNA which was fluorescently labelled and hybridized overnight on DNA microarrays for gene expression analyses.
2.5.2 Bioinformatics analysis of microarray data to mine potential pluripotency genes

Expression patterns analysis was conducted in three hESC lines; hUES3, hUES7 and MAN1, as well as in the trophoderm (TE) based on the genes’ expression score values obtained from the Affymetrix microarray data analysis. Analysis was undertaken using a Java desktop application software called Netbeans Integrated Development Environment (IDE) 6.7.1. (http://netbeans.org/). Transcription factors (TFs) were mapped to promoters 2000 base pairs either side of the transcription start site (TSS) of each gene in Ensembl v65 human GRC37 genome list (filtered for protein-coding genes only). TF datasets were from HAIB and YALE – all those sampled in hES cells – and from independent studies by Boyer et al, 2008 (cell), and Kunarso et al., 2010 (Nature genetics). Probes were ranked by t-test on log (expression value) between hES cell line and TE (outlier replicates were excluded based on anomalously high Oct4 expression). The 10% highest expressed genes (5665 probes) were retained for MAN1; top 20% genes (11330 probes) were retained for each of the hUES3 and hUES7 cell lines. This difference in stringency was due to the MAN1 dataset having high reproducibility of expression values, and showing better contrast with the TE, and better confirmation of quality and functional relevance, based on preliminary DAVID analysis and on distributions of pluripotency markers in the promoters. Only genes common to all 3 of the retained lists for MAN1, hUES3 and hUES7 were retained. Therefore, only genes that were overexpressed in all three cell lines were retained. The retained genes’ probes were then mapped to their official gene symbols and then to Ensembl. Probes mapping to multiple genes were retained multiple times, once for each corresponding gene. Also retained were probes with no corresponding gene symbol or Ensembl ID found. 3146 probes survived this procedure i.e. roughly 60% of the most highly expressed MAN1 probes relative to TE were also highly expressed relative to TE in hUES3 and hUES7 stem cell lines. The Kunarso ChIP-seq data was more extensive and gave better contrast between high and low expressed probes (genes) in hES cells compared to TE cells, so the presence of OCT4 probes from Kunarso set was used as a further filter. Also, probes with no corresponding gene symbol or Ensembl ID were removed. 520 probes survived this procedure, corresponding to 440 distinct genes. These genes were then ranked by t-statistic.
2.5.3 Functional analysis to identify potential stem cell maintenance regulatory genes

Functional analysis of identified genes was performed using DAVID, a functional annotation software (http://david.abcc.ncifcrf.gov/). GO term searches were performed on the genes by selecting the functional annotation tab on the home page of the DAVID online tool. The probeset ID numbers of the genes of interest were then copied on to the ‘upload gene list tab’ with the identifier set as ‘AFFYMETRIX_3PRIME_IVT_ID’ and list type selected as ‘gene list’. With the selection of the ‘submit list’ button, the softwares searches for different pathways and functional groups where the genes of interest are over-represented. This tool performs GO term searches on different annotation categories such as disease associations, protein-protein interactions, homologies and bio-pathways. It uses algorithm to establish if annotation terms are related to each other depending on connected genes. Therefore, the software will pull up different functional clusters where the genes are known to have specific functions. Functional clusters deemed to contain genes whose functions are related to or involved in the regulation of stem cell maintenance were identified. Only clusters that had enrichment score values of 2.0 or less were selected. These are the enrichment values considered to be significant. Enrichment scores are overall enrichment values for each functional cluster or group based on the enrichment values of each cluster member. The higher the enrichment value a cluster has, the more enriched are the constituent members of that cluster. DAVID annotation clusters are also allocated p-values which basically represent the probability that the null hypothesis is correct. The smaller the p-value, the more enriched the cluster is. The functional clusters also show Benjamini corrected enrichment scores and the lower these values are, the more enriched the clusters are. Specific genes were identified from the retained functional clusters for further experimentation.
2.6 Polymerase Chain Reaction (PCR)

2.6.1 Primer design

Primers were designed using the online database Primer 3 (http://frodo.wi.mit.edu/). Target gene cDNA sequences were acquired from NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/) and similar but non-identical sequences were identified by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Both forward and reverse primer sequences were selected from gene coding DNA sequences that did not span an intron. After their design, the primers were checked for sequence specificity using BLAST. All the primers had approximately 40-60% GC content with annealing temperature ranges of 44-64°C. The expected band sizes for each gene were calculated and noted. All the primers were purchased from Invitrogen and reconstituted to a final working concentration of 30μM. Table 2.2 shows all primers used for qualitative PCR with their annealing temperatures and cycle numbers.
<table>
<thead>
<tr>
<th>Target Gene</th>
<th>Forward Primer (5’-3’)</th>
<th>Reverse Primer (5’-3’)</th>
<th>Annealing Temperature</th>
<th>Number Of Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myst3</td>
<td>CCTCCTCCGTTCACTCCACC</td>
<td>TTCAGTGCTGCTCTCAAATGC</td>
<td>51°C</td>
<td>30</td>
</tr>
<tr>
<td>RIF1</td>
<td>TTTCCCTCTGAAGTGTTGC</td>
<td>CACCTCAGCCTCTTTCTCC</td>
<td>50°C</td>
<td>30</td>
</tr>
<tr>
<td>RBMI5</td>
<td>CTCAGCTGCTCACAGACTGC</td>
<td>GAACCTCAAAAGGTGGGAAGG</td>
<td>52°C</td>
<td>30</td>
</tr>
<tr>
<td>MYST4</td>
<td>ACATGTGCCCTGTAAGTCC</td>
<td>TTTCCGTGGAGATTTCTGG</td>
<td>48°C</td>
<td>30</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>GAAGAGGTTGGCATCGAGGC</td>
<td>GGATCATCAGCTGCTTACGC</td>
<td>52°C</td>
<td>25</td>
</tr>
<tr>
<td>TBL1XR1</td>
<td>CTCTAGCATCAGGGTCTGG</td>
<td>GCAAACCCCAGTCAGGAAACC</td>
<td>49°C</td>
<td>30</td>
</tr>
<tr>
<td>CCND1</td>
<td>GTCCCCACTCTAGGATCCAGC</td>
<td>AGTGCTTGGAATTGGAATGG</td>
<td>64°C</td>
<td>23</td>
</tr>
<tr>
<td>NME1</td>
<td>ACATGGCAGTGATTCTGG</td>
<td>TCACAGCTCAAGGCTTCC</td>
<td>51°C</td>
<td>30</td>
</tr>
<tr>
<td>YY1AP1</td>
<td>CTGGTAAAGGCCAGTCTGC</td>
<td>TGGGTACCGAGTTTCATCC</td>
<td>52°C</td>
<td>25</td>
</tr>
<tr>
<td>DGCR8</td>
<td>GGAGAATGAGCTGGATCAGG</td>
<td>CTCTGTCTGCTCTTTTTGG</td>
<td>50°C</td>
<td>30</td>
</tr>
<tr>
<td>GAPDH</td>
<td>CAATGACCCCTCATTGACC</td>
<td>TGGATTTTGGAAGGATCTCG</td>
<td>48°C</td>
<td>30</td>
</tr>
<tr>
<td>pLVTHM</td>
<td>GCATGTGCTATGTTCTCTG</td>
<td>GCACGATAACTTGTTAAATG</td>
<td>44°C</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2.2. Qualitative PCR primers and their annealing temperatures and cycle numbers. All annealing temperatures were based on and compatible with BioMix Red (Bioline), the reaction mix that was used in performing all of the qualitative PCR reactions.
2.6.2 RNA Extraction

RNA extraction was performed on highly confluent, feeder-free MAN1, hUES1, hUES3 and hUES7 cells using the QIAGEN RNeasy Mini kit according to the manufacturer’s instructions. All centrifugations were undertaken at the same speed of 13000rpm. In brief, cells were harvested by trypsinization and resuspended in 3ml hESC FF media in a 15ml tube. The suspended cells were centrifuged for 5 minutes. The supernatant was discarded and the cells were resuspended in 600μl of RTL lysis buffer. This was vortexed for a minute to thoroughly mix the cells and the buffer. The mixture was then transferred into a QIAshredder spin column and centrifuged for 2 minutes. An equal volume of ethanol was then added to the mixture and this was mixed well by pipetting up and down. The mixture was then loaded into an RNEasy spin column and centrifuged for 15 seconds. The collection tube was emptied and 700μl of RW1 wash buffer was added to the RNEasy column to remove any contaminants from the membrane bound RNA. This was centrifuged for 15 seconds. The flow through was discarded and 500μl of RPE wash buffer, which helps to ensure that there is no ethanol carry over into the RNA, was added. The mixture was centrifuged for 2 minutes. Another 500μl of RPE wash buffer was added and again the mixture was centrifuged for 2 minutes. The RNEasy spin column was then transferred into a fresh 1.5ml collection tube. RNA was eluted from the spin column by the addition of 30μl of RNase-free water. The collection tube was centrifuged for 1 minute and the eluted RNA was stored at -81°C prior to its DNase treatment.

2.6.3 DNase Treatment of RNA

DNase treatment was performed on the eluted RNA to remove any contaminating DNA that may be present in the RNA. DNase treatment was carried out using the Promega RNA DNase treatment kit according to the manufacturer’s instructions. First, a digestion reaction was setup as follows;
<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA</td>
<td>30 μl</td>
</tr>
<tr>
<td>RQ1 RNase-free DNase x10 reaction buffer</td>
<td>4 μl</td>
</tr>
<tr>
<td>RQ1 RNase-free DNase</td>
<td>4 μl</td>
</tr>
<tr>
<td>Nuclease free water</td>
<td>2μl</td>
</tr>
<tr>
<td><strong>Total reaction volume</strong></td>
<td>40μl</td>
</tr>
</tbody>
</table>

Table 2.3. Digestion reagents and their volumes used for DNase treatment of RNA

The reaction mixture was then incubated for 30 minutes at 37°C. After incubation, in order to terminate the reaction, 4μl of RQ1 DNase stop solution was added. The mixture was then incubated for 10 minutes at 65°C on a heating block. RNA was quantified on a nanodrop (ND-1000 spectrophotometer, Lab Technology International) and stored at -80°C until it was reverse transcribed.

2.6.4 cDNA synthesis

Reverse transcription for cDNA synthesis was undertaken using the Promega M-MLV-RT reverse transcription kit according to the manufacturer’s instructions. First, 2μg of RNA was added into a fresh tube.

RNA volume equalling 2μg was calculated using the formula;

\[
\text{2μg} \quad \frac{\text{}}{\text{}}
\]

\[
\text{RNA concentration (ng/μl) (Nanodrop reading)}
\]

To the RNA tube, 0.5μg of random primers was added per 1μg of the RNA. RNase-free water was added to make a 15μl total volume. The mixture was then incubated on a heating block for 5 minutes at 70°C to melt the secondary structure within the template. After incubation, the tube was cooled on ice to prevent the re-formation of the secondary structure. The tube was then briefly centrifuged to collect the solution.
at the bottom of the tube. After centrifugation, 5μl of M-MLV 5x reaction buffer was added. A dNTP master mix was prepared by combining 10μl of each of the 4 dNTPs; dATP (100mM), dCTP (100mM), dGTP (100mM) and dTTP (100mM) with 60μl of RNase-free water to a 100μl final volume. From this dNTP master mix, 1.25μl was added to the RNA mixture. Twenty five units and 200units of recombinant RNasin ribonuclease inhibitor (Promega) and M-MLVRT enzyme were then added respectively. The mixture was mixed by flicking and the tube was incubated for 1 hour at 37°C. RNase-free water was then added to make a 25μl total volume. The cDNA was then stored at -21°C until use in PCR.

2.6.5 Gel preparation

1.5 % agarose w/v gel was prepared by adding 1.5g of agarose (Lonza, Switzerland) to 100ml 50X TAE buffer (Gibco) and heating the mixture in a microwave until the agarose powder was dissolved. The gel mixture was allowed to cool and 10μl of 10000x DNA dye SybrSafe (Invitrogen) was added to the gel. The gel was poured into the gel tray (Wolf Laboratories) and a pipette was used to remove bubbles and insoluble materials in the gel. The comb was inserted and the gel was allowed to set.

2.6.6 Gel electrophoresis

After the gel had set, 50X Tris-base, Acetic acid and EDTA (TAE) buffer was poured into the electrophoresis tank until the gel was completely covered. Five μl of the DNA HyperLadder I, II or IV ladder solution (Bioline Reagents, London, UK) was added to the first well of the gel for use as a molecular weight marker. Twelve μl of each of the PCR samples was loaded onto the other wells of the gel. Electrophoresis was performed at 120 voltage, 400 MA for 25 minutes. After completion of electrophoresis, the gel was taken to the imaging machine (Bio-Doc-It 220 Imaging System, UVP, Cambridge, UK) for gel imaging on an ultraviolet (UV) illuminated base.
2.7 Quantitative Real-Time PCR (qRT-PCR)

2.7.1 Primer design

Primer sequences for quantitative RT-PCR (Table 2.4) were designed using the Roche Universal Probe-Finder software (www.roche-applied-science.com). Coding DNA sequences of the genes were uploaded from the NCBI database and loaded into the software to look for primers that did not span an intron. Primers with base pair ranges of 17-27 and annealing temperature ranges of 56-62°C were designed and checked for sequence specificity with BLAST.
<table>
<thead>
<tr>
<th>Target Gene</th>
<th>Forward Primer (5’-3’)</th>
<th>Reverse Primer (5’-3’)</th>
<th>Amplicon Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>AGCCACATCGTCAGACAC</td>
<td>AATAGCACCAAATTCGGTGACT</td>
<td>62</td>
</tr>
<tr>
<td>CCND1</td>
<td>CCTGTCTACTCCAGCCCTCA</td>
<td>GCACTTTCGAGACACTTGA</td>
<td>61</td>
</tr>
<tr>
<td>Myst3</td>
<td>TCAAGAAGATGCTTTCTGTTT</td>
<td>TTGGACACCTGGTAAACAT</td>
<td>119</td>
</tr>
<tr>
<td>NME1</td>
<td>AATGACAGGAGGCCAGAC</td>
<td>TTTCCTCTGATCGTTCCTC</td>
<td>80</td>
</tr>
<tr>
<td>RBM15</td>
<td>TCCATATGAAAAGCTGGATG</td>
<td>GATCTGGGCCCACCAAGAG</td>
<td>75</td>
</tr>
<tr>
<td>TBL1X1</td>
<td>TGCTTTAATAAGTAAGAGAAGAG</td>
<td>TGGGTCTCCAAATATGTAGTC</td>
<td>76</td>
</tr>
<tr>
<td>TCF7L2</td>
<td>CGGCGGACTGTATGGAAGAAG</td>
<td>CATTAGCTTGTGTAGGTTT</td>
<td>76</td>
</tr>
<tr>
<td>YY1</td>
<td>GAGGTGATCCTGGTGCAGA</td>
<td>GATGAGCTCTAGAAGAGGCTC</td>
<td>85</td>
</tr>
<tr>
<td>YYAP1</td>
<td>TTGTCAAGATGAACCTGGAGA</td>
<td>TCGTCTCAAGAATATCTCAG</td>
<td>60</td>
</tr>
<tr>
<td>P53</td>
<td>GCAGACACTTGGCCAAAC</td>
<td>GCCAAGGCGGACGCTTA</td>
<td>109</td>
</tr>
<tr>
<td>TGFB1</td>
<td>GTCGGAAGAAGAGGCCATGGA</td>
<td>TTCCCTCTCATAGTTCTCCTC</td>
<td>75</td>
</tr>
<tr>
<td>Oct4</td>
<td>AGACCACCTGGCCCTTTTGA</td>
<td>GCCAAGGCGGACGCTTA</td>
<td>66</td>
</tr>
<tr>
<td>Nanog</td>
<td>CCTGTGATTGTGGGCTCTG</td>
<td>GAGGTCTGCTGAGGGCCAGACT</td>
<td>87</td>
</tr>
<tr>
<td>Sox2</td>
<td>AACCAGCGCATIGAGCAAGTTAC</td>
<td>TGGTGACTGATGCTAGT</td>
<td>77</td>
</tr>
<tr>
<td>c-Myc</td>
<td>CTGGAGAGGACATGTTGGAAC</td>
<td>TCGTTGATGATGCTAGT</td>
<td>65</td>
</tr>
<tr>
<td>Nkx2.5</td>
<td>CACCTCAACAGCCTCTCTGAC</td>
<td>AATGCAAAATCCAGGGGACT</td>
<td>75</td>
</tr>
<tr>
<td>GATA4</td>
<td>CCTCTCTGCTGCTGTAATGACT</td>
<td>CGCCTCCTGAAGGCCAGGAGA</td>
<td>98</td>
</tr>
<tr>
<td>Coll1</td>
<td>GCCCTGCTGGCTCTCTGTA</td>
<td>CTGCTGTTGTTGCTAAAGA</td>
<td>82</td>
</tr>
<tr>
<td>TBX3</td>
<td>GCAGGATTAGGTATTTAACAT</td>
<td>GCCAAGGAGCGCAGCTTA</td>
<td>60</td>
</tr>
<tr>
<td>Mixl1</td>
<td>AAGCAGCACTGGCCTTCTTT</td>
<td>CCGTTCACCAAGCCTTGG</td>
<td>81</td>
</tr>
<tr>
<td>Brachyury</td>
<td>GGCGTCTAGCCCACCTGAT</td>
<td>TGAATACGACTAGGGCTATGAA</td>
<td>132</td>
</tr>
<tr>
<td>Pax3</td>
<td>TTGGCAATGGGCCTTCAC</td>
<td>AGGGGAGAGCGCTTAATC</td>
<td>65</td>
</tr>
<tr>
<td>Nkx3.2</td>
<td>TGATTGGAGATGAGGATGGGAG</td>
<td>ACCCCACCTAGCTAGTGA</td>
<td>112</td>
</tr>
<tr>
<td>Sox1</td>
<td>GCCGAGGCTCTGCTGGGAG</td>
<td>CAGGTAACACTACAAAAACTTGA</td>
<td>76</td>
</tr>
<tr>
<td>Pax6</td>
<td>CTGTGCCTGCTGCTGCTGTA</td>
<td>CCCGTGACATCGTCTGCTGCT</td>
<td>74</td>
</tr>
<tr>
<td>BMP2</td>
<td>ACCAGAAGAGGATGGGGAAAA</td>
<td>CCAACCTGTTGCTCAAGAGT</td>
<td>81</td>
</tr>
<tr>
<td>CDX2</td>
<td>GACCTCTGCCACCACCTGTA</td>
<td>CTAAGGTACTAGTCGCTACT</td>
<td>92</td>
</tr>
<tr>
<td>Sox17</td>
<td>AGAGATTGTTTTGCCCATAGTTGGGAG</td>
<td>TGGTGCTGGAACATTCAGGAAGCT</td>
<td>94</td>
</tr>
<tr>
<td>Sox7</td>
<td>GCTCTTCTCTACGTCTCCTTCT</td>
<td>GTTTCAAGCTAGCTAGTTA</td>
<td>61</td>
</tr>
<tr>
<td>FoxA2</td>
<td>TTCCAGGCCCACCCACTGTA</td>
<td>AGTCTCAACCCCCACTTCTG</td>
<td>67</td>
</tr>
<tr>
<td>Chordin</td>
<td>GGTTGCTACCCCTCTGGAAGA</td>
<td>ACAGTGTTGTTACGGAAGGAC</td>
<td>111</td>
</tr>
<tr>
<td>TBX6</td>
<td>CCGGCGAAGCCGCAGAGA</td>
<td>CCGCGCAGTTTCCCTGCTCA</td>
<td>60</td>
</tr>
<tr>
<td>EOMES</td>
<td>CTGGGGGTCTCCAGGTTTCT</td>
<td>GTGGGCCTGGAATTTTGA</td>
<td>75</td>
</tr>
<tr>
<td>FGF5</td>
<td>CAAGTGTCGGAGAGGCTTTTC</td>
<td>AGGTGACAGAAGAGGGAATCT</td>
<td>105</td>
</tr>
<tr>
<td>BMP4</td>
<td>CCGCAGCCTCACTGGAAGT</td>
<td>CCAGACTGAAAGCGCTTAAGA</td>
<td>61</td>
</tr>
<tr>
<td>Beta II Tubulin</td>
<td>GGAACATAGCGCTAATGCG</td>
<td>TCACTGTGCTGACTACCAG</td>
<td>62</td>
</tr>
<tr>
<td>Nestin</td>
<td>CAGAGGCTCGGCCAGCTC</td>
<td>GAGGCGTCTTGTTGCTCTCTC</td>
<td>61</td>
</tr>
<tr>
<td>Noggin</td>
<td>CGGCACACAGCCGACAA</td>
<td>TCAAGATAGGCTGGTGTAGTC</td>
<td>60</td>
</tr>
<tr>
<td>AFP</td>
<td>CAAACAGGGGACCATGCT</td>
<td>GAGGACGGAGGACATATG</td>
<td>73</td>
</tr>
<tr>
<td>MAP2</td>
<td>CGAAGCTTTATATTTTTTACCTCCCTTGT</td>
<td>CGGTTCTACTGCCATTTTCTC</td>
<td>66</td>
</tr>
<tr>
<td>Beta III Tubulin</td>
<td>GCAACTCAGTTGGGGCACT</td>
<td>GAGGCGTCTGTTGCTAGGACG</td>
<td>77</td>
</tr>
</tbody>
</table>

Table 2.4 qRT-PCR primers and their amplicon sizes in base pairs.
2.7.2 qRT-PCR

qRT-PCR was conducted using GAPDH as an internal standard reference gene, human genomic (hg) DNA (Promega, Southampton, UK) as a positive control and PCR water as a negative control. First, a 1:10 working stock solution was prepared for both reverse and forward primers (Invitrogen), diluting them with PCR water. A cDNA concentration of 20ng/μl was used for each PCR reaction. Ten μl of 226μg/ml hg DNA and 90μl of PCR water were combined to make 100μl of 22.6μg/ml hgDNA which falls within the range recommended by the protocol. SYBR® Green I (Applied Bio Systems) was the fluorescence dye used in all qRT-PCRs carried out. SYBR Green fluoresces when bound to double-stranded DNA during the extension step of qRT-PCR. Therefore, unbound (non-amplified) DNA molecules will not fluoresce or will emit little fluorescence and therefore will not be detected. Figure 2.3 shows illustrate different phases of PCR amplification and dissociation curves.
Figure 2.3. PCR amplification and dissociation plots. A schematic showing; a) A real-time PCR amplification curve showing the different phases of DNA amplification. Delta Rn (normalised reporter) (Y-axis) is plotted against cycle number (X-axis) and a threshold (dashed line) is set. A point at which the curve crosses the threshold generates an individual Ct value which is representative of double stranded DNA. b) A dissociation curve showing a desired product with a primer dimer. All dissociation curves with primer dimers were ignored as they signalled non-specific binding.
For each gene, a qRT-PCR master mix was prepared as follows;

<table>
<thead>
<tr>
<th>Reagent</th>
<th>10 μl Reaction</th>
<th>25 μl reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power SYBER Green</td>
<td>5 μl</td>
<td>12.5 μl</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>0.4 μl</td>
<td>1 μl</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>0.4 μl</td>
<td>1 μl</td>
</tr>
<tr>
<td>PCR H₂O</td>
<td>2.2 μl</td>
<td>9.5 μl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>10 μl</td>
<td>24 μl</td>
</tr>
</tbody>
</table>

Table 2.5.qRT-PCR master mix reagents and their volumes per single 10μl and 25μl reactions

qRT-PCR reactions were carried out at either 10 or 25 μl final reaction volumes. Each reaction for each gene was performed in triplicates for each sample. For example, for a 25μl reaction, 24 μl of the master mix was added in triplicates into the wells of a 96-well PCR plate (Applied Biosystems, Paisley, UK). To each well, 1μl of an appropriate control sample (cDNA, hgDNA, PCR water) was added, also in triplicates, to make a 25μl total reaction volume. The plate was covered with an adhesive PCR film cover and centrifuged on a plate centrifuge (Sigma) to mix and collect samples at the bottom of the plate. The plate was loaded into an ABI 7500 PCR machine (Applied Biosystems). qRT-PCRs were performed under the following thermal profile unless otherwise stated;

<table>
<thead>
<tr>
<th>PCR Step</th>
<th>Time</th>
<th>Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Denaturation</td>
<td>X cycles, 1min</td>
<td>95°C</td>
</tr>
<tr>
<td>Denaturation</td>
<td>30s</td>
<td>95°C</td>
</tr>
<tr>
<td>Annealing</td>
<td>30s</td>
<td>X°C</td>
</tr>
<tr>
<td>Elongation</td>
<td>45s</td>
<td>72°C</td>
</tr>
<tr>
<td>Final Elongation</td>
<td>5mins</td>
<td>72°C</td>
</tr>
</tbody>
</table>

Table 2.6.General qRT-PCR thermal profile. Since different genes have different annealing temperatures, X cycles and X °C refer to specific cycle numbers and annealing temperatures for specific genes.
2.7.3 Data analysis

Data analysis was performed with an ABI 7500 or ABI 7300 System SDS Software (Applied Biosystems) which develops amplification plots reflecting the level of fluorescence emitted by Sybr green when double stranded DNA is produced.

2.8 Immunofluorescence staining

2.8.1 Single gene staining

All cells that were analysed for immunofluorescence staining were cultured in 24-well plates. Staining for cell surface markers and intracellular markers were performed in the way. The only difference was that triton, a membrane permeabilizer, was not used during staining for surface markers. In brief, cells were fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature. After fixation, PFA was discarded and the cells were washed 3 times with PBS solution (PBS, 0.05% Tween, 0.1% BSA; PAA Laboratories). Seventy-five µl of blocking/permeabilization solution (PBS, 10% serum, 0.1% Triton X) was then applied for 30 minutes at room temperature if appropriate. The serum used for the blocking solution was from the species that the secondary antibody was raised in. The blocking solution was then discarded. Seventy-five µl of primary antibody solution (Table 2.7) at specified dilutions in 1% serum, 0.1% Triton X for intracellular antigens and PBS solution were applied overnight at 4°C. Primary antibody dilutions were determined by dose response curves generated in our lab. Following overnight incubation, cells were then washed 3 times with PBS solution. Then, 75µl of secondary antibody solution (Table 2.8) at specified dilutions in 1% serum, +/-0.1% Triton X and PBS were applied for 1 hour at room temperature. Cells were then washed 3 times with PBS solution. 1:50000 dilute DAPI solution (Invitrogen, Oregon, USA) was applied for 5 minutes in the dark. Cells were then washed 3 times and immediately visualized with microscopy after the 3rd wash.
2.8.2 Co-localization staining

During co-localization staining, single staining procedure (2.8.1) was done for the first gene up to the secondary antibody application stage. This was then repeated for the second gene starting at the blocking step, using a primary antibody raised in a different species and its corresponding secondary antibody with a different Alexa fluor emission to the first secondary antibody gene. Cross reactivity between the primary and secondary antibodies was checked for by application of the secondary antibody alone as a control.

2.8.3 Microscopy

All cell culture phase contrast and bright field images were captured using the Leica DM IL LED inverted microscope (Olympus, Japan) connected to a DFC295 camera and operated by the Suite Application software (all from Leica, Germany). Fixed cell imaging for immunofluorescence microscopy was performed with an Olympus IX71 inverted microscope (Olympus, Japan) connected to a Q-Imaging Retiga SRV camera (QImaging, Canada). Images were processed with QCapture Pro software (QImaging, Canada). During immunofluorescence microscopy, the same exposure time was always used for the same antigen and antibody in every experiment.

<table>
<thead>
<tr>
<th>Primary Target</th>
<th>Antibody Species &amp; Isotype</th>
<th>Concentration</th>
<th>Dilution</th>
<th>Supplying Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct4</td>
<td>Mouse IgG</td>
<td>250µg/ml</td>
<td>1:100</td>
<td>BD Biosciences</td>
</tr>
<tr>
<td>Nanog</td>
<td>Goat IgG</td>
<td>100µg/ml</td>
<td>1:100</td>
<td>R&amp;D Systems Inc.</td>
</tr>
<tr>
<td>YY1</td>
<td>Rabbit IgG</td>
<td>200µg/ml</td>
<td>1:300</td>
<td>Santa Cruz Biotech.</td>
</tr>
<tr>
<td>YY1API</td>
<td>Goat IgG</td>
<td>200µg/ml</td>
<td>1:300</td>
<td>Santa Cruz Biotech.</td>
</tr>
<tr>
<td>Tra-1-60</td>
<td>Mouse IgM</td>
<td>200µg/ml</td>
<td>1:200</td>
<td>Abcam</td>
</tr>
<tr>
<td>Tra-1-81</td>
<td>Mouse IgM</td>
<td>200µg/ml</td>
<td>1:200</td>
<td>Abcam</td>
</tr>
</tbody>
</table>

Table 2.7 Primary antibodies used for immunofluorescence staining. Their species, isotypes, concentrations and suppliers are also shown.
<table>
<thead>
<tr>
<th>Secondary Antibody Target</th>
<th>Species</th>
<th>Concentration</th>
<th>Dilution</th>
<th>Supplying Company</th>
<th>Conjugation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Mouse IgG</td>
<td>Donkey</td>
<td>2mg/ml</td>
<td>1:200</td>
<td>Invitrogen</td>
<td>Alexa Fluor 488</td>
</tr>
<tr>
<td>Anti-Rabbit IgG</td>
<td>Goat</td>
<td>2mg/ml</td>
<td>1:200</td>
<td>Invitrogen</td>
<td>Alexa Fluor 488</td>
</tr>
<tr>
<td>Anti-Mouse IgM</td>
<td>Goat</td>
<td>2mg/ml</td>
<td>1:200</td>
<td>Invitrogen</td>
<td>Alexa Fluor 488</td>
</tr>
<tr>
<td>Anti-Goat IgG</td>
<td>Donkey</td>
<td>2mg/ml</td>
<td>1:200</td>
<td>Invitrogen</td>
<td>Alexa Fluor 594</td>
</tr>
<tr>
<td>Anti-Mouse IgM</td>
<td>Goat</td>
<td>2mg/ml</td>
<td>1:200</td>
<td>Invitrogen</td>
<td>Alexa Fluor 594</td>
</tr>
</tbody>
</table>

Table 2.8. Secondary antibodies used for immunofluorescence staining. Their species, isotypes, concentrations, suppliers and conjugations are also shown.

2.9 Plasmid vector cloning

2.9.1 shRNA designation, annealing and phosphorylation

DNA oligonucleotides comprising of short hair-pin RNA sequences were designed using the Dharmacon siDESIGN Center online tool (Thermo Fisher Scientific; [http://www.dharmacon.com/designcenter/DesignCenterPage.aspx](http://www.dharmacon.com/designcenter/DesignCenterPage.aspx)). This tool designs shRNAs that target specific gene variants and have low frequency seed regions to improve target specificity. A beta-2-microglobullin shRNA was also designed for use as a negative control. Each of the forward and reverse DNA oligonucleotide sequences designed comprised of both MluI (at the 5’ end) and Clal (at the 3’ end) restriction sites. They also had the shRNA sequence, the complementary antisense strand and a loop sequence that allows the hairpin loop to form. In brief, both YY1AP1 and beta-2-microglobullin mRNA complete coding sequences (cds) were copied from the NCBI database into the siDESIGN Center online tool. Homo sapien was selected as the sequence organism and the tool was instructed to select only sequences with 50-60% GC content. This is because DNA with a higher GC content is relatively more stable than DNA with a lower GC content. The designed shRNA were then checked for sequence specificity using BLAST ([http://blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)) and then
ordered from Thermo Fisher Scientific Laboratories. Below are the shRNA sequences that were designed;

**YY1API shRNA**

**shRNA sequence:**

\[
\begin{align*}
\text{F Strand:} & \quad Cgcgtcccc TCACCAAGGCTGAGGACAAttcaagaga TTGTCCTCAGCCTTGGTGAgaga \\
\text{R Strand:} & \quad Cgatttccaaaaa TCACCAAGGCTGAGGACAAttcaagaga TTGTCCTCAGCCTTGGTGAgaga \\
\end{align*}
\]

**Beta-2- Microglobulin (Negative control) shRNA**

**shRNA sequence:**

\[
\begin{align*}
\text{F Strand:} & \quad Cgcgtcccc CTCCAAAGATTCAGGTTTAttcagaga TAAACCTGAATCTTTGGAG \\
\text{R Strand:} & \quad Cgatttccaaaaa CTCCAAAGATTCAGGTTTAttcagaga TAAACCTGAATCTTTGGAG \\
\end{align*}
\]

The forward and reverse strands of each shRNA construct were annealed by preparation of the following reaction mixture;

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>shRNA forward strand</td>
<td>1µl</td>
</tr>
<tr>
<td>shRNA reverse strand</td>
<td>1µl</td>
</tr>
<tr>
<td>10 Ligase buffer (New England Biolabs)</td>
<td>5µl</td>
</tr>
<tr>
<td>PCR H_2O</td>
<td>43µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>50µl</td>
</tr>
</tbody>
</table>

Table 2.9.shRNA annealing reagents.

The reaction mixture was incubated for 10 minutes at 85°C and then cooled for 30 minutes at room temperature.

In order to phosphorylate each shRNA, the following reaction mixture was prepared;
<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4 PNK (New England BioLabs)</td>
<td>1µl</td>
</tr>
<tr>
<td>Annealed shRNA DNA</td>
<td>2µl</td>
</tr>
<tr>
<td>10X T4 PNK buffer (New England BioLabs)</td>
<td>5µl</td>
</tr>
<tr>
<td>10nm ATP</td>
<td>5µl</td>
</tr>
<tr>
<td>PCR H₂O</td>
<td>37 µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>50 µl</td>
</tr>
</tbody>
</table>

Table 2.10. shRNA phosphorylation reagents. PNK (Polynucleotide kinase) buffer is a phosphorylation buffer.

The mixture was then incubated for 30 minutes in a 37°C incubator.

**2.9.2 Vector cloning and restriction digest**

A commercially available inducible vector pLVTHM (R&D Systems) was used to transfet shRNAs into cells. This vector has an EF1-alfa promoter for driving GFP expression and a Tet0 promoter which, upon activation by doxycycline, activates the H1 promoter, which drives shRNA expression to initiate knockdown. The pLVTHM vector also has MluI and ClaI restriction end overhangs between which the shRNA sequence inserts. Additionally, it has an ampicillin resistant gene (AmpR) for marker selection. Before cloning, LB agar media was prepared by dissolving 20g of LB in 1000ml of H₂O. This was then microwaved until it started boiling. The lid of the bottle was loosened to avoid pressure building inside. The bottle was swirled periodically to mix the contents. Once molten, the bottle was placed in a pre-warmed 50°C water bath to cool down until it could be held comfortably with both hands. To prepare LB agar plates, LB agar medium was taken to a safety cabinet where 500µl of 50mg/ml ampicillin was added to 1000ml of LB agar medium. The bottle was inverted slowly a few times to mix the agar and the antibiotic. Forty petri dishes were laid out in the
safety cabinet. Roughly 10ml of molten agar was poured into each dish, completely covering the dish surface. The medium was poured carefully to avoid formation of bubbles. The dish lids were at an angle on the dishes, leaving them slightly ajar. The dishes were left to cool for 30 minutes before being placed in bags and stored in the cold room. In order to clone the pLVTHM plasmid vector (R&D Systems) into transforming bacterial E. coli cells, 100µl of competent XL-1 blue bacterial cells and 1.7µl of beta-mercapethanol (both from Agilent Technologies,) were aliquoted into pre-chilled 1.5ml tubes and incubated for 10 minutes on ice. Then, 1µl of vector, control and ligation mixture DNA was added to the mixture before another 30 minute incubation on ice. The mixture was then heat pulsed for 45 seconds at 42°C and incubated for 2 minutes on ice. Nine hundred µl of LB medium was then added and the mixture was immediately incubated for an hour at 37°C. After incubation, the mixture was centrifuged for 3 minutes at 500xg. The supernatant was discarded and 200µl of LB medium was added to the pellet. One hundred and fifty µl of the mixture was plated on pre-prepared LB agar ampicillin plates overnight at 37°C. After overnight incubation, one colony from the plate was transferred, using a pipette tip, into 100µl of LB medium for further expansion through overnight incubation at 37°C. DNA was isolated and purified from transformed bacterial competent cells using the QIAGEN Plasmid mini kit (Qiagen, Germany) according to the manufacturer’s instructions. In brief, 25ml of the overnight incubated culture mixture was transferred into a 50ml tube and centrifuged for 30 minutes at 3000xg to harvest bacterial cells. The supernatant was discarded and 4ml of buffer P1 was added. The tube was vortexed to destroy cell clumps and 4ml of buffer P2 was added. The tube was inverted 4-6 times to mix and incubated for 5 minutes at room temperature to lyse the cells. Four ml of pre-chilled buffer P3 was then added. The tube was inverted 4-6 times to mix and incubated on ice for 15 minutes. The tube was then centrifuged for 1 hour at 3000xg. The resulting supernatant was centrifuged for 30 minutes at 3000xg. Four ml of buffer QBT was added to a QIAGEN-tip 100 and the tip column was allowed to empty. The centrifuged supernatant was then applied to the column and allowed to enter the resin by gravity flow. The QIAGEN-tip was washed 2 times with 10ml of buffer QC. Plasmid DNA was eluted from the tip into a fresh 15ml tube by the addition of 5ml of buffer QF which had been pre-warmed to 65°C. Three and a half ml of isopropanol was added to the
DNA, mixed by pipetting and centrifuged for 1 hour at 3000xg. The supernatant was carefully decanted and 70% ethanol was added to the pellet and centrifuged for 20 minutes at 3000xg. The supernatant was discarded and the pellet air-dried for 10 minutes and then dissolved in 35µl of DNAse-free water. DNA yield concentration was measured on a nanodrop spectrophotometer.

Purified empty pLVTHM plasmid vector was digested by restriction enzymes by setting up the following reaction;

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClaI restriction enzyme (New England BioLabs)</td>
<td>1µl</td>
</tr>
<tr>
<td>MluI restriction enzyme (New England BioLabs)</td>
<td>1µl</td>
</tr>
<tr>
<td>Plasmid vector DNA</td>
<td>1µl</td>
</tr>
<tr>
<td>Buffer H (New England BioLabs)</td>
<td>3µl</td>
</tr>
<tr>
<td>PCR H2O</td>
<td>4µl</td>
</tr>
<tr>
<td><strong>Total reaction volume</strong></td>
<td>20µl</td>
</tr>
</tbody>
</table>

Table 2.11. Vector restriction digest reagents.

The reaction mixture was incubated for 8 hours (or overnight) at 37°C.

**2.9.3 Gel purification and vector de-phosphorylation**

After digestion, the plasmid vector was purified by running it through 1.5 % agarose gel (see section 2.6.6. for gel preparation). Ten ml of vector DNA (mixed with loading buffer at a 1:1 dilution ratio) was loaded into the gel and ran for 25 minutes at 120 volts (as in section 2.6.7). The digested plasmid DNA was ran on the gel in parallel with the undigested vector DNA. Correct size DNA bands were visualized on a UV illuminator bed and cut out with a sharp clean scalpel into 1.5ml microfuge tubes. Long wavelengths were used during visualization to minimize DNA damage. The bands were
weighed on a microbalance and purified with the QIAGEN QIAEXII gel extraction kit (Qiagen, Germany) according to the manufacturer’s instructions. Bands were weighed in order to determine which gel extraction kit to use. Gel purification was performed the same way for each band sample. All centrifugation during gel purification was performed at the same speed of 17900xg. In brief, 3 volumes of buffer QX1 was added and vortexed for 30 seconds. Thirty µl of QIAEXII was then added and incubated for 10 minutes at 5°C in order to solubilize the agarose. The tube was vortexed every 2 minutes. The sample was centrifuged for 30 seconds and the supernatant was discarded with a pipette. The pellet was washed with 500µl of buffer QX1 and resuspended by vortexing. The sample was then centrifuged for 30 seconds and all the supernatant was removed with a pipette. The pellet was washed twice by resuspension in 500µl buffer PE, vortexing and centrifugation for 30 seconds. The resulting pellet was air dried for 10-15 minutes. DNA was eluted from the pellet by the addition of 20µl of DNase-free water, vortexing and incubation for 5 minutes at room temperature. After incubation, the tube was centrifuged for 30 seconds and the supernatant (DNA) was carefully pipetted into a clean tube and stored at -21°C.

To de-phosphorylate the purified vector, the following reaction was set-up;

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antarctic phosphatase (New England BioLabs)</td>
<td>1µl</td>
</tr>
<tr>
<td>10x Antarctic buffer (New England BioLabs)</td>
<td>2µl</td>
</tr>
<tr>
<td>Vector DNA</td>
<td>3µl</td>
</tr>
<tr>
<td>PCR H₂O</td>
<td>14µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>20µl</td>
</tr>
</tbody>
</table>

Table 2.12. Vector phosphorylation reagents.

The reaction mixture was incubated for 15 minutes at 37 °C and heat inactivated at 65 °C for 5 minutes.
2.9.4 Ligation and transformation

To ligate each of the shRNA constructs to the digested plasmid vector DNA, the following digestion reaction was setup;

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4 DNA ligase (New England BioLabs)</td>
<td>1µl</td>
</tr>
<tr>
<td>Vector DNA (50ng)</td>
<td>1µl</td>
</tr>
<tr>
<td>shRNA DNA (50ng)</td>
<td>3µl</td>
</tr>
<tr>
<td>PCR H₂O</td>
<td>5µl</td>
</tr>
<tr>
<td>2X quick ligation buffer (New England BioLabs)</td>
<td>10µl</td>
</tr>
<tr>
<td>Total reaction volume</td>
<td>20µl</td>
</tr>
</tbody>
</table>

Table 2.13.shRNA and plasmid vector ligation components.

The reaction mixture was incubated at room temperature for 5 minutes. The plasmid construct ligation mixture was then transformed in bacterial competent E.coli cells and purified with the QIAGEN mini kit as described in section 2.9.2.

2.9.5 DNA sequencing of ligated vector

An H1 sequencing forward primer (table 2.15) was ordered from Addgene. Three hundred ng of the plasmid DNA was mixed with 1µl of the H1 (20µM) forward sequencing primer (Table 2.15.). The volume was made up to 10µl by adding Milli-Q deionised water. Sequencing was done at the Manchester University DNA sequencing facility using an ABI 3100 sequencer. Analysis of sequenced shRNAs and vector was done using Vector NTI 11.5 (Invitrogen, USA).

<table>
<thead>
<tr>
<th>Sequencing Primer</th>
<th>Forward Primer (5’-3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 (pLVTHM)</td>
<td>TCGCTATGTTCTGGGAAATC</td>
</tr>
</tbody>
</table>

Table 2.14.H1 forward primer sequence.
2.10 Transfection and knockdown

2.10.1 Calcium phosphatase transfection of 293FTs

Cells to be transfected were cultured and passaged as described in section 2.1.2. The following media were prepared: MEF medium (table 2.1), transfection medium (50ml of DMEM, PAA Laboratories Inc., 1.25ml of 25nM hepes buffer, Gibco), base medium (30ml of DMEM, PAA Laboratories Inc., 750µl of 25nM hepes buffer, Gibco). Using sodium hydroxide (NaOH) and hydrochloric acid (HCl), pH levels of the transfection and base media were adjusted to 7.1 and 7.9 respectively. CaCl$_2$(1M) was prepared by dissolving 7.5g of dehydrate calcium chloride(CaCl$_2$-2H$_2$O) in 50ml of deionised water. The evening before transfection, 293FT cells were passaged using trypLE$^\text{Tm}$ Express (as in section 2.1.2) to a density of 6 x 10$^5$ cells per T75 flask. They were incubated in a 37°C, 5% CO$_2$ incubator. The following day, transfections were setup. In brief, 13.5µg of plasmid DNA and pmaxGFP control plasmid were added into 2 separate fresh 1.5ml tubes. The volume in each tube was made up to 1.29ml by the addition of transfection medium. Seventy µl of 1M CaCl$_2$ was added to each tube and the tubes were swirled to mix. The tubes were then incubated for 20 minutes at room temperature. Media was discarded from 293FT flasks. The cells were washed 2 times with PBS. Then, 5.5ml of base medium was added to each flask of cells. Plasmid and control plasmid transfection mixtures were then added to two separate 293FT cell flasks. GFP positive cells were observed at least 24 hours post transfection under an inverted microscope (Leica Microsystems, Germany).

2.10.2 Lipofectamine transfection of MCF-7s

Cells to be transfected were cultured and passaged as described in section 2.1.2. The day before the transfection, cells were passaged using trypLE$^\text{Tm}$ Express (section 2.1.2) and plated to a density of 2 x 10$^5$ cells per well of a 6-well culture plate. Only MCF-7 cells with at least 50% confluency were transfected. In brief, 2.5µg of plasmid DNA and pmaxGFP plasmid were each added to 500µl of opti-MEM medium (Gibco, USA) in two separate 1.5ml tubes. To each tube, 6.25 – 11.25µl of DNA lipofectamine RNai/Max/LTX complexes (Invitrogen, USA) was added and the mixtures mixed by
gentle pipetting. The tubes were then incubated for 25 minutes at room temperature to form lipofectamine complexes. After incubation, MEF medium was discarded from the MCF-7 cells and replaced with 2ml of fresh medium. Then, 100µl of the lipofectamine complex mixtures were directly added to two separate wells of the plate and the plate was gently rocked to mix. The plate was incubated for at least 24 hours at 37°C, 5%CO₂ before observation of GFP positive cells under an inverted microscope (Leica).

2.10.3 Nucleofection of hESCs

HUES1 and hUES3 cells to be nucleofected were cultured and harvested as described in sections 2.2.1 – 2.2.2. After harvesting, cells were counted with a haemocytometer. Then, 5 x 10⁶ cells were added to 100µl of P3 nucleofector solution (Lonza, Belgium) in two separate tubes. The nucleofector machine was switched on and 5µg of plasmid and pmaxGFP control plasmid were added to each tube. The mixtures were transferred into 100µl nucleofector curvettes. The curvettes were tapped on the table to break any bubbles. They were then transferred to the nucleofector machine and nucleofection was run on the X-Core programme. After Nucleofection, 1ml of mTESR1 medium (table 2.1), supplemented with 1µg/ml of Rho-associated protein kinase (rock inhibitor; Sigma) was added to the nucleofection curvettes and the transfected cells were transferred to pre-prepared fibronectin (section 2.2.2) coated plates containing 2ml of mTESR1 medium. ROCK inhibitor improves cell attachment and survival after nucleofection. Nucleofected cells were cultured for at least 24hours before being observed for GFP positive cells under an inverted microscope (Leica).

2.10.4 DOX-induced knockdown of MCF-7s, hESCs and iPSCs

The plasmid (pLVTHM) (Figure 2.4.) used for target gene knockdown was inducible with a Tet0 promoter in its backbone. Therefore, knockdown was activated by the addition of doxycycline (DOX; Sigma) which binds to the Tet0 promoter to activate shRNA expression at the H1 promoter. Briefly, 1µg of DOX was added to 1ml of Dimethyl Sulfoxide (DMSO; Sigma). The DOX-DMSO mixture was then added to MEF medium (for knockdown in MCF-7s) and mTESR1 medium (for knockdown in hES and iPS cells) at a ratio of 1:1000. The DOX containing respective media were then used to
feed nucleofected MCF-7, hES and iPS. DOX was maintained at a similar concentration (1μg/ml) in all media throughout the knockdown experiments.

2.10.5 EB formation from knocked down cells

Knocked-down GFP positive cells were harvested as described in section 2.2.2. and re-suspended in 3ml of MEF medium (table 2.1). The cells were counted with a haemocytometer. Then, 20μl of the cell suspension containing approximately 5000 cells was pipetted into a 10cm culture dish to form a droplet. This was repeated until about 30 droplets of cells had been pipetted onto the culture dish. The dish was covered with a lid, inverted upside down and incubated overnight at 37°C, 5% CO₂. After overnight incubation, 15ml of MEF medium was added to the culture dish and the EBs were allowed to grow for a period of 10 days with medium being changed every 4 days (section 2.1.3). RNA was collected at Days 0, 5 and 10 and treated with DNase then reversed transcribed for qRT-PCR analysis (sections 2.6.3 – 2.6.5).
Figure 2.4. pLVTHM plasmid vector. Used for YY1API shRNA nucleofection into hESCs, the vector has an EF1-alfa promoter for driving GFP expression and a Tet0 promoter which, upon Dox addition, activates shRNA expression. The plasmid also has an H1 promoter for driving shRNA expression as well as MluI and ClaI restriction sites for shRNA insertion.
2.11 Flow cytometry

2.11.1 Cell sorting

HESCs to be sorted were cultured in 6-well plates and passaged as described in sections 2.2.1 – 2.2.2. First, media was removed from the cells and they were washed 2 times with PBS- (PAA). Each well was then treated with 1ml of trypLE\textsuperscript{TM} for 2 minutes at 37\degree C, 5\% CO\textsubscript{2}. After incubation, the cells were washed off from the plates with hESC FF medium and centrifuged for 4 minutes at 600xg. The supernatant was discarded and the pellet was gently broken by hand vortexing to create a single cell suspension. Medium was added to the cells and they were counted with a haemocytometer. Then, 2 ml of 5 – 10 x 10\textsuperscript{6} cells per ml was suspended in media for each cell type and treated with penicillin/streptomycin at a 1:200 concentration ratio. Hepes buffer (25mM), which helps to maintain pH constant during sorting, was added to the media at a 1:200 ratio. The cell solution was then filtered with a 50-micron filter (BD Biosciences) and collected in 5ml polypropylene tubes. Samples were then taken for sorting and analysis using a Beckman Coulter Cyan ADP machine (Beckman Coulter, CA, USA). Output processing and analysis were performed with the Summit v4.3 program (Beckman Coulter). The sorted GFP positive cells were collected in fresh 5ml tubes and plated in pre-prepared matrigel coated 6-well plates in 3ml of hESC FF medium. Rock inhibitor was added to the medium at a 1:1000 concentration ratio. Cells were cultured for at least 24hours at 37\degree C, 5\% CO\textsubscript{2} before yield assessment through microscopic viewing of GFP positive cells.

2.12 Western Blotting

2.12.1 Harvesting of cells

At least 1 x 10\textsuperscript{6} cells were used for western blotting. These cells were typically cultured in 6-well plates. Protease (1x) and 1x phosphatase inhibitor cocktail tablets (both from Roche Applied Science) were added to Radio-Immunoprecipitation Assay (RIPA) lysis buffer (Sigma) to a 1x final concentration. The buffer was constantly kept on ice during cell harvesting. In brief, media was removed from the cells and they were washed 3
times with ice cold PBS+ (PAA Laboratories Inc.). The cells were then treated with 120μl of ice-cold RIPA lysis buffer, with swirling of the plate to spread the buffer. A cell scraper or a pipette tip was used to scrap off the cells from the plate. The cell mixture was transferred into a fresh tube and incubated for 30 minutes on ice. The cells were then centrifuged for 1 hour at maximum speed at 4°C. The supernatant (proteins) was transferred into a fresh 1.5ml tube and stored at -80°C.

2.12.2 BCA protein quantification

The Bicinchonic Acid (BCA™) total protein assay kit (Thermo Scientific) was used to determine the protein concentration of the cell lysate according to the manufacturer’s instructions. BSA was dissolved in RIPA lysis buffer at the following concentrations to make a series of protein standards; 2000μg/ml, 1000μg/ml, 750μg/ml, 500μg/ml, 250μg/ml, 125μg/ml and 0μg/ml. BCA reagent A was mixed with BCA reagent B at a 50:1 ratio to make a BCA working reagent. Then, 300μl of the working reagent was added to each protein sample and the standards. These were then incubated in the dark for 30 minutes at 37°C. After incubation, 100μl of the mixture was added in triplicates to a 96-well microplate (Thermo Scientific) and the absorbance was measured at 562nm using a microplate spectrophotometer (Bio-Tech). Basically, the difference between the average absorbance of the 0μg/ml BSA solution and the average absorbance of each sample and standard was calculated. A standard curve was then prepared by plotting the average 562nm absorbance measurement of each BSA standard against its μg/ml concentration (Figure 2.5).
Figure 2.5. BSA standard curve. A standard curve showing known standard BSA concentrations plotted against their absorbances at 562nm. The generated standard curve was used to determine unknown protein concentrations from known absorbance values.
2.12.3 Preparation of SDS gels and western blotting

First, a gel for electrophoresis was prepared. Separation and stacking gels were prepared as follows;

**Separation gel**: For making 2 gels using 1.5mm spacers.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>4%</th>
<th>5%</th>
<th>7%</th>
<th>10%</th>
<th>*12%</th>
<th>15%</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.5% acrylamide (ml)</td>
<td>2.0</td>
<td>2.5</td>
<td>3.5</td>
<td>5.0</td>
<td>7.0</td>
<td>7.5</td>
</tr>
<tr>
<td>(National Diagnostics, UK)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5M Tris pH 8.8 (ml)</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>Water (ml)</td>
<td>13.0</td>
<td>12.5</td>
<td>11.5</td>
<td>10.0</td>
<td>9.0</td>
<td>7.3</td>
</tr>
<tr>
<td>10% APS (μl) (table 2.20)</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>20% SDS (μl) (table 2.20)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>TEMED (μl) (Bio-Rad)</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 2.15 Polyacrylamide gel electrophoresis separation gels. The gels are shown at different percentages with their components volumes.

*Typically, 12% separation gels were prepared for all western blots undertaken.

For the separation gel, gel mixture was vortexed to mix and then pipetted between the glass plates 1cm below the comb positions. Water was added to the top of the glass plates to remove bubbles. The water was poured out after the gel had set and a stacking gel was prepared as follows;
Stacking gel: For making 2 gels using 1.5mm spacers.

<table>
<thead>
<tr>
<th>Percentage gel</th>
<th>4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.5% acrylamide (ml) (National Diagnostics, UK)</td>
<td>0.6</td>
</tr>
<tr>
<td>0.5M Tris pH 6.8 (ml)</td>
<td>1.5</td>
</tr>
<tr>
<td>Water (ml)</td>
<td>3.9</td>
</tr>
<tr>
<td>10% APS (μl) (table 2.20)</td>
<td>50</td>
</tr>
<tr>
<td>20% SDS (μl) (table 2.20)</td>
<td>30</td>
</tr>
<tr>
<td>TEMED (μl) (Bio-Rad)</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2.16. Polyacrylamide gel electrophoresis stacking gel components.

For the stacking gel, the gel was pipetted into the glass plates all the way to the top. The comb was immediately put in place and the gel was allowed to set. Meanwhile, 2μg of each sample protein lysate was transferred into a 1.5ml tube. An equal volume of reducing sample buffer (Bio-Rad Laboratories Inc., CA, USA) was then added. The total volume was made up to 20μl by the addition of a sample buffer and PBS- mixture at a 1:1 dilution ratio. The samples were then boiled for 10 minutes on a heating block 70°C in order to denature the proteins. The tubes were spun down to collect the condensate. Polyacrylamide electrophoresis apparatus (Bio-Rad Laboratories Inc., USA) containing the gel was then set-up. The gel was put in the tank and the inside chamber of the tank was filled up with 1x running buffer (see Table 2.20) to check for leakages. The outside chamber of the tank was then half filled. Each well of the gel was washed out with running buffer. Six μl of loading buffer (New England BioLabs) was added to the first well of the gel and samples were added to the remaining wells. All empty wells were loaded with 20μl of the sample buffer: PBS mixture in order to stop the gel from forming a smile shape. Electrophoresis was run at 100V until the dye front had run down the gel about an inch. After electrophoresis, blotting was undertaken to transfer proteins from the gel into a nitrocellulose membrane (GE HealthCare Life
Sciences, UK). A blotting sandwich was built on the black panel of the blotting apparatus. A sponge that had been pre-soaked in transfer buffer (see Table 2.20) was placed first on the panel and rolled out to remove bubbles. Two pre-soaked blotting papers (Thermo Scientific) were then placed on top of the sponge. The gel was removed from the mount and placed on top of the blotting papers. The nitrocellulose gel was placed on top of the gel followed by two soaked blotting papers and a sponge respectively. The sandwich was closed and put back into the tank with the black (negative) side of the sandwich facing the black (negative) side of the tank. The tank was filled up with transfer buffer and a stirrer was added. The tank was put in a container filled with ice. Blotting was then performed at 100V for 1 hour on a magnetic stirrer. After blotting, the membrane was stained with ponceau stain for 5 minutes to visualize the protein. The membrane was then washed with tap water and blocking was performed by the addition of a blocking solution (Odyssey, Li-Cor Biosciences, USA) for 1 hour at room temperature. After blocking, the membrane was washed 2 times with TBS-T buffer (Table 2.20). A primary antibody (Table 2.18.) was then applied at a specified dilution in blocking buffer (Li-Cor Biosciences, USA) overnight at 4°C. The membrane was then washed 3 times with TBS-T buffer and a secondary antibody (Table 2.19) was then applied at a specified dilution in blocking buffer for 1 hour at room temperature in the dark. After secondary antibody application, the membrane was washed 3 times with TBS-T buffer and taken for protein visualization with the Odyssey Infrared Imaging System machine (Li-Cor Biosciences, USA). This system has an in-built quantification system which was used during densitometry to analyse band intensity.
<table>
<thead>
<tr>
<th>Target Protein</th>
<th>Species &amp; Isotype</th>
<th>Concentration</th>
<th>Dilution</th>
<th>Supplying Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>Rabbit IgG</td>
<td>100μg/ml</td>
<td>1:1000</td>
<td>Cell Signalling Tech.</td>
</tr>
<tr>
<td>YY1API</td>
<td>Goat IgG</td>
<td>200μg/ml</td>
<td>1:500</td>
<td>Santa Cruz Biotech.</td>
</tr>
<tr>
<td>YY1</td>
<td>Rabbit IgG</td>
<td>200μg/ml</td>
<td>1:500</td>
<td>Santa Cruz Biotech.</td>
</tr>
<tr>
<td>Nanog</td>
<td>Rabbit IgG</td>
<td>100μg/ml</td>
<td>1:2000</td>
<td>Cell Signalling Tech.</td>
</tr>
<tr>
<td>P53</td>
<td>Rabbit IgG</td>
<td>100 μg/ml</td>
<td>1:1000</td>
<td>Cell Signalling Tech.</td>
</tr>
<tr>
<td>CTCF</td>
<td>Rabbit IgG</td>
<td>100 μg/ml</td>
<td>1:1000</td>
<td>Cell Signalling Tech.</td>
</tr>
<tr>
<td>Oct4</td>
<td>Mouse IgG</td>
<td>100 μg/ml</td>
<td>1:1000</td>
<td>Cell Signalling Tech.</td>
</tr>
</tbody>
</table>

Table 2.17. Primary antibodies used for western blotting.

<table>
<thead>
<tr>
<th>Target Protein</th>
<th>Species &amp; Isotype</th>
<th>Concentration</th>
<th>Dilution</th>
<th>Supplying Company</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-Goat IgG</td>
<td>Donkey</td>
<td>0.5mg/ml</td>
<td>1:3000</td>
<td>Li-Cor Biosciences</td>
</tr>
<tr>
<td>Anti-Rabbit IgG</td>
<td>Goat</td>
<td>0.5mg/ml</td>
<td>1:3000</td>
<td>Li-Cor Biosciences</td>
</tr>
<tr>
<td>Anti-Mouse IgG</td>
<td>Donkey</td>
<td>0.5mg/ml</td>
<td>1:3000</td>
<td>Li-Cor Biosciences</td>
</tr>
</tbody>
</table>

Table 2.18. Secondary antibodies used for western blotting.
<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Components</th>
<th>Preparation Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x running buffer</td>
<td>30.3g Tris Base (10x conc, 250mM)</td>
<td>Volume was made up to 900ml with de-ionised water</td>
</tr>
<tr>
<td></td>
<td>144.0g Glycine (10x conc. 1.92M)</td>
<td>pH was then adjusted to 8.3</td>
</tr>
<tr>
<td></td>
<td>10g SDS (1%)</td>
<td>Volume was made up to 1L with de-ionised water</td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 900ml with de-ionised water</td>
<td>and buffer stored at room temperature</td>
</tr>
<tr>
<td></td>
<td>100ml stock 10x running buffer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 1L with de-ionised water</td>
<td>and buffer stored at room temperature</td>
</tr>
<tr>
<td>1x transfer buffer</td>
<td>30.3g Tris base (10x conc, 250mM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>144.1g Glycine (10x conc, 1.92M)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 1L with de-ionised water</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and buffer stored at room temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100ml 10x stock transfer buffer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>200ml methanol</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1g SDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 1L with de-ionised water</td>
<td>and buffer stored at room temperature</td>
</tr>
<tr>
<td>10x TBS-T buffer</td>
<td>24.2g Tris Base</td>
<td></td>
</tr>
<tr>
<td></td>
<td>80g NaCl</td>
<td></td>
</tr>
<tr>
<td></td>
<td>900ml of deionised water was added</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH was adjusted to 7.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15ml of TWEEN 20 was added</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 1L with de-ionised water</td>
<td>and buffer stored at room temperature</td>
</tr>
<tr>
<td>4x Sample buffer</td>
<td>3.028g TRIS base</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40g glycerol (31.2ml)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 90ml with de-ionised water</td>
<td>pH was adjusted to 6.8</td>
</tr>
<tr>
<td></td>
<td>9.2g of SDS was then added</td>
<td>Volume was made up to 100ml with de-ionised water</td>
</tr>
<tr>
<td></td>
<td>Volume was made up to 100ml with de-ionised water</td>
<td>and buffer stored at room temperature</td>
</tr>
<tr>
<td></td>
<td>*Each time before use, 5% β-mercaptoethanol was added to the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>buffer before addition to the protein lysate. e.g. 50µl to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>950µl 4x sample buffer</td>
<td></td>
</tr>
<tr>
<td>10% Ammonium Per-Sulphate (APS)</td>
<td>1mg of APS was added to 10ml of de-ionised water</td>
<td>This was aliquoted into 10 x 1ml and stored at -20°C</td>
</tr>
</tbody>
</table>

Table 2.19 Western blotting reagents, buffers and their components.
2.13 Immunoprecipitation (IP)

2.13.1 Collection of samples

Immunoprecipitation was performed using the Abcam (Cambridge, UK) immunoprecipitation protocol according to the manufacturers’ instructions. A minimum of $1 \times 10^6$ cells that had been cultured in 6-well plates (section 2.2.2) were used during immunoprecipitation. To collect samples, the cell culture dish was placed on ice and washed with ice cold PBS. The PBS was then discarded and 1μl of ice cold lysis buffer (20 mM Tris HCl pH 8, 137 mM NaCl, 1% Nonidet P-40 (NP-40), 2 mM EDTA) was added to each well of cells. The cell mixture was then transferred into a pre-cooled centrifuge tube and centrifuged for 20 minutes at 12000 rpm at 4°C. After centrifugation, the tube was briefly incubated on ice and the supernatant was transferred into a fresh tube and the pellet discarded. In order to reduce non-specific binding of the proteins to the agarose beads, the lysate was pre-cleared using an irrelevant antibody to get rid of non-specific binding proteins. This was done by adding 50μl of an irrelevant antibody of the same species and Isotype as the target gene antibody to 1ml of the lysate and incubating the mixture for 1 hour on ice. Then, 100μl of protein G bead slurry was added to the lysate and this was incubated for 20 minutes at 4°C with the tube being occasionally tapped. The tube was then centrifuged for 10 minutes at 14000xg at 4°C. The supernatant was transferred into a fresh tube for use during immunoprecipitation and the pellet was discarded.

2.13.2 Running and analysis

Immunoprecipitation was performed by adding 50μg of the cell lysate to 1μg of the antibody. This was then incubated overnight at 4°C on an eppendorf rotator. Meanwhile, beads were prepared by incubating 100mg of protein-G beads in 1ml of 0.1M PBS. This was then washed for 1 hour with a wash buffer (10mM Tris; adjust to pH 7.4, 1mM EDTA, 1mM EGTA; pH 8.0, 150mM NaCl, 1% Triton X-100, 0.2mM sodium ortho-vanadate) and centrifuged. The supernatant was discarded. One ml of PBS with 0.1% BSA was added to the pellet and mixed for 1 hour on an eppendorf rotator. This was then rinsed 2 times in PBS. The supernatant was removed and 400μl
of lysis buffer containing protease inhibitors which help to slow down proteolysis was added. The slurry was mixed well and 100µl of the beads was added to each sample, keeping the samples on ice all the time. The lysate beads mixture was then incubated for 4 hours at 4°C on an eppendorf rotator. After incubation, the tube was centrifuged and the supernatant was removed from the beads and discarded. The beads were then washed with washing buffer 3 times to remove non-specific binding by mixing the beads gently with the buffer and centrifuging at 4°C and then discarding as much of the supernatant from the beads as possible. In order to elute the protein complex from the beads, 150 µl glycine elution buffer (0.2 M glycine pH 2.6) was added to 50µl of the beads and this was incubated for 10 minutes with frequent gentle agitation followed by gentle centrifugation. The eluate was then pooled together and an equal volume of Tris pH 8.0 was added. Elution of proteins from the beads was repeated 2 more times in order to collect all the eluate. In order to neutralize the beads, they were washed 2 times with 150µl of lysis buffer and the supernatant was pooled together with the eluate. The eluate samples were then ran on a western blot (section 2.12.2 – 2.12.3) to check for the precipitation of proteins.

2.14 Chromatin Immuno Precipitation Sequencing (ChIP-Seq)

2.14.1 Cross-linking

ChIP was performed on at least 20 x 10⁶ cells. Chromatin was cross-linked by the application of formaldehyde directly into the tissue culture media to a 1% final concentration. This was then left for 12 minutes at room temperature with slow agitation. Glycine was then added to a final concentration of 0.125M and this was left for 5 minutes at room temperature with slow agitation. Cell culture plates were rinsed twice with ice-cold PBS. Cells were then scraped off the plates with ice-cold PBS containing protease inhibitors (PI) and transferred into a fresh pre-chilled 15ml falcon tube. The tube was centrifuged for 5 minutes at 1650 RPM at 4°C. After centrifugation, the supernatant was discarded and the pellet was re-suspended in 4ml of cell lysis buffer (see Table 2.21) containing PI. This was then incubated on ice for 20 minutes. After ice incubation, the cells were centrifuged for minutes at 4000 RPM at 4°C. The supernatant was discarded and the pellet re-suspended in 400µl of nuclei lysis buffer
(see Table 2.21) containing PI. This was incubated on ice for 20-30 minutes. Using the Branson Sonifier 250, chromatin was sonicated to an average length of about 300bp, which is the length of de-cross-linked chromatin. During sonication, the instrument was set at 0.5 on/0.5 off, 2% amplitude. The sample was sonicated for 1 minute and then incubated on ice for 1 minute. This was repeated once. To check if sonication was successful, a 10µl aliquot was taken and added to 250µl of elution buffer and this was incubated at 65°C overnight to reverse cross-link. 1µl of Ambion RNase cocktail AM2288 was then added and this was incubated for 1hour at 37°C. After incubation, 10µl of Proteinase K (10mg/ml) was added and this was incubated for 2hours at 42°C. Extraction was done once by adding 500µl of phenol/chloroform. 900µl of freshly prepared precipitation mix (900 EtOH+1µl glycogen) was added and frozen for 30 minutes at -80°C. This was then centrifuged for 30 minutes at 14000 RPM at 4°C. After centrifugation, the supernatant was discarded and the pellet was washed with 1ml of 70% Ethanol and air-dried. The pellet was then re-suspended in 10µl of milliQ ultrapure H₂O. Fragment length was then checked on 1.5% agarose gel (as in 2.6.6 – 2.6.7). After conformation of chromatin sonication, the sample was centrifuged for 10 minutes at 14000 RPM at 4°C to eliminate debris. Chromatin was then quantified and the sample was diluted to the same concentration by the addition of nuclei lysis buffer containing PI.

2.14.2 Pre-clearing

To dilute SDS, chromatin was diluted in 10 volumes of IP dilution buffer (see Table 2.21) containing PI. Fifty µl of protein G magnetic beads (Invitrogen) was added. This was then incubated for 45mins-1hour on a rotating wheel at 4°C. Beads were then discarded by placing the tube on a magnetic rack. 1% of the chromatin was set aside for input control and stored at-80°C.

2.14.3 Immunoprecipitation, Washes and DNA Purification

During immunoprecipitation, 120-150µg of chromatin was used per IP and a mock control samples were prepared by incubation with a control IgG and incubation without an antibody. ON was incubated at 4°C on the wheel. 50µl of G magnetic beads (Invitrogen) which had been pre-equilibrated in IP dilution buffer (containing PI) was
added to the sample and this was incubated for 2 hours at 4°C on the wheel. Beads were then recovered on a magnetic rack and washed. The following washes were performed on the beads for 5 minutes at 4°C on the wheel:

- 1 low salt
- 1 high salt
- 1 low salt
- 2x TE

Reverse formaldehyde crosslink and elution of the antibody/protein/DNA complexes were performed by the addition of 250µl IP elution buffer and this was incubated overnight at 65°C. Samples were flicked every 10 minutes for the first 30 minutes of incubation. Reverse cross-linkage was also performed for the input samples. After overnight incubation, samples were briefly centrifuged and 250µl of TE 1x and 5µl of Ambion RNase cocktail buffer were added. The samples were incubated for 1 hour at 37°C. Then, 10µl of Proteinase K (10mg/ml) was added to the mixture and this was incubated for 3 hours at 42°C. After incubation, 55µl of LiCl was added to the sample. Extraction was done with 500µl of phenol/chloroform. 900µl of freshly prepared precipitation mix (900 EtOH+1µl glycogen) was added to the sample and this incubated overnight at -20°C. After incubation, the sample was centrifuged for 30 minutes at 14000 RPM at 4°C. The supernatant was discarded and the pellet was washed in 1ml of 70% ethanol and air-dried. The pellet was re-suspended in 50µl of milliQ ultrapure H₂O and the concentration of the sample was measured. Ten µg of the sample was taken for DNA sequencing per each ChIP-Seq.
2.14.4 Chip Solutions

<table>
<thead>
<tr>
<th>Cell lysis buffer</th>
<th>IP Dilution Buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 10mM Tris pH 8.0</td>
<td>- 0.01% SDS</td>
</tr>
<tr>
<td>- 0.2% NP40</td>
<td>- 1% Triton X100</td>
</tr>
<tr>
<td>- Protease Inhibitors</td>
<td>- 1.2mM EDTA</td>
</tr>
<tr>
<td>- 50mM Tris pH 8.0</td>
<td>- 16.7 mM Tris pH 8.0</td>
</tr>
<tr>
<td></td>
<td>- 167 mM NaCl</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low salt buffer</th>
<th>High salt buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 0.1% SDS</td>
<td>- 0.1% SDS</td>
</tr>
<tr>
<td>- 1% Triton X100</td>
<td>- 1% Triton</td>
</tr>
<tr>
<td>- 2mM EDTA</td>
<td>- 2mM EDTA</td>
</tr>
<tr>
<td>- 20mM Tris pH 8.0</td>
<td>- 20mM Tris pH 8.0</td>
</tr>
<tr>
<td>- 150mM NaCl</td>
<td>- 150mM NaCl</td>
</tr>
<tr>
<td>- 10mM EDTA</td>
<td>- 10mM EDTA</td>
</tr>
<tr>
<td>- 1% SDS</td>
<td>- 1% SDS</td>
</tr>
<tr>
<td>- Protease Inhibitors</td>
<td>- Protease Inhibitors</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IP elution buffer</th>
<th>TE buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>- 1% SDS</td>
<td>- 10mM Tris Base</td>
</tr>
<tr>
<td>- X100 1mM EDTA</td>
<td>- 1mMEDTA</td>
</tr>
<tr>
<td>- 10mM Tris pH 8.0</td>
<td></td>
</tr>
<tr>
<td>- 500mM NaCl</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.20. ChIP-Seq buffers and their components.

2.14.5DNA sequencing

Sequencing of immunoprecipitated lysates was done at the Faculty of Life Sciences, Genomic Technologies Core Facility Centre in the University of Manchester. Following Chromatin-immunoprecipitation (ChIP), DNA libraries were constructed according to the TruSeq® ChIP Sample Preparation Guide (Illumina, Inc.). Briefly, sample DNA (5–10 ng) was blunt-ended and phosphorylated, and a single 'A' nucleotide added to the 3' ends of the fragments in preparation for ligation to an adapter with a single-base 'T' overhang. The ligation products were then purified and PCR-amplified to enrich for fragments with adapters on both ends. The final purified product was then quantitated prior to cluster generation on a cBot instrument and the loaded flow-cell then paired-end sequenced on an Illumina HiSeq2500 instrument.
2.14.6 Results Analysis

Raw sequencing results were mapped using bowtie v2.2.3, a sequencing tool that aligns sequenced reads to reference sequences. Bowtie 2 was used to map fragments with default parameters. Fragments were mapped to the reference genome UCSC hg19 human genome assembly. Peak calling was run using MACS v2.1.0 in paired end mode (knows fragment lengths from input BAMPE file). The coverage tracks used normalised reads (by millions of non-redundant reads) which enables comparison of tracks made using different numbers of reads.
CHAPTER 3

MINING AND IDENTIFICATION OF POTENTIAL PLURIPOTENCY ASSOCIATED GENES FROM AFFYMETRIX MICROARRAY DATA

Microarray data analyses provide a great platform to dealing with large numbers of genes at the same time. They offer robust and yet effective and time efficient ways of data processing. Microarrays have been used to identify previously unknown pluripotency regulatory genes and signalling pathways of stem cells. Bioinformatics and functional genomic studies have also been used extensively in pluripotency studies (Armstrong et al. 2006; Babie et al. 2007). However, their application in the identification of novel pluripotency genes is yet to be fully exhausted. Oct4, Nanog and Sox2 are well known markers of stem cell pluripotency and self-renewal. Understanding the modes of function of these transcription factors is important in enhancing the clinical application of human embryonic stem cells. In addition to Oct4, Nanog and Sox2, there are other important transcription factors involved in stem cell maintenance. In mESCs, Klf4 maintains the self-renewal of the cells by inhibiting their differentiation. The over-expression of Klf4 has been reported to induce the up-regulation of Oct4 while inhibiting differentiation of the cells (Li et al., 2005). c-Myc has been shown to be targeted and regulated by the LIF/STAT3 signalling pathway which is involved in the regulation of stem cell self-renewal and pluripotency (Cartwright et al., 2005). Lin-28, an RNA-binding protein, has been shown to facilitate the post-transcriptional expression of Oct4 in hESCs (Qiu et al., 2010). It directly binds to Oct4 and its inhibition leads to the down regulation of Oct4 in ESCs.

Although numerous studies have been undertaken to elucidate the roles of these core transcription factors in maintaining the pluripotency and self-renewal of stem cells (Niwa et al. 2000; Chambers et al. 2003; Boyer et al. 2005), very little attention has been directed towards identifying other novel transcription factors that may also be important in stem cell maintenance. In this chapter, pre-existing microarray data from our laboratory were used to identify potential novel pluripotency associated genes. Microarray data replicates were also assessed for consistency in
MAN1, hUES3, hUES7 and TE. The results of this analysis will also be discussed here. The use of bioinformatics tools to select new candidate genes with best fit to pluripotency will also be presented in this chapter. Finally, methods that were used to create a list of genes and transcription factors with potential involvement in stem cell maintenance regulation will also be presented.

3.1 Selection of candidate genes from Affymetrix microarray data

3.1.1 Assessment of microarray replicates’ consistency and reliability

Preliminary comparison of gene expression across the different cell lines and the TE was undertaken. Fifty genes confirmed to be expressed in hUES1, hUES3 and MAN1 cell lines as well as the TE were randomly selected. A scatter plot was made by plotting the expression score values of these genes in hUES3 cells against their expression score values in hUES7 cells (Figure 3.1). There was an almost linear correlation in gene expression between the two hUES lines. This signalled how closely similar the expression of genes were between the two cell lines. For both cell lines, there was parallel increase in the expression of similar groups of genes and a similar pattern for genes expressed at low levels. It could be possible that this expression pattern was due to the fact that both hUES3 and hUES7 cell lines were derived and cultured under similar in vitro conditions and therefore underwent similar treatment. Another 50 genes with confirmed expression across the 3 cell lines were randomly selected. The expression score values of these genes in hUES3 cells were plotted against their expression values in the TE (Figure 3.2). An irregular pattern of expression was observed with genes scattered in the plot without any consistency. The plot showed that most of the selected genes were highly expressed in hUES3 stem cells than they were in the TE. This expression pattern signalled that the gene transcripts were more prevalent in pluripotent cells than in the differentiated and differentiating cells of the TE.
Figure 3.1. Gene expression distribution in hUES3 and hUES7 stem cells. Fifty genes expressed in hUES3 and hUES7 cells from Affymetrix microarray data were randomly selected using Netbeans software. The expression of the same genes in hUES3 cells correlated and was consistent with their expression in hUES7 stem cells. Same genes that were highly expressed in one cell line were also highly expressed in the other. All statistics were performed with the Netbeans software. Values in each axis represent relative gene expression score values.
Figure 3.2. Gene expression distribution in hUES3 and TE cells. Fifty genes expressed in hUES3 and hUES7 cells from Affymetrix microarray data were randomly selected using Netbeans software. The expression of the same genes in hUES3 stem cells did not correlate with that in the TE. There was inconsistent expression of the genes and no clear pattern of gene expression observed. All statistics were performed with Netbeans software. Values in each axis represent relative gene expression score values.
For comparison purposes, from the Affymetrix microarray data, the means of expression values of all the genes expressed in all the hESC and TE replicates were calculated using the Netbeans java software. The expression values were generated by a uniform cut-off of >50 expression values across all the arrays. That is, only genes with expression values of 50 or above were considered. Gene expression at this level was considered to be significant. Normalisation of microarray data is important in compensating for technical differences between array chips so that biological differences between samples can be seen clearly. The expression of genes in this study had been analysed and normalised by MAS 5.0 (Affymetrix Microarray Suite User Guide, Affymetrix, Santa Clara, version 5) method. Basically, this method calculates the intensity of each probe on the microarray to give an amount of gene expression based on the fluorescence intensity of each probe. The software calculates local background in each array square and subtracts it from the intensity of each probe. In this study, replicates were directly investigated for consistency in order to assess the reliability of the data sets. Consistency in Oct4 expression patterns was also assessed in each replicate and compared among replicates of each stem cell line or the TE. There appeared to be some inconsistency in the mean expression pattern of genes in the TE (Table 3.1). There was too much disparity between the replicates in the immunosurgically derived samples (TE7i and TE8i). Because immunosurgery involves lysing the TE, this can degrade the gene transcripts resulting in low expression of the genes. However, there was good correlation between replicate TE8i and the manually derived replicate TE9m. As expected, there was no expression of the pluripotency transcription factor Oct4 in the TE replicates (except for TE10m). This is because the TE primarily consists of differentiating and differentiated cells and would not be expected to express markers of pluripotency. The expression of Oct4 in TE10m could be a result of some contamination of the ICM cells during handling. It is also possible that the expression of Oct4 in this replicate was caused by differences in the genetic background of embryos from which the cells were derived. The cells were derived from different TEs of different blastocysts from different human genetic make-ups. There was relatively good correlation between hUES3 replicates except for replicate 3. HUES7 replicates showed slightly inconsistent expression mean values while expression mean values were almost consistent across all the MAN1 replicates.
However, unlike the other ESC lines and the TE, MAN1 had only 3 replicates instead of 4. Oct4 expression was well correlated in all the hESC replicates, except for hUES3 and hUES7 replicates 4. The low expression values of Oct4 in these two replicates could imply that there may have been a problem with amplification. It is also possible that the cells may have already started differentiating in these replicates. Overall, the expression of genes in the TE and hESC lines was satisfactorily consistent.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Replicate</th>
<th>Relative Gene Expression</th>
<th>OCT4 Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Expression (ALL GENES)</td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>7i</td>
<td>314.9</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>8i</td>
<td>100.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>9m</td>
<td>134.4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10m</td>
<td>194.4</td>
<td>77.8</td>
</tr>
<tr>
<td>hUES3</td>
<td>1</td>
<td>255.5</td>
<td>162.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>233.8</td>
<td>191.0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>153.5</td>
<td>146.1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>278.8</td>
<td>84.3</td>
</tr>
<tr>
<td>hUES7</td>
<td>1</td>
<td>169.2</td>
<td>141.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>141.9</td>
<td>180.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>211.2</td>
<td>189.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>250.0</td>
<td>74.8</td>
</tr>
<tr>
<td>MAN1</td>
<td>1</td>
<td>239.2</td>
<td>287.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>327.8</td>
<td>290.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>261.2</td>
<td>243.9</td>
</tr>
</tbody>
</table>

Table 3.1. Mean expression score values of genes expressed in hUES3, hUES7, MAN1 and TE cells. There was relative consistency among individual replicates. Values shown in red showed a high level of disparity to other replicates. There was an unexpected expression of Oct4 in one of the TE replicates (shown in green).
Further investigation of gene expression consistency was undertaken by calculating the number of genes expressed in all hESC and TE replicates using Netbeans software tool. This software was used for the statistical analysis of the array data. Generally, there was reasonable consistency in the expression of genes in the hESC lines except for hUES3 replicate 4 and hUES7 replicate 4 (Table 3.2) which showed unusually low transcript detection. Among the 3 cell lines, the greatest overall number of gene expression was observed in hUES3 cells (21,638) followed by hUES7 cells (21,090) while MAN1 cells showed the lowest number of genes expressed with a total of 19,270 genes. There was very little difference between the total number of genes expressed in hUES3 and hUES7 cells. This could be because these cell lines were derived from the same laboratory (Cowan et al., 2004) under the same laboratory conditions and are therefore likely to exhibit similar gene expression patterns. MAN1 line showed the lowest total number of genes expressed because it was derived from a different environment to the other two cell lines. It is however important to note that there were only 3 replicates analysed for this cell line because the 4th replicate was too variable from the others to analyse. There were low numbers of genes expressed in replicate TEi7 of the TE. This could be a result of the immunosurgery lysing technique explained earlier. This is in contrast to manually extracted TE cells – where no lysis induced degradation could have occurred. Overall, the trophectoderm showed the lowest number of genes expressed compared to all the stem cell lines. The findings of this analysis are shown in table 3.2 below;

<table>
<thead>
<tr>
<th>Replicate</th>
<th>hUES3 No. of genes expressed</th>
<th>hUES7 No. of genes expressed</th>
<th>MAN1 No. of genes expressed</th>
<th>TE No. of genes expressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep1</td>
<td>6066</td>
<td>5910</td>
<td>6521</td>
<td>TEi7 2317</td>
</tr>
<tr>
<td>Rep2</td>
<td>6049</td>
<td>5779</td>
<td>6765</td>
<td>TEi8 3151</td>
</tr>
<tr>
<td>Rep3</td>
<td>5833</td>
<td>5948</td>
<td>5984</td>
<td>TEm9 3822</td>
</tr>
<tr>
<td>Rep4</td>
<td>3670</td>
<td>3453</td>
<td>-</td>
<td>TEm10 3712</td>
</tr>
</tbody>
</table>

Table 3.2. Number of genes expressed in hUES3, hUES7, MAN1 and TE replicates. There was very good correlation between replicates except for values shown in red which had more disparity to other replicates of the same cell sample.
3.1.2 Selection of genes targeted by pluripotency transcription factors

As described in Chapter 2, microarray data used in this study had been prepared in our lab to compare gene expression in the human blastocyst ICMs to gene expression in hESCs isolated from those ICMs with the TE used as a differentiated cell control (see 2.5.1). In order to improve chances of selecting genes that are involved in the maintenance pluripotency and self-renewal, it is important to select genes that are targets of or interact with the core pluripotency transcription factors Oct4, Nanog and Sox2. To do this, using the Netbeans analysis tool, genes (probes) expressed in hUES3, hUES7 and MAN1 and trophectoderm (TE) cells (see 2.5.1) were ranked by their t-statistic values. Obviously, comparison of the endogenous pluripotent stem cell population with the inner cell mass would have been the best comparator to find novel pluripotency associated genes but unfortunately this could not be done as the ICM data was unavailable at the time. MAN1 replicate 4 was excluded from this analysis because it did not cluster together with the other replicates. Replicate 9m of the TE was also excluded because it has anomalously high Oct4 expression, indicating possible contamination. ‘m’ denotes mechanically derived replicates as opposed to the immunosurgically derived ones; which are denoted by ‘i’. Then for each probe, the expression levels for replicates in a stem cell line were compared with expression levels for replicates in the trophectoderm. This is because stem cells are pluripotent cells while the TE comprises of differentiating or differentiated cells. So any gene highly expressed in the TE would not be ideal for this study. The t-statistic applied was a general one 'unequal sample sizes, unequal variance'. This test, called Welch’s t-test, and is a variant of Student's t-test, and is applied when two samples of data are assumed to have unequal variance. It is applied irrespective of whether samples are the same size or not. In this case, samples were the same size, because there were 3 replicates of the hESC lines and 3 replicates of the TE. However, it could not be assumed that the variance of the expression sets would be the same in hES cell lines and in TE samples, hence the Welch's t test was used. After calculation of probe t-statistic values, only probes common to all 3 of the retained lists for hUES3, hUES7 and MAN1 cells were retained. Therefore, only probes showing evidence of overexpression in all the 3 ICM derived cell lines, relative to the TE, were retained. Then, Oct4, Nanog
and Sox2 were mapped to binding sites on the retained genes’ promoters +/- 2000 nucleotides around 5’most transcription start site (TSS) of the genes in Ensembl v65human GRC37 genome list (filtered for protein-coding genes only).

3.1.3 Comparison of gene expression trends with published data

In order to assess the efficiency of the gene selection criteria used, expression trends of genes retained from the mapping of Oct4, Nanog and Sox2 were compared to their expression trends in published data. Transcription factor datasets from the HAIB Consortium (Karachi et al., 2004 and Myers et al., 2011) and from 2 independent studies by Boyer et al., 2008 (Cell) and Kunarso et al., 2010 (Nature Genetics) were used in this analysis. From these datasets, genes retained after mapping for Oct4, Nanog and Sox2 were divided into ten 10% cohorts of equal sizes based on their t-statistic values, starting with the most expressed genes (i.e. genes with the highest t-stats values). The first cohort comprised of the top 10% most expressed genes in hUES3, hUES7 and MAN1 cell lines while the last cohort had the least 10% expressed genes in the cell lines. Therefore, genes in the first cohort were genes that were more expressed in the 3 stem cell lines than in the TE while those in the last cohort were more expressed in the TE than in the stem cell lines. From these cohorts, the total numbers of the different transcription factor binding sites were added up across all the gene cohorts. The frequency of pluripotency transcription factors to the promoter regions of genes in each cohort were plotted against the mean t-statistic values of genes contained in each cohort. Results of this analysis are presented in figure 3.3 below.
Figure 3.3. Total number of pluripotency transcription factors bound to ten 10% cohorts of genes expressed in hUES3, hUES7 and MANI stem cells. The numbers of pluripotency transcription factors targeting ten 10% cohorts of genes were counted, based on 4 previously done studies (see 2.5.2). a - c) The Kurnaso, Boyer and HAIB 1 consortium studies showed that genes that are highly expressed in stem cells were more targeted by pluripotency-associated transcription factors than those with low expression in stem cells. d) The HAIB 2 consortium study showing that another pluripotency gene, EP300, was more frequent on the promoter regions of genes highly expressed in stem cells while RAD21, a non-pluripotency gene which was used as a control, remained relatively constant.
As stated in Chapter 1, the hypothesis of this study is that since Oct4 and Nanog cannot, just the two of them, regulate the pluripotency and self-renewal of stem cells, there are other yet unidentified genes that are involved in this regulation and expressed in the same way as Oct4 and Nanog. Therefore, the number of pluripotency transcription factors targeting retained genes were then tabulated. Table 3.3. shows total numbers of pluripotency transcription factors for 10% cohorts of probes ranked by their t-stats values in MAN1 stem cells based on the 4 different studies. The same analysis was undertaken for both hUES3 and hUES7 cells but have not been shown. The same table shows ratios of total transcription factor binding sites between the most and least expressed 20% of probes in hESCs relative to the TE. Generally, there were higher ratios for MAN1 compared to hUES3 and hUES7 cells which suggests that MAN1 probes were separated out in a more functionally meaningful way by the t-test i.e. in some sense the comparison was likely more real between the array data for MAN1 and TE than between hUES3 and the TE or hUES7 and the TE. This is reflected in a more distinctive distribution of transcription factors towards more highly enriched promoters for the most expressed probes in MAN1 line. i.e. the most highly expressed probes in MAN1 line had more Oct4 and Nanog binding sites than the most highly expressed probes in hUES3 and hUES7 cells. Overall, this analysis showed that there were more pluripotency targeting of genes highly expressed in MAN1 cells. However, this targeting in MAN1 was not significantly different to the targeting of genes by pluripotency transcription factors in hUES3 and hUES7 cells. Therefore, this analysis showed all the three cell lines could be used in further analyses aimed at identifying potential pluripotency and self-renewal regulatory candidates.
Table 3.3. Total numbers of pluripotency transcription factors targeting 10% cohorts of t-stats ranked genes expressed in MAN1 cells. Based on 4 different published studies (see 2.5.2), generally, pluripotency and self-renewal regulatory transcription factors target genes that are highly expressed in hECS (MAN1, hUES3 and hUES7).

NB: H/L gives the ratio of total TF binding sites between the most and least expressed 20% of probes in ESC lines relative to the TE.
3.2 Identification of genes with best fit to pluripotency

3.2.1 Selection of target genes interacting with Oct4

In order to select genes with greatest potential to be stem cell regulators, another restriction was introduced to the data analysis. Using the Netbeans software, based on the Kunarso et al., (2010) study, only probes mapping to genes targeted to the proximal promoter by Oct4 (Figure 3.4) were retained. The Kunarso ChIP-seq data was chosen because the analysis done in this study was more extensive and gave better contrast between high and low expressed probes (genes) in hUES cells compared to TE cells. Therefore, the presence of the Oct4 potential binding site from the Kunarso study was used as a further filter. Also probes with no corresponding gene symbol or Ensembl ID were removed. This meant that genes (probes) with no identified names, or those which could not be identified were eliminated from the final retained gene list. In brief, in their study, Kunarso and colleagues investigated the genome-wide binding sites analysis of Oct4, Nanog and CTCF which are all important in the pluripotency maintenance transcription factors in mouse and human hESCs. They found that only around 5% of Oct4 and Nanog binding sites were homologously occupied between the two species compared to CTCF which was highly conserved between the species. They attributed 25% of bound sites to transposable elements. The instruction of the Netbeans software to retain only genes targeted by Oct4 led to 520 probes being retained. These probes corresponded to 440 distinct genes. This outcome was obviously due to the fact that, due to factors such as alternative splicing, some probes have multiple binding sites corresponding to multiple genes and can therefore bind to more than one gene.
Figure 3.4. Oct4 promoter on chromosome 6. A schematic showing an Oct4 proximal promoter (shown in red) which is mapped to chromosome 6 in humans and is upstream of the Oct4 transcription start site (TSS). Computer softwares also annotate an Oct4 promoter (and TSS) but this promoter has no Oct4 binding sites for interaction of Oct4 with other genes.
3.2.2 Functional clustering with DAVID annotation database

Gain of function analysis was performed on the 440 retained genes using the functional annotation online tool DAVID. DAVID provides information of the biological properties enriched in a list of genes in a specified organism. Therefore, DAVID looks for different pathways where annotated genes of interest are over-represented. It comes up with functional clusters comprising of genes with similar or related functions. Probe sets IDs of target genes were uploaded into DAVID database and functional annotation clustering was undertaken to perform go term searches on the gene list so that functionally relevant genes (i.e., genes contained in clusters related to stem cell maintenance) can be selected. This analysis pulled up a number of functional clusters. Among these clusters, the following two clusters were selected:

1. Negative regulation of stem cell differentiation
2. Negative regulation of macromolecule biosynthetic processes

Both annotation clusters had enrichment scores of above 2 which are considered significant. The cluster containing genes involved in the negative regulation of stem cell differentiation was selected because genes that negatively regulate cell differentiation are anti-differentiation and therefore likely to have roles to play in the maintenance of stem cell pluripotency. The cluster containing genes involved in the regulation biosynthetic process were selected because they regulate cellular activities, such as transcription, which are important for cell regulation. Figure 3.5 below shows an example of a DAVID functional annotation output with one of the retained functional cluster highlighted in orange. The clustering of genes using DAVID annotation tool was very valuable in that it converted the microarray data into biological meaning. DAVID summarized the gene list into shared protein domains and biological pathways, therefore enhancing the aggregation of functionally related genes. Moreover, DAVID can also search for other functionally related genes that are not in the gene list as well as identifying interacting proteins. Functional clusters created by DAVID are also directed to relevant literature, making it easier to acquire further information on the genes of interest.
Figure 3.5. DAVID functional annotation output. A screenshot of a DAVID annotation functional cluster output showing one of the retained clusters. Enrichment score is depicted at the top of the second column with each cluster name following below it. The orange bar represents enriched genes within different categories.
Further analysis was done with DAVID by clustering for genes with Oct4 binding sites (Figure 3.6). This resulted in a number of stem cell ‘themed’ functional annotation clusters.

In comparison to the previously retained 2 functional clusters, screening for Oct4 binding sites does improve enrichment in stem cell related functions of top scoring genes. However, the enrichment score for these clusters was below 2. This analysis was therefore insignificant and the clusters were not considered for further analysis. The originally retained functional clusters were maintained for further analyses.

### 3.2.3 Functional evaluation of gene clusters for assembly of a final gene list

After the acquisition of two functional clusters of genes following analysis by the DAVID annotation tool, the next task was to create a short and final list of genes (Table 3.4) for further experimentation. In order to do this, extensive literature search was undertaken on the genes contained in the two clusters. Background analyses included the use of online published literature databases such as NCBI in order to check which among the genes in the two functional clusters were most likely to have a role in stem cell regulation and maintenance. In addition, journal published papers on these genes were browsed to see what was already known about these genes. However, there were set criteria during this analysis. First, genes that were selected were those that had been previously proven to either be expressed in stem cells or have potential involvement in cell regulation. Secondly, the selected genes were those that had not been that well studied in pluripotent stem cells therefore meaning that there was a lot more to be learned about them. Literature filtering was crucial as it avoided selection
It also led to a reasonably workable number of irrelevant genes for further analysis. Moreover, it enhanced the probability of identifying pluripotency relevant genes. Table 3.4 below shows the 10 retained genes after background literature search on the 2 retained DAVID annotation functional clusters.

Table 3.4: Final target gene list. A shortlist of functionally relevant genes selected from 2 DAVID software functional analysis clusters after extensive literature search.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Literature Background</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF7L2</td>
<td>TCF7L1 previously studied and found to be expressed in pluripotent hUES3, hUES7 and MANI cells</td>
<td>Already confirmed in our lab</td>
</tr>
<tr>
<td>MYST3</td>
<td>Also called MOZ, involved in the development and self-renewal of hematopoietic stem cells</td>
<td>Katsumoto et al., 2008; Voss and Thomas, 2009</td>
</tr>
<tr>
<td>RIF1</td>
<td>A target of the pluripotency associated markers Sox2, Oct4 and Nanog in mouse ESCs</td>
<td>Loh et al., 2006; Wang et al., 2006; Zourmparo-Djayoon et al., 2011</td>
</tr>
<tr>
<td>RBM15</td>
<td>A target of c-myc, its role is yet to be fully understood. It is believed, however, to be needed for hematopoietic stem cell niche interactions and modulation of notch induced transcriptional activation</td>
<td>Ma et al., 2007; Niu et al., 2009</td>
</tr>
<tr>
<td>MYST4</td>
<td>Also called QKF and MORF, it is a marker of stem cell self-renewal and multipotency of adult neural stem cells</td>
<td>Zimmer et al., 2011; Sheilich et al., 2012</td>
</tr>
<tr>
<td>TBL1XR1</td>
<td>Also TBLR1, it is a co-repressor of transcription</td>
<td>Zhang et al., 2006</td>
</tr>
<tr>
<td>CCND1</td>
<td>It is a beta-catenin target gene involved in the regulation of the cell cycle</td>
<td>Brabletz et al., 2005; Li et al., 2012</td>
</tr>
<tr>
<td>NME1</td>
<td>A member of the NME gene family involved in cell proliferation, self renewal and differentiation</td>
<td>Hikita et al., 2008; Zhu et al., 2009; Boissan and Lacombe 2011</td>
</tr>
<tr>
<td>YYAP1</td>
<td>Also called HCCA2. Expressed in human and cancer tissues, it co-activates YY1 and is involved in cell cycle regulation</td>
<td>Wang et al., 2004; Li et al., 2007; Chechlińska et al., 2009</td>
</tr>
<tr>
<td>DGCR8</td>
<td>A gene required for the maturation of primary miRNA transcripts</td>
<td>Landthaler et al., 2004; Wang et al., 2007</td>
</tr>
</tbody>
</table>
3.3 Discussion

Microarray data used in this study was robustly analysed and assessed prior to its use in gene expression studies. In this chapter, microarray data that had been generated in the lab on hUES3, hUES7, MAN1 cell lines and the human blastocyst TE was assessed in order to identify, quantify and compare gene expression between the first differentiated cell type formed from the endogenous human pluripotent stem cell population and pluriotent human embryonic cell line. This array data was important because it enabled the identification of Oct4 and Nanog related genes. The probability of retaining stem cell maintenance genes was consistently increased by first sieving out anomalous replicates in the microarray data and then selecting only genes that were highly expressed in the stem cell lines than in the TE. Further restriction towards only Oct4 mapped and functionally significant genes added a lot of credibility to the gene selection criterion. Of course, this analysis would have been improved further if it had been possible to include ICM data because the ICM comprises of stem cells in their natural habitat. Based on these levels of analyses, there was a high chance of identifying candidate genes that are related or involved in the regulation of stem cell pluripotency. However, just because a gene is highly expressed in ESCs does not necessarily mean that it is involved in stem cell self-renewal or maintenance of pluripotency. Similarly, just because a gene forms a complex with Oct4 and/or Nanog does not mean that it is involved in stem cell self-renewal or pluripotency. In-fact, some genes whose promoters are bound by Oct4 are actually down regulated by this transcription factor (Chambers and Smith, 2004). However, as Boyer et al. (2005) found in their genome scale analysis of the three core hESC regulators; Oct4, Nanog and Sox2, these three factors can bind to promoters of many different genes which encode transcription factors that have been known to partake in stem cell regulatory mechanisms. Therefore, a gene that is both high expressed and can form a complex with Oct4 has a greater probability of being involved in stem cell self-renewal and maintenance of pluripotency.

DNA microarrays provide scientists with a platform to analyse large amounts of data at the same time. Microarrays generally comprise of gene probes that represent differentially expressed and regulated genes (Khatri and Drăghici, 2005).
Microarrays are equally important in stem cell research in that they can be used for analysis of gene expression in a specific cell line, comparison of gene expression between different stem cell lines and even for comparing gene expression patterns of two or more genes or transcription factors in one or more given stem cells lines. It is important that the reliability of any microarray data set is assessed before such data can be used for further analyses or experimentation. In this chapter, microarray data from our lab was rigorously assessed for consistency, compatibility and variability of all replicates of hUES1, hUES7, MAN1 and TE cells.

Assessment of randomly selected genes in hUES3, hUES7 and TE cells led to interesting observations. Comparison of the expression of the same genes between hUES3 and hUES7 cell lines showed a generally similar trend where genes that were highly expressed in one line were also highly prevalent in the other and those low in one line were also low in the other (Figure 3.1). HUES3 and hUES7 are members of the 17 hESC cell line of stem cells that were derived at Harvard University (Cowan et al., 2004). Even though they different genetic background, these cell lines were shown to express the same markers of pluripotency such as Oct4 and SSEA3 as well as adaptation to the same substrate and similar karyotypes (Cowan et al., 2004; Allegrucci and Young, 2007). Therefore, they are more likely to exhibit similar gene expression in vitro such as the one shown in Figure 3.1. Gene expression patterns of the same randomly selected genes between hUES3 and the TE cells showed a different picture all together. As expected, was no positive correlation in gene expression patterns between the stem cells and the TE cells (Figure 3.2). The TE is comprised of differentiating and sometimes differentiated cells. Therefore, it was not surprising to find that genes highly expressed in hUES3 cells were expressed at low levels in the TE and vice-versa. Stem cells are pluripotent cells from the ICM that expresses markers of pluripotency and whose primary function is to prevent cell differentiation. The TE, on the other hand is a mass of differentiating genes that express markers of differentiation and whose primary function is to promote differentiation. Therefore, these two cell samples orchestrate opposing gene expression patterns with negative correlation between them.

Assessment of variations between sample replicates revealed some fair level of consistency between replicates of the same cell samples (see 3.1.1). However, there
were few variations observed between same replicates of the same samples. For example, there were inconsistencies with some of the TE replicates; the TEi replicates (Table 3.2). While there were 3151 genes expressed in replicate TEi8, there were only 2317 genes expressed in replicate TEi7. It is important to note that both the hESCs and TE samples were derived from embryos donated, with written informed consent, by patients undergoing IVF treatment. Therefore, samples were derived from different genetic backgrounds which might have been a contributing factor to some of the variations observed among some of the replicates. The microarray data was also prepared by more than one person. Therefore, variations in handling techniques and general human errors cannot be ruled out with certainty.

The introduction of restrictions in the search for novel pluripotency related genes was very important during the microarray data analysis. In their study Tanaka et al. (2002) used expression profiling to compare ESCs, trophoblast stem (TS) cells and MEFS. They found that each of these cell types had a unique expression profile with some common genes between the ES and TS cells. In addition, they found that the expression of TS cells specific genes was restricted to the trophoblast. Their findings that some genes were only common between ES and TS cells could indicate that only genes that are exclusively expressed in stem cells might be involved in pluripotency.

The performance of functional analyses on retrieved gene list to select genes with functional potential was an important exercise that gave insight of the functional mechanisms of the gene contained in the final gene list. Such analyses are time effective and reduce the laborious task of analysing many genes at the same time. The use of DAVID annotation software, partnered with literature search on the gene list, indeed proved to be a very effective way of gene selection. A considerably manageable list of 10 genes was created following the functional analysis of genes contained in the 2 DAVID annotation tool functional clusters (Table 3.4). For proper analyses, functional analyses tools must have the capacity to correct for multiple experiments (Berriz et al., 2003; Shah and Fedoroff, 2004). However, DAVID annotation database lacks this capacity to make corrections. However, DAVID does indicate output enrichment levels which, although they cannot be taken as a correction for multiple experiments, indicate output significance level which brings credibility to the system.
There are some challenges associated with microarray data analysis. Hybridization and washing conditions are important for the success of microarray experiments because when optimized, they can reduce non-specific binding and enhance fluorescence measurement on the microarray (Han et al., 2006). Along with splice variance, a single probe may bind to more than one transcript depending on the prevailing biological conditions (Stalteri and Harrison 2007). One of the initial challenges that affected the bioinformatics analysis of the microarray data in this study was that of multiple probes mapping to the same gene. To solve the problem of multiple probes mapping to a single gene, Netbeans software was instructed to list all the individual probes which map to a particular gene. This corrected some errors where different probes for the same gene resulted in unpredicted outcomes in an algorithm.

In conclusion, the initial aims of this chapter, of identifying and retaining potential stem cell pluripotency and self-renewal regulatory genes were achieved. First, adequate and thorough analysis of microarray data was undertaken through rigorous comparison of replicates in the same samples as well as gene expression pattern comparisons between different sample cells. Then, using bioinformatics applications, functional analysis tools and gene background literature assessment, a reasonable small pool of potential pluripotency associated genes was generated. In subsequent chapters, genes contained in this short-list were further analysed for their expression profiles in stem cells and their effects on gene expression patterns within these cells.
CHAPTER 4

ASSESSING THE EXPRESSION OF SELECTED TARGET GENES IN hESCs

Growth factors such as Activin-A and FGF-2 are important in the in vitro maintenance of human embryonic stem cells (hESCs). It has been recently shown that FGF signalling is crucial at the early stages of definitive endoderm (DE) formation induced by Activin-A, with more DE cells formed in the presence of both FGF and Activin-A than in Activin-A alone (Vallier et al., 2005; Sui et al., 2012). In this chapter, Activin-A and FGF-2 were individually and together excluded from stem cell media to induce differentiation of the cells into embryoid bodies (EBs). The removal of these two factors results in loss of pluripotency and initiates differentiation. In chapter 3, a short list of genes with potential for stem cell regulation was generated. Here, the expression patterns of these genes were evaluated during growth factor exclusion and EB differentiation protocols. The expression of three pluripotency regulatory transcription factors (Oct4, Nanog and Sox2) during Activin-A removal differentiation in hUES7 cells was used as a positive control. For the negative control, the expression of three differentiation markers (CDX2, GATA4 and Mixl1) was observed during Activin-A removal differentiation. CDX2 is a trophoblast marker while GATA4 and Mixl1 are endodermal and mesodermal markers respectively. Results showed consistent down-regulation of most of the target genes during the different growth factor withdrawal and EB differentiation protocols in both hUES7 and MAN7 cells. All the positive control genes were steadily down regulated during Activin-A removal while all the negative controls (with the exception of CDX2) were up regulated. This signalled that cell differentiation was induced. From the 7 target genes, a transcriptional regulatory gene called YY1AP1 was selected for further experimentation. This chapter will also present results on the expression of YY1AP1 and its binding partner YY1. Expression patterns of these two genes were compared after cell differentiation and through co-localization immunofluorescence staining, immunoprecipitation and western blotting experiments in both hESCs and iPSCs. Morphological observations showed the well-established morphology of pluripotent cells which were small and dense while differentiated cells appeared to have small nuclei.
4.1 Validation of target gene expression in hESCs

4.1.1 hESC culturing and maintenance

In order to bulk up cells for experimental use, different stem cell lines were cultured on MEF feeder layers before being transferred into a feeder-free system on either matrigel or fibronectin. Basically, MEFs were derived, cultured, inactivated and thawed before being used for hESC culture (see chapter 2.1). MEFs were cultured over at least 3 passages in MEF medium on 0.1% gelatin coated culture plates (Table 2.1). After attaining at least 70% confluency, hESCs were plated on the MEFs in hES medium (Table 2.1) for at least 3 passages (chapter 2.2.1). The morphology of the hESCs was regularly assessed by phase contrast microscopy to check that they are healthy. Generally, the hESCs formed large compact colonies with defined outlines on MEF feeder layers (figure 4.1a-c). Healthy hESC colonies were then transferred and cultured in FF medium in a feeder-free system (as in chapter 2.2.2) until they reached at least 80% confluence. Assessment of cell morphology showed that the cells formed compact layers (figure 4.1d-f). The cells appeared to form consistent mono-layers unlike defined colonies observed when the cells were on MEF layers. Furthermore, the pluripotency of the cells was tested by staining the cells for both cell-surface and intra-cellular markers of pluripotency. Cells adequately expressed the nuclear localised Oct4 as well as the surface marker Tra-1-60 (figure 4.1g-l). Tra-1-81 was also expressed although at times it was difficult to obtain clear staining for this transcription factor, possibly a result of poor antibody quality (figure 4.1m-o). In addition, negative controls showed negligible secondary antibody cross-reactivity (figure 4.1p-r). In addition to immunofluorescence staining, qRT-PCR was performed on the cells for pluripotency-associated transcription factors Oct4 and Nanog and non-pluripotency factor BMP-2 (figure 4.2). Generally, there was high expression of Oct4 and Nanog, signalling stem cell maintenance while the differentiation factor BMP-2 was expressed at low levels showing that the cells had not undergone differentiation although they could have undergone alternative differentiation. Alternative differentiation refers to the differentiation towards an unexpected lineage. For example, BMP-2 promotes mesendodermal differentiation. If cells treated with BMP2 forms neurons and not mesendodermal cells, that would be deemed as alternative differentiation.
Figure 4.1. Morphological and immunofluorescence evaluation of hUES7 p29 pp12 cells. Cells were cultured as described in sections 2.2.1-2.2.2 and co-localisation immunofluorescence staining was performed as described in section 2.8.2. a-c) hESCs grown on MEF feeder layers showing round colonies. d-f) hESCs grown under feeder-free conditions showing cell monolayers with typical hESC morphology g-i) positive Oct4 staining of pluripotent hESCs. j-l) Tra-1-60 staining of hESCs. m-o) Tra-1-181 staining of hESCs. p-r) Mouse and goat immunoglobulin negative control stainings on hESCs showing no stainings. p = number of passages on MEFS; pp = number of passages on a feeder-free system. All scale bars = 100 µm. Image inserts shown in red have been zoomed in for better cell visualization.
Figure 4.2 qRT-PCR analysis of gene transcript expression of markers in pluripotent stem cells. p32 pp7 hUES1 cells were cultured (see 2.2.1-2.2.2). RNA was collected and processed as described in sections 2.6.2-2.6.4 and q-RTPCR was performed as described in section 2.7. Pluripotent stem cells showed high expression of pluripotency markers Oct4 and Nanog and low expression of the differentiation marker BMP-2. p - cell passage number on MEFs and pp - cell passage number on a feeder-free system.
4.1.2 Gene expression analysis by qualitative PCR

Analysis of our laboratory’s pre-existing microarray data was performed (see chapter 2.5.1). This microarray data was performed in order to compare gene expression in endogenous pluripotent stem cells of the ICM and gene expression in isolated hESC cell lines. For this study, interest was directed to gene expression in isolated hESCs (hUES7 and MAN7 cell lines). The expression of genes in the TE was used as a control. The TE was a suitable control because it comprises of differentiating or differentiated cells whose gene expression pattern is expected to be very different to the gene expression in the isolated hESC lines. Bioinformatics analysis of the microarray data was performed using Netbeans software. This included comparison of hUES3, hUES7, Man7 hESC lines and TE replicates from the blastocyst to test the consistency and reliability of the array data sets (see chapter 3.1). Functional analysis was then done using DAVID annotation software on genes that were commonly and highly expressed in the 3 cell lines but with low or no expression in the TE. This resulted in the selection of two functional clusters on which further analysis was done. Extensive literature search on genes contained in these clusters was performed to select ideal gene candidates for further experimentation to investigate their roles in stem cell pluripotency maintenance. A final list of 10 genes was created from the gene list on which further analysis would be performed. In order to assess these genes in further experiments, their expression in hESCs had to be confirmed. Qualitative gel electrophoresis was undertaken (as in chapter 2.6) for these genes in hUES7 stem cells (figure 4.3). From the 10 genes, 7 were clearly expressed in the stem cells. These 7 genes were therefore selected for use in further experiments. No bands were observed in the negative and non-template control (NTC) lanes, showing that there was no contamination during the qualitative PCR. There was also consistent gene expression in the positive, human genomic (hg) control lanes, showing that all the designed primers were optimally functional.
Figure 4.3 Agarose gel electrophoresis for target gene transcripts in hUES7 stem cells. Gel-electrophoresis for 10 target genes showing the expression of 7 of them in pluripotent hUES7 cells. ‘p’ denotes the number of cell passages on MEF feeder layers while ‘pp’ denotes the number of cell passages on feeder-free conditions. Note: Lane ‘d’ was PCR performed on RNA collected from hESCs. ‘hg’ was human genomic DNA which was the positive control since all gene expression was expected. No DNA was added in ‘NTC’ (No Template Control) which was the negative control; instead PCR water was added. p = cell passage number on MEFs and pp = cell passage number on a feeder-free system.

<table>
<thead>
<tr>
<th>Gene</th>
<th>d</th>
<th>hg</th>
<th>NTC</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCF7L2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYST3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIFI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBM15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYST4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBLXRI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCND1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NME1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YY1API</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGCR8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

hUES7, p28, pp5
4.2 Evaluation of the effects of hESC differentiation on the expression of target genes

4.2.1 Activin-A withdrawal induced differentiation

Because Activin-A is such an important growth factor for the survival and maintenance of human embryonic stem cells, its absence from stem cell media is always likely to have detrimental effects on the proliferation and maintenance of the hESCs. Here, during Activin-A withdrawal differentiation, hUES7 and MAN7 stem cells were cultured under normal feeder free conditions (see chapter 2.2.2) in FF medium (Table 2.1). However, Activin-A was excluded from the FF medium throughout the protocol in order to induce differentiation of the cells. RNA was collected and reverse-transcribed at days 0, 5 and 10. The resultant cDNA was used during gene expression analysis by q-RT-PCR experiments. q-RT-PCR was performed for common markers of pluripotency and differentiation respectively; in the DNA harvested from hUES7 cells undergoing Activin-A removal induced differentiation (figure 4.4a). Pluripotency markers Oct4, Nanog and Sox2, were all down-regulated as cells started to differentiate. Differentiation markers CDX2, GATA4 and Mixl1 were up regulated as cells lost pluripotency. Gene expression for the 7 target genes showed a general trend of gene down-regulation during differentiation (figure 4b-c). Some genes like TBL1XR1-in MAN7, however, showed different patterns of expression where they went up at day 5 of the differentiation protocol before going down at day 10. Others like CCND1 and TCF7L1 (both in hUES7s) showed the reverse expression patterns.
Figure 4.4. q-RT-PCR for control and target gene transcripts in hUES7 and MAN7 cells during a 10-day Activin-A withdrawal differentiation protocol. q-RT-PCR was performed as described in section 2.7 a) Pluripotency and differentiation marker expression in p28 pp5 hUES7 cells. Target gene expression in p28 pp5 hUES7 (b) and p30 pp5 MAN7 (c) cells during Activin-A withdrawal differentiation showed a general trend of down regulation of the genes. Some genes were exceptions however, such as CCND1, TBLIXR1 in hUES7 cells and RBM15 and TBLIXR1 in MAN7 cells. p - number of cell passages on MEFs. pp - cell passages on feeder-free culture conditions. n=3. Error bars= Standard deviation.
4.2.2 Differentiation through the withdrawal of FGF-2

The importance of FGF-2 in stem cells is its ability to inhibit differentiation of the cells (Dvorak et al., 2005; Vallier et al., 2005). As with Activin-A withdrawal differentiation, differentiation was induced in hUES7 and MAN7 stem cells by growing the cells in normal feeder-free conditions with the only difference being the exclusion of FGF-2 from FF medium (see Table 2.1). Gene expression was observed by q-RT-PCR on DNA collected at days 0, 5 and 10 of the differentiation protocol. There was consistent down regulation of most of the target genes in both stem cell types (figure 4.5). In addition, the expression of each target gene in hUES7 cells for both Activin-A and FGF-2 differentiation protocols was compared to its expression in MAN7 hESCs for both protocols at days 0, 5 and 10 (figure 4.6). This comparison showed that on average, there was more down regulation of gene expression in cells from the Activin-A withdrawal than those from the FGF-2 withdrawal protocol. This was the case in both hUES3 and MAN1 cells. Based on gene expression levels at day10 of the protocols, there was more gene expression down regulation during Activin-A withdrawal than during FGF-2 withdrawal for all genes except CCND1 in hUES7, Myst3 in MAN7, RBM15 in MAN7, TBL1XR1 in both lines and TCF7L2 in MAN7.
Figure 4.5. q-RT-PCR for target gene transcript in hUES7 and MAN7 cells during a 10-day FGF-2 withdrawal differentiation. q-RT-PCR was performed as described in section 2.7. a) Target gene expression in p28 pp5 hUES7 cells showed down regulation of most of the genes except for CCND1 and RBM15. b) Genes expressed in p30 pp5 MAN7 cells generally down regulated except for Myst3. p - number of cell passages on MEFs. pp - cell passages on feeder-free culture conditions. Error bars - Standard deviation.
Figure 4.6. Comparison of target gene transcript expression in Activin-A and FGF-2 withdrawal during 10-Day protocols for hUES7 and MAN-7 cells. qRT-PCR was performed as described in section 2.7. Generally, there was more down regulation of gene expression during Activin-A withdrawal protocol than during the FGF-2 withdrawal protocol. This was observed in both p28 pp5 hUES7 and p30 pp5 MAN7 hESC lines. Results are representative of ΔΔCt value averages of three separate replicate runs. p = number of cell passages on MEFs. pp = cell passages on feeder-free culture conditions. Error bars = Standard deviation.
4.2.3 Activin-A + FGF-2 withdrawal differentiation

Both individually and synergistically, the importance of Activin-A and FGF-2 in human embryonic stem cell regulation is well documented. These factors have been collectively implicated in inducing endodermal differentiation of stem cell. For example, in their earlier study, Sugi and Lough (1995) showed that during chick embryo development, Activin-A and FGF-2, at low concentrations, and in partnership with insulin, induce and regulate late cardiogenic development in the embryo. More recently, Activin-A and FGF pathways have been shown to work with the BMP signalling pathway to induce hESCs to differentiate into endodermal pancreatic cells (Xu et al., 2011) expressing markers of definitive endodermal cells such as FoxA2 and Sox17. Therefore, the cooperation between Activin-A and FGF-2 is important in the determination of stem cell lineage commitment from very early on in development.

The individual effects of Activin-A and FGF-2 in gene expression in hESCs were compared. The expressions of selected markers of all the 3 germ layers were compared at day 10 in hUES7 cells that had been differentiated through Activin-A exclusion and FGF-2 exclusion (figure 4.7). The selected endodermal markers were Gata4, Sox17, Sox7, FoxA2 and Chordin while mesodermal markers were TBX6, Mixl1, EOMES, Brachyury and FGF-5. The selected ectodermal markers were Pax6, Beta III Tubulin, Nestin and Noggin. There was a generally similar trend of very low gene expression for all the 3 germ layer markers in cells grown without FGF-2 for 10-days. However, although the expression of endodermal and ectodermal markers was equally low in Activin-A deficient cells, there was a significantly high upregulation of mesodermal markers in cells grown without Activin-A. This suggests that in pluripotent hESCs, Activin-A might be involved more in the promotion of endodermal and ectodermal differentiation of the cells than it is in promoting their differentiation towards the mesodermal lineage.
Figure 4.7. The expression of germ layer marker transcripts in hUES7 cells at Day 10 of two separate Activin-A and FGF-2 removal differentiation protocols. q-RT-PCR was performed as described in section 2.7. on cDNA harvested from p28 pp5 hUES7 and p30 pp5 MAN7 cells. There was insignificantly more expression of endodermal markers in the Activin-A exclusion protocol than in the FGF-2 protocol. Very significant upregulation of mesodermal markers was observed in cells grown in the absence of Activin-A than those in the absence of FGF-2. There was insignificant ectodermal marker expression in cells from both differentiation protocols. p = number of cell passages on MEFs. pp = cell passages on feeder-free culture conditions. Error bars= standard deviation.
Working with my colleague Tao Wang, germ layer marker expression levels were also observed in cells that had been cultured in the simultaneous absence of both Activin-A and FGF-2. Basically, pluripotent cells were cultured in FF medium (table 2.1) under normal feeder-free conditions but Activin-A and FGF-2 were withdrawn from the medium. Media was changed daily without the cells being passaged. Although RNA was collected at days 0, 2, 4, 6 and 7, gene expression analysis was performed through q-RT-PCR for cDNA collected at days 0 and 7 (figure 4.8) which are the respective start and end points of the protocol. Day 0 is representative of the pluripotent phase of the protocol. There was no expression of the endodermal markers Sox7, Sox17 and FOXA2 at either time points in the protocol and negligible expression of Sox7 at Day 0 (figure 4.8a). However, in contrast, endodermal markers GATA4 and Chordin were significantly upregulated at day 7 (figure 4.8a). There were no significant differences in the expression levels of mesodermal markers Eomes, Mixl1 and Brachyury (T) at either time points of the protocol (figure 4.8b). There was however a significant half-fold down regulation of TBX6 as well as a significant up regulation of FGF-5 (figure 4.8b). Interestingly, all the 4 ectodermal markers (Nestin, Pax6, Noggin and Beta-III-Tubulin) were all significantly upregulated (figure 4.8c) suggesting that cells were differentiating towards the endodermal lineage.
*Figure continues overleaf*
Figure 4.8. The expression of germ layer marker transcripts in hUES7 cells at Days 0 and 7 of a combined Activin-A plus FGF-2 exclusion differentiation protocol. hUES7 and MAN7 cells were cultured for 7 days in the absence of both Activin-A and FGF2. q-RT-PCR was performed as described in section 2.7 to quantify mRNA expression of different endodermal, mesodermal and ectodermal markers in pluripotent cells at day 0 and differentiated cells derived from them at day 7. a) Endodermal b) mesodermal and c) ectodermal differentiation marker transcript expression levels. All experiments were performed in triplicates and ΔΔCt values were averaged. Error bars, mean ± SEM (n = 2). A student’s t-test was used for the statistical analysis. *, P < 0.05. **, P < 0.005. ****, P < 0.0005.
4.2.4 Embryoid body differentiation

The development of pluripotent stem cells into three dimensional clusters of cells called embryoid bodies (EBs) is one of the most commonly used methods of stem cell differentiation. When stem cells are grown in the absence of substrate matrices and growth factors, they naturally cluster together to form these cell aggregates which mimic in vivo embryo development and can give rise to progenitor cells of all the three primary germ layers (Shen and Leder, 1992). Due to their ability to reproduce cell differentiation events during embryo development, EBs are very important in developmental biology (Hopfl et al., 2004). In this section, results of HESC differentiation through the formation of EBs are reported (see 2.3.2). Day 0 denotes pluripotent stem cells on the first day of the protocol. cDNA derived from these RNA samples was used in q-RTPCR experiments to investigate the expression patterns of the 7 target genes in the differentiated cells. This protocol was carried out in both hUES7 (figure 4.9a) and MAN7 (figure 4.9b) stem cells. There was an overall down regulation in the expression of the 7 target genes in EBs formed from both hUES7 (figure 4.9a) and MAN7 (figure 4.9b) stem cells. However, in comparison to the rate of gene expression down regulation during growth factor removal differentiation protocol, loss of gene expression during the EB differentiation protocol was clearly greater and more rapid.
Figure 4.9. Target gene transcript expression in HUES7 and MAN7 stem cells during a 10 day EB protocol. qRT-PCR (see section 2.7) was performed on EBs formed from p31 pp3 hUES7 and p28 pp3 MAN7 cells. a) Target gene expression in hUES7 differentiated cells showing rapid loss of most of the genes except for TBL1XR1 and TCF7L2 b) Gene expression in MAN7 differentiated cells showing massive down regulation of all the genes at day 10 except for CCND1. p - number of cell passages on MEFs. pp - cell passages on feeder-free culture conditions. Error bars- Standard deviation.
4.2.5 Gene expression relative to Oct4

It is well established that when pluripotent stem cells undergo differentiation, core stem cell maintenance factors (Oct4, Nanog and Sox2) are down regulated because the cells begin to acquire different lineages (Tai et al., 2005; Wang et al., 2009; Kallas et al., 2014). It is on this basis that new candidate potential pluripotency and self-renewal regulatory genes should follow the expression patterns of the core pluripotency regulatory factors. Oct4 is one member of the human embryonic stem cell maintenance gene triplets. As already stated in the introduction chapter, one of the hypothesis of this study is that since there are other yet to be identified genes involved in stem cell regulation which may share expression patterns with the core regulatory transcription factors of hESCs; Oct4, Nanog and Sox2. Therefore, expression levels of the 7 target genes were compared to the expression of Oct4 in pluripotent stem cells. The gene expression of each target gene in either hUES7 or MAN7 cells was normalised to Oct4 expression in the same cell line. The delta Ct value of each gene was therefore plotted relative to Oct4 Ct value which was taken as a single unit (1). Overall, although the expressions of target genes in both hUES7 (figure 4.10a) and MAN7 (figure 4.10b) cells were considerably below Oct4 expression, all the genes were found prevalent in these cell lines with the exception of TBL1XR1 in hUES7 cells (figure 4.10a) and RBM15 in MAN7 cells (figure 4.10b). It would have been interesting to see how the expression of these genes compares with Oct4 expression in differentiated cells. Unfortunately, this could not be done due to lack of cDNA.
Figure 4.10. Target gene transcript expression relative to Oct4. q-RT-PCR was performed as described in section 2.7 on cDNA harvested from p28 pp5 hUES7 and p30 pp5 MAN7 cells to determine mRNA transcript expression levels relative to Oct4 transcript expression level. a) Target gene expression in hUES7 cells showed that all target genes were expressed except TBLIXR1. b) Pluripotent MAN7 cells expressed all target genes. p = number of cell passages on MEFs. pp = cell passages on feeder-free culture conditions. Error bars = Standard deviation.
4.2.6 Microscopic evaluation of differentiated hESCs

The differentiation of human embryonic stem cells brings changes to the morphology of stem cells as the cells adapt to new culture conditions. The conditions may be the absence of growth factors in cell media or the growth of the cells in different media as is the case during growth factor removal and EB differentiation protocols respectively. To investigate the effects of Activin-A removal, FGF-2 removal and EB differentiation on the morphology of stem cells, microscopy was performed to capture phase contrast images of the cells at days 0, 5 and 10 of each protocol (figure 4.11). There were pronounced differences observed in the morphology and cell proliferation patterns of cells cultured in these various differentiation conditions. Cells grown in Activin-A deficient media in both hUES7 and MAN7 stem cells showed changes in morphology (figure 4.11a-b). Whereas the cells appeared pluripotent and evenly distributed at day 0, by day 5, some of the cells had died while others formed brownish clusters of somewhat differentiating cells (shown by red arrows). By day 10, cells became less dense and smaller and deviated more from normal stem cell morphology, a signal of cell differentiation. There were also gaps observed in between cells which were probably created by loss of cells due to death. Cells undergoing FGF-2 removal induced differentiation showed normal feeder-free morphology at the beginning of the protocol in both hUES7 and MAN7 cells. In-fact, mid-way through the protocol (day 5), most of the cells still maintained normal stem cell morphology although slight cell colour changes were observed (figure 4.11 c-d). By day 10, these cells had become lighter in colour and displayed gradual loss of stem cell morphology (shown by purple arrows). Surprisingly, most of the differentiating cells were still intact and adhered to the culture dish plate. The differentiation of both hUES7 and MAN7 stem cells in suspension led to the formation of nicely rounded embryoid bodies 10 days after differentiation (figure 4.11 e-f). The EBs were almost spherical by the final day of the protocol and showed well-defined ridges.
Figure 4.11. Phase-contrast images of hUES7 and MAN7 undergoing differentiation. Cells were cultured as described in sections 2.2.1-2.2.2 without either Actin-A or FGF-2 (see 2.3.1) or in suspension into EBs (see 2.3.2) for 10 days. Images were captured at days 0, 5 and 10. a) p28 pp5 hUES7 and b) p30 pp5 MAN7 cells undergoing a 10 day Activin-A withdrawal induced differentiation. c) p28 pp5 hUES7 and d) p30 pp5 MAN7 cells differentiated for 10 days in FGF-2 deficient media. EBs formed from the differentiation of e) p31 pp3 hUES7 and f) p28 pp3 MAN7 cells in suspension culture dishes in MEF media. Scale bars=100μm (a-d) and 200μm (e-f). p = number of times cells were passaged on MEF feeders and pp = cell passage number on feeder-free. D0 = Day0, D5 = Day5 and D10 = Day10.
4.3 Expression of YY1AP1 and YY1 in hESCs

The differentiation of 7 target genes with confirmed expression in human embryonic stem cells was important in evaluating changes in the expression patterns of these genes as stem cells differentiate. With a few exceptions, most of the genes showed steady down regulation during differentiation in both hUES7 and MAN7 stem cells (figure 4.12a-g). There was a general trend of gene expression down regulation albeit not in a consistently correlative fashion. With more than one candidate to choose from for further experimentation, the next task was to choose one gene which could be further investigated for any functions or involvement in stem cell regulation. Investigating more than one gene would have been too big a task. For this reason, a gene called YY1AP1 was selected as a candidate gene on which further experiments were performed. This selection was solely based on its consistent down regulation during the various differentiation protocols (figure 4.12g). It is worth noting that there were other potentially good candidate genes which could have been selected, for the same reason as YY1AP1, for further investigation. These include genes such as NME1 and TCF7L2 (figures 4.12c and 4.12f respectively) which were down regulated consistently during all the differentiation protocols in both hUES7 and MAN7 cells.
4.3.1 YY1AP1 is consistently down-regulated during differentiation

Figure 4.12. Combined target gene expression in hUES7 and MAN7 cells during Activin-A, FGF-2 and EB differentiation protocols. q-RT-PCR (see section 2.7) comparing transcript expression levels of target genes in hUES7 and MAN7 cells during Activin-A and FGF-2 removal as well as EB differentiation protocols. From these genes, YY1AP1 was selected for further experimentation due to its consistent down regulation during all the differentiation protocols. Error bars=Standard error.
4.3.2 Correlations between YY1AP1 and YY1 expression during differentiation

YY1AP1 is a co-transcriptional activator gene with known functions in cell cycle regulation (Li et al., 2007). It activates the transcription of its target and binding partner gene YY1 (Wang et al., 2004). Unlike YY1AP1, a lot is known about YY1, a transcriptional activator or repressor gene involved in cell proliferation, cell survival, cell cycle regulation, and cell differentiation (Gronroos et al., 2004; Sui et al., 2004; Vella et al., 2011; Gregoire et al., 2013; Morikawa et al., 2013) among many of its functions. In order to gain insight into how YY1AP1 expression compares with that of YY1 during stem cell differentiation, the expression levels of these two genes were compared in the three distinct differentiation protocols; the Activin-A removal, Activin-A plus FGF-2-removal and the chondrogenic differentiation protocol. cDNA collected at days 0, 5 and 10 from the Activin-A differentiation protocol (see 2.3.1) was used to perform q-RT-PCR for YY1AP1 and YY1 in hUES7 cells. q-RT-PCR was also performed for the two genes on cDNA collected at days 0, 1, 5, 6 and 7 of a 7 day Activin-A plus FGF-2 removal differentiation protocol in hUES7 cells. In another differentiation protocol, using cDNA (kindly donated by Dr. Aixin Cheng in our lab) from a directed differentiation chondrogenic protocol (see 2.3.3), gene expression levels of both YY1AP1 and YY1 were compared through q-RT-PCR at days 0, 4, 8 and 13. During Activin-A removal induced differentiation; both YY1AP1 and YY1 were steadily down regulated in hUES7 cells (figure 4.13a). Although the level of YY1AP1 was more than 2 times higher than that of YY1 at the start of the protocol (day 0), the down regulation patterns of these two genes were almost identical over the 10 days. During differentiation induced by the simultaneous removal of Activin-A and FGF-2 from stem cell media in hUES7 cells, YY1AP1 and YY1 were again both down regulated over 7 days (figure 4.13b). During the differentiation of hUES7 cells into chondrocytes, the level of YY1 was higher than that of YY1AP1 at the start of the protocol. However, both genes were down regulated to an almost similar at the last day of the protocol (day 13) (figure 4.23c). Overall, there was a very high similarity in expression patterns between YY1AP1 and YY1 during the three differentiation protocols.
Figure 4.13. The expression of YY1AP1 and YY1 transcripts during the withdrawal of Activin-A, Activin-A plus FGF-2 and during a Chondrogenic differentiation protocol in hESCs. Cells were cultured (see sections 2.2.1-2.2.2) and Activin-A, FGF-2 and Chondrogenic differentiation protocols (see sections 2.3.1-2.3.3) were carried out in p28 pp5 hUES7 cells (for Activin-A and FGF-2 removal protocols) and p31 ppII nUES7 cells (for the chondrogenic protocol). q-RT-PCR (see section 2.7) was performed to determine mRNA expression levels a) There were similar down regulation expression trends between YY1AP1 and YY1 during Activin-A withdrawal induced differentiation of hUES7 cells. b) YY1AP1 and YY1 were down regulated in a similar fashion when Activin-A and FGF-2 were withdrawn from stem cell media during hUES7 cell differentiation c) YY1AP1 and YY1 were both down regulated during the chondrogenic differentiation of hUES7 stem cells. n=3. Error bars = Standard deviation.
4.3.3. YY1 expression during human embryo development

YY1 is a ubiquitously expressed transcription factor in human cells (Shi et al., 1991) whose transcriptional activity is activated by YY1AP1. Human embryonic stem cells are derived from ICMs of about 5 day old blastocysts of the developing embryo. Some genes and transcription factors have been shown to be prevalent and more active at primitive stages of human embryo development. Other genes have been shown to be more active toward the later stages of embryogenesis while others remain consistently active throughout embryogenesis in humans and mice (Kimber et al., 2008). Our lab undertook a study to analyse gene expression patterns throughout human embryo development in order to identify genes involved in early and late cell fate decisions (Smith et al., in print). This study investigated all the different stages of embryogenesis from the 4 cell, 8 cell, 16 cell, blastocyst, ICM, oocyte and TE stages. The investigation was based on mRNA microarray studies that have been previously undertaken on human embryos donated, with consent from patients undergoing IVF treatment at St. Marys Hospital in Manchester. RNA samples were collected from the different stages (4 cell, 8 cell, blastocyst, ICM, oocyte and TE) of embryo development. Performed by Dr. Helen Smith in our lab, YY1 expression was compared through Poly-A PCR (figure 4.14) (see manuscript for procedure). YY1 was shown to be expressed across all the embryo developmental stages as well as in both hUES3 and hUES7 cells (figure 4.14). Although its expression was relatively constant throughout developmental stages, it was massively up regulated in the blastocyst (figure 4.14a). The expression levels of YY1 were significantly higher in both hUES3 and hUES7 cells than in any of the embryo developmental stages (figure 4.14b). Interestingly however, YY1 was expressed almost 8 fold more in hUES7 cells than in hUES3 cells (figure 4.14b)
Figure 4.14. The expression of YY1 transcript during human embryo development. RNA samples were collected at the 4 cell, 8 cell, blastocyst, ICM, oocyte and TE stages of development from embryos donated by IVF patients. Poly-A PCR was then performed (Smith et al., in print) to determine YY1 transcript expression levels at each stage of development and in hUES3 and hUES7 stem cells. a) Poly-A PCR results showing that YY1 is expressed across all the stages of human embryo development but more expressed in the blastocyst. b) YY1 is more expressed in hUES3 cells and hUES7 cells compared to the different stages of human embryo development. n=3. Gene expression was normalised to Beta-Actin.
4.3.4 Intracellular localisation of YY1AP1 with Oct4 and YY1 in hESC and iPSCs

YY1AP1 and YY1, like Oct4, are intracellular transcription factors that are located in the nuclei of hESCs. The multi-functional YY1 has lots of binding partners and therefore lots of binding sites on its promoter region. iPSCs resemble hESCs morphologically and in terms of gene expression (Takahashi et al., 2007) as well as on differentiation cell fate decisions (Hu-B-Yet et al., 2010; Guzzo et al., 2013). Therefore, there should not be too much disparity between YY1AP1 and YY1 gene expression patterns in hESCs and iPSCs. In order to check if YY1AP1 and Oct4, and YY1AP1 and YY1, occupy the same spatial positions in hESCs and iPSCs, their co-localization capacities were tested through immunofluorescence staining (see 2.8.1 – 2.8.2). In brief, this involved staining for the first gene, re-blocking the cells and performing staining for the second gene using differently fluorescent labelled secondary antibodies with different emission wavelengths. Staining results showed that YY1AP1 and Oct4 expression patterns correlate and overlap in the nuclei of hESCs (figure 4.15a). Nuclear co-localization was also observed between YY1AP1 and YY1 in hESC nuclei where there was complimentary overlap between the fluorescence indicative of these two genes (figure 4.15b). There was clear and pronounced correlation between YY1AP1 and Oct4 localization in iPSC nuclei (4.15c). Co-localization staining between YY1AP1 and YY1 was also shown in iPSCs (4.15d). However, overlap between the two genes in the iPSCs was not as pronounced as it was in hESCs. Some cells, shown by white arrows, showed stronger correlation between YY1AP1 and YY1 than other cells. To check for cross-reactivity between YY1AP1 2° Ab (anti-goat IgG) and Oct4 2° Ab (anti-mouse IgG), immunostaining was performed for Oct4 using anti-mouse IgG 2° Ab (figure 4.15e) and for Oct-4 using anti-goat IgG 2° Ab (figure 4.15f). Staining was also performed for YY1AP1 using anti-goat IgG 2° Ab (figure 4.15g) and for YY1AP1 using anti-mouse IgG Ab (figure 4.15h). These secondary antibodies were checked for cross-reactivity with the alternative primary antibody and it was found that there was no cross-reactivity between the Oct4 and YY1AP1 2° Ab as no staining was observed for either protein using the other’s 2° Ab.
Figure 4.15. YY1AP1, YY1 and Oct4 co-localisation staining in hESCs and iPSCs. Cells were cultured as in sections 2.2.1-2.2.2 and co-localisation staining was performed (see section 2.8.2) on hUES7 and iPSCs derived from human dermal fibroblasts. a) Oct4 and YY1AP1 co-localization immunostaining in hUES7, p28, pp4 cells. b) YY1AP1 and YY1 co-localization immunostaining in hUES7, p28, pp4 cells. c) Oct4 and YY1AP1 co-localization immunostaining in ZK2012L, p22, pp9 iPSCs. d) YY1API and YY1 co-localization immunostaining in ZK2012L, p22, pp9 iPSCs. e) Staining with mouse anti-Oct4 and mouse IgG 2’ Ab showed nuclear localisation. f) Staining with mouse anti-Oct4 and goat IgG 2’ Ab was negative. g) Staining with goat anti-YY1API and goat IgG 2’ Ab showed nuclear localisation. h) Staining with goat anti-YY1API and mouse IgG 2’ Ab was negative. All negative control stainings are shown as inserts. p = number of passages on MEFs, pp = number of passages on feeder-free system. Scale bars = 200 μm.
Further evaluation of YY1AP1 and YY1 expression levels was performed in iPSCs (line ZK2012L) through qRT-PCR determination of the gene expression levels. In brief, iPSCs were cultured under normal stem cell feeder-free culture conditions (see chapter 2.2.1) on culture plates that had been pre-coated with fibronectin. Little difference was observed in the expression levels of the YY1AP1 and YY1 in iPSCs. (figure 4.16). Furthermore, the expression levels of the two genes in iPSCs cells were compared with their expression levels in pluripotent hUES7 cells (figure 4.16). Compared to YY1AP1, there was more than 8 times enrichment of YY1 in hUES7 cells. However, when the expression levels of both genes in hUES7 cells were compared to their expression levels in iPSCs, there was a massive difference observed. Both genes were significantly more expressed in hUES7 cells than in the iPSCs.

Figure 4.16. YY1AP1 and YY1 expression in hUES7 hESCs and ZK2012L iPSCs. Both cell types were cultured as described in sections 2.2.1-2.2.2 and q-RT-PCR was performed (see section 2.7). Comparison of YY1AP1 and YY1 expression levels in p28 pp6 hESCs and p22 pp10 iPSCs showed that both genes are greatly more enriched in hESCs than iPSCs. n: 2. p - number of passages on MEFs, pp - number of passages on feeder-free system
4.4 Discussion

The pluripotency of stem cells has generated a lot of interest in their application to medical research. Before their use in any experiments, the expression of pluripotency-associated genes is assessed in hESCs to confirm their pluripotency. Exceptions to this are experiments which require differentiating the cells as starting material, in which case, the pluripotency of the cells is not important. Therefore, in all experiments undertaken in this study, cells were tested for proper cell morphology and pluripotency prior to their use. Qualitative PCR was used to identify and eliminate genes that were not expressed in stem cells. From the initial 10 target genes that were retained after bioinformatics, functional and literature analyses; RIF1, Myst4 and DGCR8 showed no expression in hESCs after qualitative PCR. RIF1 is a transcriptional regulatory gene with established direct protein interaction with Oct4, Nanog and Sox2 in embryonic stem cells (Zoumaro-Djayoon et al., 2011). Myst4 is a marker of pluripotency and self-renewal of neural stem cells (Sheikh et al., 2012). It was therefore expected that Myst4 and RIF1 would show expression in ESCs. There are numerous factors that could have led to their non-detection in the cells. These may include poor primer specificity, reaction tube contamination, poor primer design and damage to template DNA. DGCR8, a microRNA biogenesis gene involved in hESC self-renewal silencing (Landthaler et al., 2004; Wang et al., 2007) and Myst3, a mouse haematopoietic stem cell self-renewal and regulatory gene (Scandura et al., 2002; Di Croce 2005; Katsumoto et al., 2008) were also not detected.

Growth factor removal induced differentiation of stem cells is a common way of inducing stem cells to change lineage. Activin-A induces human embryonic stem cells to differentiate towards endodermal and mesodermal lineages (Sumi et al., 2008) while FGF-2 promotes their differentiation towards the ectodermal lineage. FGF has been shown to have the capacity to maintain stem cell proliferation in the absence of mouse embryonic fibroblast (MEFs) and conditioned medium (Wang et al., 2005). FGF can also maintain undifferentiated cell proliferation through its interaction with BMP antagonist, noggin, to supress BMP signalling activity (Xu et al., 2005). In contrast, the Activin/Nodal signalling pathway cooperates with the BMP signalling pathway to maintain stem cell self-renewal (Sumi et al., 2008). However, synergistically, the
Activin/Nodal signalling pathway regulates the FGF, Wnt and BMP pathway during hESC pluripotency maintenance (Xiao et al., 2006).

The application of Activin-A and FGF-2 differentiation protocols in this study revealed significant induction of cell differentiation and morphological changes. During Activin-A withdrawal induced differentiation in both hUES7 and MAN7 cells, Oct4, Nanog and Sox2 were abruptly down regulated, an indication of loss of pluripotency. Activin-A has been shown to be important in the maintenance of Oct4 and Nanog expression during hESC pluripotency and proliferation in an undifferentiated state (Xiao et al., 2006). Therefore, its ablation led to the negative regulation of Oct4 and Nanog. Contrastingly, Activin-A removal differentiation led to the upregulation of Gata4 and Mix1. Gata4 is mesendodermal transcription factor that is important in the regulation of gene expression during heart development (Stennard et al., 2003). Mix1 is also a mesendodermal marker of primitive streak differentiation (Sumi et al., 2008). Loss of Activin-A in hESCs therefore appear to push the cells to differentiate towards the mesendodermal lineage. CDX2, an endodermal marker of gastrointestinal differentiation was also up regulated during Activin-A removal differentiation. This is another indication of cell differentiation towards the late endodermal lineage upon the loss of Activin-A. There was massive up regulation of CDX2 at the middle of the Activin-A removal differentiation protocol (figure 4.4a) which could have been a result of intracellular growth factor dependent effects. Activin-A may be part of the regulatory complex of CDX2 or CDX2 signalling.

During the Activin-A removal differentiation in both hUES7 and MAN7 cells, only CCND1, TBL1XR1 (both in hUES7 cells) and TCF7L2 in MAN7 cells showed expression trends dissimilar to that of the other genes (figure 4.4). CCND1 is an important cell cycle progression regulatory gene in hESCs. Therefore, the lack of regulation of CCND1 was hardly surprising considering that it is a cell cycle progression regulatory gene and is needed even for the regulation of cell cycle in differentiated cells. TBL1XR1 is a transcriptional repressor gene (Zhang et al., 2006). Its increase upon loss of Activin-A might have been due to an increase in the transcriptional activities of the cells as they underwent lineage change. It is also possible that Activin- A negatively regulates TBL1XR1. Myst3 and NME1 were up regulated in MAN7 cells (figure 4.5b) during FGF-2 withdrawal induced differentiation, a potential indication that the
regulation of these genes in hESCs does not require the presence of FGF-2. Comparison of the effect of Activin-A removal to FGF-2 removal on gene expression in both hUES7 and MAN7 cell lines showed a general trend of down regulation of most of the genes with a few exceptions (figure 4.6). For most genes, there was less expression of the genes at day 10 of both protocols than there was at the beginning of these protocols. This was an indicator that these two factors are needed for the regulation most of these target genes and therefore might be involved in similar pathways with them. There was faster down regulation of gene expression in cells that were cultured in the absence of Activin-A than there was in those cultured in FGF-2 deficient conditions. During growth factor induced differentiation protocols, cells are forced to adapt to unfamiliar culture conditions. Therefore, quantitative changes that are observed during these protocols are due to intracellular growth-factor dependent effects due to the possible involvement of the removed growth factors in the regulation of the target genes.

Oct4 is an important regulator of stem cell pluripotency that is abundantly expressed in non-differentiated hESCs (Niwa et al., 2000; Boyer et al., 2005; Donohoe et al., 2009). Normalisation of target gene expression to Oct4 in both pluripotent hUES7 and MAN7 cells showed reasonably high expression of the target genes (figure 4.10). This was indicative of the potential pluripotency regulatory roles of the target genes. The hypothesis of this study was that there are other genes involved in pluripotency regulation of hESCs and that these genes are likely to be expressed in a similar fashion to Oct4 and Nanog. Oct4 normalisation of the expression of target genes correlated well with this hypothesis.

EB differentiation of cells led to the consistent down regulation of most of the target genes with complete loss of gene expression at day 10 of the protocol. This suggests that the differentiation of cells through the formation of EBs results in faster and more severe loss of gene expression when compared to both the Activin-A and FGF-2 exclusion differentiation protocols. This was more evident in the heavy loss of the expression of all the genes at day 5 of the EB protocol, even though this was merely the mid-point of the protocol.

Both Activin-A and FGF withdrawal differentiation led to no obvious changes in the expression of endodermal markers such as Sox7, Sox17 and Chordin.
Based on these findings, Activin-A and FGF-2 are therefore needed and play roles in endodermal differentiation signalling complexes of hESCs. Indeed, it has been shown that both Activin-A and FGF-2 are crucial for definitive endodermal differentiation (Vallier et al., 2005; Sui et al., 2012) and that their effects during endodermal differentiation are more potent synergistically than individually. While FGF2 withdrawal had no effect on the expression of mesodermal markers, Activin-A withdrawal led to massive up regulation of mesodermal markers such as TBX6, Mixl1 and Brachyury (figure 4.7). Activin-A is part of the P13/AKT pathway. It has been shown that the inhibition of the P13/AKT signalling pathway induces differentiation of ES cells towards the mesendodermal lineage (Armstrong et al., 2006). Therefore, the absence of Activin-A negatively regulates the P13/AKT pathway, promoting cells to differentiate towards endoderm and mesoderm. In contrast, the removal of FGF-2 does not promote endodermal or mesodermal differentiation of the cells.

The simultaneous removal of Activin-A and FGF-2 led to no significant expression of mesendodermal markers. Activin-A and FGF-2 have been previously shown to mimic anterior endodermal expression during heart formation (Sugi et al., 1995). The FGF signalling pathways also help Activin-A during the formation of the definitive endoderm from hESCs (Sui et al., 2012). Therefore, in the absence of these two growth factors, endodermal differentiation is highly unlikely. Furthermore, Activin-A cooperates with FGF-2 to promote the differentiation of hESCs towards the endodermal lineage (Xu et al., 2011). There was an up regulation of all ectodermal markers during the combined removal of Activin-A and FGF-2. This suggests that cells cultured under Activin-A and FGF-2 deficient conditions tend to differentiate toward the ectodermal lineage. Therefore, the absence of both growth factors results in loss of stem cell pluripotency. It has been shown that during the maintenance of stem cell self-renewal, FGF-2 activates some components of the Activin-A signalling pathway (Vallier et al., 2005).

After differentiation, from the 7 target genes, Yin Yang 1-Associated protein 1 (YY1AP1) was selected for further experimentation because of its consistent down regulation during differentiation (figure 4.12g). YY1AP1 has been shown to trigger the transcriptional activation of Yin Yang 1 (YY1), a transcriptional repressor gene (Wang et al., 2004). YY1 and YY1AP1 co-localize in hESC and IPSC nuclei (figure 4.15). The
expression of YY1 is evaluated during human embryo development. YY1AP1 and YY1 are binding partners who have been shown to interact directly with each other. It was therefore not surprising that, at the transcript level, both transcription factors share similar expression patterns during the differentiation of human embryonic stem cells. YY1 has been shown to be elevated during cancer progression (He et al., 2011; Zhang et al., 2011; Zhu et al., 2011), a suggestion that it is primarily involved in cell proliferation than differentiation, hence the reason why it is lost along with its binding partner during loss of cell proliferation when cells differentiate.

YY1 is widely expressed across all the stages of human embryo development as it has been shown in our lab (figure 4.14). However, it was found to be more prevalent at the blastocyst stage than any other stage of embryogenesis. Comparison of YY1 expression during embryogenesis to its expression in pluripotent hUES3 and hUES7 stem cells showed that this transcription factor is expressed at a higher level in hESCs than during embryo development. However, it is important to note that YY1 transcript level was obtained by poly-A-PCR during embryo development and by q-RTPCR in hESCs which are two slightly different PCR procedures. The level of YY1 expression in hUES3 stem cells was almost double that in the blastocyst. There was also over 10-fold more expression of YY1 in hUES7 stem cells than in the hUES3 cells. Therefore, it appears that there is more gene expression in hESCs after their isolation and maintenance in vitro than in their natural habitat of the ICM. This can be attributed to enhanced stem cell maintenance and culture methods and well as supplementation from growth factors.

Gene expression analyses showed that YY1AP1 and YY1 are expressed in iPSCs but their expression levels in hESCs are much higher than those in iPSCs (figure 4.16a-b). However, just because a gene is highly expressed in stem cells or co-localises or interacts with Oct4 does not necessarily mean that such a gene is involved in stem cell maintenance. In-fact, some genes that interact with Oct4 are actually down regulated by Oct4 or may be inhibitors (Chambers and Smith, 2004). Oct4, Nanog and Sox2 can bind to promoters of many different genes which encode transcription factors that have been known to partake in stem cell regulatory mechanisms (Boyer et al., 2005; Donohoe et al., 2009). This, therefore, means that both the high expression and Oct4 interaction of a gene increases the probability of that gene being involved in
stem cell self-renewal and maintenance of pluripotency. This therefore means that the investigation of potential novel pluripotency associated genes should involve the assessment of these potential genes with core pluripotency transcription factors. Involvement of these genes in pluripotency signalling networks would also help unveil their involvement in stem cell regulation and maintenance. It is these factors that guided the next two results chapters of this project, where the effect of YY1AP1 ablation on core pluripotency genes and the interaction of the YY1AP1 protein with proteins of these genes were investigated.
CHAPTER 5

YY1AP1 SHORT-HAIRPIN KNOCKDOWN IN hESCs

As it was shown in Chapter 4, the differentiation of human embryonic stem cells leads to a decline in YY1AP1 transcript levels. Loss of pluripotency correlated with a decline of YY1AP1 and the core stem cell maintenance transcription factors; Oct4, Nanog and Sox2. YY1AP1 is a regulator of its binding partner, YY1, which is involved in transcriptional repression and activation of target genes (Wang et al., 2004; Ohtomo et al., 2007). However, the role of YY1AP1 in stem cell pluripotency and self-renewal has not been previously studied. Its interactions with major pluripotency transcription factors and genes are also poorly understood. It has been previously reported that YY1 negatively regulates the tumour suppressor gene p53 in U2OS cells (Gronroos et al., 2004; Sui et al., 2004). This, therefore, suggests a major role for YY1 in cell proliferation which is a major property of stem cells. Whether this role extends to YY1AP1 is unknown. In this chapter, the transfection of the YY1AP1 shRNA construct into two different cell types, 293FTs and MCF-7s, will be examined better. Knockdown was carried out by the use of an inducible vector, pLVTHM, to carry the construct into the cells. This vector had a GFP promoter for driving GFP expression which enabled the identification of transfected cells. The promoter also had a Tet0 promoter which drives doxycycline expression to initiate target gene knockdown. Knockdown was initiated in the MCF-7 cell line to check the efficiency of the plasmid vector in delivering the YY1AP1 shRNA into cells and assess the efficacy of the construct in knocking down YY1AP1 before knockdown was initiated in hESCs themselves. The method of transfection used to introduce the shRNA carrying plasmid into the MCF-7s was lipofectamine. Lipofectamine was used because it is an animal-origin free formulation with low toxicity to the cells and high transfection efficiency. Because Lipofectamine is made up of positively charged lipids, it forms a complex with negatively charged DNA. This complex then fuses with the plasma membrane, allowing the DNA to enter the cell cytoplasm and ultimately the nucleus. This chapter will also report YY1AP1 construct transfection into hESCs and how a stable knock down hUES1 cell line was created.
5.1 Cloning and evaluation of a pLVTHM-YY1AP1 inducible vector system

5.1.1 Construct shRNA design and vector preparation

The success of every knockdown experiment relies heavily on the quality and effectiveness of knockdown constructs. This is achieved by the successful cloning, digestion and ligation of the knockdown construct into a vector. The structure of the plasmid vector used can be seen in section 2.10 (figure 2.4). The plasmid has an EF1-alpha promoter for driving GFP expression, a Tet0 promoter which activates shRNA expression upon Dox addition and an H1 promoter which drives shRNA expression. As described in section 2.9, shRNAs were designed, annealed and phosphorylated (2.9.1) and the plasmid vector was cloned, purified and digested (2.9.2) prior to its de-phosphorylation. De-phosphorylation enhances the ligation of the plasmid vector to the phosphorylated shRNA construct. The plasmid vector was digested with ClaI and MluI restriction endonucleases. In order to check for the efficiency of the digestion of the plasmid vector, the vector DNA was run through a 1.5% agarose gel. The digested vector DNA was separated in parallel with the undigested vector DNA. The gel electrophoresis results showed that the pLVTHM plasmid had been successfully digested. A single band was observed for the digested pLVTHM plasmid vector (green box) (figure 5.1). This is because ClaI and MluI restriction sites are very close together on the vector backbone with only 17bp between them. Prior to digestion, a plasmid vector is usually circular but digestion leads to a linearized DNA structure. Due to its circular structure, it contains both relaxed and supercoiled bands which separate and migrate at different speeds on the gel due to their molecular weight and structural differences. This is why there were two bands observed for the undigested plasmid vector (in red boxes) (figure 5.1). The positions of the plasmids on the molecular weight hyperladder are in agreement with the biological size of the pLVTHM plasmid vector (11085bp).
Figure 5.1. Gel electrophoresis of undigested and digested pLVTHM plasmid vectors. Gel electrophoresis was performed as described in sections 2.6.5-2.6.6 for the digested and undigested plasmid vectors on 1.5% agarose gel. Double bands were observed for the supercoiled undigested vector with only one band showing for the linear digested vector. The undigested (untreated) vector was used as a control. 1 = DNA hyperladder; 2 = Undigested pLVTHM plasmid and 3 = Digested (Clal + MluI) pLVTHM plasmid.
5.1.2 pLVTHM YY1AP1 plasmid transfection into 293FT and MCF-7 cells

After digestion, the plasmid vector was ligated with the shRNA constructs. Vectors ligated with both YY1AP1 and B2M constructs were then taken for sequencing. Sequencing was performed using an H1 forward strand sequencing primer. Sequencing showed complete compatibility between the complimentary sequences (figure 5.2). Michigan Cancer Foundation-7 (MCF-7) cells are some of the most commonly used cell types in cell biology. These cells were first derived from breast carcinoma cells of a patient in the early 1970s (Soule et al., 1973) and cultivated into a stable cell line which has been maintained to-date. 293FTs were derived by viral transformation of human embryonic kidney cells (Graham et al., 1977). They grow easily and rapidly. It is for these reasons that they are commonly used as test of principle cells especially in stem cell biology where stems cells are so precious. The expression of YY1AP1 was confirmed in 293FT and MCF-7 cells by gel electrophoresis. Results showed that this gene was expressed in both the 293FT cells (figure 5.3a) and the MCF-7 cells (figure 5.3b). As a proof of principle, the YY1AP1 carrying plasmid vector was transfected into 293FTs and MCF-7s. Transfection was performed using calcium phosphate in the 293FTs (see 2.10.1) and Lipofectamine in the MCF-7s (2.10.2). pmaxGFP was used as a negative transfection control. Transfection efficiency was observed 24hrs post transfection. The 293FT cells were successfully transfected and after 24hrs, they showed isolated colonies of GFP positive cells (figure 5.4a). This was also true for MCF-7 cells (figure 5.4b). In both 293FTs and MCF-7s, the transfection efficiency was very low after 24hrs. It was for this reason that cells were re-transfected for the second time in order to improve the transfection efficiency. This led to an improvement in the intensity of GFP positive cells as shown by MCF-7s 72hrs post transfection. There was a fair number of GFP positive cells observed in the negative control pmaxGFP transfected MCF-7 cells (figure 5.4c). YY1AP1 was then knocked down in MCF7s through the introduction of 0.5ug/ml of Dox into the medium. Dox was present continuously for a period of 5 days and cDNA was harvested at day0, 48hrs and 5days for qRT-PCR analysis. The expression of YY1AP1 was drastically reduced after 10 days of knockdown activation (figure 5.5a) while its expression in the control untreated cells remained relatively constant (figure 5.5b). Even though the transfection approach used
was transient, there was very high KD efficiency of YY1AP1 in the MCF7s (figure 5.5a). This was due to the fact that the cells were re-transfected (double transfection). This significantly increased the population of cells carrying the shRNA and ultimately led to a high rate of YY1AP1 RNA KD in these cells. Double transfection improves transfection efficiency and ultimately knock down efficiency in the cells. Assessment of YY1 expression in the YY1AP1 knocked down MCF7 cells showed that YY1, just like its binding partner YY1AP1, was heavily down regulated after 5 days of YY1AP1 knock down in MCF7s (5.5c). This was an indication that YY1AP1 regulates YY1. Apart from a slight elevation at the 48hr time point, the expression of YY1 in the untreated control MCF7s was relatively constant throughout the 5 day period (figure 5.5d).

Figure 5.2. Positive DNA sequencing of a) ligated YY1API shRNA and b) ligated β2M shRNA (negative control). There was 100% similarity between the sequenced DNA and the original designed shRNA sequences of YY1API and Beta-2M. This verifies insertion into the plasmid vector and that the shRNA sequences were correct.

Figure 5.3 RT-PCR showing YY1API expression in 293-FT and MCF7 cells. Qualitative PCR (see section 2.6) was performed on cDNA from 293FT and MCF7 cells on 1.5% agarose gel, which has been previously found to be the most appropriate percentage gel to use. Results showed that YY1API is expressed in a) 293FTs and b) MCF-7 cells. GAPDH was used as an internal reference gene. Human genomic (hg) DNA and Non-Template Control (NTC) were used as positive and negative controls respectively.
Figure 5.4. Transfection of YY1API plasmid into 293-FT and MCF-7 cells. 293-FTs and MCF-7s were cultured as described in sections and then transfected with the YY1API shRNA carrying plasmid by calcium phosphatase and lipofectamine respectively (see 2.10.1-2.10.2). Bright field microscopy was performed on transfected cells 24, 48 and 72hrs post transfection. Cells transfected with the pmaxGFP plasmid were used as a control a) GFP positive colonies were observed in 293-FT cells 24 hrs post transfection b) GFP positive colonies increased significantly after 72hrs of YY1API transfection into MCF-7s c) pmaxGFP control plasmid transfection in MCF-7s produced GFP positive colonies. Twenty-four, 48 and 72hrs refer to the number of hours after cells had been transfected. Scale bars, 200μm.
Figure 5.5. The expression of YY1API and YY1 in YY1API knocked down MCF-7s. YY1API knock down in transfected MCF-7s was induced by the addition Dox as described in section 2.10.4. q-RT-PCR (see section 2.7) was undertaken to assess YY1API and YY1 mRNA transcript expression levels in the knocked down cells after 0 days, 48hrs and 5 days of knockdown induction. Untransfected MCF-7 cells were used as a control. a) YY1API transcript decreased 48hrs after its knock down. b) YY1API transcript expression in untreated cells is still high after the knockdown experiment c) YY1 transcript is down regulated 48hrs post YY1API knockdown d) YY1 transcript is expressed consistently in control MCF7s. Error bars=Standard deviation.
5.2 Generation of a GFP positive hESC population

5.2.1 Nucleofection of hESCs with pLVTHM-YY1AP1 plasmid

Nucleofection involves the use of an electric shock to push the shRNA plasmid across the cell membrane into the cells. It enables the elimination of problems that are often associated with transfection such as cytotoxicity, poor cell viability and low transfection efficiencies (Zeitelhofer et al., 2007). HESCs and iPSCs were transfected by nucleofection using the Amaxa nucleofection kit and machine (see 2.10.3). Nucleofected cells were cultured in media containing ROCK inhibitor to improve cell attachment and viability. The expression of GFP positive cells was observed daily after 24hrs of transfection. The morphology of the transfected cells was also monitored under an inverted microscope. While a few GFP positive colonies were observed in the transfected hESCs after 24 hrs, these GFP positive colonies increased drastically 48hrs post nucleofection (figure 5.6a). The cell morphology of these cells was typical of normal pluripotent stem cell morphology (figure 5.6a), signalling that the nucleofection had no detrimental effect on the integrity of the cells. The YY1AP1 shRNA carrying plasmid vector was also successfully transfected into iPSCs. The transfected cells showed about 60% of transfection efficiency after 48hrs of nucleofection (figure 5.6b). The nucleofected iPSCs showed normal stem cell morphology when observed under bright field microscopy (figure 5.6b).
Figure 5.6. Nucleofection of YY1AP1 shRNA into hESCs and iPSCs. hUES1 and ZK2012L iPSC cells were nucleofected with YY1AP1 shRNA as described in section 2.10.3. Nucleofected cells were then observed under an immunofluorescent microscope 48hrs after nucleofection. Microscopy showed that GFP positive cells increased significantly after 48hrs of nucleofection in (a) hUES1 hESCs and (b) ZK2012L iPSC cells. In both cell lines, pluripotent stem cell morphology was maintained as shown by phase contrast images. A bright field enlarged image shows that cells had enlarged nuclei. Scale bars= 200μm.
5.2.2 Enrichment of a pure GFP positive cell population

In order to improve the efficiency of transfection, multiple nucleofections were performed on the same cells at-least two times. There are other alternative transfection methods. Viral transduction uses viral vectors to transfect shRNAs into cells. However, because it employs viruses, viral transduction is often associated with cell toxicity and also takes time to perform. Lipofection uses liposome reagents to transport genetic materials into cells. However, it requires extended incubation as well as the optimization of ratios between negatively charged DNA and positively charged lipids. Nucleofection was performed for YY1AP1 shRNA transfected cells as well as the beta-2-microglobulin shRNA transfected control cells and the empty vector control cells. However, results presented in this section are for YY1AP1 shRNA transfected cells. After the first transfection, cells were plated for at-least 48hrs. Nucleofection was then repeated on the cells to increase the number of GFP positive (YY1AP1 shRNA carrying) cells. Double nucleofections greatly enhanced the transfection efficiency. Figure 5.7 shows the expression of GFP in double-nucleofected hUES1 cells over 5 days. After, the second nucleofection; more than 80% of the cells were GFP positive after plating. Moreover, in contrast to cells that had been sorted by FACs and had only been transfected once, cells that were transfected multiple times maintained their GFP expression for a longer period as was evidenced by a high percentage of GFP positive cells 5 days post transfection. Also, hESCs that were transfected multiple times were able to maintain normal stem cell morphology with no changes observed in the attachment and proliferation of the cells (figure 5.7). Prior to re-transfection of the cells, in order to further enrich for a GFP positive hESC population, GFP based cell sorting (FACs) was undertaken (section 2.11). FACs enables the elimination of non-transfected cells so that a more pure GFP positive population of transfected cells can be retained. Sorted cells were then plated on matrigel coated plates and fed with media supplemented with rock inhibitor to allow the cells to quickly recover from sorting induced stress. Despite the great advantage of having a purer population of transfected cells after cell sorting, the size of this population was usually very low compared to the size of the starting population. This was made even worse by the fact that the cells sorted had only been transfected once. In one sorting experiment for the
YY1AP1 shRNA transfected population, only 57.8% of the cells were designated alive (figure 5.8). From this population of live cells, 88.8% were GFP negative while only 5.8% were GFP positive. Despite the low percentage of recovered transfected cells, these cells were able to attach on the matrigel coated plates. The cells did however take longer than usual to expand. They were cultured over at least 2 passages prior to their re-transfection and use in subsequent knockdown experiments.

Figure 5.7. Multiple transfection of hESCs. Nucleofection of hUES1 cells with the YY1AP1 shRNA carrying plasmid vector was performed by nucleofection as described in section 2.10.3. This was then repeated again after cells had been well expanded in culture. Bright field microscopy images of double transfected cells were captured to assess GFP proteins 0 days, 24hrs, 48hrs and 5 days after the initiation transfection. Images showed that double transfection of cells with the YY1API shRNA carrying plasmid vector greatly enhanced transfection efficiency over 5 days. Scale bars=200μm.
Figure 5.8. GFP-based sorting of hESCs. Cell sorting (see section 2.11.1) was performed to enrich for a purer population of GFP positive colonies. a) GFP-positive cell population shown in green (P3). b) P1 denotes a percentage of live cells in the cell population at the start of sorting. P2 shows a percentage of GFP negative cells in that population of live cells while P3 is a percentage of GFP positive cells in that population. c) A fluorescence peak of GFP negative cells. d) A fluorescence peak of GFP positive cells. After sorting of GFP positive cells, the cells were re-transfected to improve transfection efficiency.
5.3 Dox induced knockdown of YY1AP1 in pluripotent hESCs

Doxycycline (Dox) is the most commonly used of all the members of the Tetracycline antibiotic family for disease treatment (Saikali and Singh, 2003). In addition to its use in the treatment of infections, Dox is also commonly used in biological experiments; specifically tetracycline regulated transcriptional activation and repression. During the knockdown of YY1AP1 in hESCs, Dox was used to activate and drive YY1AP1 shRNA expression at the Tet-0 promoter of the plasmid vector. In order to determine the optimal concentration of Dox to use during the knockdown experiments, a Dox titration experiment was undertaken. The objective of this titration experiment was to determine the highest concentration of Dox that could be used without having any detrimental effects on the morphology and pluripotency of the cells. In brief, hUES1 cells were cultured under normal feeder free conditions (see 2.2.2) in media containing 0.1, 0.5 and 1μg/ml of Dox. Cell morphology was captured and compared under bright field on the first day of plating the cells and 24 and 48hrs later. Cells that were grown in the three different Dox concentrations showed normal stem cell morphology 48hrs after plating (figure 5.9a-c). The three different culture conditions showed no differences in cell at day0 and 48hrs with normal stem cell morphology observed. In order to check if Dox had any effect on the pluripotency of the cells, immunofluorescence staining was performed for Oct4 on cells that had been cultured in 1μg/ml of Dox. These cells were found to be pluripotent through their almost universal expression of Oct4 (figure 5.9d). After Dox titration, 1μg/ml was selected as the optimum concentration of Dox to use during knockdown experiments because, at this concentration, cells maintained good morphology and pluripotency. After the initiation of knockdown by Dox addition, knocked down cells were stained for YY1AP1 48hrs post knock down. While these cells maintained GFP expression, there was a decline in YY1AP1 expression in the cells 48hrs after knockdown (figure 5.10a-b). The expression of GFP and loss of YY1AP1 48hrs post knock down signalled that YY1AP1 shRNA transfected cells had lost YY1AP1 but maintained GFP expression because YY1AP1 is expressed from a different promoter to that of GFP. YY1AP1 and YY1 were greatly down regulated after YY1AP1 knockdown and substantially lost at day 5 with YY1 barely present (figure 5.10 c-d).
Figure 5.9. Dox titration of hUES1 cells. Cells were cultured as in sections 2.2.1-2.2.2 but in the presence of three varying concentrations (0.1, 0.5 and 1μg/ml) of Dox and microscopy (see section 2.8.3) was performed to capture their images after 0, 24 and 48 hrs of expansion in these culture conditions. a-c) Bright field microscopy showed that culturing hUES1 cells in 0.1, 0.5 and 1μg/ml of Dox had no detrimental effect on the growth and morphology of the cells. d) Mouse anti-Oct4 immunofluorescence staining in hUES1 cells cultured in 1μg/ml of Dox (see section 2.8.3) showed that the cells maintained pluripotency after 48 hrs due to their adequate expression of the pluripotency transcription factor Oct4. Scale bars=200μm.
Figure 5.10. Goat anti-YY1AP1 immunofluorescence staining and transcript levels of YY1AP1 and YY1 in GFP positive YY1AP1 knocked down hUES1s. YY1AP1 was knocked down in hUES1 cells (see section 2.10.4). Immunostaining (see section 2.8) was performed on YY1AP1 knocked down cells and images were captured 0 and 48hrs after knockdown induction. q-RT-PCR (see 2.7) was also performed on YY1AP1 knocked down cells and YY1AP1 and YY1 transcript expression levels were monitored 0hrs, 24hrs, 48hrs and 5 days post knockdown induction. a) Fluorescence microscopy and goat anti-YY1AP1 staining showed that hUES1 cells expressed both GFP and YY1AP1 at day0 of Dox induced knock down. b) hUES1 cells lost YY1AP1 expression 48hrs after the induction of its knockdown. c-d) YY1AP1 and YY1 were both down regulated after 5 days of YY1AP1 Dox-induced knockdown in the cells. Scale bars=100μm. Shown values are averages of three technical repeats.
During YY1AP1 knock down experiments, control experiments were carried out in parallel to the knockdown experiments. The negative control experiment involved the transfection of Beta-2-microglobulin (β2M) shRNA into hUES1 cells. The β2M shRNA was nucleofected using the same plasmid (pLVTHM) used for the YY1AP1 shRNA delivery. β2M shRNA expression was also initiated by the addition of 1µg/ml of Dox. YY1AP1 knock down in hUES1 cells did not have any significant effect on either the level of GFP expression or the morphology of the cells after 5 days of knock down initiation (figure 5.11a). Cells maintained steady GFP expression at day 5 and still displayed normal stem cell morphology at this time point. In another control experiment, an empty pLVTHM plasmid vector (without YY1AP1 shRNA) was nucleofected into the hUES1 cells. Transfection of the empty vector had no effect on stem cell morphology after 5 days (figure 5.11b). However, there was a steady increase in the level of GFP expression of the cells from day 0 to day 5. At day 5, there was an estimated 90% GFP positive cell colonies. The third control during the YY1AP1 knock down experiments in hUES1 cells involved the culturing of non-transfected hUES1 cells under normal feeder free stem cell conditions (see 2.2.2). These untreated cells displayed normal cell morphology and maintained confluency over a 5 day period (figure 5.12a). Cells were also cultured under normal feeder free conditions in the presence of Dox (1mg/ml) over a 5 day period to check if the presence of Dox in stem cell media had any effect on their morphology and proliferation. No observable effect on cell morphology and proliferation was seen (figure 5.12b). Gene expression analysis was performed by q-RT-PCR for YY1AP1 and YY1 in each of the samples. There were no obvious pattern indicating differences in transcript expression of the two genes in hUES1 cells transfected with β2M (figure 5.13a). In cells that had been transfected with an empty vector, there was an unusual up regulation of YY1AP1 after 48 hours and an unusually high level of YY1 in pluripotent cells at day 0. However this may be within the normal variation seen between stem cell populations. Apart from those differences, gene expression was consistent (figure 5.13b). When grown in the presence of DOX, there were no extreme differences in both genes’ transcript levels (figure 5.13c). It is worth noting though there was a dip in YY1 transcript level at 24 and 48 hours. In the last control experiment, cells were grown in normal stem cell
feeder conditions. Here, there were slight fluctuations in transcript level expressions (figure 5.13d).

Figure 5.11. GFP expression in control hUES1 cells transfected with β2M shRNA and empty vector during YY1API knock down. After two separate transfections (see section 2.10.3) of cells with β2M shRNA carrying plasmid and a non-transfected empty vector, GFP protein expression was observed (see 2.8.1) in both sets of transfected cells 0 days, 24hrs, 48hrs and 5 days after transfection. a) Cells transfected with β2M in the presence of Dox maintained GFP expression and stem cell morphology over 5 days b) Cells transfected with pLVTHM empty vector showed steadyd GFP expression and normal stem cell morphology 5 days after transfection. Scale bars=100μm
Figure 5.12. Morphology of untreated and Dox treated hUES1 cells during YY1API knock down. Untreated cells were cultured under normal feeder-free culture conditions (see sections 2.2.2.1-2.2.2) and Dox treated cells were cultured in the same way but with 1μg/ml of Dox continuously added to the cell culture medium. Bright field images of cells from both culture conditions were then captured (see 2.8.1) after 0 days, 24hrs, 48hrs and 5 days of cell culture. a) Untreated cells maintained normal stem cell morphology after 5 days in culture. b) Cells cultured in media containing 1μg/ml of Dox also maintained normal morphology after 5 days in culture with no visible detrimental effects of Dox on the cells. These cells maintained large nuclei. Scale bars=200μm
Figure 5.13. YY1AP1 and YY1 transcripts expression in control hUES1 cells during YY1AP1 knock down. q-RTPCR was performed for YY1AP1 and YY1 from cDNA harvested from cells cultured in four different control culture conditions (see section 2.7). Transcript expression levels were assessed after 0 days, 24hrs, 48hrs and 5 days of cell culture a) YY1AP1 and YY1 transcripts were steadily expressed in cells transfected with ß2M b) YY1AP1 and YY1 expression in cells transfected with an empty pLVTHM vector was also not variable c) YY1AP1 and YY1 transcripts in cells grown in the presence of Dox. d) YY1AP1 and YY1 transcripts expression in cells grown under normal feeder-free conditions. All results are representative of three experimental repeats.
Following the comparison of YY1AP1 and YY1 transcript levels in YY1AP1 knocked down cells, the expression of these two genes was assessed at a protein level in the knocked down cells. Proteins were harvested at the four time points of the YY1AP1 knock down experiment and western blotting (see 2.12) was performed for both YY1AP1 and YY1 using GAPDH as an internal protein standard. Just like at the transcript level, both YY1AP1 and YY1 were down regulated after the induction of YY1AP1 knockdown (figure 5.14). YY1AP1 seems to disappear after 48 hrs while both proteins appear to be almost completely ablated at day 5. The GAPDH protein remained highly expressed throughout the knockdown experiment (Figure 5.14).

![Figure 5.14](image.png)

Figure 5.14. YY1API and YY1 protein expressions in YY1API knocked down hUESI cells. Western blots were performed (see section 2.12) from proteins harvested in hUESI cells after YY1API had been knocked down. This was to compare how the ablation of YY1API affected the expression of both YY1API and YY1 proteins 0 days, 24hrs, 48hrs and 5 days after knockdown induction. Both YY1API and YY1 proteins were down regulated during knock down while GAPDH remained consistently expressed throughout the 5 day knockdown period. Numbers represent the densitometry of each gene protein normalised to GAPDH.
5.4 Discussion

There are different methods of shRNA transfer into cells. These include lipofection, viral transduction and electroporation. Nucleofection was selected to deliver plasmids carrying shRNAs into the cells because it is fast and fairly easy to perform. Although viral transduction lead to high transfection efficiencies, its safety and the time it takes to perform are major concerns. Among the non-viral transfection methods, nucleofection leads to the highest transfection efficiency, rapid cell growth and minimum damage to the cells during transfection because it is extremely quick (Cao et al., 2010). Therefore, nucleofection is an ideal, simple and safe method of gene transfer that has been used successfully many times in our lab.

In this chapter, it was shown that shRNA induced knockdown of YY1AP1 in MCF7 cells did not only lead to its down regulation but to the down regulation of its binding and target gene YY1. After 5 days of knockdown, both YY1AP1 and YY1 had been reduced more than 5 fold compared to their starting amounts at day 0. YY1AP1 is a direct binding partner of YY1 and has YY1 binding sites located on its promoter (Wang et al., 2001). Because YY1AP1 and YY1 are direct binding partners, cells might down regulate YY1 by reducing its regulator YY1AP1. Also, since YY1AP1 is a transcriptional regulator of YY1, its silencing and subsequent down regulation in the MCF7s meant that there were fewer YY1AP1 molecules to regulate the transcription of YY1; hence the down regulation of YY1.

The use of a plasmid vector with a GFP promoter is time efficient as it eliminates the need of a selectable marker to assess transfection efficiency. GFP expression in the pLVTHM plasmid vector was driven by an elongation factor-1 alpha (EF1-alpha promoter). Unlike cytomegalovirus (CMV), a viral GFP expression promoter, EF1-alpha is not prone to DNA methylation induced silencing and has the potential to lead up to nearly 100% GFP positive cells after transfection (Teschendorf et al., 2002). The activity of the EF1-alpha promoter can be maintained during the culture of pluripotent stem cells over a prolonged period of time while the CMV promoter has been shown to get down regulated in hESCs over approximately 50 days or so (Norrman et al., 2010). The use of this promoter was therefore greatly beneficial to the generation of a stable knock down hUES1 cell line. However, despite the efficiency and
stability of the EF1-alpha promoter, its use is not short of challenges. One such challenge was that during the YY1AP1 knock down experiments, cells were prone to gradual loss of GFP after the first transfection. Multiple nucleofections of cells were therefore undertaken to solve this problem. Re-transfecting hUES1 cells at least two times led to almost 100% GFP expressing cells. Moreover, the cells were able to maintain high stable GFP expression for more than 5 days. Of course, the ability to culture transfected cells without any effect to their proliferation and pluripotency was greatly boosted by the fact that the knockdown system used was inducible.

During YY1AP1 knock down experiments, 293FT cells and MCF7 cells were used primarily to test the transfection ability of YY1AP1 shRNA carrying plasmid vectors. These cells are commonly used because of their ability to grow rapidly in culture and the fact that they are easy to transfect. There was significantly high transfection efficiency observed in 293FT and MCF7 cells than in hUES1 cells. Twenty-four hrs after transfection, a good population of GFP positive cells (approximately 25%) was visible in transfected 293FTs and MCF7s with transfection efficiency in the MCF7s increasing drastically after 3 days. In contrast there was very low transfection efficiency observed in hUES1 cells. Interestingly, the proportion of GFP positive colonies 24hrs after transfection in 293FTs and MCF7s appeared to be the same as the proportion of GFP positive population in hUES1 cells 48hrs after transfection. This was an indication of the serious challenge of carrying out transfection in hESCs.

Transfection efficiency is partly determined by the size of the vector carrying the shRNAs. Large vectors are difficult to transfect and lead to poor transfection output. YY1AP1 and β2M shRNAs were driven by the pLVTHM plasmid vector which is 11 085 bp in size. Due to the large size of this vector, coupled with the high resistance of hESCs to transfection, it was hardly surprising that the transfection efficiency in hESCs was below par compared to the pmaxGFP vector, which was used as a positive control measure of transfection efficiency. Measuring only 2 900 kb, the pmaxGFP vector is very small and therefore can be easily driven across the cell membranes and into the nucleus.

One of the shortcomings of the YY1AP1 knockdown experiments was that no transfection efficiency assays were undertaken during nucleofection of the cells. This could have been done by counting cell populations of GFP positive cells in transfected
hESC cells and giving that as a percentage of the whole cell populations. This would have provided a precise measure of the efficiency of nucleofection. Another possible improvement to the transfection of hUES1 cells would have been to perform YY1AP1 immunofluorescence staining on the cells after their transfection with YY1AP1shRNA, β2MshRNA and the pLVTHM non-ligated empty vector. This would have allowed the identification of effects, if any, on the YY1AP1 protein. Flow cytometry could have also been performed to give quantitative data on these transfected cells. Making sure that YY1AP1 shRNA transfection and Doxycycline do not have detrimental effects on the expression of the YY1AP1 is important in understanding changes that occur after knock down is activated. This is also crucial in making sure that any changes observed after knock down are not necessarily due to the effects of nucleofection on the cells.

Sorting of transfected hUES1 cells was important in bulking up and selecting only GFP positive cells. This led to a pool of a potentially pure population of cells carrying the YY1AP1 and β2M shRNAs. The down side of sorting of the cells was that the cells took a longer time to recover after sorting. Cells were cultured in stem cell media supplemented with ROCK inhibitor, which helps to improve the survival of cells upon their dissociation. On average, it took 2 - 3 x 10^5 sorted cells at-least a week to expand and fill up a single well of a 6-well plate. However, it is also worth noting that these GFP positive colonies were themselves isolated from a small population of live cells during sorting. Nonetheless, the fact that the pLVTHM plasmid vector that was used during the transfection experiment did not carry a promoter for any selectable marker meant that sorting was the only option for selecting GFP enriched cell populations. This worked very satisfactorily as was evidenced by the homogenous populations of GFP positive cells generated. The inducible knock down system was found to be quite sufficient in generating ready-for knock down cells that were used in the ultimate Dox-induced knock down of YY1AP1 and the internal control gene β2M.

Analysis of gene expression showed that the knockdown of YY1AP1 in hUES1 cells led to a 15-fold and 10 fold down regulation of YY1AP1 and YY1 transcripts respectively; from day 0 to day 5. Both genes were also down regulated at the protein level from day 0 to day 5 of YY1AP1 knockdown. Therefore, there was strong correlation in the down regulation patterns of these two genes’ transcripts and proteins. YY1AP1 and YY1 are binding partners. In fact, it has been shown that YY1AP1
contains two YY1 binding regions (Wang et al., 2004). Because YY1AP1 regulates the transcriptional activities of YY1, both during gene expression repression and gene expression activation, its knockdown and down regulation means that there is hindrance in the activation of YY1 transcription due to the reduction in the number of YY1AP1 transcripts targeting YY1. This then lowers the YY1 transcript level. This is why there was a similar trend of down regulation for these two binding partners when YY1 is knocked down. However, the down regulation of YY1 came slightly later than that of YY1AP1.

In conclusion, it was shown in this chapter that FACs greatly enriches GFP positive cell population, increasing the number of cells carrying the YY1AP1 shRNA. Re-nucleofection of cells after FACs was shown to massively enhance transfection efficiency. The efficiency of the designed shRNA in knocking down YY1AP1 was also confirmed in MCF-7 cells. It was also shown that the YY1AP1 shRNA could be successfully transfected into both hESCs and iPSCs. The optimal concentration of Dox needed for knock down induction was determined to be 1μg/ml. When knock down was initiated in both MCF-7 and hUES1 cells, there was a down regulation of not only the YY1AP1 transcript but also that of its binding partner YY1. This was an indication that YY1AP1 regulates YY1. Both YY1AP1 and YY1 proteins were down regulated after YY1AP1 knock down in hUES1 cells showing that the loss of YY1AP1 at the protein levels induces the loss of YY1 protein too.

In the next chapter, 6, the expression patterns of YY1AP1 and YY1 transcripts and proteins will be compared with the expression patterns of other genes including endodermal markers, mesodermal markers, ectodermal markers and cell cycle regulatory genes. As has been earlier stated, the hypothesis of this study was to identify novel stem cell pluripotency-associated genes. The purpose of comparing YY1AP1 and YY1 with pluripotency and lineage markers will therefore help to find out how their expression patterns compare to that of other pluripotency genes and transcription factors. Protein interactions of YY1AP1 and YY1 with other gene proteins will also be presented in the next chapter.
CHAPTER 6

CHARACTERIZATION OF YY1AP1 KNOCKED DOWN CELLS

Following knockdown experiments, YY1AP1 knocked down cells were characterized in order to monitor gene expressions of a variety of pluripotency and differentiation markers. Characterization of the knocked down cells was first performed by nuclear immunofluorescence staining for YY1AP1. The expression patterns of transcripts of YY1AP1, YY1, pluripotency genes, endodermal, mesodermal, ectodermal and cell cycle markers were then determined by q-RTPCR in the knocked down cells over a 5 day period. The expression patterns of all these gene transcripts were also determined in control cells. These were; β2M knocked down cells, cells transfected with an empty vector, cells treated with Dox or cells grown under normal stem cell culture conditions. Protein expression patterns of YY1AP1, YY1, Oct4, Nanog and p53 during YY1AP1 knock down in hUES1 cells were then determined by western blotting. Immunoprecipitation was undertaken for YY1AP1 and YY1 in hUES1 cells to determine gene protein interactions. YY1AP1 knocked down and control cells were then differentiated over 10 days into EBs and gene transcript expression levels were determined for pluripotency and germ layer markers by qRT-PCR. It has been established in previous chapters that YY1AP1 silencing does induce loss of pluripotency in stem cells. It has been previously documented that YY1, a YY1AP1 binding partner and co-factor gene, promotes mesodermal differentiation of pluripotent stem cells for the formation of cardiac progenitor cells which forms cardiomyocytes (Gregoire et al., 2013). It has also been confirmed that YY1 is a direct target of Sox2 and binds to Oct4 through Ctcf to form a multifactor complex (Donohoe et al., 2009). However, the binding of YY1AP1 with any of the core pluripotency genes; Oct4, Nanog and Sox2, has not been confirmed. YY1AP1 plays regulatory functions in human liver cells during the inhibition of cell proliferation induced by cell cycle arrest (Li et al., 2007). Whether such a role extends to hESCs is yet to be established. In vivo, effects of YY1 knock down in mice are well documented. YY1 deficient mice undergo embryonic lethality caused by post-implantation deterioration of the embryos (Donohoe et al., 1999). Therefore, YY1 is essential during normal embryo development. In fact, it has been shown that YY1 is
required in a dosage-dependent manner during embryo development in mice (Affar et al., 2006) In this chapter, the differentiation of YY1AP1 knocked down cells was induced through the formation of embryoid bodies (EBs) and gene expression was quantitatively evaluated for pluripotency, endodermal, mesodermal, ectodermal and cell cycle markers. It was found that after knockdown, YY1AP1 expression pattern correlated with that of all pluripotency markers investigated except for Sox2. YY1AP1 was also down regulated similarly to YY1. The expressions of germ layer and cell cycle markers were inconsistent, apart from NKx2.5, Sox2 and CCND1.

6.1 Analysis of gene expression in YY1AP1 knocked down cells

6.1.1 Evaluation of cell cycle, apoptotic and differentiation marker expression in knocked down cells

Knockdown experiments are important in analysing what effects one or more genes have on the expression of other genes within a cell population. In chapter 4, it was revealed that the differentiation of hESCs through growth factor withdrawal and embryoid body formation led to the down regulation of YY1AP1. This expression pattern is synonymous with pluripotency gene expression. However, research is very scarce on the expression, regulation and potential function of YY1AP1 in human pluripotent stem cells. The effect of knocking down YY1AP1 was investigated in hUES1 stem cells by characterizing YY1AP1 knocked down cells. Knock down was confirmed by immunofluorescence staining of the YY1AP1 deficient cells 5 days post knock down induction. YY1AP1 appeared to be almost totally lost at this time point as very few cells stained positive for it (figure 6.1). Although there were variations in the expression patterns of the different germ layer and cell cycle markers, most of them were up regulated at the end of knockdown (figure 6.2-6.3). Generally, YY1AP1 knockdown led to the down regulation of itself, YY1 and pluripotency genes. Sox2 was however found to increase before being completely ablated at the end of knockdown (figure 6.2-6.3) With the exception of Sox2, target and pluripotency genes were all down regulated in a similar fashion of rapid down regulation (figure 6.3). The expression of endodermal markers Mixl1 and GATA4 remained fairly from the 24hr time point until the end of knockdown (figure 6.3). There was an up regulation of some of the mesodermal
markers like Nkx2.5 and Nkx3.2 after the loss of YY1AP1 in the cells. Interestingly, Col1a1 showed a similar expression trend to Pax3 of massive up regulation at 24hrs before declining back to consistent low levels of expression (figure 6.3). A similar expression trend was observed for TBX3. There was an up regulation of ectodermal markers (Pax6 and Sox1) and cell cycle regulators (c-Myc and p53) while CCND1 appeared to undergo delayed exponential down regulation (figure 6.3). The expression of p53 was very variable. The variability is most likely a result of different status of the cells in culture. In figure 6.2 p53 is going up and cyclin D going down. This means that the cells are withdrawing from cell cycle regulation. p53 is the major regulator of genomic stress and the most likely explanation is that the culture environment is resulting in a loss of ability to proliferate effectively and that this is correlated with some genomic stress (probably redox stress and/or imbalance in replication components) which activates p53 and inhibits proliferation.

Figure. 6.1. YY1API immunostaining in YY1API knocked down hUESI cells. To confirm knockdown, cells were immunostained (see 2.8) for YY1API 0 and 5 days after knockdown. Staining with goat anti-YY1API and goat IgG 2° Ab showed strong nuclear localisation of YY1API at day 0 and very poor staining after 5 days of knock, an indication of loss of YY1API.

Figure. 6.2. qRT-PCR for analysis of gene transcript expression in YY1API knocked down hUESI cells. q-RT-PCR was performed (see 2.7) and this graph shows transcript expression levels of members of all the different gene groups 0 days, 24hrs, 48hrs and 5 days after knockdown initiation. Results are representative of averages of ΔΔCt values from three experimental repeats. Errors bars represent standard deviation. Because GAPDH transcript expression level was used as an internal standard reference, gene expression could not be normalised to 1.
Figure 6.3. qRT-PCR for endodermal, mesodermal and ectodermal markers in YY1API knocked down hUES1 cells. q-RT-PCR was performed for germ layer marker transcript expression levels as described in section 2.7. This was performed on cDNA harvested after 0 days, 24hrs, 48hrs and 5 days of knockdown induction by provision of Doxycycline (see sections 2.10.3-2.10.4). Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In a parallel experiment to the YY1AP1 Experiment, β2M was also knocked down in hUES1 cells. The experiment was conducted under the same laboratory conditions and practices. The results showed that taken together, target genes and pluripotency gene transcripts were visibly more expressed than all the differentiation markers (figure 6.4). More importantly, gene expression of all the genes remained fairly but not strictly consistent. Unexpectedly, there was down regulation of Nanog and Sox2 24hrs after knock down initiation (figure 6.5). Apart from Col1a1 and CCND1, differentiation and cell cycle markers were very low in expression in the knockdown cells which suggests that any differences in their expression levels were insignificant (figure 6.5).

Figure 6.4.qRT-PCR for analysis of gene transcript expression in β2M knocked down hUES1 cells. q-RT-PCR (section 2.7) was performed on cDNA harvested from β2M ablated cells to assess how this would affect the overall gene transcript expression levels of pluripotency, endodermal, mesodermal, ectodermal and cell cycle genes. Gene expression was observed after 0 days, 24hrs, 48hrs and 5 days of β2M knock down initiation by provision of Doxycycline (see sections 2.10.3-2.10.4). Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.5. qRT-PCR for endodermal, mesodermal and ectodermal markers in β2M knocked down hUES1 cells. qRT-PCR, performed as described in section 2.7, was used to show the expression of YY1AP1, YY1, pluripotency genes and differentiation marker gene transcripts after 0 days, 24hrs, 48hrs and 5 days of β2M knock down induction by the addition of Doxycycline as described previously. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In addition to β2M knockdown, there were other control experiments that were run in parallel. For one control group, hUES1 cells were transfected with an empty pLVTHM plasmid vector. Therefore, cells were transfected with a vector that had not been ligated with any shRNA. Cells transfected with an empty vector had high expression of YY1AP1, YY1, Oct4 and Nanog throughout the 5 day knockdown period though the levels were highly variable. All the different germ layer markers were expressed at very low levels except for CCND1 and NKx3.2 (figure 6.6). When looked at individually, there were no pronounced differences in the expression levels of YY1AP1, YY1, Sox2, Nanog and Sox2 at the Day0, 24hrs, 48hrs and Day5 time points (figure 6.7). The expression levels of most of the differentiation markers were very low when compared to the target and pluripotency genes.

![Figure 6.6](image)

Figure 6.6. qRT-PCR for analysis of gene transcript expression in hUES1 cells transfected with a control pLVTHM empty vector. Grouped gene transcript expression levels of different markers observed at Day0, 24hrs, 48hrs and 5days after nucleofection of cells with an empty control plasmid vector that had not been ligated with a YY1AP1 shRNA. q-RT-PCR had been performed as described in section 2.7 after cells had been cultured as in 2.2.1 and 2.2.2. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.7. qRT-PCR for endodermal, mesodermal and ectodermal markers in knocked hUES1 transfected with a control pLVTHM empty vector. Individual gene transcript expression levels of different markers observed at Day0, 24hrs, 48hrs and 5days after nucleofection of cells with an empty control plasmid vector that had not been ligated with a YY1AP1 shRNA. q-RT-PCR had been performed as described in section 2.7 after cells had been cultured as in 2.2.1 and 2.2.2. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In another control experiment, cells were cultured under normal cell culture conditions. However, the stem cell media was continuously treated with 1µg/ml of doxycycline (Dox) over a five day period. As was the case with gene expression in the other controls, cells grown in the presence of Dox displayed high expression of target and pluripotency markers while germ layer and cell cycle genes were expressed at very low levels with the exception of CCND1, NKx3.2 and TGFB1 (figure 6.8). In fact, relative to target and pluripotency marker expressions, the expressions of most of the germ layer markers were negligible. A gene by gene presentation of the expression levels of the genes showed very distinct similarities in expression levels of target and pluripotency genes (figure 6.9). Almost all germ layer and cell cycle genes were expressed at very low levels throughout the 5 day cell culture period.

Figure 6.8. qRT-PCR for analysis of gene transcript expression in Doxycycline-treated hUES1 cells. Grouped expression levels of different gene transcripts after cDNA was harvested from hUES1 cells grown under normal culture conditions (see 2.2.1-2.2.2) but in the presence of 1µg/ml of Doxycycline. The cDNA was harvested after 0 days, 24hrs, 48hrs and 5 days of cell culture to perform q-RT-PCR (see 2.7). Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.9. qRT-PCR for endodermal, mesodermal and ectodermal markers in hUES1 cells treated with Doxycycline. Individual expression levels of different gene transcripts after cDNA was harvested from hUES1 cells grown under normal culture conditions (see 2.2.1-2.2.2) but in the presence of 1µg/ml of Doxycycline. The cDNA was harvested after 0 days, 24hrs, 48hrs and 5 days of cell culture to perform q-RT-PCR (see 2.7). Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In the final control experiment, cells were cultured under normal cell conditions over 5 days. Cells cultured under these conditions maintained very consistent expression of all the genes (figure 6.10). In fact, there was more consistency in gene expression levels in this control than there was in any of the other controls (figure 6.11). However, as was observed with gene expression in the other controls, target and pluripotency genes were significantly more enriched than the all germ layer and cell cycle genes although CCND1 showed profound expression (figure 6.11).

Figure. 6.10. qRT-PCR for analysis of gene transcript expression in untreated hUES1 cells. hUES1 cells were cultured under standard laboratory conditions as has been previously described in sections 2.2.1-2.2.2. Grouped gene transcript expression levels for YY1API, YY1, pluripotency genes, germ layer markers and cell cycle genes were then observed by q-RT-PCR (see section 2.7). Results are representative of averages of \( \Delta \Delta \text{Ct} \) values from three experimental repeats. Error bars represent standard deviation.
Figure 6.11. qRT-PCR for endodermal, mesodermal and ectodermal markers in untreated hUESI control cells. hUESI cells were cultured under standard laboratory conditions as has been previously described in sections 2.2.1-2.2.2. Individual gene transcript expression levels for YY1AP1, YY1, pluripotency genes, germ layer markers and cell cycle genes were then observed by q-RT-PCR (see section 2.7). Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
6.1.2 Protein level expression patterns in YY1AP1 knocked down and normal hUES1 cells

Because mRNA is translated into proteins, it is a commonly expected that there should be some correlation between mRNA and protein levels within a cellular population. However, that is not always the case. In order to determine if there was any correlation between mRNA and protein expression in YY1AP1 knocked down cells, western blotting was performed on proteins harvested at Day0, 24hrs, 48hrs and 5 days after knock down initiation. But first, YY1AP1 and YY1 protein expressions were confirmed in MCF-7 cells using GAPDH protein expression as an internal standard protein control. Both proteins were expressed in the MCF-7s (figure 6.12). Western blotting on YY1AP1 knocked down cells showed that GAPDH protein was consistently expressed at days 0, 24hrs, 48hrs and day5 of YY1AP1 knockdown. YY1AP1 and YY1 proteins were down regulated over the 5 day period (figure 6.13). Densitometry readings showed that proteins of these two genes were down regulated in a similar fashion. By 48hrs YY1AP1 had been drastically reduced and by day 5 both YY1AP1 and YY1 proteins appeared to have been completely ablated. There was also a down regulation of pluripotency proteins Nanog and Oct4 and an up regulation of the tumour suppressor gene p53 (figure 6.13). However, the loss of Oct4 was more rapid than that of Nanog, with the Oct4 protein appearing completely lost by day5. The down regulation of YY1AP1, YY1, Oct4, Nanog and the up regulation p53 mirrored the expression of mRNAs of these genes in YY1AP1 knocked down cells. It therefore appears as if the knock down of YY1AP1 in hUES1 cells led to correlated changes in mRNA and protein expressions of the same genes.

Figure. 6.12. YY1API and YY1 proteins are expressed in MCF-7s. As a test of principle, western blotting (see 2.12) was performed to confirm the expression of YY1API and YY1 proteins in MCFs that had been cultured as described in section 2.4. Both genes’ expression was normalised to internal GAPDH protein expression, which was clearly more expressed than them.
Figure 6.13. Western blotting for protein expression during YY1API knockdown in hUES1 cells. As described in section 2.12, proteins were harvested in hUES1 cells 0 days, 24hrs, 48hrs and 5 days after YY1API knockdown for western blotting to monitor gene expression variations of YY1API, YY1, Nanog, p53 and Oct4 at the said time points during knock down. Densitometry protein expression values were normalised to GAPDH expression values at each time point. Densitometry results are representative of averages from two different western blotting experiments.
Densitometry readings from two western blotting repeats for the different proteins were averaged together and normalised against the average GAPDH readings. Results showed the steady down regulation of YY1AP1, YY1, Oct4 and Nanog proteins and an up regulation of the p53 protein (figure 6.14). p53 was upregulated because YY1AP1 knock down induced cell differentiation which ultimately resulted in apoptosis. p53 is a marker of apoptosis and at day 5 of the KD, differentiation and apoptosis were occurring hence the accumulation of the p53 protein and down regulation of pluripotency gene proteins. As was the case at the mRNA level, the Oct4 protein was the most rapidly down regulated among the pluripotency genes.

![Figure 6.14. Protein expression during YY1AP1 knockdown in hUES1 cells. A plot of densitometry reading averages for different genes from western blotting as described in section 2.12. Gene proteins were harvested from hUES1 cells 0 days, 24hrs, 48hrs and 5 days after knock down initiation. Protein expression at each time point was normalised to GAPDH expression at that time point.](image-url)
6.1.3 Gene protein interaction analysis by immunoprecipitation

Protein-protein interactions are important components of the hESC maintenance networks. Immunoprecipitation and co-immunoprecipitation experiments enable researchers to evaluate these protein interactions. In order to establish the interaction of YY1AP1 and YY1 proteins, as well as the interaction of these two genes with OCT4, Nanog, CTCF and p53, immunoprecipitation (IP) experiments were undertaken and protein interactions were established by western blotting. Negative controls were western blottings performed without a protein lysate and for the secondary IgG antibody only without the primary antibody. Performance of western blotting on non-precipitated protein lysate (input) was used as a positive control. In brief, IP experiments involved the harvesting of proteins from hUES1 cells; the precipitation of the gene of interest (i.e. either YY1AP1 or YY1) from the crude protein mixture and the running of the precipitated target protein lysate on a western blot to check for its interaction with other specific gene proteins. This interaction is confirmed by the enrichment of these specific proteins in the precipitated lysate of the target protein. Co-IP involved checking for the enrichment of the YY1AP1 protein in precipitated YY1 protein lysate and vice-versa. As proof of principle, IP was first undertaken in 293FT cells. YY1AP1 and YY1 proteins were found to be enriched in precipitated YY1 protein lysate in 293FTs (figure 6.15a). No protein expression was observed for the two genes in both negative control western blottings while both proteins were enriched in the input control positive control lysate. Relative to the YY1AP1 protein, there appeared to be more enrichment of YY1 which was less surprising since the western blotting was performed on YY1 enriched lysate. Performance of reverse co-IP on the 293FTs by the precipitation of YY1AP1 showed enrichment of both YY1AP1 and YY1 proteins in YY1AP1 lysate. They were not expressed in the negative controls but in the positive input control (figure 6.15b). However, surprisingly, it appeared as if the YY1 protein was equally, if not more, enriched in the YY1AP1 precipitated lysate as was the YY1AP1 protein itself. This could be due to the ubiquitous nature of YY1. In hESCs, YY1AP1 and YY1 proteins were also shown to be enriched in immunoprecipitated YY1 lysate (figure 6.15c). However, the interaction of Oct4 and Nanog proteins with the YY1 protein could not be established. Interestingly, it was found that the CTCF protein
was enriched in YY1 precipitated lysate. CTCF is a zinc finger transcription factor that is involved in chromatin regulation through insulation (Xu et al., 2007). The interaction of YY1 with CTCF might owe to the fact that in *in vivo* ESCs, YY1 has been shown to form a complex with CTCF to bind to OCT4 and regulate ESCs (Donohoe et al., 2009). Cross co-IP showed that YY1 and YY1AP1 are both enriched in YY1AP1 lysate with no enrichment of OCT4 and NANOG observed (figure 6.15d). CTCF was also enriched in the YY1AP1 precipitated lysate. Taken together, these findings signal that YY1AP1 and YY1 are heavily linked together at the protein level. Their post-transcriptional interaction with pluripotency transcription factors Oct4 and Nanog could not be established. The enrichment of a well-known Oct4 direct interacting gene, CTCF, on both YY1AP1 and YY1 enriched lysates might be a signal that YY1AP1 and YY1 can only interact with pluripotency transcription factors in complex with other transcriptional co-factors.

![Figure 6.15](image-url)

**Figure. 6.15.** Western blotting on immunoprecipitated YY1API and YY1 proteins in 293-FT and hUES1 cells. YY1API and YY1 were co-immunoprecipitated from proteins harvested in 293-FT and hUES1 cells (see section 2.13) that had been cultured under standard laboratory conditions (see 2.2.1-2.2.2 and 2.4). Western blot experiments (see section 2.12) were performed to see which genes are enriched in precipitated lysates of YY1API and YY1 in hUES1 cells a) Protein expression in YY1 precipitated lysate in 293-FTs b) The expression of proteins in YY1API precipitated lysate in 293-FTs c) Expression of proteins in YY1 precipitated lysate in hUES1 cells d) Protein expression in YY1API precipitated lysate in hUES1 cells.
6.1.4 Differentiation of YY1AP1 knocked down cells and analysis of gene expression after differentiation

It has already been shown in this study that YY1AP1 knock down results in loss of pluripotency of cells and initiation of differentiation. This was shown by the down regulation of pluripotency transcription factors and the up regulation of differentiation markers. To check whether the differentiating YY1AP1 knocked down cells could differentiate furthermore, they were differentiated into embryoid bodies (EBs). In brief, this differentiation protocol involved the culturing of cells in MEF media over a 10 day period. cDNA samples for real-time PCR analyses were then harvested at days 0, 5 and 10 of the protocol. At the same time, the EB differentiation protocol was also performed for experimental control cell samples. These controls were; β2M knocked down cells; cells transfected with an empty vector; cells that had been cultured in 1µg/ml of Dox and untreated cells that had been cultured under normal feeder-free stem cell culture conditions. Figure 6.16a shows day0, day5 and day 10 EBs formed from YY1AP1 knocked down cells. These EBs are GFP positive because the plasmid vector carrying the YY1AP1 shRNA had an EF1 alpha promoter which drives GFP expression. Figure 6.16b shows EBs formed from cells that had been continuously treated with Dox but otherwise cultured in normal stem cell conditions. These EBs do not express GFP since the initial cells were not transfected with the plasmid vector. Figure 6.16c shows EBS formed from β2M knocked down cells that were also transfected with a β2M shRNA-carrying plasmid with a GFP promoter. EBs from untreated and untransfected cells cultured in normal stem cell conditions are shown in figure 6.16d. A day after initiation of EB formation in all cell samples (Day0), cells had already clumped up into differentially sized aggregates and by day 10, circular, almost translucent 3-D EB structures had been formed.
Figure 6.16. Knock down and control EBs formed during 10 days of differentiation of YY1API deficient hUES1 cells. Here GFP is used a reporter gene. Day 10 harvested from the hUES1 YY1API and knockdown protocol were differentiated into EBs as described in section 2.3.2 to test the differentiation capacity of the cells. These EBs were cultured for 10 days and their images (see 2.8.3) were captured at days 0, 5 and 10. Control EBs made from Dox treated cells and cells cultured normally did not express GFP because the cells were not transfected with the GFP promoter carrying plasmid. a) GFP positive EBs formed from YY1API knocked down cells b) Control EBs made from cells treated with Dox c) GFP positive control EBs from β2M knocked down cells d) Control EBs from untreated cells. All EBs showed normal morphology by day 10.
Real time PCR showed that the differentiation of YY1AP1 knocked down cells led to an all-round, almost complete, shut down of gene expression (figure 6.17). Results of gene expression in the control EBs are presented later in figures 6.20-6.27. Analyses of the individual gene expressions in EBs formed from YY1AP1 knocked down cells showed that all targets, pluripotency, endodermal and mesodermal gene transcripts were all severely ablated in the EBs after 10 days of EB growth. Gene expression down regulation was observed for all tested genes irrespective of their functional groups. Expression was however observed for Brachyury, Col1a1, TBX3 and CCND1 at day 10. (figure 6.18). Interestingly, a few number of genes (p53, TGFB1, c-Myc, NKx3.2 and NKx2.5) appear to have lost expression immediately after knockdown expression.

![Figure 6.17. qRT-PCR for gene transcript expression analysis in EBs from YY1AP1 knocked down hUES1 cells.](image)

Figure. 6.17. qRT-PCR for gene transcript expression analysis in EBs from YY1AP1 knocked down hUES1 cells. After YY1AP1 knock down (see 2.10.4), knocked down cells were tested for differentiation by culturing them in suspension into EBs (as in 2.3.2) for 10 days. cDNA was collected at days 0, 5 and 10 of differentiation to assess gene expression through q-RT-PCR (see section 2.7). The expression of all tested genes was almost shut down. For comparison, gene expression in control EBs can be found in figures 6.20 and 6.27 later on in this chapter. Due to time limitations, cell viability could not be performed during this protocol. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.18. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from YY1AP1 knocked down hUES1 cells. Individual gene transcript expression levels in EBs that were formed from day 10 hUES1 cells that had been knocked down of YY1AP1. qRT-PCR (see 2.7) was performed on cDNA harvested from EBs after 0 days, 5 days and 10 days of their formation. Generally, all genes were down regulated from day 0 to day 10. Gene expression in normal EBs is presented later in figures 6.26 and 6.27. Results are representative of averages of $\Delta \Delta Ct$ values from three experimental repeats. Error bars represent standard deviation.
In order to determine if the differentiation of YY1AP1 deficient cells induced cell differentiation bias towards any lineage, real time PCR was performed for some specific advanced differentiation markers and early mesodermal markers on cDNA collected during EB differentiation. Despite their up regulation at day 5, the expression of early mesodermal markers Sox7 and Sox17 did not appear to be regulated during the 10-day differentiation as evidenced by little disparities between their expression levels at days 0 and 10 (figure 6.19). Microtubule-associated protein-2 (MAP2) and Beta-III-Tubulin (BIIIT) are neuronal markers while Alpha-feto protein (AFP) is a hepatocellular marker gene. All these three markers were down regulated during EB formation in YY1AP1 compromised cells (figure 6.19).

Figure. 6.19. qRT-PCR showing transcript expression of differentiation markers in EBs made from YY1AP1 knocked down hUESI cells. cDNA harvested from the EBs at days 0, 5 and 10 was used during q-RT-PCR (see section 2.7) to check if the differentiation of the cells is biased towards a particular lineage. This was found untrue as no terminal marker up regulation was observed. Gene expression was normalised to GAPDH. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Gene expression was also analysed in EBs formed from the four control groups. Generally, there was a down regulation of target and pluripotency genes and a simultaneous up regulation of different germ layer markers and cell cycle regulatory genes. Collectively, β2M knocked down EBs showed consistent down regulation of YY1AP1, YY1, Oct4, Nanog and Sox2 while all differentiation markers, except Pax3 and CCND1 were enhanced (figure 6.20). Unsurprisingly, considering that it was the one knocked down in the differentiating cells, YY1AP1 was one of the quickest among the down regulated genes to go down (figure 6.21). However, surprisingly, YY1AP1 binding partner YY1 was the second slowest (after Sox2) among the down regulated genes to go down. This is probably because of the transcription regulatory role that YY1AP1 has on YY1.

![Graph showing expression trends](image)

Figure 6.20. qRT-PCR for gene transcript expression analysis in EBs from β2M knocked down hUES1 cells. β2M knocked down cells were differentiated into EBs for 10 days and cDNA harvested at days 0, 5 and 10 was used during q-RT-PCR (see section 2.7) to analyse transcript gene expression trends. Along with YY1AP1 and YY1, all pluripotency markers were down regulated while nearly all but one of the germ layer markers were up regulated. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.21. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from β2M knocked down hUES1 cells. Individual gene transcript expression levels after β2M knocked down cells were differentiated into EBs for 10 days. cDNA was harvested at days 0, 5 and 10 and used during q-RT-PCR (see section 2.7) to analyse transcript gene expression trends. Along with YY1AP1 and YY1, all pluripotency markers were down regulated while nearly all of the different germ layer markers were up regulated. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Gene expression analysis in EBs from cells that were transfected with a pLVTHM empty vector also showed that target and pluripotency genes were down regulated during differentiation while most of the different germ layer and cell cycle markers were up regulated (figure 6.22). All pluripotency markers appear completely ablated at day 10 (figure 6.23). Among the differentiation and cell cycle markers, Pax3, Gata4, CCND1 and TBX3 were surprisingly down regulated.

Figure. 6.22. qRT-PCR for gene transcript expression in EBs from hUESI cells transfected with a control pLVTHM empty vector. EBs were formed (see 2.3.2) from cells that had been transfected with an empty vector and cultured for 10 days under normal culture conditions. cDNA harvested at days 0, 5 and 10 was used to analyse different gene transcript expressions through q-RT-PCR (see section 2.7). Results are representative of averages of $\Delta\Delta$Ct values from three experimental repeats. Error bars represent standard deviation.
Figure 6.23. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from hUES1 cells transfected with a control pLVTHM empty vector. Individual gene expression from EBs that were formed (see 2.3.2) from cells that had been transfected with an empty vector and cultured for 10 days under normal culture conditions. cDNA harvested at days 0, 5 and 10 was used to analyse different gene transcript expressions through q-RT-PCR (see section 2.7). All pluripotency markers were down regulated and with the exception of CCND1, GATA4, TBX3 and Pax3, all differentiation markers were up regulated. This is indicative of cell differentiation. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In another control, the differentiation of cells that had been cultured in the presence of 1µg/ml of doxycycline led to the down regulation of both target genes and all the pluripotency genes and the up regulation of all differentiation and cell cycle genes (figure 6.24). There was an abnormally high up regulation of p53. P53 is an important regulator of cell proliferation in differentiated cells. Its absence in differentiating cells results in increased cell proliferation of the differentiating cells and reduced apoptosis. Therefore, its massive up regulation in the Dox treated EBs might have been to control the proliferative activities of the differentiating cells. p53 also maintain cell-cycle regulation by repairing damaged DNA or destroying unrepairable DNA by inducing cell cycle arrest. In the Dox treated control EBs, cells were cultured in the presence of Dox. It is likely that Dox may have had genotoxic effect on some of the EB cells prompting the up regulation of p53 to prevent damage to the cells. Interestingly, the most severe down regulation of the target and pluripotency genes and the most severe up regulation of differentiation and cell cycle markers was observed in this control EBs than in any of the other controls (figure 6.25). These observations might have been due to the effects of antibiotic activity of doxycycline on the cells. Clo1a1, TGFB1 and NKx3.2 were not enhanced while Sox2 was completely lost by day5.

Figure 6.24. qRT-PCR for gene transcript expression in EBs from hUES1 cells treated with 1µg/ml of Doxycycline. From cells cultured normally for 10 days but in continuous presence of Dox, EBs were formed as described in section 2.3.2. q-RT-PCR was performed on EBs from days 0, 5 and 10 of EB formation. There was massive up regulation of p53 which is needed for regulating the proliferation of differentiating cells and preventing DNA damage. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
Figure 6.25. qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from hUES1 cells treated with 1µg/ml of Doxycycline. Individual distinguished gene expression trends from cells cultured normally for 10 days but in continuous presence of Dox, EBs were formed from these cells as described in section 2.3.2. q-RT-PCR was performed on cDNA from the subsequent EBs at days 0, 5 and 10. There was massive up regulation of p53 which is needed for regulating the proliferation of differentiating cells and preventing DNA damage. Results are representative of averages of ΔΔCt values from three experimental repeats. Error bars represent standard deviation.
In the untreated control EBs, once again all target and pluripotency genes were up regulated and differentiation markers and cell cycle genes down regulated (figure 6.26). However, down regulation and up regulation of gene expression was much steadier in this control EBs than in all of the other controls (figure 6.27). CCND1, Pax3 and TGFB1 were the only differentiation markers not enhanced while Sox2 was again the most down regulated pluripotency marker. Taken together, these findings suggest the differentiation of YY1AP1 knock down control cells but not that of the knocked down cells themselves.

![Figure 6.26. qRT-PCR for gene transcript expression in EBs from control untreated hUES1.](image)

In these control EBs, cells were grown under normal cultured conditions (see 2.2.1-2.2.2). EBs (see section 2.3.2) were formed from these cells after 10 days of culture. cDNA samples were collected at days 0, 5 and 10 of the EB formation protocol and used for q-RT-PCR as described in section 2.7. YY1API, YY1 and all pluripotency marker transcript expression levels were down regulated and all but three differentiation marker transcripts were up regulated. This was a clear indication of loss of pluripotency and occurrence of cell differentiation. Results are representative of averages of delta Ct values from three experimental repeats. Error bars represent standard deviation.
Figure 6.27qRT-PCR for endodermal, mesodermal and ectodermal markers in EBs made from control untreated hUES1 cells. For individual gene expression in these control EBs, cells were grown under normal cultured conditions (see 2.2.1-2.2.2). EBs (see section 2.3.2) were formed from these cells after 10 days of culture. cDNA samples were collected at days 0, 5 and 10 of the EB formation protocol and used for qRT-PCR as described in section 2.7. YY1AP1, YY1 and all pluripotency marker transcript expression levels were down regulated and all but three differentiation marker transcripts were up regulated. This was a clear indication of loss of pluripotency and occurrence of cell differentiation. Results are representative of averages of delta Ct values from three experimental repeats. Error bars represent standard deviation.
6.2. ChIP-Seq analysis of gene interactions

Chromatin immunoprecipitation DNA sequencing (ChIP-Seq) experiments are important in the analysis and confirmation of gene or transcription factor interactions. In ESCs, Chip-Seq analysis has been used in studies of stem cell transcriptional regulatory circuitry (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2008; Marson et al., 2008; Kunarso et al., 2010). Here, ChIP-Seq experiments were undertaken to determine genes that directly interact with YY1. These interactions are determined by the presence of these genes’ binding sequences on the YY1 motif. In this way, genes and transcription factors that are involved in the regulation of stem cell pluripotency can be identified from the protein-DNA interactions. ChIP-Seq analysis of gene interactions generates a gene network of DNA associated proteins for any gene of interest to show how proteins and DNA interact to regulate gene expression. ChIP-Seq determines protein binding sites on the genome in an unbiased manner. On the other hand, chromatin network comprises of chromatin modifiers and histone modifications which form a signalling network that regulate the transcription and chromatin state of the genome. It has been shown that in stem cells, chromatin proteins are loosely connected to each other which allow the remodelling of chromatin for epigenetic gene regulation. Through western blotting, gene interactions in immunoprecipitated chromatin can be determined. However, this interaction of genes is not proof of regulation. Chip-Seq experiments can determine gene regulations by proving that genes bind directly to each other and share binding regions at their respective promoters. Prior to the performance of YY1 Chip-Seq on hESCs, ChIP-Seq experiments were undertaken on chromatin collected from 293FT cells. 293Fts are kidney cells derived from the human embryo. Since 293Ft cells are not pluripotent, they are expected to have a totally different gene expression network to hESCs. After the harvesting and sonication of chromatin, fragment size of the chromatin (see section 2.14) was established by gel electrophoresis (see 2.6.6). Different manuals and protocols recommend DNA fragments of between 100 and 1000bp as ideal for ChIP-Seq experiments. The closer to the 100bp size end of the range, the better. In 293FT, it was found that sonicated chromatin had sizes of between 100 and 400bp (figure 6.28). YY1 was then precipitated from the chromatin. In order to assess for the levels of both
YY1AP1 and YY1 in the precipitated YY1 lysate, real time PCR experiments were undertaken for these two genes. CollagenX, a known non-interacting gene of YY1, was used as a negative control. Gene expression in the precipitated DNA was compared to gene expression in mock cDNA collected from normal non-precipitated 293FT cells. Both YY1AP1 and YY1 were very highly expressed in YY1 precipitated DNA while CollagenX was not expressed at all (figure 6.29). Considering that it was YY1 not YY1AP1 that was precipitated, it was surprising to find that YY1AP1 was more enriched in the YY1 precipitated DNA than YY1 itself. In the mock DNA samples, no gene expression was observed.

Figure. 6.28. Agarose gel electrophoresis for fragment size determination of sheared chromatin. Sheared chromatin size was between 100 and 400 bp. This falls within the recommended range of 100 to 1000bp.

Figure. 6.29. qRT-PCR for gene transcript expression analysis in YY1 immunoprecipitated chromatin DNA and mock DNA from hUES1 cells. CollagenX was used as a non-interacting control. Mock DNA refers to normal non-enriched DNA collected from the cells. Gene expression was normalised to GAPDH.
In brief, after DNA sequencing of the YY1 enriched chromatin, reads were mapped with bowtie v2.2.3 and peak calling was done using MACS v2.1.0. Reads were mapped to the reference assembly human genome 19 (hg19) using the UCSC genome browser. Only regions with a fold enrichment of 5 or more were considered. In general, most of the genes that bound to the YY1 promoter were microRNAs, cadherins and zinc finger proteins. A lot of tumour suppressor proteins were also found on the YY1 promoter. Tumour suppressor genes were highly present in the mapped reads. Interestingly, a fairly high number of cell cycle regulatory genes were also identified in the gene list. Most of these cell cycle regulatory genes had high fold enrichment values, suggesting that they were steady binding partners of YY1. Among the cell cycle regulators found mapped on the YY1 promoter were GPC5, CYLD, E2F3 and CDC73. Although YY1 is well known for its role in cell proliferation, it can also, due to its dual functionality, repress cell growth through transcriptional repression. It is therefore not surprising that tumour suppressors were mainly found to interact with it. There were some proliferation genes identified as well which keeps in line with the function of YY1 as a proliferation and self-renewal regulatory gene in stem cells.

It is however very important to state that the analysis done on the Chip-Seq data was not conclusive. This is because the data came very late from the sequencing centre and analysis was not done to completion due to time limitations. However, data analysis continues to be done post thesis submission. I am currently running an input control Chip-Seq lysate so that gene expression in the sequenced YY1 enriched lysate can be compared to that in an un-enriched lysate for confirmation of gene expression. Also, during the Chip-Seq experiment, YY1 enrichment in 293FT cells was used as a control Chip-Seq but due to pressing time, comparison of gene expression between that control and hUES1 cells could not be completed in time before thesis completion. This is also currently being done post thesis submission. 293FTs are fibroblast cells and therefore not pluripotent. Therefore, gene expression observed in the YY1 enriched hUES1 lysate would be expected to be the opposite of gene expression in the YY1 enriched 293FT lysate.
6.3 Discussion

Results presented earlier in this chapter showed that the knockdown of YY1AP1 in pluripotent stem cells results in the differentiation of the cells as evidenced by the down regulation of pluripotency genes and the up regulation of mesodermal and ectodermal marker genes. Of course, this up regulation is not proof of the mesodermal and ectodermal differentiation of the cells. Even though more cell death was observed in knocked down than Dox control cells, the scale of cell death in the YY1AP1 knocked down cells was unknown. Dead cells can affect gene expression by inhibiting the true reflection of gene expression levels in live cells. Dead cells are sources of contamination too, something that can hinder proper gene expression in a population mix of dead and live cells. It is for these reasons that, had time allowed, detailed analysis of the apoptotic state of the cells would have been performed to determine the proportion of dead cells in the YY1AP1 compromised cells. This could have been done by flow cytometry to detect apoptotic markers in the YY1AP1 knocked down cells. Markers such as Annexin V are commonly used to evaluate and quantify apoptosis. Quantification of apoptotic cells in the population would have allowed compensation for those cells when analysing gene expression after knockdown.

The gradual ablation of YY1AP1 led to the up regulation of cell cycle regulatory genes c-Myc and p53. YY1AP1 has already been reported to form a complex with MAD2L2 and negatively regulate cell cycle progression by inducing arrest at the G0/G1 phase (Li et al., 2007). It was not surprising that its down regulation induced the transcriptional activation of cell cycle markers. In fact, the overexpression of YY1AP1 leads to the arrest of the G0/G1 phase of the cell cycle and the inhibition of cell proliferation. YY1, a YY1AP1 binding partner, has been shown to negatively regulate p53 expression (Gronroos et al., 2004; Sui et al., 2004) therefore promoting cell proliferation. The presence of YY1AP1 at low levels in the cells is likely to trigger the p53 regulated cell cycle transcription.

Because of its inability to influence gene expression and cell behaviour (Matin et al., 2004; Zafarana et al., 2009), Beta-2-microglobulin (β2M) shRNA is commonly used in knockdown experiments as a parallel standard control against which the effect of target gene(s) shRNA(s) can be compared. In this way, any gene
expression variations between experimental samples are compensated for. It is common practice during the use of β2M shRNA in knockdown experiments that at specific intervals during the knockdown experiments, levels of β2M transcript in the β2M knocked down cells are quantitatively assessed to determine the efficiency of the knockdown (Avery et al., 2010). Based on cell morphology and gene expression patterns, β2M knockdown appeared to have no apparent effect on transcript levels within the cells. It would have been beneficial to this study if β2M transcript levels had been observed at specific intervals during β2M knockdown, not only to quantify the efficiency of the knockdown but to compare it to other genes regulated by the knockdown. There was uncomfortable variation between expressions of some genes throughout the 5 day β2M knockdown. This was evidenced by high error bars for some genes such as Sox1. It is possible that β2M knockdown induced gene specific effects due to it being present in varying levels in the knockdown cells. It is also very difficult to maintain a homogenous cell population during knockdown experiments, something that can contribute to variations in gene expression levels within knocked down cell populations.

The knockdown of YY1AP1 in hUES1 cells induced cell differentiation. With the cells having lost their stem cell status, differentiation of the cells was always expected to induce even more severe effects on cells that were already poised to differentiate. Indeed, the differentiation of the YY1AP1 compromised cells not only shut down pluripotency marker expression but gene expression in the cells altogether as indicated by those genes examined. The lack of up regulation of any of the 3 primary germ layer markers signalled the inability of YY1AP1 deficient cells to differentiate into any specialised cell type. The morphological appearance of EB-like structures in these differentiated cells was therefore, unlike in the control EBs, not a marker of specialised cell formation but rather the formation of cell aggregates that signalled terminal loss of cell function due to the cells having potentially low transcription.

While it was expected that the fact that YY1AP1 knocked down cells were already partially differentiated would accelerate lineage specification of the cells, this was not the case. It was established during the EB differentiation of control cells that differentiation was induced in those cells. Therefore, YY1AP1 must have a distinct role in stem cell differentiation which is worth of further investigation.
YY1AP1 knockdown experiments in embryonic stem cells are almost non-existent. The down regulation of pluripotency transcription factors at both transcript and protein levels were unexpected. Interestingly, protein interaction analysis results in this study revealed that despite their down regulation during YY1AP1 ablation in hESCs, pluripotency genes do not interact directly with YY1AP1 in these cells. Instead, YY1AP1 was shown to interact with the chromatin regular CTCF, which is also an Oct4 co-factor. Because it clearly has an effect on the expression of pluripotency genes, YY1AP1 must be involved in the transcription or regulation of these genes either directly or through transcriptional co-factor(s). The unravelling of this complex associating YY1AP1 with the core transcriptional regulatory network of hESCs should be investigated further to understand the differentiation activation mechanism of YY1AP1. Prior to this study, no known studies had been conducted to check for the transcriptional activity of YY1AP1 in stem cell regulation. With this in mind, studies directed towards investigating a gene with little established knowledge are hard, long and strenuous. For example, there are many different pathways involved in stem cell regulation. Establishing which of these pathways specifically involve YY1AP1 would require patience and long experimental undertakings.

During ChIP-Seq experiments, the ChIP-Seq protocol stipulated that chromatin be collected from a population of 20 million cells. Due to the high cell density that chromatin was collected from, the concentration of the precipitated DNA sample for sequencing was abnormally very high (greater than 50ng/µl). Our DNA sequencing facility requested that only 5-10ng of precipitated DNA should be submitted for sequencing. This meant that my DNA sample had to be diluted prior to the sequencing. Dilution of DNA samples is a known risk factor of foreign DNA introduction which can cause contamination of an otherwise clean DNA sample. Sequencing of the diluted DNA ChIP-Seq samples resulted in low fragment size and high redundancy in mapped genes. Additionally, the number of genes with fold enrichment greater than 10 was very low. Initially, it was the objective of this study to perform the ChIP-Seq for YY1AP1. However, no company could be found that manufactured a YY1AP1 antibody suitable for application in ChIP-Seq experiments. While requesting one of manufacturing companies to design a custom made YY1AP1 ChIP compatible primary antibody was an available option, it would have been outside the timing of this project.
It is important that primary antibodies used during ChIP-Seq experiments have high specificity. The disadvantage of using custom designed antibodies is that their specificity is usually unknown and untested. Because of the predicament, it was decided that ChIP-Seq would be undertaken for YY1, which had an already available ChIP-Seq antibody and whose specificity and efficiency had already been tried and tested. Preliminary Chip-Seq results showed that cell cycle and tumour associated genes mapped to the YY1 promoter. This reiterated the well-known roles of YY1, and its binding partner, YY1AP1 in cell cycle progression and cancer.

In conclusion, the aim of this chapter of investigating the effects of YY1AP1 knock down on the overall gene expression in hESCs was achieved. Results clearly showed the negative impact that YY1AP1 ablation has on the stem cell state and its positive regulation of cell differentiation. Cells underwent cell cycle regulation as shown by transcriptional activation of cell cycle regulatory genes, signalling a potential role for YY1AP1 in cell proliferation or self-renewal.
CHAPTER 7

GENERAL DISCUSSION

Human embryonic stem cells offer great promise to the regenerative medicine community. Their clinical application potential has brought a lot of interest in their research. Pluripotency and self-renewal are two critical properties of stem cells. A lot of research has gone into understanding the functional mechanisms of key transcription factors that govern these properties, such as Oct4, Nanog and Sox2 (Chambers et al., 2003; Boyer et al., 2005; Boyer et al., 2006; Babaie et al., 2007; Fong et al., 2008; Liang et al., 2008, Kunarso et al., 2010, van den Berg et al., 2010; Wang et al., 2012; Kallas et al., 2014). However, there is very limited research on investigating and understanding novel genes and transcription factors that may also play a role in the stem cell regulatory circuitry. It is this factor that motivated this study. An extended understanding of the stem cell pluripotency and self-renewal machinery and the genes involved in this machinery is crucial in helping to identify new player transcription factor players, something that will further enhance our understanding of the stem cell regulation process.

The aim of this study was to identify and investigate the expression and regulation of novel genes that may be involved in the regulation of the self-renewal and pluripotency of human embryonic stem cells. This study was motivated by a previous study which investigated genome-wide binding sites of Oct4, Nanog and CTCF in both hESCs and mouse ESCs (Kunarso et al., 2010).

It was shown in Chapter 3 that the criteria used for selecting potential stem cell regulatory genes is very important in the whole process of mining for novel functional genes. Although the mining of these novel potential genes was performed on microarray data that was generated in our lab, it was still essential to confirm the reliability and usability of these datasets before they could be used for further experimentation. Early reliability tests carried out in this study showed consistency and correlation in targeting of genes by core pluripotency transcription factors such as Oct4, Nanog and Sox2. From our microarray data, highly expressed genes were found to be the ones more targeted by these pluripotency factors. This boded well with
similar previous published studies that found that genes that are highly expressed in stem cells are more likely to be targeted by pluripotency regulatory genes and therefore themselves more likely to be involved in pluripotency regulation (Karolchik et al., 2004; Boyer et al., 2005; Kunarso et al., 2010). This finding greatly complimented the stipulated hypothesis of the study that genes that are responsible for the regulation and maintenance of pluripotency and self-renewal of stem cells are highly expressed in the cells and are most often major targets of transcription factors that regulate these stem cell properties.

The reliability of microarray data was confirmed by the fact that there was more consistency in gene expression between the hUES3 and hUES7 lines than between each of these cell lines and the trophectoderm. Because pluripotent stem cell lines share the unique property of pluripotency, similarities in their transcriptional profiles are expected where pluripotency has been confirmed. Indeed previous studies have confirmed great commonality in gene expression profiles of pluripotent stem cells (Sperger et al., 2003; Abeyta et al., 2004). In conclusion, data obtained from the microarray data screening process in this study showed that our microarray data sets and replicates were reliable and suitable for further use in the project. One of the initial limitations of assessing consistency of the microarray data sets was that it was too laborious to work on many data spreadsheets manually. This set back was however overcome through the help of a bioinformatician and the use of a bioinformatics tool called NETBEANS. This proved to be both output and time efficient.

Differentiation protocols are efficient ways of monitoring gene expression profile changes in cells as they transition from pluripotency to differentiation and ultimate specialization. In this study, growth factor withdrawal differentiation experiments were performed over a 10 day period. This was a sensible time frame considering time limitations. However, in future experiments it would be interesting to see what happens if Activin-A and FGF-2 differentiation protocols are allowed to run beyond the 10-day period. This would allow us to determine what lineage the cells differentiate into in the absence of each growth factor. It would also allow us to determine if the differentiating cells do indeed reach terminal differentiation.
As shown in chapter 4, the potential pluripotency gene YY1AP1 was selected for further experimentation after a series of growth factor and embryoid body differentiation experiments were undertaken. It was reasoned that of the genes potentially associated with the pluripotent state, this gene was the one consistently down regulated during the FGF-2 and Activin-A withdrawal differentiation protocols as well as during differentiation induced by the formation of embryoid bodies. However, it should be noted that there were other equally qualifying candidate genes that could have been selected for further experimental evaluation. YY1AP1 is a binding factor and direct regulator of YY1, which is a transcriptional activator/repressor gene (Wang et al., 2001; Wang et al., 2004). YY1AP1 was steadily down regulated over 10 days of Activin-A withdrawal, FGF-2 withdrawal and EB formation in two different stem cells lines; hUES7 and MAN7. The role of YY1AP1 in stem cell self-renewal and pluripotency has not been investigated. However, its role in cell cycle regulation has been extensively investigated (Li et al., 2007; Chechlinska et al., 2009). YY1 has also been implicated in cell cycle progression in HeLa cells through its increased activity at the S phase during the cell cycle signalling pathway (Palko et al., 2004).

Cell cycle progression is very important in the regulation of pluripotency and self-renewal of human embryonic stem cells. The ability of stem cells to proliferate much quicker than somatic cells is because of their relatively short cell cycle progression due to a short G1 phase duration (Becker et al., 2006; Neganova et al., 2009). It is for this reason that a short G1 phase keeps stem cells self-renewing while an extension of this phase induces differentiation of the cells (Hindley and Philpott, 2013). It has been reported that the duration of the cell cycle influences the self-renewal and pluripotency of the cells with shorter cycles likely to maintain stemness while longer cycles promote differentiation (Becker et al., 2007). This is shown by elevations in the G1 phase related genes such as CDK1 and cyclin D2 (CCND2) during differentiation. Because self-renewal is promoted by the shortening of the G1 phase of the cell cycle (Becker et al., 2006), the down regulation of YY1AP1 during differentiation might have been a direct consequence of loss of self-renewal. Loss of self-renewal initiates differentiation which led to the down regulation of cell cycle regulation genes amongst which is YY1AP1.
During growth-factor withdrawal differentiation of cells, YY1, a YY1AP1 target and binding partner, was down regulated in a similar pattern to YY1AP1. These results correlate with other publications on the role of YY1 in stem cell maintenance. It has been recently shown that YY1 is absolutely crucial for the self-renewal of intestinal stem cells and its deletion promotes cell differentiation (Perekatt et al., 2014). Sui et al. (2004) showed that the ablation of YY1 in lymphoid cells slowed down proliferation rates. YY1 is also involved in the degradation of the tumour suppressor gene p53 (Sui et al., 2004). These data are consistent with our findings of the differentiation protocols that YY1 is needed for self-renewal and proliferation.

As shown in chapter 4, the down regulation of both YY1AP1 and YY1 corresponded with the down regulation of pluripotency markers Oct4, Nanog and Sox2 and the up regulation of differentiation markers GATA4 and Mixl1. This was a suggestion of potential roles of YY1AP1 and YY1 in maintaining the cells in an undifferentiated state. It is however impossible to suggest roles for these two genes in self-renewal or pluripotency maintenance merely from changes in transcript expression levels. Neither could this suggestion be made because of the genes’ nuclear co-localization or because of changes in the morphology of the cells during differentiation. Similarly an inter-regulatory relationship between YY1AP1 and YY1 cannot be established based on the observation that they are both down regulated during embryonic differentiation. It has however been shown in mouse ESCs that Oct4 forms a complex with YY1 at a protein level but only in the presence of CTCF (Donohoe et al., 2009). Currently, there have been no reports of the location of YY1AP1 or YY1 binding sites on the Oct4 promoter. Therefore, no direct interaction between these two genes and Oct4 expression has been established. However, Oct4 and Sox2 sites have been reported to be in close proximity to the CTCF-YY1 sites. Therefore, there was a need to further investigate the regulatory relationship between YY1AP1, YY1 and other stem cell maintenance factors such as Oct4. In the absence of FGF-2 and Activin-A, cell survival is somewhat compromised. In this study, ROCK inhibitor was used to improve cell survival during differentiation. It is therefore possible that ROCK inhibitor could have had an effect on the overall mRNA gene expression levels.

The role of YY1 in promoting cell proliferation has been widely documented (He and Casaccia-Bonnefil, 2008; He et al., 2011; Zhang et al., 2011; Zhu et al., 2011).
However, the role of YY1AP1 in cell proliferation is not yet understood. This study confirmed that YY1AP1 and YY1 are direct binding partners. It has been previously reported that YY1 is a prominent component of the c-Myc network that links embryonic stem cells to cancer progression (Vella et al., 2011). C-Myc is a gene involved in maintaining ESCs in an undifferentiated state. It does this through its transcriptional regulation of the cyclin-CDK complex to induce the down regulation of CDKIs (Singh and Dalton, 2009). Because YY1 and YY1AP1 reciprocally bind to each other’s promoter region, YY1AP1 might interact with c-Myc but only in complex with YY1, which is known c-Myc target gene. It has been reported that c-Myc interacts with the pluripotency associated transcription factor CTCF and that upon CTCF depletion, c-Myc is down regulated (Balakrishnan et al., 2012). Data in this thesis indicates that YY1AP1 interacts with CTCF at the protein level. It has also been reported that YY1 interacts at the protein level with Oct4 but only in complex with CTCF (Donohoe et al., 2007; Donohoe et al., 2009). Indeed, this study found through protein to protein interaction experiments that CTCF was present in YY1 immunoprecipitated lysate. However, the presence of oct4 in the YY1 lysate could not be established. Figure 7.1 show a model of the potential involvement of YY1AP1 in the stem cell self-renewal circuitry.
Figure 7.1. A proposed model of the potential role of YY1AP1 during the self-renewal of stem cells. YY1AP1 may directly or indirectly through YY1 bind to CTCF, a self-renewal transcription factor. CTCF can regulate self-renewal directly or indirectly through its association with c-Myc. C-Myc-CTCF complex binds to Cyclin CDK and this suppresses CDKIs and induces self-renewal. YY1 can bind to Oct4 only in complex with CTCF to activate self-renewal. Green lines = up regulation; Blue dotted lines = potential regulation; Red line = down regulation.
In any knockdown experiment, transfection efficiency is very important. Due to the stress put on them by nucleofection, transfected cells are more vulnerable than normal non-transfected cells. It is therefore important that in future knockdown experiments, there is a development of better culture conditions for transfected cells in order to maximise the transfection efficiency.

Knockdown and over expression research on YY1AP1 is limited in hESCs studies. It was shown in chapter 6 that knocking down YY1AP1 led to the down regulation of YY1AP1 itself and its binding partner YY1. At the same time, pluripotency transcription factors Oct4, Nanog and Sox2 were all down regulated after YY1AP1 knock down. The reduction of YY1AP1 in the cells might have initiated differentiation. MAD2L2 is a gene involved in the regulation of the cell cycle and has been shown to be a direct target of YY1AP1 (Li et al., 2007). It forms a complex with YY1AP1 and this complex negatively regulates cell cycle progression by inducing cell cycle arrest at the G0/G1 phase. Cell cycle arrest at the G1 phase leads to a decline in the proliferation capacity of the cells and results in the differentiation of the cells due to this inhibition of cell proliferation. Proliferation of the cells could also be inhibited by the transcriptional repression of G1 phase genes causing elongation of the G1 phase. This then hinders the self-renewal capacity of the cells (Li et al., 2007).

After YY1AP1 knockdown in hUES1 stem cells, there was an up regulation of; most of the mesodermal markers, both ectodermal markers Pax6 and Sox1, and cell cycle genes c-Myc and p53. These genes are may therefore be suppressed by YY1AP1. These results may suggest that YY1AP1 does not promote endodermal differentiation and is not needed for stem cell differentiation towards the mesodermal and ectodermal lineages; an indication that it is needed during stem cell self-renewal and proliferation. This might indicate that YY1 directly or indirectly represses these differentiation genes. Interestingly, after the loss of YY1AP1, CCND1 was consistently down regulated, signalling the importance of YY1AP1 to the transcriptional activity of CCND1. YY1, but not YY1AP1, has been reported to be a negative regulator of p53. It inhibits the transcriptional activity of p53 by directly binding to it and disrupting its interaction with its co-activator p300, inducing cell growth arrest or apoptosis (Sui et al., 2004). It was therefore not surprising that after the ablation of YY1AP1 from the cells, there was a massive up regulation of p53.
There were some variations with some of the transcript gene expression q-RT-PCR results from knockdown experiments. While there was differentiation after YY1AP1 knockdown in the hUES1 cells, as shown by the up regulation of the different lineage markers, this differentiation was not directed to any particular lineage. Therefore, cell differentiation through the knockdown of YY1AP1 was not expected to promote any one particular lineage. Indeed, there is no literature evidence to suggest that the ablation of YY1AP1 in hUES1 cells favours their differentiation to one particular lineage. Variations in transcript expression levels of the differentiated could also have been brought internal gene specific activities that occurred during knockdown. YY1AP1, for example, could have been involved the regulatory complexes of these genes and therefore affected their final transcript expression levels at each time point of the knockdown protocol. Variations were also observed in gene expression levels after differentiation which could have been a factor of the fact that differentiation was performed on cells that had undergone different treatments during control knockdowns (i.e. for the control EBs) and were likely of mixed populations of varying degrees leading to the inevitable differences in their transcripts levels. It is however important to point out that these experiments were repeated three times, which further strengthens the suspicion that expression variations were likely due intrinsic cellular complications than experimental mishaps. The knockdown results were very variable. This was in part due to the fact that during analysis, stem cells undergo changes such as differentiation at different rates depending on complex factors. This can be mitigated by the use of flow cytometry to select specific cell populations using pluripotency and differentiation markers. This can greatly improve experimental reliability due to the homogeneity of the final cell population.

YY1AP1 knockdown in hUES1 cells led to the down regulation of Oct4, Nanog and Sox2 and a simultaneous up regulation of differentiation makers. In future experiments, overexpression of YY1AP1 should be carried out in pluripotent stem cells to elucidate its overall effect on the gene expression profile of the cells when it is in abundance. The general expectation would be to see the reverse results of the knock down where pluripotency is enhanced and differentiation is repressed. This would reinforce the findings of the knock down experiment and strengthen the experimental design.
Due to time limitations and the intensively time consuming process of undertaking knockdown experiments, YY1AP1 knockdown was performed in hUES1 cells only, although it was performed in MCF-7 cells as a proof of principle. Moreover the knockdown and all the control knockdown experiments were repeated three times in this cell line to ascertain results consistency. Therefore, time did not allow me to repeat the whole process in a different stem cell line. However, in future knockdown experiments, it would be ideal to repeat the YY1AP1 knockdown in a different cell line such as the MAN line or even in iPSCs. This would not only strengthen the conclusions drawn from these experiments but would also confirm if the experiments can be replicated in stem cells derived from a different environment and different genetic background.

Stem cell pluripotency and self-renewal are heavily intertwined. Therefore, in future experiments on YY1AP1’s role in stem cell maintenance, it would be interesting to assess the effect of YY1AP1 in the formation of induced pluripotent cells (iPSCs) during reprogramming. This would reveal whether YY1AP1 can be used as one of the reprogramming factors or if it has detrimental effects on reprogramming. During reprogramming two pluripotency transcription factor combinations can be used. Oct4, Sox2, c-Myc and Klf4 are the most commonly used transcription factors during reprogramming. Alternatively, a combination of Oct4, Nanog, Sox2 and Lin28 can also be used to reprogram somatic cells. The experimental setup for this analysis would be to replace one of the genes in each combination (except Oct4, Nanog and Sox2) with YY1AP1 during reprogramming. For example, in the first combination, Klf4 would be replaced by YY1AP1 while in the second combination; Lin28 would be the one replaced by YY1AP1. Such an experiment will elucidate the nature of YY1AP1’s role in stem cell maintenance and investigate if its role goes beyond just having a potential regulatory role in maintaining stem cell self-renewal. Also, the roles, if any, of both YY1AP1 and YY1 in early differentiation should be investigated in order to determine if these two genes cause lineage bias.

It is also important that in future to investigate the regulatory mechanisms of both YY1AP1 and YY1 within the cell. This could be done by developing YY1AP1 and YY1 reporters in order to determine how these genes are localised in the cell; how they are translocated to the nucleus and under what control.
It is worthy that in such future studies, there is an investigation to check if there are any YY1AP1 or YY1 binding sites located on the Oct4 promoter region in order to determine if these transcription factors directly interact with Oct4. This would reveal whether Oct4 or any of the stem cell pluripotency and self-renewal transcription factors are directly regulated by YY1AP1 during stem cell maintenance. One way of doing this is performing Chip-Seq experiments for Oct4. Alternatively, where a ChIP-Seq antibody for YY1AP1 with confirmed specificity has been designed, the ChIP-Seq experiment could be performed for YY1AP1. Chip-Seq experiments could be performed for other potential Oct4 regulators such as CTCF in order to determine if they have YY1AP1 or YY1 binding sites.

In cancer studies, YY1 has been widely described in regulating carcinogenesis. YY1 is known for its dual and contradictory functions of transcriptional activation and transcriptional repression to induce or inhibit gene expression respectively. It is because of this that it has been associated with both tumour proliferation and suppression. In human breast cancer tumours, YY1 regulates carcinogenesis by directly interacting with cell cycle signalling pathway genes such as CCND1. These cell signalling genes then drive the G1 phase of cell cycle progression, resulting in the stimulation of proliferation of the tumour cells. YY1 is also involved in the regulation of cell cycle progression in mouse ESCs. Its ablation in mESCs induces cell cycle arrest while its inhibition in HeLa cells inhibits proliferation of the cells (Affar et al., 2006). It has been shown that YY1 is more highly expressed in prostate cancer cells than it is in normal human tissues (Seligson et al., 2005). Therefore, it might be hypothesized that the potential use of YY1 expression as a prognostic marker of prostate cancer might also predict tumour recurrences.

Assessment of the effect of YY1AP1 knockdown on the expression of gene proteins correlated with their expression level profiles at the transcript level. In both YY1AP1 and YY1 lysates, Oct4 and Nanog were down regulated after knockdown while p53 was up regulated. The activation of p53 and its target gene p21 in hESCs stimulate fast differentiation of the cells into the primitive endodermal and trophectodermal lineages through the closure of the S phase of cell cycle progression, causing arrest at the G0/G1 phase (Maimets et al., 2008). Because p53 is inhibited by hdm2, it is only after the degradation hdm2 that the transcriptional activity of p53 is restored and
differentiation is activated (Vousden and Lu, 2002). This therefore results in loss of pluripotency. P21 has itself been shown to negatively regulate the self-renewal of stem cells (Cheng et al., 2000). The maintenance of stem cell integrity requires a coordinated circuitry of the core pluripotency transcription factors; Oct4, Nanog and Sox2 with other co-factors such as Polycomb Repressive Complexes (Kashyap et al., 2009). This includes YY1 which is a Polycomb protein. The destruction of this coordinated regulatory circuitry initiates differentiation which leads to their subsequent down regulation.

During Chip-Seq experiments performed in this study, 20 million cells were used per sample for a single sequencing run. Generating such a high number of cells was both time consuming and labour intensive. Moreover, a lot of time was wasted while waiting for these cells to grow, something that was counterproductive. It is also imperative that improved cell culture methods are devised to speed up cell growth for experiments that require a high number of cells such as Chip-Seq.

In conclusion, understanding stem cell maintenance mechanisms will go a long in enhancing their clinical applicability. The novel contribution of this thesis was identifying the transcription factor YY1AP1 as a potentially novel player in the stem cell self-renewal and proliferation regulatory system. It was shown that YY1AP1 is lost upon stem cell differentiation. YY1AP1 ablation in stem cells led to the suppression of the transcriptional activity of pluripotency transcription factors Oct4, Nanog and Sox2. It was also that YY1AP1 did not promote any lineage specification and that it interacts with Nanog and CTCF at the protein level. Finally, it was shown that differentiation of YY1AP1 deficient cells led to total loss of gene expression. The findings of this study therefore strongly suggest a likely role of YY1AP1 in keeping pluripotent stem cells in an undifferentiated state. The findings of this research may provide a framework for future studies into human embryonic stem cell maintenance mechanisms.
REFERENCES


