Post transcriptional regulation of microRNAs in embryonic stem cells

Diana Alexieva Alexieva

Imperial College London
Department of Surgery and Cancer

Thesis submitted to Imperial College London for the degree of Doctor of Philosophy
Declaration of originality

All experiments included in this thesis were performed by me unless otherwise stated in the text.
Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.
Acknowledgements

I would like to thank my supervisor Dr Nick Dibb for giving me the opportunity to work on a project of genuine interest to me, for all the guidance and support and the scientific discussions during which many ideas contributing to this work were generated.

Many thanks to Wei Cui and everyone in her group who supplied the human stem cells used in my experiments and to Ihsan who looked after them. Thanks to Rupa, Chin and Rakhee for their guidance in the lab and for always being helpful. I would like to thank the fifth floor IRDB members, past and present, for all the help and advice.

My special thanks go to my family for the continuous support and encouragement throughout this and previous endeavours.

Special thanks to Krassi for all her support and understanding, especially in the past few years. Many thanks to the IRDB PALS without who my PhD experience would not have been the same. I would like to thank my closest friends who live in different parts of the world who never fail to show their support despite the distance. Thanks to all my friends in London for making my time here so enjoyable.
Abstract

In this project a block in processing of primary miRNA transcripts (pri-miRNAs) in human embryonic stem cells was demonstrated. Pri-miRNAs for 15 out of 24 tested intergenic miRNAs were detected by RT-PCR in the H1 cell line despite a lack of detection of their mature form in deep sequencing data. By contrast, only 2 out of 16 pri-miRNAs for which expression of the mature form was not reported were successfully amplified by RT-PCR in HEK293T cells. A reduction in the pri-miRNA levels was observed upon blocking transcription in H1 cells and the vast majority of the same pri-miRNAs were detected by RT-PCR using nascent RNA, indicative of their active transcription and degradation.

To test whether hESCs cells vigorously transcribe yet degrade more mRNA transcripts in general than differentiated cells we isolated nascent and total RNA samples from H1 and HEK293T cells and performed a microarray analysis using the Illumina HT-12 v4 platform. There were no significant differences in the number of genes transcribed in the two cell lines and across both types of RNA samples. These results suggested that a similar proportion of the coding part of the genome is actively or vigorously transcribed in both hESCs and differentiated cell types.

It is generally thought that ESCs have a higher level of background transcription than normal, due to the more open chromatin in this cell type. We therefore suggest that background transcription might be responsible for our detection of immature miRNA transcripts in hESCs instead of vigorous transcription and that enhanced background transcription has a measurable impact perhaps because pri-miRNAs are normally only transcribed at relatively low rates compared to other genes. We still need to explain why some pri-miRNA transcripts are degraded in ESCs. One possibility is that such miRNAs express both sense and antisense transcripts.
Table of Contents

Declaration of originality ........................................................................................................... 2
Copyright declaration ................................................................................................................ 3
Acknowledgements .................................................................................................................... 4
Abstract ...................................................................................................................................... 5
List of figures ............................................................................................................................ 10
List of tables ............................................................................................................................. 14
Abbreviations ........................................................................................................................... 15

Chapter 1- Introduction ........................................................................................................... 20
  1.1 The discovery of RNA interference ........................................................................... 20
  1.2 The discovery of microRNAs ..................................................................................... 21
  1.3 Genomic organisation and biogenesis of miRNAs .................................................... 22
    1.3.1 Genomic organisation of miRNAs ...................................................................... 22
    1.3.2 Nuclear processing and export of pre-miRNAs .................................................. 24
    1.3.3 Cytoplasmic processing of pre-miRNAs and target gene silencing ................... 25
    1.3.4 Non-canonical miRNAs ...................................................................................... 27
  1.4 Regulation of miRNA biogenesis ............................................................................... 28
    1.4.1 Transcriptional control ....................................................................................... 28
    1.4.2 Post transcriptional regulation of miRNA biogenesis ........................................ 34
  1.5 Embryonic stem cells- properties and uses .............................................................. 39
  1.6 The pluripotent state of ESCs .................................................................................... 40
  1.7 Transcription in ESCs ................................................................................................. 42
  1.8 MicroRNAs and ESCs ................................................................................................. 43
    1.8.1 MiRNAs are essential for maintaining self-renewal of ESCs and control
    pluripotency and differentiation ...................................................................................... 43
    1.8.2 Post transcriptional regulation of miRNAs in ESCs ............................................ 44
  1.9 Aims ........................................................................................................................... 46

Chapter 2- Materials and methods .......................................................................................... 47
  2.1 Cell culture and freezing cell lines ............................................................................. 47
    2.1.1 Cell culture of H1 cells ....................................................................................... 47
    2.1.2 Cell culture of E14 cells ...................................................................................... 48
    2.1.3 Cell culture of HEK293T cells ............................................................................. 48
2.1.4 Freezing cell lines ........................................................................................................ 49
2.2 RNA extraction and checking RNA quality .................................................................... 50
2.2.1 Total RNA extraction .................................................................................................. 50
2.2.2 Small RNA extraction ............................................................................................... 50
2.2.3 Checking RNA integrity ............................................................................................ 51
2.3 cDNA synthesis, primer design, PCR and PCR product purification ......................... 51
2.3.1 First strand cDNA synthesis .................................................................................... 51
2.3.2 MicroRNA gene selection and primer design ....................................................... 52
2.3.3 Polymerase chain reaction (PCR) ............................................................................ 52
2.3.4 PCR product elution ............................................................................................... 53
2.3.5 Quantitative PCR (qPCR) ....................................................................................... 53
2.4 Immunohistochemistry of nascent RNA ....................................................................... 54
2.5 Northern blotting .......................................................................................................... 55
2.5.1 Preparation and running a denaturing urea gel .................................................... 55
2.5.2 RNA transfer ............................................................................................................ 56
2.5.3 Labelling of DNA probes with [γ-32P] ATP .............................................................. 56
2.5.4 Hybridisation ............................................................................................................ 57
2.6 Isolation of nascent RNA ............................................................................................. 57
2.6.1 Cell labelling and biotinylation of 4sU-tagged RNA .............................................. 57
2.6.2 Magnetic purification of nascent RNA ................................................................. 58
2.6.3 Dot blot analysis of biotinylated RNA .................................................................... 58
2.7 MicroRNA pull-down .................................................................................................. 59
2.7.1 Labelling of DNA probes with [γ-32P] ATP ............................................................... 59
2.7.2 Incubation with μMACS streptavidin beads ......................................................... 59
2.7.3 Incubation with streptavidin magnetic beads (NEB) ............................................. 60
2.7.4 Denaturing gel and signal detection ...................................................................... 61
2.8 Microarray analysis .................................................................................................... 61
2.9 List of bioinformatics programs used ........................................................................... 62
2.10 Lists of primers used in chapters 3 and 5 .................................................................. 63
Chapter 3- Post transcriptional regulation of miRNAs in embryonic stem cells ............... 66
3.1 Introduction .................................................................................................................... 66
3.1.1 Previous work in our lab suggests that post transcriptional regulation of primary miRNAs is prevalent in hESCs .......................................................... 66
3.1.2 Rationale and aims ................................................................................................. 68
3.2 Results ........................................................................................................................................... 70

3.2.1 Many microRNA transcripts can be detected by RT-PCR but not by sequencing in hESCs 70

3.2.2 MicroRNA transcripts are actively turned over in hESCs ......................................................... 73

3.2.3 MicroRNA transcripts can be detected by RT-PCR but sequencing is showing a low or absent expression of their mature form in mouse ESCs ................................................. 80

3.3 Discussion ....................................................................................................................................... 82

3.3.1 Many pri-miRNAs are transcribed, but degraded in hESCs ....................................................... 82

3.3.2 Post transcriptionally regulated miRNAs play roles in cell fate determination and cancer progression .................................................................................................................. 85

Chapter 4- Optimising the isolation of nascent RNA from HEK293T and H1 hESCs ................. 87

4.1 Introduction ..................................................................................................................................... 87

4.1.1 The importance of studying changes at the mRNA level through isolating nascent RNA ................................................................................................................................. 87

4.1.2 Methods for the isolation of nascent RNA .................................................................................. 88

4.1.3 Rationale and Aims .................................................................................................................... 90

4.2 Results ........................................................................................................................................... 92

4.2.1 Optimising the duration of labelling, 4-thiouridine (4sU) concentration and precipitation of nascent RNA ................................................................................................................... 92

4.2.2 MTSEA biotin is more efficient in binding to 4sU-labelled RNA than HPDP biotin, but it is not 100% 4sU-specific .................................................................................................................. 95

4.3 Discussion ....................................................................................................................................... 99

4.3.1 Selection of a technique for analysing nascent RNA transcripts .............................................. 99

4.3.2 200 µM of 4sU, 30 minutes of labelling, HPDP biotin and isopropanol/NaCl precipitation were found to be the optimal conditions for the isolation of nascent RNA from HEK293T and H1 hESCs .................................................................................................................. 100

Chapter 5- Microarray gene expression analysis of total and nascent RNA in H1 and H293T cell lines ........................................................................................................................................ 102

5.1 Introduction ..................................................................................................................................... 102

5.1.1 Background ............................................................................................................................... 102

5.1.2 Hypothesis and aims ................................................................................................................ 103

5.2 Results ........................................................................................................................................... 104

5.2.1 The microarray controls worked well and all samples used in the experiment were included in the data analysis ......................................................................................................... 104

5.2.2 H1 hESCs transcribe and process a similar number of genes to HEK293T cells 105
5.2.3 Microarray data confirmed the identity of H1 hESCs by showing they are enriched for stem cell-specific genes .........................................................109
5.2.4 Microarray may not provide high enough sensitivity for the detection of genes expressed at a low level ..............................................................115
5.3 Discussion ................................................................................................................119
5.3.1 No significant differences were found at the steady-state transcript level between H1 and HEK293T cells .................................................................119
5.3.2 The technical performance of the microarray was good and the H1 cells used in the experiment were in an undifferentiated state.............................120
5.3.3 The microarray sensitivity is limited .................................................................122

Chapter 6- Optimising the pull-down of specific miRNAs ........................................124
6.1 Introduction .............................................................................................................124
6.1.1 Biases in miRNA representation in deep sequencing analysis ....................124
6.1.2 Rationale and aims ...........................................................................................126
6.2 Results .....................................................................................................................128
6.3 Discussion ................................................................................................................136

Chapter 7- General Discussion .....................................................................................138

Appendix ......................................................................................................................169
List of figures

Figure 1.1 Genomic organisation of miRNAs

Figure 1.2 Biogenesis of canonical miRNAs.

Figure 1.3 Autoregulatory feedback loops coordinate the expression of transcription factors (TFs) and miRNAs in development.

Figure 1.4 Transcriptional feed-forward loops (FFLs) consist of miRNAs and TFs regulating each other whilst each of them controls target gene expression.

Figure 3.1 Figure 3.1 Primer design for the detection of pri-miRNA transcripts

Figure 3.2 Post transcriptional regulation of miRNAs is more prevalent in hESCs than in NPSCs.

Figure 3.3 Deep sequencing data on 24 intergenic miRNAs (A) and RT-PCR (B) for detecting the primary transcripts of the same miRNAs in H1 hES and HEK293T cells.

Figure 3.4 Northern blots (A) were performed to confirm sequencing data for the expression of miRNAs in hESCs for which primary transcripts were detected by RT-PCR (B).

Figure 3.5 Comparison between transcript abundance of miRNAs enriched in stem cells (miR-182, miR-183) and miRNAs that are blocked in processing in H1 cells (miR-24-2, miR-27a and miR-345).

Figure 3.6 Testing the turnover of primary miRNA transcripts by actinomycin D treatment.

Figure 3.7 Immunohistochemistry staining of hESCs labelled with 5-FU showing the inhibition of transcription induced by actinomycin D.

Figure 3.8 Quantifying the reduction in the levels of primary transcripts for miR-24-2 and miR-345 by qPCR after treatment of hESCs from the H1 cell line with 1 µM actinomycin D for 4 hours.

Figure 3.9 Confirmation of the active transcription of miRNAs by RT-PCR (35 cycles) on two nascent RNA samples isolated from H1 hESCs labelled with 200 µM 4sU for 30 minutes.
Figure 3.10  Confirmation of the active transcription of miRNAs by RT-PCR (35 cycles) on two nascent RNA samples isolated from HEK293T cells labelled with 200 µM 4sU for 30 minutes

Figure 3.11  Detection of primary miRNA transcripts whose mature form is detected at a low level by sequencing of E14 mESCs small RNA.

Figure 3.12  Testing the turnover of primary miRNA transcripts by actinomycin D treatment in mouse embryonic stem cells from the E14 cell line

Figure 4.1  Schematic of the protocol used for the isolation of nascent RNA

Figure 4.2  Testing of different concentrations (A) and labelling times (B) of 4sU for the isolation of nascent RNA.

Figure 4.3  Comparison between isopropanol/NaCl precipitation and the RNeasy MinElute Cleanup Kit (Qiagen) for the recovery of nascent RNA

Figure 4.4  Isolation of nascent RNA from H1 hESCs

Figure 4.5  Comparison between MTSEA and HPDP biotin for the isolation of nascent RNA

Figure 4.6  Testing the specificity of MTSEA biotin. H1 hESCs were labelled with 200 µM 4sU for one hour

Figure 4.7  RT-PCR on nascent RNA from H1 hESCs labelled with 200 µM 4sU for 30 minutes. Nascent RNA was crosslinked to HPDP biotin

Figure 5.1  A dendrogram showing the correlation between the 12 RNA samples used in the microarray experiment

Figure 5.2  A Venn diagram showing the number of microarray probes (also expressed as a percentage) with a detection p-value <0.05 in total and nascent RNA samples in H1 hESCs and HEK293T cells

Figure 5.3  Cell type-specific Venn diagrams showing the number of microarray probes with a detection p-value <0.05 in total and nascent RNA samples in H1 hESCs and HEK293T cells (A and B)

Figure 5.4  A diagram showing the genes from our analysis involved in the maintenance of pluripotency in embryonic stem cells, the most highly upregulated pathway in H1 hESC total RNA compared to HEK293T total RNA.
Figure 5.5 Validating microarray data by RT-PCR. Cell- and sample-specific gene expression was confirmed by performing 25 cycles of RT-PCR.

Figure 5.6 Validating microarray data by RT-PCR. 35 cycles of RT-PCR were performed to screen for genes which were not detected in the microarray analysis (detection p-value>0.05).

Figure 5.7 Validating microarray data by RT-PCR. 25 cycles of RT-PCR were performed to screen for genes which were not detected in the microarray analysis (detection p-value>0.05).

Figure 6.1 Schematic of the proposed pull-down technique for the isolation of specific miRNAs.

Figure 6.2 Schematic of the protocol used for the pull-down of [γ-32P] ATP-labelled miRNA oligonucleotides.

Figure 6.3 Pull-down of [γ-32P] ATP-labelled miR-101-1 RNA oligonucleotide using the Miltenyi Biotec µMACS streptavidin kit.

Figure 6.4 Pull-down of [γ-32P] ATP-labelled miR-101-1 DNA oligonucleotide using the New England Biolabs streptavidin magnetic beads.

Figure 6.5 Figure 6. 5 Pull-down of [γ-32P] ATP-labelled miR-101-1 DNA oligonucleotide using the New England Biolabs streptavidin magnetic beads in the background of 0.45 µg of let-7a-1 DNA oligonucleotide (lanes 1-4) or 3 µg of HEK293T total cell RNA (lanes 5-8).

Figure 6.6 MiRNAs with a high similarity to miR-101-1.

Figure 6.7 Pull-down of [γ-32P] ATP-labelled miR-101-1 and [γ-32P] ATP-labelled DNA oligonucleotides with highly similar sequences to miR-101-1.

Figure 6.8 Pull-down of [γ-32P] ATP-labelled miR-101-1 and [γ-32P] ATP-labelled DNA oligonucleotides with highly similar sequences to miR-101-1 (miR-141 and miR-542).

Figure 6.9 Measuring the efficiency of miR-101 pull-down under different annealing temperatures.

Figure S5.1 Control summary plots showing the data quality generated by the Illumina HT-12 array.
Figure S5.2  Quality control guidance notes on interpreting quality control plots from Illumina HT-12 array.

Figure S5.3  Comparison between perfect match (PM) and mismatch (MM2) probe signals across the 12 samples used in the microarray experiment.

Figure S5.4  Sample quality was assessed by creating line plots of control metrics for each of the 12 samples used in the microarray experiment.

Figure S5.5  Summary bar chart showing the most highly up and downregulated pathways in H1 hESCs compared to HEK293T cells.
List of tables

Table 2.1  Forward and reverse primer pairs used in chapter 3  64

Table 2.2  Forward and reverse primer pairs used in chapter 5  65

Table 3.1  Summary table of the detection (√) or lack of detection (×) of pri-miRNA transcripts by RT-PCR in total and nascent RNA and sequencing data (high or low) of mature miRNAs in H1 hESCs and HEK293T cells.  79

Table 5.1  Numbers of significantly up and downregulated genes in HEK293T and H1 hESCs nascent RNA in comparison to cell type-specific total RNA samples (columns 2 and 3)  108

Table 5.2  Detection of all 576 miRNA probes available on the Illumina HT-12 microarray in total and nascent RNA samples in H1 and HEK293T cells.  109

Table 5.3  Differential expression of cell specific genes.  110

Table 5.4  Cell specific gene expression was shown by microarray analysis.  113

Table 5.5  Readout values for 10 genes not detected by microarray (detection p-value>0.05)  115

Table 6.1  Analysis of band intensity from figure 6.9 in ImageJ  135
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3' and 5' RACE PCR</td>
<td>3’ and 5’ rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>3’ untranslated region</td>
</tr>
<tr>
<td>4sU</td>
<td>4-thiouridine</td>
</tr>
<tr>
<td>4tU</td>
<td>4-thiouracil</td>
</tr>
<tr>
<td>5-FU</td>
<td>5-fluorouracil</td>
</tr>
<tr>
<td>ACTB</td>
<td>actin beta</td>
</tr>
<tr>
<td>actD</td>
<td>actinomycin D</td>
</tr>
<tr>
<td>ADAR</td>
<td>adenosine deaminase acting on RNA</td>
</tr>
<tr>
<td>AEL</td>
<td>acute erythroid leukaemia</td>
</tr>
<tr>
<td>AGO proteins</td>
<td>argonaute proteins</td>
</tr>
<tr>
<td>AICDA</td>
<td>activation induced cytidine deaminase</td>
</tr>
<tr>
<td>AMP</td>
<td>adenosine monophosphate</td>
</tr>
<tr>
<td>APS</td>
<td>ammonium persulfate</td>
</tr>
<tr>
<td>ASCs</td>
<td>adult stem cells</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bFGF</td>
<td>basic fibroblast growth factor</td>
</tr>
<tr>
<td>BMI1</td>
<td>B cell-specific Moloney murine leukemia virus integration site 1</td>
</tr>
<tr>
<td>BMP4</td>
<td>bone morphogenetic protein 4</td>
</tr>
<tr>
<td>BrdU</td>
<td>bromodeoxyuridine</td>
</tr>
<tr>
<td>BrU</td>
<td>5' bromouridine</td>
</tr>
<tr>
<td>BrUTP</td>
<td>5-bromouridine 5'-triphosphate</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CCR4-NOT</td>
<td>chemokine (C-C Motif) receptor 4- negative regulator of transcription 1</td>
</tr>
<tr>
<td>CD24</td>
<td>cluster of differentiation 24</td>
</tr>
<tr>
<td>CDH1</td>
<td>cadherin 1</td>
</tr>
<tr>
<td>CDKN1A</td>
<td>cyclin dependent kinase inhibitor 1A</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>cyclin dependent kinase inhibitor 2A</td>
</tr>
</tbody>
</table>
cDNA  complementary DNA
CDX2  caudal type homeobox 2
Chd1  chromodomain helicase DNA binding protein 1
COG-1  conserved oligomeric Golgi complex subunit 1
CSCs  cancer stem cells
Ct  cycle threshold
DAPI  4',6-Diamidino-2-Phenylindole, Dihydrochloride
DDX17  DEAD-Box Helicase 17
DDX5  DEAD-Box Helicase 5
DGCR8  DiGeorge syndrome critical region 8
DMEM  Dulbecco's modified eagle medium
DMF  dimethylformamide
DMSO  dimethyl sulfoxide
DNMT3B  DNA methyltransferase 3 beta
dNTPs  deoxynucleotides
DPBS  Dulbecco’s phosphate buffered saline
DPPA4  developmental pluripotency associated 4
dsRNA  double-stranded RNA
DTT  dithiothreitol
EDTA  ethylenediaminetetraacetic acid
EMT  epithelial-to-mesenchymal transition
EpiSCs  epiblast stem cells
ER  endoplasmic reticulum
EU  5-ethynyluridine
FBS/FCS  foetal bovine serum/foetal calf serum
FDR  false discovery rate
FFLs  feed-forward loops
FIT1  fat-inducing transcript 1
FOXF2  forkhead-related transcription factor 2
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde 3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GDP</td>
<td>guanosine diphosphate</td>
</tr>
<tr>
<td>GEMEM</td>
<td>Glasgow’s minimal essential medium</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>hAD-MSCs</td>
<td>human adipose tissue-derived mesenchymal stem cells</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>hESC</td>
<td>human embryonic stem cells</td>
</tr>
<tr>
<td>HIC1</td>
<td>hypermethylated in cancer 1</td>
</tr>
<tr>
<td>hMSCs</td>
<td>human mesenchymal stem cells</td>
</tr>
<tr>
<td>hnRNP A1</td>
<td>heterogeneous nuclear ribinucleoprotein A1</td>
</tr>
<tr>
<td>HOXA5</td>
<td>homeobox A5</td>
</tr>
<tr>
<td>HOXA9</td>
<td>homeobox A9</td>
</tr>
<tr>
<td>HPX</td>
<td>hemopexin</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IL1RN</td>
<td>interleukin 1 receptor antagonist</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity Pathway Analysis</td>
</tr>
<tr>
<td>IPO8</td>
<td>importin 8</td>
</tr>
<tr>
<td>iPSCs</td>
<td>induced pluripotent stem cells</td>
</tr>
<tr>
<td>KHSRP</td>
<td>KH-type splicing regulatory protein</td>
</tr>
<tr>
<td>KLF9</td>
<td>kruppel like factor 9</td>
</tr>
<tr>
<td>KO-DMEM</td>
<td>knockout DMEM</td>
</tr>
<tr>
<td>KSR</td>
<td>knockout serum replacement</td>
</tr>
<tr>
<td>L1TD1</td>
<td>LINE1 type transposase domain containing 1</td>
</tr>
<tr>
<td>LIF</td>
<td>leukaemia inhibitory factor</td>
</tr>
<tr>
<td>MEF-CM</td>
<td>mouse embryonic fibroblast-conditioned medium</td>
</tr>
<tr>
<td>MEFs</td>
<td>mouse embryonic fibroblasts</td>
</tr>
<tr>
<td>mESCs</td>
<td>mouse embryonic stem cells</td>
</tr>
<tr>
<td>miRNAs</td>
<td>microRNAs</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
</tbody>
</table>
MYOD1  myogenin and myoblast determination protein 1
NECAB1  N-Terminal EF-hand calcium binding protein 1
NMD pathway  nonsense mediated RNA decay pathway
NPSCs  neural progenitor stem cells
OR2A2  olfactory receptor family 2 subfamily A member 2
PABP  poly(A)-binding protein
PACT  protein activator of PKR
PAGE  polyacrylamide gel electrophoresis
PAN2  PAN2 Poly(A) Specific Ribonuclease Subunit
PAN3  PAN3 Poly(A) Specific Ribonuclease Subunit
PARN  polyA-specific endonuclease
P-bodies  processing bodies
PBS  phosphate-buffered saline
PBS-BT  PBS supplemented with 0.1 % Bovine serum albumin (BSA) and 0.2 % Tween 20
PDCD4  programmed cell death 4
PITX3  paired like homeodomain 3
PKR  protein kinase R
POPDC2  Popeye domain containing 2
POU5F1 aka OCT4  POU class 5 homeobox 1/octamer-binding Protein 4
pre-miRNAs  precursor miRNAs
pri-miRNAs  primary miRNAs
PTGS  post transcriptional gene silencing
qPCR  quantitative polymerase chain reaction
RISC  RNA-induced silencing complex
RNAi  RNA interference
RP11L1  retinitis pigmentosa 1 like 1
rRNA  ribosomal RNA
RT-PCR  reverse transcription polymerase chain reaction
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Official Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of the mean</td>
</tr>
<tr>
<td>SF2/ASF</td>
<td>pre-mRNA-splicing factor SF2/alternative splicing factor</td>
</tr>
<tr>
<td>siRNAs</td>
<td>small interfering RNAs</td>
</tr>
<tr>
<td>snoRNA</td>
<td>small nucleolar RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SOX2</td>
<td>sex determining region Y-box 2</td>
</tr>
<tr>
<td>SRSF3</td>
<td>serine and arginine rich splicing factor 3</td>
</tr>
<tr>
<td>SSC buffer</td>
<td>saline-sodium citrate buffer</td>
</tr>
<tr>
<td>ssRNA</td>
<td>single-stranded RNA</td>
</tr>
<tr>
<td>SSTR3</td>
<td>somatostatin receptor 3</td>
</tr>
<tr>
<td>SYCP2L</td>
<td>synaptonemal complex protein 2 like</td>
</tr>
<tr>
<td>T4 Rnl1/2</td>
<td>T4 RNA ligase 1/2</td>
</tr>
<tr>
<td>TAE buffer</td>
<td>Tris-acetate-EDTA buffer</td>
</tr>
<tr>
<td>TBE buffer</td>
<td>Tris-borate-EDTA buffer</td>
</tr>
<tr>
<td>TEMED</td>
<td>tetramethylethlenediamine</td>
</tr>
<tr>
<td>TGF-β</td>
<td>transforming growth factor beta</td>
</tr>
<tr>
<td>TRBP</td>
<td>trans-activator (TAR)-RNA-binding protein</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TUDOR SN</td>
<td>Tudor staphylococcal nuclease</td>
</tr>
<tr>
<td>UPRT</td>
<td>uracil phosphoribosyltransferase</td>
</tr>
<tr>
<td>VSMCs</td>
<td>vascular smooth muscle cells</td>
</tr>
<tr>
<td>XIST</td>
<td>X inactive specific transcript</td>
</tr>
<tr>
<td>XPO5</td>
<td>exportin 5</td>
</tr>
<tr>
<td>YAP</td>
<td>yes-associated protein</td>
</tr>
<tr>
<td>ZEB</td>
<td>zinc-finger E-box-binding homeobox</td>
</tr>
<tr>
<td>α-MHC</td>
<td>α-myosin heavy chain</td>
</tr>
</tbody>
</table>
Chapter 1- Introduction

1.1 The discovery of RNA interference

A study conducted in 1995 showed that both anti-sense and sense RNA was capable of silencing the par-1 gene of the small nematode Caenorhabditis elegans, a Serine/Threonine protein kinase required for establishing polarity in the nematode worm embryos (Guo & Kemphues, 1995). This was a highly surprising result as the sense RNA should have simply served as a negative control. RNA interference was developed as a term to distinguish the phenomenon from antisense inhibition (Rocheleau et al., 1997).

Part of the mechanism of this inhibition was revealed three years later (Fire et al., 1998). The key discovery of this paper was to show that trace amounts of contaminating double-stranded RNA (dsRNA) was responsible for the inhibitory action of both the sense and anti-sense RNAs. Indeed the double-stranded component could inhibit at concentrations significantly lower than the target mRNA, indicating that an amplification step is involved in the process (Fire et al., 1998). Furthermore, Fire et al. demonstrated that dsRNA was at least tenfold more potent in producing a phenotype than either of the single strands on its own and the effect persisted in the next generation (Fire et al., 1998).

The discovery of the small RNA responsible for post transcriptional gene silencing (PTGS) in plants was reported a year later by David Baulcombe’s group (Hamilton & Baulcombe, 1999). PTGS was induced by transgenic expression as well as viral infection and it was found to target both cellular and viral mRNAs (Hamilton & Baulcombe, 1999). Critically, the implicated determinants of the specificity of PTGS were small single-stranded RNA (ssRNA) molecules of ~25 nucleotides, now known as small interfering RNAs (siRNAs) (Hamilton & Baulcombe, 1999).

Prior to this a form of post transcriptional gene silencing (PTGS) in response to transgene sequences was reported in plants and fungi (Romano & Macino, 1992; Napoli, Lemieux & Jorgensen, 1990). Experiments in petunias aiming to increase pigment production by introducing a transgene targeting chalcone synthase with the aim of overexpression actually led to a significant decrease in colour intensity, with >40% of the plants appearing white or
variegated (Napoli, Lemieux & Jorgensen, 1990). The term cosuppression was created, denoting silencing of both the endogenous chalcone synthase as well as the transgene sequences. Two years following this discovery a similar observation was made in the fungus *Neurospora crassa* whilst investigating pigment production in this organism (Romano & Macino, 1992).

Two groups recapitulated RNAi in cell extracts from *Drosophila melanogaster* using a biochemical approach (Hammond et al., 2000; Tuschi et al., 1999). Three steps of the RNAi mechanism were identified: cleavage of a long dsRNA into shorter fragments by an enzyme named Dicer, loading of ssRNA molecules into the RNA-induced silencing complex (RISC) and degradation of the mRNA targets by RISC (Hammond et al., 2000; Tuschi et al., 1999). In some organisms an RNA polymerase extends and amplifies the siRNA molecules, which is why trace amounts of double-stranded siRNA are so effective (Tsai et al., 2015; Pak & Fire, 2007; Agrawal et al., 2003; Sijen et al., 2001).

### 1.2 The discovery of microRNAs

The first identified miRNA, *lin-4*, was detected in *Caenorhabditis elegans* (Lee, Feinbaum & Ambros, 1993). In 1993 Lee et al. described two short transcripts of *lin-4* (22 and 61 nucleotides long) complementary to a repeated sequence element in the 3’ untranslated region (3’ UTR) of the *lin-14* mRNA and suggested that this complementarity is used to regulate the temporal expression of the LIN-14 protein (Lee, Feinbaum & Ambros, 1993). They showed that *lin-4* was a non-coding sequence and demonstrated that it was partially conserved across four *Caenorhabditis* species (Lee, Feinbaum & Ambros, 1993). In an adjoining article, Wightman et al. reported that *lin-14* had seven conserved sequences (between two *Caenorhabditis* species) complementary to *lin-4* located in its 3’UTR which was necessary and sufficient to induce the temporal expression of LIN-14 regulated by *lin-4* in an antisense RNA-RNA interaction (Wightman, Ha & Ruvkun, 1993).

Subsequently, the 21-nucleotide-long RNA encoded by the gene *let-7* was also shown to regulate cell fate during *Caenorhabditis elegans* development (Reinhart et al., 2000). Importantly, the *let-7* RNA was detected in a wide range of animal species: humans, fruit flies,
chickens, frogs, zebrafish, molluscs and sea urchins, and apart from being conserved at the sequence level, a conserved temporal expression was also observed (Pasquinelli et al., 2000).

Several studies showed the abundance, evolutionary conservation and expression patterns of similar small RNA molecules, referred to as microRNAs, in *Caenorhabditis elegans*, *Drosophila melanogaster*, mouse and human, which implied that this class of molecules has an extensive and conserved regulatory function (Lee & Ambros, 2001; Lau et al., 2001; Lagos-Quintana et al., 2001).

MiRNAs are involved in every cellular process, they play an essential role in development and disrupting their expression has been associated with various pathological phenotypes (Lu et al., 2008). The number of miRNAs has rapidly expanded since their initial discovery. In the latest release of miRBase (V21, June, 2014) there are 2588 and 1915 mature miRNAs annotated in human and mouse, respectively (Ambros et al., 2003).

The same pathway responsible for mediating RNAi also generates miRNAs and enables them to exert their function in silencing target genes (Mello & Conte, 2004). However, there are additional steps in the miRNA biogenesis pathway that occur in the nucleus which are described below.

1.3 Genomic organisation and biogenesis of miRNAs

1.3.1 Genomic organisation of miRNAs

MicroRNA genes can be located within other genes or in intergenic regions (Berezikov, 2011) (figure 1.1). Since the first study on the genomic organisation of miRNAs (Rodriguez et al., 2004) evidence has been building up that the majority of miRNAs are found within the introns of protein coding genes (Hinske et al., 2014; Meunier et al., 2013; Hinske et al., 2010; Baskerville & Bartel, 2005). A recent study estimated that over 50% of all miRNAs analysed in several species, including mouse and human, locate to introns (Meunier et al., 2013). Approximately 10% have been found to lie within exonic regions and the rest—within intergenic loci (Lin & Gregory, 2015; Ha & Kim, 2014). MiRNAs can exist as ‘solo’ miRNAs or as clusters of several miRNAs in introns or intergenic regions (Berezikov, 2011) (figure 1.1).
Figure 1. Genomic organisation of miRNAs. MiRNAs are encoded between genes (intergenic, miR-A-miR-D) as well as within introns and exons of genes (intrinsic and exonic, respectively, miR-E-miR-L). They exist in clusters in the case of intergenic and intrinsic miRNAs (miR-B-miR-D and miR-F-miR-H) and as solo miRNAs (miR-A, miR-E, miR-I-miR-L). The noncanonical mirtrons are intrinsic miRNAs bypassing cleavage by the Microprocessor and are classified as conventional (miR-I), 5′-tailed (miR-J) and 3′-tailed (miR-K). Reproduced with permission from (Sen, 2014).

MiRNAs processed in the canonical pathway are transcribed as primary miRNAs (pri-miRNAs), which in most cases is mediated by Polymerase II (Lee et al., 2004; Cai, Hagedorn & Cullen, 2004). Pri-miRNAs are long transcripts that can extend to several kilobases in length (Cai, Hagedorn & Cullen, 2004; Lee et al., 2004). A small group of miRNAs interspersed among Alu repeats has been found to be transcribed by Polymerase III (Borchert, Lanier & Davidson,
Approximately 80% of intragenic miRNAs have been estimated to share the same direction of transcription as their host genes (Hinske et al., 2014; Meunier et al., 2013; Campo-Paysaa et al., 2011; Rodriguez et al., 2004).

Pri-miRNAs share the same characteristics as class II gene transcripts, they have a 5’ cap and a poly(A) tail and are spliced (Lin & Gregory, 2015; Lee et al., 2004). They consist of a 33-35 bp stem, into which the mature miRNA is embedded, a terminal loop and ssRNA sequences flanking the stem on both the 5’ and 3’ sides (Ha & Kim, 2014; Lee et al., 2004). Most miRNA genes are under the control of their own promoters (see section 1.4.1.1) (Berezikov, 2011).

1.3.2 Nuclear processing and export of pre-miRNAs

Following transcription pri-miRNAs undergo several processing steps before they turn into the active mature form (Lee et al., 2002) (figure 1.2). Within the nucleus, an RNase III- type nuclease, DROSHA, cleaves pri-miRNAs ~11 bp away from the ss-dsRNA junction to generate small hairpin molecules called miRNA precursors (pre-miRNAs), which have 3’ overhangs of either 2 or 1 nucleotides and are ~55-70 nt long (Burke et al., 2014; Han et al., 2006; Zeng & Cullen, 2005; Lee et al., 2003). DROSHA forms a complex (the Microprocessor) with a dsRNA-binding protein called DiGeorge syndrome critical region 8 (DGCR8) (Denli et al., 2004; Gregory et al., 2004) (figure 1.2). DGCR8 interacts with pri-miRNAs via ssRNA sequences as well as the pri-miRNA stem and assists DROSHA to cleave them in the precise location (Han et al., 2004; Gregory et al., 2004; Denli et al., 2004).

After processing by the Microprocessor, pre-miRNAs are exported to the cytoplasm by exportin 5 (XPOS), a dsRNA-binding protein which recognises dsRNA stems>14 bp with a short 3’ overhang (Zeng & Cullen, 2005; Bohnsack, Czapliński & Gorlich, 2004; Lund et al., 2004; Yi et al., 2003; Gwizdek et al., 2003). XPOS forms a transport complex with the GTP-bound RAN nuclear protein and the pre-miRNA (figure 1.2). After the complex is transferred into the cytoplasm via the nuclear pore, GTP is hydrolysed to GDP, the complex is disassembled and
the pre-miRNA is released into the cytoplasm (Bohsack, Czaplinski & Gorlich, 2004; Yi et al., 2003).

Figure 1. 2 Biogenesis of canonical miRNAs. miRNAs are transcribed in the nucleus as long primary miRNAs (pri-miRNAs) by RNA polymerase II. They are subsequently cleaved by the Microprocessor complex composed of the RNaseIII nuclease DROSHA and its partner protein DiGeorge syndrome critical region 8 (DGCR8) which produces 60-70 nt-long miRNA precursor hairpins (pre-miRNAs). They are recognised and exported to the cytoplasm by exportin 5 (XPO5) where another RNase III nuclease, DICER, usually working with TRBP, cleaves them to generate short dsRNA duplexes. In most cases the ‘guide’ strand from the duplex is selected to be incorporated into the miRNA-induced silencing complex (miRISC) composed of DICER and AGO proteins and it guides the complex to target mRNAs harbouring complementary sequences to the miRNAs in their 3’ UTRs. Gene suppression occurs in P-bodies and it is exerted by either translational repression or mRNA degradation. Reproduced with permission from (Lin & Gregory, 2015).

1.3.3 Cytoplasmic processing of pre-miRNAs and target gene silencing

Another RNase III enzyme, DICER, cleaves pre-miRNAs to produce a short ~22 nt-long dsRNA duplex with 2-nt 3’ overhangs (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Knight & Bass, 2001). In mammals DICER requires the binding of
trans-activator (TAR)-RNA-binding protein (TRBP) to exert its function and it is essential for the formation of miRNA-induced silencing complex (miRISC), which mediates the function of miRNAs in repressing mRNA expression (Gregory et al., 2005; Chendrimada et al., 2005) (figure 1.2). Protein activator of PKR (PACT) has also been found to associate with DICER, but its function remains unknown (Ha & Kim, 2014).

TRBP bridges DICER with Argonaute proteins (AGO1-4) to form miRISC (Lin & Gregory, 2015; Chendrimada et al., 2005). There are two steps in the formation of miRISC: the loading of the double-stranded fragment generated by DICER into an AGO protein and the duplex unwinding (Kawamata & Tomari, 2010). At the latter step usually only one of the two strands is retained in miRISC (Lin & Gregory, 2015). Human AGO proteins do not show preference towards specific sets of miRNAs, but they tend to incorporate RNA duplexes with central mismatches (Yoda et al., 2010). Strand selection is mainly determined by the thermodynamic stability of the 5’ ends of the two strands, the less stable one usually being retained (the functional or ‘guide’ strand) and the more stable one (the star strand, miRNA*) – degraded (Khvorova, Reynolds & Jayasena, 2003). Unwinding of the ds duplex is facilitated by mismatches in the guide strand at positions 2-8 and 12-15 (Yoda et al., 2010; Kawamata, Seitz & Tomari, 2009). In some tissues ‘strand switching’, characterised by the preferential use of the miRNA* strand over the guide strand, has been observed (Chiang et al., 2010; Ro et al., 2007).

After strand selection occurs miRNAs mediate translational repression or decay of target mRNAs via RISC (Kawamata & Tomari, 2010). The process occurs by recruiting members of the GW182 protein family and takes place in processing bodies (P-bodies), cytoplasmic foci involved in mRNA turnover, which are not essential for gene silencing mediated by miRNAs (Eulalio et al., 2007; Gregory et al., 2005; Liu et al., 2005) (figure 1.2). More than half of all mammalian protein-coding genes have been estimated to be regulated by miRNAs (Friedman et al., 2009). In animals target recognition usually occurs with partial complementarity via the miRNA ‘seed’ region encompassing nucleotides 2-7 or 2-8 (Ameres & Zamore, 2013; Bartel, 2009). By contrast, most of the plant miRNAs are fully complementary to their targets, which induces the endonucleolytic cleavage of mRNAs by RISC (Iwakawa & Tomari, 2015; Ameres & Zamore, 2013; Bartel, 2009; Jones-Rhoades, Bartel & Bartel, 2006). Animal RISC silences target genes by recruiting additional proteins rather than by direct cleavage of mRNAs (Iwakawa & Tomari, 2015). Typically a miRNA recognises a region in the 3’ untranslated
regions (3’ UTR) of the target mRNA (Bartel, 2009). The mechanism of inducing target mRNA decay has been shown to occur via the recruitment of several factors by the GW182 protein (Kuzuoglu-Ozturk et al., 2012; Fabian et al., 2011; Braun et al., 2011; Fabian et al., 2009). Poly(A)-binding protein (PABP), CCR4-NOT and PAN2-PAN3 deadenylase complexes trigger mRNA deadenylation and decay mediated by the 5’-to-3’ decay pathway (Iwakawa & Tomari, 2015). It has been shown that RISC itself can induce 5’ mRNA decapping by recruiting decapping factors, which can also occur in a deadenylation-independent manner (Nishihara et al., 2013). Translation has been shown to be repressed at the initiation step via several different mechanisms, some of which involve the recruitment of repressors by GW182 (Iwakawa & Tomari, 2015).

Most miRNAs control mRNA expression in the cytoplasm (Filipowicz, Bhattacharyya & Sonenberg, 2008). However, studies have also reported that miRNAs are present at abundant levels in the nucleus (Liao et al., 2010; Kim et al., 2008; Place et al., 2008; Hwang, Wentzel & Mendell, 2007). Furthermore, it has been demonstrated that they regulate mRNA processing or the expression of other miRNAs including their own in the nucleus (Tang et al., 2012; Zisoulis et al., 2012; Meister et al., 2004). Importin 8 (IPO8), a member of the karyopherin β family of receptors like XPO5, has been shown to mediate the cytoplasm-to-nucleus transport of mature miRNAs in an Argonaute 2 (AGO2)-dependent manner (Wei et al., 2014).

1.3.4 Non-canonical miRNAs

A subclass of miRNAs called ‘mirtrons’ bypasses the Microprocessor cleavage step and uses the cell’s splicing machinery instead (Berezikov et al., 2007). Upon resolving the branch point of the lariat-shaped intermediate fragment generated via splicing, either conventional (where both hairpin ends are generated by splicing without long overhangs), 3’, 5’ or two-tailed (3’ and 5’) mirtron hairpins are generated (Wen et al., 2015) (figure 1.1, mir-1-mir-K). Tailed mirtrons require exonucleolytic trimming after debranching, which ensures that the overhanging ssRNA fragments are removed and generates the pre-miRNA hairpin (Ladewig et al., 2012; Flynt et al., 2010). Other non-canonical pathways of pre-miRNA generation include, but are not limited to, DICER-independent biogenesis with one known example of pri-miR-451 which is sequentially cleaved by AGO2 and the polyA-specific endonuclease (PARN), small
nucleolar RNA (snoRNA) and transfer RNA (tRNA)-derived miRNAs and a newly discovered class of miRNAs called agotrons, which escape both DROSHA and DICER processing (Daugaard & Hansen, 2017; Babiarz et al., 2008; Ender et al., 2008).

1.4 Regulation of miRNA biogenesis

1.4.1 Transcriptional control

1.4.1.1 Regulation at the promoters of pri-miRNAs

Although it was initially thought that intronic miRNAs are under the control of the same promoter driving host mRNA transcription (Baskerville & Bartel, 2005; Bartel, 2004), there is accumulating evidence that miRNAs have independent promoters, which has been estimated to be the case for at least 50% of these miRNAs (Marsico et al., 2013; Ozsolak et al., 2008). It has been suggested that young miRNAs tend to emerge within old genes and benefit from the regulatory elements already in place with the potential to acquire independent regulation with time (França et al., 2017; França, Vibranovski & Galante, 2016). Published reports have found that clustered miRNAs are usually transcribed as a single polycitronic unit, but they can be processed individually to ensure a developmental and tissue-specific expression (Ha & Kim, 2014; Lee et al., 2002). However, a recent study found that clustered pri-miRNAs are not always cotranscribed, even if the clusters are highly conserved (Chang et al., 2015). Chang et al. found multiple examples of transcripts encoding only a subset of the expected clustered pri-miRNAs (Chang et al., 2015). Alternative promoter usage was recently found to be a common mechanism of transcriptional regulation of miRNAs and it was shown to occur in a cell-type-specific manner (Chang et al., 2015).

Transcriptional activators and repressors can positively or negatively regulate the expression of pri-miRNAs. The tumour-suppressor transcription factor p53 induces the transcription of mir-34 family members which target growth-promoting genes, thus enhancing cell cycle arrest and apoptosis (Raver-Shapira et al., 2007; Tarasov et al., 2007). The oncogenic protein MYC acts both as an activator and repressor of miRNA transcription in different types of cancer (Lin & Gregory, 2015). It stimulates the expression of the oncogenic mir-17-92 cluster in lymphoma cells and mir-9 in neuroblastoma cells, which induces cancer progression.
because these miRNAs target mRNAs involved in cell cycle regulation and angiogenesis (Dews et al., 2006; O'donnell et al., 2005). By contrast, in B cell lymphoma MYC inhibits the expression of tumour-suppressor miRNAs like mir-15a (Chang et al., 2008). Myogenin and myoblast determination protein 1 (MYOD1) bind in regions upstream of the highly-conserved muscle-specific miRNAs mir-1 and mir-133 to induce their expression during myogenesis (Rao et al., 2006; Chen et al., 2006).

Transcriptional regulation of miRNAs can also be established by epigenetic modifications of DNA and histone proteins (Lin & Gregory, 2015). The mir-203 locus has been shown to be hypermethylated, thus leading to a downregulation of miRNA expression in T-cell lymphomas in contrast with non-malignant T lymphocytes (Bueno et al., 2008). Furthermore, promoter hypermethylation of tumour-suppressor miRNAs has been found to be reversed and the miRNA expression re-activated upon pharmacological treatment with DNA-demethylating agents, resulting in a reduction of tumour growth and metastasis (Lujambio et al., 2008). Another example of epigenetically silenced miRNAs is the mir-124 family which was initially found to be dysregulated in colorectal cancer cells. Its perturbed expression was subsequently reported to occur via the same mechanism in other types of cancer like hematopoietic malignancies, gastric, liver, breast and cervical cancer (Suzuki et al., 2011b; Kunej et al., 2011; Lujambio et al., 2007). Histone modifications have also been shown to impact miRNA expression by acting in concert with DNA methylation, as is the case of mir-223 in leukaemias (Guil & Esteller, 2009).

MiRNAs can be transcribed in the antisense direction which leads to the production of entirely different mature sequences and changes in mRNA targets in both vertebrates and invertebrates (Tyler et al., 2008). In Drosophila melanogaster the iab-4 locus was reported to be transcribed in both directions resulting in the generation of the sense iab-4 and the
antisense iab-8. Both transcripts are functional, they regulate Hox genes and their ectopic expression leads to homeotic transformations (Tyler et al., 2008; Stark et al., 2008).

1.4.1.2 Regulation of miRNA expression via feedback loops

Autoregulatory feedback loops between miRNAs and mRNAs occur frequently during cell fate determination and development (Krol, Loedige & Filipowicz, 2010) (figure 1.3). In unilateral feedback loops a miRNA negatively regulates the expression of transcription factors whilst the latter induce the expression of the miRNA (Arora et al., 2013). For example, dopaminergic neuron differentiation in mice is regulated by a unilateral feedback loop between miR-133b and the transcription factor Paired Like Homeodomain 3 (PITX3) (Kim et al., 2007). PITX3 stimulates the expression of mir-133b, whereas miR-133b inhibits PITX3 in midbrain dopaminergic neurons (Kim et al., 2007) (figure 1.3). Reciprocal negative feedback loops are characterised by the mutual repression of the miRNA and the transcription factor targeted by it (Krol, Loedige & Filipowicz, 2010). Such regulation occurs between the miR-200 family of miRNAs and the zinc-finger E-box-binding homeobox (ZEB) transcription factors ZEB1 and ZEB2 which promote epithelial-to-mesenchymal transition (EMT) (Gregory et al., 2008). In tumours ZEB1 and ZEB2 bind to mir-200 promoter and promote EMT by suppressing the expression of this miRNA family (Bracken et al., 2008; Gregory et al., 2008). Another example of a reciprocal negative feedback loop is the one between miR-7 and Yan, which is observed in Drosophila melanogaster during the differentiation of photoreceptors (Li & Carthew, 2005) (figure 1.3).
**Figure 1.** 3 Autoregulatory feedback loops coordinate the expression of transcription factors (TFs) and miRNAs in development. In unilateral feedback loops the miRNA expression is stimulated by the TF, whereas the TF is repressed by the miRNA. Mutual repression occurs between TFs and miRNAs in reciprocal negative feedback loops. Double-negative feedback loops involve mRNA regulation by miRNAs which themselves are under TF control. Reproduced with permission from (Krol, Loedige & Filipowicz, 2010).

A more sophisticated double-negative feedback loop regulates left-right asymmetry of ASE chemosensory neurons in *C. elegans* (ASE left (ASEL) and right (ASER)) (Johnston et al., 2005). In ASER the Conserved oligomeric Golgi complex subunit 1 (COG-1) transcription factor stimulates the expression of miR-273, which in turn represses Dorsal Intercalation and Elongation defect 1 (*die-1*) expression (Johnston et al., 2005). By contrast, in ASEL *DIE-1* which
avoids targeting by miR-273 activates lys-6. The latter acts to suppress cog-1 and thus promotes ASEL cell fate (Johnston et al., 2005) (figure 1.3).

1.4.1.3 Regulation of miRNA expression via feed-forward loops

Feed-forward loops (FFLs) are another example of a mechanism for the transcriptional regulation of miRNAs (Arora et al., 2013) (figure 1.4). They involve miRNAs and transcription factors, which through their mutual regulation, jointly control the expression of another gene (Arora et al., 2013). In miRNA-mediated FFLs a miRNA represses both the transcription factor and target genes (figure 1.4(a)). This has been reported for miR-208 which represses GATA-4 and subsequently its target α-myosin heavy chain (α-MHC), thus inducing cardiac hypertrophy (Zhou, He & Pu, 2012). Transcription factor-mediated FFLs where miRNAs and target genes are simultaneously repressed or activated by a transcription factor result in incoherent effects on target gene expression (Arora et al., 2013) (figure 1.4(b1 and b2)). This is the case for a c-MYC-driven FFL where c-MYC activates mir-17 and mir-20a as well as E2F1. This ensures the tight control of E2F1 whose transcription is enhanced, but simultaneously its translation is indirectly limited (O’donnell et al., 2005). Transcription factor-mediated FFLs induce coherent effects on target genes when they have an opposing effect on miRNAs and their targets (Arora et al., 2013) (figure 1.4(b3 and b4)). This type of coherent loop has been observed in the heart. Transforming growth factor beta (TGF-β) positively regulates the transcription of fibrotic genes in cardiac fibroblasts upon myocardial infarction, but it represses miR-29 expression, which normally targets them, resulting in an enhanced fibrotic response (van Rooij et al., 2008). A bi-directional composite FFL is established when a miRNA acts to suppress both the transcription factor and its target genes, but the transcription factor enhances the transcription of both the miRNA and its target genes (Martinez & Walhout, 2009)(figure 1.4(c)). Such an example has been found in the relationship between MYC, the miR-17-92 cluster and the target E2F in cancer cell cycle regulation (Aguda et al., 2008).
Figure 1. 4 Transcriptional feed-forward loops (FFLs) consist of miRNAs and TFs regulating each other whilst each of them controls target gene expression. In miRNA-mediated FFLs the miRNA simultaneously represses TFs and target genes (a). Incoherent effects on target genes are observed in TF-mediated FFLs where the TF has the same effect (repressing or activating) on miRNAs and target genes (b1 and b2). Coherent effects on target gene expression occur when in TF-mediated FFLs the TF has opposite effect on miRNA and target gene expression (b3 and b4). In composite FFLs miRNAs negatively regulate TFs and target genes, whereas TFs induce miRNA and target gene expression (c). Reproduced with permission from (Arora et al., 2013).
1.4.2  Post transcriptional regulation of miRNA biogenesis

Post transcriptional regulation can occur at any stage of the biogenesis of miRNA genes following their transcription and it ensures cell, tissue and developmental stage-specific miRNA expression (Siomi & Siomi, 2010).

1.4.2.1 Regulation of miRNA expression at the level of the Microprocessor

An autoregulatory feedback loop has been reported between DROSHA and its partner DGCR8 (Han et al., 2009; Yeom et al., 2006). The Microprocessor cleaves hairpin structures in the \textit{DGCR8} mRNA, leading to its degradation, whereas DGCR8 stabilises DROSHA via protein-protein interactions (Triboulet et al., 2009; Han et al., 2009; Yeom et al., 2006). Fine-tuning the ratio between the components of the Microprocessor might be essential since having three times more DGCR8 than DROSHA was found to impair the catalytic activity of the latter \textit{in vitro} (Gregory et al., 2004). A study in which a \textit{DROSHA} knockout was performed showed that it is essential for the biogenesis of miRNAs in the canonical pathway with a 96.5% reduction in miRNA levels observed upon high-throughput sequencing analysis (Kim, Kim & Kim, 2016).

Post translational modifications have been reported to affect the activity of the Microprocessor. DROSHA phosphorylation at Ser 300 and 302 mediated by glycogen synthase kinase 3 Beta (GSK3B) is required for its nuclear localisation (Tang et al., 2011; Tang et al., 2010). Phosphorylation of DGCR8 by ERK was found to increase its stability and raise the Microprocessor levels (Herbert et al., 2013). Stabilisation of DROSHA was observed through acetylation by several acetyl transferases, which ensured its protection against degradation by the ubiquitin-proteasome pathway and ultimately was demonstrated to lead to an increase in the level of DROSHA protein (Tang et al., 2013). Deacetylation of critical lysine residues in the DGCR8 RNA-binding domains by histone deacetylase 1 (HDAC1) has been shown to increase the affinity of DGCR8 for pri-miRNAs (Wada, Kikuchi & Furukawa, 2012).

The DEAD box RNA helicases p68 (DDX5) and p72 (DDX17) were found to be associated with DROSHA and siRNA knockdown of these proteins resulted in a small decrease in mature miRNA production (Link, Grund & Diederichs, 2016; Gregory et al., 2004). Several studies have
reported an interaction between p68/p72 and accessory proteins, altering the expression of miRNAs. Transforming growth factor β (TGFβ) and bone morphogenetic protein 4 (BMP4) treatment of vascular smooth muscle cells (VSMCs) has been found to increase the contractile phenotype of these cells by increasing the processing rate of pri-miR-21 into pre-miR-21 which targets programmed cell death 4 (PDCD4), a negative regulator of contractile genes (Davis et al., 2008). It was established that SMAD proteins which accumulate upon TGFβ and BMP treatment interact with p68 and are recruited to the pri-miR-21 transcript (Davis et al., 2008). SMADs were found to bind a conserved sequence in pri-miRNA transcripts and when mutated, it resulted in a disruption in pre-miRNA accumulation and an impaired binding of the Microprocessor, implying that SMADs provide a platform for Microprocessor docking and are responsible for the more efficient cleavage by DROSHA (Davis et al., 2010). Apart from a role in transcriptional regulation of miRNA expression the tumour suppressor p53 can modulate pri-miRNA processing (Suzuki et al., 2009). The processing of several pri-miRNAs was enhanced through an interaction between p53 and the DROSHA complex mediated by p68 upon a DNA damage response in human colon cancer cells (Suzuki et al., 2009). Inactive mutant p53 was found to decrease the interaction between DROSHA and p68 and to interfere with pri-miRNA processing (Suzuki et al., 2009). However, recently it has been demonstrated that mutant p53 affects the DROSHA-p72 binding and not the assembly of DROSHA-p68 in contrast to findings from the previous study (Garibaldi et al., 2016). The tumour suppressor BRCA1 has also been found to associate with DROSHA and p68 as well as pri-miRNAs, enhancing the expression of pre- and mature miRNAs (Kawai & Amano, 2012).

High cell density has been shown to increase the global production of miRNAs, which was associated with an increased pri-miRNA processing efficiency by DROSHA as well as an enhanced incorporation of miRNAs in RISC (Hwang, Wentzel & Mendell, 2009). Studies in cancer cells established a globally decreased miRNA expression, which was further discovered to be caused by perturbations in the HIPPO-YAP pathway (Mori et al., 2014; Lu et al., 2005). Yes-associated protein (YAP), a downstream effector of the HIPPO signalling pathway, was found to sequester p72 in the nucleus at low cell density, thus repressing Microprocessor formation (Mori et al., 2014). At high cell density YAP was inactivated, which allowed binding of p72 to the Microprocessor and to an enhanced miRNA production (Mori et al., 2014).
Proteins with known function in splicing have been shown to bind directly to terminal loops of miRNAs and to regulate their processing. The KH-type splicing regulatory protein (KHSRP) binds to conserved terminal loop regions of pri-miRNAs, amongst which is pri-let-7a, and enhances their processing (Trabucchi et al., 2009). KHSRP was shown to be part of both DROSHA and DICER complexes (Trabucchi et al., 2009). Another protein implicated in RNA processing, heterogeneous nuclear ribonucleoprotein (hnRNP) A1, has been reported to bind to the terminal loop of pri-miR-18a and to induce relaxation of the stem region, thus activating the Microprocessor by creating a more favourable conformation for DROSHA cleavage (Michlewski et al., 2008; Guil & Cáceres, 2007). In addition to its role as processing activator, hnRNP A1 competes with KHSRP for binding the terminal loop of pri-let-7a-1 in somatic cell lines to inhibit its processing by DROSHA (Michlewski & Cáceres, 2010). Recently p72 and KHSRP have been shown to regulate the level of unloaded AGO2 in HeLa cells by regulating global miRNA levels (Connerty et al., 2016).

Another protein involved in the regulation of splicing, pre-mRNA-splicing factor SF2/alternative splicing factor (SF2/ASF), has been reported to bind to pri-miR-7 and induce DROSHA processing (Wu et al., 2010). Mature miR-7 was shown to target SF2/ASF mRNA, thus forming a negative feedback loop leading to a reduction in SF2/ASF expression (Wu et al., 2010).

1.4.2.2 Regulation of miRNA expression via nuclear export

Knockdown of XPO5, which mediates nuclear export of pri-miRNAs resulted in a reduction in mature miRNA levels, but not to an accumulation of pri- or pre-miRNAs, which implied that XPO5 has a role in stabilising pri-miRNA transcripts (Lund et al., 2004; Yi et al., 2003). Overexpression of XPO5 led to an increased production of mature miRNAs, which further supported that nuclear export mediated by XPOS is a rate-limiting step in miRNA processing (Yi et al., 2005). Furthermore, a genetic defect affecting the C-terminal domain of XPO5 in cancer cell lines has been shown to impair miRNA processing and to result in reduced mature miRNA levels (Melo et al., 2010). However, recently Kim et al. reanalysed the microarray data from the Melo et al. study and found that significant numbers of mature miRNAs were still expressed in these cells (Kim, Kim & Kim, 2016). Moreover, the reduction in global mature
miRNA levels in XPO5 knockout cells was only found to be modest which suggested that even though XPO5 is involved in the miRNA biogenesis pathway, it is not an essential factor in it (Kim, Kim & Kim, 2016).

1.4.2.3 Regulation of miRNA expression at the level of DICER cropping

Post transcriptional control of miRNA biogenesis can be achieved at the level of DICER processing. DICER negatively regulates its own catalytic activity through its N-terminal domain which is homologous to DExD/H-box helicases (Ma et al., 2008). An enhanced activity of DICER cleavage was observed upon deletion or mutation in this domain (Ma et al., 2008). A marked reduction in mature miRNA levels was observed upon knocking DICER out in the colorectal carcinoma HCT116 cell line (Kim, Kim & Kim, 2016). Even though the effect of ablating DICER was not as big as in DROSHA knockout cells, DICER was found to significantly contribute to the biogenesis of mature miRNAs (Kim, Kim & Kim, 2016).

One of the two DICER partner proteins in humans, TRBP, can confer changes in the regulation of pre-miRNAs (Ha & Kim, 2014). A decrease in TRBP levels leads to a reduction in pre-miRNA levels as well as DICER destabilisation, which also occurs in TRBP mutations that are reported in cancer (Krol, Loedige & Filipowicz, 2010; Melo et al., 2009). Furthermore, TRBP can alter the position of DICER cleavage, resulting in the generation of miRNA isomiRs, which differ from canonical miRNAs by a small number of bases at the 5’ and/or 3’ ends in the mature sequence (Lee & Doudna, 2012; Fukunaga et al., 2012). TRBP is stabilised by serine phosphorylation by MAPK/ERK (Paroo et al., 2009). This leads to an increase in growth-promoting miRNAs and a decrease in the expression of let-7a, a suppressor of cell proliferation (Paroo et al., 2009). Let-7a has been shown to be able to target DICER mRNA and this feedback loop has been suggested to contribute to the regulation of DICER activity (Forman, Legesse-Miller & Coller, 2008).

KSRP can also act at the level of pre-miRNA processing by binding to miRNA terminal loops and promoting DICER cleavage (Trabucchi et al., 2009). LIN28 RNA-binding proteins act at both the DROSHA and DICER levels of miRNA biogenesis by binding to let-7 miRNA terminal loops and inhibiting their processing via two distinct mechanisms in embryonic stem cells (Viswanathan & Daley, 2010) (see section 1.8.2). DICER activity has also been shown to be
altered by methylation of pre-miRNAs. The RNA methyltransferase BCDIN3D was found to methylate pre-miRNAs at the 5’ monophosphate \textit{in vivo} and \textit{in vitro}, which impairs DICER activity since it normally interacts with the 5’-terminal phosphate (Xhemalce, Robson & Kouzarides, 2012).

The MCPIP1 endoribonuclease has been reported to cleave terminal loops of pre-miRNAs and thus suppresses their biogenesis by impairing DICER activity (Suzuki et al., 2011a). An endoplasmic reticulum (ER) transmembrane kinase-endoribonuclease, IRE1α which is active during conditions of ER stress was found to antagonise DICER processing by cleaving pre-miRNAs at sites distinct from the ones used by DICER (Upton et al., 2012). During hypoxia AGO2 is phosphorylated at Tyr393 by epidermal growth factor receptor (EGFR), which reduces the amount of DICER-bound AGO2 and leads to a decrease in pre-miRNA processing (Shen et al., 2013).

1.4.2.4 Regulation of miRNAs by altering them at the level of nucleotide sequences

RNA editing of miRNAs has been shown to interfere with both DROSHA and DICER processing of miRNA precursors and alter the targets of mature miRNAs (Kawahara et al., 2007a; Yang et al., 2006). It is carried out by adenosine deaminases acting on RNA (ADARs) which convert adenosine to inosine in dsRNA segments, resulting in an Inosine-Uracil wobble base pairing (Bass & Weintraub, 1988). ADAR editing in the stem region of pri-miR-142 has been reported to result in impaired processing by DROSHA and its degradation by Tudor Staphylococcal Nuclease (TUDOR SN), a ribonuclease specifically cleaving dsRNAs containing I-U base pairs (Yang et al., 2006). Pri-miR-151 edited by ADARs results in reduced processing by DICER due to an impaired interaction with the DICER-TRBP complex (Kawahara et al., 2007a). ADAR editing can also positively impact pri-miRNA processing and lead to an accumulation of pre-miRNAs (Kawahara et al., 2008). Editing of pri-miRNAs within the ‘seed’ region has been reported to occur in a tissue-specific manner and to result in changes in the set of target mRNAs for the miR-376 family of miRNAs (Kawahara et al., 2007b).

Similarly to RNA editing a change in the primary miRNA sequence caused by a single nucleotide polymorphism (SNP) in pri-miR-16-1 has been reported to impair its processing by DROSHA in chronic lymphocytic leukaemia (CLL) patients (Calin et al., 2005). Furthermore, a
recent study showed that a single nucleotide substitution in the terminal loop of pri-miR-30c-1 altered its structure and enhanced its processing by DROSHA in breast cancer via an increased interaction with Serine and Arginine Rich Splicing Factor 3 (SRSF3), which facilitates pri-miRNA recognition and processing (Fernandez et al., 2016).

Post transcriptional regulation accounts for the production of isomiRs, miRNA variants which are derived from the same primary miRNA transcript, but differ slightly because of additions or deletions of bases at the 5’ and/or the 3’ ends or modifications changing the internal sequence of miRNAs (Neilsen, Goodall & Bracken, 2012). Different strategies for the generation of isomiRs have been reported. Amongst them are imprecise cleavage by DROSHA and/or DICER, 5’ and/or 3’ trimming, 3’ tailing and ADAR editing (Neilsen, Goodall & Bracken, 2012).

1.5 Embryonic stem cells- properties and uses

Embryonic stem cells (ESCs) were first isolated and cultured from mouse embryos at the blastocyst stage of development (Evans & Kaufman, 1981; Martin, 1981). The first stable human ESC line was derived from blastocysts in 1998 (Thomson et al., 1998) and since then the number of different ESC lines has vastly expanded, reaching >1000 in 2009 (Löser et al., 2010).

ESCs are characterised by their ability to proliferate indefinitely in culture whilst maintaining their undifferentiated state (Martin, 1981; Evans & Kaufman, 1981). They are also pluripotent, which means they have the capacity to differentiate into every type of cell of the three embryonic germ layers- ectoderm, mesoderm and endoderm (Thomson et al., 1998). In culture, hESCs maintain a stable karyotype and expression of markers characteristic of undifferentiated cells over long periods of time as well as high telomerase activity, indicative of their ability to continually undergo mitosis (Caisander et al., 2006; Heins et al., 2004).

The unique properties of ESCs make them a useful tool for studying fundamental processes of development and they hold great potential for regenerative medicine therapies (Mahla, 2016; Murry & Keller, 2008). Furthermore, the discovery of adult stem cells (ASCs) and cancer stem cells (CSCs) expanded the relevance of elucidating the mechanisms maintaining ESCs in
the pluripotent state as well as the signalling pathways leading to their differentiation (Rumman, Dhawan & Kassem, 2015; Kreso & Dick, 2014). ASCs play an essential role in maintaining tissue homeostasis and regeneration following injuries and CSCs can propagate tumours upon serial transplantation (Rumman, Dhawan & Kassem, 2015; Kreso & Dick, 2014). ASCs and CSCs have properties resembling ESCs- both cell types have a self-renewal capacity and are multipotent, meaning that they can differentiate into a more limited number of cell fates compared to ESCs (White & Lowry, 2015; Rumman, Dhawan & Kassem, 2015; Nguyen et al., 2012; Clarke et al., 2006). Another area of research where understanding the biology of ESCs is important is the generation of induced pluripotent stem cells (iPSCs) from fully differentiated somatic cells. iPSCs are highly similar to ESCs in morphology, proliferation rate, telomerase activity, and their ability to differentiate into cells from all three germ layers, but they retain the genetic characteristics of individual patients (Takahashi et al., 2007). They offer the advantages of circumventing ethical and immune rejection concerns associated with the use of ESCs in cell and regenerative therapies (Zhan et al., 2004). In addition, iPSCs have provided the possibility of studying the pathophysiology of diseases in vitro and performing drug screens as well as creating personalised therapies for autologous stem cell transplantation (Neofytou et al., 2015; Ross & Akimov, 2014).

1.6 The pluripotent state of ESCs

Depending on the cell culture medium conditions in which they are grown mouse ESCs exhibit properties of cells derived from different stages of epiblast development. In the case when two specific kinase inhibitors are present (a condition known as 2i) and no serum is added, the cells are in a pre-implantation state (naïve pluripotency) and grow slowly (ESCs) (Ying et al., 2008). If serum is added to the growth medium, stem cells become similar to early post-implantation cells from the epiblast, they are in a ‘primed’ state of pluripotency and have a much faster growth rate (Percharde, Bulut-Karslioglu & Ramalho-Santos, 2016; Kolodziejczyk et al., 2015; Nichols & Smith, 2009). They are called epiblast stem cells (EpiSCs). EpiSCs have the same differentiation capacity as ESCs and they also express pluripotency markers, but they are inefficient in forming chimeras, express low-level differentiation-specific factors and have undergone X-chromosome inactivation (Martello & Smith, 2014; Brons et al., 2007).
Human ESCs are thought to be in the primed state of pluripotency because they show requirements for the same growth factors, they also grow in flat colonies, survive poorly at the single-cell level and most of the female cell lines have undergone X-inactivation (Li & Belmonte, 2017). Recently, several groups have demonstrated the successful culture of naïve human pluripotent stem cells (Ware et al., 2014; Gafni et al., 2013; Hanna et al., 2010).

The pluripotent state of embryonic stem cells is maintained by preventing differentiation and promoting self-renewal. This is ensured by various mechanisms working at the genetic, cell signalling and epigenetic levels (Li & Belmonte, 2017; Marks et al., 2012). A detailed review of these networks is beyond the scope of this chapter. Instead, a brief overview of the main genetic circuit as well as an outline of the involvement of miRNAs (part of the epigenetic regulation, see section 1.8.1) in the maintenance of self-renewal are presented.

Human and mouse ESCs share a ‘core’ transcriptional circuit whose activity is the main determinant of the pluripotent state (Young, 2011). It is composed of three main transcription factors, OCT4, SOX2 and NANOG, which initiate a myriad of regulatory events involving numerous other transcription factors and cofactors that work together to maintain the undifferentiated state of ESCs (Huang et al., 2015; Loh et al., 2006; Boyer et al., 2005). The three core transcription factors also bind their own promoters, thus forming an autoregulatory feedback loop that controls their own expression (Li & Belmonte, 2017).

The foundation of pluripotency resides in the regulation of OCT4 and SOX2: loss of either of them leads to differentiation into trophectoderm, overexpression of OCT4 results in a mesendoderm phenotype, whereas the outcome of SOX2 overexpression is differentiation into neural ectoderm (Thomson et al., 2011; Niwa, Miyazaki & Smith, 2000). Furthermore, OCT4 and SOX2 form heterodimers and work cooperatively to bind and activate the expression of genes involved in the maintenance of pluripotency and repress genes determining lineage-specific cell fate, whilst keeping them in a ‘poised’ state, ready for expression during differentiation (Chew et al., 2005; Rodda et al., 2005; Boyer et al., 2005). Therefore, core transcription factor-bound genomic sites are a platform for the recruitment of components which help integrate the information from signalling pathways, transcriptional networks, co-activators, co-repressors, epigenetic regulators and regulatory RNAs (Hackett & Surani, 2014; Young, 2011). Retroviral transduction of human dermal fibroblasts with OCT4, SOX2, KLF4 and c-MYC led to the successful generation of iPSCs (Takahashi et al., 2007).
NANOG binding sites in the genome extensively overlap with OCT4:SOX2 occupancy (Chen et al., 2008). However, cells lacking NANOG expression can still maintain an undifferentiated state as well as the expression of genes, which are normally bound by NANOG, including pluripotency-associated factors (Silva & Smith, 2008; Chambers et al., 2007).

1.7 Transcription in ESCs

Studies on undifferentiated cells performed in the 1940s demonstrated that they have higher levels of RNA compared to mature cell types and suggested that this was due to chromatin ‘hyperfunction’ (Thorell, 1947; Brachet, 1947; Caspersson & Thorell, 1941). Hypertranscription is characterised by a global increase in nascent RNA output, often linked to rapid cell proliferation and an open chromatin state (Percharde, Bulut-Karslioglu & Ramalho-Santos, 2016).

The genome of mESCs grown in the presence of serum has been shown to be in a transcriptionally hyperactive state (Efroni et al., 2008) (also see chapter 5, Introduction). The mechanisms underlying this condition have not been extensively studied, but several factors have been implicated in its maintenance.

Shifting the growth environment of mESCs from 2i to serum-supplemented medium resulted in an upregulation of Myc family transcription factors and led to a nearly 80% increase in global transcription (Kolodziejczyk et al., 2015; Marks et al., 2012; Ying et al., 2008). MYC co-localises with Polymerase II at gene promoters (Nie et al., 2012). Furthermore, MYC has been shown to be a universal amplifier of promoter-activated genes and overexpression of a Myc dominant negative mutant resulted in the inhibition of self-renewal and the induction of differentiation of ESCs in the presence of serum (Nie et al., 2012; Cartwright et al., 2005).

ESCs are characterised by the presence of more permissive, open chromatin (euchromatin), which becomes more condensed as cells differentiate (heterochromatin) (Gaspar-Maia et al., 2011; Meshorer et al., 2006). Chromodomain helicase DNA binding protein 1 (CHD1), a chromatin remodelling factor, has been shown to be essential for the maintenance of open chromatin in mouse ESCs (Guzman-Ayala et al., 2015). The genome-wide location of CHD1 correlates with Polymerase II as well as with the activating histone mark H3K4me3 (Gaspar-
Maia et al., 2009). ESCs which do not express Chd1 have a higher content of heterochromatin and are no longer pluripotent (Guzman-Ayala et al., 2015). Furthermore, Chd1/− cells show a ~25%-30% global reduction in their transcriptional output by both Polymerase I and II, which suggested that CHD1 is required to maintain the elevated transcription in the highly proliferating ESCs (Guzman-Ayala et al., 2015).

1.8 MicroRNAs and ESCs

1.8.1 MiRNAs are essential for maintaining self-renewal of ESCs and control pluripotency and differentiation

A lack of or defects in differentiation are observed in mouse ESCs (mESCs) upon knocking out Dicer or Dgcr8, respectively, and exposing the cells to differentiating conditions (Wang et al., 2007; Kanellopoulou et al., 2005). These mutants also have a reduced proliferation rate and an increased fraction arrest in the G1 phase of the cell cycle compared to wild type cells (Wang et al., 2007). MiRNAs are thus essential for the proper differentiation of ESCs and are involved in maintaining their proliferation and cell cycle structure (Greve, Judson & Blelloch, 2013).

ESCs express a characteristic profile of miRNAs, some of which are unique to stem cells (Greve, Judson & Blelloch, 2013). Mouse ESCs uniquely express the miRNA cluster miR-290-295 and human ESCs uniquely express the miR-302-367 cluster (Leonardo et al., 2012). In both species these clusters of miRNAs contribute strongly to the total miRNA pool and many of the mouse and human miRNAs of the two clusters share the same seed sequence AAGUGC (Marson et al., 2008; Laurent et al., 2008). ESC-specific miRNAs were found to be repressed upon differentiation, which further supports their role as important regulators of the stem cell state (Houbaviy, Murray & Sharp, 2003).

OCT4, SOX2 and NANOG have been shown to co-occupy the promoters of miRNAs which are preferentially or uniquely expressed in ESCs, thus connecting the core transcriptional circuitry with miRNAs (Marson et al., 2008). Members of the miR-209-295 cluster in mESCs and miR-302 in hESCs have been shown to control cell cycle progression by repressing key G1-phase regulators and promoting the G1-S transition, which ensures their rapid proliferation (Wang
et al., 2008; Card et al., 2008). Furthermore, members of these clusters have successfully been used to reprogram somatic or cancer cells into iPSCs (Kuo et al., 2012; Judson et al., 2009).

MiRNAs are also important regulators of lineage commitment in ESCs. Certain miRNAs become upregulated upon ESC responses to differentiation signals and they have been shown to repress members of the core transcriptional circuitry maintaining pluripotency (Xu et al., 2009; Tay et al., 2008). Members of the let-7 family of miRNAs oppose the action of miR-290-295 and miR-302 clusters and their repression is considered crucial for the maintenance of the pluripotent state in ESCs (Melton, Judson & Blelloch, 2010). They are highly expressed in somatic cells and are required for differentiation to proceed.

1.8.2 Post transcriptional regulation of miRNAs in ESCs

The developmentally regulated RNA binding protein LIN28 is the best studied miRNA processing inhibitor. It has two isoforms (LIN28A and LIN28B) which block the biogenesis of let-7 family members in ESCs (Thornton & Gregory, 2012). The promoter of Lin28 has been shown to be occupied by OCT4, SOX2 and NANOG and its levels decline during differentiation (Marson et al., 2008). This in turn allows for let-7 miRNAs to accumulate and exert their functions to inhibit proliferation and to promote differentiation (Thornton & Gregory, 2012). Let-7 targets pluripotency factors amongst which is Lin28 via let-7 binding sites in its 3’ UTR (Melton, Judson & Blelloch, 2010). Reprogramming studies have demonstrated an improved efficiency of iPSCs generation by inhibiting of let-7 (Marson et al., 2008). Moreover, human iPSCs have been generated through the ectopic expression of OCT4, SOX2, NANOG and LIN28 (Yu et al., 2007). It has thus been suggested that LIN28 and let-7 act as a bistable switch which ensures the maintenance of pluripotency or differentiation (Thornton & Gregory, 2012).

Two distinct mechanisms have been reported for the LIN28-mediated repression of let-7 (Thornton & Gregory, 2012). LIN28B, predominantly localised in the nucleus, binds to the terminal loop of primary transcripts from the let-7 family and inhibits processing by DROSHA (Viswanathan, Daley & Gregory, 2008). In the cytosol LIN28A also binds to the miRNA loop, but it recruits either Zcchc11 (TUT4) or Zcchc6 (TUTase7) terminal poly(U) polymerases which induces 3’ polyuridylation of pre-let-7 and targets it for degradation by the Dis3l2 exonuclease (Chang et al., 2013; Thornton et al., 2012; Hagan, Piskounova & Gregory, 2009). However,
recent work by Rahkonen et al. challenged the established model of LIN28-let-7 regulation and expression in ESCs (Rahkonen et al., 2016). They reported that let-7 miRNAs are processed to the mature form in human ESCs and that knockdown of LIN28 proteins has no effect on the mature levels of let-7 miRNAs (Rahkonen et al., 2016).

Post transcriptional regulation of miRNAs other than let-7 has been reported in mESCs (Thomson et al., 2006). Previous work in our lab corroborated published data by showing that some miRNAs are transcribed, but blocked in processing to their mature form in hESCs, which was confirmed by RT-PCR and RNA-seq (Rupa Sarkar, unpublished).
1.9 Aims

1. To show that a substantial proportion of miRNA genes are transcribed, but not processed to mature miRNAs in ESCs and to confirm that this post transcriptional regulation is widespread in pluripotent cells

2. To optimise the technique for the isolation of nascent RNA in order to confirm that miRNAs that are blocked in processing are dynamically synthesised and degraded

3. To undertake a microarray analysis of nascent and total RNA in order to test if hESCs transcribe and rapidly degrade other types of genes.

4. To optimise a technique for the pull-down of specific mature miRNAs \textit{in vitro} using complementary biotinylated DNA oligonucleotides and subsequent purification with streptavidin-coated magnetic beads
Chapter 2- Materials and methods

2.1 Cell culture and freezing cell lines

2.1.1 Cell culture of H1 cells

2.1.1.1 Preparation of mouse embryonic fibroblast-conditioned medium (MEF-CM) for hESC cell culture

Mouse embryonic fibroblasts (MEFs) were cultured in D10 medium, which contains Dulbecco’s Modified Eagle Medium (DMEM, Life Technologies™), 10% heat inactivated foetal bovine serum (FBS, Life Technologies™), 100 U/ml of penicillin, 100 µg/ml of streptomycin and 2 mM L-Glutamine (Life Technologies™) to passage 3 or 4 depending on the speed of cell growth. Cells were trypsinised and irradiated at 40 Grays (4000 rads), centrifuged at 800 rpm for 4 minutes and plated at 18.8 x 10^6 cells onto gelatin-coated T225 flasks with D10 medium. The following day the D10 was replaced with 150 ml of Knockout Serum Replacement Medium (KSR) composed of Knockout DMEM (KO-DMEM) and knockout serum replacement and supplemented with 4 ng/ml of basic Fibroblast Growth Factor (bFGF) (PeproTech). This medium called MEF-conditioned medium (MEF-CM) was collected and replaced every day for 7 days and was stored at -80°C. When needed it was thawed in a water bath at 37°C, supplemented with 1 mM L-Glutamine, 100 U/ml of penicillin and 100 µg/ml of streptomycin and filtered before usage. Once thawed, it was kept at 4°C.

2.1.1.2 Preparation of matrigel-coated plates

5 ml of stock matrigel (Corning® Matrigel® Basement Membrane Matrix) were thawed overnight at 4°C and diluted with 5 ml of cold KO-DMEM. This solution was aliquoted as 1 ml working volume into tubes and stored at -20°C. When needed, it was thawed at 4°C and further diluted with 14 ml of cold KO-DMEM. Plates were coated with this solution (1 ml per well in a 6-well plate) and incubated at 4°C overnight. If the plates were needed urgently, they were left at room temperature for at least 4 hours before being used.
2.1.1.3 Cell culture and splitting

H1 hESCs were grown on matrigel coated plates in MEF-CM supplemented with 4-8 ng/ml of bFGF. The medium was changed daily until the cells reached the desired confluency. Cells were routinely passaged at a 1:3 dilution upon treatment with 200 U/ml collagenase IV (Invitrogen) and scraping them with a stripette.

2.1.2 Cell culture of E14 cells

Mouse embryonic stem cells (mESCs) from the E14 cell line were grown in Glasgow’s Minimal Essential Medium (GMEM, Life Technologies™), supplemented with 10% batch tested Foetal Bovine Serum (FBS), 2mM L-Glutamine (PAA), 0.5 x Penicillin/Streptomycin (PAA), 0.01 mM of Minimum Essential Medium Non-Essential Amino Acids Solution (Life Technologies™), 0.25% (w/v) Sodium Bicarbonate (Life Technologies™), 1mM Sodium Pyruvate (Life Technologies™), 0.1 mM 2-Mercaptoethanol (Life Technologies™) and 1x Leukaemia inhibitory factor (LIF) made in house.

Plates were coated with 0.1 % of gelatine solution made by diluting the 2 % stock gelatine (Sigma-Aldrich®) in Dulbecco’s Phosphate Buffered Saline (DPBS) supplemented with MgCl₂ and CaCl₂ (Sigma-Aldrich®) prior to culturing cells on them.

Upon splitting, cells were washed with DPBS without MgCl₂ and CaCl₂ (Sigma-Aldrich®) and TrypLE™ Express Enzyme (Life Technologies™) was used to detach the cells from the plate. After spinning them at 1000 rpm for five minutes they were resuspended in fresh medium and were split between 1:5 and 1:10 ratio.

2.1.3 Cell culture of HEK293T cells

Cells from the HEK293T cell line were cultured in D10 medium. Cells were cultured until they were about 80 % confluent. To split them the medium was removed, they were washed once with PBS and were subsequently incubated with an appropriate volume (1 ml per 10 cm² plate, 0.2-0.3 ml per well in a 6 well plate) of 0.25 % of Gibco® Trypsin-EDTA (Life
Technologies™) for 3 minutes at 37°C in the dark. After the cells were detached they were resuspended in D10 medium and centrifuged at 1000 rpm for 5 minutes. The medium containing trypsin was removed, the cell pellet was resuspended in fresh medium and the cells were plated at the required density.

2.1.4 Freezing cell lines

2.1.4.1 H1 cells

KO serum replacement supplemented with 10% DMSO was used to freeze cells instead of MEF-CM. Cryo vials were stored in a Mr Frosty at -80°C overnight before being transferred to liquid nitrogen for long term storage. Cells were thawed at 37°C in a water bath and resuspended in MEF-CM. They were centrifuged at 800 rpm for 5 minutes and plated on 6 well plates coated with matrigel and grown in MEF-CM supplemented with bFGF.

2.1.4.2 E14 cells

The medium used to freeze mESCs from the E14 cell line was made of 45% mESC medium, 45% FBS and 10% DMSO (Sigma-Aldrich®). Cells were resuspended in freezing medium and transferred into cryo vials which were then placed in a Mr. Frosty™ freezing container.

2.1.4.3 HEK293T cells

To freeze HEK293T cells medium was removed and cells were washed with PBS. They were detached from the plate surface with trypsin (see 2.1.3), resuspended in medium and spun down at 1000 rpm for 5 minutes. The cell pellet was resuspended in an appropriate amount of medium and DMSO was added drop by drop to a final concentration of 10% while the cells in suspension were gently shaken. They were aliquoted into 1 ml cryo vials which were transferred into a Mr Frosty™ Freezing Container (Life Technologies™) containing isopropanol at room temperature. The container was then kept at -80°C for at least 48 hours before being stored directly in liquid nitrogen.
2.2 RNA extraction and checking RNA quality

2.2.1 Total RNA extraction

Isolation of total RNA was performed with TRI Reagent® (Sigma-Aldrich) following the manufacturer’s protocol. Cells grown in a monolayer were directly lysed on the culture dish. 1 ml of TRI Reagent® was used per well in six-well plates and 5 ml were used in 10 cm² plates. After passing the cell lysate several times through a pipette to form a homogenous lysate samples were left to stand for 5 minutes at room temperature. They were transferred to 1.5 ml microcentrifuge tubes and 0.2 ml of chloroform per 1 ml of TRI Reagent® was added to each tube. After closing the lids of the tubes the samples were vigorously shaken for 15 seconds and allowed to stand for 5-10 minutes at room temperature. Centrifugation was performed at 12200 x g for 15 minutes at 4°C. The upper aqueous phase containing the RNA was transferred to a new 1.5 ml tube and mixed with 0.5 ml of isopropanol per 1 ml of TRI Reagent® used initially. After gently inverting the tubes 10 times they were left on ice for 10 minutes and subsequently centrifuged at 12200 x g for 10 minutes. The RNA pellets were washed two times with 1 ml of 75% ethanol per 1 ml TRI Reagent® used to lyse the cells at 7500 x g for 5 minutes at 4°C. The ethanol was pipetted out of the tubes and the remaining pellets were resuspended in 30 µl of RNase free water. The RNA concentration and purity were measured using the NanoDrop™ spectrophotometer.

2.2.2 Small RNA extraction

RNA samples were enriched for small RNA using the mirVana miRNA Isolation kit following the manufacturer’s instructions (Life Technologies™).
2.2.3 Checking RNA integrity

2.2.3.1 Agarose gels

RNA was run on 1 % 1 x TAE agarose gels at 4 V/cm (interelectrode distance) for one hour. Samples were mixed with formamide at a final concentration of at least 60 % as well as with 6 x DNA loading dye containing bromophenol blue and xylene cyanol and denatured at 65°C for 5 minutes before being loaded on the gel.

2.2.3.2 ©Agilent 2200 TapeStation System

RNA samples used for microarray analysis were run on the ©Agilent 2200 TapeStation following the manufacturer’s instructions. The instrument was used to quantify nascent and total RNA samples. It generated quality scores expressed as RINe (equivalent) values, indicative of the RNA integrity of total RNA samples. These values are comparable to the values produced by the ©Agilent 2100 Bioanalyzer.

2.3 cDNA synthesis, primer design, PCR and PCR product purification

2.3.1 First strand cDNA synthesis

Complementary DNA (cDNA) was made using SuperScript® IV (SSIV) Reverse Transcriptase (Life Technologies™) following the manufacturer’s protocol. In a 1.5 ml tube 500 ng of total RNA were mixed with 1 µl of 10mM dNTPs, 1 µl of 50 µM random hexamers and water up to 13 µl. The tube was heated at 65°C for 5 minutes and incubated on ice for at least 1 minute. The following components were added to the reaction: 4 µl of 5x SSIV buffer, 1 µl of 100 mM DTT, 1 µl RNaseOUT™ Ribonuclease inhibitor (40 U/µl) (Life Technologies™) and 1 µl SuperScript® IV Reverse Transcriptase (200 U/µl)( ThermoScientific™). The reaction mixture was incubated for 10 minutes at room temperature and then transferred to 55°C for 10 minutes. The reaction was terminated at 80°C for 10 minutes.
2.3.2 MicroRNA gene selection and primer design

MicroRNA genes were initially selected from our laboratory's sequencing data on the H1 hESCs cell line generated by Dr Elcie Chan. The genomic location of miRNAs whose mature form was not sequenced at all or sequenced at a really low level was checked on the NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene). Intergenic miRNAs were included in our analysis. Primers were designed for the primary miRNA transcripts so that they do not overlap with any genes in the same region to avoid detection of two different genes in the same PCR reaction. The Blat function of the UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) was used to retrieve the sequences adjacent to the mature miRNA sequence. Primers for pri-miRNAs were designed in Primer 3 (http://primer3.ut.ee/). Their specificity was confirmed by using the UCSC In-Silico PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) as well as the NCBI Primer-BLAST tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). For further details on primer design see Chapter 3.1.1.

2.3.3 Polymerase chain reaction (PCR)

Polymerase chain reaction was performed using DreamTaq Master Mix (ThermoScientific™). Expression of individual genes was checked in separate 25 µl reactions consisting of 12.5 µl of DreamTaq Master Mix (containing DreamTaq DNA polymerase, DreamTaq buffer, MgCl₂ and dNTPs), 1.25 µl of DMSO, 9.25 µl of water, 1 µl of cDNA and 1 µl of a mix of forward and reverse gene-specific primers (10 µM each). Thermo Hybaid PCR Express Thermal Cycler was used to perform RT-PCR. The instrument was set at 94°C for 3 minutes followed by 25 or 35 cycles of 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds and 1 cycle of 72°C for 10 minutes.

PCR products were run on 5% TBE polyacrylamide gels (PAGE) to confirm their sizes. The gels were prepared by mixing 6.25 ml of 40% (w/v) bisacrylamide (19:1)( Sigma-Aldrich®), 2.5 ml of 10x TBE, 41.25 ml of distilled water in a 50 ml Falcon tube. To catalyse the polymerisation of the gel 25 µl of TEMED and 500 µl of ammonium persulfate (APS) were added to the tube. The liquid was quickly aliquoted into 1 mm gel cassettes (ThermoScientific™) and 10-well combs were positioned in them.
PCR products were mixed with 6x DNA loading dye containing bromophenol blue and xylene cyanol and were run at 140 V for 40 minutes along with a 50 bp DNA ladder (New England Biolabs®).

2.3.4 PCR product elution

Bands of the expected sizes were excised from the gel and cut into small pieces with a sterile blade. 350 µl of DNA elution solution (0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1 % SDS) was added to each PCR product in a 1.5 ml microcentrifuge tube and the mixture was incubated with shaking (700 rpm) overnight at room temperature or for more than 2 hours at 37°C. It was transferred to a 0.22 µm centrifuge tube filter column (Costar® Spin-X®) and spun down for 2 minutes at 8000-10000 x g at room temperature. The remaining gel bits on top of the filter were washed with another 100 µl of DNA elution buffer in the same conditions. To precipitate the DNA from the remaining liquid 45 µl of 3 M sodium acetate (1/10 of the total volume), 1 µl of Glycobluem™ coprecipitant (ThermoScientific™, 15 mg/ml of glycogen) and 1350 µl (3 times of the total volume) of ice cold ethanol were added to the mixture and it was incubated for at least 2 hours on ice or overnight at -20°C. It was spun down at maximum speed for 20 minutes followed by a wash with 1 ml 75 % ethanol for 10 minutes. The DNA pellet was resuspended in 9 µl of water to which 1 µl of 10 µM of the forward or reverse primer specific to the expected product was added. Sanger sequencing was performed at the MRC CSC Genomics Core Laboratory at the Hammersmith Campus of Imperial College or by GenewizUK.

The obtained sequence was analysed by processing it on the Nucleotide Blast ® online tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) to confirm the genomic location of the PCR product.

2.3.5 Quantitative PCR (qPCR)

Relative quantification of primary miRNA transcripts was carried out in a 7900HT Fast Real-Time PCR System (ThermoScientific™). The components of the PCR reaction were assembled in two separate sets of PCR tubes and were aliquoted into a 96-well plate (StarLab UK) using
multichannel pipettes. One set of tubes were containing nuclease free water (9.4 µl per well) and cDNA (1 µl per well) and the other set - a mix of Power SYBR® Green PCR Master Mix (10 µl per well, ThermoScientific™) and forward and reverse primers for the miRNA of interest (0.6 µl in total at a final concentration of 0.5 µM each). After combining the cDNA, SYBR® Green master mix, water and the primers in the 96-well plate an adhesive plate seal was applied on it and it was centrifuged at 4000 rpm for 5 minutes at 4°C prior to loading it in the 7900HT Real-Time PCR System. The machine was set to:

- Stage 1: 95°C for 10 minutes
- Stage 2: 40 cycles of 95°C for 15 seconds (step 1) followed by 60°C for 1 minute (step 2)
- Stage 3: 95°C for 15 seconds (step 1), 60°C for 15 seconds (step 2), and 95°C for 15 seconds (step 3)

Fluorescence readings were performed at stage 1, at steps 1 and 2 of stage 2, half-way between steps 2 and 3 from stage 3 and during step 3 in stage 3. The PCR products were run on 5 % PAGE gels (see 2.3.3) to confirm their sizes. The results were analysed using 7900 SDS v2.4.1 software.

2.4 Immunohistochemistry of nascent RNA

Cells from the H1 hESC cell line were plated onto 13 mm² glass coverslips placed in a 24-well plate and left to adhere overnight. They were either pre-treated with 1 µM actinomycin D for 1 hour or only treated with 1 mM 5-fluorouracil (5-FU) for 30 minutes or 1 hour at 37°C in the dark. After treating the cells all steps of the protocol were performed at room temperature and in low light levels. Cells were fixed in 10 % formalin for 15 minutes and washed with PBS supplemented with 0.1 % Bovine serum albumin (BSA) and 0.2 % Tween 20 (PBS-BT). To permeabilise the cells they were treated with 0.1 Triton-X in PBS for 5 minutes. Following another wash with PBS-BT blocking solution (10 % FCS in PBS) was applied for 30 minutes. It was replaced with 1:500 monoclonal anti-BrdU antibody produced in mouse (in 10 % FCS, Sigma-Aldrich®, clone BU-33, cat number B8434) for 30 minutes. The cells were washed three times with PBS-BT and were incubated with 1:100 goat anti-mouse IgG Alexa Fluor® 488
conjugated secondary antibody in 10% FCS (Life Technologies™) for 1 hour. Cells were washed twice with PBS-BT and once with distilled water. DAPI staining was performed in 300 µl of 300 nM DAPI diluted in PBS for 1-5 minutes. Cells were washed three times with PBS, any excess buffer was drained and the coverslips were mounted on glass slides using ProLong® Gold Antifade Mountant with DAPI (Life Technologies™). Photographs were taken on a Leica SP5 Confocal Microscope.

2.5 Northern blotting

2.5.1 Preparation and running a denaturing urea gel

To prepare the gel chamber two glass plates (one plain and one notched) were cleaned with ethanol and RNaseZAP™ (Sigma-Aldrich®) and were held together with clip holders, whilst keeping them at a distance by plastic separators. Agarose was prepared at a final concentration of 2% and was melted in the microwave. While still melted, it was used to make a seal on each side of the gel chamber by applying many layers of it, thus preventing the gel from leaking once it was poured into the chamber. The gel used for Northern blotting was a 15% denaturing polyacrylamide gel (40% (w/v) bisacrylamide (19:1)) containing 7M urea and 0.5 x TBE. After the urea was dissolved in water, TBE and acrylamide (50 ml total volume), the gel polymerisation was initiated with the addition of 25 µl of TEMED and 500 µl of 10% ammonium persulfate (APS), it was poured into the chamber and a comb of the correct size was fitted in it.

After the gel set the agarose was removed from all sides of the chamber. The bottom plastic spacer and the comb were taken out. All bubbles at the bottom of the gel were removed and all wells were flushed with 0.5 x TBE prior to pre-running it in 0.5 x TBE at 100 V for 20-30 minutes to make sure the urea is evenly spread. 10-30 µg of total RNA and 1-2 µg of small RNA were denatured at 65°C for 5 minutes, incubated on ice for 1 minute and were loaded on the gel after flushing the wells to remove any excess urea. Samples were mixed with loading buffer containing 95% formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol and 5 mM EDTA. The volume of loading buffer was equal to the volume of the RNA sample mixed with DNase/RNase free water. The gel was run at 150 V until all the dye
had entered the gel and then it was run at 250 V until the bromphenol blue dye reached the bottom part of the gel.

2.5.2 RNA transfer

After separating the RNA by running it on the denaturing gel, the chamber was disassembled and the gel was briefly washed with 0.5 x TBE to remove any excess urea. Six pieces of Whatman filter paper and 1 piece of Hybond N+ nylon membrane (Amersham) measuring exactly the same size as the gel were cut and soaked in 0.5 x TBE.

The semi-dry transfer apparatus was cleaned with 70 % ethanol and RNaseZAP™ (Sigma-Aldrich®). The transfer ‘sandwich’ was assembled on the semi-dry transfer apparatus in the following order: 3 pieces of filter paper were placed on top of the apparatus, followed by the membrane, the gel and the remaining 3 pieces of filter paper. The ‘sandwich’ was rolled with a stripette to remove any bubbles. Any excess liquid was removed and the apparatus was run at 3.3 mA per cm$^2$ for 30-35 minutes without exceeding 20 V. The RNA was fixed to the membrane by UV radiation in a CL-1000 Ultraviolet Crosslinker (two times at 120000 microjoules per cm$^2$).

2.5.3 Labelling of DNA probes with [$\gamma$-$^{32}$P] ATP

Single stranded DNA probes complementary to the miRNAs of interest were designed. Between 1 and 50 pmol of the DNA oligonucleotide were incubated with 1 x Polynucleotide Kinase Buffer, 1 µl of [$\gamma$-$^{32}$P] ATP (6000Ci/mmol 10mCi/ml, Perkin Elmer®) and 20 units of T4 Polynucleotide Kinase (New England Biolabs®) for 1 hour at 37°C in a total volume of 25 µl. To precipitate the probe and remove any excess [$\gamma$-$^{32}$P] ATP 55 µl of Tris-EDTA (pH 8.0) buffer, 20 µl of ammonium acetate (10M) and 1 µl of Glycoblué™ (ThermoScientific™, 15 mg/ml of glycogen) were added to the reaction. The sample was mixed well and 250 µl of ethanol kept at -20°C was added. The sample was left on ice for 1-2 hours and was then centrifuged at maximum speed for 20 minutes at 4°C. The supernatant was transferred to a separate tube and the strength of the radiation signal was compared. An indicator of successful labelling is a ratio of 1:2- solution: pellet. The pellet was resuspended in 100 µl of nuclease free water.
2.5.4 Hybridisation

The membrane was placed in a 60 ml container with a metal cap (Sterilin™) with the RNA side facing inwards. It was washed with 2 x SSC + 0.1 % SDS for 10 minutes at room temperature. Afterwards it was prehybridised with 3 ml of hybridisation buffer preheated to 42°C consisting of 6 x SSC, 5 x Denhardt’s solution (ThermoScientific™) and 0.5 % SDS for 30 minutes with rotation in an oven at 42°C. The hybridisation buffer was replaced with 1.5-2 ml of fresh buffer and the probe was added to it. The membrane was incubated for at least 12 hours overnight in the same conditions. It was then washed 2-4 times with 2 x SSC + 0.1 % SDS at room temperature. It was wrapped in saran wrap, placed inside a cassette with intensifying screens with an x-ray film on top and left for 5 days at -80°C. The film was developed in a dark room using OPTIMAX film processor.

2.6 Isolation of nascent RNA

2.6.1 Cell labelling and biotinylation of 4sU-tagged RNA

HEK293T and hESCs grown to 80% confluency were labelled with 200 µM 4-Thiouridine (4sU) for 30 minutes. Cells to which no 4sU was added served as negative controls. Total RNA was isolated in the way described in 2.2.1.

Biotinylation of the 4sU-labelled RNA (between 70-100 µg at a final concentration of 100 ng/µl) was carried out in 1 x TE buffer (pH 7.4), 1 mM EDTA and BIOTIN-HPDP EZ-LINK (Pierce) (final concentration of 0.2 mg/ml) dissolved in dimethylformamide (DMF). The reaction was performed at room temperature with rotation for 1.5 hours. When MTSEA biotin-XX (Biotium) was used the biotinylation reaction consisted of 4sU-labelled RNA (50-70 µg), biotinylation buffer (final conditions: 10mM HEPES [pH 7.5], 1mM EDTA) and 5 µg MTSEA biotin-XX dissolved in DMF (final DMF concentration = 20%) in a total of 150 µl. Tubes were wrapped in aluminium foil and were incubated for 2 hours at room temperature with rotation.

To remove the excess biotin an equal amount of chloroform was added. The tube was vigorously shaken and left at room temperature for 3 minutes until the phases began to
separate and the bubbles began to disappear. The samples were centrifuged at maximum speed (16000 x g) for 5 minutes at 4°C and the upper aqueous phase was transferred to a pre-spun 1.5 ml Phase Lock Gel Heavy tube (5 PRIME). An equal volume of chloroform was added and the tubes were shaken vigorously and left at room temperature for 3 minutes until the phases began to separate. They were centrifuged at maximum speed (16000 x g) for 5 minutes at 4°C. RNA was precipitated from the upper phase with the addition of NaCl (5 M) at 1/10 of the volume of the aqueous phase and an equal amount of isopropanol followed by a 20-minute centrifugation at maximum speed (16000 x g) at 4°C. The pellet was washed with 75% ethanol at maximum speed (16000 x g) for 10 minutes at 4°C and was resuspended in 100 µl of water.

2.6.2 Magnetic purification of nascent RNA

The biotinylated RNA was denatured at 65°C for 10 minutes and then incubated on ice for 5 minutes. It was mixed with 200 µl of µMACS streptavidin beads (Miltenyi Biotec) and incubated at room temperature with rotation for 15 minutes. In the meantime µMACS columns were placed in a magnetic separator and were pre-run and equilibrated by applying 0.9 ml of washing buffer (100mM Tris-HCl, pH 7.4, 10 mM EDTA, 1 M NaCl, 0.1 % Tween 20) and letting it flow through. The RNA-bead solution was transferred to the columns and washed three times with 0.9 ml washing buffer heated to 65°C followed by another three washes with 0.9 ml washing buffer at room temperature. Nascent RNA was eluted from the columns by the addition of 100 µl of freshly made 100 mM DTT (dithiothreitol) followed by a second one three minutes later. A chloroform extraction and NaCl/isopropanol precipitation (see 2.6.1) or the RNeasy MinElute Spin column kit (Qiagen) were used to recover the nascent RNA from the eluted fraction.

2.6.3 Dot blot analysis of biotinylated RNA

Biotinylated RNA was analysed by performing a dot blot. RNA samples from both labelled and unlabelled cells were denatured at 65°C for 10 minutes and placed on ice for 5 minutes. Serial 10-fold dilutions of the samples were made ranging from 1 µg to 1 ng and 1 µl aliquots were
pipetted onto a positively charged Hybond- N+ nylon membrane (Amersham). The RNA was fixed to the membrane by UV radiation in a CL-1000 Ultraviolet Crosslinker (two times at 120000 microjoules per cm²). The membrane was incubated in blocking solution (PBS, 10 % SDS, 1 mM EDTA) for 20 minutes. It was then probed with 1:1000 streptavidin-horseradish peroxidase solution (Streptavidin-HRP, Millipore Cat 18-152) in blocking solution for 15 minutes. This was followed by six washes with PBS with decreasing concentrations of SDS (10 %, 1 % and 0.1 %), each of which was applied two times for 10 minutes. The membrane was dried lightly and covered with Immobilon Western Chemiluminescent HRP Substrate (Millipore Cat WBKLS0500) for 2 minutes. It was covered with a plastic wallet and any excess substrate was removed. The membrane was exposed to an x-ray film for several time points (15 s, 30 s, 1 minute and 2 minutes) to obtain the optimal image.

Nascent RNA was either visualised on an agarose gel (see section 3) or on the 2200 Tapestation Instrument (Agilent Technologies) using the Agilent High Sensitivity RNA ScreenTape System.

2.7 MicroRNA pull-down

2.7.1 Labelling of DNA probes with [γ-32P] ATP

RNA and DNA oligonucleotides were labelled with [γ-32P] ATP in the same way as in 2.5.3. After labelling water was added to adjust the total volume to 300 µl. 30 µl (1/10) of 5M NaCl, 3 µl of Glycoblué™ (15 mg/ml, Life Technologies™) and 300 µl (equal amount) of isopropanol were added to the sample to precipitate the probe. It was spun down at maximum speed at 4°C for 25 minutes and resuspended in nuclease free water.

2.7.2 Incubation with μMACS streptavidin beads

In the initial experiment when the μMACS streptavidin kit (Miltenyi Biotec) was used, the protocol for the isolation of specific transcripts/RNAs provided by the manufacturer was followed. TEN buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA, 100 mM NaCl) was added to the labelled probe to a final volume of 200 µl and the whole reaction was denatured for 5 minutes.
at 75°C. Complementary oligonucleotide to mir-101-1 which was either modified by the addition of Biotin-TEG on its 5’ end or on both the 5’ and 3’ ends was annealed to the labelled probe at either 43°C or 38°C, respectively. Annealing temperatures were calculated using this formula:

\[ T_M = 67 + 16.6 \log_{10} \left( \frac{[Na^+]}{(1.0 + 0.7[Na^+]}) + 0.8(\%GC) \right) - \frac{500}{n} \]

\([Na^+]\) is the concentration of Na+ ions in mol/l, \((%GC)\) is the percentage of GC in the duplex and \(n\) is the length of the oligonucleotide. The temperature calculated in this manner was reduced by 10°C for single Biotin-TEG-tagged oligonucleotides and by 15°C for double Biotin-TEG-tagged oligonucleotides.

The µ column was prepared by placing it into a µMACS separator and applying 100 µl of Equilibration buffer for nucleic acid applications supplied with the kit and washing it twice with 200 µl of TEN buffer (10 mM Tris/HCl pH 8.0, 1 mM EDTA, 100 mM NaCl).

After hybridisation 100 µl of µMACS streptavidin microbeads were added to the reaction and incubated for 2 minutes at the annealing temperature. It was loaded onto the magnetic column and washed 5 times with 200 µl 1 x TE buffer pH 8.0. The probe was eluted with 150 µl nuclease free water heated to 80°C. It was then precipitated with NaCl/isopropanol (see 2.6.1).

2.7.3 Incubation with streptavidin magnetic beads (NEB)

When NEB magnetic beads were used, the same denaturing and annealing steps were followed as in 2.7.2. 36µl of streptavidin beads were used per reaction. They were washed twice with 100 µl of Wash/Binding buffer (0.5 M NaCl, 20 mM Tris-HCl (pH 7.5), 1 mM EDTA) by applying a magnet to the side of the tubes for 30 seconds and removing the buffer. The reaction from the annealing step was added to the beads, the mix was vortexed and incubated for 10 minutes at room temperature with agitation by hand every 2 minutes. Upon applying a magnet and removing the supernatant the sample was washed twice with 100 µl of wash(binding buffer followed by a wash with 100 µl of low salt buffer (TEN buffer from 2.7.2).
The oligonucleotides annealed to the biotinylated DNA sequences were eluted by applying 25 µl elution buffer preheated to 70°C (10 mM Tris-HCl (pH 7.5), 1mM EDTA), vortexing and incubating at room temperature for 2 minutes. The latter step was repeated twice. The eluted products were precipitated with NaCl/Isopropanol (see 2.6.1).

2.7.4 Denaturing gel and signal detection

12 % denaturing polyacrylamide gel (40% (w/v) bisacrylamide (19:1)) containing 7M urea and 0.5 x TBE was prepared in the same way as in 2.5.1. Equal amounts of sample and 6 x DNA loading dye were loaded and run at 250 V until the bromophenol dye reached the bottom part of the gel. The gel was wrapped in saran wrap, placed inside a cassette with intensifying screens with an x-ray film on top and left overnight at -80°C. The film was developed in a dark room using OPTIMAX film processor.

2.8 Microarray analysis

The microarray experiment was performed by AROS Applied Biotechnology. Our platform of choice was the Illumina HT-12 v4 gene expression bead chip (https://www.illumina.com/products/humanht_12_expression_beadchip_kits_v4.html).

Gene expression data was analysed in GenomeStudio Software provided by Illumina. For gene expression analysis average normalisation or no normalisation was applied and the selected detection p-value cut-off was <0.05. When differential gene expression analysis was performed average normalisation and the Illumina custom error model with false discovery rate (FDR) correction were selected. The detection and differential p-value cut-off was < 0.05.
2.9 List of bioinformatics programs used

Blat (UCSC Genome Browser): https://genome.ucsc.edu/cgi-bin/hgBlat?command=start


IPA: https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/

miRBase: http://www.mirbase.org/

miRGator: http://mirgator.kobic.re.kr/


Primer3: http://primer3.ut.ee/

UCSC In-silico PCR: https://genome.ucsc.edu/cgi-bin/hgPcr

Venn Diagram: http://bioinformatics.psb.ugent.be/webtools/Venn/
### 2.10 Lists of primers used in chapters 3 and 5

#### Primers used in chapter 3

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-23a</td>
<td>GCTCTCTCTCTCTTTCTCCCC</td>
<td>TCAGGGTCGGTTGGAAATCC</td>
</tr>
<tr>
<td>hsa-miR-24-2</td>
<td>GCCTGTCTGCTGCTGCCATC</td>
<td>TGCTGAACGTGAACAGCAATGTA</td>
</tr>
<tr>
<td>hsa-miR-27a</td>
<td>CTTGTGAGCAGGGGTCCAC</td>
<td>GCAGGATGGAACAGCAAGAC</td>
</tr>
<tr>
<td>hsa-miR-132</td>
<td>TCCGGCTGTCTGCTGATCAAC</td>
<td>TGGGAATCTGGAATCCAGCG</td>
</tr>
<tr>
<td>hsa-miR-134</td>
<td>TAGCTGTAAGTACCTGGCC</td>
<td>CAGAAGTCAGTCAAGCAAGTC</td>
</tr>
<tr>
<td>hsa-miR-138-1</td>
<td>TTGGTGAATCAGGGGTCCAC</td>
<td>GTCTGTCTGAGACGGGTATAG</td>
</tr>
<tr>
<td>hsa-miR-138-2</td>
<td>TTGGTGGGAAGGTGGTTGTGTA</td>
<td>CTGGTGTCGGTAAATAGCC</td>
</tr>
<tr>
<td>hsa-miR-146b</td>
<td>ACCTGCGACTGAACCTGAA</td>
<td>AGAGTAGGCGAGTCAAAGT</td>
</tr>
<tr>
<td>hsa-miR-154</td>
<td>CGCTCTGCTAACCACCATGATG</td>
<td>AGCGGGGATTTGACTGAA</td>
</tr>
<tr>
<td>hsa-miR-299</td>
<td>GCTCAATGACGCTGCTAGAA</td>
<td>GCTCTGCTGCTGCTCAG</td>
</tr>
<tr>
<td>hsa-miR-329-1</td>
<td>TTGGGGAAGAATCAGTGGTGT</td>
<td>ACTGGAGCGAGGTACAG</td>
</tr>
<tr>
<td>hsa-miR-345</td>
<td>GTGTGTGTCATGTAATTCAG</td>
<td>ATTCGAACCCAGGCCGCTG</td>
</tr>
<tr>
<td>hsa-miR-369</td>
<td>CAGCGCGTTTCTCTGCTTCTTCTT</td>
<td>GTCTGCTGCTGCTGCTGCT</td>
</tr>
<tr>
<td>hsa-miR-376a-1</td>
<td>CTGAATGAAACGAGGTTATGCC</td>
<td>GCTGACGGCGGTATATTCAATC</td>
</tr>
<tr>
<td>hsa-miR-376a-2</td>
<td>CTAGGTTGTTGCTGTTTCCGG</td>
<td>GTATGGGTGGCTTCCAGTCC</td>
</tr>
<tr>
<td>hsa-miR-376b</td>
<td>TTGGTGCTTTTCCAGACGCCCA</td>
<td>GGCATACCCGGTTTACCTCAG</td>
</tr>
<tr>
<td>hsa-miR-376c</td>
<td>TGCTGCTTCTCTGCCATGTTGCC</td>
<td>CGACTTTTCACTTACCCTTG</td>
</tr>
<tr>
<td>hsa-miR-377</td>
<td>CATCTCGGTTGGTTTCTTGGCC</td>
<td>ATCTGTCTCATGCCGACCA</td>
</tr>
<tr>
<td>hsa-miR-382</td>
<td>GTTCGTGTTGTTGATGCTTTCCTT</td>
<td>ATGGGAAGGCTGAGATTAC</td>
</tr>
<tr>
<td>hsa-miR-409</td>
<td>GAATGCGCCAGACCTTGTG</td>
<td>GAGCAACATTCGTCGTCAG</td>
</tr>
<tr>
<td>hsa-miR-485</td>
<td>GGTACACTTGGAGAGAAGGCTGG</td>
<td>CCTCAGGAAATGCTTTGCA</td>
</tr>
<tr>
<td>hsa-miR-494</td>
<td>GTGACACCAACACTCGGCG</td>
<td>GCATGAGAACAGGAAAGTGC</td>
</tr>
<tr>
<td>hsa-miR-495</td>
<td>GGAGCTGTCTCAGTCTGTGTA</td>
<td>ACCAACCACAAAGTCAGCG</td>
</tr>
<tr>
<td>hsa-miR-654</td>
<td>GCAAAGGAGAATGGTGAGG</td>
<td>AAGCAACCCGTTTACCTG</td>
</tr>
</tbody>
</table>

63
<table>
<thead>
<tr>
<th>miRNA</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-34a</td>
<td>GGCAGTGTTATAGCTGGCTGG</td>
<td>AAGAGCTTCCGAAGTCTCTGG</td>
</tr>
<tr>
<td>hsa-miR-182</td>
<td>GCCTGGACCTCACCTAAGTGT</td>
<td>GAGAACAGCAGGTCACGCTC</td>
</tr>
<tr>
<td>hsa-miR-183</td>
<td>CGCACTGAGTGTGACTCTCTCCT</td>
<td>TTCCGACGTCTACGGACGACT</td>
</tr>
<tr>
<td>mmu-miR-201</td>
<td>CATCTATTGCTTTTCTGCTGGTTT</td>
<td>CCAAGGCTGGAATGTTCTATTAC</td>
</tr>
<tr>
<td>mmu-miR-344f</td>
<td>GAATGCACACACCTACCAAGGGAACA</td>
<td>TGGATCTGTTTCTGGGCTGAA</td>
</tr>
<tr>
<td>mmu-miR-384</td>
<td>TGCAGCAGAATGTAAAATCTGA</td>
<td>TGATCCAGAGATGTTCAGGAATG</td>
</tr>
<tr>
<td>mmu-miR-451a</td>
<td>GCCTCAACAGCCTGGCTGAC</td>
<td>CTGGGTGTCCTTTTTCTTCCA</td>
</tr>
<tr>
<td>mmu-miR-544</td>
<td>GGCCTCTCAAACCTACTCA</td>
<td>AACATGCAACATCACCTCCTCCTCA</td>
</tr>
<tr>
<td>mmu-miR-547</td>
<td>GAAATGGCAAGGGAACAGA</td>
<td>AAGCCCACTTTTCTAGGGCTCA</td>
</tr>
<tr>
<td>mmu-miR-802</td>
<td>GGCTCCCACTCTGACTCTAC</td>
<td>CGAAACAGCTCCTCTACTCTCTCT</td>
</tr>
<tr>
<td>mmu-miR-182</td>
<td>GGCCTCTCAAACCTACTCA</td>
<td>CAGCTCACGCTTACACCAT</td>
</tr>
<tr>
<td>mmu-miR-183</td>
<td>ACTCTGCTCCTGTGTAACGGC</td>
<td>GAGGTCTTGTACCTCGGCA</td>
</tr>
<tr>
<td>ACTB</td>
<td>AGAGCTAGGTGATGGGCGTACAG</td>
<td>AGGCTAGGTGATGGGCGTACAG</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AATCCCACACCTATCTTTCA</td>
<td>TGGACTCCACGCTACTCTCT</td>
</tr>
<tr>
<td>POU5F1</td>
<td>GTTGGATCTGAGCTGTGGCTGA</td>
<td>GGTTGCCCTCTACCTCGGTCTCT</td>
</tr>
<tr>
<td>SOX2</td>
<td>GCCGAGTGAGAATTTTGCTCA</td>
<td>GCAGCGTGATCTATTCTCTTCTTT</td>
</tr>
<tr>
<td>U6</td>
<td>CGCTCTGGCACCACTATAC</td>
<td>TTACGAATTTCCTGTGCAT</td>
</tr>
</tbody>
</table>

Table 2. 1 Forward and reverse primer pairs used in chapter 3
## Primers used in chapter 5

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD24</td>
<td>CGAGAGATAACCCCTGCCCG</td>
<td>CCCCCAAAAAGAAAAAGTCCGC</td>
</tr>
<tr>
<td>L1TD1</td>
<td>TTGGTTCGCTCCTCAGTCGTC</td>
<td>ACCTCAATCAAGTCCTCAGCAGT</td>
</tr>
<tr>
<td>DPPA4</td>
<td>GCAAAGGCAAGGAGTGAGGACC</td>
<td>GGGACGTITTTCCCCCTTTTCC</td>
</tr>
<tr>
<td>LIN28</td>
<td>GGAAGTTTTGAGGAGCAGG</td>
<td>GGGTGAATGTTGTGAACCCAA</td>
</tr>
<tr>
<td>CDX2</td>
<td>CTCCAACCATTGGTGCTGTCGT</td>
<td>AGGGTACATGCTACGTCT</td>
</tr>
<tr>
<td>FOXF2</td>
<td>CGCCACAATCTCTCGCTCAA</td>
<td>GCCGGCCTCATGTCGAG</td>
</tr>
<tr>
<td>HOXA9</td>
<td>TTGCACCAGACAAAGTGA</td>
<td>GCCAATGCGGGGTTTCTAG</td>
</tr>
<tr>
<td>HOXA5</td>
<td>ATGTGCGTGGAGTGCTTCTCT</td>
<td>TCCAGAGTTCGTGCAAAGGG</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AATCCCATCACCATCTTCC</td>
<td>TGGACTCCACGACGTACTCA</td>
</tr>
<tr>
<td>POU5F1</td>
<td>GTTGATCCTCGGACCTGGCTA</td>
<td>GGTGCTCTCCACTCGGGTTCT</td>
</tr>
<tr>
<td>AICDA</td>
<td>CAATAAGAAACGCGTCCAGCAG</td>
<td>GGTCTTCTAACAAAGTATCCAGCA</td>
</tr>
<tr>
<td>RP1L1</td>
<td>CGGGAAAAGGTGAGCTCGC</td>
<td>GATCACAACGGCTTGCTTCT</td>
</tr>
<tr>
<td>SYCP2L</td>
<td>GGAGGTTCTGGGATTCTGCGG</td>
<td>TGAATGCGTACAGACGAAGCA</td>
</tr>
<tr>
<td>SSTR3</td>
<td>AAGCGGTTCTGAGAGTGGGTG</td>
<td>CAGCTATTGCCCTGCCCAT</td>
</tr>
<tr>
<td>POPDC2</td>
<td>TAAGGGTGCTGCGCTGTTG</td>
<td>TCTAGAAGCCAGTATGCCACT</td>
</tr>
<tr>
<td>HPX</td>
<td>ATATCATGGCAGAGCCGC</td>
<td>GATGAGGTACAGACGGCGGAC</td>
</tr>
<tr>
<td>NECAB1</td>
<td>GTCAGCGGTCAGGTTATT</td>
<td>TTGTCTGAAGGTTGCTATTCC</td>
</tr>
<tr>
<td>OR2A2</td>
<td>GTGCACGATCCTGTGCTTCA</td>
<td>TTATGCAGGCCCTCTTTTTG</td>
</tr>
<tr>
<td>IL1RN</td>
<td>TTGCTGCAAGTCCAGAATGGA</td>
<td>CCAGATCTCAGAAGGCTTTGCAATC</td>
</tr>
<tr>
<td>FIT1</td>
<td>ACAGCAACTTCTTCACATAAAAT</td>
<td>CCATGACAGGCAGCAAAAG</td>
</tr>
</tbody>
</table>

Table 2. 2 Forward and reverse primer pairs used in chapter 5
Chapter 3- Post transcriptional regulation of miRNAs in embryonic stem cells

3.1 Introduction

3.1.1 Previous work in our lab suggests that post transcriptional regulation of primary miRNAs is prevalent in hESCs

Previous work in our lab aimed to establish the mechanism of regulation of miRNAs in human embryonic stem cells from the H1 cell line and in neural progenitor stem cells (NPSCs). Figure 3.2 shows PCR data for the detection of miRNA primary transcripts as well as sequencing data for the mature form of the same miRNAs in both cell types. Figure 3.2(A) lists the sequencing and RT-PCR results for miRNAs that are differentially expressed by H1 or NPSC cells. These examples are straightforward because the sequencing and RT-PCR results largely correlate. For example, miR-302a, a stem cell marker, has only been sequenced in hESCs and the primary transcript was only detected in these cells (figure 3.2(A)). However figure 3.2(B) shows that a set of miRNAs are transcribed, as assayed by RT-PCR, but not fully processed to their mature form in hESCs since they are either sequenced at a low level or not sequenced at all in this cell type. In total, 16 out of 37 tested miRNAs in hESCs were detected by RT-PCR but not by sequencing. This is in contrast with NPSCs, where 37 out of 37 miRNAs that were detected by RT-PCR were also detected by sequencing.

Importantly, PCR primers aimed at detecting pri-miRNA transcripts were designed so that they do not overlap with the stem-loop region of the miRNA. Primers were complementary to sequences outside of the pre-miRNAs to ensure that the desired pri-miRNA is amplified (figure 3.1).
**Figure 3.1 Primer design for the detection of pri-miRNA transcripts.** The primers, denoted by the arrows, target sequences outside of the pre-miRNA (in red), thus amplifying pri-miRNAs (in black), which also include the precursor sequence. The complication of detecting transcripts of protein coding genes is averted by selecting intergenic miRNAs.

Additionally, miRNAs investigated during this project were selected based on their genomic location. Only miRNAs which do not overlap with protein coding transcripts were included in our analysis due to technical complications arising when aiming to amplify intronic or exonic pri-miRNAs. In the latter case, if RT-PCR is used for the detection of the pri-miRNA, there are two resulting products: the protein-coding transcript and the pri-miRNA. However, it would not be possible to distinguish the two based on the product size or by performing Sanger sequencing of the amplicon because they would be the same.

A technique which can distinguish the two is 3’ Rapid Amplification of cDNA Ends PCR (3’ RACE PCR). It works by binding of an oligo(dT) adapter to the poly-A tail of genes, synthesis of cDNA and subsequent PCR with one primer specific to the adapter and the other- to the 5’ end of the gene. The overlapping transcripts have different 3’ ends and when a primer specific to the pri-miRNA is used the two resulting products can be distinguished based on the difference in their size. This technique only detects poly-A tail transcripts, therefore only the pri-miRNA and not the precursor miRNA can be seen. 3’ RACE PCR has previously been successfully used for the detection of pri-miRNAs located in introns or exons of protein-coding transcripts (figure 3.2).
3.1.2 Rationale and aims

Previous work in our lab estimated that ~40% of all miRNAs in H1 ESCs are transcribed, but not processed in this cell line. This block in processing was found to be prevalent in H1 cells after comparing them to the more differentiated NPSCs derived from these hESCs.

The aim of the initial experiments performed for this project was to confirm this observation by amplifying intergenic pri-miRNAs by RT-PCR in RNA from H1 cells and from a differentiated cell line other than NPSCs. Intergenic miRNAs were selected to avoid the complication of pri-miRNAs located within introns or exons of protein coding transcripts and to ensure their specific amplification.

A second goal was to confirm that unprocessed miRNA transcripts are dynamically transcribed and degraded and are not simply static residual transcripts. Transcription was blocked by actinomycin D and RT-PCR was performed to look for a reduction in pri-miRNA levels, indicative of active transcript degradation. Another approach was to isolate nascent RNA and perform RT-PCR to confirm the active synthesis of pri-miRNAs.
Figure 3. 2 Post transcriptional regulation of miRNAs is more prevalent in hESCs than in NPSCs. A. RT-PCR (used for the detection of intergenic miRNAs and 3’ RACE PCR (used for the amplification of intronic/exonic miRNAs and sequencing data (detecting mature miRNAs) for transcriptionally regulated miRNAs. Only miRNAs which were sequenced were also detected at the level of the primary transcript by PCR in both cell types. Primers were designed to specifically amplify pri-miRNAs by choosing sequences outside of the pre-miRNA transcripts (see 3.1.1). B. Post transcriptional regulation of miRNAs occurs in hESCs. MiRNAs that are transcribed (as seen by RT-PCR or 3’ RACE PCR) were not detected by deep sequencing in H1 cells. Work was performed by Rupa Sarkar (unpublished) and Tan et al. (Tan et al., 2014).
3.2 Results

3.2.1 Many microRNA transcripts can be detected by RT-PCR but not by sequencing in hESCs

Figure 3.3(A) shows sequencing data of miRNAs in H1 hESC and Human Embryonic Kidney (HEK293T) cell lines. We chose those miRNAs that are intergenic and are expressed at low levels in hESCs, according to our previous sequencing results (figure 3.3(A), column 2). Intergenic miRNAs were selected because their analysis is not complicated by overlapping protein coding transcripts. The HEK293T sequencing data (figure 3.3(A), column 3) was obtained from miRGator (http://mirgator.kobic.re.kr/) which is an online database that provides expression profiles (abundance and tissue specificity) of miRNAs based on publically available deep sequencing data. In figure 3.3(A) only the miRNAs in red were successfully detected by performing 35 cycles of RT-PCR (figure 3.3(B)). As expected most of these had the highest sequencing reads in HEK293T cells (figure 3.3(A), column 3). By contrast, under the same experimental conditions we were able to detect 18 miRNAs from hESCs by RT-PCR, despite rather low sequencing reads compared to other miRNAs expressed by this cell line (figure 3.3(A), (Tan et al., 2014)). Sequencing data on mir-302a (a stem cell marker) and let-7a-1 (a differentiation marker) miRNAs was included in figure 3.3(A) to confirm that the presented results were gathered from undifferentiated stem cells. Both let-7a-1 and mir-302a are located within other transcripts and RT-PCR would not be able to distinguish between the expression of the miRNAs and the overlapping transcripts. Therefore OCT4 and SOX2 were included in the PCR analysis as markers of pluripotency.
Figure 3. Deep sequencing data on 24 intergenic miRNAs (A) and RT-PCR (B) for detecting the primary transcripts of the same miRNAs in H1 hES and HEK293T cells. A. Normalised read counts for HEK293T cells were obtained from mirGator. hESC miRNA read counts were gathered in our group by Dr Elcie Chan. Highlighted in red are miRNAs whose primary transcripts were detected by RT-PCR (35 cycles from total RNA) in each cell type, as shown in (B). Data on a stem cell miRNA marker (mir-302a) and a differentiation marker (let-7a-1) was included to confirm the undifferentiated state of the hESCs used to gather sequencing data. OCT4 and SOX2 were positive controls and markers of the undifferentiated state of H1 cells and ACTB and GAPDH were positive controls in HEK293T RNA in RT-PCR experiments. Mir-34a, a previously detected miRNA transcript, was used as a positive control in both cell types.

We wanted to confirm the miRNA sequencing results for hESCs (figure 3.3(A)) by northern blotting (figure 3.4(A)). However, this technique was not sensitive enough and could only readily detect abundant miRNAs such as let-7a-1 and mir-302a (Figure 3.4(A)).
Figure 3. 4 Northern blots (A) were performed to confirm sequencing data for the expression of miRNAs in hESCs for which primary transcripts were detected by RT-PCR (B). A. Northern blots with samples enriched for small RNA (less than 200 nucleotides long) extracted from hESCs and HEK293T cells were probed for let-7a-1, mir-302a and mir-27a. B. Comparison of sequencing data for miRNAs between our and other research groups’ data deposited in miRBase and miRGator. The online miRNA expression databases miRBase and miRGator each were deposited with data on one sample of undifferentiated hESCs. Highlighted in red are miRNAs whose read count is 0 in at least two of the datasets. Data on a stem cell miRNA marker (mir-302a) and a differentiation marker (let-7a-1) was included to confirm the undifferentiated state of the hESCs and to allow the sequencing data of columns 2 to 4 to be compared.

However, our sequencing data was largely corroborated by sequencing data of two other groups (figure 3.4(B)), which was the only additional data we found. The miRNAs that are highlighted in red in column 1 of figure 3.4(B) are those which were not detected by sequencing in at least two of the three samples, yet could be detected by RT-PCR (figure 3.3). As expected, mir-302a was sequenced many times in all three samples whereas let-7a-1 was almost undetectable (figure 3.4(B)).

We compared the abundance of pri-miRNA transcripts which are blocked in processing and easily detected by RT-PCR (mir-24-2, mir-27a, mir-345, figure 3.3) to pri-miRNAs which are vigorously turned into mature miRNAs in hESCs (mir-182 and mir-183). We chose mir-182 and mir-183 because they are located in intergenic regions and are highly expressed in H1 cells.
Quantitative PCR was performed in three biological replicates (with three technical replicates from each sample) from total H1 RNA, but because of primer dimerisation fluorescence readings were not reliable and Ct values were not analysed. PCR products were run on a gel and representative images from each biological sample are shown in figure 3.5.

![Figure 3.5 Comparison between transcript abundance of miRNAs enriched in stem cells (miR-182, miR-183) and miRNAs that are blocked in processing in H1 cells (miR-24-2, miR-27a and miR-345). The bands shown here were generated by qPCR (40 cycles) on three biological samples. Each band is a representative of three technical replicates. Ct values were not usable due to primer dimers observed in some reactions. No contamination was observed in negative control samples (RT-).](image)

### 3.2.2 MicroRNA transcripts are actively turned over in hESCs

H1 hESCs were treated with 1 μM actinomycin D for four hours to test whether the primary miRNA transcripts detected by RT-PCR are being actively synthesised and degraded (figure 3.6).

In 16 out of the 18 miRNAs there was a clear reduction in the levels of their primary transcripts following actinomycin D treatment (figure 3.6). The two exceptions were mir-329-1 where the actinomycin D treated sample generated more of the miRNA primary transcript and mir-345 whose level did not seem to change between the treated and non-treated samples. Mir-34a was included as a positive control because the abundance of its primary transcript was previously shown to decrease upon actinomycin D treatment (Rupa Sarkar, unpublished).
OCT4 and SOX2, genes involved in maintaining the pluripotent state of stem cells, were used as positive controls and were easily detected after inhibiting transcription, presumably because their mRNAs are relatively stable.

Overall this experiment provided evidence that the detected primary miRNA transcripts are in general being actively transcribed and degraded since their abundance decreased upon blocking transcription with actinomycin D.

Figure 3. 6 Testing the turnover of primary miRNA transcripts by actinomycin D treatment. Human embryonic stem cells were grown in six well plates and treated with 1 μM actinomycin D for four hours. A. 35 cycles of RT-PCR were performed to detect primary miRNA transcripts whose mature form was either not or only sequenced at low levels in hESCs (see B). Control samples were treated with DMSO.
As a control we tested the effectiveness of actinomycin D by showing that it inhibited the incorporation of 5-fluorouracil (5-FU, a modified uracil base) in proliferating hESCs (compare A and B with C for the FITC row of figure 3.7).

Figure 3. 7 Immunohistochemistry staining of hESCs labelled with 5-FU showing the inhibition of transcription induced by actinomycin D. Cells were grown in 24 well plates on cover slips and were treated with: A. 1 mM 5-FU for 30 minutes. B. 1 mM 5-FU for 1 hour. C. 1 μM actinomycin D for 1 hour followed by 1 mM 5-FU for 1 hour. D. 1 mM 5-FU for 1 hour. No primary anti-BrdU antibody (recognising 5-FU) was added to the cells in D.

To quantify the changes in pri-miR transcript abundance quantitative Real-Time PCR was performed for two genes- pri-miR-24-2 and pri-miR-345 (figure 3.8). Three biological samples of H1 hESCs were used in this experiment with three technical repeats for each one. Figure 3.8 shows the mean expression values of the two pri-miR transcripts relative to GAPDH. Error bars represent standard error of the mean (SEM). Pri-miR-24-2 expression was reduced by 12.96 fold and pri-miR-345 by 1.64 fold as calculated by the ratios of actinomycin D-treated versus non-treated samples (Schmittgen & Livak, 2008).
Figure 3. Quantifying the reduction in the levels of primary transcripts for miR-24-2 and miR-345 by qPCR after treatment of hESCs from the H1 cell line with 1 µM actinomycin D for 4 hours. Bar graph showing the mean expression values (+/− SEM) of the two transcripts relative to GAPDH calculated using the $2^{-\Delta\Delta\text{Ct}}$ method with and without actinomycin D treatment. Control samples were treated with DMSO. Three biological samples were used in this experiment and three technical repeats were performed for each one.

To further confirm the active synthesis of miRNAs for which transcripts were detected in hESCs but the mature form was not (figure 3.3) nascent RNA was isolated from H1 hESCs and 35 cycles of RT-PCR were performed (figure 3.9). Nascent RNA was purified after labelling cells for 30 minutes with a modified uridine called 4-thiouridine (4sU) which incorporates into newly made RNA. Active transcription was confirmed for 15 out of the 18 miRNAs initially detected. Primary miRNA transcripts were not detected for any gene in negative control samples derived from cells to which 4sU was not added with the exception of a faint band for mir-24-2 (figure 3.9). OCT4 was detected in the negative control samples at 35 cycles of RT-PCR. However, no band was evident when the number of cycles was reduced to 25.
Figure 3.9 Confirmation of the active transcription of miRNAs by RT-PCR (35 cycles) on two nascent RNA samples isolated from H1 hESCs labelled with 200 μM 4sU for 30 minutes. Samples 1 and 2 are biological repeats. OCT4 was used as a positive control. A negative control sample (−4sU) derived from unlabelled cells treated with DMSO was included to confirm the specificity of the nascent RNA isolation technique. B. Nascent RNA images were taken after running samples on the Agilent Technologies 2200 TapeStation Instrument. No RNA is visible in the samples isolated from unlabelled cells.

The same experiment was performed after isolating nascent RNA from HEK293T cells to confirm the active transcription of miRNAs in this cell type. Cells were labelled for 30 minutes with 4sU and 35 cycles of RT-PCR were performed after the isolation of nascent RNA (figure 3.10).
Figure 3. 10 Confirmation of the active transcription of miRNAs by RT-PCR (35 cycles) on two nascent RNA samples isolated from HEK293T cells labelled with 200 μM 4sU for 30 minutes. A. Pri-miRNA transcripts that were previously detected in total RNA samples were amplified. Pri-miR-34a and GAPDH were used as positive control transcripts. A negative control sample (-4sU) derived from unlabelled cells treated with DMSO was included to confirm the specificity of the nascent RNA isolation technique. B. Nascent RNA images were taken after running samples on the Agilent Technologies 2200 TapeStation Instrument. No RNA is visible in the samples isolated from unlabelled cells.

Active transcription was confirmed for 7 out of the 12 miRNAs initially detected in HEK293T cells (figures 3.3 and 3.10). Two of these miRNAs, mir-382 and mir-154, were either not sequenced at all or sequenced at a very low level in this cell line (figure 3.3). GAPDH was used as a positive control together with pri-miR-34a. No miRNA transcripts were detected in the negative control sample. GAPDH was detected at 35 cycles of PCR in 4sU-negative samples, but its transcript was no longer present when the number of PCR cycles was reduced to 25 (figure 3.10(A)).

A table summarising RT-PCR results from total and nascent RNA samples from both H1 and HEK293T cell lines was created for a better overview of the data (table 3.1). Highlighted in red
are miRNAs whose mature form is not expressed or expressed at a low level (table 3.1, column 4) and whose primary transcript was detected in both total and nascent RNA samples (table 3.1, columns 2 and 3).

<table>
<thead>
<tr>
<th>miRNA</th>
<th>H1 hESCs</th>
<th>HEK293T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RT-PCR (total RNA)</td>
<td>RT-PCR (nascent RNA)</td>
</tr>
<tr>
<td>23a</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>24-2</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>27a</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>132</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>134</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>138-1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>138-2</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>146b</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>154</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>299</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>329-1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>345</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>369</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>376a-1</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>376a-2</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>376b</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>376c</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>377</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>382</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>409</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>485</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>494</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>495</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>654</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

Table 3. 1 Summary table of the detection (✓) or lack of detection (×) of pri-miRNA transcripts by RT-PCR on total and nascent RNA and sequencing data (high or low) of mature miRNAs in H1 hESCs and HEK293T cells. Red colour represents miRNAs for which sequencing data showed low expression and which were successfully amplified by RT-PCR in both total and nascent RNA samples.
3.2.3 MicroRNA transcripts can be detected by RT-PCR but sequencing is showing a low or absent expression of their mature form in mouse ESCs

Mouse embryonic stem cells from the E14 cell line were tested for the expression of intergenic primary miRNAs for which sequencing data shows low expression of the mature form (figure 3.11(A)). Primary transcripts were detected for six out of the seven miRNAs tested by 35 cycles of RT-PCR (figure 3.11(B)). Positive controls in the RT-PCR reaction were mir-182 and mir-183 (miRNAs known to be abundantly expressed in mESCs) and U6, a component of the spliceosome.

<table>
<thead>
<tr>
<th>A</th>
<th>mESC E14 cell line sample GSM1278485</th>
<th>mESC E14 cell line sample GSM1278486</th>
<th>mESC E14 cell line sample GSM1278487</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmu-mir-201</td>
<td>0</td>
<td>8.26</td>
<td>19.58</td>
</tr>
<tr>
<td>mmu-mir-344f</td>
<td>0</td>
<td>8.26</td>
<td>0</td>
</tr>
<tr>
<td>mmu-mir-384</td>
<td>19.68</td>
<td>16.52</td>
<td>0</td>
</tr>
<tr>
<td>mmu-mir-451a</td>
<td>68.87</td>
<td>41.29</td>
<td>78.35</td>
</tr>
<tr>
<td>mmu-mir-544</td>
<td>147.58</td>
<td>189.93</td>
<td>254.65</td>
</tr>
<tr>
<td>mmu-mir-547</td>
<td>68.87</td>
<td>24.77</td>
<td>107.74</td>
</tr>
<tr>
<td>mmu-mir-802</td>
<td>0</td>
<td>16.52</td>
<td>0</td>
</tr>
<tr>
<td>mmu-mir-182</td>
<td>5496263.23</td>
<td>5574241.67</td>
<td>6095218.64</td>
</tr>
<tr>
<td>mmu-mir-183</td>
<td>5086882.57</td>
<td>4724327.88</td>
<td>5543703.33</td>
</tr>
<tr>
<td>mmu-let-7a-1</td>
<td>3813041.78</td>
<td>4003564.27</td>
<td>3227995.28</td>
</tr>
<tr>
<td>mmu-mir-290-295 cluster</td>
<td>6266364.44</td>
<td>6796486.10</td>
<td>7857952.06</td>
</tr>
</tbody>
</table>

Figure 3.11 Detection of primary miRNA transcripts whose mature form is detected at a low level by sequencing of E14 mESCs small RNA. A. Sequencing data on the expression of miRNAs in the E14 cell line. B. RT-PCR (35 cycles) detecting primary transcripts of some of the miRNAs in A. U6, pri-miR-182 and pri-miR-183 were used as positive controls in this experiment.

To test if these miRNA transcripts are actively turned over E14 cells were treated with 1 μM actinomycin D for 4 hours and 35 cycles of RT-PCR was performed (figure 3.12). A reduction in transcript levels was visible for the 6 primary miRNAs initially detected by RT-PCR.
Figure 3. 12 Testing the turnover of primary miRNA transcripts by actinomycin D treatment in mouse embryonic stem cells from the E14 cell line. Cells were grown in six well plates and treated with 1 μM actinomycin D for four hours. 35 cycles of RT-PCR were performed to detect primary miRNA transcripts whose mature form was either not or sequenced at low levels in mESCs. The control sample was treated with DMSO.
3.3 Discussion

3.3.1 Many pri-miRNAs are transcribed, but degraded in hESCs

A large proportion (18 out of 24 analysed miRNAs) of the primary miRNA transcripts whose mature form is expressed at a low level or not expressed at all in H1 hESCs were detected in H1 cells by RT-PCR (figure 3.3). This corroborates published data showing extensive post-transcriptional regulation of miRNAs in ESCs (Thomson et al., 2006). In HEK293T cells only 5 out of 16 pri-miRNA transcripts that are not present as mature products were detected by RT-PCR (figure 3.3). However, active transcription was only confirmed for two of these pri-miRNAs (figure 3.10, table 3.1). Similarly, previous work in our lab showed little evidence of post-transcriptional blocking of miRNA maturation in neural cells derived from the H1 cell line (Sarkar, unpublished, chapter 3(Introduction)), which is in accord with the idea that the property of blocking miRNA biogenesis at the post transcriptional level is widespread in hESCs.

To further verify sequencing data gathered in our lab, northern blots of small RNA (smaller than 200 nucleotides) were performed (figure 3.4). Although they worked well for abundant miRNAs like let-7a in HEK293T cells and mir-302a, a stem cell marker, they were not sensitive enough to detect miRNAs from figure 3.3(A), which were expected to be expressed in HEK293T cells (e.g. mir-27a (figure 3.4(A))). However, other groups’ sequencing data on H1 hESCs confirmed a low or lack of expression of the miRNAs that were readily detected by RT-PCR (figure 3.4(B)). It may prove possible to enhance the detection of mature miRNAs by a different technique such as splinted ligation (Maroney et al., 2007).

There is sample and technical variation between different qPCR experiments, as evident in figure 3.5. However, the abundance of pri-miRNA transcripts of miRNAs blocked in processing which were easily detected in total RNA (figures 3.3 and 3.6) as well as nascent RNA (figure 3.9) (miR-24-2, miR-27a and miR-345) was comparable to highly expressed miRNAs in hESCs, such as miR-182 and miR-183 (figure 3.2 and figure 3.5). We could not quantify the expression of these miRNAs due to side products generated by qPCR which interfered with the fluorescence measurements, but biological repeats were consistent in showing abundant expression of the selected blocked and processed miRNAs (figure 3.5). This indicates that
primary transcripts of some of the blocked miRNAs are maintained at a similar level to miRNAs which are expressed in stem cells and strengthens the need for studying this phenomenon further.

Blocking transcription with actinomycin D resulted in a decrease in the amount of transcript present for the majority of primary miRNAs analysed (figure 3.6). This indicated that they are not dead-end products, but are actively transcribed and degraded.

Thirty-five cycles of RT-PCR were performed because some miRNAs (e.g. pri-miR-138-1) are harder to detect than others (e.g. pri-miR-24-2). Decreasing the number of cycles could lead to seeing bigger differences between actinomycin D treated and non-treated samples, possibly even in the case of mir-345 where no obvious changes between the two samples were observed.

Quantitative RT-PCR was performed for two miRNAs (figure 3.8) for which primers generated one clear product – mir-24-2 and mir-345. Pri-miRNA gene expression was quantified relative to the housekeeping GAPDH protein coding transcript because its expression was stable under conditions of inhibited transcription (results not shown). A trend of transcript reduction was observed upon actinomycin D treatment. It corroborates the results from figure 3.5- the expression of pri-mir-24-2 was reduced 12.96 times and pri-mir-345 which seemed unchanged upon actinomycin D treatment was reduced 1.64 times (figure 3.8).

Studies have shown that using global inhibitors of transcription like actinomycin D elicit a cellular stress response that may lead to the stabilisation of certain transcripts as well as to alterations in cellular localisation of long non-coding RNAs (Tani et al., 2012; Shyu, Greenberg & Belasco, 1989). This could explain why despite it being successful at blocking cellular transcription (figure 3.7), actinomycin D treatment does not result in the full degradation of primary miRNA transcripts (figure 3.6(A)).

The active transcription of 15 pri-miRNA transcripts was confirmed by performing RT-PCR on nascent RNA isolated from H1 hESCs (figure 3.9, table 3.1). Amongst them were mir-329-1 and mir-345 which previously we thought were questionable as they did not follow the same pattern of reduction in transcript levels upon actinomycin D treatment like the other miRNAs (figure 3.6). A limitation of this experiment is the lack of data for abundantly expressed miRNAs like miR-182 and miR-183, initially compared to some miRNAs which are blocked in
processing (figure 3.5). This information would demonstrate if the rate of synthesis of expressed and blocked miRNAs is similar, which would indicate if they differ in other steps of their life cycle, such as processing or degradation. Negative controls in this experiment were derived from cells to which 4sU was not added. They were used to test if the isolated nascent RNA samples are contaminated with pre-existing RNA. A faint band was observed for mir-24-2 and OCT4 was easily detected at 35 cycles of PCR. OCT4 was no longer present when 25 cycles of PCR were performed (figure 3.9(A)). The same was observed in the negative control for HEK293T cells- GAPDH was detected in 4sU-negative samples at 35 cycles of PCR but it was not amplified after 25 cycles of PCR. This observation indicates that there is some pre-existing RNA in nascent RNA samples, but this contamination is minimal.

Six out of seven mouse miRNAs which are only detected at low levels by sequencing in E14 cells were successfully amplified by RT-PCR (figure 3.11), which is consistent with the data presented on human ESCs. However, it should be noted that a large proportion of the cells used in experiment GSE52950, from which miRNA expression data was gathered, were differentiated. This is obvious from the detected levels of miR-let-7a-1, which is normally absent from stem cells (figure 3.11(A)). The ratio between the expression of the miR-290-295 cluster (the most highly expressed cluster in mESCs) and let-7a-1 varied between 1.64 and 2.43 across the three samples and can be used as an indicator for the ratios of undifferentiated to differentiated cells in this experiment.

Treating mouse embryonic stem cells with actinomycin D resulted in a visible reduction in transcript levels of the 6 pri-miRNAs detected by RT-PCR (figure 3.12). This is indicative of active transcription and degradation occurring for these pri-miRNAs in line with the data for human ESCs.

Previous work in our lab has shown that hESCs transcribe but do not process many miRNAs (Introduction, 3.1). To extend this work further the regulation of intergenic miRNAs was investigated in H1 and HEK293T cells by performing RT-PCR and comparing the result with deep sequencing data. It was confirmed that hESCs transcribe, but do not process 18 out of 24 tested pri-miRNAs, whereas for HEK293T cells between 2 to 5 (detected in nascent and total RNA samples, respectively) out of 16 were transcribed but not processed (table 3.1). This suggests that many miRNAs are regulated at a post-transcriptional level in human embryonic stem cells and also raises the possibility of such a regulation for other gene types.
3.3.2 Post transcriptionally regulated miRNAs play roles in cell fate determination and cancer progression

MiRNAs that we found to be regulated at the post transcriptional level have been implicated in various processes and disease phenotypes.

The highly conserved miR-24 and miR-27a expressed at a very low level in mouse embryonic stem cells have been found to be essential for differentiation. Biallelic double knockout of the mir-23~27~24 cluster leads to defects in mesoderm differentiation both in vitro and in vivo (Ma et al., 2015; Roy et al., 2015). The same cluster has also been shown to be part of a regulatory loop that protects mouse ESCs undergoing differentiation to epiblast stem cells (EpiSCs) from BMP4-induced apoptosis (Musto et al., 2015). In humans the same cluster is involved in different types of cancer. It has been shown to act as an oncogene by targeting the tumour-suppressor gene Hypermethylated in cancer 1 (HIC1), a transcriptional repressor that in turn was found to negatively control the expression of the miRNA cluster (Wang et al., 2016). Recently it was discovered that miR-23a, miR-27a and miR-24 can be used as therapeutic targets in human acute erythroid leukaemia (AEL) (Su et al., 2016). The miRNA cluster is downregulated in AEL and restoring its expression induces apoptosis, enhances erythropoiesis and leads to a decreased malignant proliferation (Su et al., 2016). MiR-23a inhibits the expression of E-cadherin in gastrointestinal cancer and thus promotes epithelial to mesenchymal transition (EMT) (Zheng et al., 2014). Since both human and mouse embryonic stem cells are characterised by the expression of E-cadherin, it is possible that miR-23a plays a role in differentiation through being involved in the mechanisms regulating the necessary morphological changes (Redmer et al., 2011; McLean et al., 2007).

MiR-138 has been shown to modulate the osteogenic differentiation of human mesenchymal stem cells (hMSCs) - the expression level of miR-138 is high in hMSCs and it is reduced during osteoblast differentiation (Eskildsen et al., 2011). MiR-138 was also able to inhibit osteoblast differentiation in vitro (Eskildsen et al., 2011). The same miRNA was found to inhibit adipogenic differentiation of human adipose tissue-derived mesenchymal stem cells (hAD-MSCs) (Yang et al., 2010). It acts as a tumour suppressor in ovarian cancer by preventing cancer cell invasion and metastasis and as an oncogene in malignant gliomas through inducing
growth and promoting survival of tumour-initiating glioma stem cells (Yeh et al., 2013; Chan et al., 2012).

MiR-146b has been found to inhibit glioma, breast and pancreatic cancer cell invasion and migration; it has been implicated in myogenic differentiation (Khanna, Ge & Chen, 2014; Lin et al., 2011; Xia et al., 2009; Bhaumik et al., 2008). MiR-154 is downregulated in breast cancer cells and tissues and its overexpression leads to a reduction in cell proliferation, migration and invasion (Xu et al., 2016). MiR-329 was found to be a tumour suppressor that is downregulated in gastric and breast cancer; it has also been shown to suppress angiogenesis (Kang et al., 2016; Li et al., 2015; Wang et al., 2013).

Post transcriptional regulation of miRNAs was predominantly found to occur in hESCs. Studies have shown that the miRNAs regulated in this manner are involved in the regulation of cancer progression and stem cell differentiation. Understanding the precise mechanisms of miRNA regulation in stem cells will be beneficial because it will provide new therapeutic targets as well as an insight into the mechanisms regulating cell fate, which is valuable for the generation of induced pluripotent stem cells.
Chapter 4- Optimising the isolation of nascent RNA from HEK293T and H1 hESCs

4.1 Introduction

4.1.1 The importance of studying changes at the mRNA level through isolating nascent RNA

Steady-state RNA levels in cells are determined by the interplay between RNA synthesis and decay (Dolken et al., 2008). Traditionally studies have employed total RNA to study cell response to specific signals (Friedel et al., 2009). However, the contribution of de novo synthesis and mRNA rate of degradation to the stability of RNA transcripts cannot be discerned this way (Dolken et al., 2008).

Inducing global transcriptional arrest to study RNA degradation over time has been done with inhibitors of transcription, such as actinomycin D (actD) (Sharova et al., 2009; Raghavan et al., 2002). However, this approach is cell invasive and it assumes that RNA decay occurs at its regular rate after drug treatment (Friedel & Dölken, 2009). Furthermore, due to the high stability of most transcripts even big changes in de novo transcription upon alterations in cellular conditions can go unnoticed if the total RNA fraction is used (Dolken et al., 2008). For example, we found widespread inhibition of nascent RNA splicing by spliceostatin. This same assay worked poorly with total RNA because of the masking effect of already spliced mature RNA.

Techniques to assay newly synthesised RNA have been developed over the last five decades. Some of the key steps in the progress of the development of these methods, their advantages and disadvantages are presented here.
4.1.2 Methods for the isolation of nascent RNA

Assessing *de novo* gene transcription by nuclear run-on assays has its origins in the 1970s when the advantages of using the detergent sarkosyl were noticed and applied in the studies of polyoma virus transcription complex (Green, Buss & Gariglio, 1975; Gariglio, Buss & Green, 1974; Shmookler, Buss & Green, 1974). Treating isolated cellular nuclei with sarkosyl inhibited transcription initiation and specifically stimulated the activity of RNA polymerase II which had already initiated transcription *in vitro*. The enzyme can continue transcription even after sarkosyl inactivation if nucleoside triphosphates are supplied exogenously (Hawley & Roeder, 1987).

The nuclear run-on technique has made possible the quantification of RNA polymerase density and has provided a method for comparing transcription rates and specific as well as transcriptome-wide mRNA turnover in different cellular conditions (Garcia-Martinez, Aranda & Pérez-Ortíz, 2004; Fan et al., 2002; Hirayoshi & Lis, 1999). However, this method is technically demanding, it disrupts the physiological conditions in the cell and potentially introduces experimental bias since nuclear run-on RNA is transcribed *in vitro* following the isolation of cellular nuclei (Friedel & Dölken, 2009; Hirayoshi & Lis, 1999).

Pulse labelling of newly synthesised RNA by modified uridine analogues provides a less invasive and more accurate method for studying transcriptional turnover of mRNAs (Tani & Akimitsu, 2012).

After its first successful use in labelling RNA in mammalian cells in 1959 (Eidinoff, Cheong & Rich, 1959) 5′-Bromouridine (BrU) which gets converted to 5-bromouridine 5′-triphosphate (BrUTP) in cells has been used to detect nascent RNA immunocytochemically as well as to study the number and localisation of transcriptional sites within the nucleus, the number of polymerases associated with transcriptionally active sites and the number of active polymerases at any given moment (Jackson et al., 1998; Iborra et al., 1996; Wansink et al., 1993). BrUTP labelling of human IMR90 cells combined with a nuclear run-on approach and high-throughput sequencing after capturing nascent RNA with anti-BrdU antibodies conjugated to magnetic beads mapped the position of transcriptionally active RNA polymerases at a high resolution genome-wide (Core, Waterfall & Lis, 2008). Deep sequencing of RNA samples derived from cells labelled with BrU (BrU-Seq) and comparing them with
samples which have been chased with uridine for different amounts of time (BrU-Chase Seq), allows the determination of the stability of different mRNAs (Paulsen et al., 2014). The advantages of BrU labelling are that it is not toxic to cells upon treatment and it does not cause mismatches upon cDNA synthesis when using BrU-labelled RNA as a template (Core, Waterfall & Lis, 2008). However, effective BrU incorporation in cells takes 24 hours and it has been shown to alter pre-mRNA splicing (Tani & Akimitsu, 2012; Tani et al., 2012; Wansink, Nelissen & de Jong, 1994).

Another uridine analogue, 5-ethynyluridine (EU) has successfully been used to label nascent RNA and monitor its turnover in vivo in different tissue types through a copper-catalysed azide-alkyne cycloaddition reaction ("click" chemistry or CuAAC) of fluorescent azides followed by microscopic imaging (Jao & Salic, 2008). This method induces high cellular toxicity when applied in live cell imaging due to the generation of reactive oxygen species by copper from oxygen (Uttamapinant et al., 2012). EU is not toxic to the cells and does not affect the global transcriptome (Invitrogen. Catalog number: C10365.), but culturing cells in the presence of EU for 48 hours significantly inhibits cell growth (Tani et al., 2012). Nascent RNA labelling with EU followed by "click" chemistry which crosslinks EU to biotin-conjugated azides and is then separated by an incubation with streptavidin coated magnetic beads is an efficient way of isolating newly synthesised RNA from cells and it is commercially available as the Click-iT® Nascent RNA Capture Kit (Invitrogen. Catalog number: C10365.). It can be followed up with qRT-PC, microarray or deep sequencing analysis (Invitrogen. Catalog number: C10365.).

Metabolic tagging and isolation of nascent RNA with thiol-modified uridine analogues was first successful in the late 1970s (MELVIN et al., 1978). This study compared the incorporation of 6-thioguanosine and 4-thiouridine (4sU) by a hamster kidney cell line. 6-thioguanosine was found to impair RNA and protein synthesis after long periods of exposure, whereas 4-thiouridine had minimum effect on these processes (MELVIN et al., 1978). This study isolated nascent RNA using affinity chromatography on mercurated cellulose columns. More recently the same approach was used to isolate nascent RNA synthesised in vivo after injecting a mouse model of ischaemia intraperitoneally with 4sU (Kenzelmann et al., 2007). The RNA was then subjected to microarray analysis and identified genes involved in the renal reperfusion injury response. An improvement of purifying newly synthesised RNA after 4sU labelling of cells was introduced by Dölken et al. (Dolken et al., 2008). Instead of the agarose-based
organomercurial matrix used by Kenzelmann et al., thiol-specific biotinylation of labelled RNA and subsequent incubation with streptavidin-coated beads were used to magnetically separate newly made from pre-existing RNA, a major advantage of this technique (Dolken et al., 2008). Another thiol-modified nucleoside, 4-thiouracil (4tU), can be used to isolate nascent RNA (Cleary et al., 2005). However, the uptake of 4tU requires the enzyme uracil phosphoribosyltransferase (UPRT), which is not made by mammals or insects and is not going to be discussed here (Friedel & Dölken, 2009; Cleary et al., 2005).

Effective incorporation of 4sU into nascent RNA is achieved rapidly in a variety of cell lines and this molecule can be used even when short (30 minutes) or ultrashort (< 5 minutes) labelling times are required (Rabani et al., 2011; Friedel & Dölken, 2009; Dolken et al., 2008). 4sU has been shown to be toxic if culturing cells in its presence for 48 hours (Tani et al., 2012). However, when used at 200 µM for up to two hours in murine fibroblasts no toxicity was observed in the total RNA fraction after microarray analysis was performed (Dolken et al., 2008; Kenzelmann et al., 2007).

The ability to separate pre-existing from newly synthesised RNA with high specificity, together with the total RNA fraction can be used in determining gene half-lives when large scale gene studies involving microarray or deep sequencing analysis are performed (Rabani et al., 2011; Friedel & Dölken, 2009; Dolken et al., 2008). A transcript’s half-life is closely related to its biological function (Tani & Akimitsu, 2012). Protein coding transcripts with regulatory functions tend to have short half-lives whereas long-lived mRNAs are enriched amongst housekeeping genes (Tani & Akimitsu, 2012; Dolken et al., 2008). This information is valuable when investigating specific gene regulatory networks during stable conditions in cells or when exploring their response upon the introduction of a stimulus (Dolken et al., 2008).

4.1.3 Rationale and Aims

Nascent RNA has been successfully isolated from a variety of cell lines in the past and it has been used in downstream applications such as RT-PCR, microarray and deep sequencing (see 4.1.2).
The aim of the experiments presented here was to successfully isolate nascent RNA from HEK293T and H1 hESCs and to show that it can be used for RT-PCR, necessary for the experiments in chapter 3. Another goal was to confirm that this technique is specific for purifying newly synthesised RNA, keeping the amount of pre-existing RNA to a minimum, a pre-requisite for the microarray experiment presented in chapter 5.
4.2 Results

4.2.1 Optimising the duration of labelling, 4-thiouridine (4sU) concentration and precipitation of nascent RNA

Nascent RNA was isolated from H1 hESCs after optimising the technique in HEK293T cells. The protocol was adapted from Döllken et al. (Dolken et al., 2008). The outline of the protocol is presented in figure 4.1.

![Diagram](image)

**Figure 4.1 Schematic of the protocol used for the isolation of nascent RNA.** A modified uridine, 4-thiouridine (4sU), was added to the medium of H1 hESCs for thirty minutes at a concentration of 200 μM. Total RNA was extracted following the standard TRIzol protocol and 4sU-tagged RNA was cross-linked to biotin in vitro. Magnetic beads coated with streptavidin were used to purify biotinylated RNA. The 4sU-RNA was then eluted from the beads with dithiothreitol (DTT), which reduces the disulphide bond between the biotin molecule and 4sU-labelled RNA. Nascent RNA was precipitated from the eluted solution.
Figure 4.2 (A.2) shows a titration experiment of the concentration of 4sU for RNA labelling, which led us to choose 200 µM as the standard concentration. 4sU-RNA could be purified with only 30 minutes of 4sU treatment, although increasing amounts of 4sU RNA were purified with longer labelling periods, as expected (figure 4.2 (B.2)). RNA quality was monitored during the experiments by running 1% agarose gels and checking for the integrity of the 28S and 18S ribosomal RNA (rRNA) subunits and confirming that the ratio between the two is 2:1 (Figure 4.2 (A.1, B.1)).

![Figure 4.2 Testing of different concentrations (A) and labelling times (B) of 4sU for the isolation of nascent RNA. RNA from HEK293T cells to which 4sU was not added in the culture medium was used as a negative control in these experiments. A.1 and B.1 Quality of biotinylated RNA A.2 and B.2 Purified nascent RNA](image)

Two methods were tested for the recovery of nascent RNA after it has been magnetically purified in the final step of the protocol: a column-based approach (RNeasy MinElute Cleanup Kit (Qiagen)) and isopropanol precipitation with 10 % 5M NaCl and 45 µg of glycogen. The same amount of biotinylated RNA was used in both methods and nascent RNA was resuspended in the same volume of water. Equal volumes from the final eluates of nascent RNA were loaded on the Agilent 2200 TapeStation for a precise quantitation. An image of the nascent RNA fraction as well as the quantification provided by the instrument for each sample are presented in figure 4.3.
Figure 4. 3 Comparison between isopropanol/NaCl precipitation and the RNeasy MinElute Cleanup Kit (Qiagen) for the recovery of nascent RNA. A. Image generated after running nascent RNA samples on the Agilent Technologies 2200 TapeStation. B. Nascent RNA concentration in ng/µl as calculated by the Agilent Technologies 2200 TapeStation software. RNA isolated from HEK293T cells to which 4sU was not added in the culture medium was used as a negative control.

Isopropanol/NaCl precipitation resulted in recovering 2.47 times more nascent RNA than the RNeasy MinElute Cleanup Kit (48.5 ng/µl and 19.6 ng/µl, respectively) and it was the method of choice upon preparation of samples for microarray analysis (see chapter 5).

After optimising the technique for isolating newly transcribed RNA in HEK293T cells, it was tested in the H1 hESC line (figure 4.4). Nascent RNA was successfully isolated from these cells after 30 minutes of 4sU labelling with minimum contamination with non-labelled RNA, as seen in the dot blot in figure 4.4(B).
Figure 4. 4 Isolation of nascent RNA from H1 hESCs. A. Ethidium bromide staining of 1% agarose gels showing: A.1 Biotinylated RNA. A.2 Nascent RNA. B. Dot blot showing the specificity of the biotin interaction with 4sU-labelled RNA probed with a streptavidin-horseradish peroxidase conjugate. Serial dilutions of the RNA were made and cross-linked to the membrane. RNA isolated from HEK293T cells to which 4sU was not added in the culture medium was used as a negative control.

4.2.2 MTSEA biotin is more efficient in binding to 4sU-labelled RNA than HPDP biotin, but it is not 100% 4sU-specific

Duffy et al. (Duffy et al., 2015) reported better biotinylation of 4sU-RNA with MTSEA Biotin-XX rather than EZ-Link™ HPDP Biotin used by Dölken et al.(Dolken et al., 2008).

We compared the biotinylation of 4sU-RNA with these two reagents (figure 4.5). Figure 4.5(B) indicates that MTSEA biotin may biotinylate more efficiently, as reported. A signal of similar intensity to 100 ng of HPDP biotinylated 4sU-labelled RNA is visible on the dot blot (for 1 µg of RNA) after the incubation of 4sU-free RNA with MTSEA biotin and an unclear smearing was visible upon purification of the nascent RNA fraction (figure 4.5(C)).
Figure 4.5 Comparison between MTSEA and HPDP biotin for the isolation of nascent RNA. A. Quality of biotinylated RNA. B. Dot blot showing the specificity of the biotin interaction with 4sU-labelled RNA probed with a streptavidin-horseradish peroxidase conjugate. Serial dilutions of the RNA were made and cross-linked to the membrane. RNA isolated from HEK293T cells to which 4sU was not added in the culture medium was used as a negative control and was biotinylated with MTSEA biotin. A signal of similar intensity to the one obtained from 100 ng of HPDP-biotinylated RNA is visible in the negative control. C. Purified nascent RNA.

To further test the specificity of MTSEA biotin RNA from H1 hESCs that were incubated with and without 200µM 4sU for one hour was used to isolate nascent RNA and perform RT-PCR for primary miRNA transcripts previously observed in this cell line (see chapter 3). RNA was visible in the negative control fraction after the final elution of nascent RNA, as seen in figure 4.6(A). The low specificity of crosslinking with MTSEA biotin is evident from the successful amplification of primary miRNA transcripts in the negative control sample (figure 4.6(B)).
Figure 4.6 Testing the specificity of MTSEA biotin. H1 hESCs were labelled with 200 μM 4sU for one hour. A. Agarose gel of nascent RNA isolated after magnetically separating it from pre-existing RNA. RNA is visible in the negative control sample derived from cells that were not labelled with 4sU. B. RT-PCR was performed using eluates from unlabelled (35 cycles) and 4sU-labelled (30 cycles) H1 hESCs. Detection of primary miRNA transcripts is evident in both reactions. No reverse transcriptase was added in RT- reactions to control for the presence of genomic DNA.

To avoid contamination of nascent RNA with pre-existing RNA we decided to revert to using HPDP biotin instead of MTSEA biotin. Nascent RNA was isolated after labelling H1 hESCs with 4sU for 30 minutes and samples were quantified using the Agilent 2200 TapeStation system (figure 4.7(A)). No RNA was visible in the negative control sample and its concentration was determined to be 0.1 ng/μl.
Figure 4. 7 RT-PCR on nascent RNA from H1 hESCs labelled with 200 µM 4sU for 30 minutes. Nascent RNA was crosslinked to HPDP biotin. A. Agilent 2200 TapeStation was used to quantify the amount of nascent RNA in samples derived from labelled and unlabelled cells. B. 35 cycles of RT-PCR were performed using eluates from unlabelled and 4sU-labelled H1 hESC. No reverse transcriptase was added in RT- reactions to control for the presence of genomic DNA.

To confirm the specificity of HPDP biotin 35 cycles of RT-PCR were performed on nascent RNA samples isolated from cells labelled with and without 200 µM 4sU for 30 minutes (figure 4.7(B)). In the absence of 4sU none of the tested pri-miRNAs were amplified by RT-PCR (compare No 4sU and 200 µM 4sU 30 min in figure 4.7(B)). We did, however, detect a small amount of OCT4 in the negative control sample when 35 cycles of RT-PCR were performed, but not with 25 cycles.
4.3 Discussion

4.3.1 Selection of a technique for analysing nascent RNA transcripts

Several methods have been developed for the analysis of transcription by looking at de novo RNA synthesis in the past 50 years (Tani & Akimitsu, 2012). The low toxicity, fast incorporation and the ability to efficiently and specifically isolate nascent RNA upon labelling cells with 4sU were the advantages that led us to choose the protocol developed by Dölken et al. (Dolken et al., 2008) for our experiments. This technique was used for RT-PCR and microarray analysis (chapters 3 and 5, respectively).

4sU has been reported to inhibit the primary 47S rRNA transcript synthesis and processing (Burger et al., 2013). Even though the concentrations of 4sU used in this study were similar to the ones used in our experiments (50 µM and 100µM), the labelling times were significantly higher (3-6 hours). Evidence against large-scale genomic effects induced by short labelling with 4sU was confirmed by two studies (Dolken et al., 2008; Kenzelmann et al., 2007). Since we were interested in looking for rapidly turned over RNAs (hence labelling for 30 minutes), 4sU remained the molecule of choice for these experiments.

A drawback of isolating nascent RNA with the technique presented here is the potential to introduce transcriptional bias, i.e. the preferential isolation of longer transcripts which have incorporated more 4sU residues (Duffy et al., 2015). However, Friedel et al. reported that there is no such bias when analysing short transcripts (200-500 nt) (Friedel et al., 2009). They concluded that using 4sU is problematic when ultrashort transcripts are to be analysed (<200 nt), which could be of concern when short RNA molecules such as snoRNAs or miRNAs are studied. The experiments presented as part of this thesis did not aim to analyse the expression of such transcripts. Since primary miRNAs as well as precursor mRNAs are not part of the ultrashort gene RNA fraction, using 4sU labelling and magnetic separation of nascent RNA was a valid approach.

Another disadvantage of using 4sU is that it allows base pairing with guanine instead of adenine, which can affect cDNA synthesis and subsequent results determining RNA abundance (Tani & Akimitsu, 2012). However, the aim of our experiments was to get a general
overview of transcription in hESCs and HEK293T cells rather than to focus on specific transcripts (chapter 5) and 4sU remained our molecule of choice.

A potential substitute of 4sU could have been EU due to its low toxicity, similar rate of incorporation and commercial availability of a nascent RNA isolation kit using EU (Tani & Akimitsu, 2012; Invitrogen. Catalog number: C10365.).

4.3.2 200 µM of 4sU, 30 minutes of labelling, HPDP biotin and isopropanol/NaCl precipitation were found to be the optimal conditions for the isolation of nascent RNA from HEK293T and H1 hESCs

The protocol for the isolation of nascent RNA was optimised in HEK293T cells by testing different concentrations of 4sU (figure 4.2(A)) and labelling times (figure 4.2(B)). There did not seem to be a difference in the amount of purified nascent RNA when 200 µM and 500 µM 4sU were used, but labelling with 200 µM yielded significantly more RNA than using 100 µM, so it was the concentration of choice in further experiments. Cells were labelled with 200 µM for thirty minutes, one hour and two hours and the increase in the amount of isolated nascent RNA was proportional to the length of labelling time (figure 4.2(B)). Since rapidly synthesised and degraded transcripts were the focus of this project (see chapter 5), 30 minutes was chosen as the labelling time used for the isolation of H1 hESC nascent RNA.

Precipitating RNA with isopropanol in the presence of 10% 5M NaCl and 45 µg of glycogen yielded 2.47 times more nascent RNA compared to purifying it with the RNeasy MinElute Cleanup Kit (Qiagen) and it was the method of choice when samples were prepared to be tested by RT-PCR or microarray.

Duffy et al. reported that using MTSEA biotin alleviates transcript length bias (the preferential isolation of long transcripts) induced by other molecules like HPDP biotin (Duffy et al., 2015). Despite being more efficient at crosslinking to RNA (figure 4.5(B)), MTSEA biotin was not specific to 4sU-labelled RNA, evident from the presence of RNA in the negative control fraction isolated from cells grown in the absence of 4sU (figure 4.6(A)). The lower specificity of MTSEA biotin was further confirmed by the amplification of primary miRNA transcripts when the negative control sample was used as a template for cDNA and subsequent RT-PCR (figure 4.6(B)).
Lowering the concentration of MTSEA biotin as well as reducing the biotinylation time removes some of the non-specific binding of MTSEA biotin to unlabelled RNA, but it remains problematic (L. Dölken, personal communication). Therefore, HPDP biotin was the molecule of choice in further experiments.

The Agilent Technologies 2200 TapeStation provides a precise quantification of RNA and was used to determine the amount of pre-existing RNA isolated in the final step of the protocol (figure 4.7(A)). The concentration of RNA in the negative sample was 0.105 ng/µl. Performing 35 cycles of RT-PCR detected primary miRNA transcripts in the nascent RNA sample derived from 4sU-labelled cells, but not in the negative control sample (figure 4.7(B)). OCT4 was amplified in the sample isolated from cells to which 4sU was not added. However, when the number of RT-PCR cycles was reduced to 25 OCT4 was not detected which confirmed that the amount of RNA in the negative fraction was negligible.
Chapter 5- Microarray gene expression analysis of total and nascent RNA in H1 and H293T cell lines

5.1 Introduction

5.1.1 Background

Embryonic stem cells are enriched with a more ‘open’ chromatin, known for the predominance of euchromatin over heterochromatin (Wan et al., 2013). The presence of ‘bivalent’ domains characterised by both active and repressive histone modifications at the promoters of genes regulating development poising them for activation upon a response to differentiation signals has been established in this cell type (Guenther et al., 2010). Furthermore, undifferentiated cells are known for the higher rate of proliferation compared to differentiated cells, which tend to be arrested in the G0 phase of the cell cycle (Cooper, 2000).

Previous work in our group showed that nearly 40% of tested miRNAs are transcribed yet degraded in H1 hESCs (Rupa Sarkar, unpublished, chapter 1). This corroborates published data demonstrating extensive blocking of miRNAs in mESCs at the posttranscriptional level (Thomson et al., 2006). I confirmed that this type of regulation is more prevalent in H1 hESCs than in HEK293T cells for intergenic miRNAs and showed that the primary transcripts of the selected miRNAs are actively turned over despite the lack of detection of their mature form (chapter 3).

To get an insight into transcription in hESCs and the more differentiated HEK293T cells total and nascent RNA samples from both cell lines were isolated and gene expression was compared by microarray analysis. The aim of the experiment was to test if more genes are abundantly transcribed and degraded in H1 cells compared to HEK293T cells. The reasoning behind this was to obtain a clue to the nature of the mechanism involved in blocking miRNA processing in hESCs. Significantly more abundant transcription of a big proportion of all tested protein coding transcripts in H1 cells compared to HEK293T cells would indicate that a pathway responsible for the regulation of mRNA expression is also implicated in the regulation of pri-miRNA synthesis and degradation in this cell type. By contrast, if other genes
were found to be transcribed at similar levels in both cell types, this might imply that a miRNA-specific mechanism is in place to induce the block of miRNA processing in H1 cells.

5.1.2 Hypothesis and aims

A more open chromatin has been reported in ESCs in comparison to differentiated cells and their genome has been shown to be transcriptionally globally hyperactive and to undergo large-scale silencing as cells differentiate (5.1.1, (Efroni et al., 2008)). This work was based on total RNA levels and did not account for RNA degradation. We hypothesised that transcription in hESCs occurs in a more vigorous manner than in differentiated cells, thus leading to a high-level expression of a large fraction of the genome, and that pre-mRNAs are degraded quickly before they are processed into mature transcripts. If proven right, this could explain the successful amplification of pri-miRNAs in total RNA from the H1 cell line and the lack of detection of mature miRNAs by deep sequencing and would indicate that a mechanism involved in mRNA transcription also regulates pri-miRNA expression (chapter 3). To enable the analysis of transcripts synthesised within a short period of time as a way to investigate the number and abundance of transcripts on a global level we isolated nascent and total RNA fractions from H1 and HEK293T cells.

The aim of this experiment was to test if synthesis and degradation of a bigger number and of a higher abundance of transcripts occur in hESCs from the H1 cell line in comparison with HEK293T cells. This was achieved by isolating nascent and total RNA from both cell lines and performing a microarray experiment using the Illumina HT-12 array.
5.2 Results

5.2.1 The microarray controls worked well and all samples used in the experiment were included in the data analysis.

We analysed total and nascent RNA samples isolated from H1 and HEK293T cell lines with 3 replicates of each, giving 12 samples in total. Nascent RNA was purified from cells labelled with 200 µM 4sU for 30 minutes (chapter 4).

Assessing the microarray data quality is an important step prior to using it for gene expression analysis. Following the guidance on evaluating the sample-independent and sample-dependent controls provided by Illumina any outlier samples can be identified and removed from the analysis, if necessary (see appendix).

Sample-dependent and sample-independent controls included in the Illumina HT-12 array indicated that the data is of good quality, as seen in the control summary plots in the appendix (figure S5.1). Most of the controls generated relative intensities as expected (figure S5.2). However, the high stringency control probe produced a low signal contrary to the predicted intensity (figure S5.1, see Discussion).

For the proper analysis of microarray data it is important that biological replicates have similar gene expression profiles and that any outliers are excluded. To check if the biological replicate samples isolated for our experiment cluster together, a dendrogram was generated using the GenomeStudio software (figure 5.1). Its purpose is to show how closely correlated samples are based on their expression profiles (Illumina, 2010). The higher the correlation value
between samples, the closer to 0 on the x-axis the node joining them is because the position of the node is determined by the formula (1-correlation value (r)).

![Dendrogram](image)

**Figure 5.1** A dendrogram showing the correlation between the 12 RNA samples used in the microarray experiment. A joining node positioned closer to 0 means the samples are more similar and have higher correlation values.

Biological replicates clustered together and therefore all samples were kept in the data analysis despite the lower signal intensity of samples 1 and 3 from hESC nascent RNA (figure S5.4, see Discussion).

5.2.2 H1 hESCs transcribe and process a similar number of genes to HEK293T cells

Figure 5.2 shows the distribution across cell and RNA type of the total of 26359 probes which gave a positive signal (detection p-value <0.05) from all 47231 probes on the array. The number of probes detected is also presented as a percentage from all probes with a detection p-value <0.05. The analysis was performed in GenomeStudio and average normalisation was applied to the data (Illumina, 2008).
Figure 5.2 A Venn diagram showing the number of microarray probes (also expressed as a percentage) with a detection p-value <0.05 in total and nascent RNA samples in H1 hESCs and HEK293T cells.

More than half of the probes with a positive signal (16604 or 63%) were detected in all four types of samples. There were 3054 probes (11.6%) uniquely expressed by total and/or nascent RNA in hESCs and similarly, 2260 (8.6%) in HEK293T cells (figure 5.2). Overall a comparable number of probes gave a positive signal in both cell types - 24099 in hESCs and 23305 in HEK293T cells (figures 5.2 and 5.3). To control for any biases introduced by normalising the data this step was excluded from the analysis pipeline. Omitting the average normalisation step returned the same number of probes with a positive signal distributed in the same way as in the Venn diagram in figure 5.2.

Venn diagrams showing gene expression within cell types were created (figure 5.3(A, B)). In hESCs 18790 of the probes had a detection p-value <0.05 in both total and nascent RNA samples and likewise a total of 19192 probes were detected in both total and nascent samples.
in HEK293T cells (figure 5.3(A, B)). Notably, some gene transcripts were detected in the total but not nascent RNA fractions (figure 5.3(A,B), see Discussion).

![Venn diagrams showing the number of microarray probes with a detection p-value <0.05 in total and nascent RNA samples in H1 hESCs and HEK293T cells (A and B).](image)

Figure 5. 3 Cell type-specific Venn diagrams showing the number of microarray probes with a detection p-value <0.05 in total and nascent RNA samples in H1 hESCs and HEK293T cells (A and B).

To determine if more genes are transcribed at a higher rate in nascent RNA with a reference to total RNA in hESCs compared to HEK293T cells, three differential gene expression analyses were conducted in GenomeStudio. In the first one HEK293T nascent RNA was compared to HEK293T total RNA, in the second- hESC nascent RNA was compared to hESC total RNA and in the third- hESC nascent RNA was compared to HEK293T nascent RNA (table 5.1).
Table 5.1 Numbers of significantly up and downregulated genes in HEK293T and H1 hESCs nascent RNA in comparison to cell type-specific total RNA samples (columns 2 and 3). Column 4 shows the results from differential gene expression between nascent RNA samples of H1 and HEK293T cells. Average normalisation and the Illumina custom error model with false discovery rate (FDR) correction were applied to the data in all analyses.

In the first two analyses genes expressed by both total and nascent RNA were compared in a cell type-specific manner (table 5.1, columns 2 and 3) and in the third one nascent RNA samples were compared between the two cell types for protein coding transcripts detected in both of them (table 5.1, column 4). Similar numbers of probes showed significant differences in expression in all three comparisons (table 5.1). In HEK293T cells 6762 protein coding transcripts were significantly upregulated in nascent RNA compared to total RNA and 5417 were downregulated. In H1 hESCs these numbers were 6263 and 5460, respectively. From the 18224 probes with a detection p-value <0.05 in both HEK293T and H1 hESCs nascent RNA 4551 were significantly upregulated in HEK293T cells and 2483- in H1 hESCs (table 5.1 column 4).

There were 576 miRNA probes on the Illumina HT-12 microarray chip. Their detection in the four types of samples is presented in table 5.2. The majority of miRNAs were not detected in any sample (table 5.2).
<table>
<thead>
<tr>
<th></th>
<th>H1 hESC</th>
<th>HEK293T cells</th>
<th>H1 hESCs and HEK293T cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total and nascent RNA samples</td>
<td>43</td>
<td>38</td>
<td>30</td>
</tr>
<tr>
<td>Total RNA sample only</td>
<td>24</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>Nascent RNA sample only</td>
<td>22</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>Not detected</td>
<td>487</td>
<td>498</td>
<td>474</td>
</tr>
</tbody>
</table>

Table 5. 2 Detection of all 576 miRNA probes available on the Illumina HT-12 microarray in total and nascent RNA samples in H1 and HEK293T cells.

Probes for 23 out of the 24 miRNAs investigated in Chapter 3 featured on the microarray chip. A positive signal was recorded for miR-345 in total and nascent RNA samples from both H1 and HEK293T cells. None of the other 22 miRNAs were detected in either total or nascent RNA extracted from hESCs and HEK293T cells. A positive signal was detected for let-7a-1 in the nascent RNA fraction of HEK293T cells. MiR-302a was not detected in any RNA fraction derived from H1 cells.

5.2.3 Microarray data confirmed the identity of H1 hESCs by showing they are enriched for stem cell-specific genes

To determine if the H1 hESCs used in our experiment were in an undifferentiated state we selected genes encoding for proteins known to be involved in the maintenance of pluripotency and genes expressed by differentiated cell types and checked their expression in H1 and HEK293T cells. The data is presented in table 5.3.
Table 5. 3 Differential expression of cell specific genes. Microarray readout values are presented for total and nascent RNA samples in H1 and HEK293T cells. Numbers in red represent a detection p-value >0.05, indicating that this gene is not expressed. Average normalisation and the Illumina custom error model with false discovery rate (FDR) correction was applied to the data. All values presented for DNMT3B are derived from probes specific to isoform 1.

Readout values coloured in red denote probes with a detection p-value >0.05, indicating that these genes are not expressed. Transcripts for the pluripotency-associated genes POU5F1, SOX2, NANOG, CDKN1A, CDH1, LIN28 and DNMT3B were detected only in or at a significantly higher level in H1 hESCs, as expected, whereas transcripts for the genes expressed in differentiated cell types HOXA5, HOXA9, XIST, BMI1, KLF9 and CDKN2A were more specific to HEK293T cells (table 5.3).

To further confirm the identity of the H1 cells used in our experiment Ingenuity pathway analysis (IPA, Qiagen) was performed with the genes whose expression was most highly upregulated in hESC total RNA compared to total RNA from HEK293T cells (data not shown). A total of 610 genes that were at least three times more highly expressed in hESCs than in HEK293T cells were submitted to IPA. A summary bar chart showing the pathways most highly up- and downregulated in H1 cells is presented in figure S5.5 in the appendix. The most highly upregulated pathway in hESCs was the one maintaining pluripotency in embryonic stem cells.
by OCT4, as seen in figure S5.5. It is the pathway with the highest positive z-score, denoting upregulation (figure S5.5, appendix), which corroborated the data on selected pluripotency-associated genes whose expression was found to be significantly upregulated in hESCs compared to HEK293T cells (table 5.3).

A diagram of the genes involved in this pathway that were part of the list submitted to IPA is shown in figure 5.4. Upregulated expression of the genes from the core transcriptional circuitry maintaining the pluripotent state of embryonic stem cells, such as SOX2, OCT4 and NANOG is visible (coloured in pink), indicative of the undifferentiated state of H1 cells used in the microarray experiment (figure 5.4).
A list of all 610 genes whose expression was at least three times higher in H1 compared to HEK293T cells was submitted to IPA. A summary chart of all pathways in which these genes are involved is presented in figure S5.5 in the appendix. Highlighted in pink are genes that are highly expressed in H1 compared to HEK293T cells involved in the pluripotency network, which were part of the list of genes with most highly upregulated expression in hESCs.

RT-PCR was performed to validate microarray data. Genes expressed in only one cell type were selected for this experiment and their specificity was determined by their detection p-values (table 5.4). Red colour was used to mark transcripts with a detection p-value>0.05, indicating that these genes are not expressed.
<table>
<thead>
<tr>
<th></th>
<th>H1 hESCs total RNA</th>
<th>H1 hESCs nascent RNA</th>
<th>HEK293T total RNA</th>
<th>HEK293T nascent RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD24</td>
<td>10006.7</td>
<td>9610.7</td>
<td>99.7</td>
<td>140.5</td>
</tr>
<tr>
<td>CD24</td>
<td>1204.9</td>
<td>613.8</td>
<td>92.9</td>
<td>113.1</td>
</tr>
<tr>
<td>DPPA4</td>
<td>5147.9</td>
<td>3192.2</td>
<td>89.4</td>
<td>108.4</td>
</tr>
<tr>
<td>LIN28</td>
<td>6867.1</td>
<td>4412</td>
<td>88</td>
<td>107.3</td>
</tr>
<tr>
<td>L1TD1</td>
<td>3770.7</td>
<td>3065.2</td>
<td>100.8</td>
<td>111.9</td>
</tr>
<tr>
<td>CDX2</td>
<td>96.6</td>
<td>140</td>
<td>631</td>
<td>1427.4</td>
</tr>
<tr>
<td>FOXF2</td>
<td>97.1</td>
<td>133.8</td>
<td>559.7</td>
<td>1102.6</td>
</tr>
<tr>
<td>HOXA5</td>
<td>86.2</td>
<td>120.4</td>
<td>2672.7</td>
<td>10346.2</td>
</tr>
<tr>
<td>HOXA9</td>
<td>97</td>
<td>129.2</td>
<td>1350.1</td>
<td>3883.4</td>
</tr>
<tr>
<td>HOXA9</td>
<td>86.1</td>
<td>104.4</td>
<td>121</td>
<td>178.2</td>
</tr>
<tr>
<td>GAPDH</td>
<td>6706.3</td>
<td>2484</td>
<td>4011.1</td>
<td>2225.8</td>
</tr>
<tr>
<td>GAPDH</td>
<td>11700.9</td>
<td>6659.6</td>
<td>9960.6</td>
<td>5706.7</td>
</tr>
<tr>
<td>GAPDH</td>
<td>11274.8</td>
<td>4810.7</td>
<td>7638.3</td>
<td>4158.9</td>
</tr>
</tbody>
</table>

**Table 5. 4 Cell specific gene expression was shown by microarray analysis.** Numbers in red represent a detection p-value > 0.05, indicating that this gene is not expressed. Average normalisation and the Illumina custom error model with false discovery rate (FDR) correction was applied to the data.

GAPDH was used as a positive control in all samples and 25 cycles of PCR were performed (figure 5.5). Products of the expected sizes were purified from the gel and sequenced to confirm their identity. *CD24, DPPA4, LIN28* and *L1TD1* were uniquely detected in hESCs, whereas *HOXA5* and *HOXA9* were only detected in HEK293T samples (figure 5.5(A and B)). *FOXF2* which is significantly upregulated in HEK293T cells (table 5.4 columns 5 and 6) was also detected in H1 hESC samples at a lower level, as seen from this product’s reduced band intensity in this cell line (figure 5.5(A)). *CDX2*, significantly upregulated in HEK293T cells, was not detected in either cell type (figure 5.5(A and B)). No contamination with pre-existing RNA was observed when samples isolated from DMSO-treated cells were used in the RT-PCR reactions (figure 5.5(C)).
<table>
<thead>
<tr>
<th>A</th>
<th>H1 hESCs nascent RNA</th>
<th>H1 hESCs total RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT+ RT-</td>
<td>Gene</td>
<td>RT+ RT-</td>
</tr>
<tr>
<td>CD24</td>
<td>[Image]</td>
<td>CD24</td>
</tr>
<tr>
<td>DPPA4</td>
<td>[Image]</td>
<td>DPPA4</td>
</tr>
<tr>
<td>LIN28</td>
<td>[Image]</td>
<td>LIN28</td>
</tr>
<tr>
<td>L1TD1</td>
<td>[Image]</td>
<td>L1TD1</td>
</tr>
<tr>
<td>CDX2</td>
<td>[Image]</td>
<td>CDX2</td>
</tr>
<tr>
<td>FOXF2</td>
<td>[Image]</td>
<td>FOXF2</td>
</tr>
<tr>
<td>HOXA5</td>
<td>[Image]</td>
<td>HOXA5</td>
</tr>
<tr>
<td>HOXA9</td>
<td>[Image]</td>
<td>HOXA9</td>
</tr>
<tr>
<td>GAPDH</td>
<td>[Image]</td>
<td>GAPDH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>HEK293T nascent RNA</th>
<th>HEK293T total RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT+ RT-</td>
<td>Gene</td>
<td>RT+ RT-</td>
</tr>
<tr>
<td>CD24</td>
<td>[Image]</td>
<td>CD24</td>
</tr>
<tr>
<td>DPPA4</td>
<td>[Image]</td>
<td>DPPA4</td>
</tr>
<tr>
<td>LIN28</td>
<td>[Image]</td>
<td>LIN28</td>
</tr>
<tr>
<td>L1TD1</td>
<td>[Image]</td>
<td>L1TD1</td>
</tr>
<tr>
<td>CDX2</td>
<td>[Image]</td>
<td>CDX2</td>
</tr>
<tr>
<td>FOXF2</td>
<td>[Image]</td>
<td>FOXF2</td>
</tr>
<tr>
<td>HOXA5</td>
<td>[Image]</td>
<td>HOXA5</td>
</tr>
<tr>
<td>HOXA9</td>
<td>[Image]</td>
<td>HOXA9</td>
</tr>
<tr>
<td>GAPDH</td>
<td>[Image]</td>
<td>GAPDH</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th>H1 hESCs DMSO sample</th>
<th>HEK293T DMSO sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT+ RT-</td>
<td>Gene</td>
<td>RT+ RT-</td>
</tr>
<tr>
<td>CD24</td>
<td>[Image]</td>
<td>CD24</td>
</tr>
<tr>
<td>DPPA4</td>
<td>[Image]</td>
<td>DPPA4</td>
</tr>
<tr>
<td>LIN28</td>
<td>[Image]</td>
<td>LIN28</td>
</tr>
<tr>
<td>L1TD1</td>
<td>[Image]</td>
<td>L1TD1</td>
</tr>
<tr>
<td>CDX2</td>
<td>[Image]</td>
<td>CDX2</td>
</tr>
<tr>
<td>FOXF2</td>
<td>[Image]</td>
<td>FOXF2</td>
</tr>
<tr>
<td>HOXA5</td>
<td>[Image]</td>
<td>HOXA5</td>
</tr>
<tr>
<td>HOXA9</td>
<td>[Image]</td>
<td>HOXA9</td>
</tr>
<tr>
<td>GAPDH</td>
<td>[Image]</td>
<td>GAPDH</td>
</tr>
</tbody>
</table>

Figure 5. 5 Validating microarray data by RT-PCR. Cell- and sample-specific gene expression was confirmed by performing 25 cycles of RT-PCR in: A. Nascent and total RNA samples from H1 hESCs. B. Nascent and total RNA samples from HEK293T cells. C. Negative control samples isolated from H1 and HEK293T cells treated with DMSO instead of 4sU in the purification of nascent RNA experiment.
5.2.4 Microarray may not provide high enough sensitivity for the detection of genes expressed at a low level

To further test the reliability of the microarray data RT-PCR was performed to check if transcripts with a detection p-value >0.05 can be amplified. The selected transcripts are presented in table 5.5. Red-coloured readout values represent detection p-values >0.05, indicating that these genes are not expressed.

<table>
<thead>
<tr>
<th></th>
<th>H1 hESCs total RNA</th>
<th>H1 hESCs nascent RNA</th>
<th>HEK293T total RNA</th>
<th>HEK293T nascent RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICDA</td>
<td>91.2</td>
<td>126.1</td>
<td>101.2</td>
<td>121.2</td>
</tr>
<tr>
<td>FIT1</td>
<td>94.6</td>
<td>125</td>
<td>100.3</td>
<td>121.5</td>
</tr>
<tr>
<td>HPX</td>
<td>94.8</td>
<td>127.1</td>
<td>100</td>
<td>121.6</td>
</tr>
<tr>
<td>IL1RN</td>
<td>99.3</td>
<td>132.1</td>
<td>103.1</td>
<td>125</td>
</tr>
<tr>
<td>IL1RN</td>
<td>88.1</td>
<td>125.3</td>
<td>95</td>
<td>119.4</td>
</tr>
<tr>
<td>NECAB1</td>
<td>83.1</td>
<td>109.2</td>
<td>93.1</td>
<td>100.6</td>
</tr>
<tr>
<td>OR2A2</td>
<td>100</td>
<td>127.8</td>
<td>97.8</td>
<td>118.9</td>
</tr>
<tr>
<td>POPDC2</td>
<td>92.9</td>
<td>124.1</td>
<td>100.5</td>
<td>119.1</td>
</tr>
<tr>
<td>RP1L1</td>
<td>92.9</td>
<td>129.6</td>
<td>98</td>
<td>118.4</td>
</tr>
<tr>
<td>SSTR3</td>
<td>93.3</td>
<td>128.5</td>
<td>97</td>
<td>121</td>
</tr>
<tr>
<td>SYCP2L</td>
<td>92</td>
<td>129.3</td>
<td>96.8</td>
<td>118.6</td>
</tr>
<tr>
<td>GAPDH</td>
<td>6706.3</td>
<td>2484</td>
<td>4011.1</td>
<td>2225.8</td>
</tr>
<tr>
<td>GAPDH</td>
<td>11700.9</td>
<td>6659.6</td>
<td>9960.6</td>
<td>5706.7</td>
</tr>
<tr>
<td>GAPDH</td>
<td>11274.8</td>
<td>4810.7</td>
<td>7638.3</td>
<td>4158.9</td>
</tr>
<tr>
<td>OCT4</td>
<td>2459.2</td>
<td>680.9</td>
<td>93.2</td>
<td>106.7</td>
</tr>
</tbody>
</table>

Table 5. 5 Readout values for 10 transcripts not detected by microarray (detection p-value>0.05). Numbers in red represent a detection p-value>0.05, indicating that this gene is not expressed. Average normalisation and the Illumina custom error model with false discovery rate (FDR) correction was applied to the data.

When 35 cycles of RT-PCR were performed using nascent RNA samples as a template transcripts for 8 out of the 10 selected genes from table 5.5 were detected in H1 hESCs and transcripts for 6 out of the 10 genes were detected in HEK293T cells (figure 5.6(A), the product for SSTR3 in HEK293T cells was found to be non-specific after sequencing). HPX and SYCP2L were detected in the negative control sample of H1 hESCs derived from DMSO-treated cells.
(figure 5.6(B)). HPX and GAPDH were amplified from the HEK293T negative control sample (figure 5.6(B)).

Figure 5.6 Validating microarray data by RT-PCR. 35 cycles of RT-PCR were performed to screen for genes for which transcripts were not detected in the microarray analysis (detection p-value > 0.05). Experiment was performed using: A. H1 hESC nascent RNA and HEK293T nascent RNA. B. H1 and HEK293T negative control samples treated with DMSO instead of 4sU in the purification of nascent RNA experiment.
If the number of PCR cycles was reduced to 25, transcripts for two genes were detected in each total and nascent RNA sample from both cell lines, indicative of the low expression level of the genes successfully amplified in figure 5.6 (figure 5.7(A and B)). There were no bands present in the negative control samples for the gene transcripts which were detected in nascent RNA samples at 25 cycles of RT-PCR (figure 5.7(C)). HPX and SYCP2L were detected in HEK293T total and nascent RNA samples as well as in H1 hESCs nascent RNA and H1 hESCs were found to express NECAB1 and SYCP2L in total RNA (figure 5.7(A and B)). Overall, the amplification of gene transcripts not detected by the microarray is suggestive of the low sensitivity of this technique.
Figure 5. 7 Validating microarray data by RT-PCR. 25 cycles of RT-PCR were performed to screen for gene transcripts which were not detected in the microarray analysis (detection p-value > 0.05). Experiment was performed using: A. H1 hESC and HEK293T nascent RNA. B. H1 hESC and HEK293T total RNA. C. H1 hESC and HEK293T negative control (DMSO-treated) for genes detected by RT-PCR of nascent RNA samples. We screened for gene transcripts which had a detection p-value > 0.05 (table 5.5).
5.3 Discussion

5.3.1 No significant differences were found at the steady-state transcript level between H1 and HEK293T cells

A comparable number of gene transcripts were detected in the two cell lines by the Illumina HT-12 microarray-24099 and 23305 in H1 hESCs and HEK293T, respectively (figures 5.2 and 5.3). 63% (16604) of all probes with a positive signal (26359) were detected in all four types of samples (figure 5.2). H1 hESCs were found to uniquely express 11.6% of all probes in total and/or nascent RNA samples, which was similar to the 8.6% only detected in HEK293T cells. The number of genes expressed by H1 hESCs was similar to data generated by an experiment in mouse embryonic stem cells where ~20000 gene transcripts were detected by DNA microarray in total RNA isolated from two different stem cell lines (Sharova et al., 2009).

Previous studies have reported the low sensitivity of microarrays in terms of detecting global changes in transcription due to normalisation steps which assume similar RNA levels across samples (Lin et al., 2012). Despite absolute quantification of mRNA transcripts not being an aim of this experiment, we took care with the type of normalisation used when interpreting the data because our hypothesis was that more genes were expressed in nascent RNA samples of H1 hESCs. Our microarray data was analysed after applying average normalisation in GenomeStudio (Illumina, 2008). Even though this type of normalisation does not assume equal transcript abundance across samples, it adjusts the average signal intensity of individual arrays to the global average of all array signals (Illumina, 2008). Removing this step from the analysis returned the same number of probes with a detection p-value <0.05 as presented in figure 5.2 and confirmed the overall similarity in the number of transcribed genes between the two cell lines.

Furthermore, differential gene expression analysis showed that the expression of similar numbers of genes was upregulated in nascent RNA samples compared to total RNA within cell lines (table 5.1). The same was true when looking into genes whose expression was downregulated in nascent RNA compared to total RNA fractions. The expression of 6762 genes was upregulated in HEK293T nascent RNA and the expression of 6263 genes was upregulated in nascent RNA from H1 cells when compared to cell specific total RNA samples (table 5.1).
Differential expression analysis of nascent RNA from H1 and HEK293T cells showed that more gene transcripts were downregulated (4551) than upregulated (2483) in H1 hESCs compared to HEK293T cells (table 5.1, column 4). Only genes expressed by both samples were analysed in each differential analysis and most of them (11190) were not significantly different between the two cell lines.

Our results do not corroborate data from a study showing that mouse ESCs are more open to transcription and express 20% more genes than neural progenitor cells (NPCs) derived from them (Efroni et al., 2008). However, a whole-genome tiling array was used in this study, which also covers contiguous and intergenic regions of the genome, thus enabling a more thorough analysis of transcription. By contrast, our study is limited by the use of the Illumina HT-12 array enriched for probes for well-characterised genes (Illumina, 2011).

Overall, our data indicates that H1 and HEK293T cells express a similar fraction of the genome to maintain their steady-state mRNA levels. This finding suggests that global transcriptional output is not different in H1 cells than in more differentiated cell lines and widespread abundant gene expression cannot account for the detection of pri-miRNAs in this cell type (chapters 1 and 3).

5.3.2 The technical performance of the microarray was good and the H1 cells used in the experiment were in an undifferentiated state

The 12 samples used in the microarray experiment were similar in terms of the number of gene transcripts detected within each one (figure S5.4(A), appendix). All of the sample-dependent and most of the sample-independent controls included with the array performed as per manufacturer’s recommendations (figure S5.1 and S5.2, appendix, (Illumina, 2010)). The only exception was the high stringency control, which yielded a low intensity signal even though it was expected to be high. Since the low stringency, hybridisation and biotin controls worked well and the high stringency control only consists of one probe, we considered that the microarray performed well technically.

Two out of the three H1 hESC nascent RNA samples had an overall lower signal intensity compared to the other 10 samples (figure S5.4(B,C and D), appendix). However, all biological
replicates clustered together and the two nascent RNA samples isolated from H1 hESCs with a lower overall signal (samples 1 and 3) clustered more closely between themselves than they did with the third one (sample 2), as expected (figure 5.1). The quality controls indicated the microarray experiment worked well and all samples were included in the data analysis.

The undifferentiated state of H1 cells and the differentiated state of HEK293T cells were supported by microarray and RT-PCR analysis. *POU5F1, SOX2* and *NANOG* are at the core of the regulatory circuitry that maintains the pluripotency and self-renewal properties of embryonic stem cells and were significantly upregulated in H1 cells, as shown in table 5.3 (Boyer et al., 2005). *LIN28* which selectively blocks the processing of miRNAs from the let-7 family required for differentiation was only detected in H1 cells (Hagan, Piskounova & Gregory, 2009; Viswanathan, Daley & Gregory, 2008). *DNMT3B* is a DNA methyltransferase which performs *de novo* DNA methylation. Isoform 1 of *DNMT3B*, characteristically expressed by undifferentiated hESCs (Gifford et al., 2013), was found to be significantly upregulated in H1 cells compared to HEK293T cells (table 5.3). All values presented for *DNMT3B* in table 5.3 are derived from probes specific to isoform 1.

By contrast, *HOX* genes (*HOXA5* and *HOXA9*) which are involved in the patterning of the body axis during development and are not expressed in pluripotent cells (Atkinson et al., 2008) were found to be transcribed only in HEK293T cells (table 5.3). The non-coding *XIST* RNA whose role is to perform X chromosome inactivation in female cells (Weakley et al., 2011) was expressed at a significantly higher level in HEK293T cells which are of female origin (Stepanenko & Dmitrenko, 2015). The detection of *XIST* in H1 cells, which are of male origin and were not found to express *XIST* by PCR (Hoffman et al., 2005) was close to background levels (table 5.3, figure S5.1). *KLF9*, which is upregulated upon differentiation (Ying et al., 2011) was significantly upregulated in HEK293T cells in our data (table 5.3).

Ingenuity Pathway Analysis (Qiagen, 2017) confirmed the pluripotent state of the H1 hESCs cell line used in the microarray analysis by showing that the gene transcripts most highly upregulated in stem cells are involved in pathways related to maintaining their undifferentiated state (figures 5.4, S5.5).

RT-PCR confirmed the cell type-specific gene expression as indicated by microarray data (table 5.4, figure 5.5). Genes uniquely expressed in both total and nascent RNA samples of
either H1 or HEK293T cells according to microarray analysis were selected for RT-PCR (table 5.4). Apart from LIN28, CD24, a surface marker expressed by embryonic stem cells, L1TD1, a pluripotency marker and DPPA4 which regulates the differentiation of embryonic stem cells into primitive ectoderm, were the stem cell-specific genes detected by RT-PCR (Amini et al., 2014; Närvä et al., 2012; Naujok & Lenzen, 2012). HOXA5 and HOXA9 as well as CDX2, a caudal-related homeobox transcription factor family member, and FOXF2, a transcription factor important for embryogenesis and organ development, were selected as differentiation-associated genes (Zhang et al., 2015; Niwa et al., 2005). CDX2 was not detected by RT-PCR in HEK293T cells despite the expected expression of this gene and FOXF2 was amplified in both cell types even though microarray data indicated that it was only expressed in HEK293T cells (table 5.4, figure 5.5). However, it was harder to detect it in H1 cells compared to HEK293T cells (figure 5.5).

5.3.3 The microarray sensitivity is limited

There was a large number of genes found to only be expressed by either total or nascent samples within cell types (figure 5.3). H1 cells had 3638 genes specific to total RNA and 1671-to nascent; HEK293T cells were found to express 1842 genes only in total RNA and 2271- only in nascent RNA (figure 5.3(A and B)). If a gene is only expressed in nascent RNA, it is possible that it is not turned into mature RNA due to its rapid degradation. However, if a gene is expressed as mature RNA, it must also be transcribed as pre-mRNA and should therefore be present in nascent RNA samples. The readout values for gene transcripts only detected in one type of sample were relatively low in both total and nascent RNA samples and most of them were close to the overall background signal intensity, as measured from negative probes included in the array (figure S5.1), indicative of the low sensitivity of the microarray (results not shown).

A low number of miRNAs was detected in both cell lines by microarray (table 5.2). The majority of the 576 miRNA probes had a detection p-value >0.05 which is most likely due to the bias towards longer transcripts introduced upon the isolation of 4sU-labelled nascent RNA (chapter 4). Out of the 23 miRNAs initially detected by RT-PCR (Chapter 3) for which there was a probe available on the microarray only miR-345 was detected in total and nascent RNA.
samples from both H1 and HEK293T cells. Even the abundantly expressed miR-302a in H1 cells was not detected in this experiment. Let-7a-1 was only detected in the nascent RNA fraction of HEK293T cells. This indicated the low sensitivity of the microarray. Therefore, no clear conclusion can be drawn about the miRNA biogenesis regulation or expression in either cell line based on this experiment. Gene transcripts which were not detected by microarray were successfully amplified by 35 cycles of RT-PCR, which further indicates that the sensitivity of the array is low (table 5.5, figure 5.6). Eight out of the ten tested gene transcripts were amplified in nascent RNA from hESCs and six- in nascent RNA from HEK293T cells (figure 5.6). There was slight contamination with pre-existing RNA evident in the DMSO-treated samples, but only in four of the reactions (figure 5.6). Reducing the number of cycles to 25 significantly decreased the number of detected gene transcripts in both cell lines, which confirmed that they are not expressed at a high level (table 5.5, figure 5.7).

Efroni et al. found that the tiling array used in their study was not sensitive enough to detect tissue-specific genes expressed at a low level by stem cells, despite their successful amplification by RT-PCR, similar to our findings (Efroni et al., 2008).
Chapter 6 - Optimising the pull-down of specific miRNAs

6.1 Introduction

6.1.1 Biases in miRNA representation in deep sequencing analysis

Deep sequencing is widely used to study the expression of miRNAs in various biological systems on a genome-wide scale. It offers the advantages of higher specificity and sensitivity compared to microarrays and qPCR (Zhang et al., 2013). Furthermore, it does not rely on the design of specific probes or primers and hybridisation and it can be used for the discovery of novel miRNA sequences (Zhang et al., 2013). This technique also works at high resolution and is a useful tool for distinguishing between closely related miRNAs and miRNA isomiRs which differ by small numbers of base pairs due to the additions or deletions at their 5’ and/or 3’ ends (Neilsen, Goodall & Bracken, 2012).

Performing small RNA sequencing can be achieved using different sequencing platforms, but they share a similar pipeline of sample preparation (Hafner et al., 2011). It involves preparing libraries after the isolation of small RNAs from cells. To do this adapters are annealed sequentially to the 3’ and 5’ ends of RNAs in enzyme-catalysed reactions and reverse transcription (RT) and PCR are performed to amplify them (Hafner et al., 2011).

Despite the fact that deep sequencing is sensitive to low abundance transcripts (Chiang et al., 2010; Landgraf et al., 2007), libraries generated in this manner have been shown to be biased towards representing certain miRNAs. It has been noted that this bias is mostly introduced during the sample preparation for sequencing rather than during the isolation of RNA, RT, PCR or sequencing itself (Zhuang et al., 2012; Sorefan et al., 2012; Hafner et al., 2011; Jayaprakash et al., 2011; Linsen et al., 2009). The incorrect representation of miRNAs has been noted to differ by a 1000-fold in some cases due to the poorer efficiency of cloning of some miRNAs (Zhang et al., 2013; Linsen et al., 2009).

It is important to accurately quantify the expression of specific miRNAs because, depending on the relative expression of a miRNA compared to its mRNA target, a miRNA can either work as a molecular switch or as a fine regulator of mRNA expression (Mukherji et al., 2011).
Different approaches have been tried to improve the ligation of adapters to miRNAs. Some ligation biases resulting in miRNA overrepresentation were found to be due to certain sequences co-folding with the adapter; others have been found to occur because of the generation of undesired concatamers or circular products generated through the ligation of two or more adapter or RNA molecules or by adapter self-ligation (Jackson et al., 2014; Viollet et al., 2011).

Advances in enzyme biochemistry and improvements in ligation strategies have been achieved as a solution to this problem (Jackson et al., 2014; Zhuang et al., 2012; Hafner et al., 2011; Jayaprakash et al., 2011). T4 RNA ligases 1 and 2 (T4 Rnl1 and T4 Rnl2) are used to ligate 3′-OH to 5′-PO₄ ends of RNAs and adapters in a three-step reaction sequence involving the adenylation of a 5′ phosphorylated nucleic acid donor (Viollet et al., 2011). A truncated form of T4 Rnl2 (T4 Rnl2tr) is less prone to ligating 5′-PO₄ ends together than T4 Rnl1 and it is ten times more active than the full-length T4 Rnl2 when ligating 5′ adenylated oligonucleotide adapters to 3′ ends of RNAs (Viollet et al., 2011; Ho et al., 2004). A T4 Rnl2tr with a point mutation in its C-terminal domain responsible for the transfer of AMP from the active site of the enzyme to the 5′-PO₄ of the donor has been shown to reduce the number of non-specific side-products. This was found to be due to impairing the reverse reaction of 5′-PO₄ donor adenylation, i.e. transferring AMP from adenylated adapters to 5′-PO₄ RNA molecules, resulting in unwanted products when binding to 3′-OH RNA ends (Viollet et al., 2011).

However, despite the improvements in ligation specificity, miRNA library generation was still found to be biased towards certain miRNAs because of structural changes resulting from RNAs co-folding with adapters (Zhuang et al., 2012; Hafner et al., 2011). Using a thermostable 3′ RNA ligase isolated from *Methanobacterium thermoautotrophicum* in an attempt to reduce the amount of co-folding did not alleviate bias in small RNA libraries (Jackson et al., 2014).

Another approach to reducing ligation bias was to introduce short (2-4 nt) randomised sequences at the 5′ end of the 3′ adapter and at the 3′ end of the 5′ adapter (Sorefan et al., 2012; Jayaprakash et al., 2011). It was observed that different miRNAs show different sequence preferences to the 5′ ends of 3′ adapters (Zhang et al., 2013). The last study described a library preparation technique which is reportedly nearly 100% efficient in ligating both 3′ and 5′ adapters. The improvement was due to the introduction of randomised
sequences in the adapters, optimisation of their concentration, addition of PEG8000 and optimising the annealing temperature (Zhang et al., 2013).

Here we propose a method for the isolation of specific miRNAs which enables the identification of isomiRs without the costly and time consuming deep sequencing analysis. A schematic of the technique is presented in figure 6.1. It involves a pull-down assay of miRNAs ligated to adapters at both 3’ and 5’ ends via a 5’ biotinylated DNA oligonucleotide complementary to a miRNA of interest and subsequent magnetic purification with the aid of streptavidin-coated beads. When combined with the high efficiency method for adapter ligation proposed by Zhang et al., subsequent reverse transcription and PCR will be able to distinguish isomiRs when loaded on a high percentage acrylamide gel (Zhang et al., 2013). Sanger sequencing can be performed to obtain the exact sequences of the isomiRs. In the experiments presented here radioactively $[^\gamma-32P]$ ATP- labelled miR-101-1 and DNA oligonucleotides with highly similar sequences to miR-101-1 were used to optimise the pull-down of specific miRNAs.

6.1.2 Rationale and aims

IsomiR switching of miRNAs has been shown by Northern blotting and analysis of deep sequencing data. Furthermore, in vitro work has confirmed the functional significance of different 5’ variants of miR-9 (Tan et al., 2014). A 5’ isomiR of miR-101 has been shown to be expressed and functional in different cell types (Llorens et al., 2013). Improving the cloning efficiency of miRNAs will shed more light onto the importance of isomiR expression of this miRNA.

The aim of these experiments was to demonstrate the successful isolation of $[^\gamma-32P]$ ATP-labelled miR-101-1 oligonucleotide by using a complementary biotinylated DNA oligonucleotide followed by magnetic purification and to test the specificity of this technique.
Figure 6. 1 Schematic of the proposed pull-down technique for the isolation of specific miRNAs. miRNAs which have undergone adapter ligation on both 3’ and 5’ ends are incubated with a 5′ biotinylated ssDNA (ssDNA*) complementary to a miRNA of interest (microRNA*). The double stranded DNA-RNA complexes including all non-specific adapter-ligated miRNAs are incubated with streptavidin-coated magnetic beads which bind to the 5′ end of the DNA oligonucleotide carrying a biotin modification. Following magnetic separation RT, PCR and sequencing of the PCR product can be performed to confirm the specific sequences of both the canonical and all isomiRs of any chosen miRNA. B- biotin, S- streptavidin.
6.2 Results

The outline of the technique used to isolate a specific radioactively-labelled oligonucleotide is presented in figure 6.2.

![Diagram of the protocol]

Figure 6.2 Schematic of the protocol used for the pull-down of $[\gamma$-32P] ATP-labelled miRNA oligonucleotides.

First the Miltenyi Biotec μMACS streptavidin kit used for the isolation of nascent RNA (see chapter 4) was tested in this experiment. RNA Oligonucleotides with the same sequence as known miRNAs (miR-101-1 RNA and let-7a-1 DNA oligonucleotides) were radioactively labelled and annealed to a DNA oligonucleotide complementary to miR-101-1. The latter was modified to carry a Biotin-TEG tag either on its 5’ end or on both its 5’ and 3’ ends to test if the number of biotin tags changes the efficiency of the reaction. After annealing to biotinylated DNA oligonucleotides, the radioactive probes were magnetically separated using streptavidin-coated beads and were loaded on a gel (figure 6.3).
Figure 6. 3 Pull-down of [$\gamma$-32P] ATP- labelled miR-101-1 RNA oligonucleotide using the Miltenyi Biotec µMACS streptavidin kit. A DNA let-7a-1 oligonucleotide was used as a negative control (lanes 2 and 4). A single (5' end) and a double (5' and 3' end) biotinylated DNA oligonucleotide complementary to miR-101-1 were used to pull down radioactive probes. [$\gamma$-32P] ATP- labelled oligonucleotides were used as molecular weight markers (lanes 5 and 6).

Signal was detected only in the reactions containing [$\gamma$-32P] ATP- labelled miR-101-1 and not let-7a-1 (figure 6.3). Both types of biotinylated oligonucleotides (5' and 5'/3') led to the recovery of miR-101-1 and the signal intensities from both reactions were comparable. To confirm that the signal was of the correct size a fraction of the radioactively labelled miR-101-1 and let-7a oligonucleotides were loaded on the gel (figure 6.3(lanes 5 and 6)).

To reduce the cost of this technique the experiment was repeated using New England Biolabs streptavidin magnetic beads. A signal of the correct size was only detected when radioactively labelled miR-101-1 DNA oligonucleotide was used (figure 6.4). A complementary DNA oligonucleotide with a single biotin molecule (at the 5' end) was similar in its efficiency to pull down labelled miR-101-1 probe to a double-biotinylated oligonucleotide (figure 6.4(lanes 1 and 2)). The 5' biotinylated oligonucleotide was used in the rest of the optimisation reactions.
Figure 6. 4 Pull-down of [γ-32P] ATP- labelled miR-101-1 DNA oligonucleotide using the New England Biolabs streptavidin magnetic beads (1-2). A DNA let-7a-1 oligonucleotide was used as a negative control (3-4). A single (5’ end) and a double (5’ and 3’ end) biotinylated DNA oligonucleotide complementary to miR-101-1 were used to pull down radioactive probes. [γ-32P] ATP- labelled oligonucleotides were used as molecular weight markers (lanes 5 and 6).

To test whether the pull-down of miR-101-1 oligonucleotide would work in conditions where specificity of the annealing reaction might be abrogated by the presence of abundant transcripts the experiment was performed in the background of 0.45 μg of let-7a-1 oligonucleotide or 3 μg of total RNA from HEK293T cells (figure 6.5). Serial dilutions of the radioactively labelled miR-101-1 oligonucleotide were used to determine the efficiency of the pull-down in the two types of background and upon varying the ratios of non-specific to specific transcripts.
Figure 6. 5 Pull-down of [γ-32P] ATP- labelled miR-101-1 DNA oligonucleotide using the New England Biolabs streptavidin magnetic beads in the background of 0.45 µg of let-7a-1 DNA oligonucleotide (lanes 1-4) or 3 µg of HEK293T total cell RNA (lanes 5-8). Serial 10-fold dilutions of the [γ-32P] ATP- labelled miR-101-1 oligonucleotide were performed. 5’ biotinylated complementary DNA oligonucleotide was used to pull down [γ-32P] ATP- labelled miR-101-1. [γ-32P] ATP- labelled miR-101-1 was used as a molecular weight marker (lane 9).

To further verify the specificity of the biotinylated DNA oligonucleotide the annealing reaction was performed in the presence of oligonucleotides sharing a high percentage of similarity with miR-101-1. The Fuzzy Lookup Add-in for Excel was used to identify the most highly similar miRNAs to miR-101-1 from all naturally occurring miRNAs in humans. The search retrieved four miRNAs which are more than 50% similar to miR-101-1, 63% being the lowest and 85% the highest percentage of similarity (figure 6.6).

<table>
<thead>
<tr>
<th>microRNA</th>
<th>% similarity to miR-101-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;hsa-miR-141-3p MIMAT0000432</td>
<td>83%</td>
</tr>
<tr>
<td>UAACACUGUCUGGUAAAGAUGG</td>
<td></td>
</tr>
<tr>
<td>&gt;hsa-miR-542-3p MIMAT0003389</td>
<td>85%</td>
</tr>
<tr>
<td>UGUGACAGAUUGUAACUGAAA</td>
<td></td>
</tr>
<tr>
<td>&gt;hsa-miR-4797-5p MIMAT0019972</td>
<td>63%</td>
</tr>
<tr>
<td>GACAGAGUCCACUACUGAAA</td>
<td></td>
</tr>
<tr>
<td>&gt;hsa-miR-7161-5p MIMAT0028232</td>
<td>83%</td>
</tr>
<tr>
<td>UAAAGACUGUAGGCAACUGGU</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6. 6 MiRNAs with a high similarity to miR-101-1. The Fuzzy Lookup Add-in for Excel was used to identify miRNAs from the complete list of miRNAs occurring in the human species which share over 50% similarity with miR-101-1. Four miRNAs were returned with a range of similarity 63-85%.
The pull-down experiment was performed with all four oligonucleotides highly similar to miR-101-1 (figure 6.7). [γ-32P] ATP- labelled oligonucleotides were loaded on the gel to control for size. Signal was detected only when miR-101-1 oligonucleotide was used.

![Figure 6.7 Pull-down of [γ-32P] ATP- labelled miR-101-1 and [γ-32P] ATP- labelled DNA oligonucleotides with highly similar sequences to miR-101-1. The eluate from each reaction (labelled ‘pull-down’) and the [γ-32P] ATP- labelled oligonucleotides (labelled ‘probe’) used in each reaction were loaded on the gel. [γ-32P] ATP- labelled miR-101-1 was successfully purified. No signal was detected from the rest of the reactions after overnight exposure of the gel.](image)

During the magnetic purification of miRNA probes three washes are performed before the final elution step - two low stringency washes (with high salt buffer) and one high stringency wash (with low salt buffer). To test if the specific isolation of miR-101-1 in figure 6.7 was due to sequence specificity or it was conferred by the washes the experiment was repeated by varying the number and types of washes for the two miRNAs most highly similar to miR-101-1, miR-141 and miR-542 (figure 6.8). Three types of washing were performed: two low stringency washes and one high stringency wash (all washes), two low stringency washes (high salt washes) or none of them (no washes) (figure 6.8(lanes 5-10)). The product from the
miR-101-1 pull-down in which all washes were performed was used as a positive control. The radioactive probes were run on the gel to control for size (figure 6.8, lanes 1-3).

![Figure 6.8](image)

**Figure 6. 8 Pull-down of [γ-32P] ATP- labelled miR-101-1 and [γ-32P] ATP- labelled DNA oligonucleotides with highly similar sequences to miR-101-1 (miR-141 and miR-542).** [γ-32P] ATP- labelled oligonucleotides used in each reaction were loaded as molecular weight markers (labelled ‘probe’, lanes 1-3). A miR-101-1 reaction in which all washes were performed was used as a positive control (lane 4). Two high salt washes and one low salt wash (all washes), two high salt washes or no washes were performed in miR-141 and miR-542 reactions (lanes 5-10).

A strong signal was detected from the miR-101-1 reaction (figure 6.8(lane 4)). A weak signal was observed in the lanes containing purified products from miR-141 and miR-542 reactions in which no washes were performed (figure 6.8(lanes 9 and 10)). A faint band was visible after eluting the product from the miR-542 reaction in which only the low stringency washes were performed (lane 8).

The efficiency of the pull-down technique was determined over a range of annealing temperatures by analysing band intensities from all tested conditions (figure 6.9, table 6.1). The image was inverted, as recommended by ImageJ, to obtain the most accurate measurements of band intensities (ImageJ, 2017). A reaction which was not incubated with streptavidin-coated magnetic beads was used as a control to which all other intensities were compared (figure 6.9, table 6.1).
Figure 6. 9 Measuring the efficiency of miR-101 pull-down under different annealing temperatures. A. Image showing bands generated by the pull-down products from the reactions performed under different conditions (see table to the right of A and B). In 1 the reaction was set up without proceeding with bead incubation and was used as a control for measuring the efficiency of the pull-down in the other conditions. All reactions were performed using a complementary DNA oligonucleotide biotinylated at the 5’ end and denatured for 5 minutes at 75°C with the exception of 5, marked with *. The arrow indicates the position of [γ-32P] ATP- labelled miR-101-1 oligonucleotide used as a size control (not shown). B. Inverted image used to measure the intensity of the bands in ImageJ, as recommended in the dot blot analysis protocol of ImageJ (ImageJ, 2017).

We followed the protocol for analysing dot blots in ImageJ and obtained values for the area of the band, the mean, minimum and maximum grey value intensities as well as a value for integrated density, calculated as a product of area and mean grey value (ImageJ, 2017). We used the control reaction (1 in figure 6.9 and table 6.1) as a representation of 100% efficiency and calculated the efficiency of the rest of the reactions as a percentage from 1 using the integrated density values (table 6.1). The highest efficiency (15.7%) was achieved when annealing the complementary biotinylated DNA oligonucleotide at 45°C (table 6.1).
<table>
<thead>
<tr>
<th>Area</th>
<th>Mean grey value</th>
<th>Min grey value</th>
<th>Max grey value</th>
<th>Integrated density</th>
<th>% from 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>32122</td>
<td>246.624</td>
<td>130</td>
<td>255</td>
<td>7922060</td>
</tr>
<tr>
<td>2</td>
<td>4105</td>
<td>219.935</td>
<td>190</td>
<td>246</td>
<td>902832</td>
</tr>
<tr>
<td>3</td>
<td>1048</td>
<td>147.986</td>
<td>133</td>
<td>169</td>
<td>155089</td>
</tr>
<tr>
<td>4</td>
<td>5124</td>
<td>242.926</td>
<td>111</td>
<td>255</td>
<td>1244754</td>
</tr>
<tr>
<td>5</td>
<td>2574</td>
<td>171.193</td>
<td>115</td>
<td>192</td>
<td>440652</td>
</tr>
<tr>
<td>6</td>
<td>5134</td>
<td>185.227</td>
<td>147</td>
<td>211</td>
<td>950957</td>
</tr>
<tr>
<td>7</td>
<td>6193</td>
<td>183.801</td>
<td>151</td>
<td>217</td>
<td>1138277</td>
</tr>
</tbody>
</table>

**Table 6.1 Analysis of band intensity from figure 6.9 in ImageJ.** Selecting the brightest fields from the bands, as seen in figure 6.9 (B), generated information about the area in square pixels, mean grey value, minimum and maximum grey levels and integrated density (the product of area and mean grey density). The efficiency of the reactions presented in figure 6.9 was calculated by comparing the values of integrated density between the different bands and presenting them as percentage from reaction 1.
6.3 Discussion

Using both the Miltenyi Biotec µMACS streptavidin kit as well as the New England Biolabs streptavidin magnetic beads led to the successful purification of [γ-32P] ATP- labelled miR-101-1 (figures 6.3 and 6.4). The specificity of the reaction was confirmed by the lack of signal when [γ-32P] ATP- labelled let-7a-1 oligonucleotide was used instead of miR-101-1 (figure 6.3(lanes 2 and 4), figure 6.4(lanes 3 and 4)).

Increasing the number of biotin-TEG tags on the complementary DNA oligonucleotide used to anneal to radioactively-labelled probes did not change the efficiency of the pull-down (figures 6.3 and 6.4). An oligonucleotide modified to carry a single biotin molecule on its 5’ end was used for the rest of the experiments.

To further test the specificity and efficiency of this technique it was performed in the presence of either let-7a-1 oligonucleotide or HEK293T total RNA (figure 6.5). Performing the pull-down of miR-101-1 in both conditions resulted in purifying products with similar signal intensities.

Four human miRNAs are more than 50% similar to miR-101-1 by the Fuzzy Lookup Add-in for Excel (figure 6.6). Even though three of them were over 80% similar to miR-101-1, only radioactively-labelled miR-101-1 was efficiently pulled down (figure 6.7).

Most of the specificity of the technique was ensured by the sequence of the probe rather than by the washing steps (figure 6.8). When no washing steps were performed during the magnetic purification step in pull-down reactions with miR-141 and miR-542, only a weak signal was detected (figure 6.8, lanes 9 and 10). When all or only the low stringency washes were performed there was no signal detected in any reaction (figure 6.8, lanes 5-7). The only exception was miR-542, where a faint signal was observed in the pull-down performed with two low stringency washes (figure 6.8, lane 8). MiR-542 is the most highly similar miRNA to miR-101-1 (85%) and this result is likely to account for it.

Despite the lack of or minimum signal in the cases where all washes or two high salt washes were performed it is possible that a fraction of the biotinylated oligonucleotide can be bound to miRNAs different from miR-101-1. To ensure maximum specificity in future experiments all washes will be performed.
The highest efficiency of the pull-down (15.7%) was achieved by performing the annealing step with the complementary biotinylated DNA oligonucleotide at 45°C, as measured in ImageJ (figure 6.9 and table 6.1). This is the temperature nearest to the one used in the previous optimisation steps (43°C) which was calculated according to the formula provided by the Miltenyi Biotec protocol for the pull-down of specific transcripts (Materials and methods, 2.7.2).

A pull-down efficiency of 15.7% would probably be enough to amplify a sufficient amount of product by RT-PCR to determine the isomiR composition of a specific miRNA.

Future work will involve optimising adapter annealing as proposed by Zhang et al. and RT-PCR (Zhang et al., 2013). Testing the efficiency of the biotinylated oligonucleotide when attempting a pull-down of longer oligonucleotides (miRNAs annealed to adapters) will be performed before proceeding to RT-PCR amplification. Another possibility is designing a longer biotinylated oligonucleotide which is complementary to one of the adapters as well as the miRNA of interest. Applying the method proposed here in cell lysates may lead to the need of further optimisation due to the lower concentrations of miRNAs of interest. However, we have shown that the pull-down assay is highly specific and efficient enough to encourage future experiments. This technique will save time and will reduce costs when projects are focused on a limited number of miRNAs.
Chapter 7- General Discussion

In this project we showed that post transcriptional blocking in miRNA processing is more prevalent in hESCs than in HEK293T cells. We detected primary transcripts by RT-PCR for 15 out of the 24 tested miRNAs whose mature form was not detected or was present at a low level in H1 cells as shown by deep sequencing data. HEK293T cells were only found to express primary transcripts for 2 out of the 16 miRNAs for which the mature form was not sequenced at all or was sequenced at a low level (table 3.1). We demonstrated that these pri-miRNAs are actively transcribed by performing RT-PCR using RNA samples from actinomycin D-treated cells and nascent RNA samples isolated after 30 minutes of 200 µM 4sU labelling (figures 3.6, 3.9 and 3.10). Using nascent RNA we studied transcription in H1 and HEK293T cells. Our microarray data suggests that transcription is not wasteful in H1 cells, i.e. the number of vigorously transcribed genes is similar between hESCs and HEK293T cells (figures 5.3, 5.4). We thus excluded the possibility that gene expression in H1 cells is dysregulated on a global scale which would lead to disproportionate transcription from most parts of the genome.

miRNAs are regulated at the post transcriptional level in a wide range of biological systems (chapter 1). Members of the let-7 family are known to be transcribed but blocked for processing in ESCs (Thornton & Gregory, 2012). In addition, it has been reported that ESCs transcribe but do not process miRNAs from several other families, which was demonstrated by real-time RT-PCR designed to amplify both the primary and mature forms of these miRNAs in a single reaction (Thomson et al., 2006). The ratios of expression of primary to mature transcript were similar between let-7 family members and 5 out of 9 other tested miRNAs. Thomson et al. analysed mostly pri-miRNA transcripts which overlap with mRNA genes by real-time RT-PCR, thus not distinguishing the two. However, previous work in our group accounted for the genomic organisation of miRNAs and showed by comparing the expression of primary transcripts (by performing 5’ and 3’ RACE PCR for intragenic miRNAs and RT-PCR for intergenic miRNAs) and sequencing data for the mature miRNAs that 40% of all tested miRNAs are blocked in processing in hESCs compared to the more differentiated NPSCs (Rupa Sarkar, unpublished).
Primary transcripts of miRNA genes which are not processed to the mature form in H1 cells were found to be expressed at a similar level to pri-miRNAs which are processed and sequenced at a high level in this cell type. This corroborated our data showing active transcription of these pri-miRNAs (figures 3.5, 3.6 and 3.9).

It is possible that due to poor cloning efficiency (Zhang et al., 2013) certain mature miRNAs might be less readily detected by sequencing. However, this does not explain the marked high percentage of unprocessed miRNA transcripts that we detect for hESCs compared to other cell types (chapter 3). We assumed that the lack of detection of mature miRNAs for which primary transcripts were successfully amplified is not due to a rapid turnover of the mature miRNAs in hESCs. MiRNAs are usually very stable and for this reason are used as biomarkers in blood samples (Brase et al., 2010; Mitchell et al., 2008). Furthermore, the minority of unstable miRNAs, such as those involved in cell cycle control can still be detected by sequencing (Rissland, Hong & Bartel, 2011).

Published data indicated that ESCs are in a state of hypertranscription, that their chromatin is in a more open state than in differentiated cells and that low levels of cell lineage specification genes can be detected in primed ESCs (Gaspar-Maia et al., 2011; Efroni et al., 2008; Meshorer et al., 2006). Furthermore, this data was derived from stable RNA fractions and so did not account for the possible contribution of RNA synthesis that might be masked by degradation (Efroni et al., 2008).

We compared nascent RNA fractions from hESCs and HEK293T cells in a microarray experiment to check if pluripotency entails vigorous transcription of an extensive part of the transcriptome (chapter 5). Despite the possibly low sensitivity of the microarray selected for our work, it still allowed the detection of abundant transcripts. We observed similar numbers of expressed genes in the two cell types in both total and nascent RNA (figure 5.2). Furthermore, a similar number of genes was found to be upregulated in nascent RNA compared to total RNA in each cell type as well as when nascent RNA samples from H1 and HEK293T cells were compared to each other, which further supported the idea that vigorous transcription is limited to a fraction of the genome in both cell lines (table 5.1).

Our microarray data did not provide an absolute quantification of mRNA levels in total and nascent RNA fractions, but it seems unlikely that we have missed detecting a global
upregulation of transcriptional output from hESCs. As previous reports have shown, global elevation in transcription can be easily missed due to limitations in analysis methods (Percharde, Bulut-Karslioglu & Ramalho-Santos, 2016; Lin et al., 2012). However, hypertranscription does not result in an equal elevation in transcription across the whole genome or in a linear increase in expression of all actively transcribed genes (Percharde, Bulut-Karslioglu & Ramalho-Santos, 2016). Therefore, any significant elevation in transcription of a fraction of the genome would be easily detected by microarray. This statement corroborates our data which showed minor differences in the number of mRNA genes expressed by hESCs and the more differentiated HEK293Ts. The microarray results indicate that it is unlikely that pri-miRNAs that are blocked in processing in H1 cells are expressed due to vigorous transcription of a big fraction of the genome in these cells and raised the possibility that a more specific miRNA-dedicated mechanism might be operating to achieve this.

RNA binding proteins have been demonstrated to block the processing of precursor miRNA molecules by DROSHA and/or DICER in a variety of biological settings (chapter 1). A recent study performed a large-scale in vitro screen of interactions between miRNA hairpins and proteins (Treiber et al., 2017). 72 miRNA precursors were immobilised and incubated with lysates from 11 different cell lines, amongst which was the NTERA-2 teratocarcinoma cell line. Surprisingly, despite the in vitro approach, most of the interactions were highly specific, with one or a small subset of proteins interacting with each miRNA hairpin in a cell-type-specific manner (Treiber et al., 2017). Nearly 90% of the identified proteins were annotated RNA-binding proteins. Kwon et al. defined the mRNA interactome of mESCs and reported that it is composed of 555 proteins, from which 294 were preferentially expressed in mESCs compared to differentiated cell types (Kwon et al., 2013). 151 proteins were found to be RNA-binding proteins. It is thus possible that the pri-miRNA block in processing in hESCs is mediated by RNA-binding proteins at least in some cases.

The sensitivity of the microarray was not very high and indeed genes not expressed according to the analysed data were successfully amplified by RT-PCR (table 5.5, figure 5.6). This led us to speculate that the detection of pri-miRNAs which are blocked in processing might also be due to a low level of transcription resulting from the more open chromatin in hESCs. It seems possible that such low level of transcription might influence our miRNA transcription results.
because it is conceivable that miRNA genes are in general relatively weakly transcribed compared to most genes, as indicated by the relatively high number of PCR cycles required to detect them as well as the fact that they are underrepresented in RNA-Seq data sets (Chang et al., 2015; Assis et al., 2014). It has been reported that there is a low level of background transcription in ESCs compared to more differentiated cell types (Efroni et al., 2008). To test if this property may account for the discrepancy in expression we see for miRNA genes, it might be informative to use RT-PCR to analyse other genes whose transcription is likely less than average. If our hypothesis is correct, the transcripts from such genes should be more easily detected in nascent RNA from hESCs than HEK293T cells.

The latter could also be addressed by altering the duration of 4sU labelling for the isolation of nascent RNA and selecting earlier time points. Shorter labelling time combined with RNA-seq would reveal differences in the rates of RNA degradation between the two cell lines. This approach would also be more likely to result in the detection of genes synthesised at a high rate in case they get degraded within the 30 minute time point that we used.

Lastly, it would be interesting to explore the possibility of miRNA expression regulation by antisense transcription. Antisense transcription has been known to occur for an extensive part of the mammalian genome (Engström et al., 2006; Katayama et al., 2005). Oct4, Sox2 and Nanog have been reported to produce antisense transcripts in mESCs (Zhao et al., 2014). Furthermore, the Sox2 antisense transcript has been found to regulate the expression of SOX2 at both the protein and mRNA level, raising the possibility that antisense transcription might be involved in regulating pluripotency (Zhao et al., 2014). Performing strand-specific RT-PCR or strand-specific RNA-seq using nascent and total RNA for hESCs might reveal if pri-miRNAs are transcribed in just one or both directions and whether bidirectional transcription is a feature of miRNAs whose primary transcripts are transcribed but not processed (Li et al., 2013; Ho, Donaldson & Saville, 2010).
References


153


Appendix

Sample-independent controls are oligonucleotides spiked into the hybridisation solution and measure the technical performance of the microarray experiment. They consist of hybridisation controls, low stringency controls and biotin and high stringency controls (figures S5.1, S5.2). There are 6 hybridisation control probes found at three different concentrations on the array (high, medium and low) with corresponding Cy3-labelled oligonucleotides present in the hybridisation buffer, thus yielding a gradient hybridisation response. There are 4 low-stringency control probes with two mismatches (MM2) within their sequences corresponding to the Cy3-labelled hybridisation controls and are present at medium and high concentrations. The purpose of the low-stringency control is to test the stringency of the hybridisation. If it is low, the signal intensity generated from the MM2 probes is similar to their perfect match (PM) counterparts used as hybridisation controls. The biotin controls consist of two probes complementary to biotin-tagged oligonucleotides in the hybridisation buffer, which are detected by secondary staining. There is one probe part of the high stringency control with a high G+C content, which should anneal to its complementary oligonucleotide even in high-stringency conditions. The aim of this control is to detect highly stringent hybridisations which would be evident if it is the only one producing a signal out of all the sample-independent controls.

Sample-dependent metrics controls consist of negative controls, gene intensity (housekeeping genes) and artificial RNA spike probes, which control for labelling. Approximately 800 negative control probes designed so that they are not complementary to any part of the genome are used to define the non-specific background signal level. The latter could result from non-specific binding of dye or non-specific hybridisation of the samples. There are 14 probes for housekeeping genes, which should be expressed in any type of cell and are expected to yield a high intensity signal. There is an option to include artificial spike RNA probes (labelling) in each sample, but this step was omitted in our experiment.

To further test the specificity of the experiment PM and MM2 probe signals from the hybridisation controls were plotted for each sample (figure S5.3). The PM signal was consistently higher than the MM2 signal, as per manufacturer’s recommendations, which indicated that the hybridisation specificity is good across all samples (Illumina, 2010).
The GenomeStudio software provided by Illumina generates a detection p-value for each probe on the array, representing the probability that it is detected above background levels defined by the negative controls described above. For our analysis we selected 0.05 as a detection p-value threshold, so any value below that means that a gene is detected.

The number of genes with a p-value <0.05 was similar across the 12 samples used (figure S5.4(A)). However, when the 95th intensity percentile (p95), which is a measure of the high-end signal intensity, and signal-to-noise ratios calculated by dividing high-end to low-end intensity values and high-end intensity values to background intensity values of the negative controls (p95/p05 and p95/background) were plotted for each sample, it was obvious that two of the hESC nascent RNA samples (samples 4 and 6) had lower overall signal intensity compared to the other 10 (figure S5.4(B,C and D)).
Figure S5.1 Control summary plots showing the data quality generated by the Illumina HT-12 array. For guidance on the interpretation of the plots see figure S5.2.
Figure S5. 2 Quality control guidance notes on interpreting quality control plots from Illumina HT-12 array.
Figure S5. 3 Comparison between perfect match (PM) and mismatch (MM2) probe signals across the 12 samples used in the microarray experiment.
Figure S5. 4 Sample quality was assessed by creating line plots of control metrics for each of the 12 samples used in the microarray experiment. The plots represent: A. The number of detected genes. B. Signal-to-noise ratio measured by the ratio of high-end intensity signal (p95) to low-end intensity signal (p05). C. 95th intensity percentile (p95). D. Signal-to-noise ratio measured by the ratio of p95 to background signal.
Figure S5.5 Summary bar chart showing the most highly up and downregulated pathways in H1 hESCs compared to HEK293T cells. IPA analysis was performed with the 610 most highly upregulated genes in H1 cells. All genes included in the analysis were at least three times more highly expressed in H1 than in HEK293T total RNA.