Understanding and exploiting naïve human pluripotency



Benjamin Thomas Dodsworth

Sir William Dunn School of Pathology

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Abstract

Naïve human pluripotent stem cells (hPSC) represent an earlier stage in human development compared to conventional "primed" hPSC. They resemble the inner cell mass of an embryonic day (E) 5 human blastocyst and are considered developmentally equivalent to mouse embryonic stem cells (mESC). Access to this new cell type facilitates interrogation of early human development. In this thesis, I have shown that the Theunissen *et al.* protocol for generating naïve hPSC yields cells that most closely approximate the human E5 inner cell mass and is robustly reproducible across several induced pluripotent stem cell lines. miRNAs are essential in mouse embryogenesis for differentiation from the naïve to the primed state, but the miRNAome of the human E5 inner cell mass has yet to be determined. I have established the naïve to primed differentiation and identified the miRNAome of naïve, differentiating and primed human pluripotent stem cells by miRNAseq. This novel dataset has revealed that the largest difference in the miRNAome between mouse and the human is due to cell state rather than species. The mouse miR-290 family are the most highly expressed miRNAs in mESC and the absence of the human equivalent homologs miR-371-373 in conventional hPSC was thought to be a difference in species. Using the miRNAseq dataset, I have shown that miR-371-373 are highly upregulated in naïve hPSC, confirming that naïve hPSC more closely resemble mESC. I also observed high expression of miR-182-5p, which downregulates the homology directed repair (HDR) pathway. This surprising finding contrasts with previous publications which suggest that naïve hPSC genomes are more efficiently manipulated by HDR. In support of this finding, I show that naïve hPSC repair Cas9 induced double strand breaks less often by HDR than primed hPSC and demonstrate that a higher proportion of naïve hPSC are in G1 phase of the cell cycle which is likely a contributor to the lower rates of HDR.

Acknowledgements

First, I express my gratitude to Dr Sally Cowley. She has been the most incredible supervisor and her support has been a significant source of motivation throughout the DPhil. Sally has found the perfect and rare combination of both guiding and empowering her students. Should I find myself in a similar role, I will endeavour to emulate her approach.

I would like to extend my thanks to my supervisors Dr Rowan Flynn and Dr Claas Meyer, who have both been a tremendous support. I would particularly like to thank Claas for being my mentor in business and industry. My collaborators have also been of great help. I am grateful to Dr Klas Hatje for his excellent work and his patience with my requests. Many thanks to Dr Maria Rostovskaya for her help with the HNES lines and discussions. I must also thank my funders, the BBSRC and Roche, who made this DPhil possible.

I am also grateful to members, friends and alumni of the James lab. In particular, many thanks to Jane and Cathy for their patience in training me. Thanks to William James, Julian, Cecilia, Walther, Sharat, Hazel, Liv, Jackie, Kenny, Gwen, Alun, Kourosh and Becky for your coffee breaks, discussions and help. I am also thankful to the department and the GSA for creating an excellent working environment and to all my friends and colleagues at Roche who made my time there a fantastic experience.

My sanity is indebted to Anna, Tim & Raadiyah, as well as my Mum, Dad, Fritz, Hanne, Granny, Carolyn & Mike. Finally, I would like to take this opportunity to thank my dearest friends, who have stuck with me despite my minor preoccupation with a DPhil: The Darmstadters: Frieder, Arne, Kim, Thiemo, Andy, Marc, Patrick, Esther and Gordia. The Sheffielders: Sam N., Peter, Jamie, Amy & Sam D., Alex, Sam A. The Oxfordians: Fede, Chris, Neele, Lisa & David, Kirk, Javian, Kate, Polly, Peny, Roma. And the Baslers: Timo & Julia, Valerio, Baran, JP, Sarah, Tilly, Andy. And apologies to anyone I have forgotten – I just wrote a DPhil.

Table of contents

1 INTRODUCTION 1

1.1. Naïve and primed human pluripotent stem cells 1
1.1.1. Harnessing pluripotency 1
1.1.2. Naïve human pluripotent stem cells 3
1.1.3. Evidence for naivety 6
1.1.4. Naïve transcription factor network and gene expression 6
1.1.5. Generating naïve hPSC 7
1.1.6. Comparing naïve hPSC generation protocols 13
1.1.7. Naïve to primed differentiation 16
1.1.8. Summary 17

1.2. miRNAs 17
1.2.1. Introduction 17
1.2.2. miRNAs in development 19
1.2.3. *Dgcr8* and *Dicer1* knockout murine studies 21
1.2.4. Summary 23

1.3. Gene editing and DNA repair in PSC 23
1.3.1. Introduction 23
1.3.2. Homologous recombination and non-homologous end joining 24
1.3.3. Short single stranded DNA as repair templates 27
1.3.4. Summary 28

1.4. Motivation and aims 28

1.5. Summary of thesis 29

2 MATERIALS AND METHODS 31

2.1. Cell culture 31
2.1.1. Cell counts 31
2.1.2. Primed iPS cell culture 31
2.1.3. Preparing irradiated mouse embryonic fibroblasts 32
2.1.4. Naïve media 33
2.1.5. Naïve conversions & maintenance 35
2.1.6. Priming naïve cells 37
2.1.7. Freezing and thawing 37

2.2. Microscopy 37

2.3. Primers and probes 37

2.4. RT-qPCR 40

2.5. Karyotyping 42
2.5.1. Measuring copy number variation 42

2.6. Flow cytometry 43

2.7. miRNA analysis **45**
2.7.1. Total RNA extraction 45
2.7.2. miRNA qPCR 46
2.7.3. miRNAseq 48

2.8. Electroporation of mimics and inhibitors **48**

2.9. Assay to measure HDR & NHEJ **49**
2.9.1. Preparation of ribonucleoproteins (RNPs) 49
2.9.2. Electroporation of cells with RNPs 49
2.9.3. Droplet digital PCR 51

2.10. *DGCR8* knockout **52**
2.10.1. Generating the knockout 52
2.10.2. On plate DNA extraction 53
2.10.3. Screening by PCR 53
2.10.4. Cloning of lentiviral, dox inducible *DGCR8* 54

2.11. Statistical analysis **57**

3 GENERATING AND OPTIMISING THE CULTURE OF NAÏVE HUMAN PLURIPOTENT STEM CELLS 61

3.1. Introduction and aims **61**

3.2. Results **62**
3.2.1. Reviewing published RNAseq data of naïve conversion protocols 62
3.2.2. Comparison of nine naïve protocols 66
3.2.3. Gene expression profiling of cells generated by naïve protocols 70
3.2.4. Karyotype analysis 75
3.2.5. Optimisation of electroporation of naïve hPSC 78
3.2.6. Optimisation of flow cytometry markers for naïve vs. primed cells 80

3.3. Summary and discussion **82**
3.3.1. Comparison of published RNAseq data 82
3.3.2. Experimental comparison of naïve protocols 83

4 MIRNAS INVOLVED IN NAÏVE-PRIMED DIFFERENTIATION 86

4.1. Introduction **86**
4.1.1. Small non-coding RNAs 86
4.1.2. Aims 87

4.2. Results **88**
4.2.1. Human naïve hPSC can be differentiated to the primed state 88
4.2.2. miRNAseq reveals candidate miRNAs 95
4.2.3. Gene set enrichment analysis 97
4.2.4. Determining miRNAs of interest 101
4.2.5. Functional analysis of miRNAs 103
4.2.6. Future perspective: knockout of *DGCR8* 107

4.3. Summary and discussion **112**
4.3.1. Naïve to primed differentiation 112
4.3.2. ESCC miRNAs with AAGUGC seed 112
4.3.3. miRNAs in the naïve state 114

4.3.4. Differentiating and primed state miRNAs 115
4.3.5. Unchanged and highly expressed miRNAs 115
4.3.6. Expected but not expressed miRNAs 116
4.3.7. miRNA inhibition experiment 116
4.3.8. *DGCR8* knockout 118
4.3.9. Summary 118

5 GENE EDITING IN NAÏVE HPSC 120

5.1. Introduction 120
5.1.1. Gene editing in naïve hPSC 120
5.1.2. Droplet digital PCR for measuring rates of NHEJ and HR 121
5.1.3. Aims 122

5.2. Results 122
5.2.1. HDR and NHEJ can be measured by droplet digital PCR 122
5.2.2. HDR occurs only during the first 24h after electroporation of Cas9, guide RNAs and template oligo 126
5.2.3. The ratio of HDR vs. NHEJ repaired DNA is lower in naïve hPSC 127
5.2.4. The proportion of cells in G_1 phase are higher in naïve vs. primed cells 128

5.3. Summary and discussion 132
5.3.1. Cas9 kinetics 132
5.3.2. Proportion of HR repaired DSBs differs between naïve and primed hPSC 135
5.3.3. Regulation of HDR by cell cycle and miRNAs 137
5.3.4. Summary 137

6 GENERAL DISCUSSION 138

6.1. General discussion and future perspectives 138
6.1.1. Protocols to generate naive hPSC produce variable results 138
6.1.2. miRNA expression profiles distinguish naïve and primed pluripotent states and suggest close resemblance to mouse ESC 139
6.1.3. Naïve hPSC accumulate in G_1 stage of the cell cycle, affecting HDR 140
6.1.4. miRNAs affecting cell cycle and HR 141
6.1.5. Further directions 142
6.1.6. Expectations and hurdles for naïve hPSC 143

6.2. Concluding remarks 146

7 APPENDIX 148

7.1. Supplementary methods 148
7.1.1. Pseudotime analysis for single cell transcriptomics data of human preimplantation embryos integrated with bulk transcriptomics data of embryonic stem cells 148
7.1.2. miRNA processing 150
7.1.3. Gene set enrichment 151

7.2. Bibliography 169

List of Figures

Figure 1.1: Comparison of human and mouse embryogenesis. 4
Figure 1.2: Pathways involved in naïve hPSC and inhibitors used in naïve conversion protocols. 8
Figure 1.3: Strategies to convert to or derive naïve hPSC 10
Figure 1.4: Schematic of the miRNA biogenesis pathway showing *DGCR8* and *DICER1*. 18
Figure 1.5: DNA repair pathways and their regulation by the cell cycle. 25
Figure 3.1: Comparison of RNAseq data from published naïve protocols to in vivo embryogenesis single cell RNAseq datasets. 63
Figure 3.2: Representative morphology of cells treated with different naïve protocols. 71
Figure 3.3: Fold change in gene expression of marker genes in cells generated by different naive conversion protocols to their primed control. 72
Figure 3.4: The expression of 7 markers (*KLF4, TBX3, DNMT3L, NANOG, OCT4, CER1, OTX2*) relative to their housekeeping genes (*B2M, RPL13A*) were reduced to two constituents by principal component analysis. 74
Figure 3.5: Karyotypic abnormalities seen after converting to or deriving naïve hPSC 76
Figure 3.6: Optimising electroporation. 79
Figure 3.7: Validation of flow cytometry markers. 81
Figure 4.1: Gene expression during naive-primed differentiation 89
Figure 4.2: Effects of feeder-free culture on differentiation and survival. 91
Figure 4.3: Priming causes immediate gain of CD90 and slow acquisition of CD24 protein expression. 93
Figure 4.4: The miRNAome distinguishes naïve, differentiating, and primed cells. 96
Figure 4.5: Validation of miRNA expression by qPCR 98
Figure 4.6: The control miRNA mimic but not inhibitor have a dose-dependent relationship to their miRNA target gene. 105
Figure 4.7: Inhibition of miRNAs selected due to their upregulation during differentiation does not affect the differentiation gene expression pattern in these conditions. 108
Figure 4.8: DGCR8 knockout strategy. 109
Figure 4.9: Analysis of DGCR8 heterozygotes. 111
Figure 4.10: Summary miRNAs expressed in naïve and primed cells. 113
Figure 5.1: An assay to measure homology directed repair (HDR) and non-homologous end joining (NHEJ) by droplet digital PCR. 124
Figure 5.2: A time course showing emergence and repair of a Cas9 induced double strand break in iPSC shows repair is completed by 48 hours after electroporation. 125
Figure 5.3: DNA repaired by the HDR pathway is observed less often in naïve cells than primed. 129
Figure 5.4: The proportion of cells in G_1 phase are higher in naïve versus primed cells. 130
Figure 5.5: A summary of the findings of this chapter. 136

Supplementary Figure S1: Inducible DGCR8 vector design 152

Supplementary Figure S2: Alignment of naïve cells generated by different protocols across human development. 153
Supplementary Figure S3: Assessing genes suitable as endogenous controls for qRT-PCR. 154
Supplementary Figure S4: Comparison of gene expression between primed controls and naïve cells. 155
Supplementary Figure S5: Comparison of gene expression of unused protocols. 156
Supplementary Figure S6: Cells in a mixture of naïve medium and E8 differentiate and die. 158
Supplementary Figure S7: The miRNA endogenous controls are consistently expressed across the differentiation time course. 159
Supplementary Figure S8: Sequencing data correlates with qPCR results. 160
Supplementary Figure S9: Further analysis of miRNAseq data. 163
Supplementary Figure S10: Detailed design of probes. 164
Supplementary Figure S11: Cas9 induced double strand breaks were predominantly repaired by NHEJ. 165
Supplementary Figure S12: Relationship between %NHEJ and %HDR measurements allows calculation of a slope using linear regression. 167

List of Tables

Table 1.1: Comparison of naïve media. 11
Table 1.2: Evidence published for available protocols. 14
Table 2.1: List of TaqMan assays used in this thesis. 38
Table 2.2: List of probes, primers and oligos. 39
Table 2.3: List of antibodies and dilutions used in this thesis. 44
Table 2.4: List of miRNA inhibitors used. 44
Table 2.5: List of CRISPR guide RNAs used. 44
Table 2.6: List of reagents and suppliers. 58
Table 3.1: Gene weightings of principal components PC1 and PC2. 67
Table 4.1: *TBX3* is the only target gene, which has an enriched for differentiation day 2 upregulated miRNAs. 102
Table 4.2: miRNAs chosen for further testing have similar seed sequences. 104

Supplementary Table S1: Datasets used for in silico protocol comparison. 168

Abbreviations

2A	2A self-cleaving peptide
2i	CHIR99021 and PD0325901
2iL	2i + LIF
2iLF	2i + LIF + bFGF
alt-EJ	alternative end joining
AMO	Anti miRNA oligonuceotide
AO	Acridine orange
bFGF	basic fibroblast growth factor
CDKs	cyclin-dependent kinases
CHIR	CHIR99021
CHO	Chinese hamster ovary
C-NHEJ	canonical non-homologous end joining
DAPI	4′,6-diamidino-2-phenylindole
ddH_2O	double distilled water
ddPCR	droplet digital PCR
DiM	(S)-(+)-Dimethindene maleate
DMEM	Dulbecco`s Modified Eagle Media
dox	doxycycline
DSB	double strand break
dsDNA	double stranded DNA
E	embryonic day
E. Coli	Escherichia coli
EB	Embryoid body
EpiSC	epiblast stem cells
ES FBS	Embryonic Stem Cell Qualified Fetal Bovine Serum
ESCC	ESC-specific cell cycle–regulating
FDR	false discovery rate
FGFR1	FGF receptor 1
GMP	Good Manufacturing Practice
Go	Gö6983
GPCRs	G protein–coupled receptors
gRNA	guide RNA
GSEA	gene set enrichment analysis
HDAC	histone deacetylases
HDACi	inhibitor of histone deacetylases
HDR	homology directed repair
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HPRT1	hypoxanthine phosphoribosyltransferase 1
hPSC	human pluripotent stem cells
HR	homologous recombination
indels	insertions and deletions
iPSC	induced pluripotent stem cell

KO	knockout
KSR	Knockout Serum Replacement
LB	Lysogeny broth
LIF	Leukemia inhibitory factor
LNA	locked nucleic acid
logFC	log_2Fold Change
mEpiSC	mouse epiblast stem cells
MiH	Minocycline hydrochloride
miRNA	Micro RNA
MMEJ	microhomology mediated end joining
mPSC	mouse pluripotent stem cells
MRE	miRNA response element
NEAA	Non essential amino acids
NES	normalized enrichment score
NOD/SCID	Nonobese diabetic/severe combined immunodeficiency
NOM p-val	nominal p-value
nt	nucleotides
OSN	OCT4, SOX2 and NANOG
PARP	poly ADP ribose polymerase
PBS	Phosphate-buffered saline
PCA	principal component analysis
PD03	PD0325901
PD17	PD173074
PGCLC	primordial germ cell-like cells
PI	propidium iodide
PKC	Protein kinase C
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA transcript
PTO	Phosphorothioate Oligonucleotides
RFP	Red Fluorescent Protein
RNAseq	RNA sequencing
RNP	ribonucleoprotein
RPA	replication protein A complex
RT-qPCR	quantitative reverse transcription polymerase chain reaction
SB43	SB431542
SB59	SB590885
SI	Separation index
SP6	SP600125
SSA	single strand annealing
ssDNA	single stranded DNA
SU	SU5402
TLS	DNA translesion synthesis
UTR	untranslated region
Vitamin C	L-ascorbic acid 2-phosphate
XaXi	one active, one inactive X chromosome
Xist	X-inactive specific transcript

1 Introduction

1.1. Naïve and primed human pluripotent stem cells

1.1.1. Harnessing pluripotency

Pluripotency describes the ability of a cell to generate all lineages of a developing and adult organism. Pluripotent cells occur naturally, albeit transiently, in the blastocyst of early developing embryos. In 1981, the first pluripotent cells were cultured *in vitro*: Embryonic stem cells were derived from mouse blastocysts and retained their pluripotency. They could also self-renew indefinitely without the need for genetic transformation (Evans and Kaufman, 1981, Martin, 1981). Human embryonic stem cells (hESC) followed (Thomson et al., 1998), as well as induced pluripotent stem cells (iPSC) originating from somatic cells in both mouse and human (Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Yu et al., 2007). IPSC are derived by reprogramming somatic cells to a pluripotent state, which is achieved by exogenously introducing transcription factors which are highly expressed in conventional ESC, until the endogenous pluripotency machinery becomes autonomous (Takahashi and Yamanaka, 2006, Takahashi et al., 2007, Yu et al., 2007). Pluripotent stem cells (PSC) is an umbrella term referring to both ESC and iPSC.

Mouse epiblast stem cells (mEpiSC), resembling the postimplantation epiblast, were derived in 2007 and revealed that pluripotent stem cells exist in a spectrum of pluripotency, with mESC resembling an earlier stage than mEpiSC (Brons et al., 2007, Tesar et al., 2007). In addition, it was quickly realised that mEpiSC closely resemble hESC. When mEpiSC were successfully converted to

mESC, the question arose if human PSC (hPSC) are also able to be converted to an earlier 'naïve' stage (Guo et al., 2009).

The motivation for discovering new pluripotent states and the excitement about harnessing pluripotency comes from their many applications. Mouse embryonic stem cells can be genetically manipulated and reintroduced into a developing mouse blastocyst, efficiently creating transgenic mice. Knockouts of most protein coding genes have been achieved in mouse models, generating extraordinarily rich datasets of genotype-phenotype studies. These studies have vastly improved our knowledge of the mammalian genome and protein functions (Brown et al., 2018). Both human and mouse ESC are powerful tools for disease modelling and therefore also drug discovery. This is particularly due to their ability to indefinitely self-renew and differentiate into disease specific tissues. Disease modelling by genetic modification was first achieved by disruption of the *hypoxanthine phosphoribosyltransferase* (*HPRT*) gene, which allows modelling and study of the Lesch-Nyhan syndrome (Zwaka and Thomson, 2003, Urbach et al., 2004). In addition, iPSC technology has allowed the derivation of cell lines from patients with genetic diseases. Since it is possible to generate iPSC lines from several patients, results are more robust due to different genetic backgrounds. Disease causing mutations can be corrected or introduced and provide isogenic controls, which differ in a single genetic lesion (Avior et al., 2016). Finally, regenerative medicine and cell therapies are starting to make use of PSC, specifically their ability to generate vast numbers of cells and differentiate into the cell types and tissues of interest. Current clinical trials include cell replacement therapies for neurodegenerative diseases, macular degeneration, cardiac failure, and type I diabetes mellitus (Trounson and DeWitt, 2016), with many more disease areas expected to be added in the near future (Nguyen et al., 2018).

Unfortunately, many applications and protocols which work efficiently with mESC, do not work well with conventional hPSC. These include gene editing, single cell cloning and many differentiation protocols, such as to primordial germ cells. This may be due to a difference in the

developmental stage, rather than a species difference and may therefore be overcome by naïve hPSC. (Dodsworth et al., 2015, Ware, 2016, Manor et al., 2015).

1.1.2. Naïve human pluripotent stem cells

Naïve hPSC are cultured cells resembling the inner cell mass of an embryonic day (E) 5 preimplantation blastocyst (Huang et al., 2014, Boroviak et al., 2014, Marks et al., 2012, Hackett and Surani, 2014) (Figure 1.1). Conventional hPSC are not considered naïve but instead are referred to as primed, since they do not resemble the inner cell mass but rather the epiblast of an E5-7 blastocyst (De Los Angeles et al., 2012). This reported resemblance to epiblast cells consists of several strands of evidence: direct comparison to *in vivo* embryogenesis, including the observations of global hypermethylation (Smith et al., 2014, Guo et al., 2014, Okae et al., 2014), and the high expression of genes associated with the epiblast stage and lineage-associated genes such as *FGF5*, *LEFTY1*, *CER1* or *OTX2* (Buecker et al., 2014, Hackett and Surani, 2014, Joo et al., 2014). In addition, assumptions have been made from working with mouse PSC. Mouse pluripotent stem cells (mPSC) can be cultured as primed epiblast stem cells (mEpiSC) by using a medium containing basic fibroblast growth factor (bFGF)/Activin A, whereas in naïve mESC this usually causes differentiation. Naïve mESC require LIF for self-renewal, but mEpiSC do not respond to LIF (Hackett and Surani, 2014). Similarly, conventional human ES cells require bFGF signalling and do not respond to LIF (Dahéron et al., 2004) and even morphologically resemble primed murine EpiSC (Brons et al., 2007, Tesar et al., 2007). Finally, mouse ESC rely on oxidative phosphorylation, whereas both mEpiSC and hPSC are almost entirely glycolytic with low respiration capacity (Zhou et al., 2012b).

Naïve hPSC promise advantages over conventional hPSC: Differentiation protocols which function for mESC but not for conventional hPSC may now work with naïve hPSC. The most prominent example being the differentiation to primordial germ cells which has now been achieved using naïve hPSC (Irie et al., 2015). Other beneficial properties of mESC compared to hPSC include

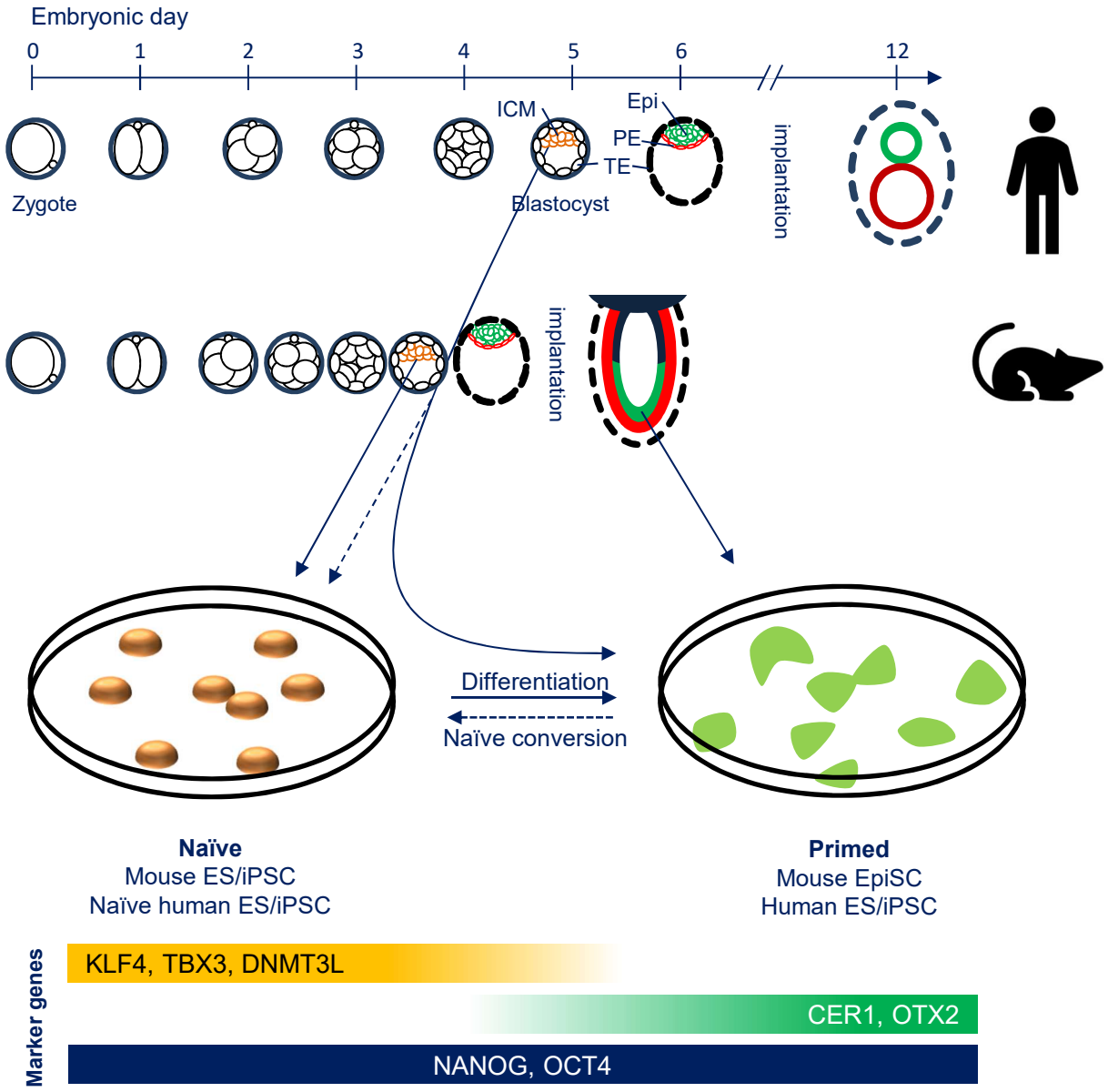


Figure 1.1: Comparison of human and mouse embryogenesis. The blastocyst forms at E3.5 in the mouse but only at E5 in humans. Both human and mouse preimplantation blastocysts consist of Trophectoderm (TE) and inner cell mass (ICM). TE cells give rise to the placenta, whereas the ICM turns into the epiblast (Epi) and primitive endoderm (PE). The epiblast will give rise to all tissues of the foetus and the PE to the extra embryonic endoderm eventually forming the yolk sac. (Niakan *et al.* 2012).
To derive new embryonic stem cell lines, cells from the inner cell mass (ICM) are dissected from the blastocyst and cultured *in vitro*. Mouse ES cells cultured in 2i+LIF conditions are maintained in the naïve state resembling cells from the ICM. However, conventional human ES cells continue developing *in vitro* and established hES lines resemble epiblast cells (Epi). To make human naïve pluripotent stem cells (hPSC), conventional (primed) hES/hiPSC can be converted to the naïve state (Theunissen *et al*. 2014) or directly derived from an E5 ICM (Guo *et al*. 2016) into naïve medium. Markers of the naïve state include KLF4, TBX3 and DNMT3L. Shared pluripotency markers are NANOG and OCT4, whereas markers of the primed state include CER1 and OTX2. Embryogenesis figure adapted from Niakan, *et al.* (2012).

shorter doubling times and easier genetic manipulation. In addition, the differentiation of mESC to mEpiSC has successfully been used as a model to allow dissection of early mouse development. These properties may now be unlocked with naïve hPSC.

Cells resembling earlier stages of pluripotency have also been discovered. In particular, cells resembling a 2-cell stage blastocyst have been revealed in mESC populations. It has been shown that mESC populations are made up of cells occupying a spectrum of pluripotency and that each individual mESC occasionally cycles between a 2C-like state back to the naïve epiblast stage in conventional conditions (Macfarlan et al., 2012, Baker and Pera, 2018, Morgani et al., 2013, Rodriguez-Terrones et al., 2018). Mouse expanded potential stem cells have been derived using media conditions designed to inhibit blastomere differentiation by targeting MAPKs, Src and Wnt/Hippo signalling. These cells were able to contribute to extra-embryonic lineages as well as just the somatic lineages and germ line (Yang et al., 2017a). Genome editing techniques have revealed differences in species. Removal of the OCT4 gene in zygotes compromised human blastocyst development, whereas mouse blastocyst development still occurred despite compromised mESC maintenance (Fogarty et al., 2017).

The naïve state was discovered in the mouse first, which means that the naïve state in humans was actively searched for. This makes it more likely that an entirely artificial state is created which only resembles *in vivo* embryogenesis in the attributes which were used to select the culture conditions (Pera, 2014, Dodsworth et al., 2015). However, any cultured pluripotent stem cells are inherently artificial, since they are maintained in a state which is transient *in vivo*. Culturing naïve cells as a means to an end (e.g. to make primordial germ cells) does not require perfect reproductions of the *in vivo* naïve state. However, if naïve hPSC are used to dissect early human embryogenesis, a close resemblance is necessary. For this reason, evidence for close resemblance should be carefully examined.

1.1.3. Evidence for naivety

The evidence brought forward for the naïve state varies in reliability and relevance. Collier *et al.* (2017) identified surface markers by screening antibodies which bound only to either naïve cells converted using the 4iLA Theunissen *et al.* protocol or to primed cells (Collier et al., 2017). However, this assumes that Theunissen *et al.* cells are indeed naïve and is therefore recursive evidence, which can be very useful for monitoring these cells but cannot be used as an argument for naivety. Other evidence brought forward for naivety compares human naivety to the phenotype of mouse naïve ESC in culture and assumes little or no difference between species. Unfortunately, many differences in embryogenesis exist between mouse and human (Theunissen and Jaenisch, 2017), which limit the relevance of these comparisons. The predictions include expected shorter doubling times, domed colony morphology and a metabolism based on oxidative phosphorylation compared to glycolytic primed cells (Zhou et al., 2012b, Takashima et al., 2014). Direct comparisons to *in vivo* data are more relevant but can vary in reliability. For example, it has been observed *in vivo*, that cells of the inner cell mass are globally hypomethylated in contrast with hypermethylated epiblast cells (Smith et al., 2014). However, it is simple to induce hypomethylation in many cell types by adding ascorbic acid (Arzenani et al., 2011, Blaschke et al., 2013). So, despite being relevant, hypomethylation on its own is not a reliable marker. The most reliable and relevant data is gene expression, since several datasets of human *in vivo* embryogenesis now exist to which naïve cells can be compared (Yan et al., 2013, Petropoulos et al., 2016). It is important to note that RNAseq datasets are large and the way the data is processed, or which genes are chosen to be displayed in a publication might influence how naïve cells look (Consortium, 2014). Notably, none of the original protocols compare their RNAseq datasets directly to *in vivo* data. This analysis will be performed in Chapter 3 of this thesis.

1.1.4. Naïve transcription factor network and gene expression

The core transcription factor network of both naïve and primed hPSC consists of OCT4, SOX2 and NANOG (OSN). They have been shown to co-occupy hundreds of sites (including their own

promoters) and roughly half of their targets are actively transcribed in hPSC (Boyer et al., 2005). OSN act mostly by repressing differentiation. OCT4 regulates the BMP4 pathway which specifies major cell fates. NANOG is a repressor of neuroectoderm and neural crest lineages, whereas SOX2 represses primitive streak differentiation (Wang et al., 2012). It has been shown that targets are poorly conserved between mouse and human ESC (Loh *et al*. 2006) but this may be due to a difference in state rather than species. OSN recruit and act synergistically with other factors which differ between naïve and primed states: Together with OSN the primed pluripotency network relies on factors OTX2 and ZIC2/3, DUSP6. Whereas in the naïve state, OSN acts together with TFCP2L1, DPPA3/5, KLF2/4/5/17 (Takashima et al., 2014, Theunissen et al., 2014, Theunissen et al., 2016, Collier and Rugg-Gunn, 2018, Chan et al., 2013, Ye et al., 2014, Dunn et al., 2014, Hackett and Surani, 2014, Niwa, 2014). TBX3 has also been shown to be required for maintaining the naïve state in mouse ESC and loss of TBX3 causes differentiation. TBX3 targets DPPA3 and is downstream of the Wnt/βCatenin pathway (Waghray et al., 2015) (Figure 1.2). The differential expression of several key members of these pluripotency transcription factor networks are used as markers for naïve and primed states in this thesis.

1.1.5. Generating naïve hPSC

The best optimised conditions to maintain naïve murine ESC is by simply culturing cells in medium containing LIF and inhibitors of GSK3-β and MEK1/2 (CHIR99021 and PD0325901, respectively; collectively referred to as 2i). However, this culture condition applied to human PSC causes neural differentiation (Hirano et al., 2012, Theunissen et al., 2014). This is likely due to the withdrawal of bFGF, which is required by human but not mouse ESC for maintenance of the pluripotent state, and the further inhibition of this pathway by PD0325901 (PD03), causing differentiation towards the ectoderm lineage (Lee et al., 2017).

An alternative approach is required. From 2013 onwards, there has been a plethora of protocols published for deriving naïve hPSC. To generate naïve human PSC, protocols developed by Chan *et*

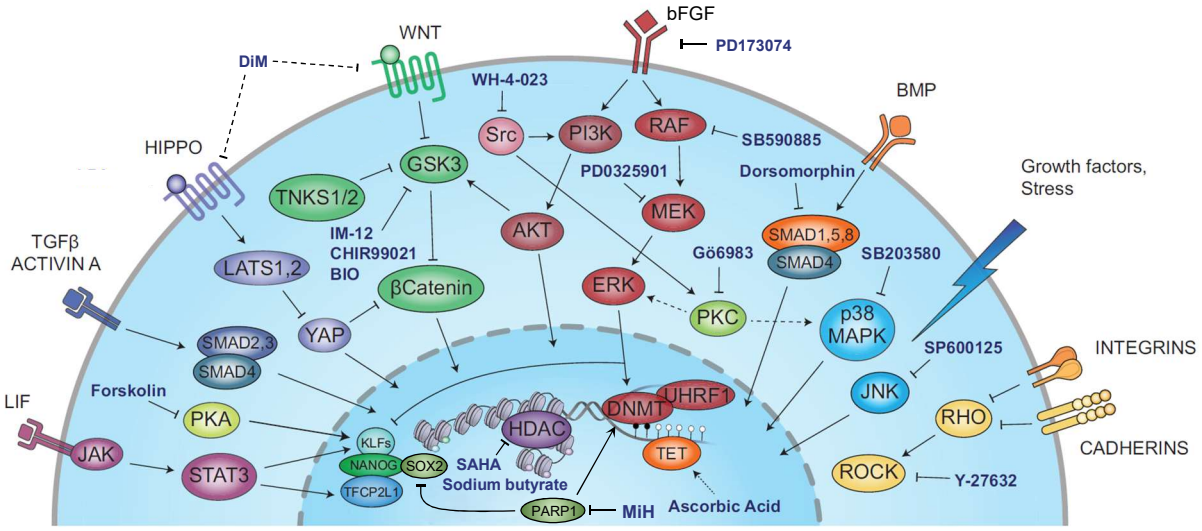


Figure 1.2: Pathways involved in naïve hPSC and inhibitors used in naïve conversion protocols. This figure shows signalling pathways involved in naïve human pluripotency. This is non-exhaustive but illustrates where inhibitors and growth factors of the different protocols act (Table 1.1). Dashed lines indicate a likely interaction. It is likely that the inhibitors may affect other components. Abbreviations: MiH = minocycline hydrochloride; DiM = (S)-(+)-Dimethindene maleate; KLFs = KLF2, KLF4, KLF5; SAHA= suberanilohydroxamic acid. Figure adapted from Collier et al. 2018

al., Duggal *et al.*, Gafni *et al.*, Theunissen *et al.* and Yang *et al.* use a media formulation to either convert conventional hPSC or derive naïve lines directly from a human blastocyst (Figure 1.3). Takashima *et al.* (2014) use the transient transfection of NANOG and KLF2 to reprogram primed cells to the naïve state. Valamehr *et al.* (2014) developed a platform for direct reprogramming from fibroblasts to naïve hiPSC using a sequence of several media formulations and single cell sorting. Ware *et al.* (2014) use a reverse toggle approach where an open chromatin state is induced by adding HDAC inhibitors which cause acetylation of histones, before changing to a naïve medium formulation. They also used this medium to derive the naïve cell line Elf1 directly from a human blastocyst.

The media formulations often contain similar components (Table 1.1) which act on several pathways to stabilise the naïve state (Figure 1.2). Importantly, all protocols require human LIF, all but the 4iLA Theunissen *et al.* condition require a GSK-3β inhibitor and all but the Yang *et al.* protocol require PD03 (MEK inhibitor). In addition, most protocols require the naïve cells to be grown on a layer of feeder cells (Table 1.1). This suggests that the optimal conditions for naïve hPSC culture have not yet been identified. Agents promoting open chromatin structure, by DNA demethylation such as ascorbic acid or histone acetylation using inhibitors against histone deacetylases (HDACi), are added in several protocols. Notably, bFGF is also added in several protocols. From the understanding of mouse naïve pluripotency, bFGF promotes differentiation to the primed state and is upstream of the MEK/ERK pathway inhibited by PD03. However, the addition of bFGF together with PD03 may promote signalling via PI3K/AKT rather than the MEK/ERK pathway (Figure 1.2) (Wigger et al., 2017, Leeb et al., 2014, Zhou et al., 2010). The protocol by Takashima *et al.* uses an FGF receptor 1 (FGFR1) inhibitor transiently after bFGF withdrawal for several days until naïve hPSC have been established. The species differences for bFGF signalling, require further research to be fully understood. In summary, media formulations (Table 1.1) contain small molecules and growth factors, which reinforce the naïve pluripotency transcription factor network (Figure 1.2).

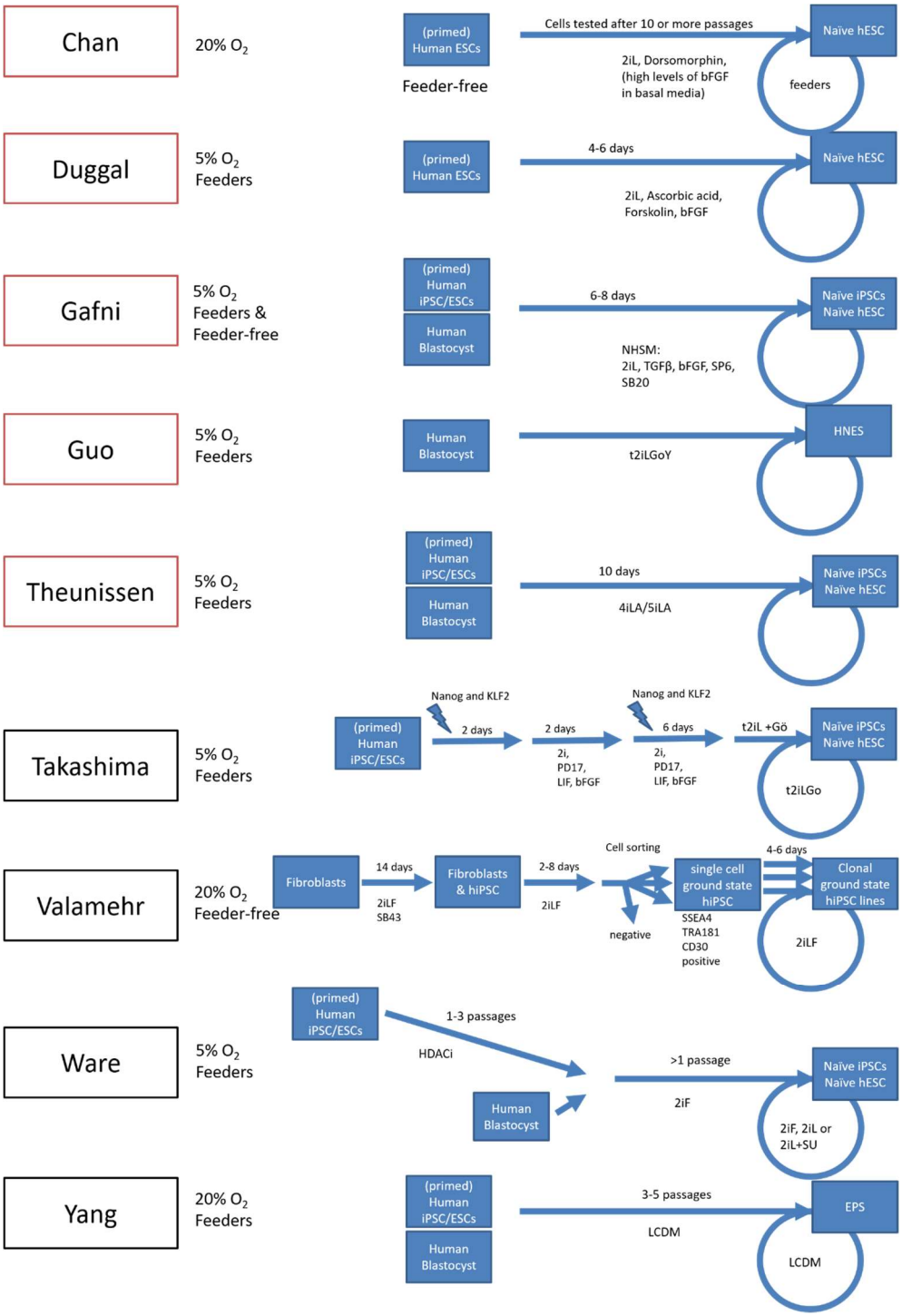


Figure 1.3: Strategies to convert to or derive naïve hPSC. Takashima *et al.* requires transient transfections, whereas other protocols are transgene-free. Guo *et al.*, Ware *et al.* and Yang *et al.* established cell lines directly from human blastocysts. The remaining protocols use hES or hiPSCs and convert using media formulations. Protocols used in this thesis are highlighted in red. Abbreviations: CHIR99021 + PD0325901 (2i), hLIF (L), bFGF (F), naïve human stem cell medium (NHSM), SP6 (SP600125), BIRB796 or SB203580 or SB202190 (p38i), titrated (t), Gö6983 (Go), Y-27632 (Y), human naïve embryonic stem (HNES) cells, PD0325901 + SB590885 + WH-4-023 + ROCKi + hLIF + Activin A (4iLA), 4iLA + IM12 (5iLA), PD173074 (PD17), SB431542 (SB43), histone deacetylase inhibitors (HDACi): sodium butyrate and suberoylanilide hydroxamic acid (SAHA), SU5402 (SU), hLIF + CHIR99021 + (S)-(+)-Dimethindene maleate + Minocycline hydrochloride (LCDM), extended pluripotent stem (EPS). Summary of media components shown in Table 1.1.

	Component	Function	Chan *et al.*	Duggal *et al.*	Gafni *et al.* NHSM1A	NHSM1B	ENHSM	Guo *et al.*	Theunissen *et al.*	Takashima *et al.*	Valamehr *et al.*	Ware *et al.*	Yang *et al.*
Small molecules	CHIR-99021 BIO IM-12	GSK-3βi	✓	✓	✓	✓ 20%	✓ 50%	✓ 33%	(✓)	✓	✓	✓	✓ (33%)
	PD0325901	MEKi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	Thiazovivin/Y-27632	ROCKi	initially		(✓)	✓	✓		✓		✓		(✓)
	Go6983	PKCi			(✓)	✓ 40%	✓ 40%	✓		initially			
	BIRB/SB203580/SB202190	p38i			✓	✓	✓						
	IWR1	WNT/β-catenin pathway inhibitor				✓	✓						(✓)
	Ascorbic acid	Tet-dependant demethylation		✓				✓					
	SP600125	JNKi, MAPKi			✓		✓ 50%						
	Dorsomorphin	BMPi	✓										
	Forskolin	stimulates cAMP production		✓									
	SB590885	BRAFi							✓				
	WH-4-023	LCKi/SRCi							✓				
	PD173074	FGFR1i								initially			
	SB431542	ALKi									initially		
	SAHA	HDACi										preculture	
	Sodium butyrate	HDACi										preculture	
	SU5402	FGFRi/VEGFRi										(✓)	
	DiM	inhibits G protein coupled receptors											✓
	MiH	PARP1i											✓
Signaling Proteins	hLIF	Activates JAK/STAT and MAPK cascades	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	bFGF/FGF2	FGF signalling	*	✓	✓				(✓)	initially	✓	✓	
	Activin A	Activation of SMAD2/3			✓	✓	✓		✓				
	Insulin	Metabolic and catabolic regulator			✓	✓	✓						
	TGFb	Activation of SMAD2/3	*		✓								
Growth conditions	Feeders		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
	Feeder-free			(✓)	✓	✓	✓				✓		
	Oxygen levels		20%	5%	20%	20%	20%	5%	5%	5%	20%	5%	20%
	Basal media		TeSR1	KO-DMEM, KSR, NEAA, Glutamine	KO DMEM, AlbuMAX1, NEAA, N2	DMEM/F12, Neurobasal, BSA, KSR, NEAA, Glutamine, N2, B27	DMEM/F12, Neurobasal, BSA, KSR, NEAA, Glutamine, N2, B27	DMEM/F-12, Neurobasal, B27, L-Glutamine, N2	DMEM/F12, Neurobasal, BSA, KSR, NEAA, Glutamine, N2, B27	DMEM/F-12, KSR, NEAA, Glutamine. Maintenance in N2B27 medium	DMEM/F12, KSR, NEAA, Glutamine, N2, B27. (No N2 or B27 for maintenance)	DMEM/F12, KSR, NEAA, Glutamax, sodium pyruvate	DMEM/F12, Neurobasal, BSA or KSR, NEAA, Glutamax, N2, B27

Table 1.1: Comparison of naïve media. Different components required for derivation and/or maintenance of naïve hPSCs are shown here. An asterisk (*) indicates high levels of this component in the basal media. Components in brackets are optional. Percentages next to ticks specify the component has been titrated to a lower concentration than the other media formulations. Components are ordered according type and frequency of usage across protocols.

To arrive at each media formulation, different groups used different strategies. Takashima *et al.* used *KLF2* and *NANOG* transfection to reset primed hPSC to the naïve state. They found that maintenance without these transgenes was not possible in a simple 2iL medium. The authors then tested candidate pathway inhibitors and found that the protein kinase C (PKC) inhibitor Gö6983 (Go), together with titrated 2iL, was able to maintain naivety without transgene expression. However, Takashima and colleagues did not reveal how many or which other candidate pathway inhibitors were tested (Takashima et al., 2014). Ware *et al.* reasoned that the transgene approach could be replaced by epigenetically resetting hPSC first and then keeping the cells in 2i+LIF+bFGF (2iLF). Although they demonstrated this in their publication, they did not disclose how they arrived at their conclusion (Ware et al., 2014). Several groups performed screens to find compounds to increase naïve features: Chen *et al*. screened 20 compound combinations for higher NANOG expression (Chan et al., 2013). Gafni *et al*. and Duggal *et al*. screened 16 and 5 compound/growth factor combinations, respectively, for increased OCT4 expression (Gafni et al., 2013, Duggal et al., 2015). Both NANOG and OCT4 are essential for pluripotency but are not significantly differentially expressed between states in mouse (Brons et al., 2007, Tesar et al., 2007) and human (Theunissen et al., 2014). Nevertheless, conditions were found which enhanced other naïve features (Table 1.2). A more relevant and comprehensive approach was performed by Theunissen *et al*., who designed a reporter system based on the differential enhancer usage of *OCT4*. It has been conclusively shown that *OCT4* expression is driven from a distal enhancer in mESC and a proximal enhancer in mEpiSC (Tesar et al., 2007). Theunissen *et al*. validated this with a doxycycline (dox) inducible *KLF2* and *NANOG* naïve cell line. They found, as previously revealed by Takashima *et al.*, maintenance without transgene expression in just 2i+LIF (2iL) was not possible. Theunissen and colleagues performed a large screen of several hundred conditions containing 2iL and individual kinase inhibitors. Doxycycline was used as a positive control and a highly active distal enhancer was found for several compounds. They tested these compounds for viability and found SB590885 (SB59) and WH-4-023 very effective in terms of maintaining naivety and viability. This approach

was repeated by Yang *et al.*, this time using primed cells as a starting point and on a larger scale. They found DiM ((S)-(+)-Dimethindene maleate) and MiH (Minocycline hydrochloride) effective at generating naïve hPSC and reported that these compounds together with LIF and CHIR allowed culture of naïve cells without PD03 mediated MEK inhibition and functioned even with TGFβ signalling inhibited. They were able to demonstrate that MiH acts via poly ADP ribose polymerase (PARP) inhibition but how DiM, an inhibitor of G protein–coupled receptors (GPCRs), enhances naivety is yet to be shown (Yang et al., 2017b).

1.1.6. Comparing naïve hPSC generation protocols

Many different strategies and protocols to generate naïve hPSC have now been published (Chan et al., 2013, Duggal et al., 2015, Gafni et al., 2013, Guo et al., 2016, Theunissen et al., 2014, Takashima et al., 2014, Valamehr et al., 2014, Ware et al., 2014, Yang et al., 2017b). Such a plethora of methods to generate naïve hPSC demands that the reported evidence for naivety is reviewed and is subsequently listed in Table 1.2. Assessed are whether the cells are indeed pluripotent by their ability to differentiate. This has been measured in different ways: *in vitro* spontaneous differentiation can be induced by making embryoid bodies (EB) and measuring markers of each germ line, for example by quantitative reverse transcription polymerase chain reaction (RT-qPCR). Alternatively, each germ line can be accessed using a directed differentiation protocol. Finally, the differentiation ability *in vivo* can be assessed by teratoma formation. This method involves subcutaneous injection of cells into non-obese diabetic/severe combined immunodeficiency (NOD/SCID) mice, allowing teratomas to form and confirming the presence of all 3 germ layers. A more stringent method was put forward by Hanna and colleagues, who suggest injecting naïve cells into the developing mouse embryo and assessing if human cells are present in each tissue type (Gafni et al., 2013). Due to the difficulty of the technique, only one group has used it since (Yang et al., 2017b) but concerns about lack of reproducibility and false positives have been raised (Theunissen et al., 2016). An additional report showed success by using pig

		in vitro differentiation		teratoma formation into all 3 germ layers	XaXa	Hypo-methylated	oxidative phosphorlyation	media used to isolate mESC
		spontaneous (EB)	directed					
Chan *et al.*	2013	✓	N.D.	✓	N.D.	N.D.	N.D.	N.D.
Duggal *et al.*	2015	✓	✓	N.D.	✓	✓	N.D.	✓
Gafni *et al.*	2013	✓	N.D.	✓	✓	✓	N.D.	N.D.
Guo *et al.*	2016	✓	✓	N.D.	N.D.	✓	✓	N.D.
Theunissen *et al.*	2014	N.D.	✓	✓	b	b	N.D.	N.D.
Takashima *et al.*	2014	✓	✓	✓	✓	✓	✓	N.D.
Valamehr *et al.*	2014	✓	✓	✓	✓	✓	N.D.	N.D.
Ware *et al.*	2014	N.D.	N.D.	✓	✓	✓	a	N.D.
Yang *et al.*	2017	✓	N.D.	✓	✓	N.D.	N.D.	✓

		transgene-free	normal karyotype	mES morphology	no cell sorting required	naïve specific RNA profile		
						RNA-seq	array	qPCR
Chan *et al.*	2013	✓	✓	✓	✓	✓	N.D.	✓
Duggal *et al.*	2015	✓	✓	✓	✓	b	✓	✓
Gafni *et al.*	2013	✓	✓	✓	✓	b/c	✓	✓
Guo *et al.*	2016	✓	✓/X	✓	✓	✓	N.D.	✓
Theunissen *et al.*	2014	✓	X	✓	✓	b	✓	✓
Takashima *et al.*	2014	X	✓	✓	✓	✓*	✓	N.D.
Valamehr *et al.*	2014	✓	✓	X	X	N.D.	✓	N.D.
Ware *et al.*	2014	✓	✓	✓	✓	b	N.D.	N.D.
Yang *et al.*	2017	✓	✓	✓	✓	✓	N.D.	N.D.

Table 1.2: Evidence published for available protocols. Ticks (✓) indicate that positive evidence has been provided, crosses (X) specify evidence against that category and N.D. that no evidence was brought forward. An asterisk (*) indicates that this evidence has been reproduced by an independent group. “a” means (less stringent) mitochondrial morphological assessment only. “b” indicates that no evidence was brought forward in the initial publication but has subsequently been published. “c” means that no evidence was included in the initial publication but evidence against naivety of these cells has now been brought forward by a different group. An objective comparison of all published RNAseq datasets is shown in Figure 3.1 and S2. Normal karyotype is determined by G-banding or more stringent methods. Protocols above the dotted line have been tested in this thesis.

instead of mouse blastocysts (Wu et al., 2017a), but the technique has not been adopted more widely, at least in part due to the ongoing ethical debate around blastocyst chimera experiments. In female cells, the inactivation of the second X chromosome occurs after exit from the naïve state (Minkovsky et al., 2012). Therefore, both X chromosomes being active is a marker of naivety and Table 1.2 shows protocols which have brought forward the appropriate evidence. Of note, Theunissen *et al.* (2014) originally reported that their naïve cells had one inactivated X chromosome (XaXi) when first publishing their work. They concluded this because they observed high expression of X-inactive specific transcript (Xist) and measured apparently dosage-compensated expression of X-linked genes. However, it has since been shown that Xist expression is also high in early *in vivo* embryogenesis and expression from the X chromosome in female cells is dosage compensated but biallelic (Petropoulos et al., 2016, Okamoto et al., 2011). This led the authors to re-examine their naïve cells, using a different fluorescent reporter on each X chromosome. This more stringent assay found both alleles to be expressed in naïve cells and only one in primed or differentiated cells (Theunissen et al., 2016). This was corroborated by analysing SNPs in RNAseq data (Sahakyan et al., 2017).

Unlike primed cells, naïve cells of the inner cell mass have been shown to be hypomethylated (Smith et al., 2014). Most published protocols show evidence for a hypomethylated genome (Table 1.2). In addition, oxidative phosphorylation has been shown to be the primary metabolic pathway used in (naïve) mESC (Zhou et al., 2012b). Takashima *et al.*, Ware *et al.* and Guo *et al.* observed this in their naïve human PSC and use it as further evidence for naivety, since human primed cells are almost entirely glycolytic (Zhou et al., 2012b). However, none of the other groups measured their cells metabolism.

Finally, two groups have used their naïve human PSC medium to derive naïve mouse ES cells equivalent to standard 2iL conditions. This does not add to the evidence that the human cells are

indeed naïve, however it is important in at least one aspect: Despite the presence of bFGF in the Duggal *et al.* medium, mouse ES cells can be derived in the naïve state (Duggal et al., 2015).

Table 1.2 also shows that protocols have different disadvantages. Takashima *et al.* requires transgenes, Theunissen *et al.* and Guo *et al.* have issues with genome integrity and Valamehr *et al.* requires an extensive platform including cell sorting for their protocol. Finally, mouse ESC morphology has been reported for all but the Valamehr protocol.

Gene expression has been assessed for cells from all protocols. In several initial publications RNAseq data was not included but added retrospectively either by other groups or in follow up publications. These protocols include Duggal, Ware and Gafni (Warrier et al., 2017) and cells generated using Takashima, Gafni and Theunissen protocols (Theunissen et al., 2016). Each group shows successful naïve gene expression but often by their own metrics. Despite being the most important evidence, it seems the most difficult to interpret and the easiest to misinterpret. Table 1.2 shows which protocols have evidence for or against naïve gene expression but does not comment on the quality or credibility of the result. Instead, the available RNA sequencing (RNAseq) datasets will be analysed in parallel and compared to *in vivo* human embryogenesis in chapter 3 of this thesis.

1.1.7. Naïve to primed differentiation

Naïve to primed differentiation in mESC has successfully been used as a model for preimplantation development (Pandolfini et al., 2016, Wu and Izpisua Belmonte, 2015, Buecker et al., 2014, Factor et al., 2014, Zhou et al., 2012a). Both *in vivo* and *in vitro* there is a dismantling of the naïve pluripotency network before an upregulation of lineage specification genes (Van Bortle and Corces, 2014). Large screens were performed to identify key players and pathways required for the exit from mouse naïve pluripotency. These studies identified *Dicer* and *Dgcr8* (key components of the miRNA system), in addition to many genes associated with the ERK and/or GSK3-β pathway as well as epigenetic modifiers, such as genes for histone demethylases *Kdm6a*

and *Kdm5b* and Polycomb group proteins *Suz12* and *Jarid2* (Yang et al., 2012, Leeb et al., 2014, Betschinger et al., 2013, Pandolfini et al., 2016). The polycomb group of proteins act in two Polycomb repressive complexes (PRC1 & PRC2) to mediate heritable silencing of developmental regulators. *Jarid2* and *Suz12* act together with *EeD* and *Ezh2* in PRC2 and catalyses the trimethylation of histone H3 at lysine 27 (H3K27). PRC1 contains *Cbx2*, *Cbx4* and *Cbx8*, *Mel18* and *Bmi1*, *Phc1/2/3*, as well as *Ring1A/B*. *Ring1A/B* represses developmental regulators in mESC enabling maintenance of ES cell identity. Many PRC1 target genes are also repressed by OCT3/4 and polycomb silencing has been thought to act downstream of the core transcriptional circuit (Endoh et al., 2008). Finally, forced expression of *OTX2* has been shown to induce exit from the human naïve state (Buecker et al., 2014), but no further studies on human exit of naïve pluripotency have been performed so far.

1.1.8. Summary

Naïve hPSC are an exciting new cell type which promise to overcome major bottlenecks in hPSC culture and applications. However, it is unclear which protocol enables derivation of the most physiological naïve hPSC. In addition, the naïve to primed differentiation has not been established in hPSC. This will be performed and further explored in Chapter 4.

1.2. miRNAs

1.2.1. Introduction

Due to the requirement of DICER1 and DGCR8 for the naïve to primed differentiation, miRNAs have been identified as being of potential interest in pluripotency. Micro RNAs (miRNAs) are a class of small non-coding RNAs (~22 nucleotides) which inhibit complementary mRNAs by binding to the 3' untranslated region (UTR) and either flagging it for degradation or blocking translation. To produce miRNAs, regions of the genome are transcribed by RNA polymerase II (Figure 1.4). Alternatively, they reside within an intron and are processed following splicing of the intron. Both can produce a long primary transcript (pri-miRNAs) containing at least one hairpin. Still in the

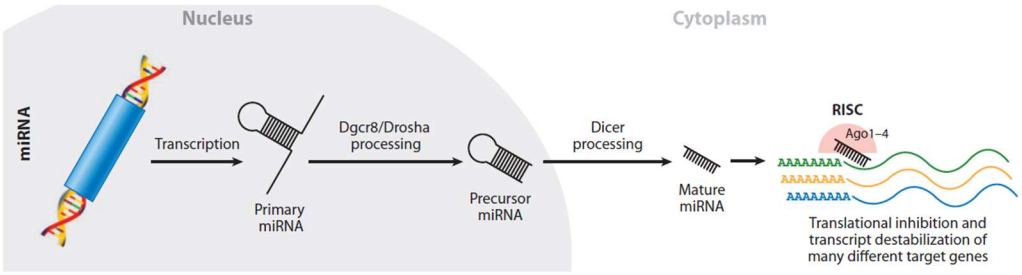


Figure 1.4: Schematic of the miRNA biogenesis pathway showing DGCR8 and DICER1. To produce miRNAs, regions of the genome are transcribed by RNA polymerase II. Alternatively, they reside within an intron and are processed following splicing of the intron. Both can produce a long primary transcript (pri-miRNAs) containing at least one hairpin. Still in the nucleus, DGCR8 processes it into precursor miRNA (pre-miRNA) by cleaving the hairpin structure off the rest of the transcript. These pre-miRNAs are transported to the cytoplasm, processed further by DICER1 which removes the loop structure at the end of the hairpin, leaving double stranded mature miRNA. The mature miRNA is then loaded onto AGO, one strand is removed, and the miRNA can perform its functions in gene regulation (Bartel, 2009, Fabian and Sonenberg, 2012). Figure adapted from Greve *et al.* (2013).

nucleus, DGCR8 processes it into precursor miRNA (pre-miRNA) by cleaving the hairpin structure off the rest of the transcript. These pre-miRNAs are transported to the cytoplasm, processed further by DICER1 which removes the loop structure at the end of the hairpin, leaving double stranded mature miRNA. The mature miRNA is then loaded onto AGO, one strand is removed, and the miRNA can perform its functions in gene regulation (Bartel, 2009, Fabian and Sonenberg, 2012).

The 6-8 bases near the 5' end of miRNA are largely responsible for its target specificity. This region is referred to as the seed region. The miRNA mediated inhibition tolerates imperfect pairing for the rest of the miRNA, which makes *in silico* target prediction both challenging and inaccurate (Bartel, 2009). Families of miRNAs are miRNAs which share the same seed sequence and will often bind the same mRNA targets. Clusters are miRNAs from the same locus, which can be from different families and are transcriptionally co-regulated (Friedman et al., 2009). miRNAs are commonly viewed as fine tuners of gene expression which cause only small changes but over a wide range of targets. A single miRNA can effect hundreds of targets (Farh et al., 2005). In addition, it has been observed that miRNAs often act synergistically (Xu et al., 2011), although master regulator miRNAs have also been discovered (Rajman and Schratt, 2017).

miRNAs represent an additional layer of post-transcriptional regulation. They regulate more than a third of all protein coding genes (Bentwich et al., 2005) and their importance as potential cell fate regulators is emphasised by the observation that transcription factors are over represented as targets of miRNAs (Cui et al., 2006).

1.2.2. miRNAs in development

Early miRNAseq experiments in mouse ES cells led to the definition of so called ESC-specific cell cycle-regulating (ESCC) miRNAs which have the seed sequence AAGUGC and dominate miRNA expression with 20-50% of all miRNAs in mESC (Jouneau et al., 2012, Marson et al., 2008, Bar et al., 2008, Morin et al., 2008, Calabrese et al., 2007). ESCC miRNAs are downregulated during

differentiation and upregulation has been associated with cancer (Wang et al., 2008). These miRNAs generally are not expressed in somatic cells.

ESCC clusters in both mouse and human are miR-302-367, miR-17-92b, miR-106a-363, miR-106b-25 and either miR-290-295 in the mouse or the human homolog miR-371-373. In *in vivo* mouse early embryogenesis, the miR-290 family is heavily expressed until the emergence of extraembryonic structures at E7.5. Then it is exclusively expressed in these tissues such as the placenta and yolk sac. miR-290 expression is therefore a marker of early pre-implantation embryogenesis. In contrast, the miR-302 family is only expressed from E5.5 in the epiblast, where it is co-expressed with miR-290 until E7.5. Here, expression of miR-302 drops until E10.5, where it is last expressed localised to anterior structures. So miR-302 expression is a marker of the postimplantation epiblast. (Houbaviy et al., 2003, Greer Card et al., 2008, Parchem et al., 2015, Parchem et al., 2014, Tang et al., 2007)

ESCC miRNAs are important but act redundantly in embryogenesis. Knockout of miR-290 leads to partially penetrant embryonic lethality and germ cell defects but an adult mouse can still be formed (Medeiros et al., 2011), whereas deletion of mir-302a-d leads to increased and early neuronal progenitor proliferation and failure in neural tube closure. Deletion of both miR-290 and miR-302a-d leads to early embryonic lethality (Parchem et al., 2015).

In vitro the miR-290 family are the most highly expressed miRNAs in naïve mouse ESC (Greve et al., 2013) and do not exist in the human. The human homolog, the miR-371-373 cluster, is reported to be variably, if at all, expressed in conventional human PSC lines (Wu et al., 2014a, Morin et al., 2008, Bar et al., 2008). However, it was shown to be upregulated in human naïve 5iLAF (Theunissen *et al.*) cells but only in one cell line that was used for the analysis (Faridani et al., 2016).

The miR-302/367 family is the most abundant miRNA family in human primed ES cells and mouse EpiSC (Suh et al., 2004, Morin et al., 2008, Bar et al., 2008). Despite the miR-302 cluster being in

an intron (of the gene LARP7), it was discovered to have its own transcriptional unit. It is transcribed not only as an intron but also independently from its host gene as a separate polycistronic transcript containing five mature miRNAs: miR-302a/b/c/d and miR-367 (Barroso-delJesus et al., 2008). In addition to the evidence from miR-302 knockouts, the miR-302/367 cluster has been shown to repress neural differentiation also *in vitro* (Rosa et al., 2009, Lipchina et al., 2011) and its expression is driven at least in part by OCT4 and SOX2 (Marson et al., 2008).

Without a functional model of human pre-implantation, most experiments on miRNAs involved in pluripotency have been performed in the mouse. The exceptions include the discovery that miR-145 downregulates *OCT4, SOX2, KLF4* in the human (Xu et al., 2009) and Let-7, also antagonises pluripotency and is blocked by LIN28 in both the mouse and human (Viswanathan et al., 2008, Newman et al., 2008, Rybak et al., 2008). A single-cell small-RNA sequencing experiment of naïve and primed hPSC showed no variability between cells but a large difference between naïve and primed states. However, this analysis was limited, since only one cell line was used and the single-cell small-RNA sequencing only included 23 and 11 cells for naïve and primed states, respectively (Faridani et al., 2016).

1.2.3. *Dgcr8* and *Dicer1* knockout murine studies

DICER1 and DGCR8 proteins are required for miRNA biogenesis. Knockout of the *Dicer1* gene has been achieved *in vivo* by crossing two heterozygous knockout mice. *Dicer1* null mutants develop normally until E5.5, where gastrulation is delayed by a day due to increased apoptosis. At E7.5 the null embryos are morphologically abnormal but both embryonic and extraembryonic structures are visible, whereas from E8.5, embryos have an empty or necrotic epithelium (Spruce et al., 2010, Bernstein et al., 2003). These studies suggest that miRNAs are required for postimplantation and not for preimplantation development. In these experiments, an important confounding factor is potentially left over maternally-contributed DICER1 protein which might mask miRNA requirements up to E5.5. However, mouse *Dgcr8* knockout oocytes have identical gene expression

to wild type cells. They also develop entirely normally (generating healthy offspring) when fertilised with wild type sperm, indicating that miRNAs are indeed not important in very early stages of development (Suh et al., 2010, Ma et al., 2010). This is in line with the miRNA biogenesis pathway being less active during pre-implantation (Garcia-Lopez and del Mazo, 2012).

The naïve state of mESC is equivalent to E4.5 and the primed mEpiSC to E6.5, spanning the time points at which *Dicer1* null mutants show a phenotype. This has allowed investigation of miRNA requirements *in vitro*. Knocking out *Dicer1* or *Dgcr8* has been achieved in mESC but not mEpiSC. It causes the mESC to be blocked in the naïve state. When differentiation cues are presented, the naïve cells do not fully downregulate the naïve pluripotency programme and do not acquire lineage specific genes (Wang et al., 2007, Kanellopoulou et al., 2005). This suggests that miRNAs are not required for maintenance of the naïve state but are essential for the exit from naivety. This phenotype can be partially rescued by the addition of either miRNA-290-3p or miR-302-3p mimics. In addition, knockout of both miR-290 and miR-302 clusters, but not either alone, blocks the silencing of the naive pluripotency program (Gu et al., 2016). This is in line with an observation that unlike wild type cells, *Dicer1* null mutant mESC are unable to methylate the *Oct4* locus when differentiating conditions are applied. An effect which can be rescued again by the addition of miR-290 cluster miRNAs (Sinkkonen et al., 2008). Finally, during conventional 2i culture these knockout cells have a slight growth disadvantage with more cells in G_1 phase. This can be overcome by removing the GSK3-β inhibitor (Yan et al., 2017) or again by transient transfection of members of the miR-290 family (Wang et al., 2008). Intriguingly, the miR-290 family does not exist in the human (Greve et al., 2013) and it will remain to be seen to what extent the homologous human miR-371-373 cluster assumes the same roles.

So far no *DICER/DGCR8* knockout in hPSC cells exist and it has been shown to be an essential gene in primed hPSC (Yilmaz et al., 2018). However, a knockdown of *DROSHA* or *DICER1* using shRNA in hESC (2.3-3.6 fold and 5.3-6.3 fold mRNA downregulation, respectively) results in similar cell cycle

defects as seen in the mouse which can also be restored by miR-372 (a homolog of the mouse miR-290 family discussed above), as well as members of the miR-302-367 and the miR-106b-25 cluster (Qi et al., 2009).

Overall expression levels of miRNAs increase from week 6 of human development with many of the key developmental miRNAs also involved in tumorigenesis (Lin et al., 2013). Cumulatively, these experiments indicate that miRNAs are not required in mouse preimplantation development but are required for postimplantation development. The naïve to primed differentiation has been used skilfully and appropriately in this context. Until the naïve state was described in the human system, this has not been possible and to date, the miRNA requirements of human development have not been fully explored. It remains unclear if human preimplantation development also does not require miRNAs.

1.2.4. Summary

In summary, in murine cells both *in vivo* and *in vitro*, there is a switch in dominant miRNA expression in pluripotent cells from the preimplantation miR-290 family to the postimplantation miR-302 family. miRNAs in general seem dispensable in the preimplantation blastocyst but required for postimplantation development. This has yet to be studied in the human system. In this thesis, the miRNAome of early human development will be examined in chapter 4.

1.3. Gene editing and DNA repair in PSC

1.3.1. Introduction

Mouse naïve ES cells are easier to genetically manipulate through homologous recombination (HR) than either EpiSC or human PSC (Zwaka and Thomson, 2003). It has been hypothesised and early published results suggest that naïve human PSC may offer significant improvements in gene editing efficiency versus their primed counterparts, because of their less condensed, hypomethylated state which allows for greater chromatin accessibility (Wu et al., 2014b, Dodsworth et al., 2015). Alternatively, they may possess a more active HR pathway.

1.3.2. Homologous recombination and non-homologous end joining

When a DNA double strand break (DSB) occurs, it is crucial for the cell to repair it as lesions can lead to chromosomal rearrangements, genomic instability, cell death or tumorigenesis. Several repair mechanisms exist in mammalian cells, with the most prominent pathways being canonical non-homologous end joining (C-NHEJ) and homologous recombination (HR) (Jackson and Bartek, 2009). Both HR and C-NHEJ are high fidelity repair pathways which contrast with two lesser used alternative and highly mutagenic pathways: microhomology mediated end joining (MMEJ) and single strand annealing (SSA; Figure 1.5A)(Her and Bunting, 2018, Ceccaldi et al., 2016a).

HR is an error-free DNA repair pathway which uses a sister chromatid as a template. At the outset of homology directed repair (HDR), end clipping generates small 20 bp 3' ssDNA overhangs. This is achieved by the CtIP/MRN complex (containing MRE11, Rad50 and NBS1): It cuts one strand close to the break site using its 5'-3' endonuclease activity and then degrades the same strand by 3'-5' exonuclease activity. This MRE11-mediated end processing prevents NHEJ by displacing the Ku70/80 heterodimer required for NHEJ (Shibata et al., 2014). The resulting short single stranded DNA (ssDNA) overhangs can either be substrate for the mutagenic MMEJ or can be further resected by helicases and exonucleases (such as EXO1, BML, DNA2 and CtIP) (Kowalczykowski, 2015, Sturzenegger et al., 2014, Symington and Gautier, 2011) generating 3' ssDNA overhangs up to 3500 nucleotides in length (Zhou et al., 2014). The ssDNA resulting from end resection is protected from degradation by the binding of the Replication Protein A (RPA) complex, which also prevents secondary structure formation. RPA is removed by BRCA2 and RAD51 polymerises on the exposed ssDNA overhangs, forming the presynaptic filament (Zhao et al., 2015). Strand invasion occurs when the presynaptic filament invades the unbroken sister chromatid and finds the region of homology where it provides a template for repair. The invading 3' strand anneals and is extended primarily by DNA polymerase δ (Li et al., 2009), creating a large region of homology to the resected strand. The break is resolved by the re-annealing of the invading strand to its original complementary sequence, gaps are filled by DNA synthesis and the repair is

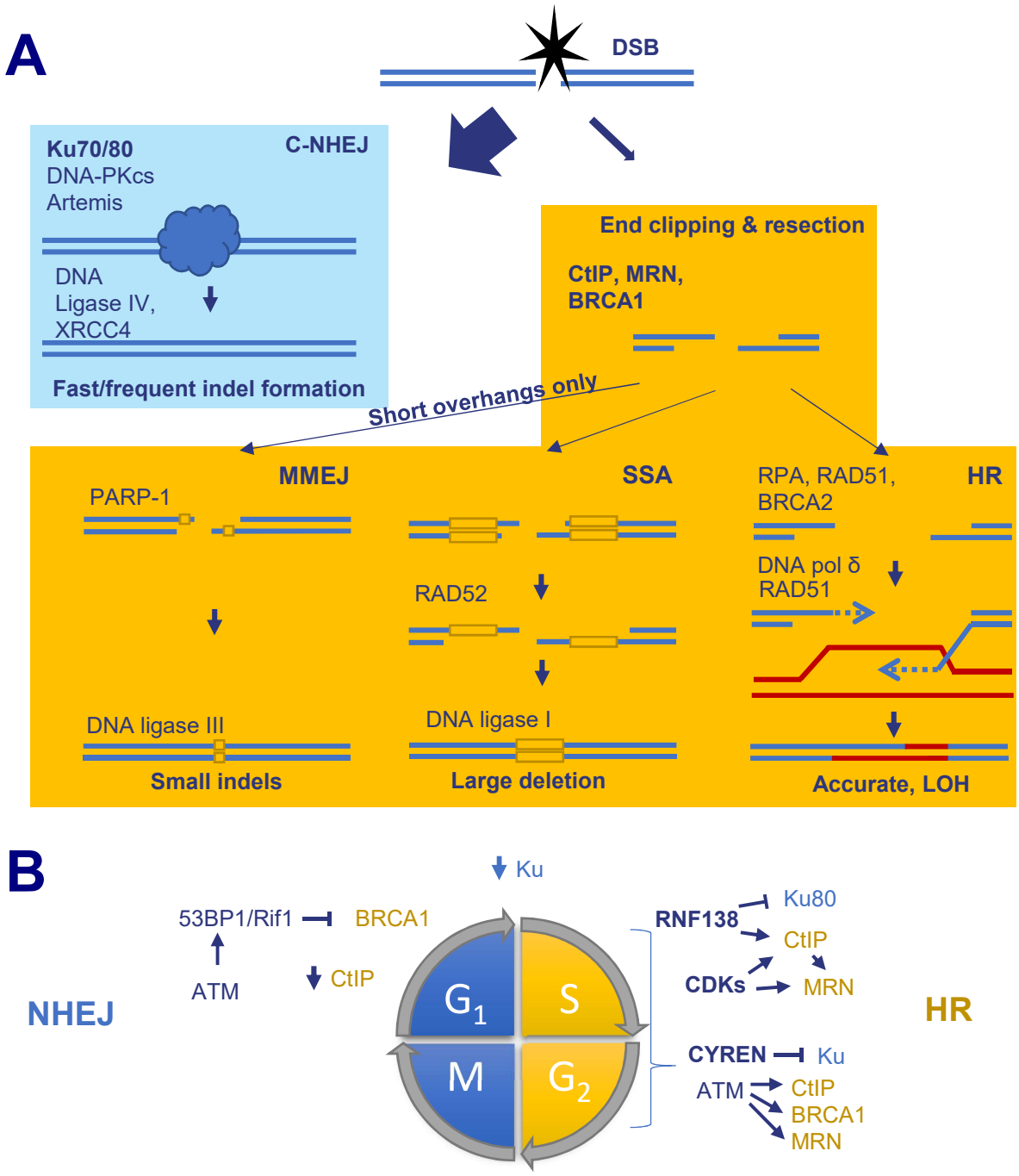


Figure 1.5: DNA repair pathways and their regulation by the cell cycle. (A) Most double strand breaks are swiftly repaired by C-NHEJ which requires Ku70/80 heterodimer binding to the break site. Alternatively, end resection is mediated by CtIP, BRCA1 and the MRN complex. End resection leads to: homologous recombination (HR) mediated by RPA, RAD51 and BRCA2; single strand annealing (SSA) mediated by RAD52; or microhomology mediated end joining (MMEJ) mediated by PARP-1. In physiological conditions, HR uses a sister chromatid as a template and is therefore accurate but leads to loss of heterozygosity (LOH). SSA connects two homologous regions which causes deletion of one of the regions and any sequence in between. Finally, MMEJ connects microhomologies and leads to small deletions of the sequence in between. (B) NHEJ and HR are regulated by the cell cycle, since HR is the most accurate pathway but requires access to a sister chromatid only present in S and G_2. The control is exerted on members of the first step of the pathway, namely end resection or binding of Ku70/80 to the break site. Figure adapted from Bunting *et al.* (2018) and Ceccaldi *et al.* (2016)

completed by ligation. This dominant form of resolving HR is the synthesis-dependent strand annealing (SDSA) pathway (Her and Bunting, 2018, Ceccaldi et al., 2016a). Other mechanisms to resolve the HR pathway are used especially during meiosis which is not in the scope of this thesis and are reviewed elsewhere (Bizard and Hickson, 2014).

The C-NHEJ pathway repairs breaks with little or no homology. Several publications suggest that C-NHEJ is attempted first, before end resection and alternative pathways are utilised (Beucher et al., 2009, Karanam et al., 2012, Mao et al., 2008, Shahar et al., 2012). This is based on several strands of evidence. Studies using fluorescent reporters showed that while HR requires several hours to repair, C-NHEJ is much faster and takes approximately 30 minutes (Mao et al., 2008). In addition, C-NHEJ is the most commonly used repair pathway with 80% of DSB induced by radiation in G_2 repaired by NHEJ in human cell lines. It has therefore been proposed that, due to the fast kinetics of C-NHEJ, it protects from catastrophic chromosomal translocations (Beucher et al., 2009, Her and Bunting, 2018). Complexity of the break also determines pathway choice. Blunt end breaks or nuclease induced cuts are simple to repair compared to infrared radiation which requires removal of affected nucleotides (Reynolds et al., 2012, Shibata et al., 2011).

When microhomologies are present, i.e. when overhangs are perfectly compatible, NHEJ can repair accurately. However, loss of nucleotides occurs frequently, especially with blunt ends or when overhangs are not compatible, which leads to insertions and deletions (indels). C-NHEJ is initiated by Ku70/80 heterodimer binding to the DSB. Ku70/80 form a scaffold for further co-factors, such as DNA-PKcs, to facilitate repair. End processing occurs if ends are incompatible and final ligation is performed by DNA ligase IV and XRCC4 (Her and Bunting, 2018, Ceccaldi et al., 2016a).

Microhomology-mediated end joining (MMEJ), also referred to as alternative end joining (alt-EJ), is thought to function as a backup to C-NHEJ (Guirouilh-Barbat et al., 2007, Chiruvella et al., 2013, Bentley et al., 2004). However, it has been shown to compete with HR (Mateos-Gomez et al.,

2015, Ceccaldi et al., 2015). MMEJ acts after initial end clipping but before long resection. It works in a Ku independent manner and PARP-1 has been proposed to recognise short regions of homology and allow co-factor binding. Co-factors include DNA translesion synthesis (TLS) polymerases which facilitate the repair (Audebert et al., 2008, Chan et al., 2010).

The SSA pathway competes with HR and causes end joining between nucleotide repeats. After nucleotide resection and RPA binding, SSA requires reannealing of RAD52. Nucleotide repeats are annealed and one repeat as well as the region between repeats is lost during the repair (Heyer et al., 2010).

HR requires a sister chromatid in close proximity, which only occurs during S and G_2 phase of the cell cycle. Predictably, the cell cycle plays a central role in regulating repair pathway activity (Figure 1.5B). The 5'-3' end resection is a commitment to repair by HR and is inhibited during G_1 phase by the Ku70/80 heterodimer, thereby favouring NHEJ (Shao et al., 2012, Sun et al., 2012). Conversely, increased end resection is observed during S and G_2 phases. This occurs through the action of cyclin-dependent kinases (CDKs). During S phase, CDKs phosphorylate and activate CtIP and members of the MRN complex. In addition, RNF138 increases CtIP activity and inhibits Ku by ubiquitinylation (Huertas and Jackson, 2009, Falck et al., 2012). During both S and G_2 phase CYREN is expressed, which inhibits Ku binding (Arnoult et al., 2017). Without stabilisation by CDKs, CtIP degrades and is only present in low levels. 53BP1 is present at higher levels during G_1 and sequesters BRCA1, further reducing end resection and increasing activity of C-NHEJ. During S phase 53BP1 is diluted, increasing the available BRCA1 (Bunting et al., 2010). The cell cycle, therefore, is a key regulator of DNA repair pathway activity and will be assessed in chapter 5.

1.3.3. Short single stranded DNA as repair templates

HDR occurs after a double strand break (DSB) and in physiological contexts (such as meiotic homologous recombination) using a double stranded DNA template as a donor. HDR with the use of single stranded oligos as a repair template utilises the single strand template repair (SSTR)

pathway. Here, instead of RAD51 mediated strand invasion, the oligo acts as a template for DNA synthesis by annealing to the resected ends. SSTR is RAD51 independent and requires the Fanconi anemia pathway. (Richardson et al., 2018, Ceccaldi et al., 2016b, Davis and Maizels, 2016, Kan et al., 2017)

1.3.4. Summary

There is reason to believe that gene editing efficiencies may be different between naïve and primed cells. This may be due to improved survival in naïve hPSC or a more active HR pathway. In addition, HR and NHEJ pathways are under the control of the cell cycle. Both DNA repair pathways and the cell cycle will be investigated in naïve and primed hPSC in chapter 5.

1.4. Motivation and aims

The introductions to the main topics of this thesis illustrate several gaps in our understanding. Namely, we do not know which naïve hPSC generation protocols objectively yield the most authentic cells. We do not know the miRNAome of the human E5 inner cell mass. We do not know if the naïve to primed differentiation, used successfully as a model of pre-implantation embryogenesis in the mouse, also functions in the human, nor do we know what miRNAs are involved in that differentiation. Finally, we do not know how DNA repair pathway activity compares between naïve and primed hPSC, nor how the cell cycle dynamics compare.

This raises the following questions: Are we able to generate human pluripotent stem cells, which closely resemble the inner cell mass of an E5 blastocyst? What is the miRNAome of these human naive cells and do they have different miRNA expression profiles to conventional hPSC, as we would expect from the murine system? Does the naïve to primed differentiation also work in the human system? What miRNAs are involved? What are the rates of HR in naïve and primed hPSC and are the proportion of cells in G_1 phase different?

I intend to address these questions by framing the following hypotheses and using the following methodologies. “There are one or more naïve hPSC generation protocols producing authentic

naïve hPSC" and "Naïve hPSC resemble the cells from inner cell mass of an E5 blastocyst" will be addressed by *in silico* comparison of published *in vivo* human embryogenesis single cell RNAseq data to published RNAseq datasets from naïve hPSC and generation of naïve hPSC following published protocols and comparison by RT-qPCR. The hypotheses "The naïve to primed differentiation can be replicated in hPSC" as well as "The miRNAome of naïve, differentiating and primed hPSC are different" will be addressed by small-RNA sequencing on naïve, differentiating and primed cells, as well as RT-qPCR and miRNA qPCR. Finally, "Naïve and primed hPSC have different rates of HR" will be interrogated by measuring repair of Cas9 induced DSB by ddPCR and "Naïve and primed hPSC have the same cell cycle dynamics" by propidium iodide (PI) staining and flow cytometry.

1.5. Summary of thesis

In chapter 2, I describe the experimental reagents and methods used in this thesis.

In chapter 3, I compare published datasets to find the closest approximation of human *in vivo* E5 ICM cells and experimentally generate and compare naïve cells of most protocols. I find the 4iLA modified Theunissen *et al.* protocol the most convincing, as well as the direct derivation HNES lines. Finally, I adapt techniques to work with naïve hPSC to ensure the best conditions for chapters 4 and 5.

In chapter 4, I set up naïve to primed differentiation as a model for early embryogenesis and investigate miRNA expression in naïve, differentiating and primed states. I find that despite different genomic locations, the human miR-371-373 cluster shows similar expression to the homologous mouse miR-290 family. To further investigate miRNA requirements in early human embryogenesis, I end the chapter by designing a strategy to knock out *DGCR8*.

In chapter 5, I uncover that naïve hPSC repair DSB less often by HDR than primed cells. I at least partially explain this by finding that naïve hPSC reside significantly more in G_1 phase than primed cells.

In chapter 6, I summarise the results of the thesis and discuss their broader context. I extrapolate from my results and results generated by others in the field, to discuss the future of naïve hPSC culture as well as the hurdles faced in the short and long term.

2
Materials and methods

2.1. Cell culture

2.1.1. Cell counts

A NucleoCounter® NC-3000™ (chemometec) was used for cell counting. 19 µL of single cell suspension was added to 1 µL of solution 13 (910-3013; chemometec), mixed and added to an A8-slide (942-0003; chemometec). Solution 13 contains Acridine orange (AO) which stains all cells and 4',6-diamidino-2-phenylindole (DAPI) which stains only dead cells. The slide was then placed into the NucleoCounter® NC-3000™ which automatically images and counts cells. Live cell counts (ignoring DAPI+ cells) were used for all experiments.

2.1.2. Primed iPS cell culture

Stocks of primed iPSC were grown in a humidified incubator (Galaxy 170 R CO2 Incubator; New Brunswick) at 37°C, 5% CO2 on geltrex (A1413302; ThermoFisher Scientific) in E8 (A15169-01; ThermoFisher Scientific) or mTeSR™1 (5850; Stem Cell Technologies) medium in 6 well plates (657160; Greiner) or 10cm dishes (430167; Corning). To keep conditions consistent with naïve hPSC, primed cells used for any experiment were cultured at 7% CO_2 and 5% O_2. Cells were fed daily by aspirating all medium and replacing it with fresh, room temperature medium. The volume of medium used per well/dish: 96 well plate =100 µL, 24 well plate = 500 µL, 12 well plate = 1 mL, 6 well plate = 2 mL, 10 cm dish = 10 mL. Routine expansions in 6 well format were performed as follows: at approximately 80% confluency cells were passaged by aspirating medium, washing with 2 mL Phosphate-buffered saline (PBS; BE17-516F24; Lonza) and adding 1 mL PBS

supplemented with 0.5mM EDTA (0.5M stock; 15575-020 ThermoFisher Scientific) for 5 minutes at 37°C. The EDTA PBS was then carefully aspirated and cell aggregates were lifted by adding 6 mL of medium and by gently mixing until cell aggregates were of uniform size of approximately 10-20 cells (Beers et al., 2012). Cell aggregates were then split 2 mL each into fresh wells which were precoated with 1 mL of geltrex for 1-24h at 37°C and had the geltrex removed by aspiration just before adding the cell aggregates. Cells were thawed from large, karyotyped batches and kept in culture only for the duration of an experiment.

2.1.3. Preparing irradiated mouse embryonic fibroblasts

Irradiated mouse embryonic fibroblasts (MEFs) can be used as feeders to maintain PSC. MEFs were acquired from Millipore (PMEF-CFL) and thawed into 35 mL MEF medium (500 mL Advanced DMEM (12491-015; ThermoFisher Scientific) supplemented with 56 mL FBS (F4135; Sigma), 5.6 mL Glutamax (35050038; ThermoFisher Scientific) and 560 µL 50 mM beta mercaptoethanol (31350010; ThermoFisher Scientific)) into a T175 flask and were kept in a humidified incubator at 37°C, 5% CO_2 and normoxic conditions. MEFs were fed every other day with a full medium change of MEF medium and split every 4 days 1:4 as follows: Flasks were washed with PBS and 3 mL of TryPLE (12604013; ThermoFisher Scientific) was added. After 5 minutes at 37°C, the cells were dissociated by firmly tapping each flask and washing cells off with 9 mL MEF medium and rinsing again with an extra 12 mL MEF medium. Cells were then pooled and divided into 4 times as many flasks as before, each containing 30 mL prewarmed MEF medium. After 3 splits (16 flasks) and when the MEFs just reached confluency, they were dissociated as described above but pooled into PBS. MEFs were then counted and spun at 400 x *g* for 5 minutes and resuspended in a total of 15 mL MEF medium and transferred to a low attachment flask. After irradiating (30 gy) by Jane Vowles, they were further diluted in MEF medium to 2, 4 and 8 million cells per mL. 1 volume of 2X ice cold freeze medium (20% DMSO (D2650; Sigma), 20% KO-DMEM (10829-018; ThermoFisher Scientific), 60% Fetal Bovine Serum (F4135-500ML; Sigma)) was added to each and batches of 1 mL of 1, 2 and 4 million cells were frozen in cryovials by transferring to precooled (4°C) cryo-bombs

(Nalgene, “Mr Frosty”, containing Isopropanol (24137-2.5L-R; Sigma)) and storing at -80°C overnight. The following day the vials were transferred to -196°C nitrogen vapour storage.

2.1.4. Naïve media

Aliquots for growth factors human leukaemia inhibitory factor (LIF; 200 µg/mL in 0.1% BSA in water; 300-05 Peprotech) and Activin A (0.2 mg/mL in 0.1% BSA in PBS, 120-14 Peprotech) were stored at -80°C until thawed and then stored for up to 1 week at 4°C. Master mixes of inhibitors solved in DMSO were prepared for each media formulation and aliquots stored at -80°C. Inhibitor aliquots underwent less than 4 freeze-thaws. The basal media (without growth factors and inhibitors) were kept for up to a week at 4°C. All other components were stored according to the manufacturer’s recommendation.

Chan *et al*. uses mTeSR-1 as a basal medium. Daily, the following was added to 10 mL mTeSR-1: 0.5 µL LIF, 1 µL PD0325901 (PD03; 10 mM stock in DMSO; 1408; Axon Medchem), 2 µL BIO (10 mM stock in DMSO; B1686 Sigma) and 4 µL Dorsomorphin (5 mM stock in DMSO; P5499; Sigma) (Chan et al., 2013).

100 mL of Duggal *et al*. basal medium was made by adding 1 mL Non-Essential Amino Acids (NEAA; 11140-035; ThermoFisher Scientific), 1 mL Glutamax (35050038; ThermoFisher Scientific), 100 µL Beta Mercaptoethanol (50 mM stock dilution; 31350010; ThermoFisher Scientific), 100 µL ascorbic acid (50 µg/mL stock dilution in water; A5960 Sigma), 20 mL Knockout Serum Replacement (KSR; 10828-028; ThermoFisher Scientific), 160 µL bFGF (stock dilution 10 µg/mL in 0.1% BSA in PBS; 4114-TC-01M; R&D Systems Europe Ltd) to 76 mL Knockout-DMEM (KO DMEM; 10829-018; ThermoFisher Scientific). Daily, the following was added to 10 mL basal medium: 1 µL LIF, 1 µL CHIR99021 (CHIR; 30 mM stock in DMSO; 1386; Axon Medchem), 1 µL PD0325901 (PD03; 10 mM stock in DMSO; 1408; Axon Medchem) and 1 µL Forskolin (100 mM stock in DMSO; Sigma) (Duggal et al., 2015).

100 mL of RSeT™ Medium (05970; Stem Cell Technologies) was made by adding 20 mL of 5X supplement to 79.7 mL basal medium. 200 µL of 500X supplement was then added, followed by 100 µL of 1000X supplement. This medium was stored for up to 1 week at 4°C.

100 mL of NHSM1A was made by adding 1 mL NEAA, 1 mL Glutamax, 1 mL N2 supplement (17502048 ThermoFisher Scientific), 100 µL beta mercaptoethanol, 80 µL bFGF, 10 µL recombinant human TGFβ1 (10 mg/mL stock in 10 mM Citric Acid, pH 3.0 and 0.1% BSA in ddH_2O; 100-21 Peprotech), 80 µL recombinant human insulin (10 mg/mL stock; I9278 Sigma) and 10 mL AlbuMAX1 (100 mg/mL stock; 11020021; ThermoFisher Scientific) to 85 mL KO DMEM. Daily, the following was added to 10 mL basal medium: 1 µL LIF, 1 µL CHIR, 1 µL PD03, 1 µL Gö6983 (GO; 50 mM stock in DMSO; 2285; TOCRIS), 1 µL BIRB796 (2 mM stock in DMSO; 1358; Axon Medchem), 1 µL SP600125 (SP6; 10 mM stock in DMSO; 1496; TOCRIS) (Gafni et al., 2013).

100 mL of NHSM1B was made by adding 1 mL NEAA, 1 mL Glutamax, 1 mL N2 supplement, 2 mL B27 supplement (17504044; ThermoFisher Scientific), 100 µL beta mercaptoethanol, 1 mL KSR, 125 µL recombinant human insulin, 100 µL L-ascorbic acid 2-phosphate (Vitamin C; 50 mg/mL stock in water; A8960 Sigma), 70 µL BSA Fraction V 7.5% solution (15260037; ThermoFisher Scientific), 400 µL Defined Lipid Concentrate (11905031; ThermoFisher Scientific) to 46 mL DMEM/F12 (11320; ThermoFisher Scientific) and add 46 mL Neurobasal (21103; ThermoFisher Scientific). Added daily to 10 mL basal medium: 1 µL LIF, 1 µL Activin A, 0.2 µL CHIR, 1 µL PD03, 0.4 µL GO, 1 µL BIRB796, 2.5 µL IWR1 (20 mM in DMSO; I0161; Sigma).

100 mL of ENHSM was made by adding 1 mL NEAA, 1 mL Glutamax, 1 mL N2 supplement, 2 mL B27 supplement, 100 µL beta mercaptoethanol, 200 µL defined lipid concentrate, 75 µL recombinant human insulin, 700 µL sodium pyruvate (100 mM stock; 11360070; ThermoFisher Scientific), 100 µL Vitamin C, 70 µL BSA Fraction V 7.5% solution, 500 µL KSR to 46 mL DMEM/F12 and add 46 mL Neurobasal. Added daily to 10 mL of basal medium: 1 µL LIF, 1 µL Activin A, 0.5 µL CHIR, 1 µL PD03, 0.4 µL GO, 1 µL BIRB 796, 5 µL IWR1, 0.5 µL SP6.

100 mL of 5iLAF was made by adding 1 mL NEAA, 1 mL L-Glutamine (100mM stock in ddH_2O; G6392; Sigma), 1 mL N2 supplement, 2 mL B27 supplement, 500 µL KSR, 100 µL BSA (50 mg/mL stock in ddH_2O; A7979; Sigma) and 80 µL bFGF to 48 mL DMEM/F12 and add 48 mL Neurobasal. Added daily to 10 mL of basal medium: 1 µL LIF, 1 µL Activin A, 1 µL PD03, 1 µL IM-12 (10 mM stock in DMSO; SML0084; Sigma), 1 µL SB590885 (SB59; 5 mM stock in DMSO; 2650; TOCRIS), 1 µL WH-4-023 (10 mM stock in DMSO; 2381; Axon Medchem) and 10 µL Y-27632 Rho kinase inhibitor (ROCKi; 1 mM stock in ddH_2O; ab 120129; Abcam)(Theunissen et al., 2014).

100 mL of 4iLA was made by adding 1 mL of NEAA, 500 µL of L-Glutamine, 1 mL of N2 supplement, 2 mL of B27 supplement, 200 µL of beta mercaptoethanol and 16.7 µL of BSA to 48 mL DMEM/F12 and add 48 mL Neurobasal. Added daily to 10 mL of basal medium: 1 µL LIF, 1 µL Activin A, 1 µL PD03, 1 µL SB59, 1 µL WH-4-023 and 10 µL ROCKi. (Theunissen et al., 2016, Sahakyan et al., 2017)

100 mL of t2iLGoY was made by adding 1 mL L-Glutamine, 500 µL home-made N2 (100 µg/mL apo-transferrin (ABC2553; eBioscience), 3 µM sodium selenite (S5261; Sigma), 1.6 mg/mL putrescine (P6024; Sigma), 2 µg/mL progesterone (P6149; Sigma) in DMEM/F12), 1 mL B27 supplement, 200 µL beta mercaptoethanol, 88 µL ascorbic acid and 100 µL recombinant human insulin to 49 mL of DMEM/F12 and add 49 mL Neurobasal. Added daily to 10 mL of basal medium: 1 µL LIF, 1 µL PD03, 0.33 µL CHIR, 0.5 µL Go. (Guo et al., 2016)

2.1.5. Naïve conversions & maintenance

Cells were cultured in a humidified incubator (Galaxy 170 R CO2 Incubator; New Brunswick) at 37°C, 7% CO2 and 5% O_2. At 80% confluency primed iPSC were dissociated to single cell suspension by washing each well of a 6 well plate with 2 mL PBS, adding 1mL of accutase (A1110501; ThermoFisher Scientific) for 5 minutes at 37°C and gently mixing with a P1000 pipette tip. Accutase was then inactivated by diluting the cell suspension 1:5 in 10% embryonic stem cell qualified fetal bovine serum (ES FBS; 16141-079; ThermoFisher Scientific) in KnockOut™ DMEM (KO DMEM; ThermoFisher Scientific). Cells were counted as described above and spun at 400 x *g* for 5 minutes.

After aspirating the supernatant, the pellets were gently resuspended in mTeSR1 + ROCKi and 50k and 150k (5.5k and 16.7k cells/cm^2) were each plated in 2 mL mTeSR1 onto 13 wells of irradiated MEFs. The MEFs were plated 24-48h before by pre-treating each well with 1 mL of 0.1% Gelatin (G1393; Sigma) in PBS for >15 min at room temperature, thawing a fresh vial of irradiated MEFs, spinning them in 5 volumes of MEF medium and resuspending them to 150k cells/mL MEF medium, to seed 2mL per well of a 6 well plate (33.3k cells/cm^2). After 24h the medium was changed to naïve media. Each of the two cell concentrations had one well each of the following media: Chan *et al*., Duggal *et al*., RSeT (5970; Stem Cell Technologies), NHSM 1A, NHSM 1B, ENHSM, 5iLAF and mTeSR1. Cells were fed daily with their naïve medium and were split as above (resuspending in their naive medium) every 2-4 days using the most appropriate of the two cell densities and always splitting again to 50k and 150k cells per well (5.5k and 16.7k cells/cm^2). When not already part of the formulation, ROCKi was added to the medium when cells were split. Cells remaining after the split were pelleted and frozen at -80°C. After 23 days the pellets were used for RNA extraction and qRT PCR. After one further passage, remaining cells were frozen.

Conversion of cells with 4iLA was achieved by dissociating and plating cells as explained above. Due to expecting high cell death, more cells were plated down: 100k and 200k per well (11.1k and 22.2k cells/cm^2). After 24 hours medium was changed to 4iLA and on subsequent days cells were fed daily with a 70% media change of 4iLA. Cells were split 1:1-1:3 every 3-5 days cells by PBS washing, adding 1 mL of Accumax (A7089; Sigma), incubating at 37°C for 4 minutes and dissociating by pipetting with a P1000. Naïve cells were stickier and hardier than conventional primed cells and needed to be vigorously dissociated. Cells were then diluted in 10% ES FBS in KO DMEM to inactivate Accumax and spun at 400 x *g* for 5 minutes, resuspended in the appropriate volume of 4iLA and plated down on irradiated MEFs, prepared as discussed above.

HNES cells were grown on MEFs and maintained in t2iLGoY. Medium was replaced by full media change daily and split every 2-4 days 1:2-1:4, the same as 4iLA cells.

2.1.6. Priming naïve cells

To prime naïve cells, 1 – 5 x 10^4 cells/cm^2 were plated onto a well of a six well plate (precoated with geltrex for 1-24h). Unless indicated otherwise, medium was switched to E8 after 24 hours and from then on, a full media change performed every 24h.

2.1.7. Freezing and thawing

To freeze, naïve cells were dissociated with Accumax, primed with accutase as discussed above. After spinning, cells were resuspended in their regular medium at 2 x 10^6 cells/mL for primed cells and 0.75 mL per dissociated well for naïve cells. Then an equal volume of ice-cold 2X freeze medium was added. 2X freeze medium consisted of 20% DMSO (D2650; Sigma), 20% KO-DMEM (10829-018; ThermoFisher Scientific), 60% ES FBS. Cells were briefly and gently mixed and aliquoted into 1 mL in pre-cooled (4°C) cryovials. These were then transferred into pre-cooled (4°C) cryo-bombs (containing isopropanol; C1562-1EA; Sigma) and stored overnight at -80°C. The following day the vials were transferred to -196°C nitrogen vapour storage.

To thaw cells, a vial was swirled in a 37°C stirred water bath (GR150 Grant) until only a pea sized piece of ice was left. This was then transferred to at least 4 volumes of PBS, spun at 400 x *g* for 5 minutes and resuspended in the appropriate medium containing 10 µM ROCKi. Primed cells were then seeded onto 2 wells pre-treated with geltrex (1-24h at 37°C) with 80% cells in one and 20% cells in the other well. Naïve cells were plated on 1 well of MEFs prepared as described above.

2.2. Microscopy

Phase contrast images were taken either using an EVOS FL auto or EVOS XL Core (AMAFD1000; AMEX1000; ThermoFisher Scientific) and image analysis performed using ImageJ.

2.3. Primers and probes

For gene expression analysis, only TaqMan Gene Expression Assays (4331182; ThermoFisher Scientific; Table 2.1) that bridge exon-exon junctions were chosen, to ensure no contaminating genomic DNA was detected. Primers and probes used in this thesis are listed in Table 2.2.

TaqMan® Advanced miRNA Assays	
hsa-miR-106a-5p	478225_mir
hsa-miR-135a-5p	478581_mir
hsa-miR-142-5p	477911_mir
hsa-miR-18b-3p	478734_mir
hsa-miR-199a-5p	478231_mir
hsa-miR-205-3p	478760_mir
hsa-miR-205-5p	477967_mir
hsa-miR-20b-3p	478764_mir
hsa-miR-20b-5p	477804_mir
hsa-miR-214-3p	477974_mir
hsa-miR-28-3p	477999_mir
hsa-miR-29c-5p	478005_mir
hsa-miR-363-3p	478060_mir
hsa-miR-363-5p	478840_mir
hsa-miR-4417	478896_mir
hsa-miR-4497	479836_mir
hsa-miR-486-5p	478128_mir
hsa-miR-515-5p	478147_mir
hsa-miR-519d-5p	478987_mir
hsa-miR-653-5p	479134_mir
hsa-miR-766-3p	478342_mir
hsa-miR-767-5p	479176_mir
hsa-miR-7704	480576_mir
hsa-miR-877-3p	477856_mir
hsa-miR-92a-2-5p	479206_mir

TaqMan® Gene Expression Assays	
KLF4	Hs00358836_m1
NANOG	Hs04399610_g1
DNMT3L	Hs01081364_m1
DPPA3/STELLA	Hs01657068_g1
POU5F1/OCT4	Hs01895061_u1
TBX3	Hs00195612_m1
TFCP2L1/LBP9	Hs00232708_m1
CER1	Hs00193796_m1
OTX2	Hs00222238_m1
B2M	Hs99999907_m1
RPL13A	Hs01578912_m1
18s	Hs99999901_s1
DPPA5	Hs00988349_g1
KLF17	Hs00702999_m1
ZFP57	Hs01378213_g1
HMGA2	hs04397751_m1
TWF1	Hs00702289_s1
GRPEL2	Hs01128446_m1
C5orf51	Hs03037752_m1
ABT1	Hs01092221_g1
ATP6V1G1	Hs00895280_g1

TaqMan® Copy Number Assays	
NANOG	Hs02550507_cn
RNase P control	#4403328

Table 2.1: List of TaqMan assays used in this thesis.

Probes		
Name	**Sequence (5'->3')**	**Fluor-Quencher**
GRN Assay reference probe	CAGGAACATAATGCCATTCTGTGC	FAM-Zen
GRN Assay HDR probe	CAGGATCAAGCCTTCACG	FAM-Zen
GRN Assay Cas9 F1/R2 NHEJ	CTGCGAGAAGGAAGTGGT	HEX-Zen
GRN Assay Dark probe for Cas9	CAGGATCGAGCCTTCACG	None. 3' phosphate
RBM20 Assay reference probe	TGGGAGGTGTGAAGATTCTAAATC	FAM-Zen
RBM20 Assay HDR probe	CCGCGGTCTAGTAGTCC	FAM-Zen
RBM20 Assay Cas9 F1/R2 NHEJ probe	AGAGTGACCGGCTCAC	HEX-Zen
RBM20 Assay dark probe	CCGCGGTCTCGTAGTCC	None. 3' phosphate
ATP7B Assay Reference	TGGTGGTTGCTGTGGCT	FAM-Zen
ATP7B Assay HDR	TGGGCCTGTGGCTG	FAM-Zen
ATP7B Assay Cas9 F1/R2 NHEJ	CTGGGCCGGTGGCTG	HEX-Zen
ATP7B Assay Dark probe for Cas9	CTGGGCCGGTGGCT	None. 3' phosphate

Primers and oligos	
Name	**Sequence (5'->3')**
GRN Assay forward primer	CTGGATAGGGGAGCTAAG
GRN Assay reverse primer	GTCTGGTTATCATGGCAG
GRN Assay HDR template	CGGCTGGCTACACCTGCAACGTGAAGGCTTGATCCTGCG AGAAGGAAGTGGTCTCTGCCC
RBM20 Assay forward primer	CTGTGTGTGGGTGGGGT
RBM20 Assay reverse primer	AGGAGGTGAAGCTGGGAG
RBM20 HDR template	ACAGATATGGCCCAGAAAGGCCGCGGTCTAGTAGTCCGG TGAGCCGGTCACTCTCCCCGA
ATP7B Assay forward primer	TGCTTATGTTTATTCTCTGGTCATC
ATP7B Assay reverse primer	CCTGAAGCTGCTGTTACCTT
ATP7B Assay HDR template	CATGCTCTTTGTGTTCATTGCCCTGGGCCTGTGGCTGGA ACACTTGGCAAAGGTAACAGC
ATP7B Assay HDR template + PTO modification	C*A*TGCTCTTTGTGTTCATTGCCCTGGGCCTGTGGCTG GAACACTTGGCAAAGGTAACA*G*C
BD131	CCCCTACGAGGTATGTTGGC
BD132	ACTCCTTTGCTGCCTGTCTG

Table 2.2: List of Probes, primers and oligos.

2.4. RT-qPCR

RNA extractions from cell pellets (equal or less than 1 x 10^6 cells) were performed using either the RNeasy micro kit (for < 1 x 10^5 cells; 74004; Qiagen) or the RNeasy mini kit (> 1 x 10^5 cells; 74104; Qiagen) following the manufacturers protocol. Briefly, pellets were thawed at room temperature and 350 µL of RLT buffer was added. Cells were lysed and homogenised by vortexing for 1 minute. 1 volume of 70% ethanol (diluted in ddH_2O; 32221-2.5L-M Sigma) was added, mixed by pipetting, and the whole sample was transferred to a spin column. Columns were spun at >8000 x *g* for 15s and flow-through was discarded. 350 µL Buffer RW1 was added to the column, the column spun at >8000 x *g* for 15s and the flow-through discarded. To remove contaminating DNA, 10 µL DNase I stock solution (79254; Qiagen) was added to 70 µL buffer RDD per sample and mixed by inverting. 80 µL DNase mix was applied to each column and incubated at room temperature for 15 min. 350 µL RW1 was then added, columns were spun at >8000 x *g* for 15s and the flow-through was discarded. (When using the micro kit, the column was transferred to a new collection tube.) 500 µL Buffer RPE was then added to each column and spun for >8000 x *g* for 2 minutes (15s for micro kit). An extra wash of 500 µL 80% Ethanol was included only the micro kit, after which the column was spun at > 8000 x *g* for 2 minutes. Columns were then placed into a new collection tube and spun empty at full speed for 1 min to dry the membrane and again, placed into a new collection tube. 30 µL RNase free water (14 µL for the micro kit) was added and tubes were spun at > 8000 x *g* for 1 min to collect the purified RNA. If not used immediately, RNA was stored at -80°C. Note: RNA from samples used for both miRNAseq and qRT-PCR used a phenol-chloroform based extraction to include small RNAs (method in miRNAseq section below).

RNA levels were measured using a NanoDrop 2000/2000c (ND-2000C; ThermoFisher Scientific) and then normalised across all samples within an experiment by diluting in RNase free water. To make cDNA the High Capacity RNA to cDNA kit (4387406; ThermoFisher Scientific) was used: 1 µg (except where samples were limiting) of RNA in 9 µL was added to 11 µL premixed master mix and enzyme. The premixed master mix contains 1 µL enzyme and 10 µL of master mix per sample.

Samples were then incubated at 37°C for 1 hr, heated to 95°C for 5 minutes and kept at 4°C until further use or stored at -20°C. This kit utilises both random octamers, and oligo dT-16 for cDNA generation.

Regular qPCR was performed by making a mastermix containing 1 volume of TaqMan primer-probes (Table 2.1), 10 volumes TaqMan™ Gene Expression Master Mix (4369016; ThermoFisher Scientific) and 2.33 volumes qPCR grade H_2O. cDNA was diluted 1:40 -1:100 in qPCR grade H_2O and 1 volume of cDNA was added per 2 volumes mastermix. Samples were run either in 96 well format (15 µL per well, in duplicate or triplicate) on a StepOnePlus (4376600; ThermoFisher Scientific) or in 384 well format (10 µL per well, quadruplets) on a Lightcycler 480 (05015243001; Roche Life Science). Plates were sealed with MicroAmp™ Optical Adhesive Film (4311971; ThermoFisher Scientific) and spun at 1000 x *g* for 2 minutes before running. Care was taken to avoid or remove any bubbles and 384 well plates were spun longer and faster if required.

Analysis was performed by normalising against the average Ct of at least two endogenous control housekeeping genes. Housekeeping genes used were *B2M*, *RPL13A* and *18s*. If the difference to a control was of interest data is expressed as fold change ($2^{-\Delta\Delta ct}$) and can be compared to any other data expressed equally. Alternatively, data is expressed as $2^{-\Delta ct}$ where a comparison between samples of the same experiment but not between genes or experiments can be made. The $2^{-\Delta ct}$ method normalises Ct values of the transcript of interest to control housekeeping genes. This removes variation introduced by the cDNA reaction but assumes that the housekeeping genes are stably expressed across all samples and that the primer efficiencies are approximately 100%. (Livak and Schmittgen, 2001). This method was used in this thesis, since expression of the housekeeping genes was assessed and the primer efficiencies of TaqMan primer/probes have been validated.

2.5. Karyotyping

DNA was extracted from pellets containing approximately 1 x 10^6 cells using the DNeasy blood and tissue kit (69506; Qiagen) following the manufacturer's instructions. Briefly, pellets were thawed and resuspended in 200 µL PBS. To lyse the cells and remove RNA, 20 µL of proteinase K and 4 µL of 100 mg/mL RNase A was added and the sample vortexed thoroughly. After 2 min incubation at room temperature, 200 µL buffer AL was added and mixed thoroughly by vortexing. After 10 min incubation at 56°C, 200 µL of ethanol (Sigma) was added and mixed by vortexing. The lysed sample was then transferred to a spin column and spun at >6000 x *g* for 1 minute. The collection tube was replaced with a new tube and 500 µL buffer AW1 was added. The column was spun at > 6000 x *g* for 1 min and the collection tube replaced with a fresh tube. 500 µL buffer AW2 was added and the column centrifuged for 1 min at > 6000 x g. To dry the membrane, the collection tube was replaced again, and the column spun for 3 min at 20,000 x g. The column was then transferred to a new collection tube and DNA was eluted by adding 100 µL qPCR grade H_2O, incubating for 1 min at room temperature, and spinning for 1 min at > 6000 x g.

DNA samples (500 ng in 10 µL) were processed by the Wellcome Trust Centre for Human Genetics, Oxford, using Illumina Human CytoSNP-12v2.1 beadchip array (300,000 SNPs) or OmniExpress24 array (700,000 SNPs). Resulting image data files were analysed by GenomeStudio and Karyostudio software (Illumina) to assess genome integrity.

2.5.1. Measuring copy number variation

Copy number was investigated by qPCR and was performed by making a mastermix containing 1 volume of NANOG (FAM) and 1 volume of RNase P (VIC) TaqMan primer-probe (Table 2.1; 4400291; ThermoFisher Scientific), 10 volumes TaqMan™ Genotyping Master Mix (4371355; ThermoFisher Scientific) and 1.33 volumes qPCR grade H_2O. Genomic DNA was diluted to 5 ng/µL and 1 volume was added to 2 volumes of mastermix. Samples were run in 96 well format, with 15 µL per well and in triplicate and plates were sealed with MicroAmp™ Optical Adhesive Film and

spun at 1000 x *g* for 2 minutes before running on a StepOnePlus. Thresholds were set the same for both RNase P and NANOG and equal ct's for both was observed for a healthy wild type control. Data is expressed as $2^{-\Delta ct}$ or copies of *NANOG* per *RNase P*.

2.6. Flow cytometry

All centrifugation in this section was performed at 400 x *g* for 5 minutes. Antibodies used in flow cytometry are listed in Table 2.3 with their recommended dilution and with their used dilution after titration. Antibody staining was performed by lifting the cells by washing with PBS and adding 1 mL accutase per well of a 6 well plate (1-2 wells), incubating for 4 minutes at 37°C and washing them off with a P1000 pipette. They were then passed through a 40 µm Nylon Cell Strainer (352340; SLS Ltd) to remove cell clumps and diluted 1:5 in 10% ES FBS. Cells were counted, spun, and resuspended to 5 x 10^6 cells/mL in PBS. Per stain, 100 µL of cells were added to separate microfuge tubes (spike controls were made by mixing 50 µL naïve and 50 µL primed cells) and centrifuged. The supernatant was removed, and cells were labelled by resuspension in 100 µL FACs buffer (10 µg/mL human IgG (I2511; Sigma), 1% Fetal Bovine Serum (FCS; 10500064 ThermoFisher Scientific), 0.01% sodium azide (769320 Sigma) in PBS) containing the antibodies (Table 2.3). Cells were incubated at 4°C for 30 mins, washed twice by spinning, removing supernatant and resuspending in 2% FBS in PBS. Finally, cells were resuspended in FACs buffer with DAPI (0.5 µg/mL) to stain nuclei, transferred to a FACs tube and ran through the flow cytometer. Gating was set by using spike controls. (Collier et al., 2017)

Propidium Iodide (PI) staining for cell cycle analysis was performed by lifting and passing the cells through the cell strainer into 10% FBS as above. Cells were transferred to a 15 mL conical tube, counted, spun and resuspending in 1.5 x 10^6 in PBS. 1 mL was transferred to a new conical tube, spun and the supernatant discarded. Cells were then thoroughly resuspended in cold PBS, ensuring no cell clumps. Cells were fixed by adding 5mL ice cold (-20°C) 100% ethanol dropwise to the tube while gently vortexing. Fixed cells were kept at -20°C for up to several weeks, until several samples were ready for staining. Staining of the DNA was performed by spinning the cells,

Antibodies						
Species	**Target**	**Conjugate**	**Manufacturer**	**Catalog#**	**Recommended dilution**	**Optimised dilution**
Mouse anti human	CD130	PE	BD Biosciences	555757	1:5	1:5
Mouse anti human	CD57	FITC	BD Biosciences	555619	1:5	1:5
Mouse anti human	CD24	FITC	BD Biosciences	560992	1:5	1:20
Mouse anti human	CD90	APC	BD Biosciences	561971	1:20	1:20
Rat anti mouse	CD90.2	Pacific Blue	BioLegend	105323	1:50	1:100

Table 2.3: List of antibodies and antibody dilutions used in this thesis.

miRNA inhibitors	
Name	**inhibited sequence**
hsa-miR-106a-5p	AAAAGUGCUUACAGUGCAGGUAG
hsa-miR-20b-5p	CAAAGUGCUCAUAGUGCAGGUAG
hsa-miR-363-3p	AAUUGCACGGUAUCCAUCUGUA
hsa-miR-4417	GGUGGGCUUCCCGGAGGG
hsa-miR-205-5p	UCCUUCAUUCCACCGGAGUCUG
hsa-miR-4488	AGGGGGCGGGCUCCGGCG
hsa-miR-7704	CGGGGUCGGCGGCGACGUG
hsa-miR-28-3p	CACUAGAUUGUGAGCUCCUGGA
hsa-miR-486-5p	UCCUGUACUGAGCUGCCCCGAG
hsa-miR-92b-3p	UAUUGCACUCGUCCCGGCCUCC
hsa-miR-561-5p	AUCAAGGAUCUUAAACUUUGCC
hsa-miR-3178	GGGGCGCGGCCGGAUCG
hsa-miR-7974	AGGCUGUGAUGCUCUCCUGAGCCC
hsa-miR-519d-5p	CCUCCAAAGGGAAGCGCUUUCUGUU
hsa-miR-92a-3p	UAUUGCACUUGUCCCGGCCUGU
hsa-miR-302a-3p	UAAGUGCUUCCAUGUUUUGGUGA
hsa-miR-302b-3p	UAAGUGCUUCCAUGUUUUAGUAG
hsa-miR-302c-3p	UAAGUGCUUCCAUGUUUCAGUGG
hsa-miR-302d-3p	UAAGUGCUUCCAUGUUUGAGUGU

Table 2.4: List of miRNA inhibitors used.

CRISPR guide RNAs	
Name	**Target sequence (5'->3') excluding PAM**
GRN gRNA-F1	GAAGGCTCGATCCTGCGAGA
GRN gRNA-R2	AGAGACCACTTCCTTCTCGC
RBM20 gRNA-F1	GGTCTCGTAGTCCGGTGAGC
RBM20 gRNA-R2	GGGGAGAGTGACCGGCTCAC
ATP7B gRNA-F1	CATTGCCCTGGGCCGGTGGC
ATP7B gRNA-R2	TTGCCAAGTGTTCCAGCCAC
DGCR8 left 1	GGGATGCAGGGCCTACCCAT
DGCR8 right 1	TTGCAGGCAGGGTAAACCTG
DGCR8 left 2	TGCAGGAGCATGAGCGCCAG
DGCR8 right 2	CAGGGTCAGGGCCACTTGGT

Table 2.5: List of CRISPR guide RNAs used.

removing the supernatant, and resuspending the cells in 100 µL PI staining solution (50 µg/mL PI + 100 µg/mL RNAse A in PBS; 556463; BD Biosciences). Cells were filtered using a 40 µm filter (352235; BD Biosciences) to remove cell clumps and incubated at 4°C in the dark for 20 min. They were then analysed on a flow cytometer (BD LSRFORTESSA X-20; BD Biosciences). (Krishan, 1975)

2.7. miRNA analysis

2.7.1. Total RNA extraction

To extract RNA including the very short mature miRNAs, total RNA was extracted using the miRNeasy mini kit (217004; Qiagen) which combines TRIzol extraction with column purification. Pellets of 1 x 10^6 cells were disrupted and homogenised by adding 700 µL QIAzol and vortexing for 1 minute. After 5 minutes incubation at room temperature 140 µL chloroform was added, the tube vigorously shaken for 15s and incubated at room temperature for 2-3 minutes. The tubes were then centrifuged for 15 minutes at 12,000 x *g* at 4°C to separate the sample into three phases. The upper aqueous phase containing RNA was carefully transferred to a new tube and 1.5 volumes of 100% Ethanol was added, the sample thoroughly mixed and transferred to a RNeasy Mini spin column. After spinning at >8000 x *g* for 15s, the flow-through was discarded and 350 µL Buffer RWT was added. The column was then centrifuged for 15s at > 8000 x *g* and flow-through discarded. To remove contaminating DNA, 10 µL DNase I stock solution was added to 70 µL Buffer RDD per sample and 80 µL was added directly onto each column membrane. Columns were incubated at room temperature for 15 minutes and 350 µL of Buffer RWT was added, the columns centrifuged for 15s at > 8000 x *g* and the flow-through was discarded. The membrane was washed by adding 500 µL of Buffer RPE into the spin column and centrifuging the columns for 15s at >8000 x g. Flow-through was discarded and a further 500 µL Buffer RPE was added. After spinning for 2 minutes at > 8000 x *g* the spin column was transferred to a new collection tube and centrifuged at full speed for 1 minute to dry the membrane. Finally, the columns were transferred to new collection tubes and 30 µL RNase-free water was added directly to elute the RNA. After centrifuging at > 8000 x *g* for 1 minute, the RNA levels were measured using a nanodrop (as

described in section 2.4). The RNA was then used for miRNAseq, conventional reverse transcription and qPCR as outlined above, and miRNA qPCR described below. Remaining RNA was stored at -80°C.

2.7.2. miRNA qPCR

Since miRNAs are only ~20 nt in length, conventional amplification and qPCR methods are impossible. To circumvent this, poly(A) tailing on the 3' end of the miRNA together with adaptor ligation on the 5' end add length and allow complementary 3p and 5p miRNAs to be distinguished from each other. Reverse transcription was performed using a universal primer binding the poly(A) tail and an amplification (miR-Amp) reaction increases abundance in signal by using two universal primers. The TaqMan advanced miRNA assay cDNA synthesis kit (A28007; ThermoFisher Scientific) was used to achieve this. To avoid RNA degradation, the protocol was performed swiftly, and samples were kept on ice. Each master mix was made with an extra 10% overage and particular care was given to ensure they were entirely mixed.

Poly(A) tailing was performed by making a master mix containing 0.5 µL 10X Poly(A) Buffer, 0.5 µL ATP, 0.3 µL Poly(A) Enzyme and 1.7 µL RNase-free water per sample. The reaction mix was vortexed and spun. All RNA samples were diluted to 5 ng/µL and 2 µL of RNA was added to 3 µL reaction mix. After vortexing and spinning, the tubes were incubated in a thermal cycler (MJ Research PTC-200 DNA Engine; Bio-Rad): Polyadenylation proceeded at 37°C for 45 minutes, the reaction was stopped by heating to 65°C for 10 minutes and samples were then cooled to 4°C before proceeding immediately to adapter ligation.

The adapter ligation reaction mix was prepared containing 3 µL 5X DNA ligase buffer, 4.5 µL 50% PEG 8000, 0.6 µL 25X Ligation Adaptor, 1.5 µL RNA Ligase and 0.4 µL RNase-free water per reaction. The reaction mix was thoroughly mixed and 10 µL added to each sample. The tubes were vortexed and spun before performing the ligation at 16°C for 60 minutes. Samples were then cooled to 4°C before moving on to the reverse transcription.

A reverse transcription master mix was made by mixing 6 µL of 5X RT Buffer, 1.2 µL dNTP Mix, 1.5 µL 20X Universal RT Primer, 3 µL 10X RT Enzyme Mix and 3.3 µL RNase-free water. The reaction mix was vortexed, spun and 15 µL of the reaction mix was added to each sample and the samples mixed. Reverse transcription was induced by incubating for 42°C for 15 minutes and the reaction stopped by heating to 85°C for 5 minutes. The samples were cooled to 4°C before proceeding to the miR-Amp reaction.

The miR-Amp master mix contained 25 µL 2X miR-Amp Master Mix, 2.5 µL miR-Amp Primer Mix and 17.5 µL RNase-free water per sample. The three constituent parts of the 45 µL of the reaction mix were added to a new tube each and 5 µL of the RT reaction product was added (remaining pre-Amp cDNA was frozen at -80°C). The tubes were then vortexed and spun. Amplification was performed in a thermal cycler with the following settings: 95°C for 5 minutes, 14 cycles of: 95°C for 3s and 60°C for 30s, a 10 minute 99°C stop reaction and finally the samples were cooled to 4°C and the undiluted and amplified cDNA was stored at -20°C until further use.

To perform the qPCR, the cDNA was diluted 1:18 and the master mix was prepared by adding 1 volume of 20X TaqMan Advanced miRNA Assay (A25576; ThermoFisher Scientific) to 10 volumes of 2X TaqMan Fast Advanced Master Mix (4444557; ThermoFisher Scientific). Most reactions were performed in quadruplets in 384 well format in a 7900HT Fast Real-Time PCR System with 384-Well Block Module (4329001; ThermoFisher Scientific) by pipetting 5.5 µL master mix per well and adding 4.5 µL diluted cDNA. Alternatively, qPCR was performed in 96w format in duplicate by adding 6.75 µL diluted sample to 8.25 µL master mix in a StepOnePlus (4376600; ThermoFisher Scientific). The TaqMan miRNA assays used in this thesis are listed in Table 2.1.

Analysis was performed by normalising against the average Ct of the endogenous controls hsa-miR-191-5p, hsa-miR-361-5p, hsa-miR-423-5p, hsa-miR-26a-5p and hsa-miR-186-5p. Data is expressed as $2^{-\Delta ct}$. (Livak and Schmittgen, 2001)

2.7.3. miRNAseq

Purified total RNA was sent to EA | Q^2 Solutions (5927 S. Miami Blvd., Suite 100, Morrisville, NC 27560, US) who performed library preparation and miRNA sequencing. Dr Klas Hatje performed the miRNA processing and gene set enrichment (Supplementary Methods 7.1.2 and 7.1.3). Further analysis was performed by myself using Orange (Demsar J, 2013), Excel 2016 (Microsoft) and GraphPad Prism 7 (GraphPad Software).

2.8. Electroporation of mimics and inhibitors

Exiqon power locked nucleid acids (LNA; 339130; Qiagen) and MirVana miRNA mimics and inhibitors (4464066; 4464084; ThermoFisher Scientific) were resuspended 100 µM in ddH_2O. miRNA inhibitors from Integrated DNA Technologies (Table 2.4) were resuspended 100 µM in the supplied IDTE, pH 8 buffer. Delivery of miRNA inhibitors/mimics was achieved by electroporation using the Neon Electroporation System (MPK5000; ThermoFisher Scientific). Cells were dissociated using accutase as described above and passed through a 40 µm cell strainer to remove aggregates. The accutase was inactivated by diluting 1:5 in 10% FBS in KO DMEM. Cells were resuspended in 1 mL PBS, counted (diluting 1:10), spun at 400 x *g* for 5 minutes and resuspended in 1 x 10^7 cells/mL in Buffer R. For the titration of inhibitors and mimics, 9 µL cells were added to 1 µL of either: neat (final 10 µM), 1:4 (2.5 µM), 1:20 (500 nM) or 1:100 (100 nM) of the inhibitor/mimic. Cells were gently mixed, taken up into the 10 µL electroporation pipette tip (MPK1025; ThermoFisher Scientific) and electroporated using programme A3 unless specified otherwise and plated into 24 well format. For time course experiments, 90 µL cells were added to 10 µL mimic/inhibitor, electroporated in a 100 µL electroporation pipette tip (MPK10025; ThermoFisher Scientific) and added to 4 mL medium (containing ROCKi) and split across 4 wells of a 12 well plate. For the inhibitor pool experiment cells were also electroporated with a 100 µL electroporation tip, added to 4 mL medium (containing ROCKi) and split between 2 wells of a 6 well plate.

2.9. Assay to measure HDR & NHEJ

The assay to measure homology directed repair (HDR) and non-homologous end joining was based on Miyaoka *et al.* (Miyaoka et al., 2016), adapted here to work with Alt-R® CRISPR-Cas9 System procured from Integrated DNA Technologies and applied to hPSC. The assay requires the introduction of a Cas9 induced double strand break, extraction of genomic DNA and uses a droplet digital PCR based method for HDR and NHEJ quantification.

2.9.1. Preparation of ribonucleoproteins (RNPs)

Alt-R® CRISPR-Cas9 crRNA (Table 2.5) and Alt-R® CRISPR-Cas9 tracrRNA (1072532; Integrated DNA Technologies) were resuspended in 100 µM in the supplied duplex buffer. To allow binding of crRNA and tracrRNA, 2 volumes of each were added to 1 volume of duplex buffer. The mixture was heated to 95°C for 5 minutes, removed from heat and allowed to cool on the bench. This guide RNA mixture was used directly or stored at -20°C for future use. Alt-R® S.p. HiFi Cas9 Nuclease (1081060; Integrated DNA Technologies) was diluted by adding 3 volumes to 2 volumes of buffer R (from the Neon electroporation kit). Diluted Cas9 was then mixed in equal volumes with the prepared guide RNAs and allowed to incubate for 20 minutes. The single stranded template oligo (Table 2.2) and the non-targeting oligo control (Alt-R® Cas9 Electroporation Enhancer; 1075915; Integrated DNA Technologies) were resuspended in duplex buffer to 100 µM. They were used directly and stored at -20°C for further use. The oligos were further diluted in buffer R to 10.8 µM and 1 volume of ribonucleoprotein (RNP) was mixed with 2 volumes of the template oligo or the control Alt-R® Cas9 Electroporation Enhancer. 3 µL (for 10 µL electroporation tip) or 30 µL (100 µL tip) of the mixture was added to a sterile tube ready for electroporation.

2.9.2. Electroporation of cells with RNPs

Cells were lifted by washing with PBS and adding 1 mL Accutase per well of a 6 well plate or 3 mL per 10 cm dish. After 4-5 minutes of incubation, the cells were washed through a 40 µm cell

strainer and diluted 1:5 with 10% ES FBS in KO DMEM to inactivate the enzyme. They were then spun at 400 x *g* for 5 minutes and the supernatant was subsequently discarded. Cells were resuspended in 1 or 2 mL PBS, counted (diluted 1:10 for counting), 1 mL transferred into an Eppendorf and spun. After discarding the supernatant, cells were resuspended to 2.2 x 10^7 cells/mL in buffer R.

For the time course experiment 5 tubes of 30 µL RNP & template were prepared as above, 90 µL primed cells added to each and electroporated using the 100 µL neon electroporation tip using the high trans protocol (1400V, Pulse width 20, Pulse number 1). After electroporation, cells were immediately pooled into 19.5 mL mTeSR + ROCKi. (The no template control was performed using just one tube and transferred to 8 mL mTeSR + ROCKi.) For early time points (15 min, 30 min, 1h, 2h, 4h, 6h, 12h, 24h), 2 mL (1 x 10^6) electroporated cells were plated onto a well each of a 6 well plate precoated with geltrex (1.1 x 10^5 cells/cm^2). For time points 48h and 96h, cells were diluted 1:2 before plating in 6 well format on geltrex precoated wells. The wild type and no template controls were plated at 5 x 10^5 cells per well onto geltrex precoated wells (5.5 x 10^4 cells/cm^2). The 15 min time point was not plated but instead placed in a tube ready for pelleting. At each other time point, cells were lifted with accutase as previously explained, pelleted by spinning for 1 min at 1000 x *g* and frozen at -80°C. For early timepoints up to 24h the supernatant was included in the tube, so any floating cells were included in the pellet.

To measure differences in repair pathways between naïve and primed cells, 9 µL cells were added to the 3 µL cas9 RNP + template prepared above and electroporated in a 10 µL electroporation tip. Naïve cells were electroporated with the programme A3 (1150V, Pulse width 30, Pulse number 2), primed with high trans (1400V, Pulse width 20, Pulse number 1). Cells were immediately transferred into 12 well plates (either pre-treated with geltrex for primed cells or containing MEFs for naïve) containing the appropriate medium with ROCKi. After 4 days, cells were pelleted and DNA extracted using the same method implemented for karyotyping discussed above.

2.9.3. Droplet digital PCR

Droplet digital PCR was used to measure if DNA was repaired with an indel formation or using a template strand.

For each locus, a 20x mix of primers and probes was prepared. 100 µL of the mix was prepared by combining 18 µL each of 100 µM forward and reverse primers (in ddH_2O). Probes were resuspended to 100 µM in ddH_2O and 5 µL each of HDR, NHEJ and reference probes were added. For the RBM20 and GRN assay, 10 µL of a dark probe was also included. A dark probe is an oligo with 3' phosphate (to avoid priming) which competes with the HDR probe to prevent false positives. To complete the 20x mix the volume made up to 100 µL with ddH_2O. (Table 2.2)

A master mix containing all but the gDNA was prepared for each locus: Per sample, 12.5 µL supermix for probes (Bio-Rad) was added to 1.25 µL 20x primer/probe mix (prepared above) and 1 µL restriction Enzyme. The restriction enzymes were *Hind* III-HF (R3104; NEB) for the loci ATP7B and RBM20 and *MseI* I (R0525; NEB) for GRN. 100 ng of gDNA was added to each reaction by diluting the genomic DNA to 9.76 ng/µL and adding 10.25 µL to 14.75 µL master mix. Each 25 µL reaction was well mixed by pipetting and 20 µL was transferred to the sample well of a DG8™ cartridge (1864007; Bio-Rad) taking care not to introduce bubbles. Each cartridge was processed with all 8 samples wells used. 70 µL droplet generation oil for probes (1863005; Bio-Rad) was added to each oil well, the gasket (1864007; Bio-Rad) attached, and the droplets generated in a QX200 Droplet Generator (1864002; Bio-Rad).

Droplets were transferred to a PCR plate using a multichannel pipette. The plate was sealed using non-adhesive, pierceable foil (1814040; Bio-Rad) and the PX1™ plate sealer (1814000; Bio-Rad). The plate was then placed in a C1000 Touch™ Thermal Cycler (1851197; Bio-Rad) and the following PCR programmes were used with all steps ramped by 2 °C/s: For GRN and ATP7B: Step 1, 95 °C for 10 min; step 2, 94 °C for 30 s; step 3, 58 °C for 1 min; step 4, 72 °C for 2 min; repeat

steps 2, 3, and 4 39 times; step 5, 98 °C for 10 min. For RBM20: step 1, 95 °C for 10 min; step 2, 94 °C for 30 s; step 3, 59 °C for 1 min; repeat steps 2 and 3 39 times; step 4, 98 °C for 10 min.

The plate was transferred to a QX200™ Droplet Reader (1864003; Bio-Rad) to count the droplets and the data was analysed with the QuantaSoft Software. Gating was performed by using the wild type and no-template controls. Percentages of NHEJ or HDR were calculated by dividing by all positive droplets.

2.10. *DGCR8* knockout

2.10.1. Generating the knockout

To generate a knockout of *DGCR8* the Alt-R® CRISPR-Cas9 system was used and guides against DGCR8 (Table 2.5) were electroporated as RNPs as described above in 10 µL format into both naïve and primed SFC856-03-04 cells. After 4 days, naïve and primed cells were split to single cells and plated onto MEFs in 3 wells at low cell numbers (222, 444 and 889 cells/cm^2) to allow individual colonies to be picked after a further 6 days in culture. Colonies were then picked by scraping with a P200 pipette tip, taking the colony up and ejecting it into a 96 well containing MEFs and the appropriate medium. 96 well plates were split after 3-5 days by aspirating the medium, washing with 200 µL PBS and adding 20 µL Accutase to lift the cells. After 6 minutes incubation at 37°C, 200 µL medium was added and the cells dissociated by pipetting. The accutase was then diluted out further by adding an extra 100 µL of medium to each well and then transferring 100 µL of the cell suspension to three new plates each containing 100 µL medium and MEFs. 3-5 days after the split, 1 plate was frozen for DNA extraction by washing with PBS and freezing the plates dry by sealing them with adhesive foil (E2796-9792; StarLab) and storing them at -80°C. The remaining two plates were frozen for further expansion by aspirating the medium, washing with 200 µL PBS and adding 20 µL Accutase to lift the cells. After 6 minutes, cells were washed off by pipetting up and down in 100 µL 1X freeze medium. The plate was then sealed with

adhesive foil, placed in a Styrofoam box, and stored overnight at -80°C. The following day the plate was transferred to -196°C nitrogen vapour storage.

2.10.2. On plate DNA extraction

To extract DNA, the frozen plates were removed from -80°C and left for 5-10 minutes on the bench to allow them to thaw. 50 µL lysis buffer was then added to each well. Lysis buffer was prepared by making the following solution: 10 mM tris pH7.5 (15567027; ThermoFisher Scientific), 10 mM EDTA (15575-020; ThermoFisher Scientific), 10 mM NaCl (S7653; Sigma), 0.25% Triton X100 (X100; Sigma). The solution was filter sterilised by passing through a 0.22 µm filter (SCGPU02RE; Merck) and stored at 4°C. Just before use, proteinase K (50 mg/mL stock; BP1700-100; ThermoFisher Scientific) was added to 1 mg/mL. Each plate was then incubated overnight at 55°C in a sealed plastic bag containing dampened paper towels (humidified). The following day the DNA was precipitated out of the lysed cell solution by adding 100 µL ice cold (-80°C) precipitation solution per well. The precipitation solution was made by adding 15 µL of 5M NaCl (S7653; Sigma) per mL of 100% Ethanol and storing at -80°C. Each plate was then incubated at room temperature for 2 hours and the DNA became visible as a stringy substance in the bottom of the wells. Plates were inverted to discard the liquid and washed 3 times with 100 µL 70% Ethanol. After the last Ethanol wash, each plate was inverted and thoroughly blotted on paper towels. Plates were spun at 800 x *g* for 1 minute to ensure the DNA was at the bottom of the well. The plates were then kept at room temperature and airdried for several hours until the wells were dry. The DNA was resuspended by adding 100 µL ddH_2O and incubating at 4°C overnight to ensure that the DNA had fully gone into solution.

2.10.3. Screening by PCR

The plate of extracted gDNA was screened by PCR. The amount of genomic DNA in each well was measured by Nanodrop and diluted to 5 ng/µL. Per 20 µL reaction, the following 2X master mix was prepared (with an extra 10% overage): 4 µL 5X Phusion HF buffer (B0518S; NEB), 0.4 µL 10mM

dNTPs (18427013; ThermoFisher Scientific), 1 µL 10mM forward primer BD131 (Table 2.2; Sigma), 1 µL 10mM reverse primer BD132 (Table 2.2; Sigma), 0.2 µL HS Phusion (M0535S; NEB) and 3.4 µL ddH_2O. 10 µL of 5 ng/µL gDNA was then added to 10 µL master mix and incubated on a thermal cycler with the following programme: Step 1, 98°C for 30s; step 2, 98°C for 7s; step 3, 69°C for 20s; step 4, 72°C for 70s; repeat steps 2, 3, and 4 29 times; step 5, 72°C for 10 min; step 6, 12°C hold.

4 µL of 6X purple loading dye (B7024; NEB) was added to each well and 15 µL ran on a 1% agarose gel at 60V for approximately 1 hour. The loading buffer contains Ficoll to increase the density of the sample, SDS to denature DNA binding enzymes as well as two dyes which run at two different speeds and allow estimation of how far the DNA has migrated. The agarose gel was prepared by adding 1% w/v agarose (AGD1-500G/A4-0700; Geneflow Ltd) to buffer TAE and heating in a microwave until the agarose was fully dissolved. 50X TAE was prepared by mixing 242g Tris-base (000000010708976001; Sigma), with 57.1 mL 100% acetic acid (33209-M; Sigma), 100 mL 0.5M EDTA (15575-020; ThermoFisher Scientific) and making up the volume to one litre with dH_2O. 1X TAE was made by adding 20 mL 50X TAE to 980 dH_2O. The agarose was then cooled to 50°C and 10 µL SYBR Safe DNA Gel Stain (S33102; ThermoFisher Scientific) per 100 mL agarose was added. It was then poured into a casting tray containing a well comb and allowed to set for approximately 30 minutes. The well comb was removed, the gel was placed into a gel electrophoreses unit (1704502; Bio-Rad) containing buffer TAE and loaded with hyperladder 1kb (BIO-33053; Bioline) and samples.

2.10.4. Cloning of lentiviral, dox inducible *DGCR8*

To generate a lentiviral, dox inducible *TurboRFP-2A-DGCR8* an existing plasmid containing the tet-inducible promoter, *TurboRFP* and lentiviral backbone was used as a starting vector. The original plasmid was digested using *Bsu36* I (R0524S; NEB) and *Mlu* I (R0198S; NEB), creating non-compatible sticky ends. The digestion was performed by adding 1 µg original vector in 1.56 µL to 5 µL 10X buffer 3.1 (B7203S; NEB), 1 µL *Bsu36* I, 1 µL *Mlu* I and 41.44 µL ddH_2O. The tube was

mixed and incubated at 37°C in a thermal cycler for 15 minutes. 10 µL 6X loading dye was added and the whole reaction mix was run across several wells on a 1% agarose gel as described above for 2-3 hours. The double cut plasmid was then cut out of the agarose and purified.

The purification was performed using the Monarch DNA Gel extraction kit (T1020S; NEB): the cut-out agarose piece was weighed and 4 volumes of gel dissolving buffer added. The sample was incubated at 50°C in a heat block (dri-block db100/2; Techne) for 5-10 minutes while regularly vortexing until the gel slice was completely dissolved. Then an extra 1.5 volumes of ddH_2O was added since the empty vector was a large fragment which binds tighter to the spin column. The sample was then added to the spin column, spun for 1 minute at 16,000 x *g* and the flow-through was discarded. 200 µL wash buffer was added and the column spun again for 1 minute and the flow-through was discarded. The spin column was added to a fresh tube. To improve the elution of the large vector, the elution buffer was warmed to 50°C and 10 µL added to the spin column. After 1 minute incubation, the column was spun for 1 minute to elute the DNA.

A double stranded DNA fragment (gBlock; Supplementary Figure S1) was ordered and synthesised by Integrated DNA technologies. It contained cut sites for *Mlu* I and *Bsu36* I on either side together with 10-15bp extra overhang, an *RFP* fragment, a *2A* sequence and the *DGCR8* cDNA. The *2A* sequence is translated to a combination of peptides which cause a break in the polypeptide during translation leading to two products from the same reading frame (Kim et al., 2011). The gBlock was resuspended to 10 ng/µL by adding 100 µL ddH_2O to 1 µg DNA. The tube was then vortexed and incubated at 50°C for 20 minutes followed by more vortexing to ensure the gBlock had gone into solution. The concentration was confirmed by nanodrop and the gBlock was digested with *Bsu36* I and *Mlu* I as described above but adding 43 µL (430 ng) of DNA and no ddH_2O. After the 15-minute incubation at 37°C, the digested gBlock was purified using a PCR purification kit (28104; Qiagen): 5 volumes of Buffer PB were added, mixed and the mixture added onto a spin column. After spinning for 30s at 17,900 x *g* the flow-through was discarded and the column was washed

by adding 750 µL buffer PE and spinning again as before. The flow-through was discarded and the column dried by spinning for an additional 1 minute. The column was then placed into a clean tube, 30 µL elution buffer was added to the centre of the column membrane and was left for 1 minute before spinning for an additional minute to elute the purified DNA. The concentration was measured by nanodrop.

Ligation of the complementary DNA fragments was performed by adding 2 µL 10X ligase buffer (B0202S; NEB) to 1.16 µL vector (at 43 ng/µL), 3.68 µL gBlock (9.5 ng/µL) and 12.16 µL ddH_2O. Finally, 1 µL T4 Ligase was added (M0202S; NEB), the tube was then mixed and incubated at room temperature for 10 minutes and transferred onto ice for transformation. The no-insert control was performed by replacing the insert with ddH_2O.

Transformation was performed using competent Stabl3 *Escherichia coli* (*E.Coli*; C737303; ThermoFisher Scientific): Bacteria were thawed on ice, 2 µL ligation mix added and gently mixed. The vials were then incubated on ice for 30 minutes and heat-shocked for 45 seconds in a waterbath at 42°C without shaking. The vials were then placed on ice for 2 minutes and 250 µL room temperature S.O.C. Medium was added. The vials were then shaken horizontally in a shaking incubator (Eppendorf New Brunswick™ Innova 44/44R) at 37°C for 1 hour at 180 rpm. 25 µL and 100 µL bacteria were added to 100 µg/mL ampicillin (A0166-5G; Sigma) bacterial agar plates (prepared in house). The plates were inverted and incubated overnight at 37°C and colonies were picked the following day.

A starter culture was prepared by adding a colony to 3 mL 100 µg/mL Ampicillin Lysogeny broth (LB) and incubating in a shaking incubator for 7h. Then 80 µL of the starter culture was added to 50 mL 100 µg/mL ampicillin LB and incubated overnight in a shaking incubator at 37°C. The following morning a midiprep was performed using the PureLink™ HiPure Plasmid Filter Midiprep Kit (K210014; ThermoFisher Scientific): 15 mL EQ1 buffer was added to the filtration cartridge for equilibration. Each overnight culture was added to a 50 mL disposable tube and spun at 4000 x *g*

for 10 minutes (in an Eppendorf 5804 R centrifuge). Medium was removed, and the pellet resuspended by adding 10 mL buffer R3 (with RNase A). Once the suspension was homogenous, 10 mL lysis buffer was added and gently mixed by inversion. After 5-minute incubation at room temperature 10 mL precipitation buffer was added and the tube was mixed by inversion. The lysate was then transferred into the column and allowed to drain. 10 mL wash buffer was then added and allowed to drain. The inner filtration cartridge was then removed, and 20 mL wash buffer added to the column. After allowing it to drain, a sterile 15 mL tube was placed under the column and the plasmid eluted by adding 5 mL elution buffer to the column. The DNA was then precipitated out of solution by adding 3.5 mL isopropanol to the eluate, incubating for 2 minutes at room temperature and then spinning the tube at 4000 x *g* at 4°C for 1h. The supernatant was then removed and 3 mL 70% ethanol was added to wash the pellet. The tube was then centrifuged for 10 minutes at 4°C and the supernatant removed. The pellet was air dried until no liquid was visible and resuspended in 100 µL ddH_2O. The concentration was measured by nanodrop and the DNA was stored at -20°C.

2.11. Statistical analysis

Statistical analysis was performed using GraphPad Prism 7. Error bars indicate the standard deviation and statistical tests performed are detailed in the figures. Unless otherwise stated, missing error bars indicate that the error bars are smaller than the symbol used for the mean. Statistical significance was defined as ns = $P > 0.05$; * = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$; **** = $P < 0.0001$.

Product	Supplier	Catalogue number
0.22 µm filter	Merck	SCGPU02RE
100mm TC dish	Corning	430167
10mM dNTPs	ThermoFisher Scientific	18427013
10x ligase buffer	NEB	B0202S
12 well TC plates	Greiner	665180
24 well TC plate	Greiner	662160
2-ME	ThermoFisher Scientific	31350010
40 µm filter	BD Biosciences	352235
5X Phusion HF buffer	NEB	B0518S
6 well TC plates	Greiner	657160
6X purple loading dye	NEB	B7024
96w plate	Greiner	655180
Accumax	Sigma	A7089
Accutase Cell Dissociation Reagent	ThermoFisher Scientific	A1110501
Acetic acid	Sigma	33209-M
Activin A	Peprotech	120-14
Agarose Hi Res Standard	Geneflow	AGD1-500G/A4-0700
AlbuMAX1	ThermoFisher Scientific	11020021
Aluminium StarSeal (PCR), Self-Adhesive	StarLab	E2796-9792
Ampicillin	Sigma	A0166-5G
apo-transferrin	eBioscience	ABC2553
Ascorbic acid	Sigma	A5960
B27 Supplement (50X)	Life Tech	17504044
Bacterial agar plates	Dunn School Services	Amp100
bFGF (500ul aliquot 1000X)(rh-FGF)	R&D Systems	4114-TC-01M
BIO	Sigma	B1686
BIRB796	Axon Medchem	1358
BSA	Sigma	A7979
BSA Fraction V 7.5% solution	ThermoFisher Scientific	15260037
C1000 Touch™ Thermal Cycler	Bio-Rad	1851197
Cell counting NC Slides A8	Chemometec	942-0003
Cell Strainers 40um	SLS Ltd	352340
CF1 mef	Millipore	PMEF-CFL
CHIR99021	Axon Medchem	1386
cryo-bombs	Sigma	C1562-1EA
Cryovials	Starlab	E3090-6222
Cryovials (internal thread)	Greiner	122263
Cytotune 2 Sendi reprogramming (12/pack)	Invitrogen (Life Technologies)	A16517
Defined Lipid Concentrate	ThermoFisher Scientific	11905031
DG8 Cartridge & gaskets	Bio-Rad	1864007
DMEM/F12	ThermoFisher Scientific	11320
DMSO	Sigma	D2650-100ML
Dnase	Qiagen	79254
Dneasy Blood and tissue kit	Qiagen	69506

Table continued on next page

Product	Supplier	Catalogue number
dNTPs 10mn SET OF 4	NEB	N0446S
dNTPs mix	NEB	N0447S
Dorsomorphin	Sigma	P5499
droplet generation oil for probes	Bio-Rad	1863005
EDTA 0.5M	ThermoFisher Scientific	15575-020
ES- FBS	ThermoFisher Scientific	16141-079
Essential 8	ThermoFisher Scientific	A15169-01
Ethanol	Sigma	32221-2.5L-M
FBS	Sigma	F4135-500ML
Fetal Bovine Serum	ThermoFisher Scientific	10500064
Gel Extraction Kit (50)	Qiagen	28704
Gelatin 2%	Sigma	G1393-100ML
Geltrex	ThermoFisher Scientific	A1413302
Glutamax	ThermoFisher Scientific	35050038
Gö6983	TOCRIS	2285
High Capacity RNA to cDNA kit	ThermoFisher Scientific	4387406
HindIII-HF	NEB	r3104
HiPure Plasmid Filter Midiprep Kit-25	ThermoFisher Scientific	K210014
HS Phusion	NEB	M0535S
Human IgG	Sigma	I2511
IM-12	Sigma	SML0084
Insulin Solution, Human recombinant	Sigma-Aldrich	I9278
Iso-propanol	Sigma	24137-2.5L-R
IWR1	Sigma	I0161
KO DMEM	ThermoFisher Scientific	10829-018
KO SR	ThermoFisher Scientific	10828-028
Ladder 1 KB	Bioline	BIO-33053
L-ascorbic acid 2-phosphate	Sigma	A8960
LB broth	Dunn School Services	LB
L-Glutamine	Sigma	G6392
LIF	Peprotech	300-05
MEM Non Essential amino acids	ThermoFisher Scientific	11140-035
MicroAmp™ Optical Adhesive Film	ThermoFisher Scientific	4311971
Midi prep kit	ThermoFisher Scientific	K210014
miRNeasy mini kit	Qiagen	217004
MseII	NEB	R0525
mTeSR	Stemcell Technologies	5850
N2 Supplement 100X	ThermoFisher Scientific	17502048
NaCl	Sigma	S7653
NEAA	ThermoFisher Scientific	11140035
Neon Transfection System 100 Ml Kit	ThermoFisher Scientific	MPK10025
Neon Transfection System tubes	ThermoFisher Scientific	MPT100
Neon™ Transfection System 10 µL Kit	ThermoFisher Scientific	MPK1025
Neon™ Transfection System 100 µL Kit	ThermoFisher Scientific	MPK10025

Table continued on next page

Product	Supplier	Catalogue number
Neurobasal	ThermoFisher Scientific	21103
non-adhesive, pierceable foil	Bio-Rad	1814040
Paraformaldehyde 4%	Alfa Aesar	J61899
PBS	Lonza	BE17-516F24
PCR Cleanup Kit	Qiagen	28106
PCR purification kit	Qiagen	28104
PCR tubes 8 strip +attached lid	Greiner	608281
PD0325901	Axon Medchem	1408
PD0325901	Axon Medchem	1408
PI staining solution	BD Biosciences	556463
progesterone	Sigma	P6149
proteinase K	ThermoFisher Scientific	BP1700-100
putrescine	Sigma	P6024
qRT PCR MicroAmp 96 well plate	ThermoFisher Scientific	4346906
Rnase A 100mg/ml	Qiagen	19101
Rnase-free Dnase set (50)	Qiagen	79254
RNEasy micro kit	Qiagen	74004
Rneasy mini kit	Qiagen	74104
RSeT medium	Stemcell Technologies	5970
SB590885	TOCRIS	2650
Sodium Azide	Sigma	769320
sodium pyruvate	ThermoFisher Scientific	11360070
Sodium Selenite	Sigma	S5261
Solution 13	Chemometec	910-3013
SP600125	TOCRIS	1496
Sterile reservoirs	Star lab	E2310-1010
Sybr safe DNA gel stain	ThermoFisher Scientific	S33102
T4 Ligase	NEB	M0202S
TaqMan advanced miRNA assay	ThermoFisher Scientific	A25576
TaqMan advanced miRNA assay cDNA synthesis kit	ThermoFisher Scientific	A28007
Taqman Copy number assay	ThermoFisher Scientific	4400291
TaqMan Fast Advanced Master Mix	ThermoFisher Scientific	4444557
TaqMan Gene Expression Assays	ThermoFisher Scientific	4331182
TaqMan Gene Expression Master mix	ThermoFisher Scientific	4369016
TaqMan™ Genotyping Master Mix	ThermoFisher Scientific	4371355
tris pH7.5	ThermoFisher Scientific	15567027
Tris-base	Sigma	10708976001
Triton X100	Sigma	X100
TrypLE Express	ThermoFisher Scientific	12604013
Water, distilled	ThermoFisher Scientific	15230089
WH-4-023	Axon Medchem	2381
Y-27632 Rho kinase inhibitor	ABCam Plc	ab 120129-10mg

Table 2.6: List of reagents and suppliers.

3

Generating and optimising the culture of naïve human pluripotent stem cells

3.1. Introduction and aims

To date, 9 protocols suggest methods of reversing primed cells to the naïve state (Chan et al., 2013, Duggal et al., 2015, Gafni et al., 2013, Guo et al., 2016, Theunissen et al., 2014, Takashima et al., 2014, Valamehr et al., 2014, Ware et al., 2014, Yang et al., 2017b). Each use a different media formulation (Table 1.1) and a variety of cell lines are used across different labs. A direct and objective comparison of available protocols is required both functionally and in assessing available published evidence. This will allow identification of the most robustly and universally functional protocols, eliminating media formulations which only work with certain cell lines or only in certain labs. Available RNAseq data can be benchmarked against *in vivo* single cell RNAseq data from human embryogenesis to find which naïve cells most accurately resemble the inner cell mass of an E5 blastocyst.

The aims for this chapter are:

- Objective review of reported evidence (3.2.1)
- Side by side comparison of available naïve conversion protocols (3.2.2 & 3.2.3)
- Generation of 3 karyotypically healthy naïve lines using the most convincing protocol (3.2.4)
- Adapting flow cytometry and electroporation methods to naïve cells (3.2.5 & 3.2.6)

3.2. Results

3.2.1. Reviewing published RNAseq data of naïve conversion protocols

The most stringent, and least artificial evidence of naïve pluripotency is a comparison of gene expression to published *in vivo* RNAseq data. To assess gene expression of the naïve conversion protocols I collated the relevant RNAseq datasets to compare to two *in vivo* single cell RNAseq datasets of human embryogenesis E-MTAB-3929 (Petropoulos et al., 2016) and GSE36552 (Yan et al., 2013). Dr Klas Hatje normalised, filtered (for variable genes), log-transformed, batch corrected and scaled all datasets to compare these protocols and perform a principal component analysis (PCA; Supplementary Methods 7.1.1). The PCA analysis was chosen over other dimensionality reduction techniques such as t-SNE, since global trends are more accurately represented and distances on the plot correlate with how different individual samples are. This method is therefore most suitable for the comparison of samples and embryonic stages.

PCA can be used to reduce high dimensional data to fewer dimensions by feature extraction (and subsequent feature elimination). This functions by finding which variables are associated with one another in a covariance matrix. The PCA accomplishes this by fitting straight lines to the data points. There are as many principal components as there are dimensions, however the most variance is explained by the first. The value of the first principal component (PC1) is where each datapoint sits along the best fitting straight line. The second principal component (PC2) is orthogonal to the first (and independent). The third is orthogonal to both first and second and so on.

In Figure 3.1A the PCA effectively separates the different embryonic stages. E0, E1 and E2 cells (all from GSE36552) cluster together, and starting next to them, E3-E7 cells form a continuous transition, with slight overlap but correct temporal order, ending in the latest developmental stage of the dataset: E7. The two datasets do not overlap perfectly, in that E5 cells from GSE36552 have similar PC2 but are shifted to the left on PC1. hESC derived as part of this dataset cluster

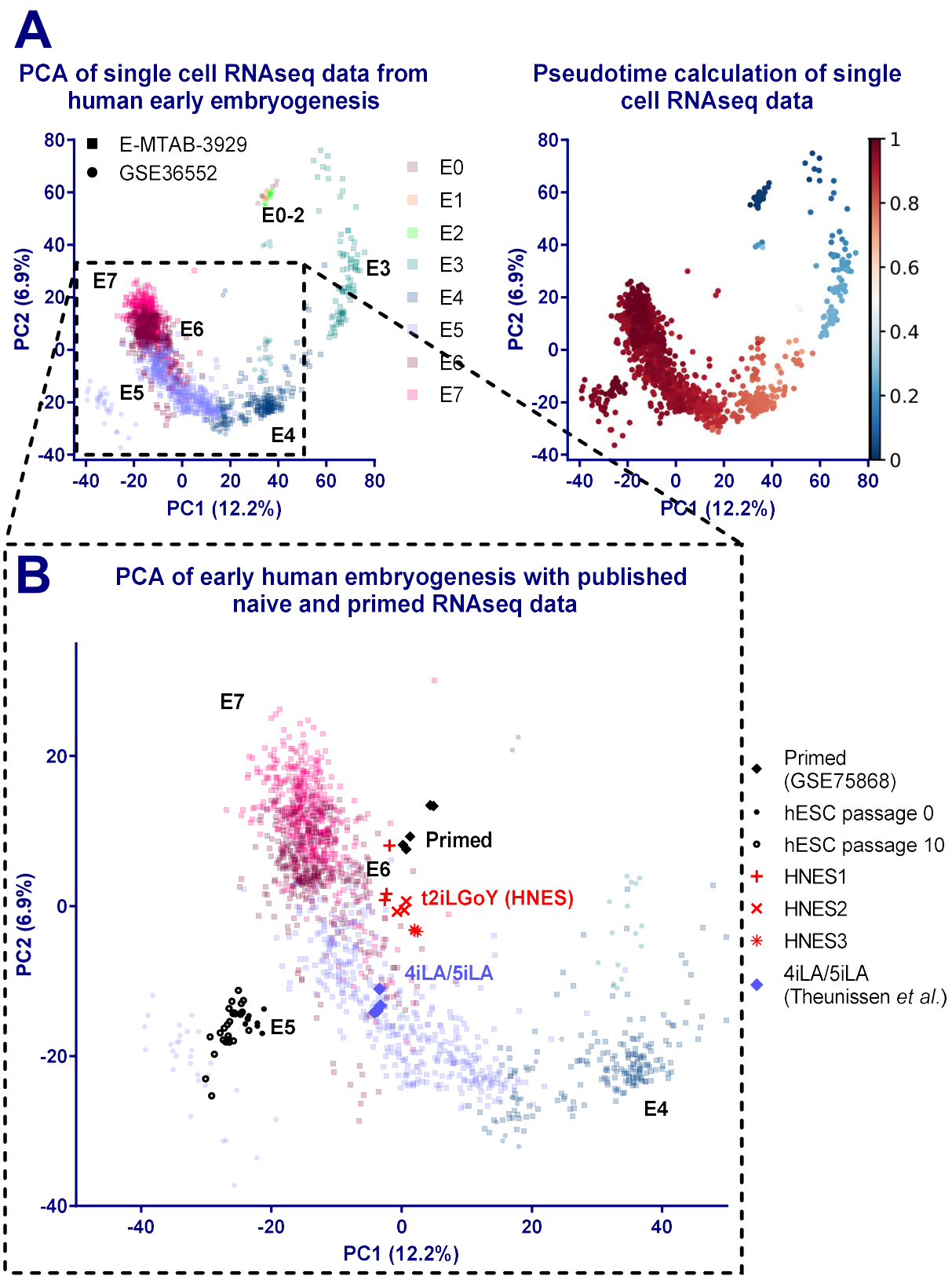


Figure 3.1 continued on next page

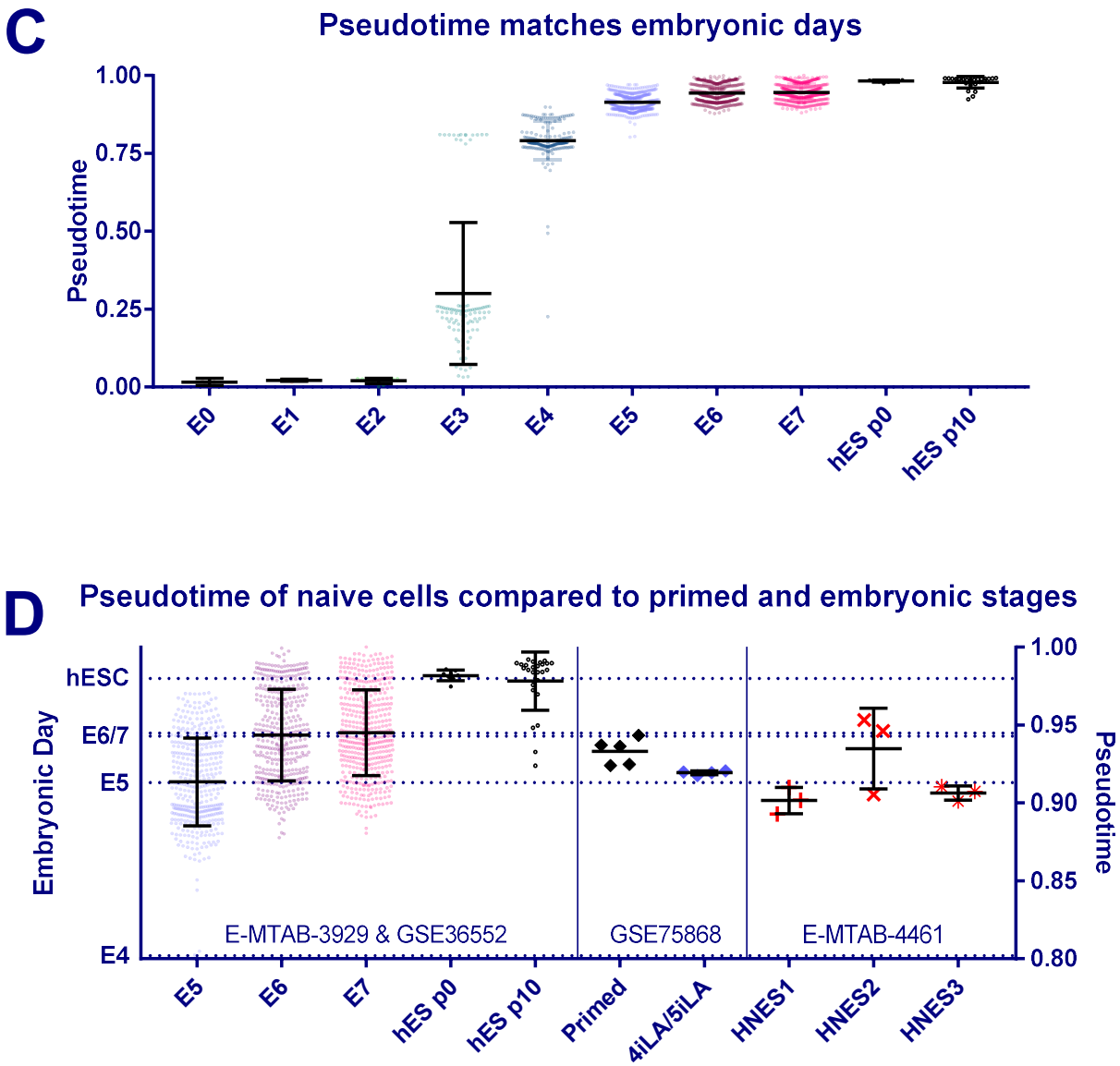


Figure 3.1: Comparison of RNAseq data from published naïve protocols to *in vivo* embryogenesis single cell RNAseq datasets. All datasets were normalised and processed by Dr Klas Hatje. Normalisation was performed first by each sample's library size followed by a filtering step to include variable genes used for PCA and pseudotime analysis. A scaling step was incorporated ensuring equal scatter of counts around the mean in each sample. These normalisations were performed in all samples independently. In order to compare datasets, an unbiased method to remove batch effects was used (Supplementary Method 7.1). (A) The two *in vivo* single cell RNAseq datasets (E-MTAB-3929, GSE36552) were then merged and a principle component analysis performed. The E0 oocyte was then designated as the starting point and an unbiased pseudotime analysis applied. The two datasets largely overlap, with the exception of cells at E5 from GSE36552 clustering below the E-MTAB-3929 dataset. (B) To evaluate where naïve and primed hPSC from key protocols align during early human development, published bulk RNAseq datasets were compared and plotted. The single cell GSE36552 dataset includes freshly derived hESC lines at p0 and p10 which cluster with the E5 cells below the E-MTAB-3929 dataset. Naïve cells converted by Theunissen *et al.* cluster with E5 cells, whereas primed control cells are closer to E6/E7 cells. The HNES cell lines derived by Guo *et al.* cluster in between. (C) The pseudotime matches embryonic days and can be used as a single comparative metric. The error bars represent the standard deviation. (D) The pseudotimes of developmental stages E5, E6 and E7 are compared to naïve and primed cells. Protocols not in the focus of this thesis are shown in Figure S2. Both 4iLA/5iLA and HNES cells resemble an earlier time point than primed cells. The pseudotime analysis alone might underestimate the difference between 4iLA/5iLA and HNES cells, since a clear difference between them is visible in the PCA plot.

with these cells and not with other published primed datasets, showing the limitations of comparing separate datasets. The main caveat of this analysis is the difficulty of comparing bulk and single cell datasets. This requires batch correction which may introduce bias or may not sufficiently correct the technical differences between datasets. Variability between datasets is also an issue, illustrated by the hESC from the single cell dataset not clustering with conventional primed cells and the two embryogenesis datasets not entirely overlapping.

The protocols used in this thesis were plotted onto the PCA plot containing the *in vivo* embryogenesis data (Figure 3.1B). Cells generated by Theunissen *et al.* cluster perfectly with mid E5 cells - the *in vivo* equivalent of naïve PSC. Primed cells are closer to E6/E7 and HNES cells used in this thesis are in between. All other protocols with published RNAseq datasets were plotted in Supplementary Figure S2A. Only Takashima *et al.* and Theunissen *et al.* occupy the mid E5 space, suggesting these protocols might be the closest to *in vivo* naïve cells.

Pseudotime analysis uses single cell RNAseq data of a process or differentiation and orders each cell/sample according to the nearest neighbour, starting from a defined starting cell/sample. This alignment is an *in-silico* approximation of the order of gene expression changes during the process. The pseudotime value represents the relative order of each sample and is thought to represent temporal separation during the process. (Wolf et al., 2018)

To perform a pseudotime analysis, Dr Klas Hatje used the PCA data to calculate a neighbourhood graph. An oocyte from dataset GSE36552 was chosen as the starting point, the trajectory was calculated based on the shortest distance and a pseudotime value between 0 and 1 is reported for each sample (Figure 3.1A, C; Supplementary Methods 7.1.1). Some information is lost when focusing only on the pseudotime since only one dimension is plotted. Most naïve conversion cells have an earlier pseudotime than their primed counterparts (Figure 3.1D, Supplementary Figure S2B), but this analysis is unable to resolve the differences between protocols as clearly as seen in the PCA analysis.

Analysis of the individual gene weightings of the PCA is performed in Table 3.1. The gene expression data fed into the PCA analysis is scaled to 0 mean. This is achieved by subtracting the values from the mean (Supplementary Methods 7.1). Higher expression therefore leads to more negative values. So (counterintuitively), high expression of "positively" weighted genes therefore leads to a lower PC. The individual gene weightings revealed that primed marker CD24 contributes "negatively" to both PC1 and PC2. Naïve marker KLF4 contributes "positively" to both PC1 and PC2, whereas naïve markers DNMT3L and TBX3 both contribute "positively" to PC2. This is corroborated by what is observed in the PCA analysis. Naïve hPSC have a lower PC2 and slightly lower PC1 compared to their primed counterparts.

Several naïve conversion protocols were chosen for testing based on the following criteria: Any protocol requiring no direct derivation from blastocysts or reprogramming from fibroblasts. Protocols requiring the use of transgenes were also omitted, since vectors may integrate and cause artefacts. The Yang *et al.* protocol was published after the experiment in 2017 and was therefore not included in the comparison (Table 1.2).

3.2.2. Comparison of nine naïve protocols

To find a suitable protocol to revert iPSC to the naïve state, initially 7 different protocols were tested for their ability to generate naïve cells. The tested protocols were: Chan *et al.* (Chan et al., 2013), Duggal *et al.* (Duggal et al., 2015), RSeT (commercially available from Stem Cell Technologies, based on NHSM 1A), NHSM 1A (Gafni et al., 2013), NHSM 1B, ENHSM and Theunissen *et al.* (Theunissen et al., 2014). NHSM 1B and ENHSM are unpublished media formulations based on NHSM 1A and were developed by the same lab and shared with us via their website. A comparison of components is listed in Table 1.1. Theunissen *et al.* in its original form includes IM-12 which was subsequently shown to be unnecessary (Sahakyan et al., 2017, Theunissen et al., 2016). The media containing IM-12 is referred to as 5iLA and without 4iLA. An "F" such as in 5iLAF indicates addition of the optional bFGF.

A Contribution of genes of interest to principal components PC1 and PC2

Gene	also known as	PC1	PC2	Marker for
CD24		**-0.0438**	**-0.0221**	primed
DNMT3L		-0.0053	**0.0360**	naïve
KLF4		**0.0334**	**0.0286**	naïve
TBX3		-0.0116	**0.0325**	naïve
POU5F1	*OCT4*	**0.0275**	**0.0319**	shared
IL6ST	*CD130*	-0.0033	0.0189	naïve
NANOG		0.0075	0.0122	shared
OTX2		0.0118	-0.0023	primed
DICER1		-0.0086	0.0108	
THY1	*CD90*	0.0015	0.0030	primed

B Top 10 positively weighted genes in PC1

Gene	PCs1	PCs2	Function
FAM151A	0.0470	0.0071	protein coding gene with unknown function
ATG3	0.0462	0.0056	ubiquitin-like-conjugating enzyme, esponsible for the E2-like covalent binding of phosphatidylethanolamine to the C-terminal Gly of ATG8-like proteins (including *GABARAPL1* also in top 10)
BAK1	0.0460	-0.0025	apoptosis inducing BCL2 family protein
PSORS1C2	0.0451	-0.0069	secreted protein
SNHG15	0.0444	0.0024	snoRNA, miRNA sponge for miR-211-3p (no detectable expression in naïve/primed hPSC)
SOAT2	0.0436	-0.0142	acyl coenzyme A:cholesterol acyltransferase
TRIM61	0.0427	-0.0092	Protein Coding gene associated with Premature Ovarian Failure 1.
CCPG1	0.0424	-0.0109	Acts as an assembly platform for Rho protein signaling complexes. May be involved in cell cycle
GABARAPL1	0.0423	-0.0131	Ubiquitin-like modifier that increases cell-surface expression of kappa-type opioid receptor through facilitating anterograde intracellular trafficking of the receptor. Involved in formation of autophagosomal vacuoles.
ETNPPL	0.0422	-0.0125	Catalyzes the pyridoxal-phosphate-dependent breakdown of phosphoethanolamine

C Top 10 negatively weighted genes in PC1

Gene	PCs1	PCs2	Function
GATA2	-0.0449	-0.0003	Transcription factor involved in pluripotent stem cells
SLC7A8	-0.0448	0.0005	amino acid transport
KRT19	-0.0447	-0.0228	keratin family/structural
CD24	-0.0438	-0.0221	sialoglycoprotein, CD24 used as a marker in flow cytometry experiments
PTN	-0.0435	-0.0321	Secreted growth factor, Binds anaplastic lymphoma kinase (ALK) which induces MAPK pathway activation
ABCG2	-0.0431	-0.0151	ABC proteins transport, expressed in placenta, involved in drug resistance
HINT1	-0.0427	-0.0153	hydrolyzes purine nucleotide phosphoramidates substrates, tumour supressor
SLC38A1	-0.0427	-0.0304	sodium-dependent amino acid transporter, Glutamine
DLX3	-0.0422	-0.0204	Homeobox transcription factor
TPM4	-0.0414	-0.0026	member of the tropomyosin family of actin-binding proteins

D Top 10 positively weighted genes in PC2

Gene	PCs1	PCs2	Function
ANPEP	0.0027	0.0533	Broad specificity aminopeptidase
MX1	0.0175	0.0524	This gene encodes a guanosine triphosphate (GTP)-metabolizing protein that participates in the cellular antiviral response.
CD81	-0.0095	0.0510	This encoded protein is a cell surface glycoprotein that is known to complex with integrins. This protein appears to promote muscle cell fusion and support myotube maintenance. Also it may be involved in signal transduction.
RAB15	0.0189	0.0510	GTP binding and GTPase activity
TUBB4A	0.0144	0.0508	This gene encodes a member of the beta tubulin family. Microtubules/structural
NFKB2	0.0094	0.0503	This gene encodes a subunit of the transcription factor complex nuclear factor-kappa-B. endpoint of a series of signal transduction including inflammation, immunity, differentiation, cell growth, tumorigenesis and apoptosis.
PVRL2	0.0086	0.0496	This protein is one of the plasma membrane components of adherens junctions
CHST2	0.0144	0.0493	This locus encodes a sulfotransferase protein. The encoded enzyme catalyzes the sulfation of a nonreducing N-acetylglucosamine residue,
PYGB	0.0292	0.0479	The activity of this enzyme is positively regulated by AMP and negatively regulated by ATP, ADP, and glucose-6-phosphate. This enzyme catalyzes the rate-determining step in glycogen degradation
UACA	0.0187	0.0476	Inhibits NFKB1.This gene encodes a protein that contains ankyrin repeats and coiled coil domains and likely plays a role in apoptosis.

E Top 10 negatively weighted genes in PC2

Gene	PCs1	PCs2	Function
RRM2	0.0131	-0.0428	This reductase catalyzes the formation of deoxyribonucleotides from ribonucleotides
PAPOLA	-0.0140	-0.0419	required for the addition of adenosine residues for the creation of the 3'-poly(A) tail of mRNAs
ITGA5	0.0013	-0.0417	The product of this gene belongs to the integrin alpha chain family. Integrins function in cell surface adhesion and signaling.
PLIN2	0.0011	-0.0412	belongs to the perilipin family, members of which coat intracellular lipid storage droplets.
LGMN	0.0151	-0.0401	This gene encodes a cysteine protease that has a strict specificity for hydrolysis of asparaginyl bonds. Overexpression of this gene may be associated with the majority of solid tumor types.
SLC2A3	0.0127	-0.0385	Facilitative glucose transporter that can also mediate the uptake of various monosaccharides across the cell membrane (PubMed:9477959, PubMed:26176916).
CCNB1	0.0245	-0.0379	Regulatory protein involved in mitosis. The gene product complexes with p34(cdc2) to form the maturation-promoting factor (MPF). The encoded protein is necessary for proper control of the G2/M transition phase of the cell cycle
C12orf5	-0.0073	-0.0378	This gene is regulated as part of the p53 tumor suppressor pathway. Functions by blocking glycolysis and directing the pathway into the pentose phosphate shunt.
ANXA5	-0.0198	-0.0371	Annexin 5 is a phospholipase A2 and protein kinase C inhibitory protein with calcium channel activity and a potential role in cellular signal transduction, inflammation, growth and differentiation.
WBP5	-0.0066	-0.0362	This gene encodes a WW domain binding protein. This gene also encodes a domain with similarity to the transcription elongation factor A, SII-related family.

Table 3.1: Gene weightings of principal components PC1 and PC2. All genes are ordered according to their weighting. Since expression values are centred by subtracting from the mean, high gene expression leads to negative values. Therefore, high expression of "positively" weighted genes leads to lower PC1/PC2 values and expression of "negatively" weighted genes leads to higher PC1/PC2 values. (A) Gene weightings of genes of interest used in this thesis. Larger contributions are marked in bold. (B, C) Top 10 positively and negatively weighted genes on PC1 are listed here. (D,E) Top 10 positively and negatively weighted genes on PC2 are listed here.

To test the conversion protocols, SBAD3-01 primed iPS cells were cultured in the 7 different naïve media for 23 days and split every 3-4 days. The naivety of the cells was assessed by morphology (Figure 3.2A) and gene expression (Figure 3.3A). Cells cultured in Theunissen *et al.* 5iLAF medium did not survive the conversion. Naïve morphology is described as round, domed colonies with refractive edges and was observed in RSeT, NHSM 1B and to a lesser extent in ENHSM media. The elongated cells in the background are MEF feeder cells, required by the naïve conversion protocols. Despite naïve morphology in RSeT, NHSM 1B and ENHSM, gene expression profiling discussed in the next section did not show naïve gene expression and further protocols needed to be investigated.

Since none of the above protocols were entirely satisfactory, the naïve cell line HNES1 (derived directly from the inner cell mass of a human blastocyst) was obtained from the originating lab (Guo et al., 2016), to continue the pursuit of the optimal protocol for further experiments. The medium used to grow HNES cells, is referred to as t2iLGoY. In addition, the revised medium 4iLA by Theunissen *et al.* (Table 1.1) was tested on iPSC. Both HNES1 and 4iLA cells exhibited naïve morphology (Figure 3.2B) and all generated cells were used for gene expression profiling.

3.2.3. Gene expression profiling of cells generated by naïve protocols

Since gene expression is the most relevant indicator of naivety, it was measured using a panel of marker genes, as follows. *KLF4*, *TBX3* and *DNMT3L* are genes known to be highly upregulated in naïve cells (Blakeley et al., 2015). NANOG and OCT4 are key transcription factors maintaining pluripotency and are expressed in both naïve and primed states (Wang et al., 2012, Boyer et al., 2005). CER1 and OTX2 are markers of the *in vivo* epiblast in mouse (Joo et al., 2014, Buecker et al., 2014) and human (Yan et al., 2013, Blakeley et al., 2015), and therefore appear downregulated in naïve cells. To find suitable endogenous controls, the expression of established housekeeping genes *RPL13A*, *B2M*, *18s* and *GAPDH* (Vossaert et al., 2013) were tested across a time course of naïve to primed differentiation (Supplementary Figure S3). *GAPDH* was unsuitable since it was

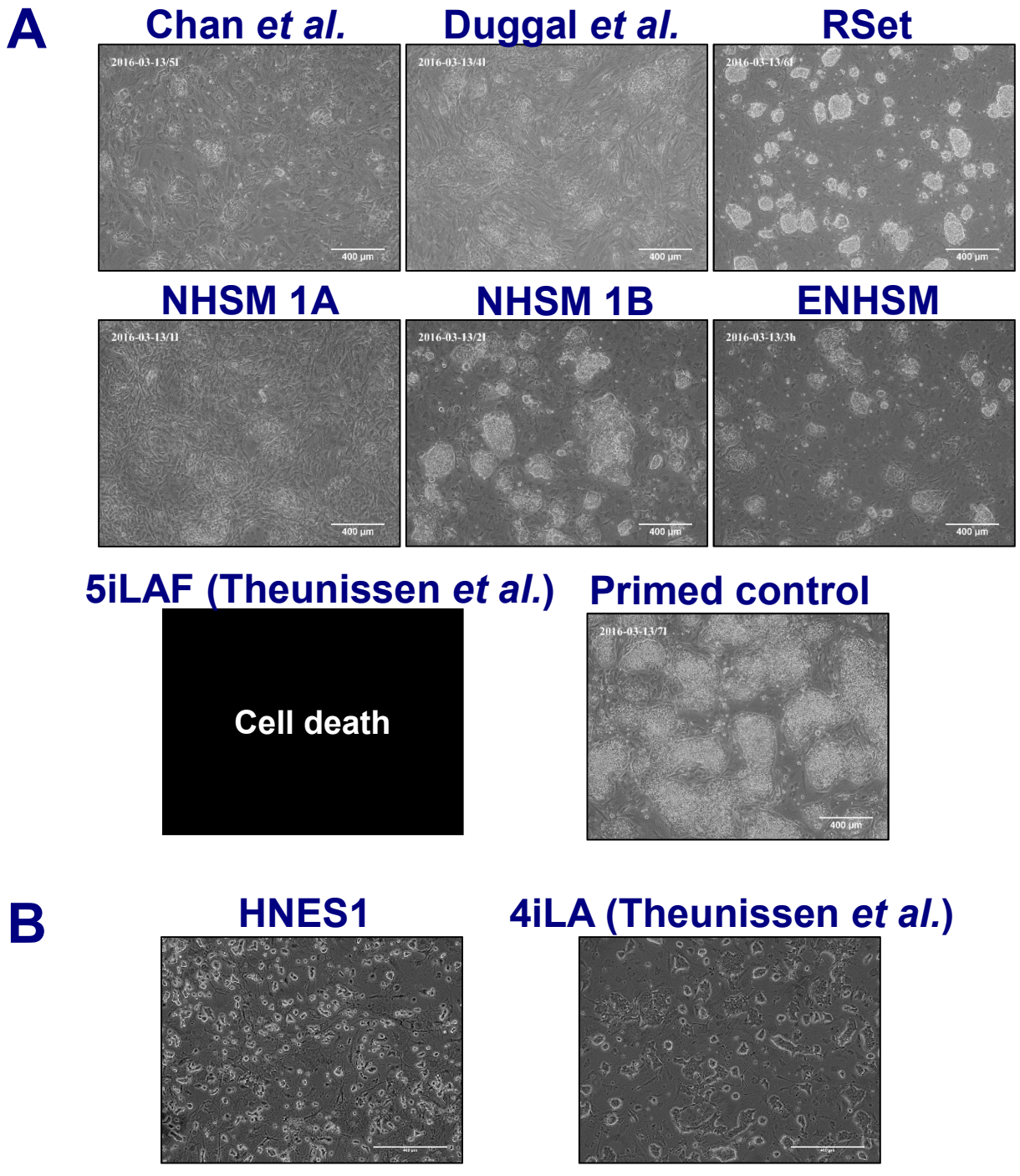


Figure 3.2: Representative morphology of cells treated with different naïve protocols. (A) 7 protocols were directly compared. Expected naïve morphology are round, domed colonies with refractive edges, as seen in cells treated with RSeT or NHSM1B. Cells in Theunissen *et al.* medium did not survive this initial experiment. SBAD3-01 cells were used for all protocols. (B) HNES1 cells were derived by Guo *et al.* directly from a blastocyst into naïve medium. Theunissen *et al.* protocol was repeated with several adaptations using cell line SFC180-01-01 and was successful.

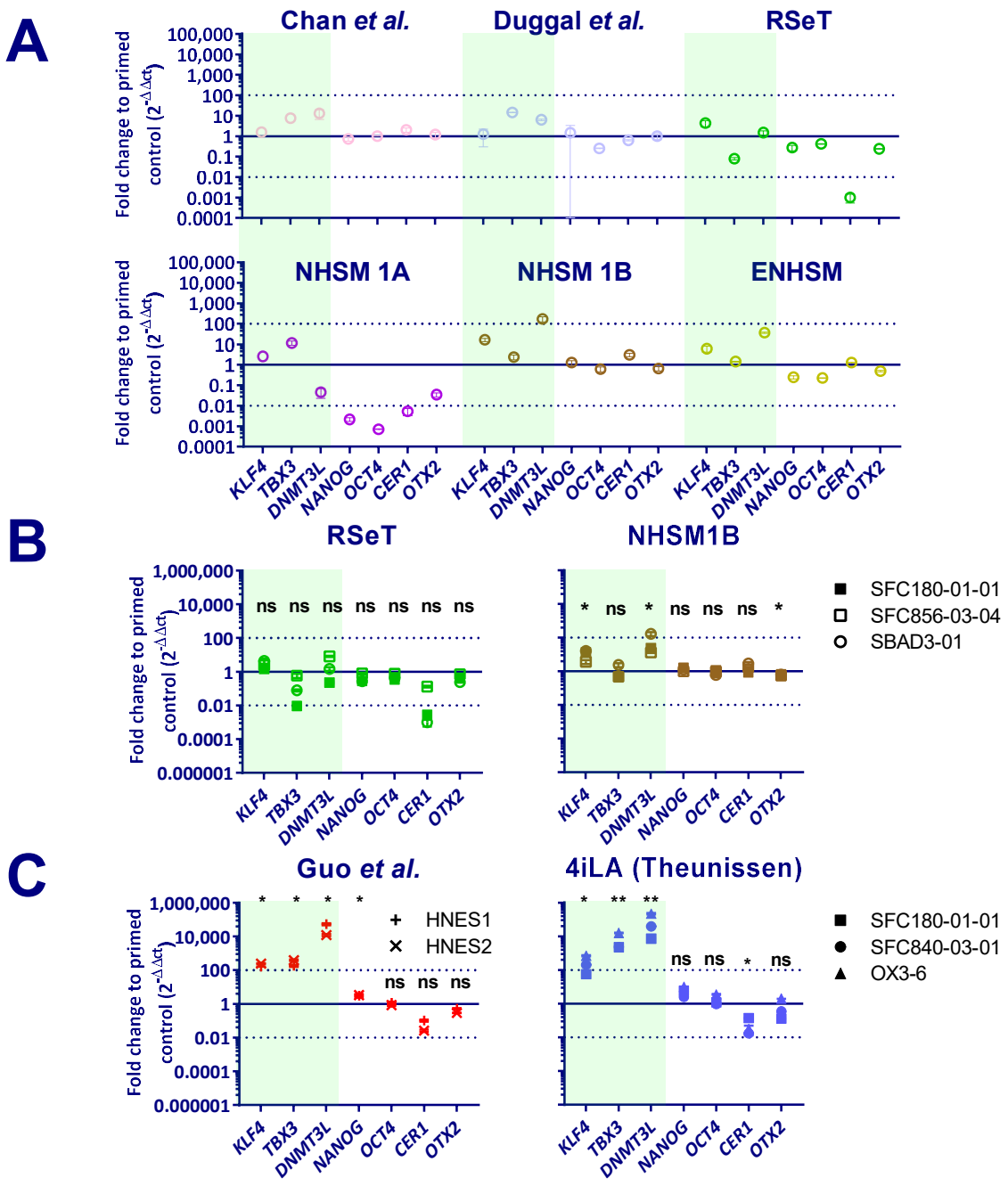


Figure 3.3: Fold change in gene expression of marker genes in cells generated by different naive conversion protocols to their primed control. *KLF4*, *TBX3*, *DNMT3L* (marked with green background) are markers of the naïve state. *NANOG* and *OCT4* (*POU5F1*) are shared marker genes of pluripotency expressed in both cell types and *CER1* and *OTX2* are markers of the primed state. The shape of the symbols indicates which cell line was used and each protocol is represented by a different colour. (A) None of the initially tested 6 protocols showed high upregulation of naïve markers. Only one cell line (SBAD3-01) was used in this initial comparison. (B) The protocols with the most favourable gene expression (NHSM 1B) and most naïve morphology (RSeT) were then tested on 2 more cell lines which reproduced the results. (C) Two lines directly derived from a human blastocyst into naïve media (HNES1/HNES2) and cells generated using the Theunissen *et al.* protocol show improved naïve gene expression. Data normalised first to housekeeping genes (*B2M*, *RPL13A*), then to the primed control using the $2^{-\Delta\Delta ct}$ method. A two-tailed ratio paired t-test of the $2^{-\Delta ct}$ values of biological replicates informed significance (ns = $P > 0.05$; * = $P \leq 0.05$; ** = $P \leq 0.01$). Statistical significance is observed for Guo *et al.* despite only 2 cell lines used. If karyotypically abnormal cells generated by the Theunissen *et al.* protocol are used as further replicates, p values decrease to $p < 0.0001$ for *KLF4*, *TBX3*, *DNMT3L* and $p < 0.001$ for *NANOG* and *CER1*. *OCT4* and *OTX2* remain non-significant.

differentially expressed showing higher expression in primed cells. Primed cells switch to glycolytic metabolism and *GAPDH* is an essential part of the pathway (Ercolani et al., 1988), so this is not a surprising result. *RPL13A* and *B2M* were chosen for normalisation, as they were not differentially regulated between naïve and primed cells. In Figure 3.3A the initial 6 protocols were compared. NHSM1B and to a lesser extent ENHSM show upregulation of naïve genes. Downregulation of primed genes whilst maintaining *NANOG* and *OCT4* expression is only seen in the RSeT condition. *NANOG* and *OCT4* are heavily downregulated in the NHSM 1A condition, suggesting the cells may have differentiated into non-pluripotent cell types. The data in Figure 3.3 is expressed as fold change to primed controls, and the underlying $2^{-\Delta ct}$ values are plotted in Supplementary Figure S2.

RSeT and NHSM 1B conditions were used on two more cell types to test if they could robustly generate naïve cells (Figure 3.3B). *KLF4* and *DNMT3L* were significantly upregulated in NHSM1B cells compared to primed, however *TBX3* is not significantly affected. The RSeT condition led to naïve morphology, but no significant differences in gene expression compared to the primed control. This indicates that naïve-like gene expression (the more physiologically relevant and more stringent readout), does not necessarily correlate with morphology.

The naïve cell line HNES1, and cells that were converted using 4iLA, exhibit strong naïve gene expression (Figure 3.3C). Unlike cells from any of the previous protocols, naïve marker genes including *TBX3* are all significantly upregulated. *KLF4* and *DNMT3L* are upregulated only 10-100 fold in NHSM1B cells but 100-10,000 fold in both HNES and 4iLA cells.

Given that the above experiments showed that 4iLA was the protocol which yielded the most naïve cells, efforts were made to generate further naïve lines with this protocol. To generate the 3 karyotypically normal 4iLA converted naïve cell lines, 6 conversions were performed in total. All were assessed according to their gene expression and a principal component analysis (PCA) was performed to reduce the expression of 7 markers to two principal components (Figure 3.4). The karyotypically normal lines (OX3-6, SFC180-01-01, SFC840-03-01) are plotted together with

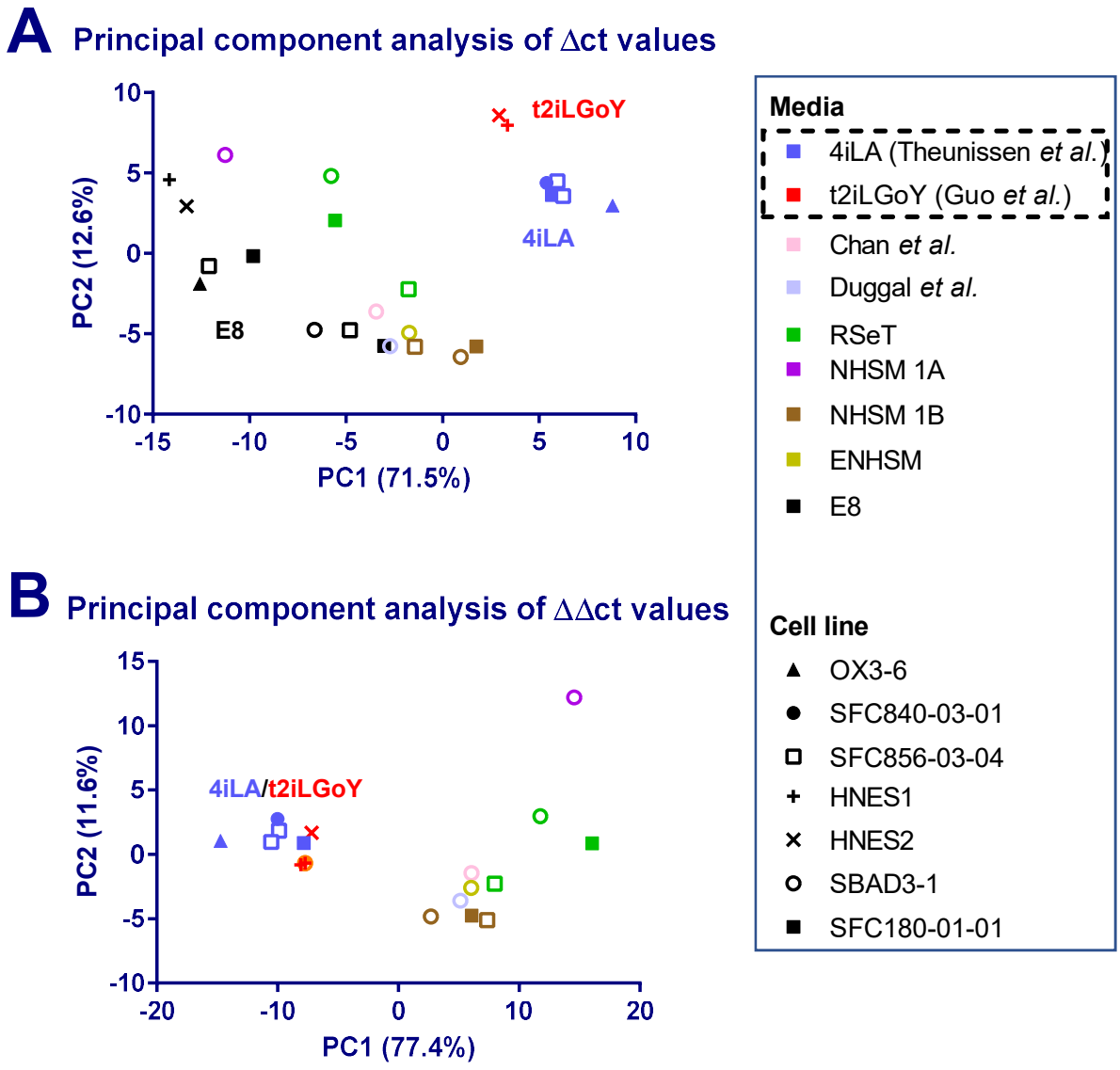


Figure 3.4: The expression of 7 markers (*KLF4, TBX3, DNMT3L, NANOG, OCT4, CER1, OTX2*) relative to their housekeeping genes (*B2M, RPL13A*) were reduced to two constituents by principal component analysis. The colours of each symbol represent the cell culture medium used to generate the naïve cells. The shape represents the cell line used. (A) Despite data from separate experiments, analysis of the Δct values shows a clear distinction between successful protocols (t2iLGoY & 4iLA) and primed controls (E8). Unsuccessful media formulations cluster between E8 and 4iLA/t2iLGoY conditions. (B) All samples were tested alongside a primed control. Therefore normalisation to the primed control (ΔΔct values, equivalent to log2FoldChange) removes variation between experiments. Again, there is a clear separation between successful and unsuccessful protocols (below). The 4iLA converted cell line SFC840-03-02 was omitted in this analysis, due to an incomplete dataset (Figure S5A).

abnormal lines (2x SFC856-03-04) and all other protocols (which were tested but not used in further experiments). The karyotypically abnormal 4iLA SFC840-03-02 cell line was omitted in the PCA analysis since expression of one marker (*DNMT3L*) was not measured due to shortage of the primer/probe. Figure 3.4A shows clustering of the protocols away from E8 controls across all experiments performed. Each experiment included the appropriate primed control, so normalised data is plotted in Figure 3.4B, which shows that differentiated NHSM 1A seen in Figure 3.3 cluster away from both successful and unsuccessful protocols. HNES1 and HNES2 cells (cultured in t2iLGoY medium) cluster with 4iLA cells. The underlying $2^{-\Delta ct}$ values are plotted in Supplementary Figure S4, and those of karyotypically abnormal cells and unused protocols are plotted in Supplementary Figure S5. Together, these results show that HNES cells in t2iLGoY and 4iLA converted cells have naïve-like gene expression and they were therefore used for all further experiments in this thesis. Naïve cells cluster tightly together, whereas cells grown in E8 show a larger spread. This may be due to the tight culture conditions used in naïve protocols compared to the possibly more permissive E8 condition.

3.2.4. Karyotype analysis

The karyotype of all successfully converted 4iLA cells and HNES1 cells were determined by SNP analysis. None of the primed starting cell lines demonstrated any gross karyotypic abnormalities by this analysis, however several converted 4iLA cells were rejected due to large chromosomal rearrangements (Figure 3.5A). The 4iLA SFC841-03-01 line had a duplication of chromosome 5 and both lines derived from SFC856-03-04 had a duplication of the short arm of chromosome 12, which includes relevant pluripotency genes such as *NANOG* and *DPPA3* (in addition to *RAD52*, which is relevant for later chapters). The second conversion of SFC856-03-04 also had an additional duplication of chromosome 9, which includes *KLF4*. Since both conversions of SFC856-03-04 had the same duplication on chromosome 12, it is likely that an undetectable sub-population existed in the original primed batch, which had a selection advantage during naïve conversion. To further elucidate the extent of abnormal cells, genotyping of SFC856-03-04 4iLA

A
SFC180-01-01
SFC840-03-01
OX3-6
Naïve
Primed
SFC856-03-04
SFC841-03-01
A
B
HNES1
Naïve
SBAD3-01
Primed

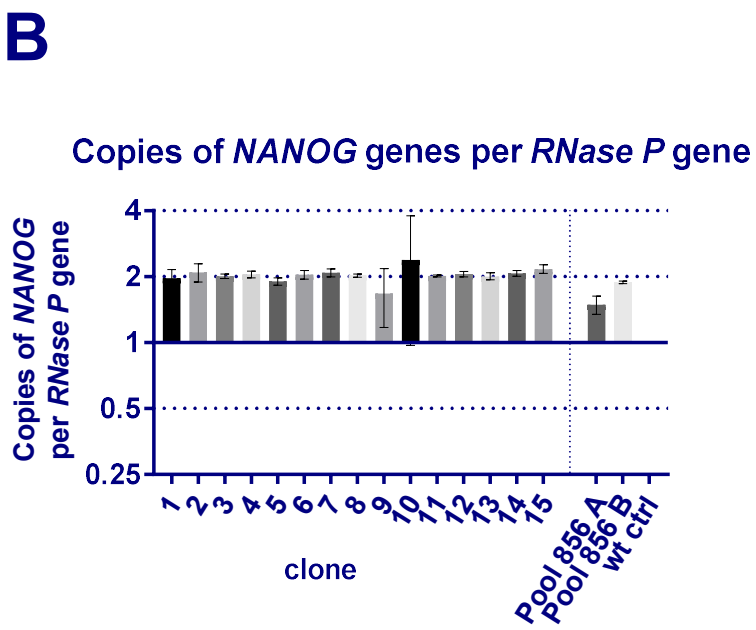


Figure 3.5: Karyotypic abnormalities seen after converting to or deriving naïve hPSC. (A) SNP profiling is shown here. Duplications are marked by green bars next to the genomic location, whereas deletions are shown by orange bars. The X and Y chromosome are marked according to the sex of the cell line. The profiling shows a duplication of Chromosome 4 in naïve SFC841-03-01. In two separate conversions of SFC856-03-04 one arm of chromosome 12 is duplicated (containing naïve genes such as *DPPA3* and *NANOG*). In the second conversion B an additional duplication in chromosome 9 has been detected. Lines SFC180-01-01, SFC840-03-01 and OX3-6 were successfully converted and the SNP analysis showed no gross abnormalities. (B) Copy number variation analysis by qPCR was performed on SFC856-03-04 A (Pool 856 A). The graph shows the mean and standard deviation of 3 technical replicates. SFC856-03-04 A (Pool 856 A) is a mixed population containing approximately half karyotypically normal cells. However, subcloning (clone numbers 1-15) did not identify any of these, possibly due to the large growth advantage of abnormal cells. The copy number of *NANOG* gene compared to the *RNase P* control gene is shown here.

converted cells by qPCR was performed. It showed a homogenous karyotypically abnormal population for conversion B, but a mixed population with approximately half wild type cells for conversion A (Figure 3.5B). To obtain karyotypically normal 4iLA SFC856-03-04 cells, 32 individual colonies were picked for clonal expansion. However, only 15 survived expansion and all had the same duplication of the *NANOG* gene, indicating a substantial growth advantage for abnormal cells. Before the SNP analysis, the karyotypically abnormal 4iLA SFC856-03-04 B cells were used for setting up many experiments but were then replaced by karyotypically healthy lines after receiving the SNP results. The HNES1 cells had two large deletions (in chromosome 10 and 18) and a small duplication on chromosome 12, also containing *NANOG* and *DPPA3* (and *RAD52*). HNES1 cells were used for parts of this thesis but replaced by karyotypically normal 4iLA lines for any experiments performed after the SNP analysis in March 2017. The naïve to primed differentiation experiment was performed with HNES1 and HNES2 cells which had been karyotyped by G-banding prior to the experiment, and no gross karyotypic abnormalities were observed.

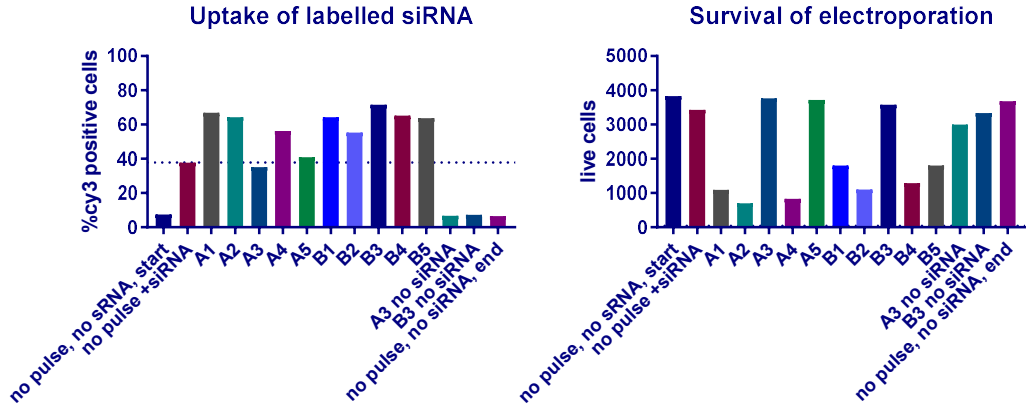
3.2.5. Optimisation of electroporation of naïve hPSC

The uptake of ribonucleoprotein (RNP) and anti-miRNA oligonucleotides (AMOs) was required for techniques used in this thesis. Naïve hPSC do not have a standardised delivery system and are difficult to transfect. To optimise electroporation, different parameters were tested based on programmes used on mouse and human ESC (Figure 3.6A). HNES1 cells were transfected using these protocols with a cy3 labelled siRNA, and fluorescence was measured by flow cytometry after 24h (Figure 3.6B). A clear difference was visible between cells taking up the siRNA (e.g. A1), and cells with siRNA either stuck to the outside of the cell or taken up passively (e.g. no pulse + siRNA control). The survival was measured also by flow cytometry and has therefore accrued large technical variability. Some programmes may not have correctly functioned, since overwhelming cell survival similar to the no pulse control is observed together with unsuccessful transfections (e.g. A3 and A5). The highest transfection and survival was observed with protocol B3, however

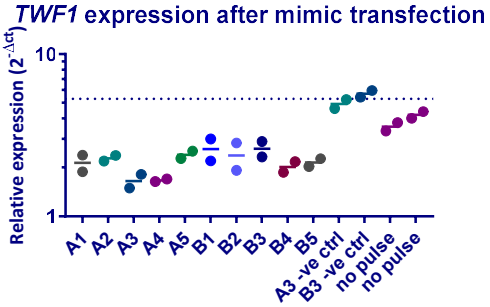
A

Name	Voltage	Pulse width	Pulse number
A1	1050	30	2
A2	1100		
A3	1150		
A4	1200		
A5	1250		
B1	1150	20	1
B2	1200		
B3	1250		
B4	1300		
B5	1350		
h.via	1000	40	1
h.trans	1400	20	1

B



C



D

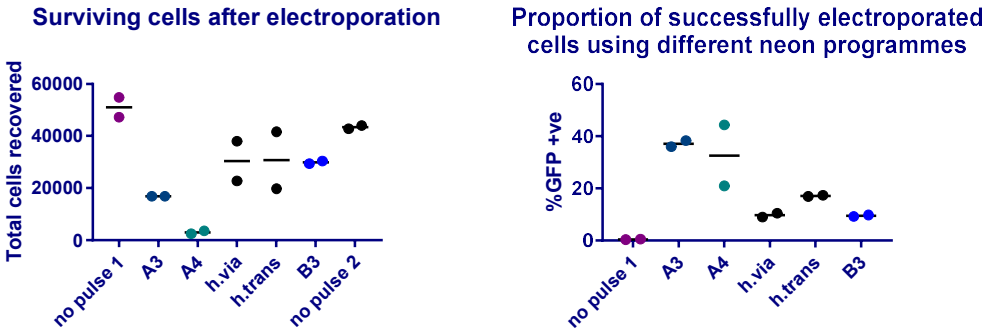


Figure 3.6: Optimising electroporation. (A) A series of different electroporation settings were tested using HNES1 cells. (B) A Cy3 labelled siRNA was introduced by electroporation and uptake measured by flow cytometry (no replicates). Cell survival was not measured immediately but downstream by flow cytometry leading to considerable variation. The no pulse +siRNA control is partially positive likely due to the siRNA either sticking to the outside of the cell or being passively taken up. (C) Electroporation efficiency was also measured by the effect of a control mimic knocking down *TWF1*. Displayed are both technical replicates as dots and the mean as a line. (D) 4iLA cells were tested in replicate by GFP expression from a transfected plasmid. Overall setting A3 was used for further experiments. Gene expression data is normalised to the mean of housekeepers *B2M* and *RPL13A*.

due to unsuccessful electroporations and a lack of replicates, a further set of experiments was performed in order to test the uptake of miRNA mimics by measuring target gene expression. The miRNA mimic of miR-1 reliably downregulated *TWF1* mRNA and has been widely used as a positive control (Kojima et al., 2012, Xu et al., 2009). *TWF1* was downregulated in all electroporation settings (Figure 3.6C). The highest downregulation is achieved in A3 and A4. The A3, A4 and B3 programmes were also tested on 4iLA SFC856-03-04 cells by GFP expression plasmid. Survival was measured by cell counting and GFP expression by flow cytometry. H.via and h.trans were used for electroporation of primed cells and were included in the test. For the naïve SFC856-03-04 line the best compromise between survival and GFP expressing cells was programme A3 which was used for all further electroporations of naïve cells.

3.2.6. Optimisation of flow cytometry markers for naïve vs. primed cells

Antibodies staining hPSC cell surface markers (Collier et al., 2017) were titrated for flow cytometry (Figure 3.7). Naïve 4iLA and primed SFC856-03-04 cells were mixed to have both a negative and positive population to find the best antibody concentration. Separation between negative and positive populations can be calculated using the separation index (SI)(Martin, 2007). The highest SI indicates the best antibody concentration to maximise positive signal and minimise background. The gating strategy avoids doublets and is shown in Figure 3.7A. SI values are plotted in Figure 3.7B. The SI value is calculated by comparing median fluorescent intensities of the positive and negative peaks and normalising by the right-hand side (84th percentile) of the negative distribution as shown below.

$$SI = \frac{(Median_{pos} - Median_{neg})}{\left(\frac{(84\%_{neg} - Median_{neg})}{0.995}\right)}$$

The ideal antibody concentration for CD24 is 1:4 and for CD90.2 1:2 of the manufacturer's recommended dilution (listed in Table 2.3). CD90, CD130 and CD57 have not plateaued at the manufacturers recommended dilution and this will be used for further experiments. The SI of

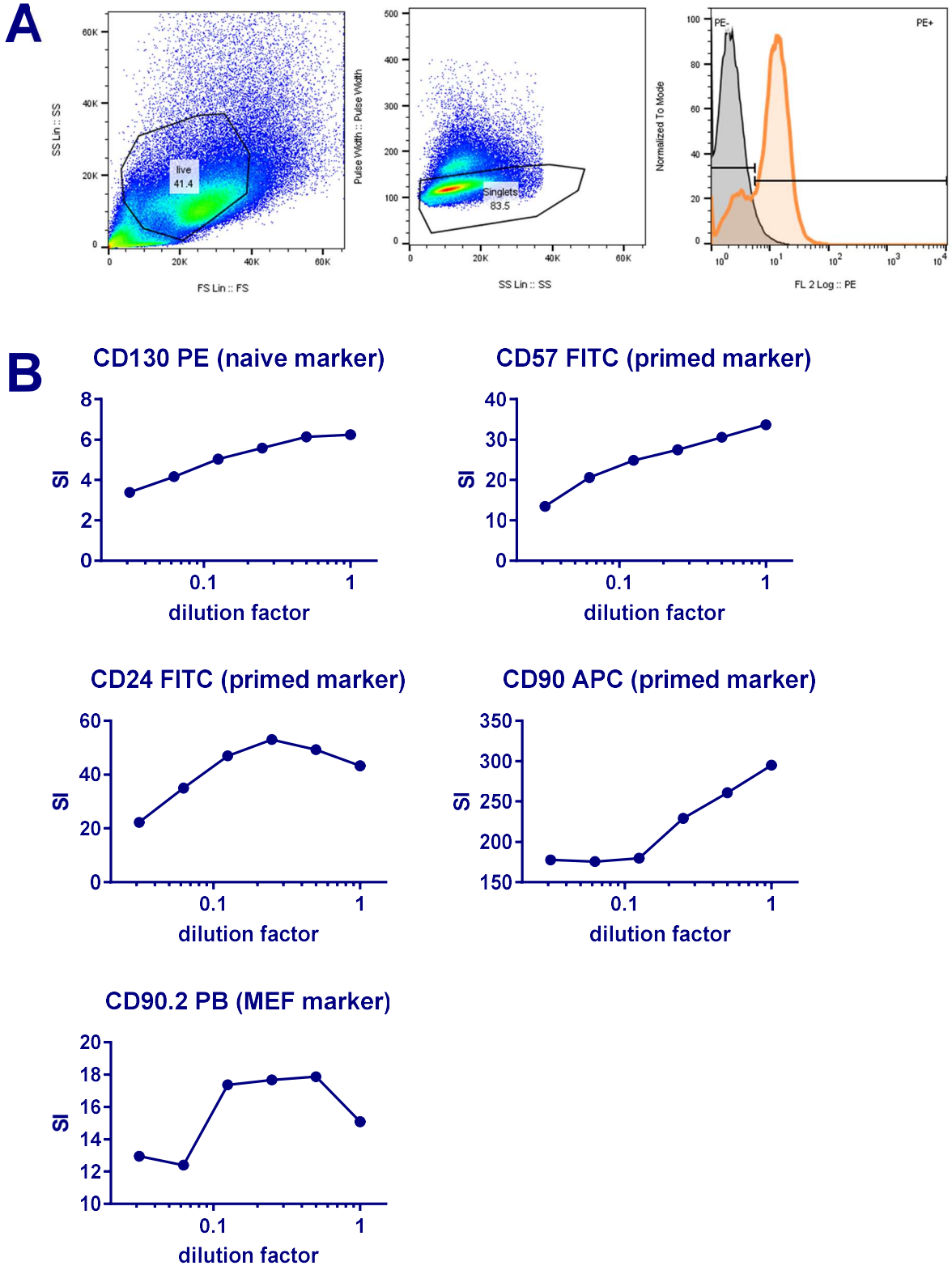


Figure 3.7: Validation of flow cytometry markers. The optimal dilutions for antibodies staining naïve or primed cells were calculated by mixing naïve and primed cells and staining with different dilutions of antibody (no replicates). (A) The gating strategy selects live cells in singlets. (B) The separation index (SI) is calculated by comparing negative and positive populations. The starting dilutions were recommended dilutions by the antibody manufacturer listed in Table 2.3. Abbreviations: APC = Allophycocyanin; FITC = Fluorescein isothiocyanate; PB = Pacific Blue; PE = Phycoerythrin

CD90 flattens at the lower concentrations due to a technical artefact: the median of the negative population reached the axis limit and a higher laser voltage should have been used to avoid this. This does not affect the higher doses of the antibody, so it has not influenced the outcome of this optimisation experiment. With the antibody concentrations optimised, this marker panel was ready to be used for further experiments.

3.3. Summary and discussion

3.3.1. Comparison of published RNAseq data

With many protocols recently published that claim to convert cells to the naïve state, an objective comparison was required. Re-analysing the published RNAseq data revealed that only naïve cells converted using Theunissen *et al.* and Takashima *et al.* protocols showed very similar gene expression to cells of the inner cell mass of a human E5 blastocyst. This is in line with analyses performed by two different groups:

The authors of the first study (Huang et al., 2014) compared the gene expression of naïve cells generated by Takashima *et al.*, Theunissen *et al.*, Gafni *et al.*, Ware *et al.* and Chan *et al.* to *in vivo* cell stages, using a weighted gene co-expression network analysis (WGCNA). The limitations of WGCNA are the inherent subjectivity of the analysis as well as the requirement for many biological replicates for a robust result. They did not have access to the by far most extensive and largest *in vivo* single cell RNAseq dataset by Petropoulos *et al.* (Petropoulos et al., 2016), that has been used in this thesis, but they exploited only the Yan *et al.* dataset (Yan et al., 2013) together with microarray data (Vassena et al., 2011). Not all protocols had been published at that point, and for Gafni *et al.* and Ware *et al.* only the microarray data was available. To perform cross-platform comparisons and remove batch effects, each dataset was processed with different parameters chosen by the authors. They also had to amplify the results by raising their correlation matrix to a user-defined power. They found that cells by Takashima *et al.* and Theunissen *et al.* most closely

resembled cells from the human ICM. This result alone should be viewed with caution due to the many user interventions which may have caused bias.

In 2018, a more refined analysis was performed by Stirparo *et al.*, together with the Smith lab (who devised the HNES and t2iLGoY conditions). Unlike Huang *et al.*, the authors used only RNAseq datasets and relied heavily on the Petropolous *et al.* publication. With less user intervention, they used a WGCN analysis to reveal that cells generated using protocols by Theunissen *et al.*, Takashima *et al.* and Guo *et al.* were the only ones with a high degree of similarity to the human ICM. Their analysis is certainly more powerful than Huang *et al.* but since the Authors developed the t2iLGoY condition themselves, their analysis may not be as impartial (Stirparo et al., 2018).

The approach in this thesis was both impartial and used a powerful analysis. We avoided any potential bias which may have been introduced by user defined parameters and used only RNAseq data. Rather than comparing generally to the ICM, we used the Petropolous *et al.* dataset to have a higher resolution and separate cells resembling different embryonic days. Our principal component analysis focusses on genes which vary during development since the analysis was performed on cells at different developmental stages. With our impartial analysis we observed that naïve cells by Takashima *et al.* and Theunissen *et al.* closely resemble cells from the inner cell mass of an E5 blastocyst using published datasets.

3.3.2. Experimental comparison of naïve protocols

Next, each protocol was experimentally tested side by side. Protocols by Chan *et al.*, Duggal *et al.* and multiple variations of Gafni *et al.* did not lead to cells with naïve gene expression, supporting the results from the RNAseq comparison. The published 5iLAF medium by Theunissen *et al.* initially lead to cell death. When repeated with more cells and without bFGF and the GSK-3β inhibitor as seen in more recent publications (Sahakyan et al., 2017, Theunissen et al., 2016),

successful conversion was repeatedly observed. Both HNES and 4iLA cells showed naïve gene expression measured by qPCR.

Many of the 4iLA cell lines generated were revealed to have large karyotypic abnormalities. These abnormalities are frequently observed with naïve cells (Kilens et al., 2018, Pastor et al., 2016, Guo et al., 2016, Theunissen et al., 2014). The inhibitor PD03 might be at least partially responsible for genomic instability. The inhibition of MEK/ERK signalling by PD03 is required for mouse ES self-renewal and pluripotency. However, a knockout of ERK1/2 in mouse ESC causes genomic instability and loss of pluripotency (Chen et al., 2015). The same was observed in a knockdown in human ES cells (Ma et al., 2016b) indicating that the pharmacological inhibition of the MEK/ERK pathway might require careful titration (Di Stefano and Hochedlinger, 2018). An additional factor might be the state of hypomethylation observed in naïve PSC. Unrelated studies have linked hypomethylation to cancer and genomic instability (Gaudet et al., 2003), suggesting that the accumulation of karyotypic abnormalities may be an inherent property of naïve hPSC. However, further research (not in the scope of this thesis) is required to validate and quantify these observations. Importantly, despite the propensity for abnormalities, 3 karyotypically normal naïve lines have been established.

Since naïve hPSC were now available for further work and since they are a novel cell type, methods for manipulation and evaluation needed to be optimised for these cells. Delivery of RNPs or miRNA inhibitors was enabled by empirically determining the optimal electroporation parameters. Cell surface markers for naïve and primed cells for flow cytometry were validated and the antibodies titrated. All published markers by Collier *et al.* tested in this thesis successfully distinguished naïve and primed cells (Collier et al., 2017).

In summary, this chapter identifies the best protocol for naïve hPSC generation. This protocol was then used to generate 3 karyotypically normal naïve hPSC lines. Protein and nucleic acid delivery by electroporation were refined, and naïve and primed cell surface marker antibodies were

titrated for optimal use. Therefore, optimal conditions have been established to study miRNA expression profiles in chapter 4 and interrogate DNA repair pathways in chapter 5.

4

miRNAs involved in naïve-primed differentiation

4.1. Introduction

Having established in chapter 3 which are the best available conditions under which iPSC can be cultured as naïve cells, we can now assess their miRNAome.

4.1.1. Small non-coding RNAs

Micro RNAs (miRNAs) are a class of small non-coding RNAs (~22 nucleotides) which inhibit complementary mRNAs by binding to the 3' untranslated region (UTR) and either flagging it for degradation or blocking translation. The miRNA biogenesis pathway begins with a long primary transcript (pri-miRNA) either transcribed directly by RNA polymerase II or from a spliced out intron. The pri-miRNA contains at least one hairpin which is cleaved by DGCR8 in the nucleus. This forms a precursor miRNA (pre-miRNA) which is transported to the cytoplasm and the loop structure at the end of the hairpin is removed by DICER1. The resulting double stranded mature miRNA is loaded onto AGO, one strand is removed, and the miRNA can perform its functions. The miRNA complex causes silencing either by inhibiting translation or by degradation of the target mRNA. To perform its silencing functions, the miRNA does not require perfect base pairing and often targets many genes. (Bartel, 2009, Fabian and Sonenberg, 2012)

Small interfering RNA (siRNA) describe a similar species of small non-coding RNA with several subtle but important differences. siRNA are produced from double stranded RNA (dsRNA)

structures. These can originate from endogenous transcripts forming intra or intermolecular duplexes or from exogenous material such as viral RNA or transfection of artificial siRNA. The guide strand of the siRNA duplex is loaded onto the AGO and causes degradation of its targets by perfect base pairing. By contrast to miRNA, these interactions tend to be highly specific. (Carthew and Sontheimer, 2009)

Piwi-interacting RNA (piRNA) is a class of animal specific, non-coding RNA molecules which form RNA-protein complexes with piwi proteins. These complexes are involved in post-transcriptional silencing, especially of transposons or in viral defence. They differ from miRNAs in size (26-31 nucleotides) and low sequence conservation. The biogenesis pathway is distinct from miRNA and siRNA, as piRNAs are produced from long single stranded RNA. In birds and mammals, long non-coding RNA (lncRNAs) are the starting point for piRNA production. (Ozata et al., 2018)

4.1.2. Aims

miRNAs are non-essential during mouse early pre-implantation development but are required for further differentiation from E5.5 (Spruce et al., 2010, Bernstein et al., 2003). The naïve to primed differentiation has been successfully used in mouse cells as a model for early embryogenesis to uncover miRNA requirements (Wang et al., 2007, Kanellopoulou et al., 2005). This has not yet been achieved in the human system and it is unclear what the miRNA requirements for early human embryogenesis are. The aims of this chapter are to:

- Establish naïve to primed differentiation using human PSC (4.2.1)
- Perform miRNA profiling on naïve, transitioning and primed cells in (4.2.2)
- Dissect which miRNAs are important for the naïve to primed differentiation in (4.2.2 & 4.2.5 & 4.2.5)

4.2. Results

4.2.1. Human naïve hPSC can be differentiated to the primed state

To investigate miRNAs involved in the naïve to primed differentiation in humans, the differentiation itself needed to be established *in vitro*. The differentiation was investigated by separately culturing three replicates each of HNES1 and HNES2 cells, plating them on geltrex and switching medium to E8 after 48 hours (Figure 4.1A). Due to restrictions on the availability of the cell lines at the time of this experiment, cell culture was performed by Dr Maria Rostovskaya in Prof. Austin Smith's lab, and lysed pellets were sent to me for downstream processing. The time points were chosen to capture the exit from naïve pluripotency and a final late time point to ensure complete priming of the cells. Day 42 was chosen for the final time point since the cells had regained the ability to differentiate using standard primed protocols (data from the Smith lab - not shown). Cells were plated 48h before switching to E8 medium to remove variability introduced by splitting the cells.

When HNES cells were differentiated to the primed state, naïve marker genes were downregulated but not all simultaneously (Figure 4.1B). *KLF4* is steadily significantly downregulated immediately and continuously. *TBX3* is significantly downregulated rapidly within 24 hours and maintains these levels until a further downregulation when the cells have reached the fully primed state. *DNMT3L* is consistently and slowly significantly downregulated during the first 5 days and is heavily downregulated by the time the primed state has been acquired. *DPPA3* levels do not change during the first 5 days but is heavily and significantly downregulated when the cells are fully primed. Markers of shared pluripotency (*NANOG* and *OCT4*), active in both states, do not change much during the differentiation (Figure 4.1C). The primed marker *CER1* is upregulated within the first 5 days, whereas *OTX2* is only heavily upregulated upon acquisition of the primed state (Figure 4.1D).

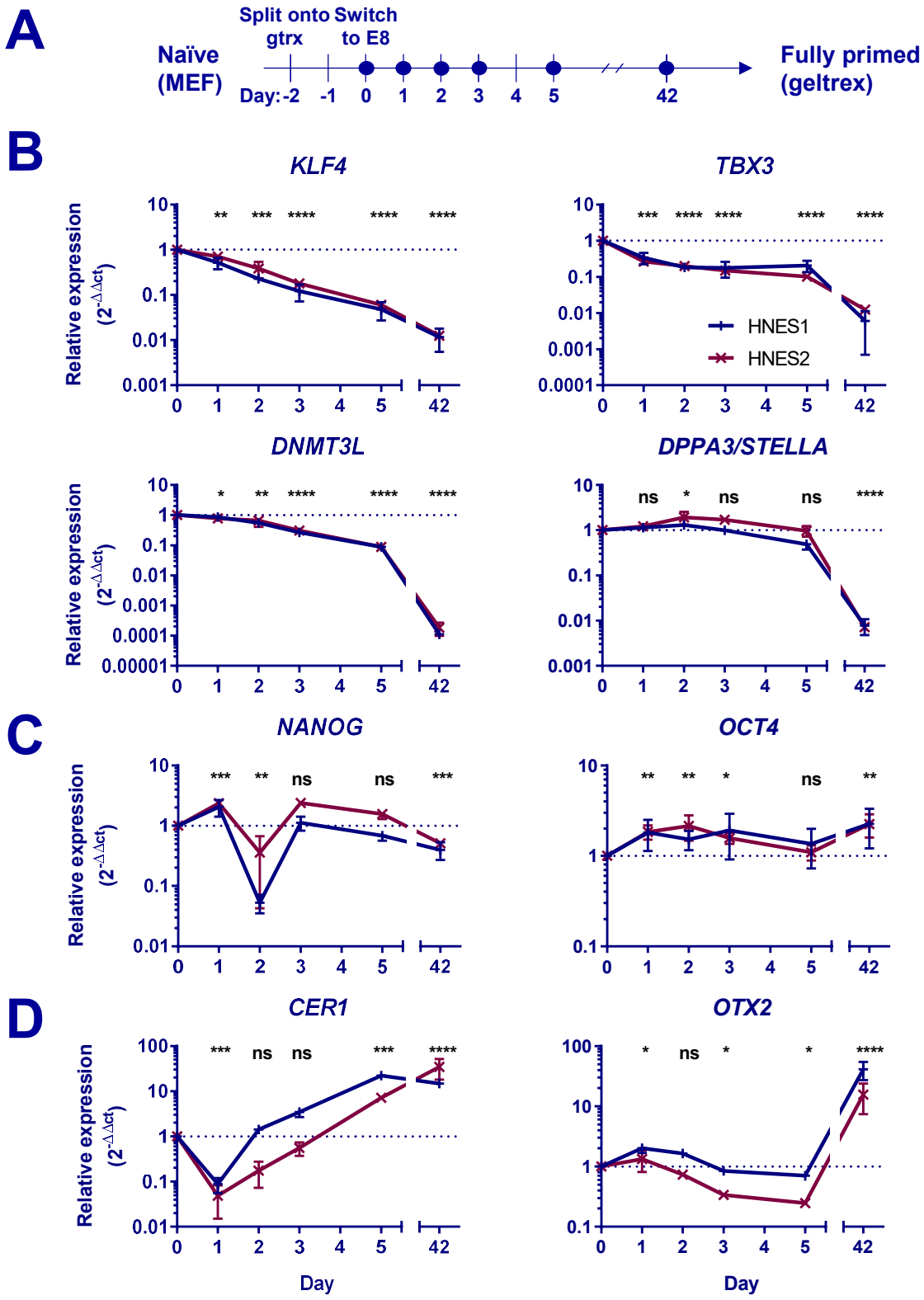


Figure 4.1: Gene expression during naive-primed differentiation. (A) Schematic of the experiment, dots represent when samples were taken. (B) Markers of the naïve state are downregulated rapidly, (C) shared pluripotency markers are expressed throughout and (D) genes associated with the primed state are upregulated. Data normalised first to housekeeping genes (*B2M, RPL13A*), then to day 0 using the $2^{-\Delta\Delta ct}$ method. Error bars indicate the standard deviation between 3 biological replicates for each cell line. A two-tailed ratio paired t-test of the $2^{-\Delta ct}$ values of all 6 replicates informed significance (ns = P > 0.05; * = P ≤ 0.05; ** = P ≤ 0.01; *** = P ≤ 0.001; **** = P ≤ 0.0001). HNES1 is marked blue, HNES2 red.

The differentiation worked robustly by plating cells onto geltrex and switching to E8. E8 contains bFGF, which is required for maintaining primed hPSC and could feasibly drive exit from the naïve state. Alternatively, releasing the cells from inhibitors used in the naïve medium or the removal of MEFs may be sufficient to allow differentiation. Finally, growing naïve hPSC in feeder-free conditions would aid comparisons to primed or differentiating cells and would remove the potential for contaminating MEFs affecting downstream applications. To address these possibilities, HNES1 cells were differentiated under several conditions (labelled 1-6) and their gene expression and survival were assessed at different time points (Figure 4.2A). Naïve cells kept in naïve medium (t2iLGoY) showed the same expression of naïve marker genes when plated onto geltrex (condition 1) or MEFs (condition 2; Figure 4.2B). In addition, when plated onto geltrex fewer cells survived the initial plating but recovered to the same cell number seen with naïve cells on MEFs (Figure 4.2C). This indicates that naïve cells can be kept in feeder-free conditions at least for one passage. Naïve HNES1 cells downregulate naïve markers at the same rate in E8 with or without MEFs (conditions 6 & 4). In addition, switching to basal medium containing no inhibitors or growth factors also caused differentiation equivalent to switching to E8 (condition 5; Figure 4.2B). Finally, a large loss of cells was observed when switching to feeder-free conditions and E8 at the same time (condition 4; Figure 4.2C). However, this can be circumvented by delaying the switch to 24h after plating onto geltrex (condition 3). Cell numbers are then equivalent to MEF conditions and the differentiation (as measured by the downregulation of naïve markers) is delayed by 24h (Figure 4.2B). Due to low cell numbers, only one timepoint of the delayed E8 switch has been assessed. However, this is sufficient to inform further experiments. In conclusion, plating naïve hPSC onto geltrex in naïve medium and switching to E8 medium after 24h causes differentiation, while minimising cells lost during plating.

Cell surface markers measurable by flow cytometry of naïve and primed states have recently been identified and published (Collier et al., 2017) and allow confirmation of our results. HNES1 cells were kept in several conditions and assessed by flow cytometry after 3, 6 and 9 days. Control

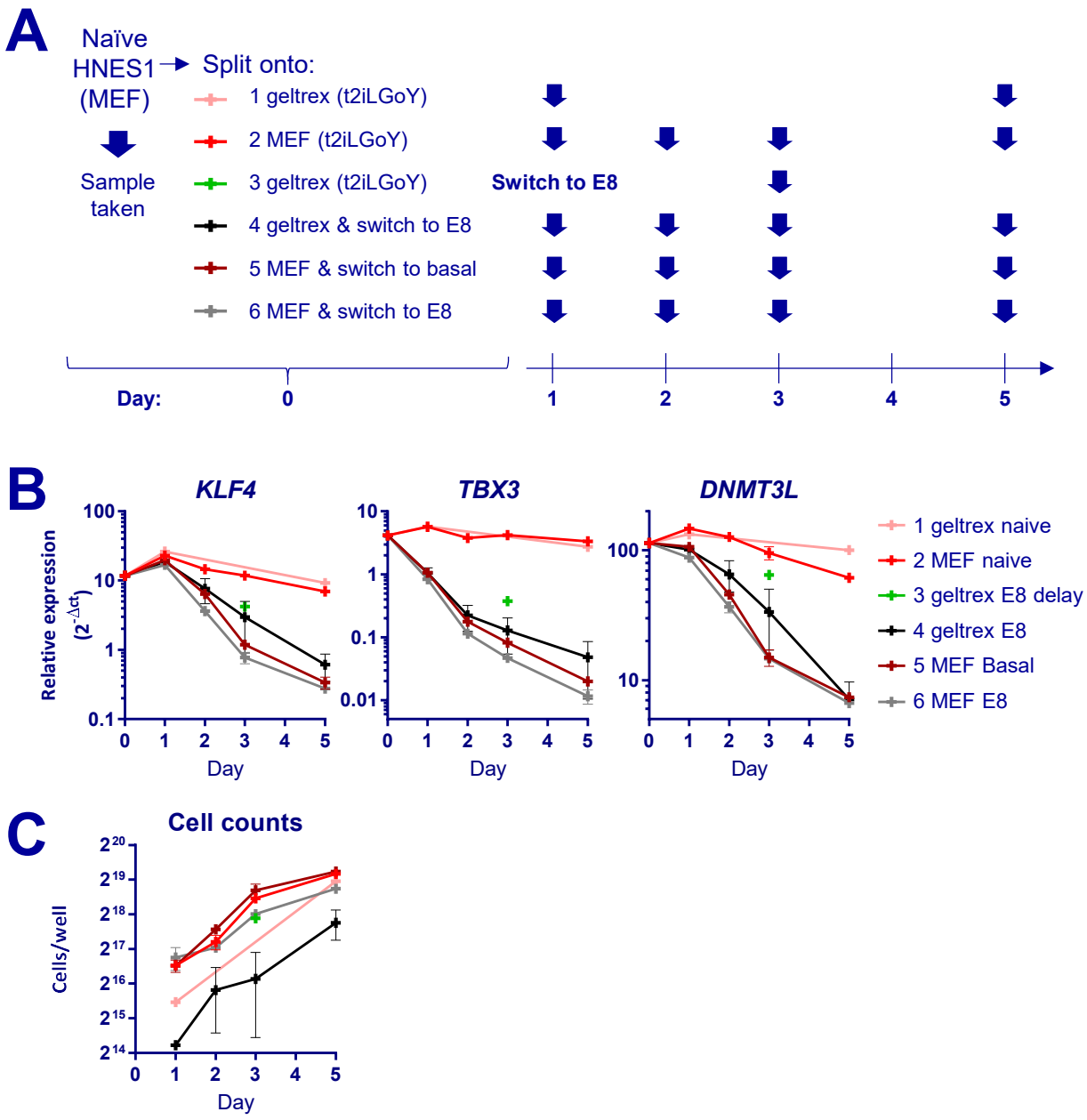
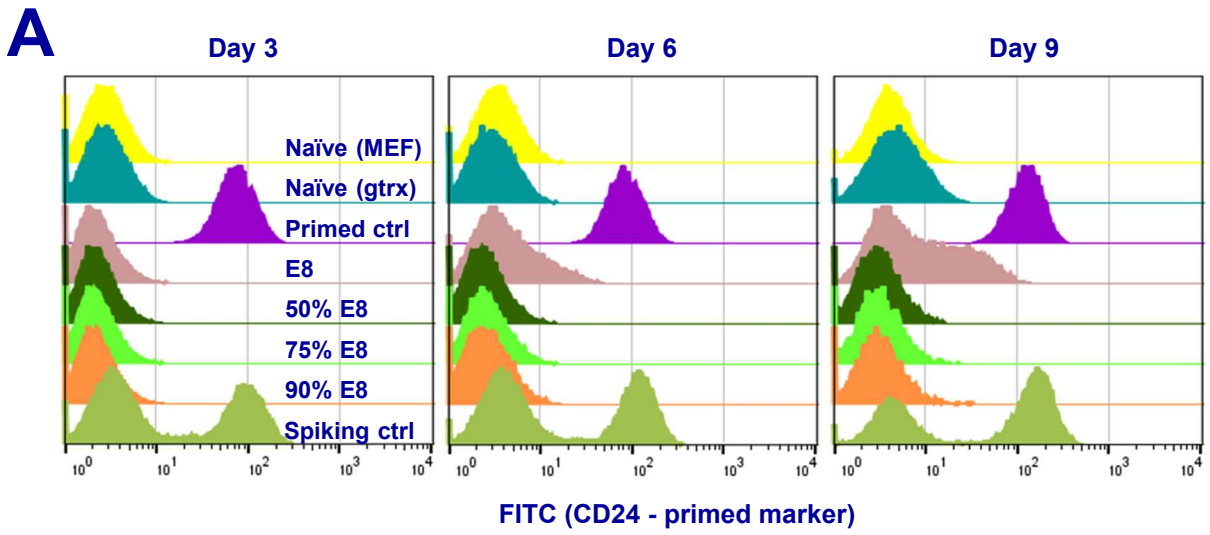


Figure 4.2: Effects of feeder-free culture on differentiation and survival. (A) Schematic of the experiment shows that not every time point was collected for conditions 1 and 3. The conditions used are marked by number and colour coded as shown. (B) Naïve medium retains naïve gene expression across this time course (2) even in feeder-free conditions (1). Delaying the switch from naïve medium to E8 (3) delays the differentiation compared to direct switch to E8 (4). Withdrawal of inhibitors and growth factors from the naïve medium causes immediate differentiation (5) equivalent to adding E8 containing bFGF (6). (C) Seeding naïve cells directly into E8 (4) comes at a viability cost which can be circumvented by switching to E8 a day later (3).

conditions were: cells kept naïve on MEFs (equivalent to condition 2), a primed control and a spike control where both were mixed to allow easier gating. Naïve cells on geltrex (equivalent to condition 1) were included to confirm that at least transient culture of naïve cells in feeder free conditions is possible. To test the differentiation, an E8 condition was included equivalent to condition 3 above. Finally, to answer the question if an intermediate condition allowing both naïve and primed cell survival exists, 3 titrations of naïve medium into E8 were also examined. Cells were plated on day 0 and medium switched to differentiating conditions after 24h.

CD24 is a marker of the primed state and the primed control is heavily positive (Figure 4.3A), whereas the naïve control remains negative across the time course. The differentiating E8 condition heterogeneously starts expressing CD24 at day 6 and increases expression at day 9 but not reaching the levels of the positive primed control. No other conditions caused the acquisition of CD24 signal.

The naïve marker CD130 was measured together with the primed marker CD90. Controls to ensure correct gating are shown in Figure 4.2B. Single stain controls were performed on the spiking control (a mixture of naïve and primed control cells) and informed gating. Naïve and primed control cells were stained against both markers and are overlaid to show effective separation by the choice of markers. Figure 4.2C shows the tested conditions. Naïve cells maintained on geltrex resemble those on MEFs, again confirming that transient feeder-free culture can be achieved for naïve cells. The differentiating E8 condition leads to an immediate homogenous gain of CD90 expression at the earliest time point day 3. The naïve marker CD130 seems unaffected for the first 6 days. Only at the final time point day 9 is there a small homogenous shift toward lower expression. Complete loss of this marker may require further culture in E8. Even if this marker is no longer expressed, the protein may have a long half-life. An additional condition testing the effect of LIF removal showed no difference to the naïve control. Titration of naïve medium into E8 caused cell death (Supplementary Figure S6) and less cells were



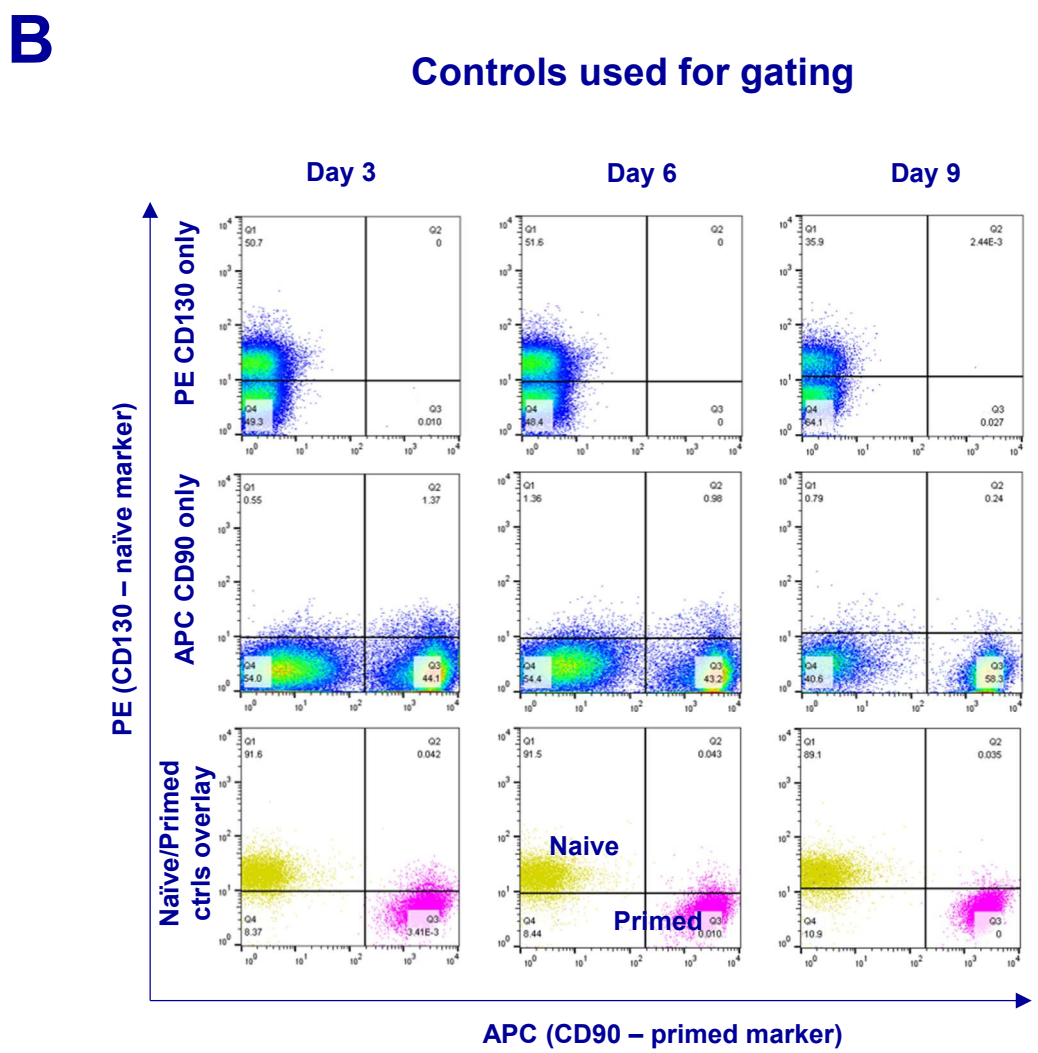


Figure 4.3 continued on next page

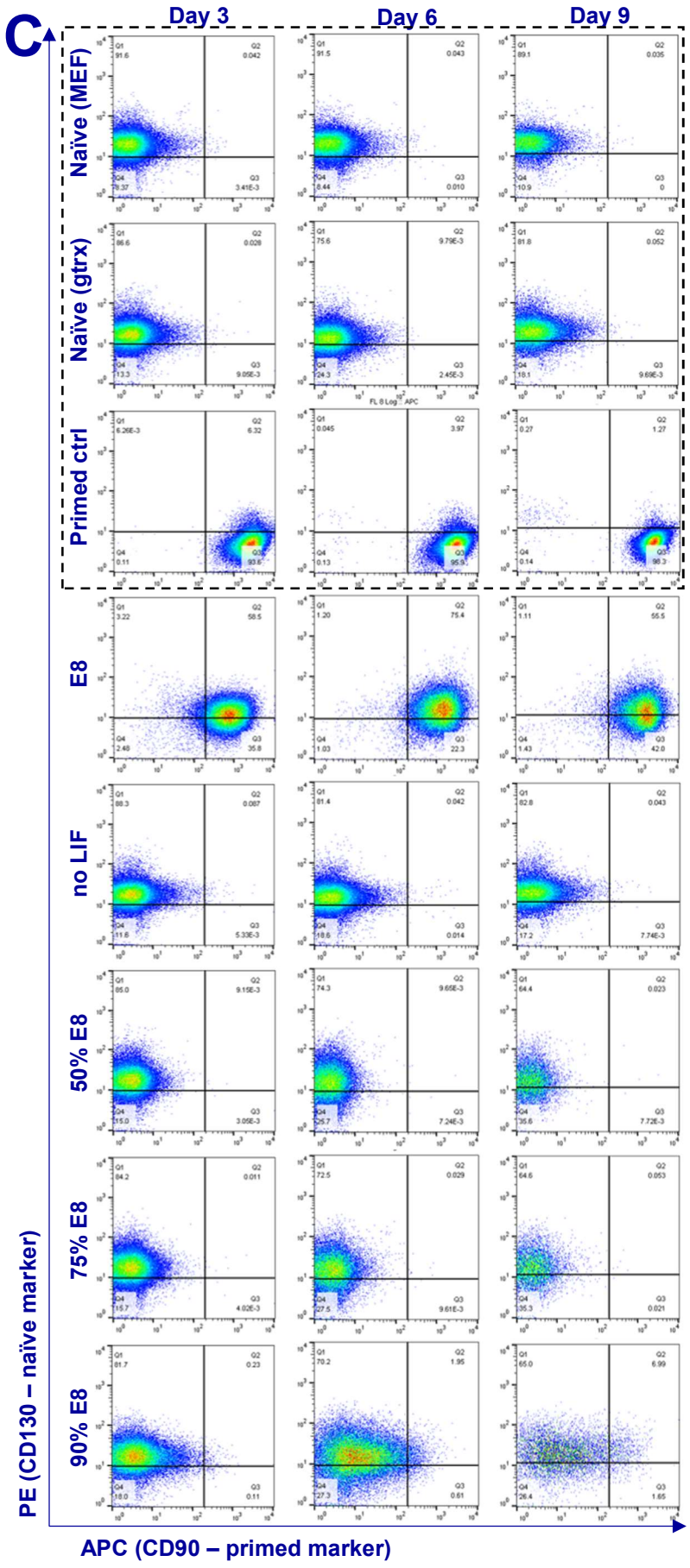


Figure 4.3: Priming causes immediate gain of CD90 and slow acquisition of CD24 protein expression. (A) Histogram of CD24 expression is shown here for each different condition across a time course. CD24 is gained gradually and heterogeneously over the entire time course of 9 days after naïve cells are switched to E8 medium. (B) CD130 and CD90 levels were measured simultaneously. Single stain controls and the overlay of control naïve and primed cells show the panel distinguishes both cell types well. (C) CD130 and CD90 levels of all samples are shown here. CD90 protein is detected homogenously in naïve cells switched to E8 medium from the earliest time point at day 3, whereas a slow loss of naïve marker CD130 is only starting to occur at day 9 or later. Cell death while maintaining naïve protein signatures was observed when naïve medium was titrated into E8 (Supplementary Figure S6). Only in 90% E8 is there a heterogeneous and low gain of primed marker CD90. This suggest the naïve medium inhibitors and growth factors maintain cells tightly in the naïve state. An additional condition of naïve medium excluding LIF behaves the same as the naïve controls.

available for the flow cytometry experiments. Only the lowest concentration of 10% naïve medium caused heterogenous gain of CD90 and only after 6 days compared to the instant homogenous acquisition in the E8 condition.

In summary, the naïve-to-primed differentiation works robustly by plating cells into feeder-free conditions, removing naïve medium and replacing with E8 medium after 24h. This has been demonstrated by changes in marker gene expression and cell surface marker expression. Naïve cells can be cultured in feeder-free conditions at least for one passage. Withdrawal of inhibitors and growth factors from naïve medium is sufficient to cause differentiation. Mixing of naïve and primed medium is toxic to naïve cells.

4.2.2. miRNAseq reveals candidate miRNAs

miRNAs involved in the exit from naïve pluripotency may be upregulated during the exit to dismantle the naïve pluripotency network. To find miRNAs upregulated during the differentiation, samples were taken across a differentiation time course of HNES1 and HNES2 cells (Figure 4.1A). Successful differentiation was confirmed by marker gene expression (Figure 4.1B, C, D). RNA samples from naïve cells (day 0), differentiating (day 2) and fully primed cells (day 42) were sent for miRNAseq. Each HNES1 and HNES2 cell line was cultured separately in biological triplicates. Data pre-processing was performed by Dr Klas Hatje (Supplementary Method 7.1.2). Principal component analysis shows a clear separation between naïve, differentiating, and primed cells, indicating that the miRNAome alone is sufficient to distinguish between states (Figure 4.4A). Cells at day 0 and day 2 form separate tight clusters, whereas primed cells at day 42 are more diffuse. This is likely due to separate culturing for 42 days. The top ten expressed miRNAs in each state have been plotted in Figure 4.4B, C, D. miRNAs of interest investigated further are marked as bold. As shown in the PCA plot the expression of these miRNAs vary between the states. miRNA-22-3p, miRNA-143-3p and miRNA-512-3p are expressed highly in naïve but not primed cells. Conversely, miRNA-92b-3p, miRNA-302a-5p and miRNA-302b-3p are heavily expressed in primed but not

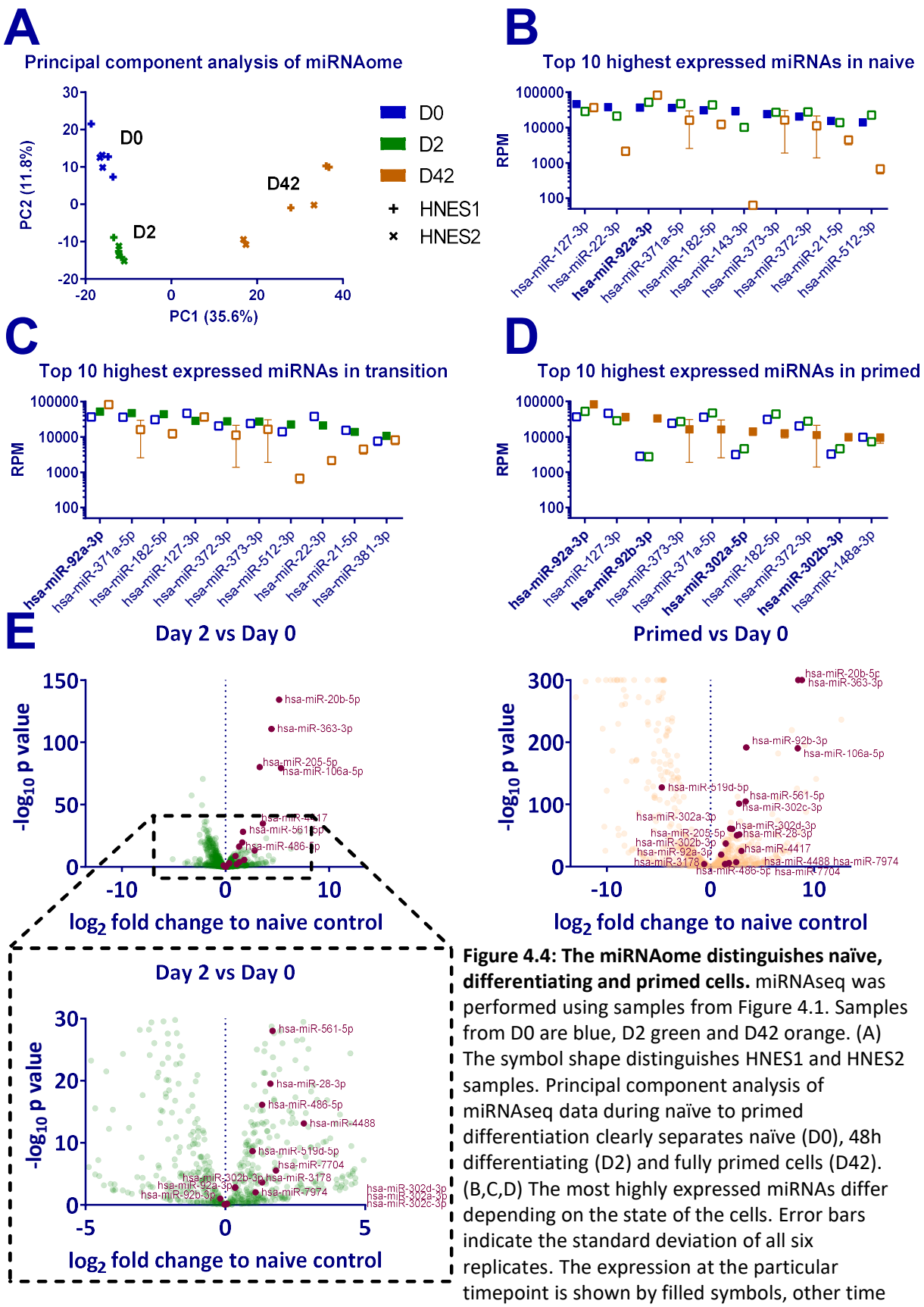


Figure 4.4: The miRNAome distinguishes naïve, differentiating and primed cells. miRNAseq was performed using samples from Figure 4.1. Samples from D0 are blue, D2 green and D42 orange. (A) The symbol shape distinguishes HNES1 and HNES2 samples. Principal component analysis of miRNAseq data during naïve to primed differentiation clearly separates naïve (D0), 48h differentiating (D2) and fully primed cells (D42). (B,C,D) The most highly expressed miRNAs differ depending on the state of the cells. Error bars indicate the standard deviation of all six replicates. The expression at the particular timepoint is shown by filled symbols, other time points are included for comparison (empty symbols). (E) Volcano plots show miRNAs differentially expressed between different cell stages. Those miRNAs chosen for further investigation (Table 4.2) are annotated in red (and marked bold in B, C, D).

naïve cells. The day 2 condition is predictably more similar to day 0 but remarkably has already noticeably changed its miRNA expression profile. To visualise these changes, volcano plots comparing confidence ($-\log_{10}$ p value) and the change in expression between time points ($\log_2$ fold change) are plotted in Figure 4.4E. miRNAs chosen for further investigation are marked in red. Of particular note are miRNA-20b-5p and miRNA-363-3p which are highly upregulated after exit of naivety and the miRNA302 family which have been shown in the literature to be key markers of (primed) human ES and iPS cells (Suh et al., 2004, Bar et al., 2008, Morin et al., 2008) and mouse EpiSCs (Jouneau et al., 2012).

Validation of the miRNAseq results was performed by qPCR using Life Technologies TaqMan stem-loop primers and normalising miRNA expression to the average of 5 endogenous control miRNAs (Supplementary Figure S7). This also allowed insight into the expression of specific miRNAs over the entire time course (Figure 4.5). The validated miRNAs are grouped into clusters (shared location) and families (similar sequence). miRNAs in the same cluster and family often have similar expression patterns. The validated miRNAs are grouped broadly into 4 categories: upregulated, transiently upregulated, unchanged/undetected expression and downregulated. miRNAs of interest are underlined and investigated further. A comparison of qPCR and sequencing data is performed in Supplementary Figure S8.

4.2.3. Gene set enrichment analysis

Many miRNAs cause only a minor difference individually but synergistically with other miRNAs can downregulate genes effectively (Xu et al., 2011). To uncover which genes are most heavily regulated by miRNAs during the exit from naïve pluripotency, a gene set enrichment analysis (GSEA) was performed by Dr Klas Hatje (Supplementary Method 7.2.1). Briefly, each gene was assigned all miRNAs which bind to its mRNA *in silico*. The gene set enrichment analysis then lists genes which have an enrichment of binding sites for miRNAs upregulated at day 2 of the differentiation compared to naïve cells. Due to the large amount of *in silico* binding partners of

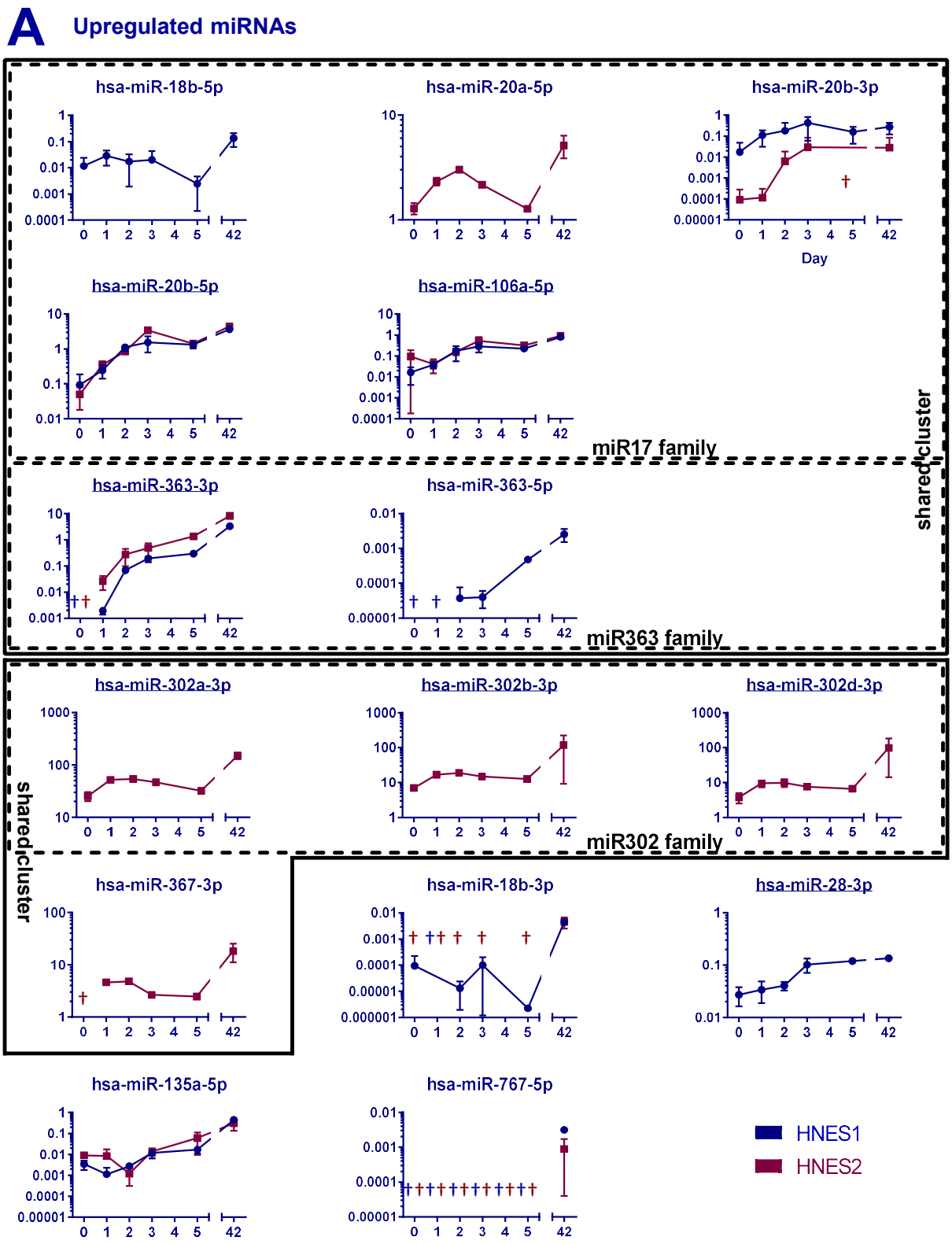


Figure 4.5 continued on next page

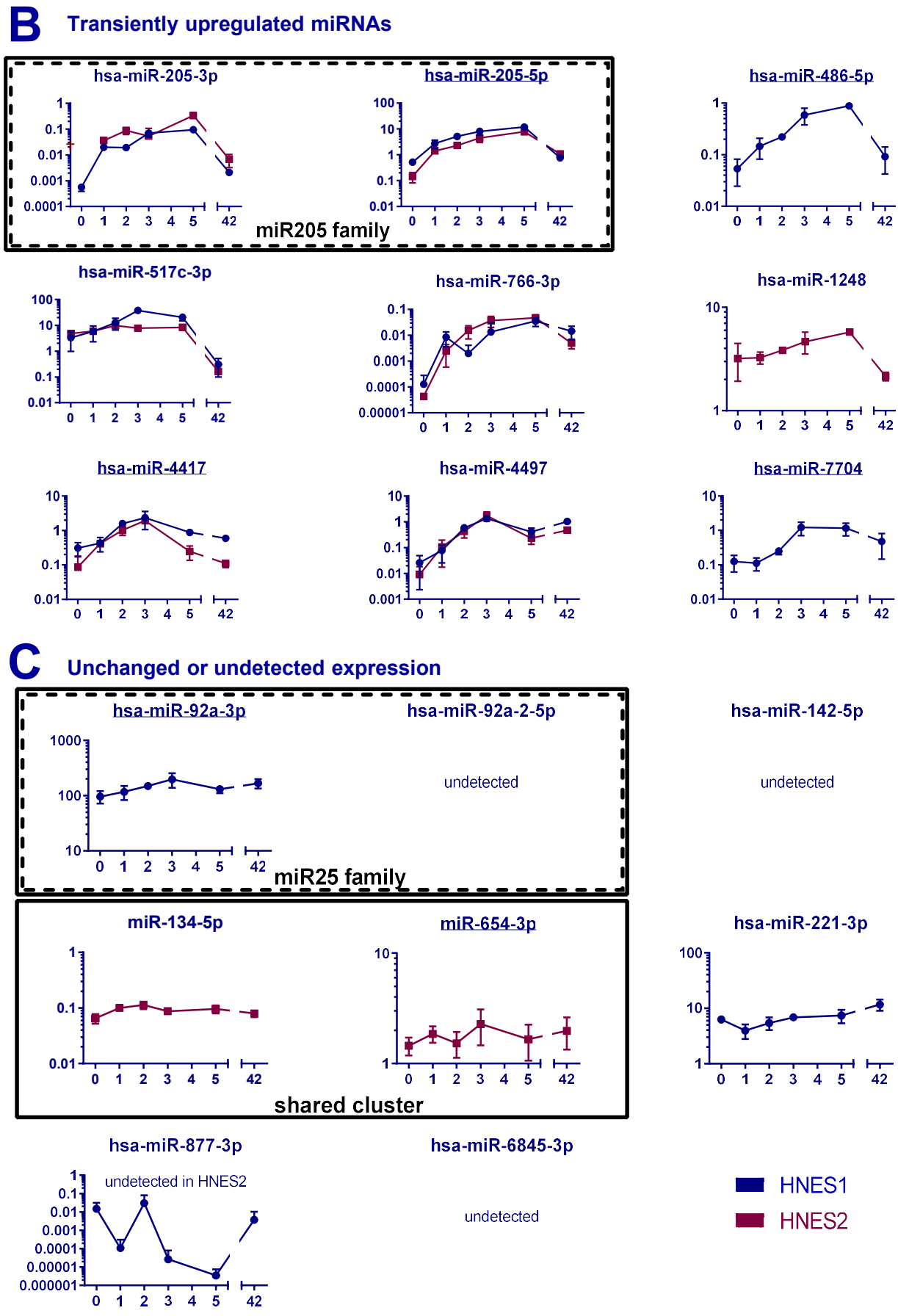


Figure 4.5 continued on next page

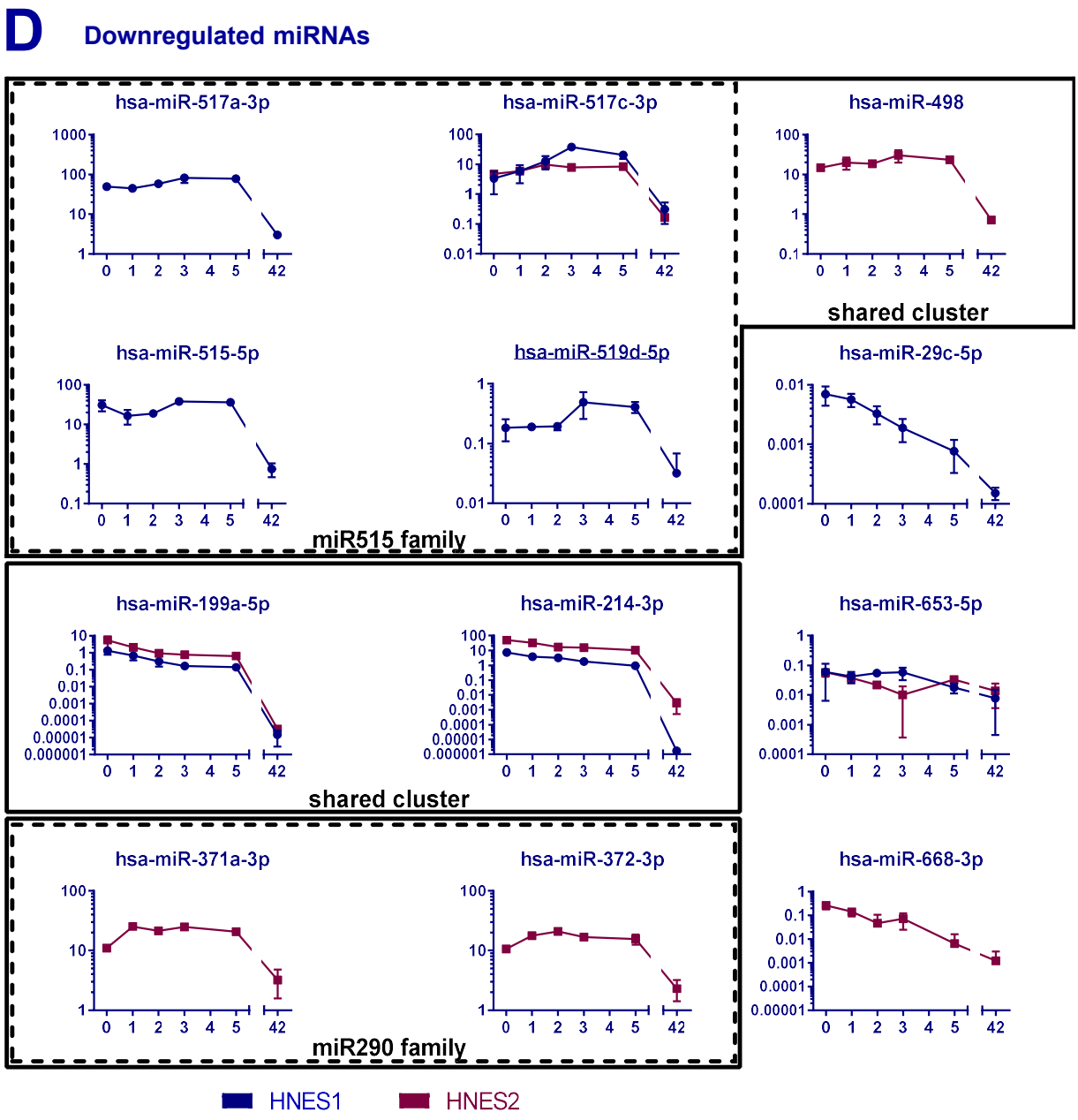


Figure 4.5: Validation of miRNA expression by qPCR. HNES1 samples are blue, HNES2 red. Data normalised to 5 stably expressed miRNAs (miR-26a-5p, miR-186-5b, miR-191-5p, miR-361-5p, miR-423-5p; Supplementary Figure S7) using the $2^{-\Delta ct}$ method. Either or both HNES1 and HNES2 lines were used for validation. Clusters are grouped in boxes, miRNA families are in dashed boxes. The dagger (†) indicates no detectable expression. Underlined miRNAs were chosen for further investigation (Table 4.2). Error bars are the standard deviation of 4 or 2 technical replicates. Results grouped into (A) Upregulated miRNAs, (B) transiently upregulated miRNAs, (C) unchanged or undetected miRNAs and (D) downregulated miRNAs. Comparison to miRNAseq data is performed in Supplementary Figure S8.

miRNAs and high frequency of false positives, the analysis is statistically not very powerful and the threshold for discovery is commonly set at a false discovery rate of 0.25 and below (Subramanian et al., 2005). In this analysis, the only gene which was below the discovery threshold was *TBX3* (Table 4.1). I have previously shown that *TBX3* is heavily downregulated during the first days of differentiation (Figure 4.1 & Figure 4.2), suggesting that miRNAs might be involved in dismantling the naïve pluripotency network.

4.2.4. Determining miRNAs of interest

To find miRNAs involved in naïve to primed differentiation, candidate miRNAs had to be determined to test functionally. Since in the mouse miRNAs are not necessary to maintain the naïve state but instead are required for exit from the naïve state, the focus was on miRNAs which are upregulated during the differentiation. Candidate miRNAs were determined according to the following inclusion/exclusion criteria:

1. Most highly upregulated miRNAs likely important for exit from naïve state
    - Top 10 miRNAs with >10 reads at D2 ordered according to fold change to naïve
2. Most highly upregulated miRNAs likely important for acquisition/maintenance of primed state
    - Top 10 miRNAs with >50 reads at D42 ordered according to fold change to naïve
3. Raw expression of miRNAs is important
    - Top 10 miRNAs with logFC >1 and >2 $-\log_{10}$p value ordered according to raw expression at day 2
4. TBX3 binding miRNAs might be important for dismantling the naïve pluripotency network
    - miRNAs predicted to bind TBX3 with either
        - >10 reads at D2, logFC >1 and >2 $-\log_{10}$p value
        - >50 reads at D42, logFC >1 and >2 $-\log_{10}$p value
5. miRNAs previously reported in the literature may be involved
    - Add key members of miR302 family

NAME	SIZE	NES	NOM p-val	FDR	Pass
TBX3_MIRNA BINDING_INHIBITION_UPSTREAM	14	1.771	1.47E-03	**0.242**	**Yes**
AKAP13_MIRNA BINDING_UNSPECIFIED_UPSTREAM	17	1.724	3.36E-03	0.334	No
ABCG4_MIRNA BINDING_INHIBITION_UPSTREAM	10	1.732	4.46E-04	0.345	No
EGLN3_MIRNA BINDING_UNSPECIFIED_UPSTREAM	12	1.634	1.68E-02	0.434	No
VAPA_MIRNA BINDING_INHIBITION_UPSTREAM	9	1.635	1.87E-02	0.436	No
MAP3K8_MIRNA BINDING_UNSPECIFIED_UPSTREAM	8	1.637	1.84E-02	0.439	No
FGD1_MIRNA BINDING_UNSPECIFIED_UPSTREAM	7	1.635	1.14E-02	0.443	No
HSPA1B_MIRNA BINDING_UNSPECIFIED_UPSTREAM	22	1.638	1.11E-02	0.444	No
ATP1A2_MIRNA BINDING_UNSPECIFIED_UPSTREAM	9	1.655	8.99E-03	0.450	No
TNRC6A_MIRNA BINDING_UNSPECIFIED_UPSTREAM	6	1.628	2.15E-03	0.451	No

Table 4.1: *TBX3* is the only target gene, which has an enriched for differentiation day 2 upregulated miRNAs. The amount of miRNAs which bind the target gene *in silico* are listed in "size". The nominal p-value (NOM p-val) estimates significance only for a single gene set. Unlike the p-value, the false discovery rate (FDR) is adjusted for gene set size and multiple hypothesis testing and is therefore a more relevant measure. The only gene to pass the false discovery rate of below 0.25 is TBX3. The normalised enrichment score (NES) accounts for differences in gene set size and in correlations between gene sets and the expression dataset, and is a measure of enrichment (Subramanian et al. 2005).

Due to overlaps between these lists, there are 19 miRNAs of interest. Strikingly, many of these miRNAs had similar seed sequences and were grouped into pools accordingly (Table 4.2).

4.2.5. Functional analysis of miRNAs

Two strategies were devised to test which miRNAs are functionally involved in naïve to primed differentiation: Measuring the effect of miRNA inhibition on the differentiation, or supplementing miRNA mimics to restore the ability of a *DGCR8* knockout to differentiate. The advantage of using mimics with a knockout is the clear readout, however a knockout is yet to be successfully established. Using inhibitors is slightly more physiologically relevant, but the readout potentially less clear. It is likely that exogenously adding miRNA mimics leads to an excess of the miRNA at unphysiologically high levels. Both strategies require functioning miRNA inhibitors or mimics.

To test mimics and inhibitors, validated mirVANA control mimics and inhibitors were electroporated into HNES1 cells and expression of their target genes were measured after 48h. The validated control mimic resembles miR-1 and effectively downregulates the miR-1 target *TWF1* in a dose-dependent manner 48h after electroporation (Figure 4.6A). However, the expression in the negative scrambled control mimic is higher than the no mimic control or no pulse control. A reason for this became clear when the *TWF1* expression over time was investigated: Although *TWF1* expression recovered back to the negative scrambled control level over 5 days, the expression of *TWF1* in the negative control decreased suggesting an inverse correlation of *TWF1* expression with confluency (Figure 4.6B). This inverse correlation may explain that due to less cells surviving the electroporation and slight toxicity of the scrambled control, confluency in these conditions is lower and the expression of *TWF1* is higher. Since *TWF1* expression is affected by confluence, use of this mimic as a positive control requires a carefully matched negative control with similar confluency. *TWF1* is an actin-binding protein which inhibits polymerisation (Rainger et al., 2017).

	average RPM			log$_2$ Fold Change			
Name	Naïve	Transition	Primed	Transition	Primed	Pool	miRNA sequence
hsa-miR-3178	26	73	70	1.32	1.41	1	GGGGCGCGGCCGGAUCG
hsa-miR-4417	32	405	249	3.61	2.97	1	GGUGGGCUUCCCGGAGGG
hsa-miR-4488	71	486	187	2.81	1.77	1	AGGGGGCGGGCUCCGGCG
hsa-miR-7704	147	566	417	1.81	1.44	1	CGGGGUCGGCGGCGACGUG
hsa-miR-302a-3p	2097	2387	8606	0.03	1.90	2	UAAGUGCUUCCAUGUUUUGGUGA
hsa-miR-302b-3p	3231	4601	9837	0.35	1.46	2	UAAGUGCUUCCAUGUUUUAGUAG
hsa-miR-302c-3p	567	621	4194	-0.03	2.76	2	UAAGUGCUUCCAUGUUUCAGUGG
hsa-miR-302d-3p	1977	2275	9190	0.06	2.09	2	UAAGUGCUUCCAUGUUUGAGUGU
hsa-miR-363-3p	2	48	977	4.43	8.81	3	AAUUGCACGGUAUCCAUCUGUA
hsa-miR-92a-3p	36918	52558	82754	0.35	1.03	3	UAUUGCACUUGUCCCGGCCUGU
hsa-miR-92b-3p	2814	2758	33231	-0.18	3.44	3	UAUUGCACUCGUCCCGGCCUCC
hsa-miR-106a-5p	1	31	247	5.35	8.40	4	AAAAGUGCUUACAGUGCAGGUAG
hsa-miR-20b-5p	2	78	745	5.17	8.45	4	CAAAGUGCUCAUAGUGCAGGUAG
hsa-miR-205-5p	65	698	405	3.29	2.54	5	UCCUUCAUUCCACCGGAGUCUG
hsa-miR-486-5p	2797	7843	2144	1.31	-0.62	5	UCCUGUACUGAGCUGCCCCGAG
hsa-miR-28-3p	331	1164	2414	1.62	2.78	6	CACUAGAUUGUGAGCUCCUGGA
hsa-miR-519d-5p	428	938	18	0.97	-4.71	6	CCUCCAAAGGGAAGCGCUUUCUGUU
hsa-miR-561-5p	15	54	176	1.69	3.38	6	AUCAAGGAUCUUAAACUUUGCC
hsa-miR-7974	140	261	411	1.08	2.47	6	AGGCUGUGAUGCUCUCCUGAGCCC

Table 4.2: miRNAs chosen for further testing have similar seed sequences. The miRNAs listed here were chosen due to their expression level and degree of upregulation during the differentiation process. Most chosen miRNAs have similar seed sequences and were grouped into 6 pools accordingly. The consensus sequences are highlighted in red and the seed sequence has a grey background.

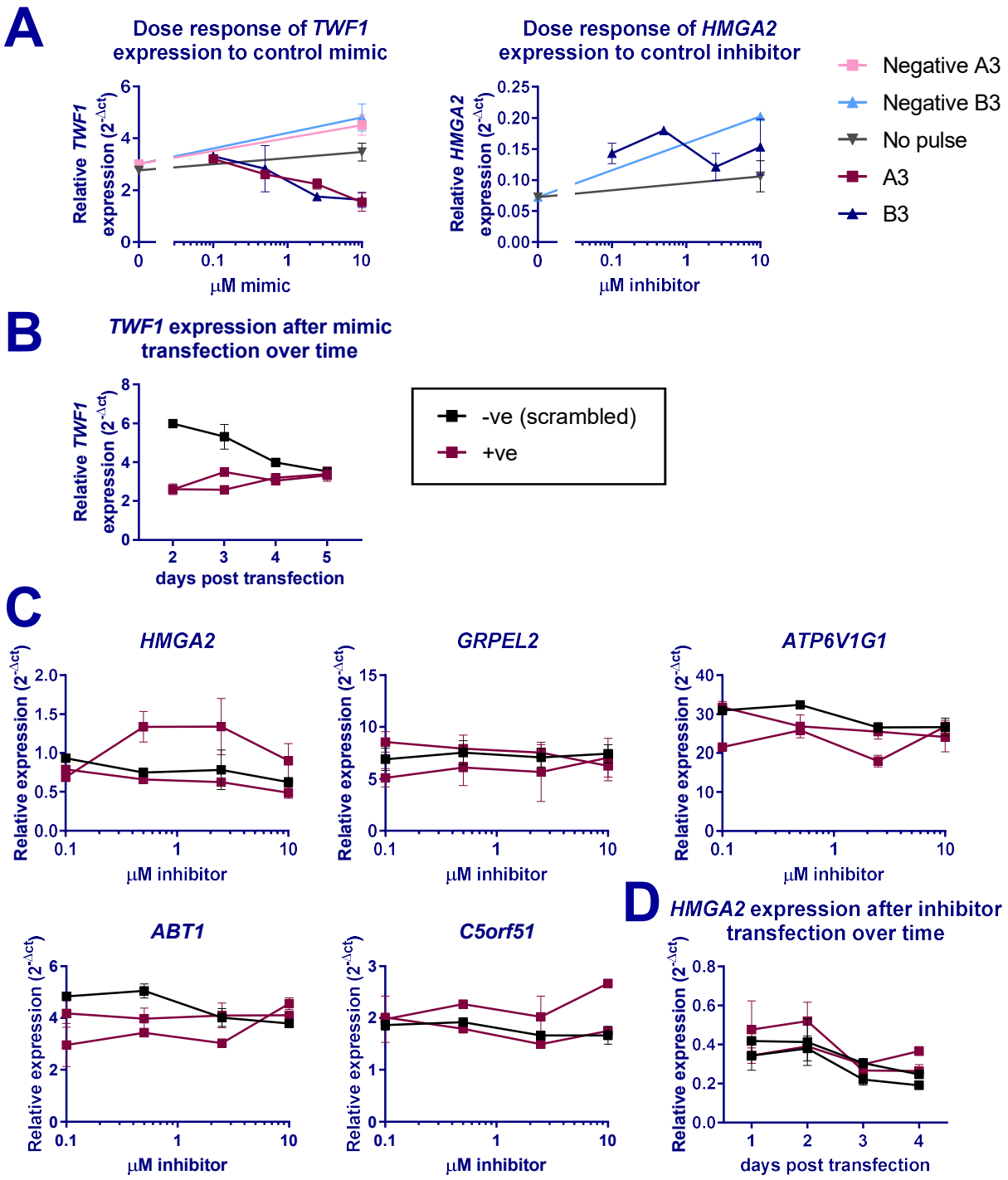


Figure 4.6: The control miRNA mimic but not inhibitor have a dose dependent relationship to their miRNA target gene. Scrambled mimics/inhibitors of the same chemistry are used as negative controls. (A) A range of concentrations of control mimic or inhibitor were electroporated into HNES1 cells and target gene expression was measured after 48h. Programme A3 (red) and B3 (blue) were used. Negative controls were 10µM scrambled mimic and a no mimic control as well as a no pulse control. The mimic of miR-1 downregulates its target gene *TWF1* in a dose-dependent manner. Inhibition of let-7c should lead to more of its target *HMGA2*, however *HMGA2* mRNA levels were not affected by the let-7c inhibitor. (B) 10 µM of miR-1 mimic was electroporated into HNES1 cells at day 0 in duplicates (programme A3) and samples were taken across a 5 day time course. The effect of the mimic on *TWF1* is lost over the course of 5 days. Basal expression of TWF1 seems linked to cell number and confluency. (C) An alternative LNA based inhibitor for let-7c was electroporated into naïve 4iLA SFC856-03-04 cells in duplicate (programme A3) but failed to increase the abundance of several let-7c target genes including *HMGA2*. Electroporation of 10 µM of the LNA at day 0 did not change *HMGA2* expression over a time course of 4 days.

The control inhibitor binds let-7c and should therefore prevent let-7c mediated downregulation of its target *HMGA2* (Shi et al., 2009). A dose-dependent increase of *HMGA2* was expected but was not observed (Figure 4.6A), despite a high abundance of let-7c determined by miRNAseq (4261 RPM). *HMGA2* itself is not highly expressed in naïve cells and may have too low turnover to see any effect. Alternatively, the proprietary mirVANA inhibitors may not be potent enough. An antisense phosphorothioate backbone locked nucleic acid (LNA) of let-7c was introduced by electroporation into naïve 4iLA SFC856-03-04 at a range of concentrations and after 48h the expression of several validated target genes was measured. The additional target genes (*GRPEL2, ATP6V1G1, ABT1, C5orf51*) have higher expression than *HMGA2* in naïve cells but an upregulation was not observed in any of the tested genes (Figure 4.6C). To avoid the possibility of missing the effect due to wrong timepoints, 10 μM of the LNA was electroporated into cells and no effect was observed on *HMGA2* across a time course of 4 days (Figure 4.6D).

No useable positive control miRNA inhibitor was able to be identified, despite best efforts. This might be due to a problem with the delivery, the inhibitor not functioning, the assay to measure its effect or the biology. Without a functioning positive control, only a clear positive result of a miRNA inhibitor preventing differentiation can be interpreted, since a negative result could be due to experimental settings. Despite this caveat, anti-miRNA oligonucleotides (AMOs) against miRNA candidates were grouped into pools according to their target miRNAs seed sequence (Table 4.2). In addition, a pool containing all 19 AMOs was tested. Each pool was electroporated into naïve 4iLA SFC180-01-01 at two concentrations (3 and 10 μM) and then each of these seeded into two wells. After 24h one well was maintained in naïve conditions, the other switched to priming E8 conditions. After 4 days of differentiation RNA was extracted and the naïve pluripotency network examined. The negative controls were no inhibitor and 10 μM of negative (scrambled) inhibitor. Comparing both the priming and naïve conditions of these controls confirmed that the differentiation had successfully begun, evidenced by the downregulation of

naïve markers (Figure 4.7A). Successful inhibition of miRNAs required for naïve to primed differentiation should cause a retention or an increase of naïve marker gene expression. The expression was normalised to the negative control of the appropriate condition (naïve/priming) thus a retention of naïve markers would be indicated by upregulation. No convincing upregulation is detected (Figure 4.7B). In fact, a slight dose-dependent downregulation of TBX3 is observed in priming conditions for all but pool 5. This may be an artefact from the experiment or may have different reasons, such as higher toxicity from increased inhibitor concentrations. Unfortunately, no conclusion can be drawn from this experiment since, due to the lack of a robust positive control, it cannot be concluded that these miRNAs are not involved in the differentiation.

4.2.6. Future perspective: knockout of *DGCR8*

The idea that miRNAs are required for the naïve to primed differentiation assumes that human naïve PSC resemble mouse naïve PSC with regards to their miRNA requirements. A way to confirm or refute this is by knocking out the gene *DGCR8*. The protein DGCR8 is a key component required for miRNA biogenesis. It is non-essential in naïve mouse PSC but required for exit from the naïve state. To the best of our knowledge, no *DGCR8* knockout has been generated in hPSC, and it has been shown to be an essential gene in primed hPSC (Yilmaz et al., 2018). The assumption that miRNAs are not required in the naïve state can be challenged by failure to generate naïve *DGCR8* knockout hPSC.

To generate a *DGCR8* KO, the essential exon 3 of *DGCR8* was cut out using CRISPR/Cas9 targeted to both flanking introns. All known splice variants include this exon, and removal causes a frame shift and early stop codons in the following exon (Figure 4.8). Guide RNAs (gRNAs) were designed using the CCTop software (Stemmer et al., 2015, Labuhn et al., 2018). gRNAs were chosen which were central in the introns to avoid removal of any region important for splicing. To prevent off-target effects, the gRNAs were chosen to also have least 3 mismatches in the closest off-target sites (GRCh38/hg38) except for gRNA right 2 which has one predicted off target site with 2

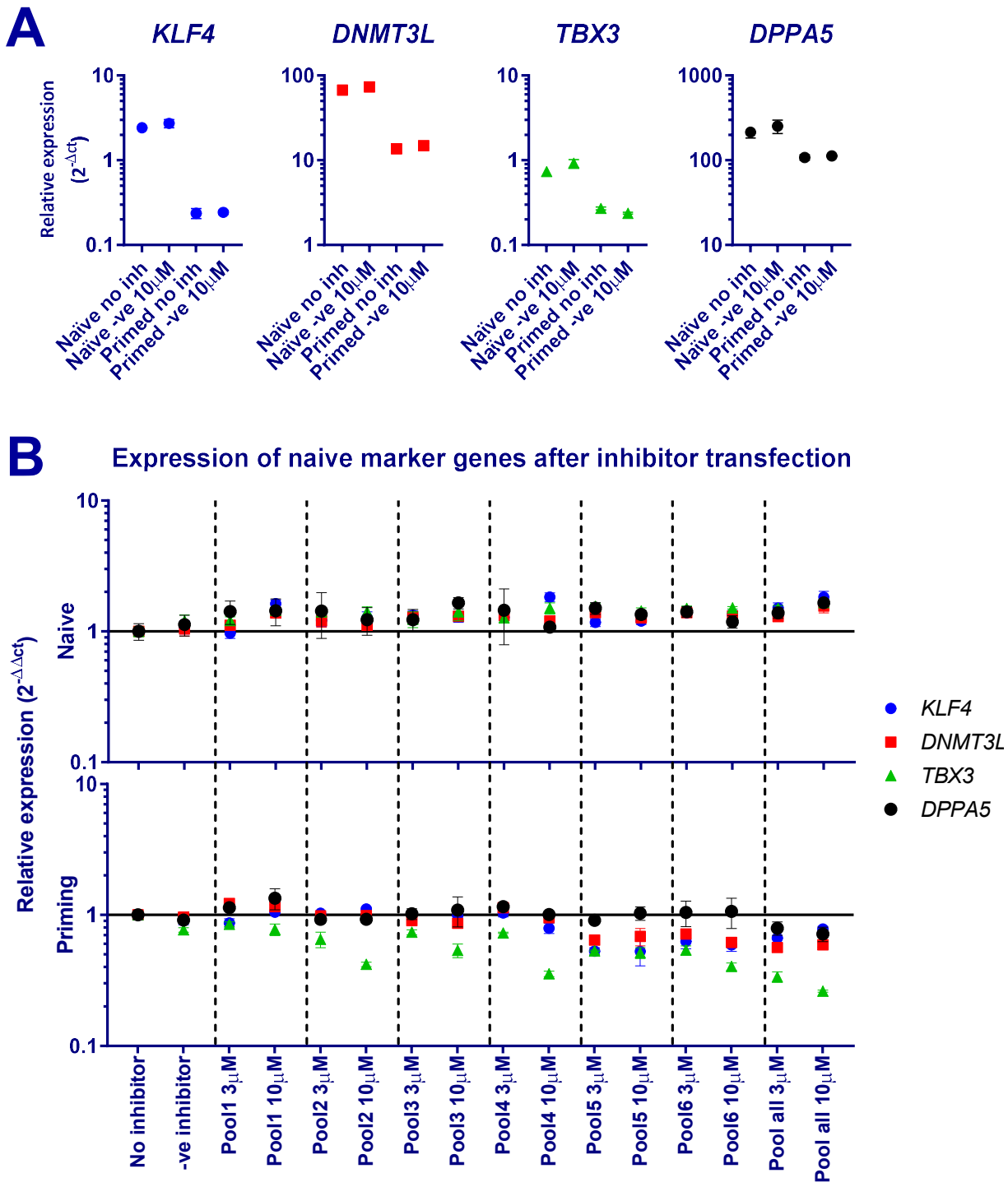


Figure 4.7: Inhibition of miRNAs selected due to their upregulation during differentiation does not affect the differentiation gene expression pattern in these conditions. The average and standard deviation of 3 technical replicates is shown per datapoint. (A) Expression level of naïve marker genes of control conditions either kept in naïve or transferred to priming conditions are shown. The negative control inhibitor does not affect the expression of naïve marker genes and the differentiation itself has been successful, evidenced by their downregulation. (B) Change in gene expression of cells transfected with pools of miRNA inhibitors compared to the no inhibitor control. An upregulation of naïve genes in comparison to the negative control would be expected if the miRNAs inhibited are functionally involved in the exit from naïve pluripotency. However, the pools of inhibitors (Table 4.2) do not affect the expression of the naïve pluripotency network. Due to the lack of a positive control, no conclusion can be drawn. Data normalised to endogenous controls *B2M* and *18s*.

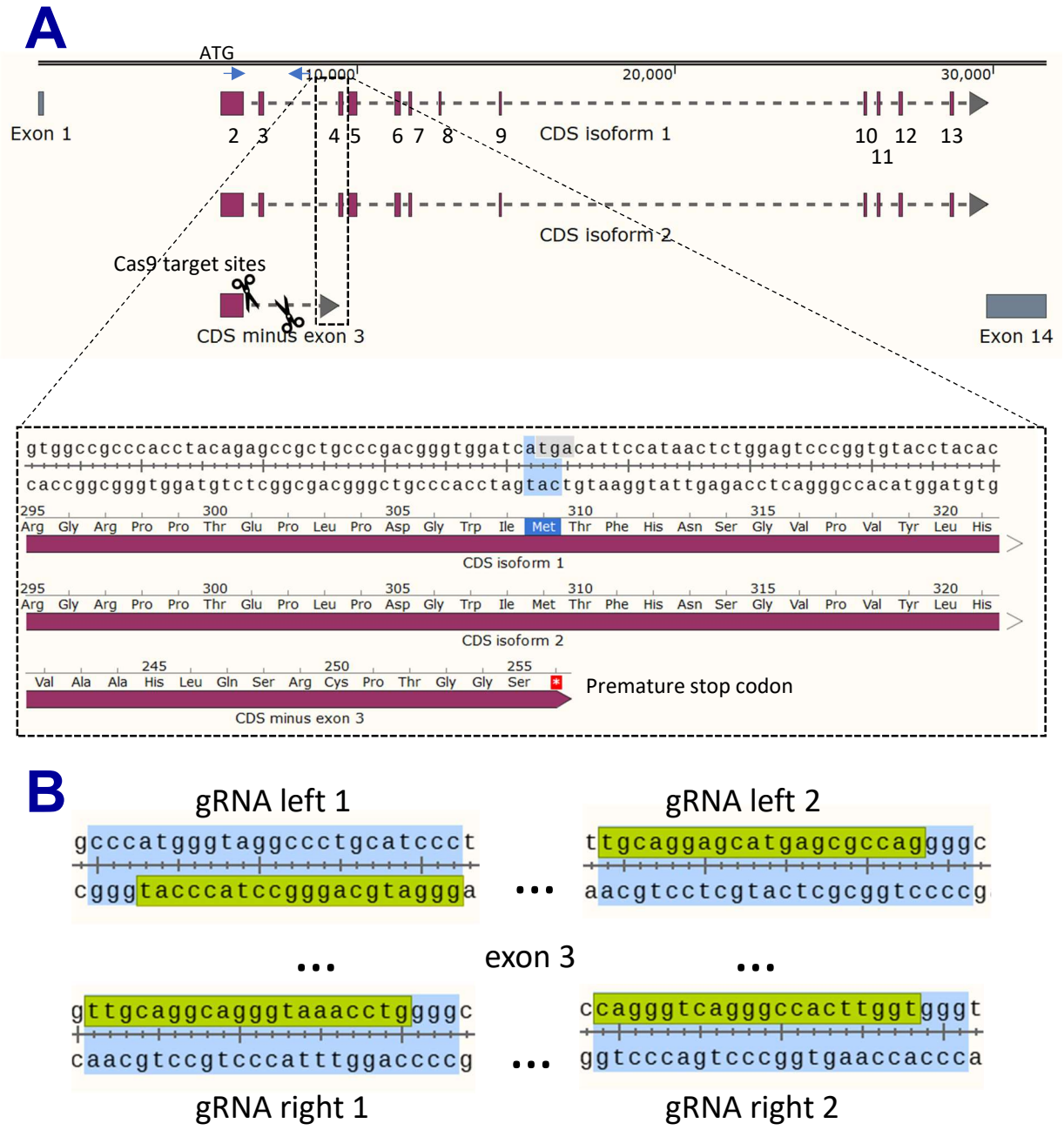


Figure 4.8: *DGCR8* knockout strategy. (A) Two different isoforms of *DGCR8* are expressed and their genomic origin are shown here. Removal of exon 3 causes a frameshift and premature stop codons. Cas9 cut sites to the left and right of exon 3 are marked by scissors and the primer binding sites for BD131 and BD132 are marked as arrows. Exon 3 was chosen for deletion since it is not the first coding exon (which, when deleted, can lead to alternative start sites and functional truncated proteins). Loss of exon 3 leads to a frame shift and premature stop codons, likely causing nonsense mediated decay. (B) 4 different guide RNAs were designed using the CCTop Software (Stemmer *et al.*, 2015) to avoid off-target effects. 2 guides were designed on the left and 2 on the right side of exon 3. Primer and gRNA sequences are listed in Table 2.2 and Table 2.5.

mismatches in an intergenic region (chr5:179387434-179387456). In addition, the regions were sequenced and any gRNAs containing variable SNPs were excluded. The gRNAs (complexed with Cas9 protein and tracrRNA) were transfected into both primed and naïve SFC856-03-04 cells. The method to knock out genes in primed cells is well established and out of 74 surviving clones, 26 heterozygotes were generated using primed SFC856-03-04 cells. However, since naïve hPSC behave very differently, only 23 naïve clones survived (10 heterozygotes) and no homozygous knockout was established (Figure 4.9). The interpretation is either that *DGCR8* knockout cells have a substantial growth disadvantage, or that it is in fact an essential gene even in the naïve state.

To get a useful readout even if *DGCR8* is essential, a lentiviral vector was designed containing a TetON inducible *RFP-2A-DGCR8*. This allows ectopic expression of DGCR8 during the knockout procedure, alleviating any growth disadvantage and allowing knockout (KO) cells to proliferate even if the gene is essential. The RFP confirms successful transduction and expression. A 2A sequence causes skipping of a peptide bond during translation yielding two separate proteins (Kim et al., 2011). After clonal selection of KO cells, doxycycline (activating the inducible TetON promoter) can be withdrawn and the effect on cell survival measured. The absence of RFP is an important indicator ensuring minimal leaky transgene expression. If KO hPSC survive, the effects on the differentiation can be measured and miRNA mimics can be supplemented to try to reverse the phenotype. If DGCR8 is essential even in the naïve state, an important species-species difference between mouse and human has been uncovered and the relevance of the miRNAome of the naïve state emphasised. This increases the impact of our miRNAseq data and profiling of naïve and primed states. The vector has been generated, by cloning a gBlock containing the *2A-DGCR8* into an existing lentiviral backbone containing the inducible promoter, *RFP* and all required auxiliary components (Supplementary Figure S1).

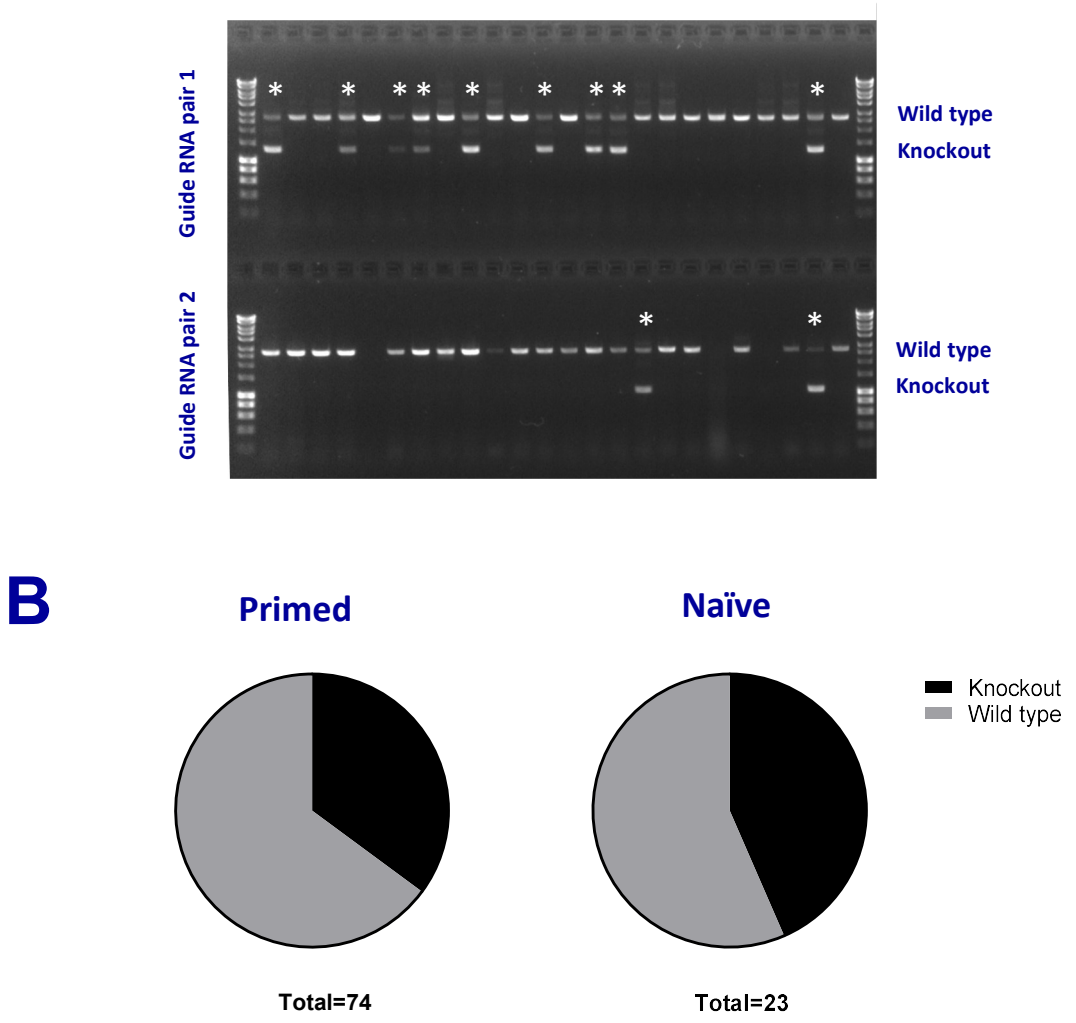


Figure 4.9: Analysis of *DGCR8* heterozygotes. (A) The guide RNA pair 1 generated many more heterozygous knockouts than guide RNA pair 2. Pair 1 is therefore used for all further experiments. Heterozygous deletions are marked by an asterisk (*). (B) No homozygous knockouts were generated but would be expected if no growth disadvantage is associated with the knockout.

4.3. Summary and discussion

4.3.1. Naïve to primed differentiation

The naïve to primed differentiation has been used as a model in the mouse before (Pandolfini et al., 2016, Wu and Izpisua Belmonte, 2015, Buecker et al., 2014, Factor et al., 2014, Zhou et al., 2012a) but not in the human system. Here, we saw that in the human system the naïve marker genes and cell surface markers are downregulated with different kinetics across a time course of differentiation. This is unlikely due to the addition of primed medium (containing e.g. bFGF) but instead the release from inhibitors and growth factors, since simply removing inhibitors and LIF from the medium was enough to induce differentiation (Figure 4.2).

4.3.2. ESCC miRNAs with AAGUGC seed

For the first time the human miRNAome in naïve, differentiating and primed cells was analysed (summarised in Figure 4.10). This novel dataset has revealed that the miRNAome is very different between the 3 stages. Sufficiently so, that there is a large separation on the principal component analysis (Figure 4.4) indicating that the expression of miRNAs alone can distinguish the 3 states. Many miRNAs from the ESC-specific cell cycle-regulating (ESCC) family containing the sequence AAGUGC (Supplementary Figure S9) are highly enriched, comprising 15% of the total miRNAome in naïve, 21% in differentiating conditions and 17% primed. Just like in the mouse (Greve et al., 2013) we see a switch of AAGUGC miRNA expression. The miR371-373 cluster (the human homolog of the miR290 family in the mouse) dominate expression in the naïve state, whereas the miR302 family become dominant in primed cells (Supplementary Figure S9A). This explains a phenomenon which has puzzled scientists studying the evolution of miRNAs (Wu et al., 2014a). The miR371-373 cluster is indeed heavily expressed – just not in the primed hPSC previously assessed by small RNA sequencing. Further validation of our miRNAseq results come from a single-cell small-RNA sequencing experiment using Theunissen et al. 5iLAF cells (Faridani et al., 2016)(Supplementary Figure S9B). Expression of the miR-371-373 cluster becomes variable

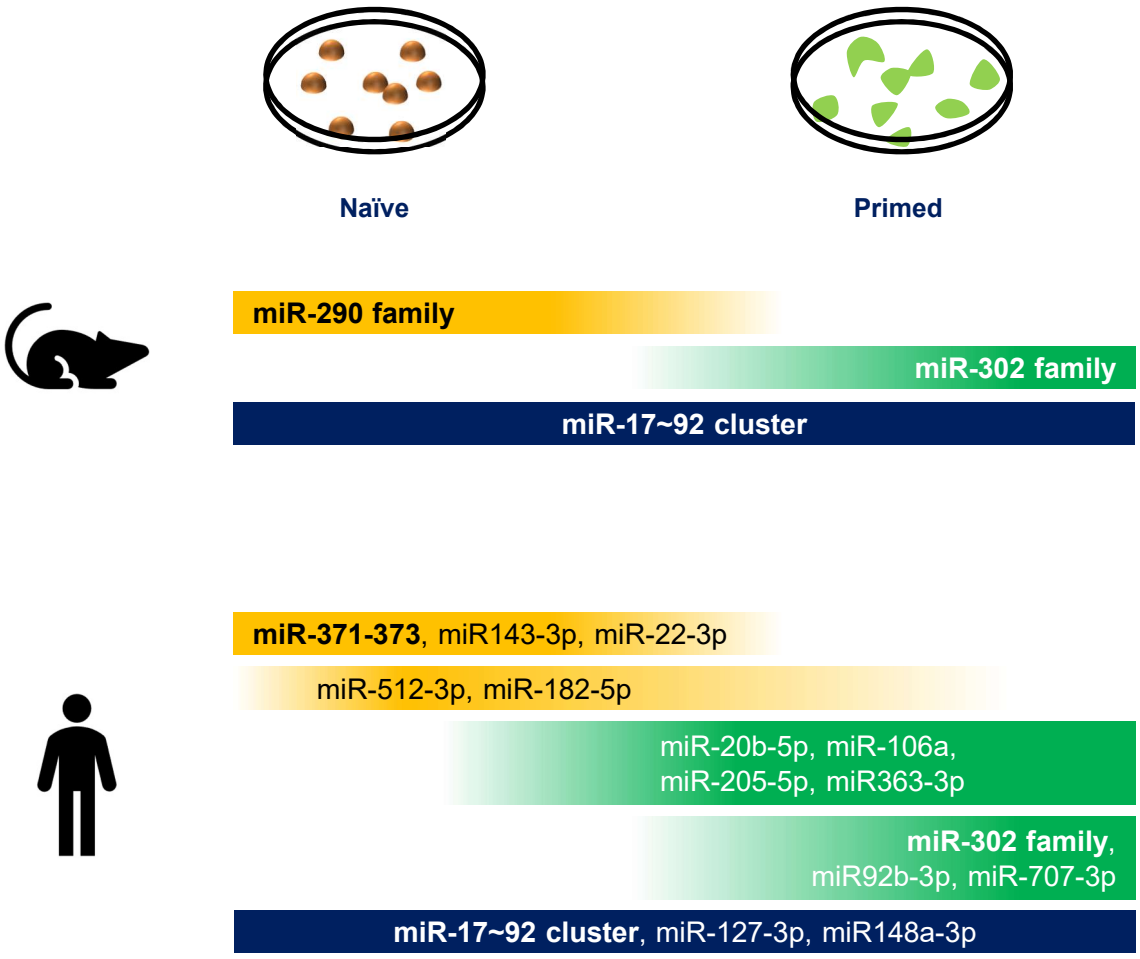


Figure 4.10: Summary of miRNAs expressed in naïve and primed cells. Known markers of naïve and primed cells from mouse studies are also active in the human system and are written in bold. New miRNA markers of naïve and primed states uncovered in this thesis are also summarised here.

between lines in the primed state (with HNES1 showing high expression of 4-6% contrasting to HNES2 showing only 0.7-1%). This variability in expression of the miR-371-373 cluster has been previously reported for conventional primed ES cell lines in the literature (Wu et al., 2014a) with some publications implying no expression (Lipchina et al., 2011, Bar et al., 2008). Our data confirms that the cluster is variable in primed cells but is universally highly expressed in naïve. Notably, high expression of miR-372-3p has been linked to cancer (Dieckmann et al., 2017). The other highly expressed AAGUGC containing miRNA is miR-512-3p with very high expression in the differentiating state and low expression in the primed and has not been linked to development but is reportedly a tumour suppressor (Chen, 2010, Zhu et al., 2015).

4.3.3. miRNAs in the naïve state

Apart from miRNAs from the miR-371-373 cluster and miR-512-3p discussed above, several other miRNAs were highly expressed in naïve cells but not in primed: miR-143-3p which has been found to be a tumour suppressor and is often expressed in mesenchymal like cells (Kent et al., 2014). The highly expressed miR-22-3p is an oncomiR which inhibits apoptosis and regulates proliferation (Lv et al., 2018) but has not been linked to development. miR-182-5p is highly expressed in the naïve state and at day 2 of the differentiation and is an oncomiR which is upregulated in many cancers (Krishnan et al., 2012, Hirata et al., 2012, Hirata et al., 2013). Intriguingly, it has been shown to target the homologous recombination pathway (Krishnan et al., 2012). OncomiRs can act via the cell cycle, allowing faster growth by inhibiting check points and affecting HDR (Yu et al., 2017). In addition, oncomiRs may also inhibit differentiation, as de-differentiation is a common feature of tumorigenesis (Chaudhuri et al., 2012). Most miRNAs from the let-7 family, which reportedly antagonise pluripotency (Viswanathan et al., 2008, Newman et al., 2008, Rybak et al., 2008, Melton et al., 2010) were more highly expressed in naive cells (cumulatively approximately 25,000 RPM) and expression in primed cells was significantly lower (590 RPM). This may comprise a species-species difference since the let7 family has been observed to accumulate during

differentiation and is more highly expressed in mouse primed EpiSC than ES cells (Viswanathan et al., 2008, Jouneau et al., 2012).

4.3.4. Differentiating and primed state miRNAs

Possible markers for the primed state are the ESCC miR-302 family explained above but also miR-92a-3p. It is more heavily expressed in absolute terms, although this miRNA is also expressed highly in naïve cells. The miR-92b-3p miRNA on the other hand is only heavily expressed in primed cells. There are several miRNAs which are not amongst the most highly expressed but are very differentially expressed between states (Figure 4.4E). For example, the novel miRNA miR-707-3p may be a marker for the primed state. It is expressed in both cell lines at low level (154 RPM) but was not at all detected at day 0 or day 2. No proposed function has been reported at all in the literature. The miRNAs miR-20b-5p and miR-106a-5p (both members of the miR-17 family and in the ESCC cluster) are highly upregulated compared to the naïve state even at day 2 and their seed sequence is very similar to the 302 and 371-373 families. Both have the seed sequence "AAAGUG(C…)" which is just shifted from 302/290's seed "AAGUGC". miR-20b-5p has been linked to myoblast differentiation and repressed myoblast proliferation (Luo et al., 2016), whereas miR-106a-5p has been shown to be a tumour suppressor (He et al., 2016). Another highly upregulated miRNA at day 2 is miR-363-3p which is a tumour suppressor by inhibiting epithelial to mesenchymal transition (Hu et al., 2016) and miR-205-5p, which is reportedly also a tumour suppressor (Yamada et al., 2017).

4.3.5. Unchanged and highly expressed miRNAs

Some miRNAs are highly expressed in all states: The miR-17~92 cluster has been linked to embryogenesis (Concepcion et al., 2012, Mendell, 2008) and upregulation of the cluster to cancer (Hayashita et al., 2005, Lu et al., 2007, Olive et al., 2010). Members of this cluster are miR-17, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR92a. The miR-92a is the most highly expressed miRNA in the primed state and the third highest in naïve, corroborating previous publications indicating

that it is likely an important regulator in embryogenesis (Concepcion et al., 2012, Mendell, 2008). Dysregulation of this cluster leads to developmental defects such as Feingold syndrome (Mogilyansky and Rigoutsos, 2013). Finally, miRNAs which are highly expressed in all states include miR-127-3p, a tumour suppressor (Chen et al., 2013) and miR-148a-3p also tumour suppressor acting via DNMT1 (Wang et al., 2016).

4.3.6. Expected but not expressed miRNAs

The miRNAs miR-145 which was reported to downregulate OCT4, SOX2, KLF4 in the human (Xu et al., 2009, Cordes et al., 2009) was expressed at low levels in naïve cells (50-80 RPM) and not expressed at all in primed cells. Although this is in line with previous observations that it is only highly upregulated when differentiating from the human primed state (Xu et al., 2009), it does contrast with another publication which suggests that it is co-expressed with the highly expressed miR-143 (Cordes et al., 2009).

4.3.7. miRNA inhibition experiment

The approach to inhibit miRNAs and measure the effect on the differentiation should allow insights into which miRNAs are functionally involved in the differentiation process, providing the correct controls are in place. Several types of inhibitors of miRNAs were tested before the experiment and none of the control inhibitors worked. It was unclear if this was due to the control inhibitor, the assay or the biology.

Possible reasons for the control inhibitors not working are plentiful: The delivery or the inhibitor itself may not be functional. Alternatively, it could be that the target let7c, which is expressed at intermediate levels in HNES cells may not be expressed in 4iLA cells. The GSE81287 dataset (Faridani et al., 2016), which is a single cell miRNAseq dataset of 5iLAF and primed hPSC, only includes the most highly abundant miRNAs and did not detect let7c. It is unclear if it is not expressed at all or if the single cell miRNAseq is the limitation. Measuring the expression by other means in 4iLA cells, for example by qPCR would give insight. It is also possible that the assay and

the biology do not overlap. The assay was devised to detect a difference in target mRNA levels since the canonical pathway for miRNA mediated repression is by destabilisation of the mRNA by deadenylation (Guo et al., 2010, Hendrickson et al., 2009, Baek et al., 2008, Bazzini et al., 2012, Wu et al., 2006, Eulalio et al., 2009, Djuranovic et al., 2012, Eichhorn et al., 2014). This pathway explains 66% to >90% of the steady-state repression mediated by mammalian miRNAs (Eichhorn et al., 2014) with repression of translation usually being a minor but measurable component. However, research on zebrafish and *C. Elegans* showed that miRNAs may act primarily by translational inhibition in embryonic cells and by mRNA destabilisation in post-embryonic cells (Bazzini et al., 2012, Subtelny et al., 2014). Deadenylation of mRNAs in early zebrafish embryos causes dramatic translational inhibition but no changes in mRNA stability, whereas post-embryonic cells act in exactly the opposite fashion (Subtelny et al., 2014). This has not been reported in conventional pluripotent stem cells (mouse and human) but naïve human PSC have yet to be studied in this context. It is therefore possible, that the RT-qPCR based assay is unsuitable and protein levels must be measured to detect the effect of miRNA inhibitors. Another possible explanation is that the miRNA biogenesis pathway may be less active during pre-implantation (Garcia-Lopez and del Mazo, 2012, Ma et al., 2010) and as *DGCR8* knockouts show, not required in mouse ES cells (Wang et al., 2007). However, an argument against both hypotheses is that a change in mRNA expression level has been observed for miRNA mimics, indicating that the pathways are present and functional for ectopic miRNAs (Figure 4.6). However, the miRNA mimic used in Figure 4.6 was added in unphysiologically high levels.

Unless the delivery, or the inhibitor itself, does not work, the inhibition of key miRNAs may have yielded positive results. This is because the pluripotency network consists of transcription factors. A miRNA induced difference on this network – also if purely translational – should cause an effect on the expression of naïve specific mRNAs. This approach has been used in the mouse system successfully before, albeit using *DGCR8* knockouts and mimic supplementation (Gu et al., 2016). Therefore, pools of miRNAs inhibitors were tested on differentiating naïve hPSC, despite no

functioning positive control (Figure 4.7). Unfortunately, none of the inhibitors showed a retention of naïve marker gene expression. Since, despite best efforts, no positive control miRNA inhibitor was found, this result does not allow any further conclusions.

An alternative giving clearer results would be supplementation of mimics to rescue the ability to differentiate in *DGCR8* KO cells. In this approach, the miR-371-373 cluster would also be included, due to the slight upregulation at day 2 of the differentiation but also the involvement of the mouse miR-290 homolog in the differentiation (Gu et al., 2016).

4.3.8. *DGCR8* knockout

DGCR8 or *DICER1* knockouts in mESC are unable to differentiate to the primed state, suggesting that miRNAs are essential for this differentiation (Wang et al., 2007, Kanellopoulou et al., 2005). However, a knockout has not been performed in the human system. To investigate if miRNAs are also not required for the naïve state but essential for the differentiation in the human, a knockout of *DGCR8* was attempted. Despite generating many heterozygous knockouts, no homozygotes were found. Numerically, homozygotes would have been expected if knockouts have no disadvantage. Therefore, and since DGCR8 may be essential, a lentiviral DGCR8 supplementation construct has been designed and cloned. This allows exogenous supplementation of DGCR8 during the knockout procedure and withdrawal of DGCR8 for further experiments. Inability of the knockout to differentiate is evidence for the involvement of miRNAs in this process. Rescue experiments using miRNA mimics give a clear readout and mimics (unlike inhibitors) work well in naïve hPSC (Figure 3.6C).

4.3.9. Summary

In this chapter the naïve to primed differentiation was successfully performed in human PSC. A large miRNA sequencing experiment was completed across this differentiation time course and uncovered that miRNA expression signatures alone are enough to distinguish human naïve, differentiating and primed states. This dataset also uncovered that the human miR-371-373

cluster is highly expressed in naïve cells and therefore acts similarly to the mouse homolog, the miR-290 family. This suggests that mouse and human preimplantation development are more similar in their miRNA expression profile than previously assumed. Validation of the dataset was performed for many miRNAs by qPCR and uncovering functional roles for miRNAs was attempted using miRNA inhibitors. Finally, the need for a homozygous *DGCR8* knockout was discussed and the lentiviral vectors required for ectopic supplementation of DGCR8 designed and cloned.

Intriguingly, miRNA miR-182-5p is highly expressed in the naïve state and at day 2 of the differentiation and has been shown to target the HR pathway (Krishnan et al., 2012). Differences in activity of the HR pathway will be discussed in chapter 5.

5
Gene editing in naïve hPSC

5.1. Introduction

Naïve mESC can be more readily genetically manipulated by HR than mEpiSC or conventional, primed human PSC (Zwaka and Thomson, 2003). On the other hand, the miRNA miR-182-5p is highly expressed in the human naïve state as shown in chapter 4, and has been shown to target the homologous recombination pathway (Krishnan et al., 2012). It is therefore unclear, whether the HR pathway is likely to be more or less active in naïve hPSC.

5.1.1. Gene editing in naïve hPSC

To test genome editing efficiency in naïve hPSC, Buecker *et al.* first generated human naïve-like cells by reprogramming fibroblasts using stable transgene expression of *OCT4, SOX2, KLF4* and *cMyc* (Buecker et al., 2010) and then measured the efficiency of random insertion of 10-20kb cassettes containing a fluorescent marker and a drug resistance gene. They reported a 200-fold increase in insertion frequency in their naïve-like cells compared to primed cells. However, they measured the insertion frequency by simply counting how many fluorescent colonies survived selection and the experiment lacked sufficient controls to take into account other factors affecting insertion frequency, such as how readily DNA is taken up by the cells and their ability to survive both the electroporation procedure and as single cells. A simple no-template control would have provided a baseline survival measure but was not included. This would have been particularly important since the efficacy of the approach they used was likely biased towards mESC owing to

their research background, therefore yielding better results in naïve over primed human cells. In a separate experiment the same researchers targeted the *HPRT* gene in a nuclease-free approach using a puromycin cassette containing 4-4.5 kb homology arms. Of the puromycin selected cells, they reported HDR targeting rates as high as 1%, though they did not provide a comparison to primed cells (Buecker et al., 2010). In conclusion, the authors showed that their gene editing protocols function more efficiently with naïve cells but failed to address why this might be, in part due to missing controls.

Gafni *et al.* measured rates of correct insertion in the two endogenous loci *COL1A* and *OCT4*, again without aid of a nuclease. These loci were targeted with puromycin selection cassettes containing homology arms of 2.1-2.5 kb and 4-4.5 kb in length, respectively. After puromycin selection, surviving clones were screened for correctly inserted targeting cassettes. The authors reported very high on-target integration rates of 11-14.5% in their naïve cells, compared to very low rates in primed cells (0-0.3%). This result suggests either higher rates of random integration in primed cells or indeed improved HR mediated gene editing in naive. It could also be due to different sensitivities to puromycin - the dose used was not specified and two very different cell types were compared (Gafni et al., 2013). In a third paper a proof of principle that Cas9 mediated gene editing works in naïve hPSC was performed but no assessment of efficiency or comparison to primed cells was made (Jacobs et al., 2017).

Publications so far that address whether naïve hPSC may be more readily genetically manipulated, do not have adequate controls and do not directly explore whether the rates of HR are different between these cell types.

5.1.2. Droplet digital PCR for measuring rates of NHEJ and HR

Miyaoka *et al.* developed a droplet digital PCR assay which accurately measures whether a nuclease induced DSB is repaired using a template strand via HDR or with an indel formation caused by C-NHEJ (or MMEJ) (Miyaoka et al., 2016). It was designed for use with TALENs and

plasmid-based Cas9 delivery. Since this publication, higher gene editing efficiencies have been made possible using Cas9 protein instead of plasmids (Jacobi et al., 2017). In this chapter, Miyaoka *et al*.'s assay will be modified to work with Cas9 protein and will be used to measure repair pathway frequencies in naïve and primed hPSC.

5.1.3. Aims

The aims for this chapter are:

- Adapt established HDR/NHEJ measuring methods to function with Cas9 protein and naïve and primed hPSC (5.2.1)
- Determine whether rates of HDR differ between naïve and primed hPSC (5.2.3)
- If HDR rates do indeed differ between the cell types, investigate the mechanistic basis (5.2.4)

5.2. Results

5.2.1. HDR and NHEJ can be measured by droplet digital PCR

Miyaoka *et al*. developed a droplet digital PCR based assay to measure both HDR and NHEJ at the same time (Miyaoka et al., 2016). This was achieved by introducing a double strand break and observing if a template oligo was used for repair by HR or if indels were formed. Indels can be formed by either C-NHEJ or MMEJ, although C-NHEJ is the predominant pathway. The assay works by transfecting cells with a nuclease and a template oligo containing a single base pair change. After allowing sufficient time for the turnover of the nuclease and all breaks to be repaired, genomic DNA is extracted and analysed by droplet digital PCR (ddPCR). In ddPCR, the PCR reaction mix is segregated into 20,000 oil stabilised droplets in which individual PCR reactions occur. The genomic DNA in the PCR reaction mix is diluted such that each droplet on average contains less than one allele, thus statistically ensuring no more than a single target for amplification in each. If the locus is present in a droplet it is amplified, and probes are hydrolysed and activated in the process. If the droplet contains a wild type sequence, both the reference probe (FAM) close to the

cut site and the NHEJ probe spanning the cut site (HEX) will bind (Figure 5.1A). This leads to a double positive FAM+HEX droplet. If, however, the double strand break yielded an indel, the NHEJ probe is unable to bind leading to a single FAM positive population. Finally, if the template was used for repair and a single nucleotide change is introduced, the HDR probe (FAM) will be able to bind. This leads to a FAM++ droplet with higher signal amplitude. To prevent false positives, the HDR probes are either competing with a dark probe (oligos with 3' phosphate to avoid priming) which contains the wild type sequence or directly against the NHEJ probe (Supplementary Figure S10). Primers are designed to bind outside the region containing the oligo to ensure that any contaminating template strand does not interfere with the assay. When measuring events in tens of thousands of droplets, frequencies of HDR and NHEJ can be accurately calculated. Mock data is shown in Figure 5.1B for illustration. Wild type cells exclusively have double negative or double positive droplets. If Cas9 and the guide RNA are added, a single positive population appears. Finally, addition of the template strand gives rise to the FAM++ positive population. To increase efficiency in the experiment, HiFi Cas9 protein replaced the plasmid based TALENs/Cas9 used in the original publication. The HiFi Cas9 differs from the wild type by a R691A mutation. This has been shown to reduce off-target effects, while maintaining high on-target efficiency. The reduction in off-target effects is likely due to inhibiting the transition from the inactive to active state in the presence of mismatches between guide RNA and target DNA (Vakulskas et al., 2018).

HDR and NHEJ were detected at GRN and RBM20 loci (Supplementary Figure S11). Though NHEJ was observed at the ATP7B locus, no HDR was detected. The main difference between RBM20/GRN and ATP7B is the design of their respective probes. Unlike RBM20 and GRN, the Cas9 cut site in ATP7B is within the HDR probe binding site (Supplementary Figure S10). Repair by HDR introduces only a 1 bp change which may be insufficient to prevent Cas9 from cutting. This is supported by evidence discussed in the next section, where Cas9 was observed to be active for up to 48h, though no further HDR was observed after 24h (Figure 5.2). Efforts were made to improve HDR efficiency by PTO modifying the template strand to increase its half-life, but still no

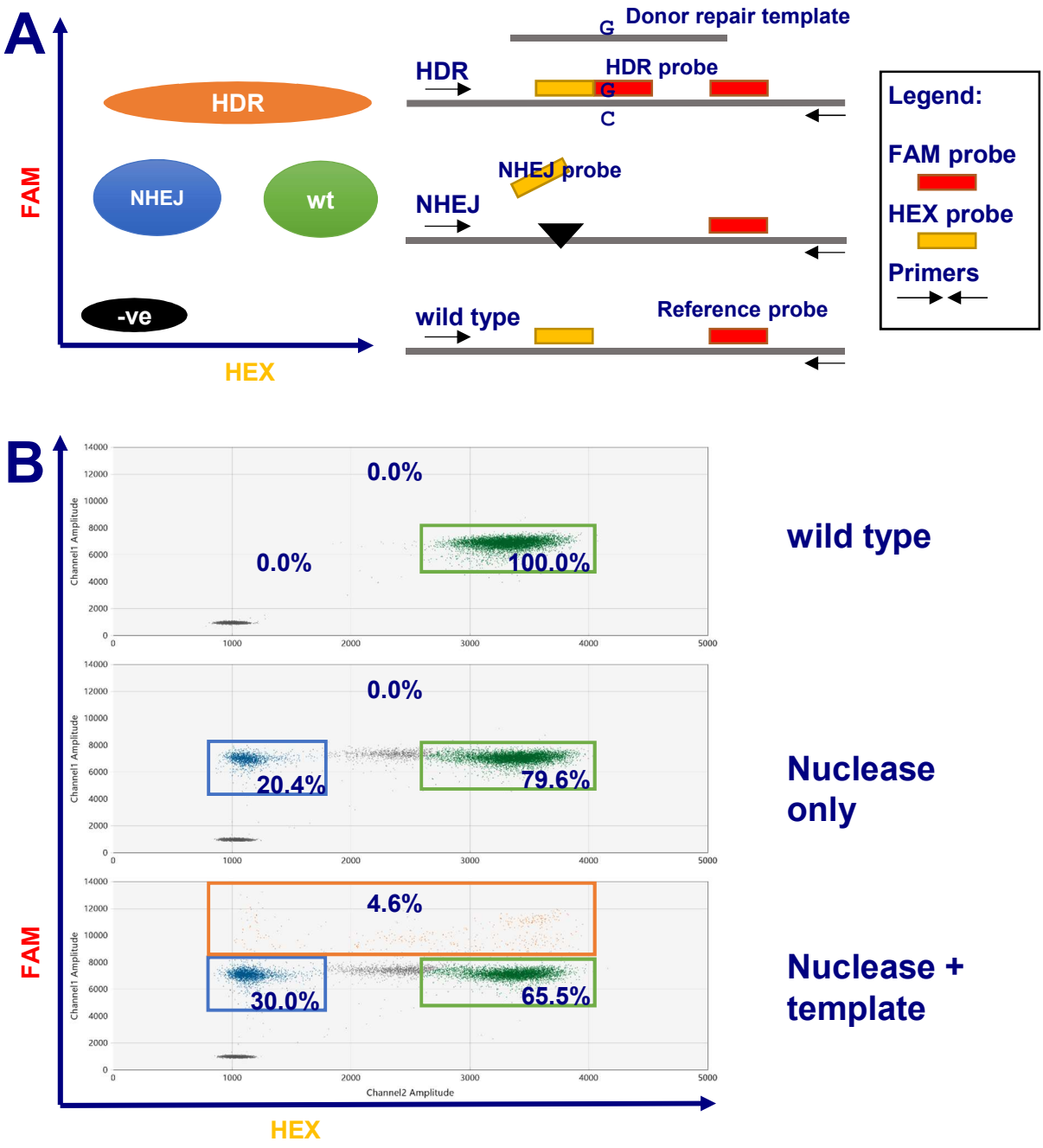


Figure 5.1: An assay to measure homology directed repair (HDR) and non-homologous end joining (NHEJ) by droplet digital PCR. (A) The reference probe distinguishes droplets containing the amplicon from negative droplets. Unedited amplicons allow binding of both the NHEJ and reference probe and are the double positive FAM+HEX population. Indel formation due to NHEJ prevents the NHEJ probe from binding, resulting in the single positive FAM population. When the template strand is used for repair, a single nucleotide difference is introduced. This allows binding of the HDR probe and leads to the FAM++ population. Exact design of the probes are shown in Supplementary Figure S10. (B) Example data (primed SFC856-03-04, locus *GRN*, guide F1) illustrates the gating used to distinguish different populations. The first plot shows unedited wild type cells lead to a FAM+HEX population. Adding the nuclease (and guide) causes cutting and repair by NHEJ and leads to an additional population of single positive FAM droplets. Only if the template oligo and nuclease are introduced the HDR FAM++ population appears. Uptake of the RNP is often improved when an oligo is present, explaining the increase in the NHEJ population.

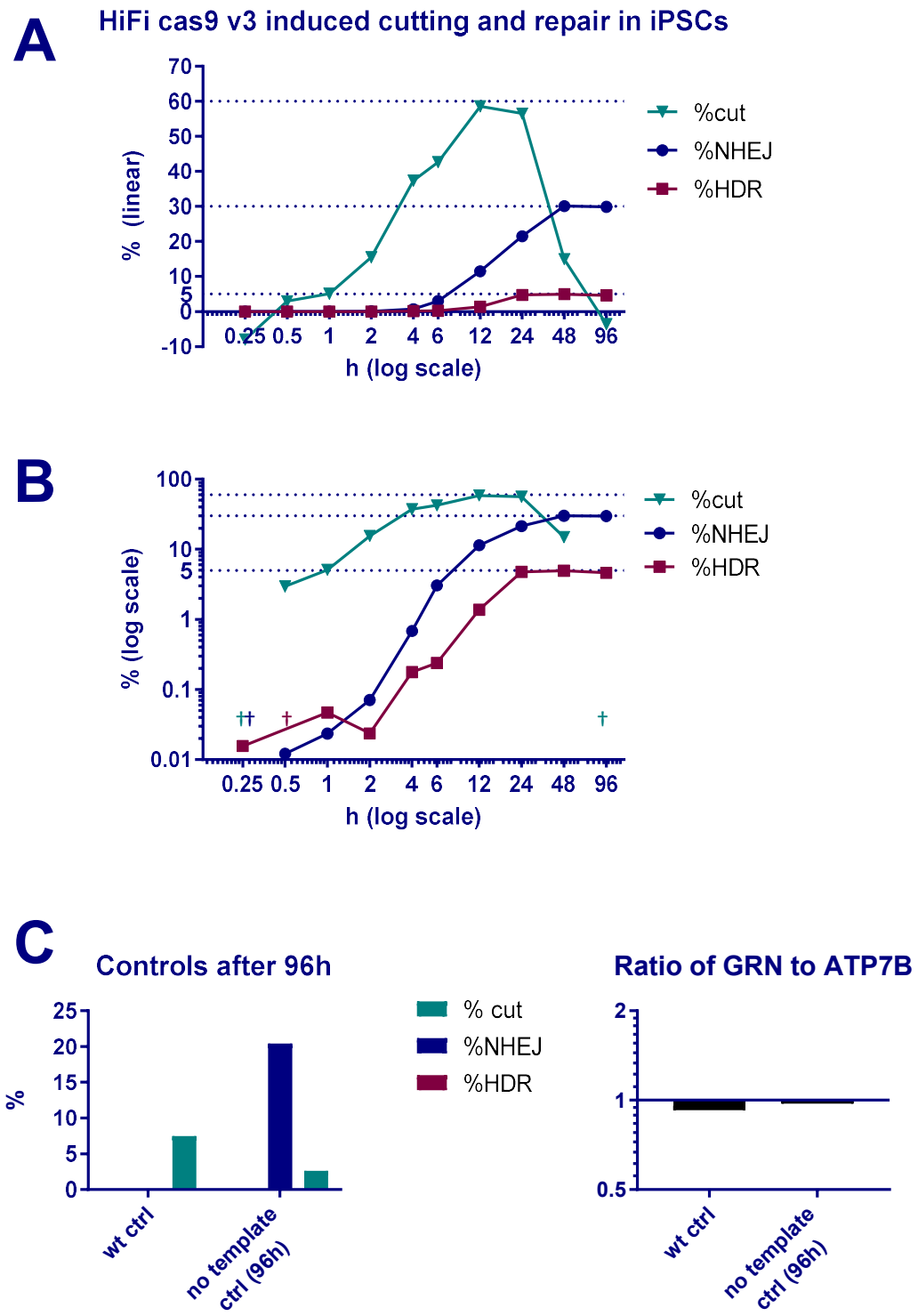


Figure 5.2: A time course showing emergence and repair of a Cas9 induced double strand break in iPSC shows repair is completed by 48 hours after electroporation. (A) The proportion of alleles repaired by homology directed repair (HDR) increases most predominantly after 6 hours and plateaus after 24h. The time point 0 is the time of electroporation of the RNP complex. The locus *GRN* was edited using guide RNA F1 in primed SFC856-03-04. No technical replicates were performed and further biological replicates are required to test if other loci and other cell lines behave differently or the same. (B) The same data but with the Y axis in log scale is plotted here, to show earlier time points. The dagger (†) indicates a data point not plotted due to its value being 0 or negative. (C) Controls show that the %cut is an inaccurate measurement. The ratio of GRN to ATP7B which informs the %cut measurement is also shown.

HDR was detected (Supplementary Figure S11B) and loci GRN and RBM20 are used for the remainder of the thesis.

5.2.2. HDR occurs only during the first 24h after electroporation of Cas9, guide RNAs and template oligo

Miyaoka *et al.* used the assay with plasmid based (pX330) Cas9. This thesis shows that with only minor alterations, HiFi Cas9 protein can be used. This improves the efficiency of the experiment and, unlike the original published method, allows very accurate measurement of repair over time. To find out when repair first occurs post electroporation and for how long new repair events continue to occur, primed SFC856-03-04 cells were electroporated with HiFi Cas9, guide RNAs and repair template oligo for the GRN locus. Pellets of electroporated cells were recovered over a 96h time course, DNA was extracted, and the type and extent of repair determined by droplet digital PCR (Figure 5.2).

To measure cut, but not yet repaired DNA a second reaction was set up for each sample measuring the concentration of an additional locus (ATP7B). Cut DNA is not amplified so the ratio of the concentrations of GRN over ATP7B measures the %cut for each sample. Since this relies on two separate measurements of the sample, it is less accurate than a multiplexing assay. This inaccuracy is made clear by the negative values at the 15 minute and 96h time points but also the positive values of the wild type control (Figure 5.2C). However, despite this caveat, it is clear that over the 96h time course, the %cut rises sharply within the first 12h and begins to decrease after 24 hours. After 96h, no cut DNA remains (Figure 5.2A). Repair with indel formation occurs at low levels as early as 30 minutes after transfection, reaching 1% after 4 hours and plateauing at 30% after 48h (Figure 5.2A, B). The measurement of the repair using the template strand seems variable at very low levels, but a clear increase is seen at 4 hours and beyond. After 24h, HDR levels plateau at 5% (Figure 5.2A, B).

Although there are no technical or biological replicates, it should be noted that there is inherent statistical confidence in this experiment due to the very large number of droplets analysed and the many samples taken across the time course. However, this experiment needs to be repeated with several iPSC lines and different loci to assess if these repair kinetics are universal or whether they vary.

Importantly, this experiment has shown that all repair has been completed and no further Cas9-mediated cleavage of DNA occurs after 96h. For all further experiments using this assay, cells were assessed after 96h.

5.2.3. The ratio of HDR vs. NHEJ repaired DNA is lower in naïve hPSC

To measure and compare rates of HDR in two very different cell types, a robust internal control is required for normalisation. By normalising rates of HDR to NHEJ, variation introduced by, for example, altered Cas9 uptake efficiency, is removed. Normalisation was performed by first plotting all HDR and NHEJ data collected for each of the loci RBM20 and GRN (Supplementary Figure S12). A least squares linear regression was used to fit lines separately for naïve and primed cells for both loci. The slope is not significantly different between naïve and primed cells (ANCOVA, RBM20 $p=0.85$, GRN $p=0.80$). The combined slope of both naïve and primed lines was used for normalising each locus (RBM20 0.043, GRN 0.025). Normalisation was performed using the formula

normalised HDR = %HDR - (%NHEJ x slope).

The normalised HDR value is therefore equivalent to the intersect on the Y axis, which represents the theoretical HDR rate at 0% NHEJ.

Naïve cells established using the Theunissen 4iLA protocol were used in this experiment. Naïve and their corresponding primed cell lines were transfected with Cas9, guide RNAs (Table 2.5) and template strands (Table 2.2) for either GRN or RBM20. Each transfection was performed in

duplicate and an extra non-specific template control using an unspecific oligo ("electroporation enhancer") instead of the template was included. After 4 days, DNA was extracted from the cells and the ratio of HDR to NHEJ was measured using the droplet digital PCR assay. For both loci, the ratio of HDR to NHEJ is significantly lower (0.4-0.5 fold, $p<0.0001$, two-tailed students t-test) in naïve cells (Figure 5.3B). The experiments were performed with 3 karyotypically normal lines, as well as SFC856-03-04, which is karyotypically abnormal in the naïve condition.

As expected, neither NHEJ nor HDR is observed in the wild type controls and HDR is not observed in the nuclease only control (Supplementary Figure S11). Generally (except for naïve OX3-6/GRN), the R2 guide RNA shows less NHEJ than the nuclease control. This effect is exclusively observed for R2 guide RNAs, indicating that design may have caused the difference. R2 and F1 guide RNAs differ in the strand that they target. The sequence of the R2 guide RNA for both loci is complementary to the template oligo (Supplementary Figure S10), which might be sequestering the guide RNA away from the RNP complex. This may affect the experiment and R2 guide RNAs were excluded from further data analysis. In future gene editing experiments the anti-sense oligo should be used with R2 guide RNAs.

5.2.4. The proportion of cells in G_1 phase are higher in naïve vs. primed cells

Since HDR occurs only in late S and G_2 phases, changes in cell cycle dynamics affects rates of HDR (Shao et al., 2012, Sun et al., 2012, Huertas and Jackson, 2009, Falck et al., 2012). Naïve and primed cells were fixed (each replicate on a different day), stained with propidium iodide (PI) and their DNA content measured by flow cytometry. Cell debris and duplets were removed by gating and the Dean-Jett-Fox method was used to calculate proportions of cells in different stages of the cell cycle (Figure 5.4A,B)(Fox, 1980). The proportion of cells in G_1 phase, where no HDR takes place, is significantly higher in naïve cells (Figure 5.4C). This is likely the reason, or at least a contributor to the lower rate of HDR observed before.

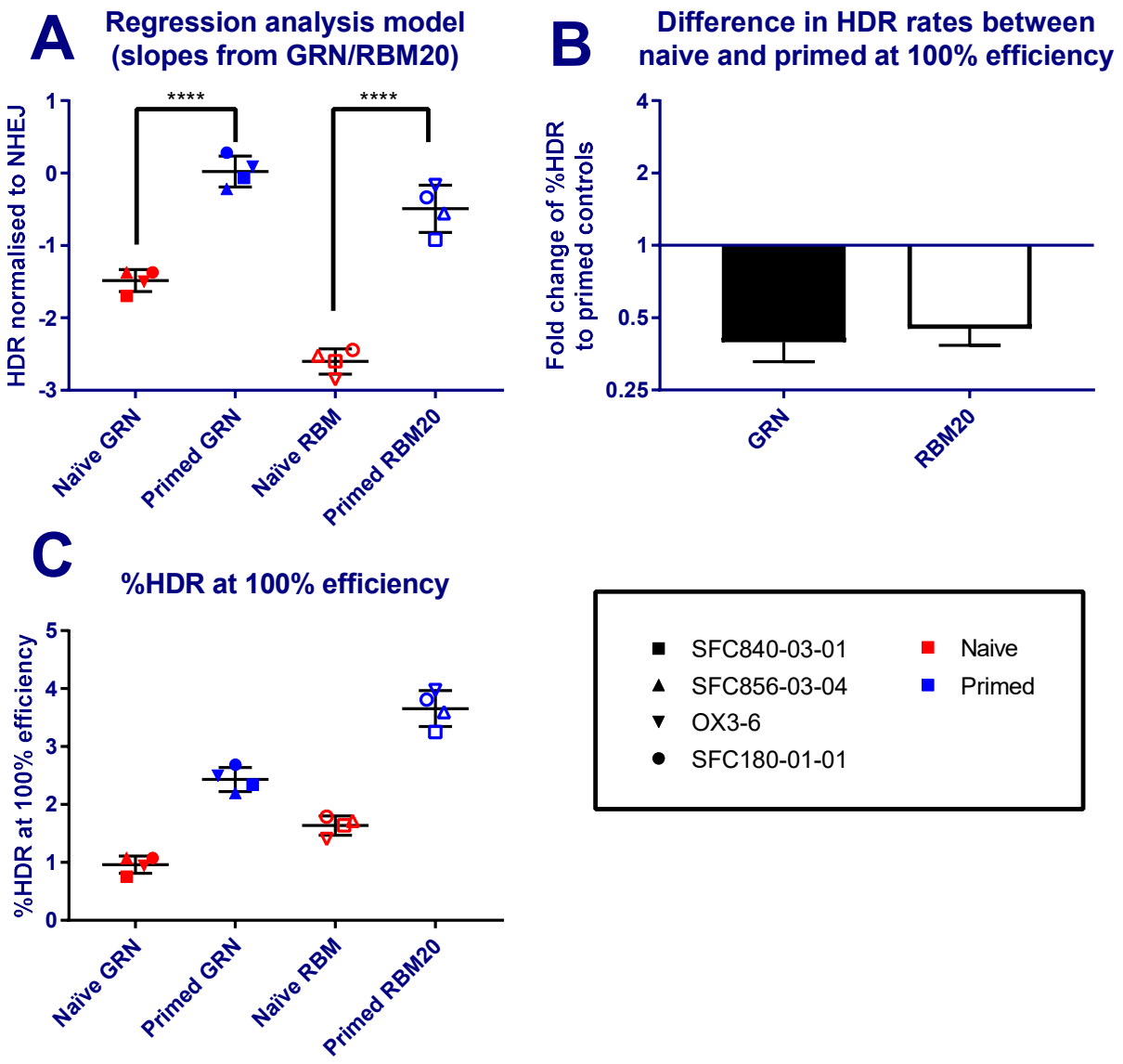


Figure 5.3: DNA repaired by the HDR pathway is observed less often in naïve cells than primed. (A) Each symbol is the mean of 2 electroporations of the same cell line and error bars show the standard deviation of the 4 biological replicates. The rate of HDR is normalised to NHEJ to remove variation introduced by electroporation efficiency. Normalisation is performed by regression analysis using the formula

normalised HDR = %HDR - (%NHEJ x slope).

The normalised HDR rate can be negative as it is equivalent to the theoretical baseline HDR rate at 0% NHEJ. The slope is calculated in Supplementary Figure S12. The underlying data before normalisation is shown in Supplementary Figure S11. A two-tailed students t-test informed significance (**** $P \leq 0.0001$). (B) The fold change between naïve and primed cells is approximately 0.4-0.5 for both loci. (C) The theoretical HDR rates at 100% cutting efficiency are displayed here.

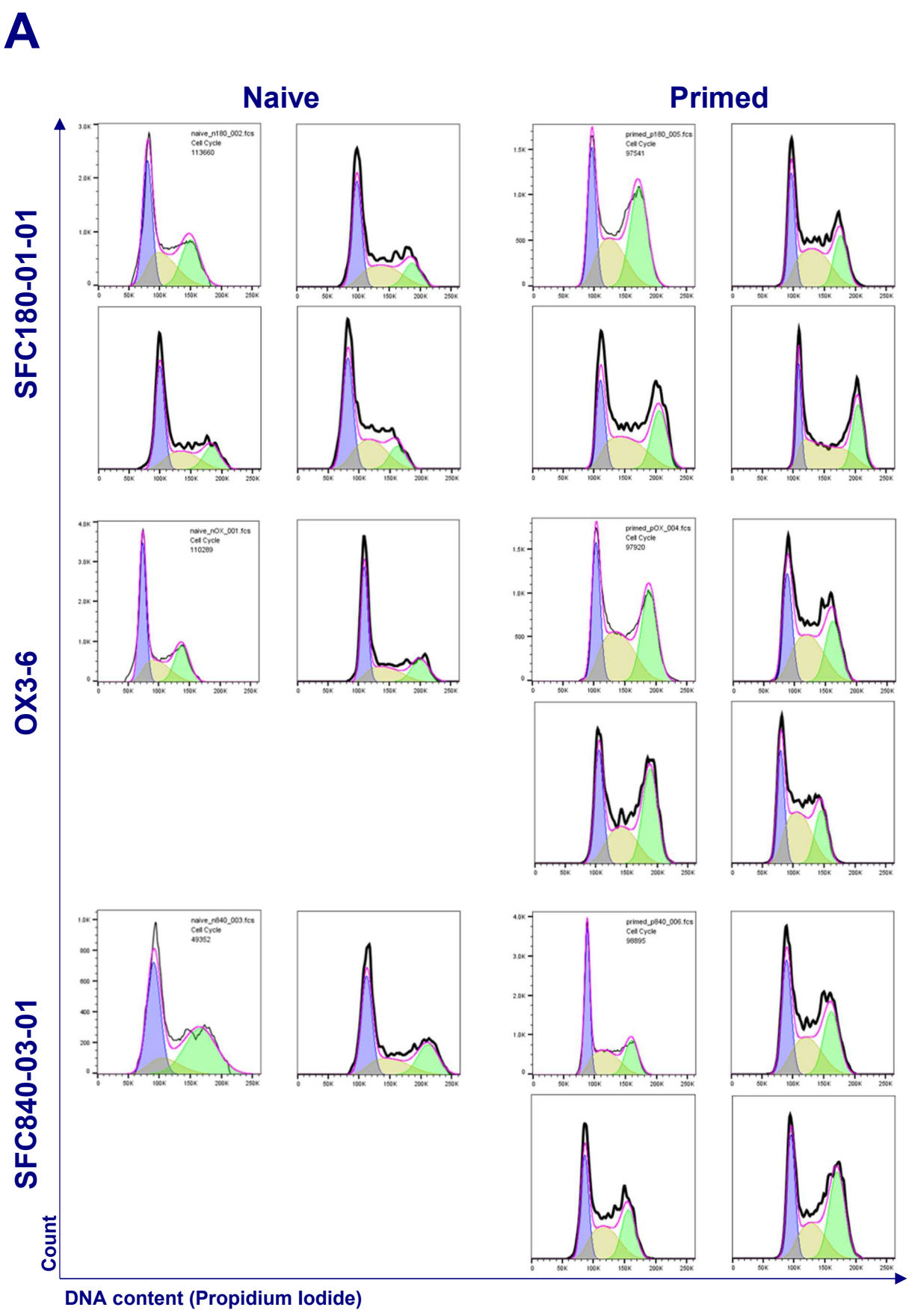
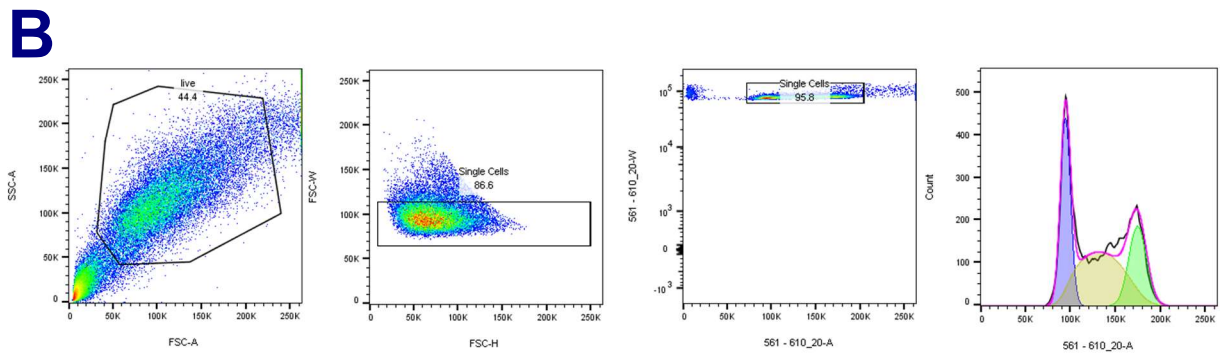


Figure 5.4 continued on next page



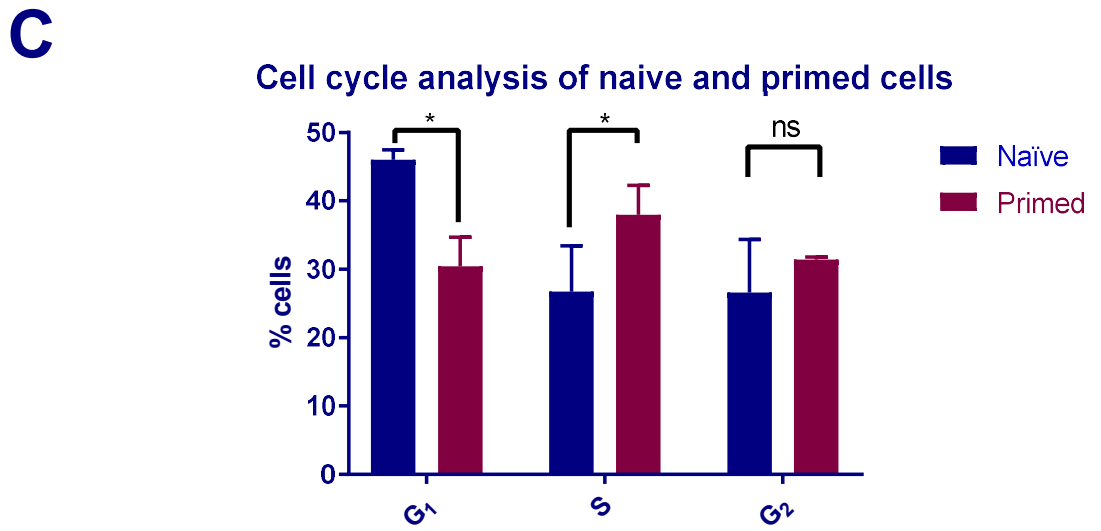


Figure 5.4: The proportion of cells in G_1 phase is higher in naïve versus primed cells. 3 cell lines (SFC180-01-01, OX3-6, SFC840-03-01) each in naïve and primed states were fixed (2-4 separate replicates, fixed on different days) and stained with propidium iodide (PI) to analyse the proportion of cells in each stage of the cell cycle. (A) PI staining shows how much DNA each cell contains. Histograms of the number of cells compared to their DNA content are shown here. The DNA content of a cell is doubled during S phase. The Dean-Jett-Fox method was used to calculate proportions of cells in each stage of the cell cycle. In this mathematical model, cells predicted to be in G_1 are marked in purple, yellow are in S phase and green in G_2. The pink line shows the sum of these populations and should follow the actual measured DNA content (black line) as closely as possible. Bold plots were analysed by flow cytometry together. (B) The gating strategy for PI staining ensures only single cells are analysed. (C) Cell cycle analysis shows a higher proportion of naïve cells are in G_1 than primed cells. Significance informed by a two-tailed students ratio paired t-test on 3 biological replicates (averages of 2-4 replicates; ns = $p > 0.05$; * = $p \leq 0.05$).

5.3. Summary and discussion

5.3.1. Cas9 kinetics

In this chapter, the assay described by Miyaoka *et al.*, which measures repair by HR/NHEJ of a nuclease induced DSB (Miyaoka et al., 2016), was adapted to function with HiFi Cas9 protein rather than plasmid based Cas9 or TALENs. Unlike the original assay, this now allows accurate measurements over time, and was used to measure cutting and repair across a 96h time course (Figure 5.2). To the best of our knowledge, this is the first dataset investigating Cas9 kinetics in hiPSC.

The accuracy of the percentage cut (%cut) measurement is low, as evidenced by the controls and first and final time points (Figure 5.2). This is likely due to technical error and could be improved by increasing replicates or multiplexing. To measure %cut, the frequency of amplified droplets from one locus is compared to an unrelated locus. The difference allows the %cut to be calculated, since a cut region will not amplify. However, without multiplexing, two separate wells are required, and additional error is introduced. Two strategies for multiplexing could be applied to overcome this: Primers and the FAM reference probe for the GRN locus mixed with either ATP7B primers and the HEX NHEJ probe or the RNase P Taqman copy number control, both used in this thesis (Table 2.1, Table 2.2).

The time course data fits well with previously published results. DSBs caused by ionising radiation are repaired in under 1h in mammalian cells such as Chinese hamster ovary (CHO), normal human skin fibroblasts and human malignant glioma cells (Metzger and Iliakis, 2009, Stenerlöw et al., 2003, Wang et al., 2006). However, and as our data suggests, it was previously observed that repair of Cas9 induced breaks takes significantly longer (Brinkman et al., 2018, Rose et al., 2017). This could be due to Cas9 remaining tightly bound to the DNA after cutting (Richardson et al., 2016, Sternberg et al., 2014, Shibata et al., 2017). Catalytically inactive Cas9 has been shown to bind to DNA *in vivo* (Knight et al., 2015) and remain bound for 2h (Ma et al., 2016a). *In vitro*, wild

type Cas9 has been shown to bind indefinitely until harsh denaturing conditions are applied (Brinkman et al., 2018). Another factor affecting the perceived time taken for repair, is that correct repair, resulting in restoration of wildtype sequence after cleavage, cannot be measured. This has led to the hypothesis that C-NHEJ might be a very accurate pathway and that several rounds of cutting and perfect repair occur before an incorrect repair (Betermier et al., 2014). However, using a DNA sequencing and modelling based approach, Brinkman *et al*. estimated perfect repair of Cas9 induced DSB to be very rare and repair with indel formation considerably more frequent (Brinkman et al., 2018). In this thesis, a maximum of ~60% cut DNA was observed after introduction of Cas9, but only 35% was repaired by erroneous NHEJ or HDR (using the template oligo). This suggests that a large proportion is indeed repaired correctly – at least ~25% but possibly more which is then re-cut. This ~25% correct repair contrasts slightly with Brinkman *et al*.'s findings, since they calculated perfect repair to be very rare. They estimated perfect repair occurring at a frequency of 0.01%, although lack of robustness in their model lead them to a more conservative approximation of a frequency of at most 28%. Our observation is still in line with their prediction, but certainly is at the higher end of the scale. Crucially, their experiments were performed differently with their data stemming from only one locus and a different Cas9 protein. Most importantly, the cells they used are the human myeloid leukaemia cell line K562. This cancer cell line harbours many oncogenic mutations including TP53 (Law et al., 1993) – more erroneous repair is therefore not unexpected.

Previous attempts at measuring Cas9 kinetics *in vivo* have been largely limited by varying Cas9 expression levels from inducible or plasmid-based systems. This was addressed by Rose *et al.* who developed a chemically activated Cas9 (Rose et al., 2017) and by Brinkman *et al.* who used a K562 cancer cell line constitutively expressing an unstable Cas9, which can be stabilised by the addition of a ligand (Brinkman et al., 2018). The results in this thesis are a proof of principle, requiring more replicates in different cell lines and loci. Encouragingly, Brinkman *et al*. found the highest percentage of cut DNA to be 12-16h after induction in their K562 cells - the same as we have

observed with iPSC. Using high throughput sequencing, they found mostly the same +1 insertions or -7 deletions. Using inhibitors of C-NHEJ they found that the +1 indels were due to repair by C-NHEJ and -7 deletions by MMEJ. They also found that MMEJ is used less frequently due to a delayed onset (Brinkman et al., 2018). Their results support the hypothesis that MMEJ may act as a backup to C-NHEJ (Guirouilh-Barbat et al., 2007). The assay used in this thesis is easier to analyse but cannot distinguish between the two pathways. It is certainly possible to sequence the DNA generated from this experiment to interrogate the repair further. However, Brinkman *et al*.'s finding of differential repair patterns is likely due to the specific microhomologies present in their chosen locus.

MMEJ is a driver of mutations because it is error prone and leads to deletion of the region between the homologies. Recent research has shown that repair by MMEJ often leads to predictable outcomes based on local microhomologies (Allen et al., 2018). MMEJ is linked to the plasticity of genomes and the pathway is used for example, during class-switch recombination in developing lymphocytes. (Sfeir and Symington, 2015, Ottaviani et al., 2014)

Finally, repair by SSA would likely lead to a deletion of the reference probe binding site. Repair by SSA would therefore be measured as cut but not repaired DNA. If the accuracy of the %cut measurement can be improved, we might discover a fraction of cells where the locus is entirely deleted.

Our results concur with current estimates of Cas9 efficiency and represent a useful guide for setting up experiments in hiPSC. To calculate accurate kinetics of Cas9 mediated cutting, replicates with different cell lines and different loci are required. The advantage of Cas9 protein plus gRNA/tracr RNA/template DNA electroporation, as used here, is a prompt and homogenous start to the process. The droplet digital PCR assay itself is also highly accurate. However, one disadvantage is that we have not yet characterised the degradation of the Cas9 protein, gRNA/tracr RNA or template strand and this may affect the results, if kinetic measurements are

of interest. The degradation of these components could be measured by western blot for Cas9 and qPCR or ddPCR for the template strand. As well as degradation, dilution of components may occur due to cell division. Normalising to a loading control for western blot and an endogenous locus for qPCR allows quantification of loss of Cas9 and template strand both through dilution and degradation. Finally, loss of gRNA/tracr RNA can be measured by using a fluorescently labelled gRNA/tracr RNA and assessing their presence by flow cytometry.

5.3.2. Proportion of HR repaired DSBs differs between naïve and primed hPSC

Naïve hPSC were observed to have repaired Cas9 induced DSBs using HDR significantly less frequently than primed cells (Figure 5.3; Figure 5.5). Naïve and primed hPSC behave very differently and the comparison between them was achieved by normalising rates of HDR to NHEJ. This was particularly important since different optimal electroporation programmes were found, and the total repair frequencies were also different between the two cell types (Supplementary Figure S11). Optimisation of the electroporation was performed in chapter 3 to ensure high viability as well as high uptake. These optimisations indicate that naïve cells survive harsher electroporation conditions, which allows for higher uptake efficiency (60-80% compared to 20-60% for primed cells; Supplementary Figure S11). This is likely what was observed by Buecker *et al.* who measured significantly more surviving naïve than primed colonies after electroporation (Buecker et al., 2010). Gafni *et al.* showed that naïve, but not primed hPSC, can spontaneously integrate a targeting vector with large homology arms into the correct locus by homologous recombination (Gafni et al., 2013). Unlike the experiments in this thesis, no DSB was introduced to induce the repair, meaning that they were not measuring how frequently a repair pathway is used but instead how genetically unstable their cells were. On the other hand, they do demonstrate another similarity to mouse ESC. Mouse ESC can also be edited using large homology arms and no exogenously-delivered nuclease (Mansouri, 2018).

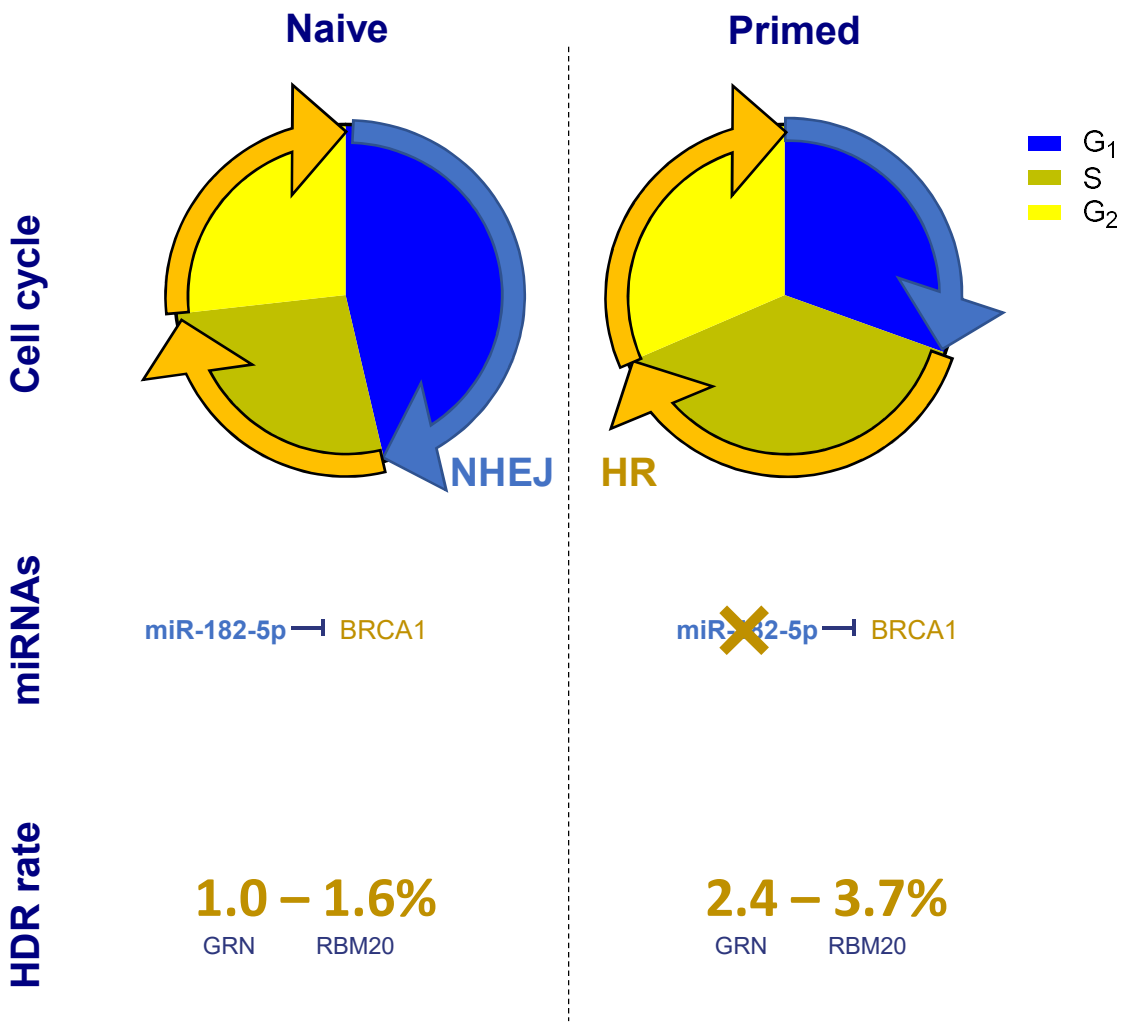


Figure 5.5: A summary of the findings regarding DNA repair pathway use and cell cycle. The proportion of naïve hPSC in S and G_2 phase is lower than in primed cells. In addition, miR-182-5p has been shown to target members of the HR pathway (including BRCA1) and is expressed in naïve but not primed cells. Finally, the average rate of HDR is lower in naïve cells. The HDR rate is expressed as the %HDR at 100% efficiency and is an average of all replicates of the locus GRN and RBM20.

5.3.3. Regulation of HDR by cell cycle and miRNAs

Naïve cells repair DSBs less often by HDR than primed cells. The repair pathways are under the control of the cell cycle and less naïve cells are present in S and G_2 phases where HDR takes place (Figure 5.4; Figure 5.5). The link between activity of HDR and stage of cell cycle is well established (Shao et al., 2012, Sun et al., 2012, Huertas and Jackson, 2009, Falck et al., 2012) and more cells in S and G_2 phase would be expected to result in higher levels of HDR, which was observed in this thesis. Further experiments are required to determine if this is the sole reason for differential use of DNA repair pathways. It is certainly possible that other regulators play a role. For example, in Chapter 4 we observed that miR-182-5p is upregulated in naïve hPSC (Figure 4.4B) which has been shown to target multiple components of the homologous recombination pathway. Of note, *BRCA1* (Liu et al., 2012, Moskwa et al., 2011, Krishnan et al., 2012) as well as its upstream regulator *CHEK2* (Krishnan et al., 2012) were shown to be directly regulated by the miRNA. Functionally, overexpression of miR-182 causes sensitivity to PARP inhibitors due to the downregulation of *BRCA1* (Moskwa et al., 2011). The sensitivity to PARP inhibitors is a phenotype linked to defects in the homologous recombination pathway often via BRCA1/2 (Patel et al., 2011). Therefore, it is possible that the high expression of this miRNA might contribute to the lower use of the HDR pathway.

5.3.4. Summary

In this chapter, an assay to measure both HDR and NHEJ was adapted to function with naïve and primed hPSC and Cas9 protein. The timing of DSB repair over a time course of 96h was investigated in primed hPSC and revealed different kinetics of HR and NHEJ repair. In addition, this assay was used to demonstrate for the first time that naïve hPSC have a significantly less active HDR pathway compared to primed. Possible reasons for this were found. An analysis of the cell cycle revealed that significantly more naïve hPSC are in G_1 phase and the high expression of miR-182-5p in naïve hPSC may lower HDR activity (Figure 5.5).

6
General discussion

6.1. General discussion and future perspectives

The aims of this thesis were to establish which naïve hPSC protocols yield cells most closely resembling the *in vivo* naïve state, find out which miRNAs are involved in naïve, primed and differentiating cells, and finally, dissect whether and why rates of HDR are different between naïve and primed hPSC. This thesis has addressed all these questions and sheds light on several aspects of the novel naïve hPSC cell type. In this chapter, the broader significance of the results and future perspectives are discussed.

6.1.1. Protocols to generate naive hPSC produce variable results

Revisiting published RNAseq datasets and comparing to *in vivo* data showed a close resemblance of Theunissen *et al.* and Takashima *et al.* naïve hPSC to cells from human E5 blastocysts. This was corroborated by the finding that only the Theunissen *et al.* protocol out of 7 tested naïve cell generation protocols produced convincingly naïve hPSC in my hands. The Takashima *et al.* protocol was not tested, but instead directly derived HNES1 and HNES2 lines were obtained from the same lab and showed similar naïve marker gene expression to Theunissen *et al.* cells. The observation that naïve hPSC conversion protocols are of variable quality has been reported by other researchers attempting to replicate results (Warrier et al., 2017, Theunissen et al., 2016) and the current requirement for feeder cells (Table 1.1) also suggests lack of robustness in culture conditions. In this thesis we have compared culture conditions and found Theunissen *et al.* most

promising, albeit requiring further adaptations to address the hurdles discussed later in this discussion.

6.1.2. miRNA expression profiles distinguish naïve and primed pluripotent states and suggest close resemblance to mouse ESC

miRNA expression profiles differ greatly between naïve, differentiating and primed states. The miRNAome alone is sufficient to distinguish these states (Figure 4.4). In mESC, the miR290 family is highly expressed, dominating the miRNAome. In contrast, the miR290 family as such does not exist in hPSC and the miR371-373 homolog is variably, if at all expressed in conventional hPSC (Wu et al., 2014a, Lipchina et al., 2011, Bar et al., 2008). Instead, the miR302 family dominates the miRNAome of conventional hPSC. Initially this was thought to be a species difference, though the discovery of low expression of miR290 and high expression of miR302 family miRNAs in mEpiSC questioned this conclusion (Jouneau et al., 2012). Since mEpiSC are equivalent to conventional hPSC, the high miR302 expression is indeed not unexpected and due to a later developmental time point. This leads to the question whether the human homologs of the mouse miR290 family, miR371-373, are also highly expressed in naïve cells. Indeed, in this thesis we have demonstrated that miR371-373 are highly upregulated in naïve cells. Excitingly, this expression pattern was also observed for miR290 family miRNAs *in vivo* in developing mouse embryos, corroborating that naïve hPSC may indeed be a physiological representation of human early embryogenesis, rather than an entirely artificial state.

The high expression of miR371-373 in naïve hPSC has further implications. Conventional hPSC cannot directly be differentiated to primordial germ cell-like cells (PGCLC). To generate hPGCLC, primed hPSC need to be converted to an intermediate state by bFGF withdrawal, supplementation of Activin A, and GSK3β inhibition using CHIR. After several days, differentiation to PGCLC is induced by adding a cocktail of cytokines including LIF (Sasaki et al., 2015, Yamashiro et al., 2018). Intriguingly, the generation of the intermediate state resembles the conversion to naïve hPSC. In

addition, the capacity to generate hPGCLC is generally low and varies considerably between cell lines. Lower PGCLC differentiation capacity of lines was linked to increased expression of the primed marker bFGF (Yokobayashi et al., 2017). Using naïve hPSC as a starting point has been suggested to allow direct and more efficient differentiation to PGCLC and has been a key driver of naïve hPSC research (Hayashi et al., 2011, Irie et al., 2015). In mouse preimplantation blastocysts, miR290 family miRNAs expression is silenced in postimplantation epiblast cells but remains active in cells designated to become germ cells (Heard et al., 2008, Medeiros et al., 2011). This indeed suggests that conventional hPSC may have passed the avenue for germ cell differentiation and that naïve hPSC, still highly expressing miR371-373, may be able to access this differentiation route more efficiently. In fact, primordial germ cells have been generated using naïve hPSC (Irie et al., 2015) and this approach is verified by our finding of miR-371-373 expression. It would be interesting to find out if the capacity to generate hPGCLC from primed cells is also linked to miR-371-373 expression, considering its variable expression in primed cells seen in this thesis and in the literature (Wu et al., 2014a).

6.1.3. Naïve hPSC accumulate in G_1 stage of the cell cycle, affecting HDR

To compare rates of HDR between two different cell types, an assay was used which also measured rates of erroneous NHEJ/MMEJ as well as HDR (Miyaoka et al., 2016). Comparison of the HDR rates between the two cell types was possible by normalising to rates of NHEJ, eliminating any differences due to, for example, uptake efficiency. Using this strategy, lower rates of HDR were uncovered in naïve hPSC. The reason, or at least a contributor, is that naïve hPSC accumulate in G_1 phase, where HDR does not take place.

Mouse ESC divide rapidly with a doubling time of only 10-14h compared to 18h for mEpiSC. Human cells require longer per cell division both *in vivo* and *in vitro* with doubling times of conventional hPSC of 20-30h (Figure 1.1A) (Panyutin et al., 2017, Tesar et al., 2007, Brons et al., 2007). Doubling times of naïve Takashima *et al.* hPSC are not significantly different, around 24h (Guo et al., 2017).

It is unclear if this is a difference in species or if the culture conditions for naïve hPSC have yet to be optimised to allow faster growth. The accumulation of cells in G_1 phase observed in this thesis may indicate the latter.

6.1.4. miRNAs affecting cell cycle and HR

Several miRNAs are known to modulate HR efficiencies or the cell cycle. For example, in chapter 4 we observed that miR-182-5p is highly upregulated in naïve hPSC (Figure 4.4B). This miRNA has been shown to target multiple components of the homologous recombination pathway, including *BRCA1* (Liu et al., 2012, Moskwa et al., 2011, Krishnan et al., 2012) and may contribute to the lower rates of HR.

The effects of miRNAs on the cell cycle have been well studied. Unlike somatic cells, PSC spend significantly less time in G_1 phase, since they are rapidly dividing. Swift progression through the G_1-S checkpoint is enabled in part by miRNAs. In the mouse, knockouts of *Dgcr8* and *Dicer* have longer doubling times and accumulate in G_1 phase (Wang et al., 2007, Kanellopoulou et al., 2005). This growth disadvantage can be rescued by supplementing ESC-specific cell cycle-regulating (ESCC) miRNAs which act redundantly. ESCC miRNAs have been shown to target expression of checkpoint proteins directly and indirectly to allow faster progression through the G_1-S checkpoint. They directly downregulate Rbl2, a protein which prevents cell cycle progression. Rbl2 is further downregulated by its repressor Cdk2-Cyclin E. Cdk2-Cyclin E expression is also increased by miRNA mediated downregulation of its repressors Lats2 and Cdkn1a (Wang et al., 2008). Global expression of ESCC miRNAs in mESC are reported to vary greatly with reports suggesting 24% (Jouneau et al., 2012) to 89% (Marson et al., 2008) of the total miRNA expression belonging to this category. Our results show 15% in naïve, 20% in differentiating and 20% or 12% in primed for HNES1 and HNES2, respectively. ESCC miRNAs are thought to be at a saturated level in wild type conditions (Wang et al., 2008, Greve et al., 2013) and primed hPSC do not show the G_1

accumulation, suggesting that these miRNAs may not play a role in the slower progression through the cell cycle observed in human naïve PSC.

6.1.5. Further directions

Further research is required to elucidate questions which have arisen during this thesis. In particular, a complete knockout of *DGCR8* or *DICER1* in naïve hPSC can elucidate miRNA requirements in early human development. If DGCR8/DICER1 protein is essential in naïve hPSC, this will uncover a key difference between human and mouse development and will ignite research into miRNAs as drivers of evolutionary change. If the knockout is successful, it should also be unable to differentiate to the primed state. Supplementing miRNA mimics can then identify which miRNAs are responsible for the differentiation process.

To find out how miRNAs are exerting control over the differentiation process, different techniques can be used to uncover miRNA response elements (MREs), which are miRNA binding sites in the 3' UTR of transcripts that allow miRNA mediated repression (Bassett et al., 2014, Wu et al., 2017b). Prediction of MREs *in silico* is error prone (Rajewsky, 2006) and was used in this thesis for the GSEA. More stringent approaches which allow quantification of the impact of each MRE in the cell type of interest can be used. For example, a plasmid containing a miRNA mimic and a fluorescent reporter can be transfected into the cells. Cell sorting allows the separation according to brightness, which is a surrogate measure for miRNA mimic dose. RNAseq then allows the identification of dose-dependent negative changes in transcript abundance, which indicates a functional MRE. Alternatively, a GenERA screen can be used (Wu et al., 2017b). Here, predicted MREs are disrupted using Cas9 and the abundance of mutant sequences can be correlated between transcriptome and genome. If deletions of the MRE are overrepresented in the transcriptome, this indicates that this region of the 3'UTR is indeed involved in downregulation of the transcript under physiological conditions. Both methods are possible in naïve hPSC, since electroporation of plasmids and Cas9 RNPs has been optimised in Chapter 3.

The lower rates of HR in naïve hPSC may be exclusively due to the accumulation of cells in G_1 phase. Alternatively, there may be more mechanisms involved. Two possible experiments can address this question by investigating rates of HDR only in cells in S or G_2 phase. This can be achieved using the HDR/NHEJ assay optimised in this thesis and either synchronising cells before, or by cell sorting an asynchronous population after the gene editing event. Flow cytometry sorting of cells which are in S and G_2 phase after electroporation is possible by supplementing the cells with EdU, a thymidine nucleotide analogue. Cells in S phase incorporate these and are made fluorescent by fixing the cells and binding a fluorescent tag to the EdU by a click reaction. Cells can then be sorted by flow cytometry, DNA extracted and rates of HDR and NHEJ measured by ddPCR. This approach will require careful timing, so our measurements of rates of HDR over time performed in chapter 5 will help determine optimal EdU supplementation time points. If the cell cycle differences are the only reason for differing rates of HR, rates of HR should not differ between these sorted cell populations.

6.1.6. Expectations and hurdles for naïve hPSC

The expectations for naïve hPSC have arisen by extrapolating differences seen between mEpiSC and mESC to the human system. These assume that many differences between mESC and conventional hPSC are due to cell state and not species difference. These expectations include that naïve hPSC should differentiate more widely and without lineage bias, allow for higher gene editing efficiencies and grow both faster and more robustly than their primed counterparts, all while maintaining a healthy karyotype (De Los Angeles et al., 2012). The results in this thesis indeed confirm some of these aspects but also emphasise several hurdles which need to be overcome for naïve hPSC to reach their potential.

Increased differentiation potential may be a feature of naïve hPSC. Advances in generating primordial germ cells have been made using naïve hPSC (Irie et al., 2015) and the finding of high miR-371-373 expression in this thesis suggests that this is a viable differentiation route. However,

conventional differentiation protocols have not been successful on naïve hPSC and instead differentiation to the primed state is required before conventional differentiation protocols can be successfully applied (Warrier et al., 2017, Guo et al., 2017). The expectation that mESC differentiation protocols may also work with naïve hPSC has yet to be investigated but may not be as likely considering the differing requirements in culture conditions. Removal of lineage bias may also be possible by conversion to the naïve state and epigenetic erasure but is yet to be investigated.

Faster growth and higher yields of naïve hPSC have not been observed in the literature (Guo et al., 2017) or in this thesis, although the doubling time of naïve hPSC was not formally assessed and quantified. Slower growth may be due to current media formulations not yet being fully optimised and very tightly holding naïve cells in their state. Indeed, the high occurrence of karyotypical abnormalities in naïve hPSC cultures observed here and in the literature (Theunissen et al., 2014, Di Stefano and Hochedlinger, 2018, Guo et al., 2016) suggests that genetic adaptations to these stringent culture conditions provide drastic growth advantages. The genomic instability observed in naïve hPSC cultures is a crucial hurdle which needs to be overcome to allow confident use of these cells. Addressing this problem by identifying media components which increase the occurrence of karyotypic abnormalities, and altering the media formulation by titrating or removing them, could improve both genomic stability and yields. Early but promising results have been achieved by titration of the MEK inhibitor PD03, thereby improving genomic stability (Di Stefano and Hochedlinger, 2018). In addition, the requirement for MEFs indicates non-optimal media formulations. In fact, the requirement for MEFs is also a hurdle to overcome for naïve hPSC culture. Feeder-free culture increases yields dramatically, removes MEFs as a contaminating cell type and allows development of Good Manufacturing Practice (GMP) compliant cell lines. Again, some pioneering work has already been performed. Titration of CHIR in t2iLGo medium to even lower levels (10% of original 2i medium; termed tt2iLGo) and the use of growth factor free geltrex spiked into the naïve medium directly while plating, has been

discovered as a functional feeder-free method (Guo et al., 2017). The CHIR equivalent GSK3β inhibitor BIO used in the original 5iLA Theunissen *et al.* protocol has been omitted in the revised 4iLA already, suggesting the feeder-free method using growth factor free geltrex may also work for Theunissen *et al*. cells. In summary, the correct media formulation may allow the key hurdles of genomic instability, low yields and requirement of feeders to be overcome.

Gene editing efficiency is a multifaceted property. It combines the ability to survive as single cells for clonal selection steps, the rates of HR and finally, the efficiency and survival of available transfection procedures. Previous work showed that naïve hPSC survive transfection and clonal selection procedures more robustly than their primed counterparts (Buecker et al., 2010). This has been indirectly observed in this thesis. The optimisation of electroporation programmes indicates a higher tolerance of naïve hPSC to more aggressive settings, although a controlled experiment has not been performed. Naïve hPSC therefore have some properties which facilitate gene editing but have lower HR activity. Again, this hurdle may be overcome by the right media formulation, if this removes the accumulation of cells at G_1 phase.

A final promise of naïve hPSC is that access to both naïve and primed hPSC may allow study of human embryogenesis *in vitro*. This has been fulfilled in this thesis, since the naïve to primed differentiation has, similarly to the mouse, been used as a model for human embryogenesis. This opens a new avenue for understanding human pre-implantation biology. Experiments which require many cells, such as the miRNAseq performed in this thesis, cannot be performed using *ex vivo* developing blastocysts due to limiting cell numbers. The naïve to primed differentiation model is therefore an excellent alternative.

A final hurdle is that a new cell type requires adaptations of existing protocols. The adaptation of electroporation programme performed in this thesis can be applied to any other laboratory using naïve hPSC. Similar work will be performed and published in time, increasing the available methods which are compatible with this novel cell type. Naïve hPSC may indeed become a

valuable cell type in our arsenal of human pluripotent stem cells, if optimal culture conditions can be found, their genetic stability increased, and protocols can be adapted to function with these cells.

6.2. Concluding remarks

This thesis has uncovered which naïve hPSC protocols yield cells most closely resembling the *in vivo* naïve state, revealed which miRNAs are involved in naïve, primed and differentiating cells and finally dissected whether and why rates of HDR are different between naïve and primed cells.

Naïve hPSC have only recently been successfully cultured *in vitro*. This thesis has identified the Theunissen *et al.* 4iLA protocol as the most reproducible protocol and showed that cells generated with this protocol most closely resemble *in vivo* naïve cells. However, the multiple hurdles of low yields, requirements for MEFs, frequent karyotypic abnormalities and low rates of HR still impair the use of these cells. Further optimisation of the media may improve all these aspects and indeed has already incrementally done so.

In this thesis, we have added to the debate around the authenticity of these cells. The high expression of miR-371-373 in naïve hPSC suggests that they indeed closely resemble the inner cell mass of an E5 blastocyst. This finding also gives hope for advances in finding a differentiation protocol for germ cells, a key promise of naïve hPSC.

This thesis has identified the low activity of HR as a feature of naïve hPSC. The finding of cells accumulating in G_1 phase is a likely reason or contributor. Naïve hPSC may still be beneficial for gene editing, due to greater survival of electroporation and cloning. Generation of knockouts do not require HR and can certainly benefit from the features of naïve hPSC. Synchronising the culture and precisely timing the Cas9 delivery has significantly improved gene editing efficiency in hESC (Lin et al., 2014) and could be employed to overcome low HR rates in naïve hPSC. Finally, if the G_1 accumulation is solved by optimisation of media, naïve hPSC may be a better choice for

gene editing than conventional hPSC, once protocols have been adapted to work with this new and challenging cell type.

7 Appendix

7.1. Supplementary methods

Dr Klas Hatje performed the following analyses:

7.1.1. Pseudotime analysis for single cell transcriptomics data of human preimplantation embryos integrated with bulk transcriptomics data of embryonic stem cells

Multiple publicly available single cell transcriptomics datasets from human preimplantation embryos and bulk transcriptomics dataset from embryonic stem cells were processed individually and then integrated to perform a pseudotime analysis. All datasets were downloaded from NCBI's Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) or Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) repositories. The datasets and related citations are listed in Supplementary Table S1.

Read mapping and counting for gene expression quantification was performed as following: Single cell and bulk RNA-sequencing reads were mapped to the human or mouse genome (hg19, mm10) by using the short-read aligner GSNAP (Wu and Watanabe, 2005) and RefSeq transcript annotations. Gene count matrices (including a single gene count value per gene and cell or sample) were derived by counting all reads mapped to any transcript variant of a gene. For E-MTAB-3929 the gene count matrix downloaded from GEO was used directly. The microarray dataset GSE29397 was normalized using the Robust Multi-array Average (RMA) method.

Gene counts from all datasets were merged into a single matrix. The analysis was performed using Scanpy (Wolf et al., 2018) and additional python packages in the following versions: scanpy 1.0.4, anndata 0.6.1, numpy 1.13.3, scipy 0.19.1, pandas 0.22.0, scikit-learn 0.19.1, statsmodels 0.8.0, python-igraph 0.7.1, louvain 0.6.1. The following pre-processing and analysis steps were performed:

1. The merged gene count matrix was loaded as AnnData object together with the sample annotation.
2. For each single cell (or bulk sample, respectively) the number of counts was normalized to 10,000 counts per cell (or sample) using the scanpy.api.pp.normalize_per_cell function.
3. The normalized gene count matrix was filtered for variable genes based on average gene expression cut-offs (min_mean=0.1, max_mean=3) and minimal dispersion cut-off (min_disp=0) using the scanpy.api.pp.filter_genes_dispersion function.
4. The normalized and filtered gene count matrix was log-transformed using the scanpy.api.pp.log1p function.
5. To account for technical differences between the multiple datasets a linear regression method using the dataset name as categorical confounding variable was applied to the pre-processed gene count matrix using the scanpy.api.pp.regress_out function.
6. The normalized, filtered, log-transformed, and regressed-out gene count matrix was scaled to unit variance and zero mean using the scanpy.api.pp.scale function.
7. A principal component analysis was performed based on the pre-processed gene count matrix using the scanpy.api .tl.pca function.
8. A neighborhood graph was calculated based on the pre-processed gene count matrix with a local neighborhood (n_neighbors=30) and a Gaussian neighborhood function using the scanpy.api.pp.neighbors function.
9. A trajectory was calculated based on the neighborhood graph by applying diffusion pseudotime (DPT) that infers progression of cells through geodesic distance along the

graph using the scanpy.api.tl.dpt function without branching. The root was set to an oocyte cell from dataset GSE36552.

10. A pseudotime value between 0 and 1 is reported for each single cell (and bulk sample, respectively).

For the reported pseudotime analysis, only the human single cell and bulk RNA-sequencing datasets were used as indicated in Supplementary Table S1 because the microarray dataset GSE29397, the murine dataset GSE45719, and murine samples from dataset GSE80732 are too different to provide conclusive results in the integrated analysis.

7.1.2. miRNA processing

The sequences received in fastq format were processed using sRNAbench version 10/14 (http://bioinfo2.ugr.es:8080/ceUGR/srnabench/) with the following parameters: microRNA=hsa; recursiveAdapterTrimming=true; adapter=TGGAATTCTCGGGTGCCAAGGAACTCCAGTCAC; predict=true; isoMiR=true; noMM=0; alignType=n; adapterMinLength=10; maintainOrig=true; removeBarcode=0; minRC=2; adapterMM=1; seed=20; minReadLength=15. The processing included mapping to the human genome version hg19 with Bowtie (http://bowtie-bio.sourceforge.net), quantification of known microRNAs as annotated in miRBase (http://www.mirbase.org), and prediction of novel microRNAs. Sequence reads per microRNA were normalized by reads per million mapped reads (RPM) with the total number of genome mapped reads per sample. Quality control was applied including read statistics and unbiased clustering of counts and RPMs.

The software package edgeR (https://bioconductor.org/packages/release/bioc/html/edgeR.html) (Robinson et al., 2010) was used to identify differentially expressed miRNAs comparing the states primed to naïve, transitioning to naïve, and primed to transitioning. The cell line, HNES1 or HNES2, was included as co-variate in the design matrix.

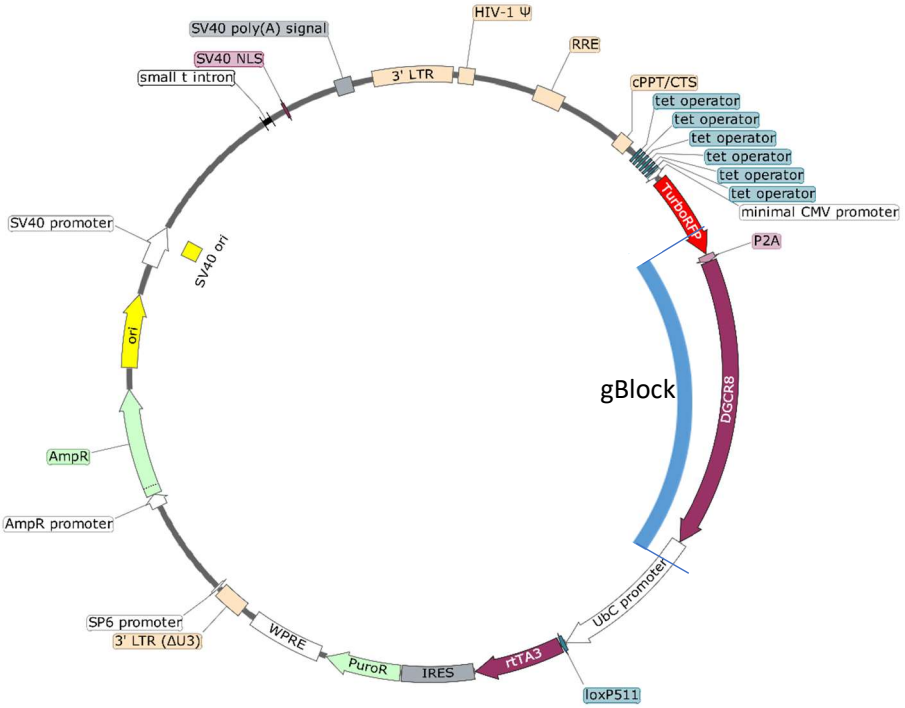
The same method was applied to the publicly available single cell sequencing experiment GSE81287 (Faridani et al., 2016). We compared the differential expression results between primed and naïve cells from this dataset and our own experiment by applying Pearson's correlation coefficient to the log2-fold-changes of both datasets.

Differential gene expression analysis using edgeR was also applied to the publicly available RNA-sequencing dataset E-MTAB-2857 (Takashima et al., 2014) comparing human pluripotent cells, which were reset to ground state pluripotency, to conventional pluripotent stem cells. For each miRNA quantified in our experiment the fraction of up- or down-regulated target genes in E-MTAB-2857 as well as the median log2-fold-changes of those target genes were calculated. MiRNA target genes were extracted from Thomson Reuters MetaBase v6.28. In addition, we extracted 10k clusters and miRNA families from miRBase release 21 (Kozomara and Griffiths-Jones, 2014).

7.1.3. Gene set enrichment

For each target gene we defined a gene set containing all miRNAs binding to that target gene as annotated in Thomson Reuters MetaBase v6.28. Gene set enrichment analysis (GSEA) (Subramanian et al., 2005, Mootha et al., 2003) was applied to these gene sets using differentially expressed miRNAs for three contrasts comparing the states primed to naïve, transitioning to naïve, and primed to transitioning. As input to GSEA the differentially expressed genes were ranked by signed log-p-values derived from the edgeR results. Default parameters were used with the exception that the minimal gene set size was set to two to account for target genes with few miRNA binders.

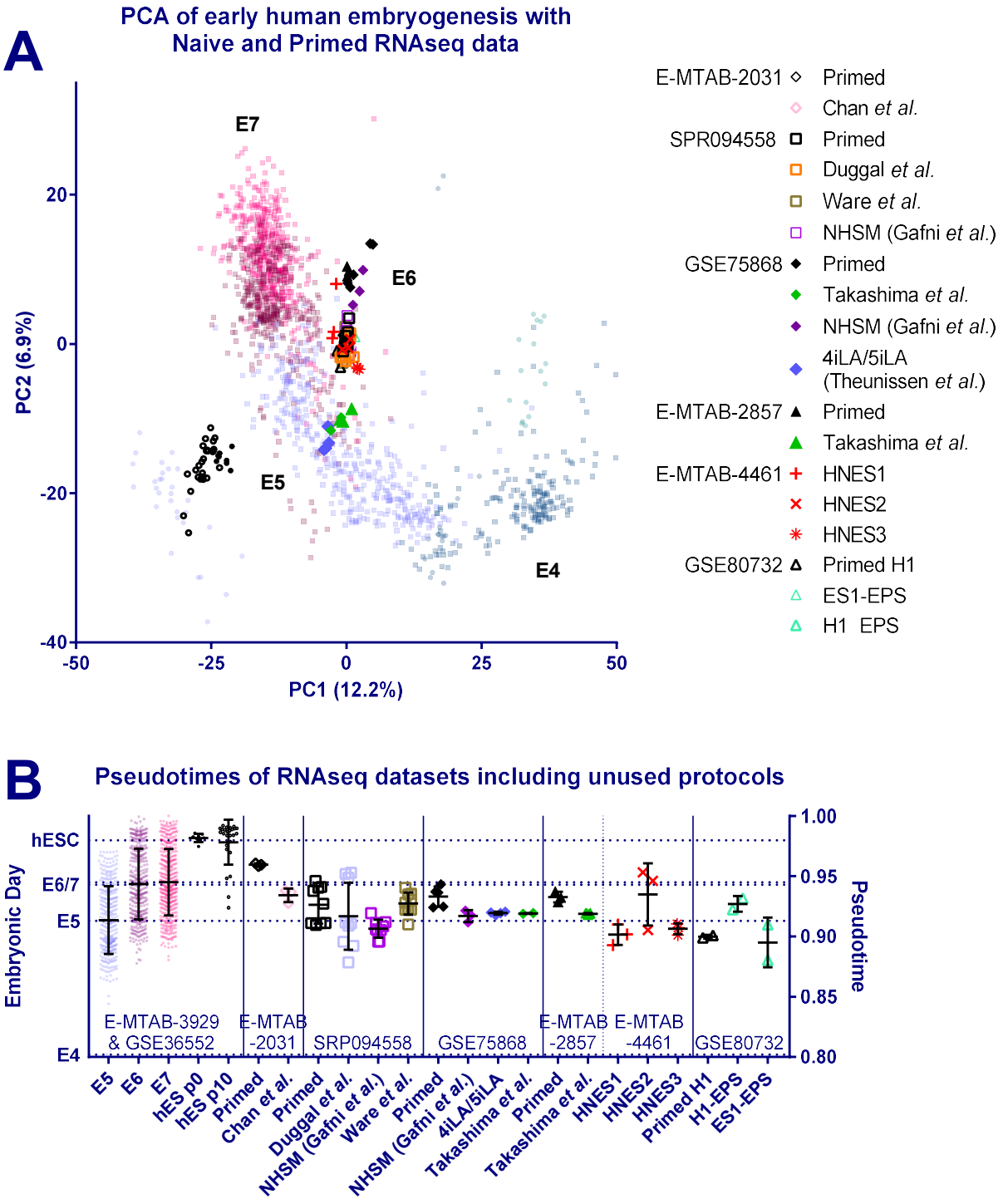
In addition, to obtain functional enrichment, we applied GSEA using the MiEAA web-interface (Backes et al., 2016) including the same three contrasts.



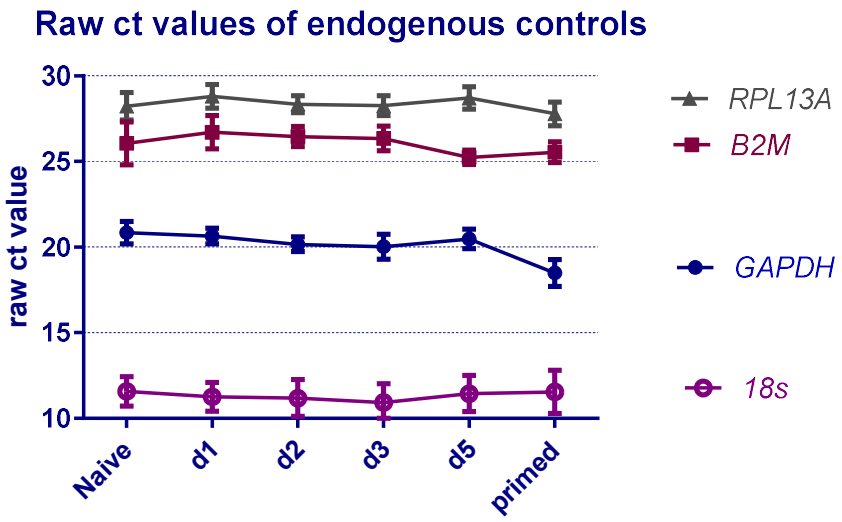
gBlock sequence:

agatctgctcgagggtcgac**cctgagg**gcttcacatgggagagaatcaccacatacgaagacgggggcgtgctgaccgctacccagga
caccagcttccagaacggctgcatcatctacaacgtcaagatcaacggggtgaacttcccatccaacggccctgtgatgcagaagaaa
acacgcggctgggaggccaacaccgagatgctgtaccccgctgacggcggcctgagaggccacagccagatggccctgaagctcgtgg
gcgggggctacctgcactgctccttcaagaccacatacagatccaagaaacccgctaagaacctcaagatgcccggcttccacttcgt
ggaccacagactggaaagaatcaaggaggccgacaaagagacctacgtcgagcagcacgagatggctgtggccaagtactgcgacctc
cctagcaaactggggcacagaggatctggagcaacaaacttctcactactcaaacaagcaggtgacgtggaggagaatcccgggcct**a**
tggagacagatgagagcccctctccgctcccgtgtgggcccgcaggagaagcggtgatggagagccgagctcgccccttccaagcgct
gccccgtgagcagtctccaccacctcccctgcaaacgtccagtggtgcagaggtaatggacgttggctctggtggtgatggacagtcc
gaactccctgctgaggacccctcaacttctacggagcttctcttctctccaaaggatccttctctaagggccgcctcctcatagacc
cgaactgtagtggccacagcccgcgcaccgcccggcacgcacctgcggtccggaagttctcccctgaccttaagttgcttaaggatgt
aaagattagcgtgagctttaccgagagctgcaggagtaaggacaggaaggtgctgtacacaggagcagagcgcgacgtgcgggcggag
tgcggtctgctccttagccctgtcagtggggacgtgcatgcttgtccctttggcgggagtgttggtgacggggtaggcataggggtg
agagtgctgataagaaggatgaggagaatgagctggatcaggaaaagagagtggagtatgcagtgctcgatgagttagaagattttac
tgacaatttggagctagatgaagaaggagcaggcgggttcacggctaaagcaatcgttcagagagacagagtggatgaagaggccttg
aatttcccctacgaggatgactttgacaacgatgtggatgctctgctggaagaaggccttttgtgcccccaaaaagaggcgaacagagg
aaaaatatggcggagacagcgaccatccgtccgatggagagacaagtgtgcagccgatgatgaccaagattaaaacagtgctcaaaag
tcgtggccgcccacctacagagccgctgcccgacgggtggatcatgacattccataactctggagtcccggtgtacctacacagagag
tctcgggtggtcacctggtccaggccatacttcttgggaacgggaagcatacggaaacacgaccctcctctgagtagcatcccttgtc
tgcattataagaaaatgaaggacaacgaggaacgggagcaaagcagtgacctcacccctagtggggatgtgtccccgtcaagcccct
gagccgatctgcagagctggagtttcccctggatgagcctgactctatgggtgctgacccggggcccccggacgagaaagacccacta
ggggctgaggcagcccctggggccctggggcaggtgaaggccaaagtcgaggtgtgcaaagatgaatccgttgatctcgaggaatttc
gaagctacctggagaagcgttttgactttgagcaagttactgtgaaaaaattcaggacttgggctgagcggcggcaattcaatcggga
aatgaagcggaagcaggcggagtccgagaggcccatcttgccagccaatcagaagctcattactttatcagtgcaagatgcacccaca
aagaaagagtttgttattaaccccaacgggaaatccgaggtctgcatcctgcacgagtacatgcagcgtgtcctcaaggtccgccctg
tctataatttctttgaatgtgagaacccaagtgagccttttggtgcctcggtgaccattgatggtgtgacttacggatctggaactgc
aagcagcaaaaaacttgcgaagaataaagctgcccgagctacactggaaatcctcatccctgactttgttaaacagacctctgaagag
aagcccaaagacagtgaagaactcgagtattttaaccacatcagcatcgaggactcgcgggtctacgagctgaccagcaaggctgggc
tgttgtctccatatcagatcctccacgagtgccttaaaagaaaccatgggatgggtgacacgtctatcaagtttgaagtggttcctgg
gaaaaaccagaagagtgaatacgtcatggcgtgtggcaagcacacagtgcgcgggtggtgtaagaacaagagagttggaaagcagtta
gcctcacagaagatccttcagctgctgcacccacatgtcaagaactgggggtctttactgcgcatgtatggccgtgagagcagcaaga
tggtcaagcaggagacatcggacaagagtgtgattgagctgcagcagtatgccaagaagaacaagcccaacctgcacatcctcagcaa
gctccaagaggagatgaagaggctagctgaggaaagggaggagactcgaaagaagcccaagatgtccattgtggcgtccgcccagcct
ggcggtgagcccctgtgcaccgtggacgtgtga**acgcgt**Gctcgaggctagcagatct

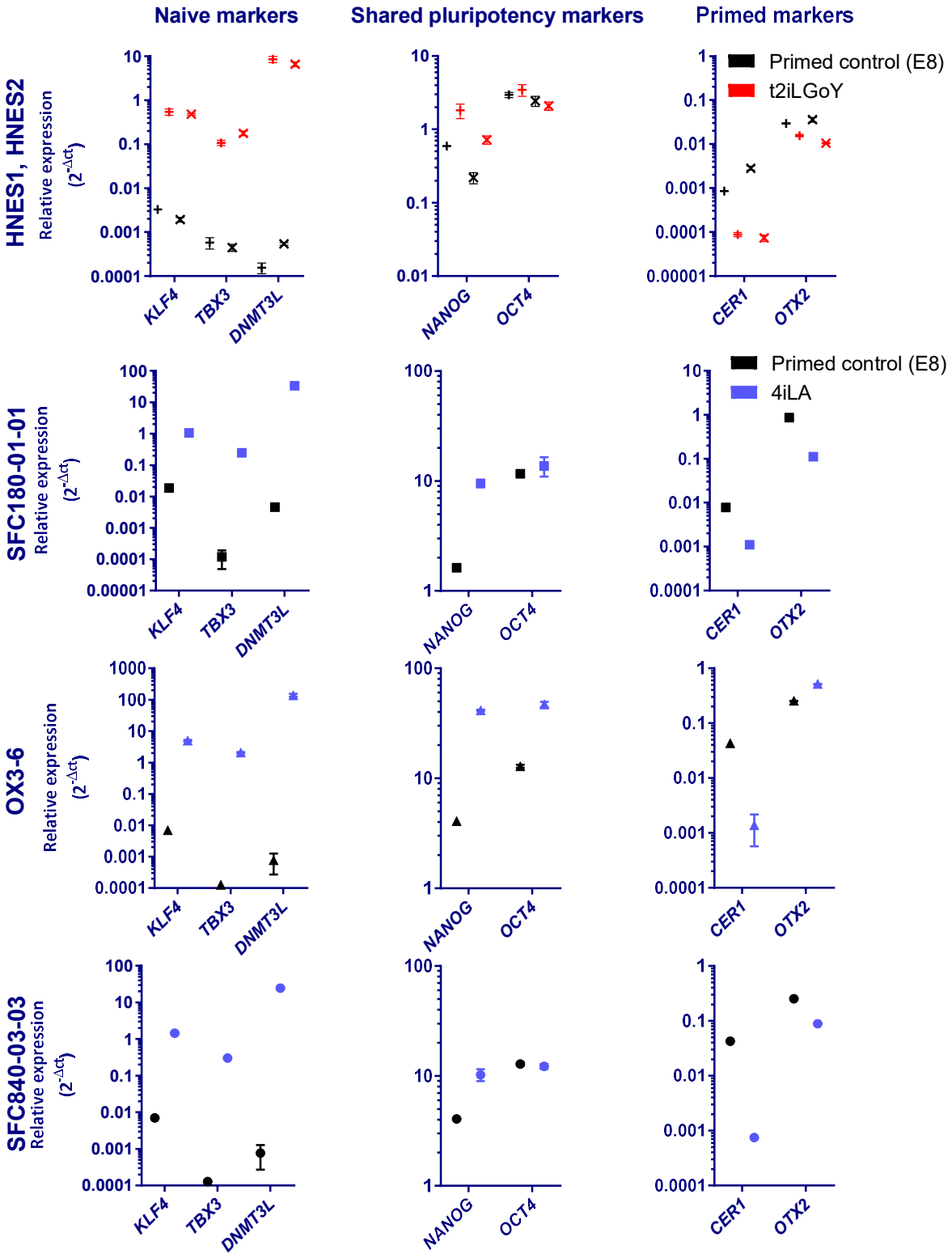
Supplementary Figure S1: Inducible *DGCR8* vector design. This lentiviral construct contains an inducible promotor, *TurboRFP* connected to a *2A* sequence and *DGCR8*. The gBlock has been designed to minimise cloning steps and contains a *Bsu36* I cut site (bold), the *TurboRFP* fragment (red), P2A sequence (blue), *DGCR8* (brown) and a *Mlu* I cut site (bold). Both sides contain overhangs to allow efficient cutting.



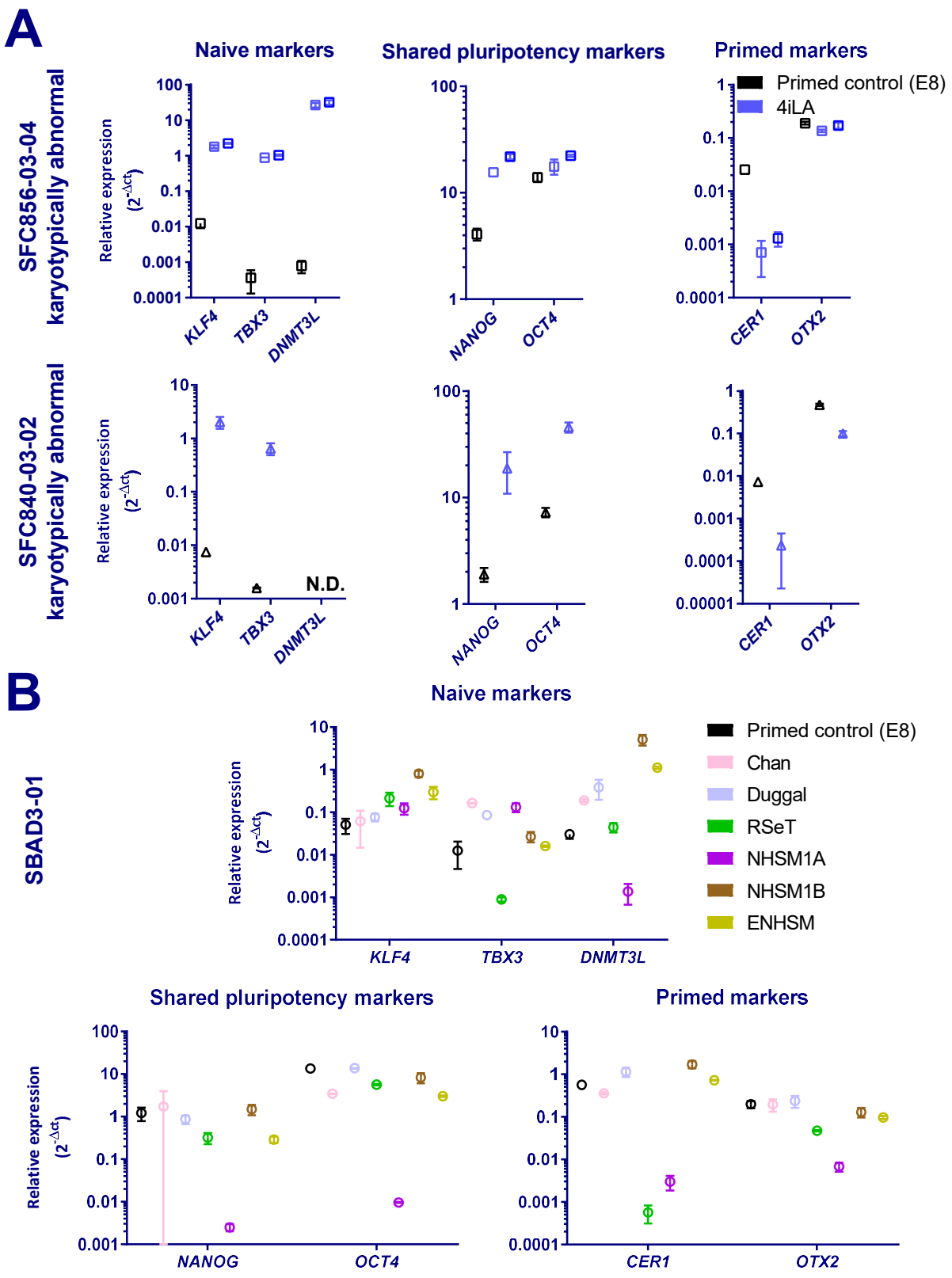
Supplementary Figure S2: Alignment of naïve cells generated by different protocols across human development. (A) Naïve cells are plotted according to their principle components (Figure 3.1). (B) Comparison of pseudotime of protocols with available RNAseq data.



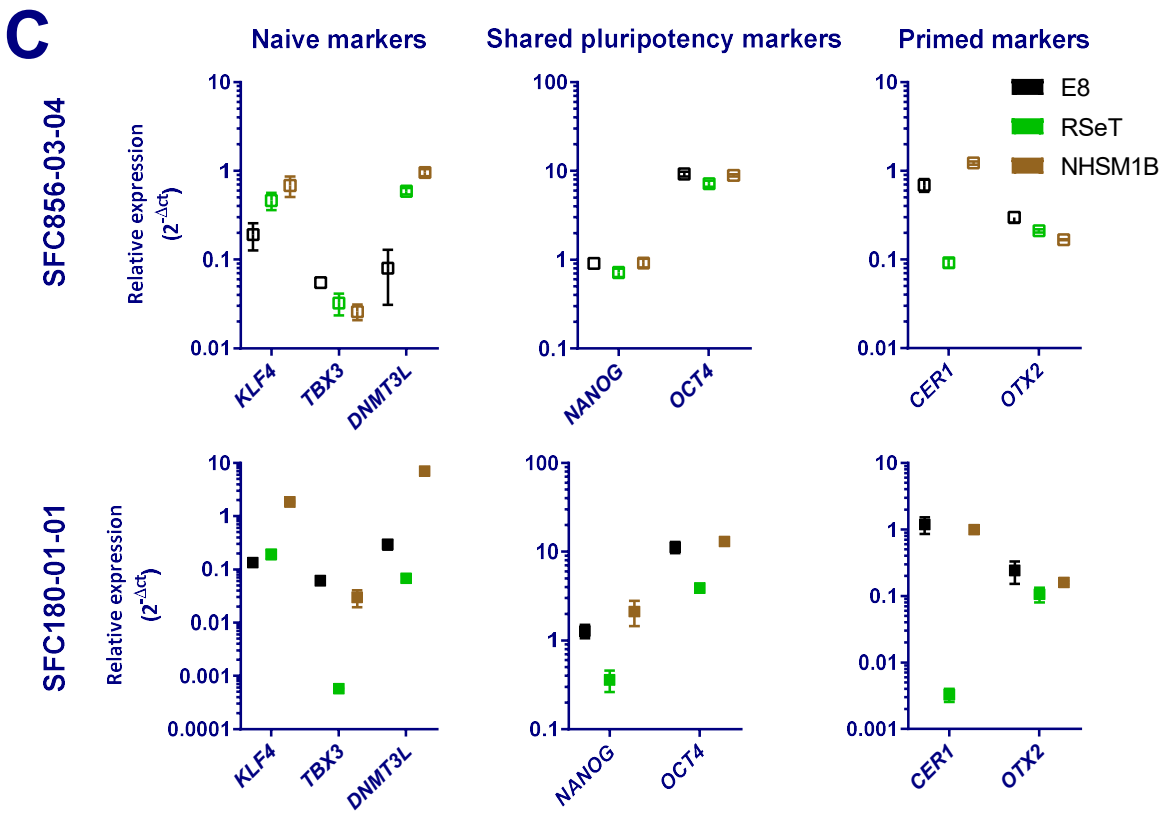
Supplementary Figure S3: Assessing genes suitable as endogenous controls for qRT-PCR. A suitable endogenous control is a housekeeping gene with stable expression which does not vary in the cell types of interest. Several common endogenous controls were tested across a time course of differentiation to exclude any which are differentially expressed. *GAPDH* has a lower Ct value in primed cells indicating an upregulation and will therefore not be used as an endogenous control. *RPL13A*, *B2M* and *18s* do not change their expression depending on cell type and are suitable endogenous controls used in this thesis. Each datapoint shows the average of 6 biological replicates (3x HNES1, 3x HNES2), which are each an average of 4 technical replicates. The error bars show the standard deviation of the 6 biological replicates.



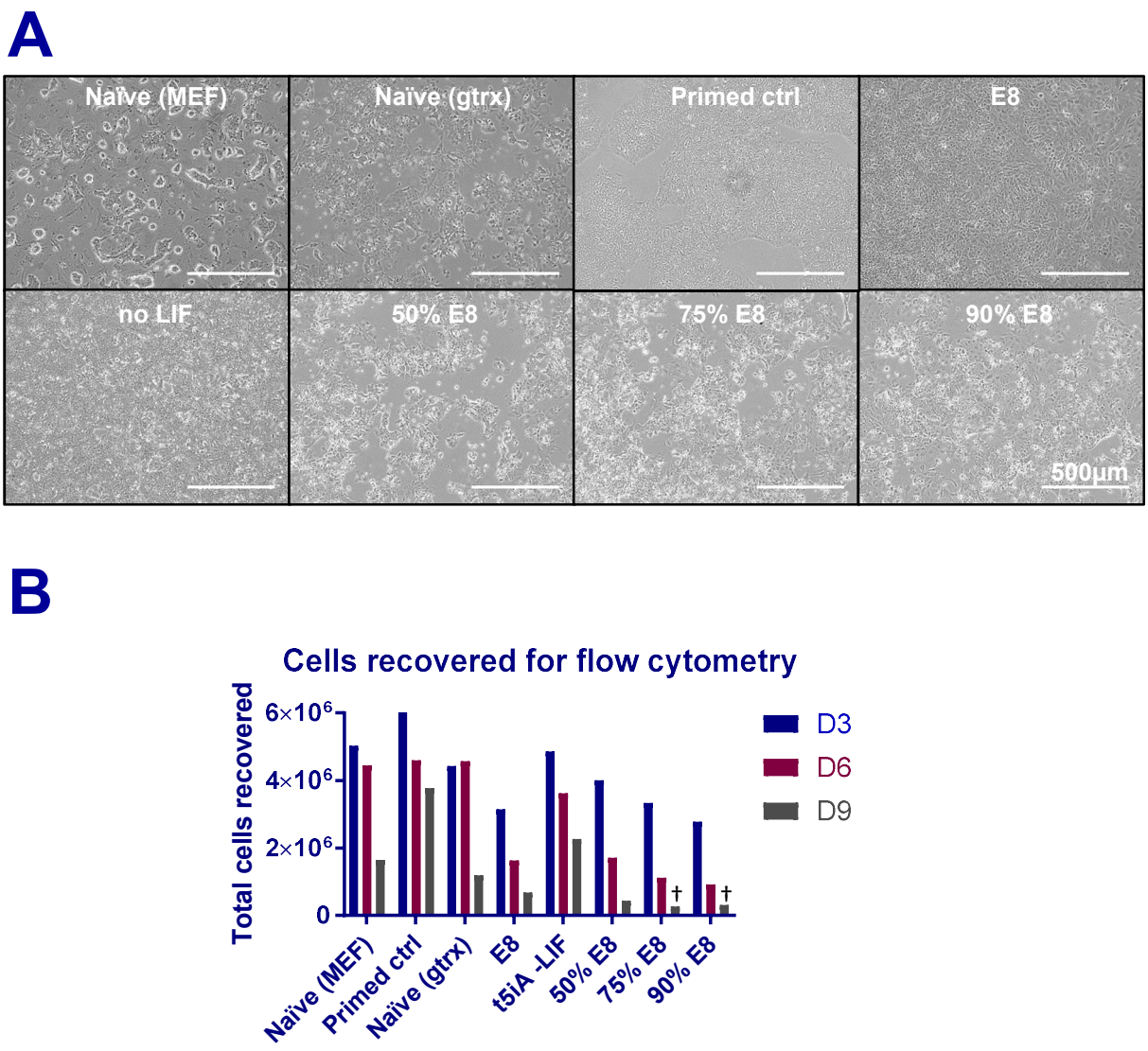
Supplementary Figure S4: Comparison of gene expression between primed controls and naïve cells (Figure 3.3). Data expressed as relative expression to housekeeping genes (*B2M, RPL13A*).



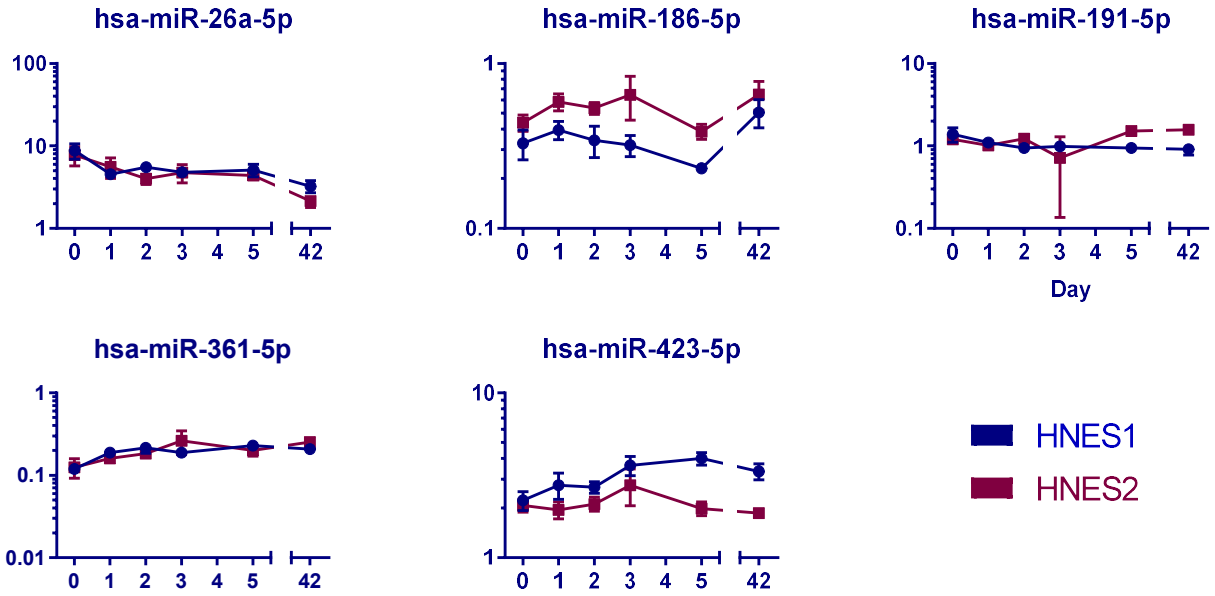
Supplementary Figure S5 continued on next page



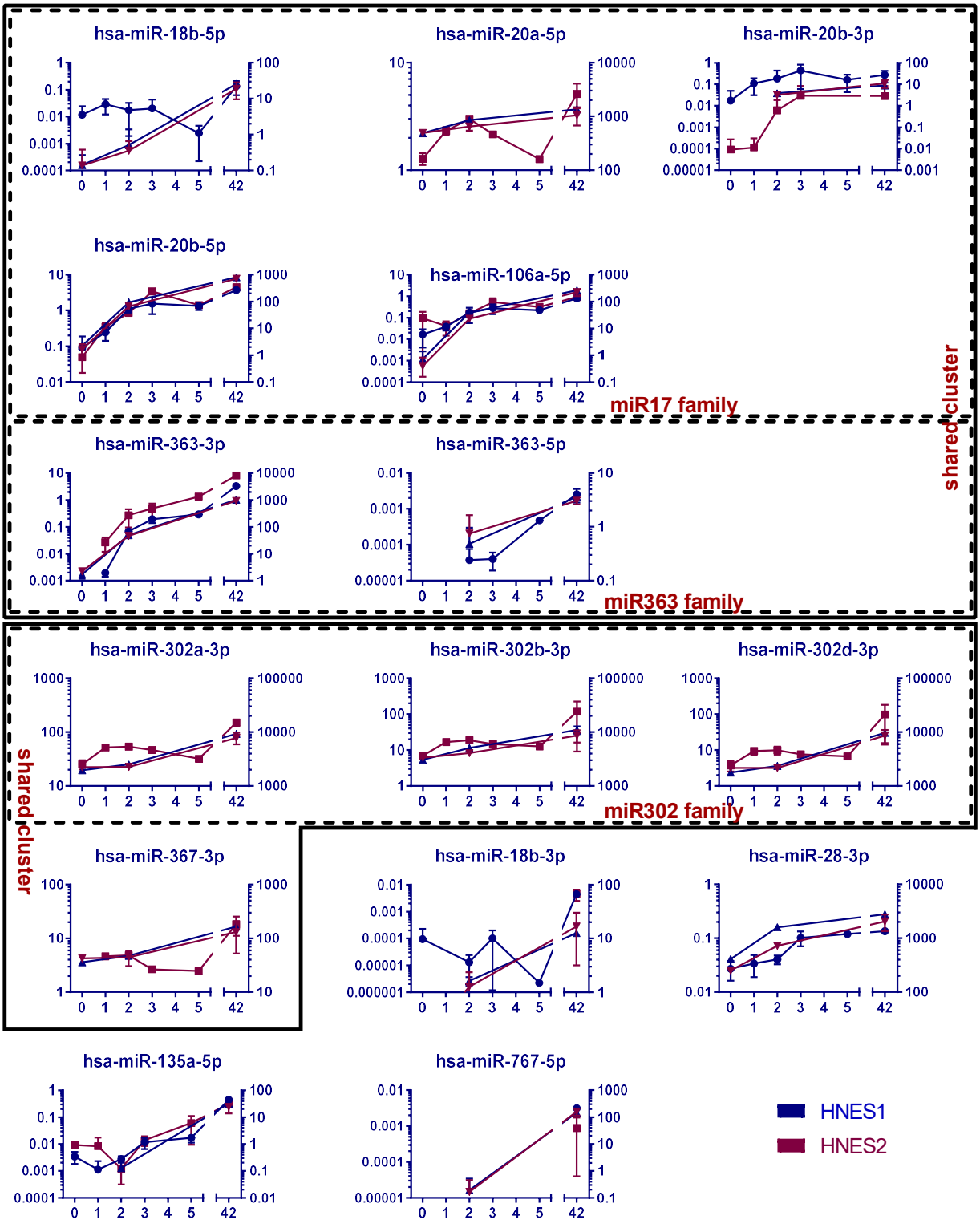
Supplementary Figure S5: Comparison of gene expression of unused protocols. Further information to Figure 3.3. Colour of the symbols refers to which media condition was used. (A) Gene expression of naive cells generated by Theunissen *et al.* protocol not used for further experiments due to abnormal karyotype. The abnormal SFC840-03-02 was only tested with an incomplete set of markers (*DNMT3L* was excluded due to a shortage of the primer/probe). (B) Comparison of naive conversion protocols not used for further experiments. (C) Replicates generated for RSeT and NHSM1B protocols. Relative expression to housekeeping genes (*B2M*, *RPL13A*).

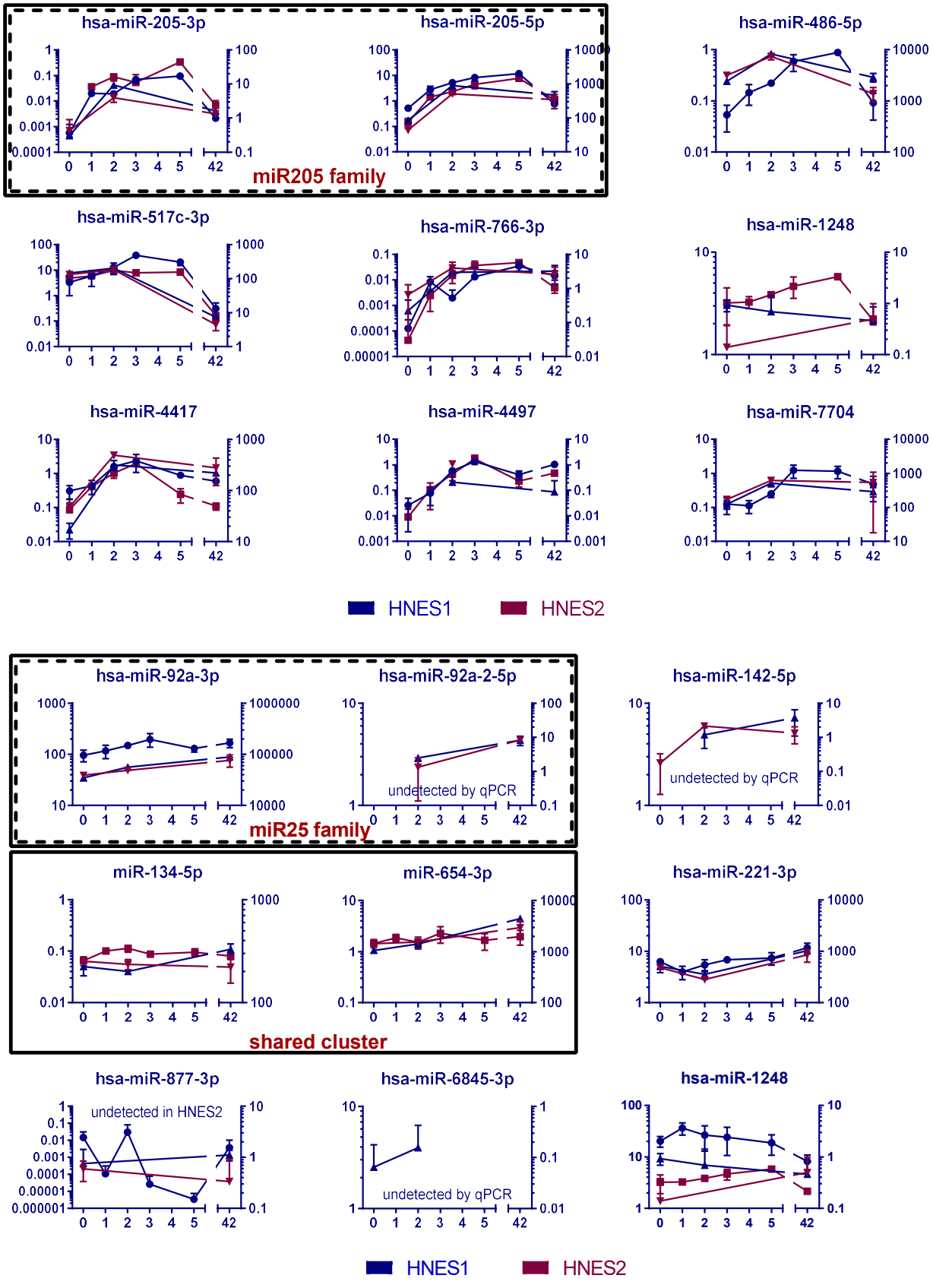


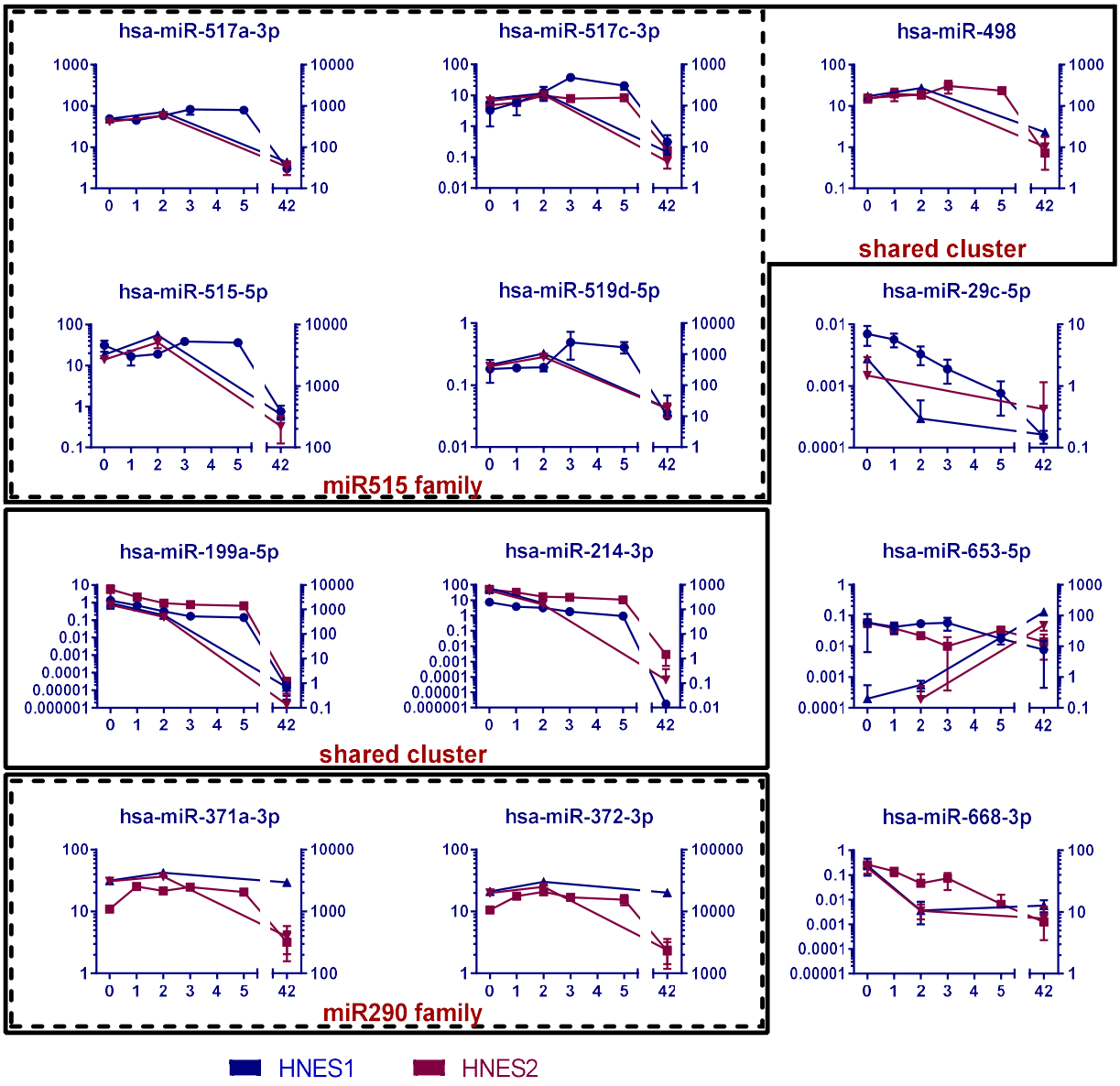
Supplementary Figure S6: Cells in a mixture of naïve medium and E8 differentiate and die. (A) Microscopy of all conditions at the final day 9 shows that naïve cells on geltrex appear flatter than their counterparts on MEFs. Cells in E8 medium are acquiring primed morphology. Floating dead cells and differentiating cells are visible in the E8/naïve medium mixtures. (B) Cell counts show low survival in E8 and mixtures of E8. The dagger (†) indicates inaccurate cell numbers due to too few cells for the cell counter.



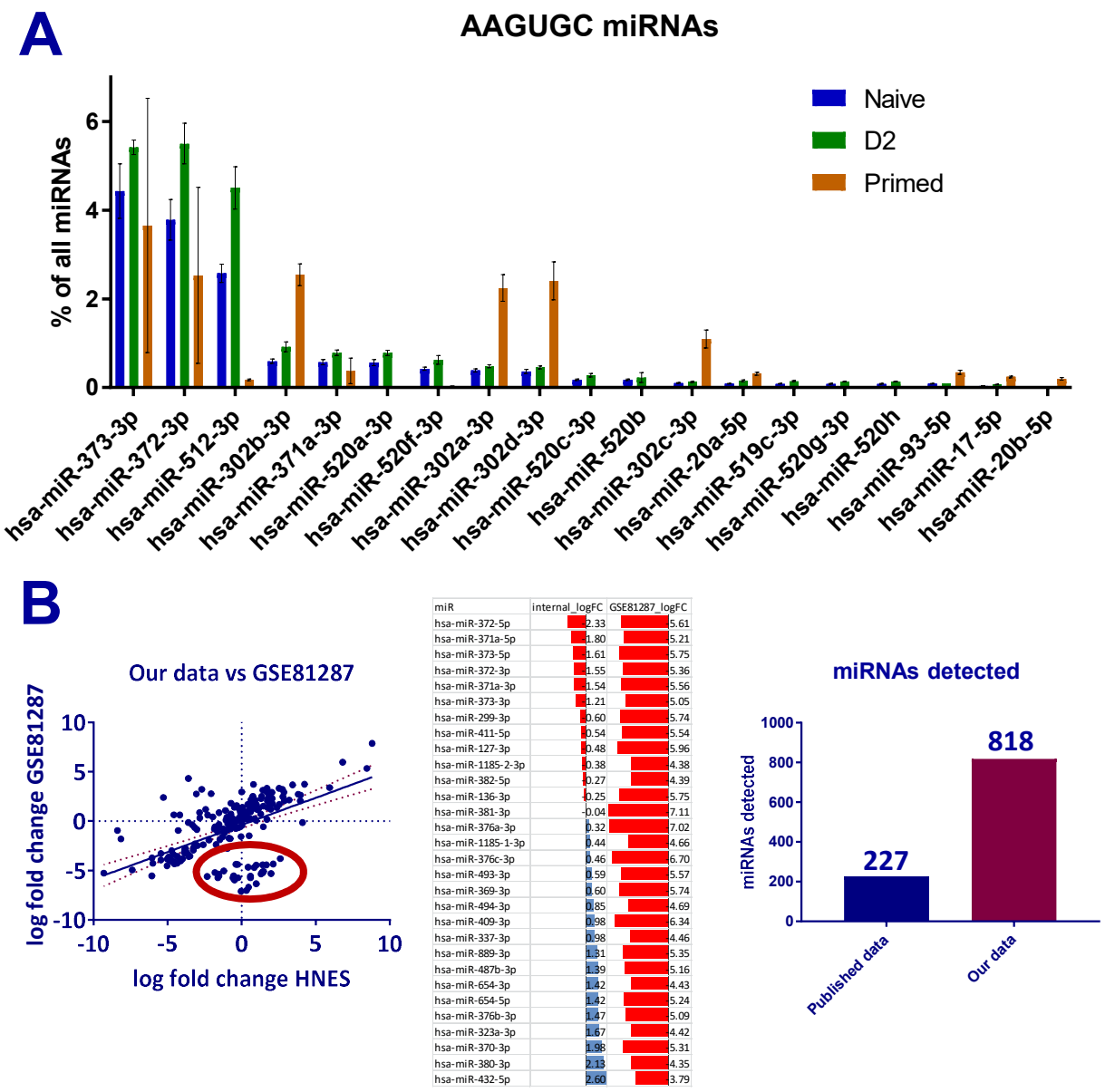
Supplementary Figure S7: The miRNA endogenous controls are consistently expressed across the differentiation time course. Expression of miRNAs normalised to the average of all five control miRNAs.


hsa-miR-18b-5p
hsa-miR-20a-5p
hsa-miR-20b-3p
hsa-miR-20b-5p
hsa-miR-106a-5p
miR17 family
shared cluster
hsa-miR-363-3p
hsa-miR-363-5p
miR363 family
hsa-miR-302a-3p
hsa-miR-302b-3p
hsa-miR-302d-3p
miR302 family
shared cluster
hsa-miR-367-3p
hsa-miR-18b-3p
hsa-miR-28-3p
hsa-miR-135a-5p
hsa-miR-767-5p
HNES1
HNES2


hsa-miR-205-3p
hsa-miR-205-5p
miR205 family
hsa-miR-486-5p
hsa-miR-517c-3p
hsa-miR-766-3p
hsa-miR-1248
hsa-miR-4417
hsa-miR-4497
hsa-miR-7704
HNES1
HNES2
hsa-miR-92a-3p
hsa-miR-92a-2-5p
undetected by qPCR
miR25 family
hsa-miR-142-5p
undetected by qPCR
miR-134-5p
miR-654-3p
hsa-miR-221-3p
shared cluster
hsa-miR-877-3p
undetected in HNES2
hsa-miR-6845-3p
undetected by qPCR
hsa-miR-1248
HNES1
HNES2

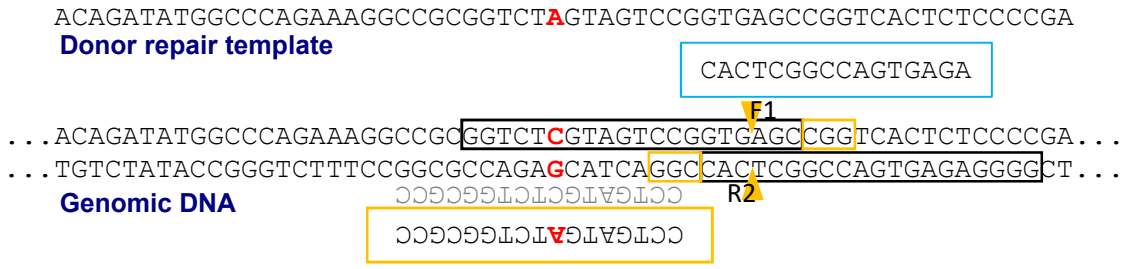


Supplementary Figure S8: Sequencing data correlates with qPCR results. The left Y axis represents $2^{-\Delta Ct}$ values obtained by qPCR, the right Y axis sequencing results expressed as reads per million (RPM) displayed as triangles. Sequencing was performed on samples from day 0, 2 and 42. Missing datapoints represent undetectable expression. The miRNAs hsa-miR-142-5p, hsa-miR-6845-3p, hsa-miR-92a-2-5p were not detected by qPCR but were picked up by sequencing. Most qPCR results correlate well with the sequencing data with the exceptions of the miRNAs hsa-miR-653-5p, hsa-miR-18b-5p, hsa-miR-18b-3p, hsa-miR-877-3p.

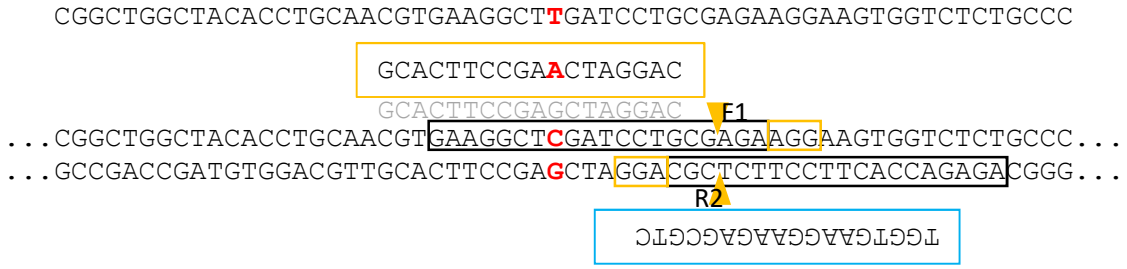


Supplementary Figure S9: Further analysis of miRNAseq data. (A) Different miRNAs with the seed sequence AAGUGC are highly expressed in naïve and primed states. The miRNA-373-3p and miRNA-372-3p are stably and highly expressed in naïve hPSC but very variably expressed in primed. miRNAs of the 302 family are expressed lowly in naïve cells and highly in primed. miRNAs with expression in all states below 0.1% are not included. The total percentage of ESCC miRNAs vary depending on state: in naïve HNES cells, ESCC miRNAs make up 15% of the total miRNAome. In differentiating cells, this increases to 20%, whereas primed cells are variable with HNES1 cells at 20% and HNES2 cells only 13%. (B) The comparison (performed by Klas Hatje) between our dataset and the published single cell RNAseq dataset using Theunissen *et al.* 5iLAF naïve cells largely shows an overlap with a Pearson's correlation coefficient of 0.52 indicating only minor differences between 5iLAF and HNES cells in terms of miRNA expression. The miRNAs which are not entirely in agreement are shown in the table. The log fold changes between naïve and primed hPSC are compared. The miRNAs not in agreement may seem artificially highly dysregulated due to the single cell small RNAseq method used by the authors, or it may be a difference between cell lines or naïve pluripotency protocols. Overall, our data is considerably more powerful than the previously published data. (Faridani 2016)

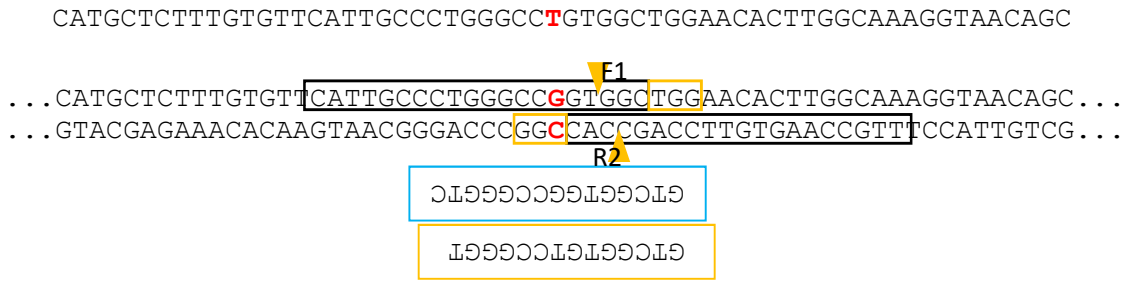
RBM20



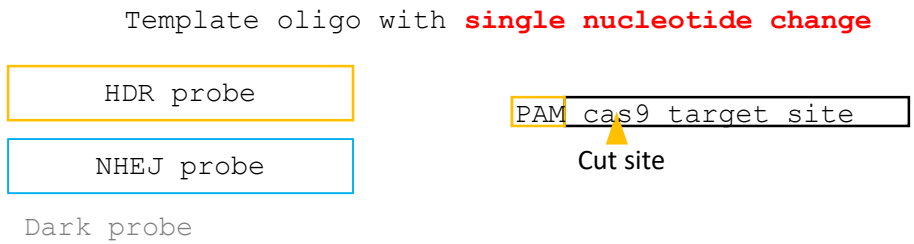
GRN



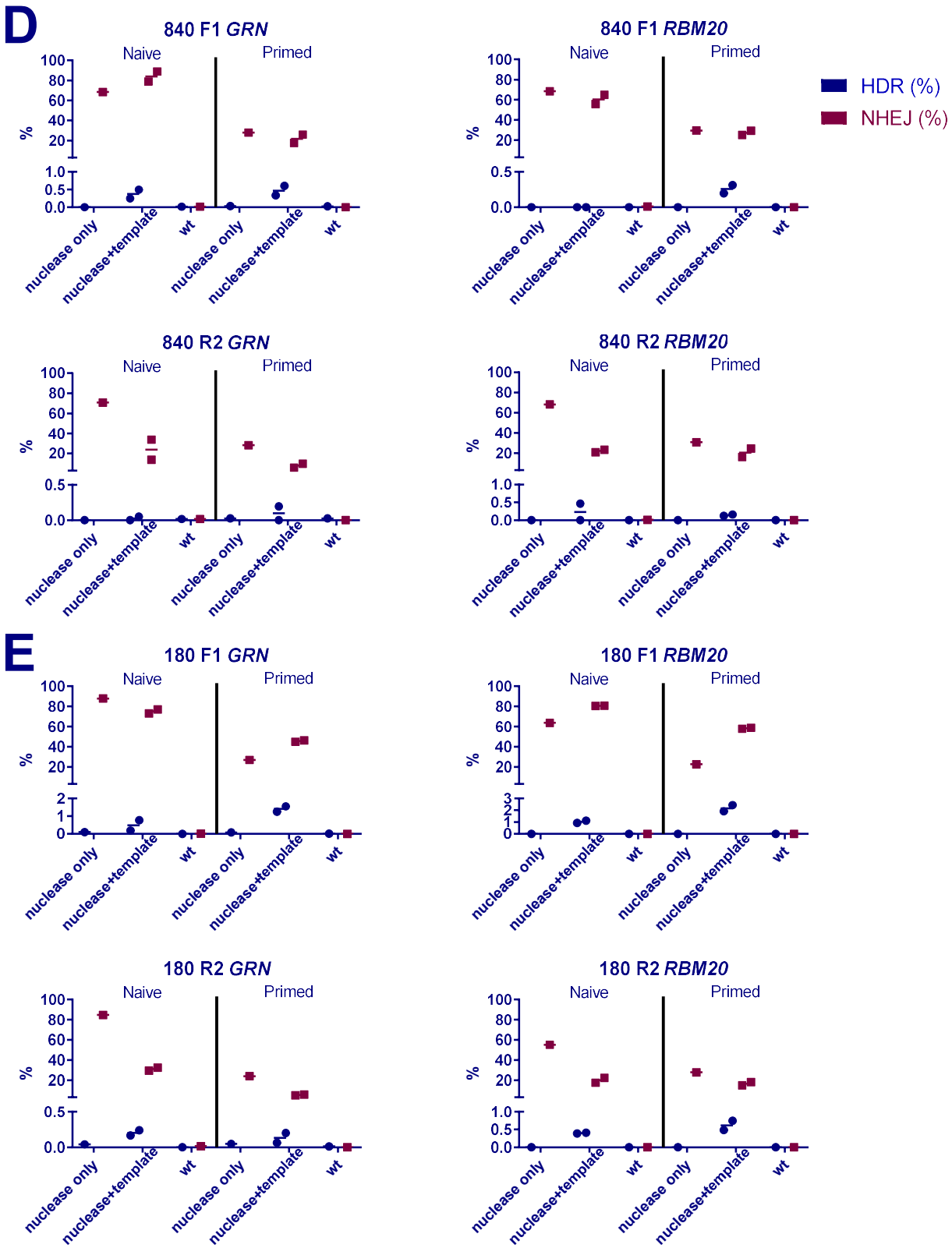
ATP7B



Legend

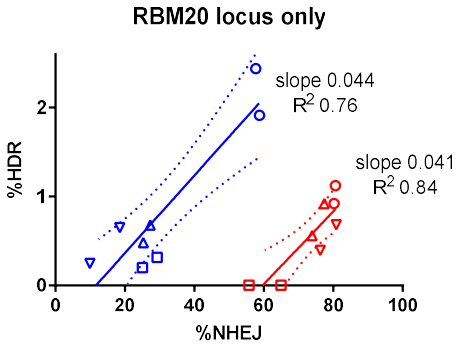


Supplementary Figure S10: Detailed design of probes. A dark probe is an oligo with a 3' phosphate to avoid priming and is used for *RBM20* and *GRN* to prevent false positive HDR probe binding but is not necessary for *ATP7B* due to the NHEJ probe competing for the same region. NHEJ probes span the Cas9 cut sites and are unable to bind after indel formation. The template oligos are 60 nucleotides in length with the single nucleotide change in the centre. Reference probes and primers (not shown) are further away from the central region and cannot bind the template strand.

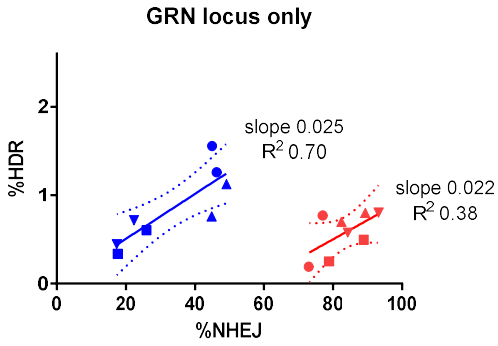


Supplementary Figure S11: Cas9 induced double strand breaks were predominantly repaired by NHEJ. (A) No HDR was observed for the ATP7B locus, even after increasing the stability of the template strand by PTO modification. Therefore this locus was not used for further experiments. (B,C,D,E) The underlying data used to calculate the normalised HDR rate is shown. The electroporation of nuclease + template conditions were performed in duplicate with the exception of OX3-6 R2 GRN due to cell death. Abbreviations: 856 (SFC856-03-04), 840 (SFC840-03-01), 180 (SFC180-01-01)

Are the slopes different between naive and primed cells?

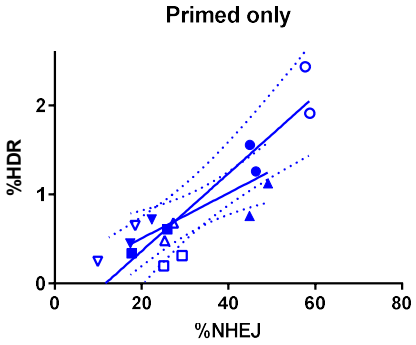


No, for RBM20 the slopes are not significantly different (p = 0.853). The pooled slope is 0.043

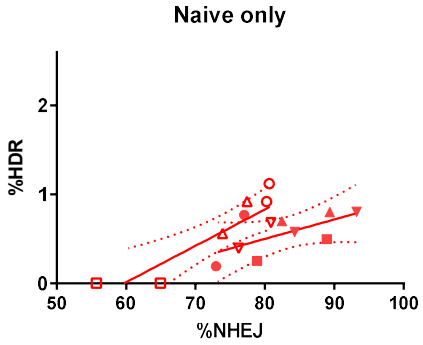


No, for GRN the slopes are not significantly different (p = 0.800). The pooled slope is 0.025.

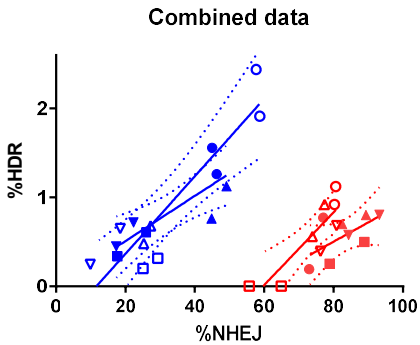
Are the slopes different between the loci RBM20 and GRN?



No, for primed cells the slopes are not significantly different (p = 0.118). The pooled slope is 0.037.



No, for naive cells the slopes are not significantly different (p = 0.224). The pooled slope is 0.034.



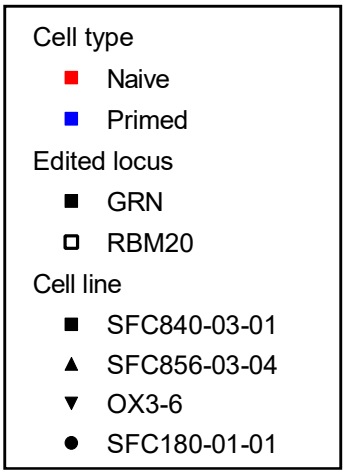


Figure S12: Relationship between %NHEJ and %HDR measurements allows calculation of a slope using linear regression. The colour of each symbol represents cell state, the shape indicates the cell line and the fill shows the locus. The pooled slope for RBM20 is 0.043 and for GRN is 0.025.

Dataset	Species	Cell type	In vivo / in vitro	Technology	Platform	N samples	Used for pseudotime	Citation	Description
E-MTAB-3929	Human	Primary cells	In vivo	Single-cell RNA-Seq	RNA-Seq (Illumina HiSeq 2000)	1529	x	Petropoulos et al. 2016	Single-cell RNA-seq of 1,529 cells obtained from 88 human preimplantation embryos ranging from embryonic day 3 to 7.
GSE36552	Human	Primary cells	In vivo	Single-cell RNA-Seq	RNA-Seq (Illumina HiSeq 2000)	124	x	Yan et al. 2013	Tracing pluripotency of human early embryos and embryonic stem cells by single cell RNA-seq.
GSE29397	Human	Primary cells	In vivo	Microarray	Microarray (Affymetrix Human Gene 1.0 ST Array)	24		Vassena et al. 2011	Gene expression profiling of human pre-implantation was followed with biological triplicates of oocyte to blastocyst stages, and compared with ESCs.
GSE45719	Mouse	Primary cells	In vivo	Single-cell RNA-Seq	RNA-Seq (Illumina HiSeq 2000)	317		Deng et al. 2014	We performed RNA-Sequencing on individual cells, from zygote to the cells of the late blastocyst, and also individual cells from the adult liver.
GSE75868 (GSE83765 + GSE85689)	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq (Illumina HiSeq 2500)	20	x	Theunissen et al. 2016	RNA-seq of 4 naive ES samples in 4i/LA, 3 naive ES samples in 5i/LA, 2 transgene-dependant naive ES cell samples, 5 primed ES cell samples (in hESM), 3 naive ES samples in NHSM (Gafni et al.) and 3 primed ES cell samples (in hESM)
E-MTAB-4461	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq (Illumina HiSeq 2500)	9	x	Guo et al. 2016	Transcriptional activity of HNES1, HNES2 and HNES3 cell lines was assessed with RNA-seq.
E-MTAB-2031	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq	6	x	Chanet al. 2013	We use a chemical approach (3iL: 3 small molecule inhibitor and cytokine) to induce an expression signature that more closely resembles native pluripotent cells. To better understand the difference between this hESC state and the conventional hESC state, we generated the global expression profiles using RNA-seq.
GSE80732_human	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq (Illumina HiSeq 2500, Illumina HiSeq 2000)	10	x	Yang et al. 2017	Global transcriptional analysis of human extended pluripotent stem cells, human primed pluripotent stem cells, mouse extended pluripotent stem cells, and mouse embryonic stem cells by RNA-seq.
GSE80732_mouse	Mouse	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq (Illumina HiSeq 2500, Illumina HiSeq 2000)	10		Yang et al. 2017	Global transcriptional analysis of human extended pluripotent stem cells, human primed pluripotent stem cells, mouse extended pluripotent stem cells, and mouse embryonic stem cells by RNA-seq.
SRP094558	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq (NextSeq 500)	36	x	Warrier et al. 2017	Several culture conditions have demonstrated successful conversion of primed human embryonic stem cells (hESC) towards a naïve state without pre-ectopic expression of naïve markers, such as naïve human conversion media (NHSM) or reverse toggle protocol (RT) and our published in-house developed naïve conversion media (NCM) as well using minimal growth factors/cytokines. We report here a direct comparative analysis of three in-house derived hESCs converted in the aforementioned naïve media.
E-MTAB-2857	Human	Cell line	In vitro	Bulk RNA-Seq	RNA-Seq	6	x	Takashima et al. 2014	Human pluripotent cells were reset to ground state pluripotency by transient overexpression of NANOG and KLF2 and subsequent inhibition of ERK and protein kinase C

Supplementary Table S1: Datasets used for *in silico* protocol comparison.

7.2. Bibliography

ALLEN, F., CREPALDI, L., ALSINET, C., STRONG, A. J., KLESHCHEVNIKOV, V., DE ANGELI, P., PALENIKOVA, P., KHODAK, A., KISELEV, V., KOSICKI, M., BASSETT, A. R., HARDING, H., GALANTY, Y., MUNOZ-MARTINEZ, F., METZAKOPIAN, E., JACKSON, S. P. & PARTS, L. 2018. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*.

ARNOULT, N., CORREIA, A., MA, J., MERLO, A., GARCIA-GOMEZ, S., MARIC, M., TOGNETTI, M., BENNER, C. W., BOULTON, S. J., SAGHATELIAN, A. & KARLSEDER, J. 2017. Regulation of DNA repair pathway choice in S and G2 phases by the NHEJ inhibitor CYREN. *Nature,* 549**,** 548-552.

ARZENANI, M. K., ZADE, A. E., MING, Y., VIJVERBERG, S. J., ZHANG, Z., KHAN, Z., SADIQUE, S., KALLENBACH, L., HU, L., VUKOJEVIC, V. & EKSTROM, T. J. 2011. Genomic DNA hypomethylation by histone deacetylase inhibition implicates DNMT1 nuclear dynamics. *Mol Cell Biol,* 31**,** 4119-28.

AUDEBERT, M., SALLES, B. & CALSOU, P. 2008. Effect of double-strand break DNA sequence on the PARP-1 NHEJ pathway. *Biochem Biophys Res Commun,* 369**,** 982-8.

AVIOR, Y., SAGI, I. & BENVENISTY, N. 2016. Pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Mol Cell Biol,* 17**,** 170-82.

BACKES, C., KHALEEQ, Q. T., MEESE, E. & KELLER, A. 2016. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res,* 44**,** W110-6.

BAEK, D., VILLEN, J., SHIN, C., CAMARGO, F. D., GYGI, S. P. & BARTEL, D. P. 2008. The impact of microRNAs on protein output. *Nature,* 455**,** 64-71.

BAKER, C. L. & PERA, M. F. 2018. Capturing Totipotent Stem Cells. *Cell Stem Cell,* 22**,** 25-34.

BAR, M., WYMAN, S. K., FRITZ, B. R., QI, J., GARG, K. S., PARKIN, R. K., KROH, E. M., BENDORAITE, A., MITCHELL, P. S., NELSON, A. M., RUZZO, W. L., WARE, C., RADICH, J. P., GENTLEMAN, R., RUOHOLA-BAKER, H. & TEWARI, M. 2008. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells,* 26**,** 2496-505.

BARROSO-DELJESUS, A., ROMERO-LOPEZ, C., LUCENA-AGUILAR, G., MELEN, G. J., SANCHEZ, L., LIGERO, G., BERZAL-HERRANZ, A. & MENENDEZ, P. 2008. Embryonic stem cell-specific miR302-367 cluster: human gene structure and functional characterization of its core promoter. *Mol Cell Biol,* 28**,** 6609-19.

BARTEL, D. P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell,* 136**,** 215-33.

BASSETT, A. R., AZZAM, G., WHEATLEY, L., TIBBIT, C., RAJAKUMAR, T., MCGOWAN, S., STANGER, N., EWELS, P. A., TAYLOR, S., PONTING, C. P., LIU, J. L., SAUKA-SPENGLER, T. & FULGA, T. A. 2014. Understanding functional miRNA-target interactions in vivo by site-specific genome engineering. *Nat Commun,* 5**,** 4640.

BAZZINI, A. A., LEE, M. T. & GIRALDEZ, A. J. 2012. Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science,* 336**,** 233-237.

BEERS, J., GULBRANSON, D. R., GEORGE, N., SINISCALCHI, L. I., JONES, J., THOMSON, J. A. & CHEN, G. 2012. Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions. *Nat Protoc,* 7**,** 2029-40.

BENTLEY, J., DIGGLE, C. P., HARNDEN, P., KNOWLES, M. A. & KILTIE, A. E. 2004. DNA double strand break repair in human bladder cancer is error prone and involves microhomology-associated end-joining. *Nucleic Acids Res,* 32**,** 5249-59.

BENTWICH, I., AVNIEL, A., KAROV, Y., AHARONOV, R., GILAD, S., BARAD, O., BARZILAI, A., EINAT, P., EINAV, U., MEIRI, E., SHARON, E., SPECTOR, Y. & BENTWICH, Z. 2005. Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet,* 37**,** 766-70.

BERNSTEIN, E., KIM, S. Y., CARMELL, M. A., MURCHISON, E. P., ALCORN, H., LI, M. Z., MILLS, A. A., ELLEDGE, S. J., ANDERSON, K. V. & HANNON, G. J. 2003. Dicer is essential for mouse development. *Nat Genet,* 35**,** 215-7.

BETERMIER, M., BERTRAND, P. & LOPEZ, B. S. 2014. Is non-homologous end-joining really an inherently error-prone process? *PLoS Genet,* 10**,** e1004086.

BETSCHINGER, J., NICHOLS, J., DIETMANN, S., CORRIN, P. D., PADDISON, P. J. & SMITH, A. 2013. Exit from pluripotency is gated by intracellular redistribution of the bHLH transcription factor Tfe3. *Cell,* 153**,** 335-47.

BEUCHER, A., BIRRAUX, J., TCHOUANDONG, L., BARTON, O., SHIBATA, A., CONRAD, S., GOODARZI, A. A., KREMPLER, A., JEGGO, P. A. & LOBRICH, M. 2009. ATM and Artemis promote homologous recombination of radiation-induced DNA double-strand breaks in G2. *EMBO J,* 28**,** 3413-27.

BIZARD, A. H. & HICKSON, I. D. 2014. The Dissolution of Double Holliday Junctions. *Cold Spring Harbor Perspectives in Biology,* 6.

BLAKELEY, P., FOGARTY, N. M. E., DEL VALLE, I., WAMAITHA, S. E., HU, T. X., ELDER, K., SNELL, P., CHRISTIE, L., ROBSON, P. & NIAKAN, K. K. 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development,* 142**,** 3151-3165.

BLASCHKE, K., EBATA, K. T., KARIMI, M. M., ZEPEDA-MARTINEZ, J. A., GOYAL, P., MAHAPATRA, S., TAM, A., LAIRD, D. J., HIRST, M., RAO, A., LORINCZ, M. C. & RAMALHO-SANTOS, M. 2013. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature,* 500**,** 222-6.

BOROVIAK, T., LOOS, R., BERTONE, P., SMITH, A. & NICHOLS, J. 2014. The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification. *Nat Cell Biol,* 16**,** 516-28.

BOYER, L. A., LEE, T. I., COLE, M. F., JOHNSTONE, S. E., LEVINE, S. S., ZUCKER, J. P., GUENTHER, M. G., KUMAR, R. M., MURRAY, H. L., JENNER, R. G., GIFFORD, D. K., MELTON, D. A., JAENISCH, R. & YOUNG, R. A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell,* 122**,** 947-56.

BRINKMAN, E. K., CHEN, T., DE HAAS, M., HOLLAND, H. A., AKHTAR, W. & VAN STEENSEL, B. 2018. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell,* 70**,** 801-813 e6.

BRONS, I. G. M., SMITHERS, L. E., TROTTER, M. W. B., RUGG-GUNN, P., SUN, B., CHUVA DE SOUSA LOPES, S. M., HOWLETT, S. K., CLARKSON, A., AHRLUND-RICHTER, L., PEDERSEN, R. A. & VALLIER, L. 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature,* 448**,** 191-195.

BROWN, S. D. M., HOLMES, C. C., MALLON, A. M., MEEHAN, T. F., SMEDLEY, D. & WELLS, S. 2018. High-throughput mouse phenomics for characterizing mammalian gene function. *Nat Rev Genet,* 19**,** 357-370.

BUECKER, C., CHEN, H.-H., POLO, J. M., DAHERON, L., BU, L., BARAKAT, T. S., OKWIEKA, P., PORTER, A., GRIBNAU, J., HOCHEDLINGER, K. & GEIJSEN, N. 2010. A Murine ESC-like State Facilitates Transgenesis and Homologous Recombination in Human Pluripotent Stem Cells. *Cell Stem Cell,* 6**,** 535-546.

BUECKER, C., SRINIVASAN, R., WU, Z., CALO, E., ACAMPORA, D., FAIAL, T., SIMEONE, A., TAN, M., SWIGUT, T. & WYSOCKA, J. 2014. Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. *Cell Stem Cell,* 14**,** 838-853.

BUNTING, S. F., CALLEN, E., WONG, N., CHEN, H. T., POLATO, F., GUNN, A., BOTHMER, A., FELDHAHN, N., FERNANDEZ-CAPETILLO, O., CAO, L., XU, X., DENG, C. X., FINKEL, T., NUSSENZWEIG, M., STARK, J. M. & NUSSENZWEIG, A. 2010. 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell,* 141**,** 243-54.

CALABRESE, J. M., SEILA, A. C., YEO, G. W. & SHARP, P. A. 2007. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc Natl Acad Sci U S A,* 104**,** 18097-102.

CARTHEW, R. W. & SONTHEIMER, E. J. 2009. Origins and Mechanisms of miRNAs and siRNAs. *Cell,* 136**,** 642-55.

CECCALDI, R., LIU, J. C., AMUNUGAMA, R., HAJDU, I., PRIMACK, B., PETALCORIN, M. I., O'CONNOR, K. W., KONSTANTINOPOULOS, P. A., ELLEDGE, S. J., BOULTON, S. J., YUSUFZAI, T. & D'ANDREA, A. D. 2015. Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. *Nature,* 518**,** 258-62.

CECCALDI, R., RONDINELLI, B. & D'ANDREA, A. D. 2016a. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol,* 26**,** 52-64.

CECCALDI, R., SARANGI, P. & D'ANDREA, A. D. 2016b. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol,* 17**,** 337-49.

CHAN, S. H., YU, A. M. & MCVEY, M. 2010. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet,* 6**,** e1001005.

CHAN, Y.-S., GOEKE, J., NG, J.-H., LU, X., GONZALES, K. A. U., TAN, C.-P., TNG, W.-Q., HONG, Z.-Z., LIM, Y.-S. & NG, H.-H. 2013. Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. *Cell Stem Cell,* 13**,** 663-675.

CHAUDHURI, A. A., SO, A. Y., MEHTA, A., MINISANDRAM, A., SINHA, N., JONSSON, V. D., RAO, D. S., O'CONNELL, R. M. & BALTIMORE, D. 2012. Oncomir miR-125b regulates hematopoiesis by targeting the gene Lin28A. *Proc Natl Acad Sci U S A,* 109**,** 4233-8.

CHEN, F. Z., HAI-HONG; ZHOU, LIN-FU; WU, SHAN-SHAN; WANG, JING; CHEN, ZHI 2010. Inhibition of c-FLIP expression by miR-512-3p contributes to Taxol-induced apoptosis in hepatocellular carcinoma cells. *Oncology Reports,* 23.

CHEN, H., GUO, R., ZHANG, Q., GUO, H., YANG, M., WU, Z., GAO, S., LIU, L. & CHEN, L. 2015. Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. *Proceedings of the National Academy of Sciences,* 112**,** E5936-E5943.

CHEN, J., WANG, M., GUO, M., XIE, Y. & CONG, Y. S. 2013. miR-127 regulates cell proliferation and senescence by targeting BCL6. *PLoS One,* 8**,** e80266.

CHIRUVELLA, K. K., LIANG, Z. & WILSON, T. E. 2013. Repair of double-strand breaks by end joining. *Cold Spring Harb Perspect Biol,* 5**,** a012757.

COLLIER, A. J., PANULA, S. P., SCHELL, J. P., CHOVANEC, P., PLAZA REYES, A., PETROPOULOS, S., CORCORAN, A. E., WALKER, R., DOUAGI, I., LANNER, F. & RUGG-GUNN, P. J. 2017. Comprehensive Cell Surface Protein Profiling Identifies Specific Markers of Human Naive and Primed Pluripotent States. *Cell Stem Cell*.

COLLIER, A. J. & RUGG-GUNN, P. J. 2018. Identifying Human Naive Pluripotent Stem Cells - Evaluating State-Specific Reporter Lines and Cell-Surface Markers. *Bioessays,* 40**,** e1700239.

CONCEPCION, C. P., BONETTI, C. & VENTURA, A. 2012. The microRNA-17-92 family of microRNA clusters in development and disease. *Cancer J,* 18**,** 262-7.

CONSORTIUM, S. M.-I. 2014. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol,* 32**,** 903-14.

CORDES, K. R., SHEEHY, N. T., WHITE, M. P., BERRY, E. C., MORTON, S. U., MUTH, A. N., LEE, T. H., MIANO, J. M., IVEY, K. N. & SRIVASTAVA, D. 2009. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature,* 460**,** 705-10.

CUI, Q., YU, Z., PURISIMA, E. O. & WANG, E. 2006. Principles of microRNA regulation of a human cellular signaling network. *Mol Syst Biol,* 2**,** 46.

DAHÉRON, L., OPITZ, S. L., ZAEHRES, H., LENSCH, W. M., ANDREWS, P. W., ITSKOVITZ-ELDOR, J. & DALEY, G. Q. 2004. LIF/STAT3 Signaling Fails to Maintain Self-Renewal of Human Embryonic Stem Cells. *STEM CELLS,* 22**,** 770-778.

DAVIS, L. & MAIZELS, N. 2016. Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep,* 17**,** 1872-1881.

DE LOS ANGELES, A., LOH, Y. H., TESAR, P. J. & DALEY, G. Q. 2012. Accessing naive human pluripotency. *Current Opinion in Genetics & Development,* 22**,** 272-282.

DEMSAR J, C. T., ERJAVEC A, GORUP C, HOCEVAR T, MILUTINOVIC M, MOZINA M, POLAJNAR M, TOPLAK M, STARIC A, STAJDOHAR M, UMEK L, ZAGAR L, ZBONTAR J, ZITNIK M, ZUPAN B 2013. Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research,* 14**,** 2349-2353.

DI STEFANO, B. & HOCHEDLINGER, K. 2018. Reduced MEK inhibition preserves genomic stability in naïve human ES cells.

DIECKMANN, K. P., RADTKE, A., SPIEKERMANN, M., BALKS, T., MATTHIES, C., BECKER, P., RUF, C., OING, C., OECHSLE, K., BOKEMEYER, C., HAMMEL, J., MELCHIOR, S., WOSNIOK, W. & BELGE, G. 2017. Serum Levels of MicroRNA miR-371a-3p: A Sensitive and Specific New Biomarker for Germ Cell Tumours. *Eur Urol,* 71**,** 213-220.

DJURANOVIC, S., NAHVI, A. & GREEN, R. 2012. miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. *Science,* 336**,** 237-240.

DODSWORTH, B. T., FLYNN, R. & COWLEY, S. A. 2015. The Current State of Naïve Human Pluripotency. *STEM CELLS,* 33**,** 3181-3186.

DUGGAL, G., WARRIER, S., GHIMIRE, S., BROEKAERT, D., DEROO, T., VAN DER JEUGHT, M., LIERMAN, S., DEROO, T., PEELMAN, L., VAN SOOM, A., CORNELISSEN, R., MENTEN, B., MESTDAGH, P., VANDESOMPELE, J., ROOST, M., SLIEKER, R. C., HEIJMANS, B. T., DEFORCE, D., DE SUTTER, P., DE SOUSA LOPES, S. C. & HEINDRYCKX, B. 2015. Alternative routes to induce naïve pluripotency in human embryonic stem cells. *Stem Cells*.

DUNN, S. J., MARTELLO, G., YORDANOV, B., EMMOTT, S. & SMITH, A. G. 2014. Defining an essential transcription factor program for naive pluripotency. *Science,* 344**,** 1156-1160.

EICHHORN, S. W., GUO, H., MCGEARY, S. E., RODRIGUEZ-MIAS, R. A., SHIN, C., BAEK, D., HSU, S. H., GHOSHAL, K., VILLEN, J. & BARTEL, D. P. 2014. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell,* 56**,** 104-15.

ENDOH, M., ENDO, T. A., ENDOH, T., FUJIMURA, Y., OHARA, O., TOYODA, T., OTTE, A. P., OKANO, M., BROCKDORFF, N., VIDAL, M. & KOSEKI, H. 2008. Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development,* 135**,** 1513-24.

ERCOLANI, L., FLORENCE, B., DENARO, M. & ALEXANDER, M. 1988. Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. *Journal of Biological Chemistry,* 263**,** 15335-41.

EULALIO, A., HUNTZINGER, E., NISHIHARA, T., REHWINKEL, J., FAUSER, M. & IZAURRALDE, E. 2009. Deadenylation is a widespread effect of miRNA regulation. *RNA,* 15**,** 21-32.

EVANS, M. J. & KAUFMAN, M. H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature,* 292**,** 154-156.

FABIAN, M. R. & SONENBERG, N. 2012. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol,* 19**,** 586-93.

FACTOR, D. C., CORRADIN, O., ZENTNER, G. E., SAIAKHOVA, A., SONG, L., CHENOWETH, J. G., MCKAY, R. D., CRAWFORD, G. E., SCACHERI, P. C. & TESAR, P. J. 2014. Epigenomic Comparison Reveals Activation of "Seed" Enhancers during Transition from Naive to Primed Pluripotency. *Cell Stem Cell,* 14**,** 854-863.

FALCK, J., FORMENT, J. V., COATES, J., MISTRIK, M., LUKAS, J., BARTEK, J. & JACKSON, S. P. 2012. CDK targeting of NBS1 promotes DNA-end resection, replication restart and homologous recombination. *EMBO Rep,* 13**,** 561-8.

FARH, K. K.-H., GRIMSON, A., JAN, C., LEWIS, B. P., JOHNSTON, W. K., LIM, L. P., BURGE, C. B. & BARTEL, D. P. 2005. The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. *Science,* 310**,** 1817-1821.

FARIDANI, O. R., ABDULLAYEV, I., HAGEMANN-JENSEN, M., SCHELL, J. P., LANNER, F. & SANDBERG, R. 2016. Single-cell sequencing of the small-RNA transcriptome. *Nat Biotech,* advance online publication.

FOGARTY, N. M. E., MCCARTHY, A., SNIJDERS, K. E., POWELL, B. E., KUBIKOVA, N., BLAKELEY, P., LEA, R., ELDER, K., WAMAITHA, S. E., KIM, D., MACIULYTE, V., KLEINJUNG, J., KIM, J.-S., WELLS, D., VALLIER, L., BERTERO, A., TURNER, J. M. A. & NIAKAN, K. K. 2017. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature*.

FOX, M. H. 1980. A model for the computer analysis of synchronous DNA distributions obtained by flow cytometry. *Cytometry,* 1**,** 71-77.

FRIEDMAN, R. C., FARH, K. K., BURGE, C. B. & BARTEL, D. P. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res,* 19**,** 92-105.

GAFNI, O., WEINBERGER, L., MANSOUR, A. A., MANOR, Y. S., CHOMSKY, E., BEN-YOSEF, D., KALMA, Y., VIUKOV, S., MAZA, I., ZVIRAN, A., RAIS, Y., SHIPONY, Z., MUKAMEL, Z., KRUPALNIK, V., ZERBIB, M., GEULA, S., CASPI, I., SCHNEIR, D., SHWARTZ, T., GILAD, S., AMANN-ZALCENSTEIN, D., BENJAMIN, S., AMIT, I., TANAY, A., MASSARWA, R., NOVERSHTERN, N. & HANNA, J. H. 2013. Derivation of novel human ground state naive pluripotent stem cells. *Nature,* 504**,** 282-+.

GARCIA-LOPEZ, J. & DEL MAZO, J. 2012. Expression dynamics of microRNA biogenesis during preimplantation mouse development. *Biochim Biophys Acta,* 1819**,** 847-54.

GAUDET, F., HODGSON, J. G., EDEN, A., JACKSON-GRUSBY, L., DAUSMAN, J., GRAY, J. W., LEONHARDT, H. & JAENISCH, R. 2003. Induction of Tumors in Mice by Genomic Hypomethylation. *Science,* 300**,** 489-492.

GREER CARD, D. A., HEBBAR, P. B., LI, L., TROTTER, K. W., KOMATSU, Y., MISHINA, Y. & ARCHER, T. K. 2008. Oct4/Sox2-Regulated miR-302 Targets Cyclin D1 in Human Embryonic Stem Cells. *Molecular and Cellular Biology,* 28**,** 6426-6438.

GREVE, T. S., JUDSON, R. L. & BLELLOCH, R. 2013. microRNA Control of Mouse and Human Pluripotent Stem Cell Behavior. *Annual Review of Cell and Developmental Biology,* 29**,** 213-239.

GU, K. L., ZHANG, Q., YAN, Y., LI, T. T., DUAN, F. F., HAO, J., WANG, X. W., SHI, M., WU, D. R., GUO, W. T. & WANG, Y. 2016. Pluripotency-associated miR-290/302 family of microRNAs promote the dismantling of naive pluripotency. *Cell Res,* 26**,** 350-66.

GUIROUILH-BARBAT, J., RASS, E., PLO, I., BERTRAND, P. & LOPEZ, B. S. 2007. Defects in XRCC4 and KU80 differentially affect the joining of distal nonhomologous ends. *Proc Natl Acad Sci U S A,* 104**,** 20902-7.

GUO, G., VON MEYENN, F., ROSTOVSKAYA, M., CLARKE, J., DIETMANN, S., BAKER, D., SAHAKYAN, A., MYERS, S., BERTONE, P., REIK, W., PLATH, K. & SMITH, A. 2017. Epigenetic resetting of human pluripotency. *Development,* 144**,** 2748-2763.

GUO, G., VON MEYENN, F., SANTOS, F., CHEN, Y., REIK, W., BERTONE, P., SMITH, A. & NICHOLS, J. 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports*.

GUO, G., YANG, J., NICHOLS, J., HALL, J. S., EYRES, I., MANSFIELD, W. & SMITH, A. 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development,* 136**,** 1063-9.

GUO, H., INGOLIA, N. T., WEISSMAN, J. S. & BARTEL, D. P. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature,* 466**,** 835-40.

GUO, H., ZHU, P., YAN, L., LI, R., HU, B., LIAN, Y., YAN, J., REN, X., LIN, S., LI, J., JIN, X., SHI, X., LIU, P., WANG, X., WANG, W., WEI, Y., LI, X., GUO, F., WU, X., FAN, X., YONG, J., WEN, L., XIE, S. X., TANG, F. & QIAO, J. 2014. The DNA methylation landscape of human early embryos. *Nature,* 511**,** 606-10.

HACKETT, J. A. & SURANI, M. A. 2014. Regulatory Principles of Pluripotency: From the Ground State Up. *Cell Stem Cell,* 15**,** 416-430.

HAYASHI, K., OHTA, H., KURIMOTO, K., ARAMAKI, S. & SAITOU, M. 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell,* 146**,** 519-32.

HAYASHITA, Y., OSADA, H., TATEMATSU, Y., YAMADA, H., YANAGISAWA, K., TOMIDA, S., YATABE, Y., KAWAHARA, K., SEKIDO, Y. & TAKAHASHI, T. 2005. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer Res,* 65**,** 9628-32.

HE, Q. Y., WANG, G. C., ZHANG, H., TONG, D. K., DING, C., LIU, K., JI, F., ZHU, X. & YANG, S. 2016. miR-106a-5p Suppresses the Proliferation, Migration, and Invasion of Osteosarcoma Cells by Targeting HMGA2. *DNA Cell Biol,* 35**,** 506-20.

HEARD, E., HAYASHI, K., CHUVA DE SOUSA LOPES, S. M., KANEDA, M., TANG, F., HAJKOVA, P., LAO, K., O'CARROLL, D., DAS, P. P., TARAKHOVSKY, A., MISKA, E. A. & SURANI, M. A. 2008. MicroRNA Biogenesis Is Required for Mouse Primordial Germ Cell Development and Spermatogenesis. *PLoS ONE,* 3**,** e1738.

HENDRICKSON, D. G., HOGAN, D. J., MCCULLOUGH, H. L., MYERS, J. W., HERSCHLAG, D., FERRELL, J. E. & BROWN, P. O. 2009. Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol,* 7**,** e1000238.

HER, J. & BUNTING, S. F. 2018. How cells ensure correct repair of DNA double-strand breaks. *J Biol Chem,* 293**,** 10502-10511.

HEYER, W. D., EHMSEN, K. T. & LIU, J. 2010. Regulation of homologous recombination in eukaryotes. *Annu Rev Genet,* 44**,** 113-39.

HIRANO, K., NAGATA, S., YAMAGUCHI, S., NAKAGAWA, M., OKITA, K., KOTERA, H., AINSCOUGH, J. & TADA, T. 2012. Human and mouse induced pluripotent stem cells are differentially reprogrammed in response to kinase inhibitors. *Stem Cells Dev,* 21**,** 1287-98.

HIRATA, H., UENO, K., SHAHRYARI, V., DENG, G., TANAKA, Y., TABATABAI, Z. L., HINODA, Y. & DAHIYA, R. 2013. MicroRNA-182-5p promotes cell invasion and proliferation by down regulating FOXF2, RECK and MTSS1 genes in human prostate cancer. *PLoS One,* 8**,** e55502.

HIRATA, H., UENO, K., SHAHRYARI, V., TANAKA, Y., TABATABAI, Z. L., HINODA, Y. & DAHIYA, R. 2012. Oncogenic miRNA-182-5p targets Smad4 and RECK in human bladder cancer. *PLoS One,* 7**,** e51056.

HOUBAVIY, H. B., MURRAY, M. F. & SHARP, P. A. 2003. Embryonic Stem Cell-Specific MicroRNAs. *Developmental Cell,* 5**,** 351-358.

HU, F., MIN, J., CAO, X., LIU, L., GE, Z., HU, J. & LI, X. 2016. MiR-363-3p inhibits the epithelial-to-mesenchymal transition and suppresses metastasis in colorectal cancer by targeting Sox4. *Biochem Biophys Res Commun,* 474**,** 35-42.

HUANG, K., MARUYAMA, T. & FAN, G. 2014. The Naive State of Human Pluripotent Stem Cells: A Synthesis of Stem Cell and Preimplantation Embryo Transcriptome Analyses. *Cell Stem Cell,* 15**,** 410-415.

HUERTAS, P. & JACKSON, S. P. 2009. Human CtIP mediates cell cycle control of DNA end resection and double strand break repair. *J Biol Chem,* 284**,** 9558-65.

IRIE, N., WEINBERGER, L., TANG, W. W., KOBAYASHI, T., VIUKOV, S., MANOR, Y. S., DIETMANN, S., HANNA, J. H. & SURANI, M. A. 2015. SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. *Cell*.

JACKSON, S. P. & BARTEK, J. 2009. The DNA-damage response in human biology and disease. *Nature,* 461**,** 1071-8.

JACOBI, A. M., RETTIG, G. R., TURK, R., COLLINGWOOD, M. A., ZEINER, S. A., QUADROS, R. M., HARMS, D. W., BONTHUIS, P. J., GREGG, C., OHTSUKA, M., GURUMURTHY, C. B. & BEHLKE, M. A. 2017. Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. *Methods,* 121-122**,** 16-28.

JACOBS, E. Z., WARRIER, S., VOLDERS, P. J., D'HAENE, E., VAN LOMBERGEN, E., VANTOMME, L., VAN DER JEUGHT, M., HEINDRYCKX, B., MENTEN, B. & VERGULT, S. 2017. CRISPR/Cas9-mediated genome editing in naive human embryonic stem cells. *Sci Rep,* 7**,** 16650.

JOO, J. Y., CHOI, H. W., KIM, M. J., ZAEHRES, H., TAPIA, N., STEHLING, M., JUNG, K. S., DO, J. T. & SCHOLER, H. R. 2014. Establishment of a primed pluripotent epiblast stem cell in FGF4-based conditions. *Sci Rep,* 4**,** 7477.

JOUNEAU, A., CIAUDO, C., SISMEIRO, O., BROCHARD, V., JOUNEAU, L., VANDORMAEL-POURNIN, S., COPPÉE, J.-Y., ZHOU, Q., HEARD, E., ANTONIEWSKI, C. & COHEN-TANNOUDJI, M. 2012. Naive and primed murine pluripotent stem cells have distinct miRNA expression profiles. *RNA,* 18**,** 253-264.

KAN, Y., RUIS, B., TAKASUGI, T. & HENDRICKSON, E. A. 2017. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res,* 27**,** 1099-1111.

KANELLOPOULOU, C., MULJO, S. A., KUNG, A. L., GANESAN, S., DRAPKIN, R., JENUWEIN, T., LIVINGSTON, D. M. & RAJEWSKY, K. 2005. Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev,* 19**,** 489-501.

KARANAM, K., KAFRI, R., LOEWER, A. & LAHAV, G. 2012. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol Cell,* 47**,** 320-9.

KENT, O. A., MCCALL, M. N., CORNISH, T. C. & HALUSHKA, M. K. 2014. Lessons from miR-143/145: the importance of cell-type localization of miRNAs. *Nucleic Acids Res,* 42**,** 7528-38.

KILENS, S., MEISTERMANN, D., MORENO, D., CHARIAU, C., GAIGNERIE, A., REIGNIER, A., LELIEVRE, Y., CASANOVA, M., VALLOT, C., NEDELLEC, S., FLIPPE, L., FIRMIN, J., SONG, J., CHARPENTIER, E., LAMMERS, J., DONNART, A., MAREC, N., DEB, W., BIHOUEE, A., LE CAIGNEC, C., PECQUEUR, C., REDON, R., BARRIERE, P., BOURDON, J., PASQUE, V., SOUMILLON, M., MIKKELSEN, T. S., ROUGEULLE, C., FREOUR, T., DAVID, L. & MILIEU INTERIEUR, C. 2018. Parallel derivation of isogenic human primed and naive induced pluripotent stem cells. *Nat Commun,* 9**,** 360.

KIM, J. H., LEE, S.-R., LI, L.-H., PARK, H.-J., PARK, J.-H., LEE, K. Y., KIM, M.-K., SHIN, B. A. & CHOI, S.-Y. 2011. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. *PLoS ONE,* 6**,** e18556.

KNIGHT, S. C., XIE, L., DENG, W., GUGLIELMI, B., WITKOWSKY, L. B., BOSANAC, L., ZHANG, E. T., EL BEHEIRY, M., MASSON, J.-B., DAHAN, M., LIU, Z., DOUDNA, J. A. & TJIAN, R. 2015. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science,* 350**,** 823-826.

KOJIMA, S., CHIYOMARU, T., KAWAKAMI, K., YOSHINO, H., ENOKIDA, H., NOHATA, N., FUSE, M., ICHIKAWA, T., NAYA, Y., NAKAGAWA, M. & SEKI, N. 2012. Tumour suppressors miR-1 and miR-133a target the oncogenic function of purine nucleoside phosphorylase (PNP) in prostate cancer. *Br J Cancer,* 106**,** 405-13.

KOWALCZYKOWSKI, S. C. 2015. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. *Cold Spring Harb Perspect Biol,* 7.

KOZOMARA, A. & GRIFFITHS-JONES, S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res,* 42**,** D68-73.

KRISHAN, A. 1975. Rapid flow cytofluorometric analysis of mammalian cell cycle by propidium iodide staining. *The Journal of Cell Biology,* 66**,** 188-193.

KRISHNAN, K., STEPTOE, A. L., MARTIN, H. C., WANI, S., NONES, K., WADDELL, N., MARIASEGARAM, M., SIMPSON, P. T., LAKHANI, S. R., GABRIELLI, B., VLASSOV, A., CLOONAN, N. & GRIMMOND, S. M. 2012. MicroRNA-182-5p targets a network of genes involved in DNA repair. *Rna,* 19**,** 230-242.

LABUHN, M., ADAMS, F. F., NG, M., KNOESS, S., SCHAMBACH, A., CHARPENTIER, E. M., SCHWARZER, A., MATEO, J. L., KLUSMANN, J.-H. & HECKL, D. 2018. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Research,* 46**,** 1375-1385.

LAW, J. C., RITKE, M. K., YALOWICH, J. C., LEDER, G. H. & FERRELL, R. E. 1993. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leukemia Research,* 17**,** 1045-1050.

LEE, J. H., LARONDE, S., COLLINS, T. J., SHAPOVALOVA, Z., TANASIJEVIC, B., MCNICOL, J. D., FIEBIG-COMYN, A., BENOIT, Y. D., LEE, J. B., MITCHELL, R. R. & BHATIA, M. 2017. Lineage-Specific Differentiation Is Influenced by State of Human Pluripotency. *Cell Rep,* 19**,** 20-35.

LEEB, M., DIETMANN, S., PARAMOR, M., NIWA, H. & SMITH, A. 2014. Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell,* 14**,** 385-93.

LI, X., STITH, C. M., BURGERS, P. M. & HEYER, W. D. 2009. PCNA is required for initiation of recombination-associated DNA synthesis by DNA polymerase delta. *Mol Cell,* 36**,** 704-13.

LIN, S., STAAHL, B. T., ALLA, R. K. & DOUDNA, J. A. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife,* 3**,** e04766.

LIN, Y., ZENG, Y., ZHANG, F., XUE, L., HUANG, Z., LI, W. & GUO, M. 2013. Characterization of microRNA expression profiles and the discovery of novel microRNAs involved in cancer during human embryonic development. *PLoS One,* 8**,** e69230.

LIPCHINA, I., ELKABETZ, Y., HAFNER, M., SHERIDAN, R., MIHAILOVIC, A., TUSCHL, T., SANDER, C., STUDER, L. & BETEL, D. 2011. Genome-wide identification of microRNA targets in human ES cells reveals a role for miR-302 in modulating BMP response. *Genes Dev,* 25**,** 2173-86.

LIU, Z., LIU, J., SEGURA, M. F., SHAO, C., LEE, P., GONG, Y., HERNANDO, E. & WEI, J. J. 2012. MiR-182 overexpression in tumourigenesis of high-grade serous ovarian carcinoma. *J Pathol,* 228**,** 204-15.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods,* 25**,** 402-8.

LU, Y., THOMSON, J. M., WONG, H. Y., HAMMOND, S. M. & HOGAN, B. L. 2007. Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Dev Biol,* 310**,** 442-53.

LUO, W., LI, G., YI, Z., NIE, Q. & ZHANG, X. 2016. E2F1-miR-20a-5p/20b-5p auto-regulatory feedback loop involved in myoblast proliferation and differentiation. *Sci Rep,* 6**,** 27904.

LV, K. T., LIU, Z., FENG, J., ZHAO, W., HAO, T., DING, W. Y., CHU, J. P. & GAO, L. J. 2018. MiR-22-3p Regulates Cell Proliferation and Inhibits Cell Apoptosis through Targeting the eIF4EBP3 Gene in Human Cervical Squamous Carcinoma Cells. *Int J Med Sci,* 15**,** 142-152.

MA, H., TU, L. C., NASERI, A., HUISMAN, M., ZHANG, S., GRUNWALD, D. & PEDERSON, T. 2016a. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *J Cell Biol,* 214**,** 529-37.

MA, J., FLEMR, M., STEIN, P., BERNINGER, P., MALIK, R., ZAVOLAN, M., SVOBODA, P. & SCHULTZ, R. M. 2010. MicroRNA activity is suppressed in mouse oocytes. *Curr Biol,* 20**,** 265-70.

MA, X., CHEN, H. & CHEN, L. 2016b. A dual role of Erk signaling in embryonic stem cells. *Exp Hematol,* 44**,** 151-6.

MACFARLAN, T. S., GIFFORD, W. D., DRISCOLL, S., LETTIERI, K., ROWE, H. M., BONANOMI, D., FIRTH, A., SINGER, O., TRONO, D. & PFAFF, S. L. 2012. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature,* 487**,** 57-63.

MANOR, Y. S., MASSARWA, R. & HANNA, J. H. 2015. Establishing the human naive pluripotent state. *Current opinion in genetics & development,* 34**,** 35-45.

MANSOURI, A. 2018. Mouse Knockouts: Modifying the Mouse Genome by Using Embryonic Stem Cells. 1-10.

MAO, Z., BOZZELLA, M., SELUANOV, A. & GORBUNOVA, V. 2008. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst),* 7**,** 1765-71.

MARKS, H., KALKAN, T., MENAFRA, R., DENISSOV, S., JONES, K., HOFEMEISTER, H., NICHOLS, J., KRANZ, A., STEWART, A. F., SMITH, A. & STUNNENBERG, H. G. 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell,* 149**,** 590-604.

MARSON, A., LEVINE, S. S., COLE, M. F., FRAMPTON, G. M., BRAMBRINK, T., JOHNSTONE, S., GUENTHER, M. G., JOHNSTON, W. K., WERNIG, M., NEWMAN, J., CALABRESE, J. M., DENNIS, L. M., VOLKERT, T. L., GUPTA, S., LOVE, J., HANNETT, N., SHARP, P. A., BARTEL, D.

P., JAENISCH, R. & YOUNG, R. A. 2008. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell,* 134**,** 521-33.

MARTIN, B. 2007. Separation Index: An Easy-to-Use Metric for Evaluation of Different Configurations on the Same Flow Cytometer. *Current Protocols in Cytometry,* 40**,** 1.21.1-1.21.6.

MARTIN, G. R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences,* 78**,** 7634-7638.

MATEOS-GOMEZ, P. A., GONG, F., NAIR, N., MILLER, K. M., LAZZERINI-DENCHI, E. & SFEIR, A. 2015. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. *Nature,* 518**,** 254-7.

MEDEIROS, L. A., DENNIS, L. M., GILL, M. E., HOUBAVIY, H., MARKOULAKI, S., FU, D., WHITE, A. C., KIRAK, O., SHARP, P. A., PAGE, D. C. & JAENISCH, R. 2011. Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. *Proc Natl Acad Sci U S A,* 108**,** 14163-8.

MELTON, C., JUDSON, R. L. & BLELLOCH, R. 2010. Opposing microRNA families regulate self-renewal in mouse embryonic stem cells. *Nature,* 463**,** 621-6.

MENDELL, J. T. 2008. miRiad roles for the miR-17-92 cluster in development and disease. *Cell,* 133**,** 217-22.

METZGER, L. & ILIAKIS, G. 2009. Kinetics of DNA Double-strand Break Repair Throughout the Cell Cycle as Assayed by Pulsed Field Gel Electrophoresis in CHO Cells. *International Journal of Radiation Biology,* 59**,** 1325-1339.

MINKOVSKY, A., PATEL, S. & PLATH, K. 2012. Concise review: Pluripotency and the transcriptional inactivation of the female Mammalian X chromosome. *Stem Cells,* 30**,** 48-54.

MIYAOKA, Y., BERMAN, J. R., COOPER, S. B., MAYERL, S. J., CHAN, A. H., ZHANG, B., KARLIN-NEUMANN, G. A. & CONKLIN, B. R. 2016. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep,* 6**,** 23549.

MOGILYANSKY, E. & RIGOUTSOS, I. 2013. The miR-17/92 cluster: a comprehensive update on its genomics, genetics, functions and increasingly important and numerous roles in health and disease. *Cell Death Differ,* 20**,** 1603-14.

MOOTHA, V. K., LINDGREN, C. M., ERIKSSON, K.-F., SUBRAMANIAN, A., SIHAG, S., LEHAR, J., PUIGSERVER, P., CARLSSON, E., RIDDERSTRÅLE, M., LAURILA, E., HOUSTIS, N., DALY, M. J., PATTERSON, N., MESIROV, J. P., GOLUB, T. R., TAMAYO, P., SPIEGELMAN, B., LANDER, E. S., HIRSCHHORN, J. N., ALTSHULER, D. & GROOP, L. C. 2003. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics,* 34**,** 267.

MORGANI, S. M., CANHAM, M. A., NICHOLS, J., SHAROV, A. A., MIGUELES, R. P., KO, M. S. & BRICKMAN, J. M. 2013. Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep,* 3**,** 1945-57.

MORIN, R. D., O'CONNOR, M. D., GRIFFITH, M., KUCHENBAUER, F., DELANEY, A., PRABHU, A. L., ZHAO, Y., MCDONALD, H., ZENG, T., HIRST, M., EAVES, C. J. & MARRA, M. A. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res,* 18**,** 610-21.

MOSKWA, P., BUFFA, F. M., PAN, Y., PANCHAKSHARI, R., GOTTIPATI, P., MUSCHEL, R. J., BEECH, J., KULSHRESTHA, R., ABDELMOHSEN, K., WEINSTOCK, D. M., GOROSPE, M., HARRIS, A. L., HELLEDAY, T. & CHOWDHURY, D. 2011. miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. *Mol Cell,* 41**,** 210-20.

NEWMAN, M. A., THOMSON, J. M. & HAMMOND, S. M. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *RNA,* 14**,** 1539-49.

NGUYEN, H. T., JACOBS, K. & SPITS, C. 2018. Human pluripotent stem cells in regenerative medicine: where do we stand?

NIWA, H. 2014. The pluripotency transcription factor network at work in reprogramming. *Current opinion in genetics & development,* 28**,** 25-31.

OKAE, H., CHIBA, H., HIURA, H., HAMADA, H., SATO, A., UTSUNOMIYA, T., KIKUCHI, H., YOSHIDA, H., TANAKA, A., SUYAMA, M. & ARIMA, T. 2014. Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet,* 10**,** e1004868.

OKAMOTO, I., PATRAT, C., THEPOT, D., PEYNOT, N., FAUQUE, P., DANIEL, N., DIABANGOUAYA, P., WOLF, J. P., RENARD, J. P., DURANTHON, V. & HEARD, E. 2011. Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature,* 472**,** 370-4.

OLIVE, V., JIANG, I. & HE, L. 2010. mir-17-92, a cluster of miRNAs in the midst of the cancer network. *Int J Biochem Cell Biol,* 42**,** 1348-54.

OTTAVIANI, D., LECAIN, M. & SHEER, D. 2014. The role of microhomology in genomic structural variation. *Trends Genet,* 30**,** 85-94.

OZATA, D. M., GAINETDINOV, I., ZOCH, A., O'CARROLL, D. & ZAMORE, P. D. 2018. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet*.

PANDOLFINI, L., LUZI, E., BRESSAN, D., UCCIFERRI, N., BERTACCHI, M., BRANDI, R., ROCCHICCIOLI, S., D'ONOFRIO, M. & CREMISI, F. 2016. RISC-mediated control of selected chromatin regulators stabilizes ground state pluripotency of mouse embryonic stem cells. *Genome Biol,* 17**,** 94.

PANYUTIN, I. V., HOLAR, S. A., NEUMANN, R. D. & PANYUTIN, I. G. 2017. Effect of ionizing radiation on the proliferation of human embryonic stem cells. *Sci Rep,* 7**,** 43995.

PARCHEM, R. J., MOORE, N., FISH, J. L., PARCHEM, J. G., BRAGA, T. T., SHENOY, A., OLDHAM, M. C., RUBENSTEIN, J. L., SCHNEIDER, R. A. & BLELLOCH, R. 2015. miR-302 Is Required for Timing of Neural Differentiation, Neural Tube Closure, and Embryonic Viability. *Cell Rep,* 12**,** 760-73.

PARCHEM, RONALD J., YE, J., JUDSON, ROBERT L., LARUSSA, MARIE F., KRISHNAKUMAR, R., BLELLOCH, A., OLDHAM, MICHAEL C. & BLELLOCH, R. 2014. Two miRNA Clusters Reveal Alternative Paths in Late-Stage Reprogramming. *Cell Stem Cell,* 14**,** 617-631.

PASTOR, WILLIAM A., CHEN, D., LIU, W., KIM, R., SAHAKYAN, A., LUKIANCHIKOV, A., PLATH, K., JACOBSEN, STEVEN E. & CLARK, AMANDER T. 2016. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell*.

PATEL, A. G., SARKARIA, J. N. & KAUFMANN, S. H. 2011. Nonhomologous end joining drives poly(ADP-ribose) polymerase (PARP) inhibitor lethality in homologous recombination-deficient cells. *Proc Natl Acad Sci U S A,* 108**,** 3406-11.

PERA, M. F. 2014. In Search of Naivety. *Cell Stem Cell,* 15**,** 543-545.

PETROPOULOS, S., EDSGÄRD, D., REINIUS, B., DENG, Q., PANULA, SARITA P., CODELUPPI, S., PLAZA REYES, A., LINNARSSON, S., SANDBERG, R. & LANNER, F. 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*.

QI, J., YU, J. Y., SHCHERBATA, H. R., MATHIEU, J., WANG, A. J., SEAL, S., ZHOU, W., STADLER, B. M., BOURGIN, D., WANG, L., NELSON, A., WARE, C., RAYMOND, C., LIM, L. P., MAGNUS, J., IVANOVSKA, I., DIAZ, R., BALL, A., CLEARY, M. A. & RUOHOLA-BAKER, H. 2009. microRNAs regulate human embryonic stem cell division. *Cell Cycle,* 8**,** 3729-41.

RAINGER, J., WILLIAMSON, K. A., SOARES, D. C., TRUCH, J., KURIAN, D., GILLESSEN-KAESBACH, G., SEAWRIGHT, A., PRENDERGAST, J., HALACHEV, M., WHEELER, A., MCTEIR, L., GILL, A. C., VAN HEYNINGEN, V., DAVEY, M. G., UK10K & FITZPATRICK, D. R. 2017. A recurrent de novo mutation in ACTG1 causes isolated ocular coloboma. *Hum Mutat,* 38**,** 942-946.

RAJEWSKY, N. 2006. microRNA target predictions in animals. *Nat Genet,* 38 Suppl**,** S8-13.

RAJMAN, M. & SCHRATT, G. 2017. MicroRNAs in neural development: from master regulators to fine-tuners. *Development,* 144**,** 2310-2322.

REYNOLDS, P., ANDERSON, J. A., HARPER, J. V., HILL, M. A., BOTCHWAY, S. W., PARKER, A. W. & O'NEILL, P. 2012. The dynamics of Ku70/80 and DNA-PKcs at DSBs induced by ionizing radiation is dependent on the complexity of damage. *Nucleic Acids Res,* 40**,** 10821-31.

RICHARDSON, C. D., KAZANE, K. R., FENG, S. J., ZELIN, E., BRAY, N. L., SCHAFER, A. J., FLOOR, S. N. & CORN, J. E. 2018. CRISPR-Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat Genet,* 50**,** 1132-1139.

RICHARDSON, C. D., RAY, G. J., DEWITT, M. A., CURIE, G. L. & CORN, J. E. 2016. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol,* 34**,** 339-44.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

RODRIGUEZ-TERRONES, D., GAUME, X., ISHIUCHI, T., WEISS, A., KOPP, A., KRUSE, K., PENNING, A., VAQUERIZAS, J. M., BRINO, L. & TORRES-PADILLA, M. E. 2018. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat Genet,* 50**,** 106-119.

ROSA, A., SPAGNOLI, F. M. & BRIVANLOU, A. H. 2009. The miR-430/427/302 family controls mesendodermal fate specification via species-specific target selection. *Dev Cell,* 16**,** 517-27.

ROSE, J. C., STEPHANY, J. J., VALENTE, W. J., TREVILLIAN, B. M., DANG, H. V., BIELAS, J. H., MALY, D. J. & FOWLER, D. M. 2017. Rapidly inducible Cas9 and DSB-ddPCR to probe editing kinetics. *Nat Methods,* 14**,** 891-896.

RYBAK, A., FUCHS, H., SMIRNOVA, L., BRANDT, C., POHL, E. E., NITSCH, R. & WULCZYN, F. G. 2008. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol,* 10**,** 987-93.

SAHAKYAN, A., KIM, R., CHRONIS, C., SABRI, S., BONORA, G., THEUNISSEN, T. W., KUOY, E., LANGERMAN, J., CLARK, A. T., JAENISCH, R. & PLATH, K. 2017. Human Naive Pluripotent Stem Cells Model X Chromosome Dampening and X Inactivation. *Cell Stem Cell,* 20**,** 87-101.

SASAKI, K., YOKOBAYASHI, S., NAKAMURA, T., OKAMOTO, I., YABUTA, Y., KURIMOTO, K., OHTA, H., MORITOKI, Y., IWATANI, C., TSUCHIYA, H., NAKAMURA, S., SEKIGUCHI, K., SAKUMA, T., YAMAMOTO, T., MORI, T., WOLTJEN, K., NAKAGAWA, M., YAMAMOTO, T., TAKAHASHI, K., YAMANAKA, S. & SAITOU, M. 2015. Robust In Vitro Induction of Human Germ Cell Fate from Pluripotent Stem Cells. *Cell Stem Cell,* 17**,** 178-94.

SFEIR, A. & SYMINGTON, L. S. 2015. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends Biochem Sci,* 40**,** 701-714.

SHAHAR, O. D., RAGHU RAM, E. V., SHIMSHONI, E., HARELI, S., MESHORER, E. & GOLDBERG, M. 2012. Live imaging of induced and controlled DNA double-strand break formation reveals extremely low repair by homologous recombination in human cells. *Oncogene,* 31**,** 3495-504.

SHAO, Z., DAVIS, A. J., FATTAH, K. R., SO, S., SUN, J., LEE, K. J., HARRISON, L., YANG, J. & CHEN, D. J. 2012. Persistently bound Ku at DNA ends attenuates DNA end resection and homologous recombination. *DNA Repair (Amst),* 11**,** 310-6.

SHI, G., PERLE, M. A., MITTAL, K., CHEN, H., ZOU, X., NARITA, M., HERNANDO, E., LEE, P. & WEI, J. J. 2009. Let-7 repression leads to HMGA2 overexpression in uterine leiomyosarcoma. *J Cell Mol Med,* 13**,** 3898-905.

SHIBATA, A., CONRAD, S., BIRRAUX, J., GEUTING, V., BARTON, O., ISMAIL, A., KAKAROUGKAS, A., MEEK, K., TAUCHER-SCHOLZ, G., LOBRICH, M. & JEGGO, P. A. 2011. Factors determining DNA double-strand break repair pathway choice in G2 phase. *EMBO J,* 30**,** 1079-92.

SHIBATA, A., MOIANI, D., ARVAI, A. S., PERRY, J., HARDING, S. M., GENOIS, M. M., MAITY, R., VAN ROSSUM-FIKKERT, S., KERTOKALIO, A., ROMOLI, F., ISMAIL, A., ISMALAJ, E., PETRICCI, E., NEALE, M. J., BRISTOW, R. G., MASSON, J. Y., WYMAN, C., JEGGO, P. A. & TAINER, J. A.

2014. DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol Cell,* 53**,** 7-18.

SHIBATA, M., NISHIMASU, H., KODERA, N., HIRANO, S., ANDO, T., UCHIHASHI, T. & NUREKI, O. 2017. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Commun,* 8**,** 1430.

SINKKONEN, L., HUGENSCHMIDT, T., BERNINGER, P., GAIDATZIS, D., MOHN, F., ARTUS-REVEL, C. G., ZAVOLAN, M., SVOBODA, P. & FILIPOWICZ, W. 2008. MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat Struct Mol Biol,* 15**,** 259-67.

SMITH, Z. D., CHAN, M. M., HUMM, K. C., KARNIK, R., MEKHOUBAD, S., REGEV, A., EGGAN, K. & MEISSNER, A. 2014. DNA methylation dynamics of the human preimplantation embryo. *Nature,* 511**,** 611-5.

SPRUCE, T., PERNAUTE, B., DI-GREGORIO, A., COBB, B. S., MERKENSCHLAGER, M., MANZANARES, M. & RODRIGUEZ, T. A. 2010. An early developmental role for miRNAs in the maintenance of extraembryonic stem cells in the mouse embryo. *Dev Cell,* 19**,** 207-19.

STEMMER, M., THUMBERGER, T., DEL SOL KEYER, M., WITTBRODT, J. & MATEO, J. L. 2015. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLOS ONE,* 10**,** e0124633.

STENERLÖW, B., KARLSSON, K. H., COOPER, B. & RYDBERG, B. 2003. Measurement of Prompt DNA Double-Strand Breaks in Mammalian Cells without Including Heat-Labile Sites: Results for Cells Deficient in Nonhomologous End Joining. *Radiation Research,* 159**,** 502-510.

STERNBERG, S. H., REDDING, S., JINEK, M., GREENE, E. C. & DOUDNA, J. A. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature,* 507**,** 62-7.

STIRPARO, G. G., BOROVIAK, T., GUO, G., NICHOLS, J., SMITH, A. & BERTONE, P. 2018. Integrated analysis of single-cell embryo data yields a unified transcriptome signature for the human pre-implantation epiblast. *Development,* 145.

STURZENEGGER, A., BURDOVA, K., KANAGARAJ, R., LEVIKOVA, M., PINTO, C., CEJKA, P. & JANSCAK, P. 2014. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *J Biol Chem,* 289**,** 27314-26.

SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A,* 102**,** 15545-50.

SUBTELNY, A. O., EICHHORN, S. W., CHEN, G. R., SIVE, H. & BARTEL, D. P. 2014. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature,* 508**,** 66-71.

SUH, M.-R., LEE, Y., KIM, J. Y., KIM, S.-K., MOON, S.-H., LEE, J. Y., CHA, K.-Y., CHUNG, H. M., YOON, H. S., MOON, S. Y., KIM, V. N. & KIM, K.-S. 2004. Human embryonic stem cells express a unique set of microRNAs. *Developmental Biology,* 270**,** 488-498.

SUH, N., BAEHNER, L., MOLTZAHN, F., MELTON, C., SHENOY, A., CHEN, J. & BLELLOCH, R. 2010. MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr Biol,* 20**,** 271-7.

SUN, J., LEE, K. J., DAVIS, A. J. & CHEN, D. J. 2012. Human Ku70/80 protein blocks exonuclease 1-mediated DNA resection in the presence of human Mre11 or Mre11/Rad50 protein complex. *J Biol Chem,* 287**,** 4936-45.

SYMINGTON, L. S. & GAUTIER, J. 2011. Double-strand break end resection and repair pathway choice. *Annu Rev Genet,* 45**,** 247-71.

TAKAHASHI, K., TANABE, K., OHNUKI, M., NARITA, M., ICHISAKA, T., TOMODA, K. & YAMANAKA, S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell,* 131**,** 861-72.

TAKAHASHI, K. & YAMANAKA, S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell,* 126**,** 663-76.

TAKASHIMA, Y., GUO, G., LOOS, R., NICHOLS, J., FICZ, G., KRUEGER, F., OXLEY, D., SANTOS, F., CLARKE, J., MANSFIELD, W., REIK, W., BERTONE, P. & SMITH, A. 2014. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell,* 158**,** 1254-69.

TANG, F., KANEDA, M., O'CARROLL, D., HAJKOVA, P., BARTON, S. C., SUN, Y. A., LEE, C., TARAKHOVSKY, A., LAO, K. & SURANI, M. A. 2007. Maternal microRNAs are essential for mouse zygotic development. *Genes Dev,* 21**,** 644-8.

TESAR, P. J., CHENOWETH, J. G., BROOK, F. A., DAVIES, T. J., EVANS, E. P., MACK, D. L., GARDNER, R. L. & MCKAY, R. D. G. 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature,* 448**,** 196-U10.

THEUNISSEN, T. W., FRIEDLI, M., HE, Y., PLANET, E., O'NEIL, R. C., MARKOULAKI, S., PONTIS, J., WANG, H., IOURANOVA, A., IMBEAULT, M., DUC, J., COHEN, M. A., WERT, K. J., CASTANON, R., ZHANG, Z., HUANG, Y., NERY, J. R., DROTAR, J., LUNGJANGWA, T., TRONO, D., ECKER, J. R. & JAENISCH, R. 2016. Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell,* 19**,** 502-515.

THEUNISSEN, T. W. & JAENISCH, R. 2017. Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development,* 144**,** 4496-4509.

THEUNISSEN, T. W., POWELL, B. E., WANG, H., MITALIPOVA, M., FADDAH, D. A., REDDY, J., FAN, Z. P., MAETZEL, D., GANZ, K. & SHI, L. 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell,* 15**,** 471-487.

THOMSON, J. A., ITSKOVITZ-ELDOR, J., SHAPIRO, S. S., WAKNITZ, M. A., SWIERGIEL, J. J., MARSHALL, V. S. & JONES, J. M. 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science,* 282**,** 1145-1147.

TROUNSON, A. & DEWITT, N. D. 2016. Pluripotent stem cells progressing to the clinic. *Nat Rev Mol Cell Biol,* 17**,** 194-200.

URBACH, A., SCHULDINER, M. & BENVENISTY, N. 2004. Modeling for Lesch-Nyhan Disease by Gene Targeting in Human Embryonic Stem Cells. *STEM CELLS,* 22**,** 635-641.

VAKULSKAS, C. A., DEVER, D. P., RETTIG, G. R., TURK, R., JACOBI, A. M., COLLINGWOOD, M. A., BODE, N. M., MCNEILL, M. S., YAN, S., CAMARENA, J., LEE, C. M., PARK, S. H., WIEBKING, V., BAK, R. O., GOMEZ-OSPINA, N., PAVEL-DINU, M., SUN, W., BAO, G., PORTEUS, M. H. & BEHLKE, M. A. 2018. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med,* 24**,** 1216-1224.

VALAMEHR, B., ROBINSON, M., ABUJAROUR, R., REZNER, B., VRANCEANU, F., LE, T., MEDCALF, A., LEE, T. T., FITCH, M., ROBBINS, D. & FLYNN, P. 2014. Platform for Induction and Maintenance of Transgene-free hiPSCs Resembling Ground State Pluripotent Stem Cells. *Stem Cell Reports,* 2**,** 366-381.

VAN BORTLE, K. & CORCES, V. G. 2014. Lost in Transition: Dynamic Enhancer Organization across Naive and Primed Stem Cell States. *Cell Stem Cell,* 14**,** 693-694.

VASSENA, R., BOUÉ, S., GONZÁLEZ-ROCA, E., ARAN, B., AUER, H., VEIGA, A. & BELMONTE, J. C. I. 2011. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development,* 138**,** 3699-3709.

VISWANATHAN, S. R., DALEY, G. Q. & GREGORY, R. I. 2008. Selective Blockade of MicroRNA Processing by Lin28. *Science,* 320**,** 97-100.

VOSSAERT, L., O'LEARY, T., VAN NESTE, C., HEINDRYCKX, B., VANDESOMPELE, J., DE SUTTER, P. & DEFORCE, D. 2013. Reference loci for RT-qPCR analysis of differentiating human embryonic stem cells. *BMC Molecular Biology,* 14**,** 21.

WAGHRAY, A., SAIZ, N., JAYAPRAKASH, A. D., FREIRE, A. G., PAPATSENKO, D., PEREIRA, C. F., LEE, D. F., BROSH, R., CHANG, B., DARR, H., GINGOLD, J., KELLEY, K., SCHANIEL, C., HADJANTONAKIS, A. K. & LEMISCHKA, I. R. 2015. Tbx3 Controls Dppa3 Levels and Exit from Pluripotency toward Mesoderm. *Stem Cell Reports,* 5**,** 97-110.

WANG, M., WU, W., WU, W., ROSIDI, B., ZHANG, L., WANG, H. & ILIAKIS, G. 2006. PARP-1 and Ku compete for repair of DNA double strand breaks by distinct NHEJ pathways. *Nucleic Acids Res,* 34**,** 6170-82.

WANG, X., LIANG, Z., XU, X., LI, J., ZHU, Y., MENG, S., LI, S., WANG, S., XIE, B., JI, A., LIU, B., ZHENG, X. & XIE, L. 2016. miR-148a-3p represses proliferation and EMT by establishing regulatory circuits between ERBB3/AKT2/c-myc and DNMT1 in bladder cancer. *Cell Death Dis,* 7**,** e2503.

WANG, Y., BASKERVILLE, S., SHENOY, A., BABIARZ, J. E., BAEHNER, L. & BLELLOCH, R. 2008. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet,* 40**,** 1478-83.

WANG, Y., MEDVID, R., MELTON, C., JAENISCH, R. & BLELLOCH, R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet,* 39**,** 380-5.

WANG, Z., ORON, E., NELSON, B., RAZIS, S. & IVANOVA, N. 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell,* 10**,** 440-54.

WARE, C. B. 2016. Concise Review: Lessons from Naïve Human Pluripotent Cells. *STEM CELLS***,** n/a-n/a.

WARE, C. B., NELSON, A. M., MECHAM, B., HESSON, J., ZHOU, W., JONLIN, E. C., JIMENEZ-CALIANI, A. J., DENG, X., CAVANAUGH, C., COOK, S., TESAR, P. J., OKADA, J., MARGARETHA, L., SPERBER, H., CHOI, M., BLAU, C. A., TREUTING, P. M., HAWKINS, R. D., CIRULLI, V. & RUOHOLA-BAKER, H. 2014. Derivation of naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 4484-4489.

WARRIER, S., VAN DER JEUGHT, M., DUGGAL, G., TILLEMAN, L., SUTHERLAND, E., TAELMAN, J., POPOVIC, M., LIERMAN, S., CHUVA DE SOUSA LOPES, S., VAN SOOM, A., PEELMAN, L., VAN NIEUWERBURGH, F., DE CONINCK, D. I. M., MENTEN, B., MESTDAGH, P., VAN DE SOMPELE, J., DEFORCE, D., DE SUTTER, P. & HEINDRYCKX, B. 2017. Direct comparison of distinct naive pluripotent states in human embryonic stem cells. *Nat Commun,* 8**,** 15055.

WIGGER, M., KISIELEWSKA, K., FILIMONOW, K., PLUSA, B., MALESZEWSKI, M. & SUWINSKA, A. 2017. Plasticity of the inner cell mass in mouse blastocyst is restricted by the activity of FGF/MAPK pathway. *Sci Rep,* 7**,** 15136.

WOLF, F. A., ANGERER, P. & THEIS, F. J. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol,* 19**,** 15.

WU, J. & IZPISUA BELMONTE, J. C. 2015. Dynamic Pluripotent Stem Cell States and Their Applications. *Cell Stem Cell,* 17**,** 509-25.

WU, J., PLATERO-LUENGO, A., SAKURAI, M., SUGAWARA, A., GIL, M. A., YAMAUCHI, T., SUZUKI, K., BOGLIOTTI, Y. S., CUELLO, C., MORALES VALENCIA, M., OKUMURA, D., LUO, J., VILARINO, M., PARRILLA, I., SOTO, D. A., MARTINEZ, C. A., HISHIDA, T., SANCHEZ-BAUTISTA, S., MARTINEZ-MARTINEZ, M. L., WANG, H., NOHALEZ, A., AIZAWA, E., MARTINEZ-REDONDO, P., OCAMPO, A., REDDY, P., ROCA, J., MAGA, E. A., ESTEBAN, C. R., BERGGREN, W. T., NUNEZ DELICADO, E., LAJARA, J., GUILLEN, I., GUILLEN, P., CAMPISTOL, J. M., MARTINEZ, E. A., ROSS, P. J. & IZPISUA BELMONTE, J. C. 2017a. Interspecies Chimerism with Mammalian Pluripotent Stem Cells. *Cell,* 168**,** 473-486 e15.

WU, L., FAN, J. & BELASCO, J. G. 2006. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A,* 103**,** 4034-9.

WU, Q., FERRY, Q. R. V., BAEUMLER, T. A., MICHAELS, Y. S., VITSIOS, D. M., HABIB, O., ARNOLD, R., JIANG, X., MAIO, S., STEINKRAUS, B. R., TAPIA, M., PIAZZA, P., XU, N., HOLLANDER, G. A., MILNE, T. A., KIM, J. S., ENRIGHT, A. J., BASSETT, A. R. & FULGA, T. A. 2017b. In situ functional dissection of RNA cis-regulatory elements by multiplex CRISPR-Cas9 genome engineering. *Nat Commun,* 8**,** 2109.

WU, S., AKSOY, M., SHI, J. & HOUBAVIY, H. B. 2014a. Evolution of the miR-290–295/miR-371–373 Cluster Family Seed Repertoire. *PLOS ONE,* 9**,** e108519.

WU, T. D. & WATANABE, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics,* 21**,** 1859-75.

WU, X., SCOTT, D. A., KRIZ, A. J., CHIU, A. C., HSU, P. D., DADON, D. B., CHENG, A. W., TREVINO, A. E., KONERMANN, S., CHEN, S., JAENISCH, R., ZHANG, F. & SHARP, P. A. 2014b. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology,* 32**,** 670-+.

XU, J., LI, C. X., LI, Y. S., LV, J. Y., MA, Y., SHAO, T. T., XU, L. D., WANG, Y. Y., DU, L., ZHANG, Y. P., JIANG, W., LI, C. Q., XIAO, Y. & LI, X. 2011. MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. *Nucleic Acids Res,* 39**,** 825-36.

XU, N., PAPAGIANNAKOPOULOS, T., PAN, G., THOMSON, J. A. & KOSIK, K. S. 2009. MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell,* 137**,** 647-58.

YAMADA, Y., NISHIKAWA, R., KATO, M., OKATO, A., ARAI, T., KOJIMA, S., YAMAZAKI, K., NAYA, Y., ICHIKAWA, T. & SEKI, N. 2017. Regulation of HMGB3 by antitumor miR-205-5p inhibits cancer cell aggressiveness and is involved in prostate cancer pathogenesis. *J Hum Genet*.

YAMASHIRO, C., SASAKI, K., YABUTA, Y., KOJIMA, Y., NAKAMURA, T., OKAMOTO, I., YOKOBAYASHI, S., MURASE, Y., ISHIKURA, Y., SHIRANE, K., SASAKI, H., YAMAMOTO, T. & SAITOU, M. 2018. Generation of human oogonia from induced pluripotent stem cells in vitro. *Science*.

YAN, L., YANG, M., GUO, H., YANG, L., WU, J., LI, R., LIU, P., LIAN, Y., ZHENG, X. & YAN, J. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology,* 20**,** 1131-1139.

YAN, Y., YANG, X., LI, T.-T., GU, K.-L., HAO, J., ZHANG, Q. & WANG, Y. 2017. Significant differences of function and expression of microRNAs between ground state and serum-cultured pluripotent stem cells. *Journal of Genetics and Genomics*.

YANG, J., RYAN, D. J., WANG, W., TSANG, J. C., LAN, G., MASAKI, H., GAO, X., ANTUNES, L., YU, Y., ZHU, Z., WANG, J., KOLODZIEJCZYK, A. A., CAMPOS, L. S., WANG, C., YANG, F., ZHONG, Z., FU, B., ECKERSLEY-MASLIN, M. A., WOODS, M., TANAKA, Y., CHEN, X., WILKINSON, A. C., BUSSELL, J., WHITE, J., RAMIREZ-SOLIS, R., REIK, W., GOTTGENS, B., TEICHMANN, S. A., TAM, P. P. L., NAKAUCHI, H., ZOU, X., LU, L. & LIU, P. 2017a. Establishment of mouse expanded potential stem cells. *Nature*.

YANG, S. H., KALKAN, T., MORRISROE, C., SMITH, A. & SHARROCKS, A. D. 2012. A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation. *PLoS Genet,* 8**,** e1003112.

YANG, Y., LIU, B., XU, J., WANG, J., WU, J., SHI, C., XU, Y., DONG, J., WANG, C., LAI, W., ZHU, J., XIONG, L., ZHU, D., LI, X., YANG, W., YAMAUCHI, T., SUGAWARA, A., LI, Z., SUN, F., LI, X., LI, C., HE, A., DU, Y., WANG, T., ZHAO, C., LI, H., CHI, X., ZHANG, H., LIU, Y., LI, C., DUO, S., YIN, M., SHEN, H., BELMONTE, J. C. & DENG, H. 2017b. Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell,* 169**,** 243-257 e25.

YE, S., LIU, D. & YING, Q.-L. 2014. Signaling pathways in induced naïve pluripotency. *Current opinion in genetics & development,* 28**,** 10-15.

YILMAZ, A., PERETZ, M., AHARONY, A., SAGI, I. & BENVENISTY, N. 2018. Defining essential genes for human pluripotent stem cells by CRISPR-Cas9 screening in haploid cells. *Nat Cell Biol,* 20**,** 610-619.

YOKOBAYASHI, S., OKITA, K., NAKAGAWA, M., NAKAMURA, T., YABUTA, Y., YAMAMOTO, T. & SAITOU, M. 2017. Clonal variation of human induced pluripotent stem cells for induction into the germ cell fate. *Biol Reprod,* 96**,** 1154-1166.

YU, H., XU, W., GONG, F., CHI, B., CHEN, J. & ZHOU, L. 2017. MicroRNA-155 regulates the proliferation, cell cycle, apoptosis and migration of colon cancer cells and targets CBL. *Exp Ther Med,* 14**,** 4053-4060.

YU, J., VODYANIK, M. A., SMUGA-OTTO, K., ANTOSIEWICZ-BOURGET, J., FRANE, J. L., TIAN, S., NIE, J., JONSDOTTIR, G. A., RUOTTI, V., STEWART, R., SLUKVIN, I. I. & THOMSON, J. A. 2007. Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science,* 318**,** 1917-1920.

ZHAO, W., VAITHIYALINGAM, S., SAN FILIPPO, J., MARANON, D. G., JIMENEZ-SAINZ, J., FONTENAY, G. V., KWON, Y., LEUNG, S. G., LU, L., JENSEN, R. B., CHAZIN, W. J., WIESE, C. & SUNG, P. 2015. Promotion of BRCA2-Dependent Homologous Recombination by DSS1 via RPA Targeting and DNA Mimicry. *Mol Cell,* 59**,** 176-87.

ZHOU, H., LI, W., ZHU, S., JOO, J. Y., DO, J. T., XIONG, W., KIM, J. B., ZHANG, K., SCHOELER, H. R. & DING, S. 2010. Conversion of Mouse Epiblast Stem Cells to an Earlier Pluripotency State by Small Molecules. *Journal of Biological Chemistry,* 285**,** 29676-29680.

ZHOU, W., CHOI, M., MARGINEANTU, D., MARGARETHA, L., HESSON, J., CAVANAUGH, C., BLAU, C. A., HORWITZ, M. S., HOCKENBERY, D. & WARE, C. 2012a. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *The EMBO journal,* 31**,** 2103-2116.

ZHOU, W., CHOI, M., MARGINEANTU, D., MARGARETHA, L., HESSON, J., CAVANAUGH, C., BLAU, C. A., HORWITZ, M. S., HOCKENBERY, D., WARE, C. & RUOHOLA-BAKER, H. 2012b. HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J,* 31**,** 2103-16.

ZHOU, Y., CARON, P., LEGUBE, G. & PAULL, T. T. 2014. Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic Acids Res,* 42**,** e19.

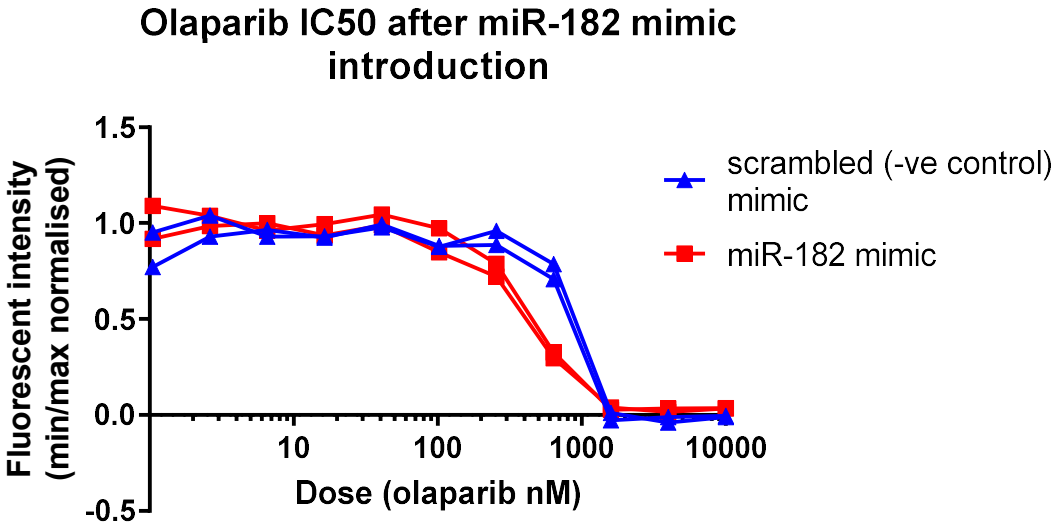
ZHU, X., GAO, G., CHU, K., YANG, X., REN, S., LI, Y., WU, H., HUANG, Y. & ZHOU, C. 2015. Inhibition of RAC1-GEF DOCK3 by miR-512-3p contributes to suppression of metastasis in non-small cell lung cancer. *Int J Biochem Cell Biol,* 61**,** 103-14.

ZWAKA, T. P. & THOMSON, J. A. 2003. Homologous recombination in human embryonic stem cells. *Nat Biotechnol,* 21**,** 319-21.

7.2. Appendix II: Updates since submission

Since submission Messmer *et al.* (2019) published a comprehensive single cell RNAseq dataset of naïve (t2iLGö) and primed hPSC. They found their cells to be mostly homogenous, with the exception of a small subpopulation (9/480 cells) of naïve cells which resemble the naïve state but also have features of primed cells. To compare their single cell RNAseq data to data from in vivo embryogenesis of several species, the authors defined a naïve-primed axis (containing all genes with cut-offs for maximum FDR, minimum presence in individual cells, minimum differential expression for naïve and primed states). Using this approach, they found that this expression signature is conserved across mouse, cynomolgus monkey and human development, albeit with different timings. Their analysis suggests that the primed state is not established in the human by E7 and that cells become more primed from E13 in cynomolgus monkey. This contrasts slightly with our observed clustering of naïve cells with E5 and primed cells with E6/E7 stages of embryogenesis.

Messmer, T., et al. (2019). "Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution." Cell Rep **26(4): 815-824 e814.**



Appendix II: Preliminary data from post-submission experiments. In chapter 4, miR-182-5p was shown to be expressed in naïve but not primed cells. In chapter 5, naïve cells were revealed to have lower rates of HDR. miR-182-5p has been shown to target BRCA1 and downregulates HR in literature. Here, a mimic of miR-182-5p was introduced into primed cells by electroporation and the cells were plated at 66.6k cells per cm^2 onto 96 well plates. Different doses of the PARP inhibitor olaparib were added to the cells the following day, for 5 days. Cells were then stained with NucBlue™ Live ReadyProbes™ Reagent according to the manufacturers instructions. The cell density was measured as overall fluorescent intensity which was normalised by min/max normalisation.

Olaparib selectively kills cells with deficiencies in HR and the addition of the miR-182-5p mimic sensitises the primed cells to olaparib, confirming that addition of miR-182-5p in primed iPSC also causes a decrease in HR. This experiment was performed in two technical replicates (same transfection, different wells) in the cell line SFC180-01-01.