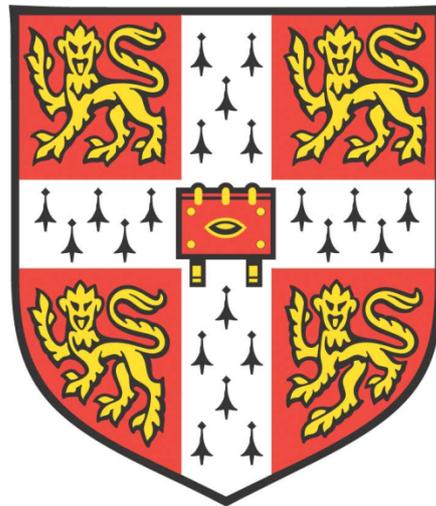


*ACTIVIN/NODAL SIGNALLING CONTROLS
THE EPIGENOME AND EPITRANSCRIPTOME
OF HUMAN PLURIPOTENT STEM CELLS*



Alessandro Bertero

St Catharine's College
Department of Surgery
University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

June 2016

To Anna and Elisa, the women of my life

“Dicebat Bernardus Carnotensis nos esse quasi nanos gigantium humeris insidentes, ut possimus plura eis et remotiora videre, non utique proprii visus acumine, aut eminentia corporis, sed quia in altum subvehimur et extollimur magnitudine gigantean” - Iohannes Saresberiensis

In memory of Professor Guido Tarone, a titan whose shoulders upheld me

ACKNOWLEDGEMENTS

The path that led me to this work was by far my toughest climb yet. Thankfully, this journey was eased by a fantastic group of people and institutions. If I reached the top of this mountain it is thanks to the help of my PhD mentor Prof. Ludovic Vallier, my colleagues, my collaborators, my College, the University, my founding body, my friends, and my family. This dissertation is the fruit of their support as much as it is of my efforts.

Ludovic, you have been a terrific adviser during my studies. I have been truly privileged to benefit from your teachings, guidance, patience, trust, and continuous support. You have been a masterful example of the scientist I aim to become, and I will do my very best to put this treasure to good use in the years to come.

Will, Miguel, Daniel, and Matthias, you have been much more than peers and colleagues: I am lucky to call you among my dearest friends. Thanks for keeping up with my many flaws, and for your help both inside and outside of the lab. Fotis, Anna, Loukia, Joana, Elisa, and the countless other marvellous LRM people, thanks for having made our institute such a wonderful place to both work and play. I will sorely miss the long days (and nights!) spent side by side in the lab, but I will miss even more teatimes, cinema nights, pub outings, game nights, and all the other great times together. Stephanie, you have been a tremendous support for almost all of my projects, and a trusted advisor in many times of technical difficulties or mental stress. Pedro, you have been my dependable bioinformatics wizard, as well as an inspiration for me to learn the basics of such dark arts. Kirsten and James, you have been a fantastic asset with our ongoing project, and I am truly thankful for your hard work. Sasha, Siim and Nick, your teachings at the beginning of my PhD were a massive aid in my studies, and you have shown me the way to go for my next career stage. Dr Sanjay Sinha, Prof. Roger Pedersen, Dr Mark Kotter, and all the members of their respective laboratories, thanks for creating such a scientifically fervent and collaborative environment in our institute.

Antonella, Nina, Dan, Inmaculada, and all my other collaborators both in and outside of the UK, your expertise has been pivotal with many crucial aspects of my work. I have been extremely lucky to work with and learn from such an amazing group of specialists.

St Catharine's College has been a unique place to study. Being involved in the MCR community was a tremendous help in my "social" transition to the UK. The Fellows, Senior Tutors, and Graduate Administrator have all been instrumental in my success, and the economical support that I received from the College during my training was truly appreciated. The University of Cambridge and the Department of Surgery are fabulous institutions, and I have been blessed by the opportunity to pursue my PhD in such a wondrous environment.

The British Heart Foundation most generously supported my PhD studies, and I will be eternally grateful to have been given such a marvellous opportunity. Prof. Martin Bennett, thanks for your guidance and support whilst leading this fantastic PhD program. Dr Anthony Davenport, Dr Amer Rana, and Prof. Nick Morrell, your supervision during the MRes that preceded my PhD was a unique opportunity for personal and scientific growth.

Vicky, Laura, Glenn, Ib, John, Golnessa and Ciara, you have been the best friends I could hope for during my time in Cambridge. Thanks for not giving up on me, and for having always made me feel part of the gang. Fagiani, through your friendship and comradery you mitigated my nostalgia for our sweet home during these long years in foreign land. Maury, Fra, Gorze, Chiara, Davide, Luca, and all the other friends from back home, thanks for having been always so close despite the physical distance.

Anna and Sergio, mum and dad, you have always been my greatest supporters. Thanks for having given me everything a son can ever ask for, even when I did not truly deserve it. While you paid the highest price in order to see me thrive and succeed, you always backed my decisions without a flinch. I owe you my life, I owe you my happiness, and I owe you my future.

Elisa, my soulmate and bride to be, I met you at the beginning of this long journey and we travelled together ever since. I would not be here without you, and I would not be the man I am without you. I look forward to helping you through the last steps of your own PhD, and then share with you all the other climbs that life will present us. I will always be at your side. I love you.

DECLARATION OF ORIGINALITY

I hereby declare that this dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except as specified in the text.

Furthermore, this dissertation is not substantially the same as any that I have submitted or is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

Alessandro Bertero

June 2016

Cambridge

STATEMENT OF LENGTH

As stipulated by the Degree Committee of Clinical Medicine, this dissertation does not exceed 60,000 words in length (excluding figures, photographs, tables, appendices and bibliography).

LIST OF PUBLICATIONS AND PRESENTATIONS

Work from this dissertation has contributed to the following scientific publications and presentations.

Publications

Bertero A, Madrigal P, Galli A, Hubner NC, Moreno I, Burks D, Brown S, Pedersen RA, Gaffney D, Mendjan S, et al. 2015. Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev* **29**: 702–17.

Bertero A, Pawlowski M, Ortmann D, Snijders K, Yiangou L, Cardoso de Brito M, Brown S, Bernard WG, Cooper JD, Giacomelli E, et al. Optimized inducible shRNA and CRISPR/Cas9 platforms to study human development. Manuscript in review at *Development*.

Bertero A, Brown S, Madrigal P, Hubner NC, Osnato A, Ortmann D, Pawlowski M, Farnell E, Stunnenberg HG, Mendjan S, et al. SMAD2/3 controls both the epigenome and epitranscriptome of human pluripotent stem cells. Manuscript to be submitted

Bertero A, Vallier L. 2015. Fucci2a mouse upgrades live cell cycle imaging. *Cell Cycle* **14**: 948–9.

Pauklin S, Madrigal P, Bertero A, Vallier L. 2016. Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. *Genes Dev* **30**: 421–433.

Pawlowski M, Ortmann D, Bertero A, Pedersen RA, Vallier L, Kotter M. Inducible and deterministic forward programming of human pluripotent stem cells. Manuscript in review at *Stem Cell Reports*.

Sampaziotis F, Cardoso de Brito M, Madrigal P, Bertero A, Saeb-Parsy K, Soares FAC, Schrupf E, Melum E, Karlsen TH, Bradley JA, et al. 2015. Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nat Biotechnol* **33**: 845–852.

Oral Presentations

- Bertero A. (October 2015). Control of human cardiovascular development by TGFbeta superfamily signalling (invited presentation). *BHF Oxbrige Centre for Regenerative Medicine annual meeting, The Royal Society, London, UK.*
- Bertero A, Madrigal P, Mendjan S, Galli A, Moreno I, Burks D, Hubner NC, Pauklin S, Vallier L. (April 2014). Activin/Nodal signalling controls the epigenetic landscape of human pluripotent stem cells (abstract selected for oral presentation). *Keystone Symposia meeting on "Stem Cells and Reprogramming", Squaw Creek, CA, USA.*
- Bertero A, Madrigal P, Mendjan S, Galli A, Moreno I, Burks D, Hubner NC, Pauklin S, Vallier L. (February 2014). Activin/Nodal signalling controls the epigenetic landscape of human pluripotent stem cells (invited presentation). *Cambridge Stem Cell Club, Cambridge, UK.*
- Bertero A, Pauklin S, Mendjan S, Hubner NC, Cader Z, Pedersen RA, Vallier L. (September 2013). Activin signalling regulates H3K4 trimethylation in pluripotent stem cells (invited presentation). *Anne McLaren Laboratory for Regenerative Medicine annual meeting, Cambridge, UK.*
- Bertero A, Pauklin S, Mendjan S, Cader Z, Pedersen RA, Vallier L. (November 2012). The role of novel SMAD2/3-binding epigenetic regulators during early cell-fate decisions in human embryonic stem cells (abstract selected for oral presentation). *International Society of Differentiation conference on "Stem Cells, Development and Regulation", Amsterdam, The Netherlands.*

Poster Presentations

- Bertero A, Pawlowski M, Ortmann D, Cardoso de Brito M, Brown S, Yiangou L, Bernard WG, Giacomelli E, Gambardella L, Hannan NRF, et al. (June 2015). OPTiKD: an inducible gene knockdown system to study human development using pluripotent stem cells. *International Society for Stem Cell Research annual meeting, Stockholm, Sweden.*
- Bertero A, Mendjan S, Pawlowski M, Ortmann D, Hubner NC, Brown S, Stunnenberg HG, Kotter M, Vallier L. (April 2015). Dissecting the role of TGFbeta signalling during human cardiovascular development by using an optimized inducible gene knockdown system in differentiating pluripotent stem cells. *BHF Cambridge Centre of Excellence annual meeting, Cambridge, UK.*
- Bertero A, Mendjan S, Hubner NC, Stunnenberg HG, Vallier L. (December 2014). Identification and functional characterization of the SMAD2/3 interactome in hPSCs. *MRC-Wellcome Trust Cambridge Stem Cell Institute annual meeting, Cambridge, UK.*
- Bertero A, Madrigal P, Mendjan S, Galli A, Moreno I, Burks D, Hubner NC, Pauklin S, Vallier L. (September 2014). Activin/Nodal signalling controls the epigenetic landscape of human pluripotent stem cells. *The Company of Biologists workshop: "From Stem Cells to Human Development", Dorking, UK.*
- Bertero A, Pauklin S, Mendjan S, Hubner NC, Cader Z, Pedersen RA, Vallier L. (July 2013). SMAD2/3 cooperates with NANOG to direct H3K4 trimethylation epigenetic marks in pluripotent stem cells. *MRC-Wellcome Trust Cambridge Stem Cell Institute annual PhD meeting, Cambridge, UK.*
- Bertero A, Pauklin S, Mendjan S, Hubner NC, Cader Z, Pedersen RA, Vallier L. (April 2013). SMAD2/3 and NANOG-dependent H3K4 trimethylation epigenetic marks regulate human embryonic stem cell self-renewal and differentiation. *BHF 4-Year PhD meeting, Oxford, UK.*
- Bertero A, Pauklin S, Mendjan S, Cader Z, Pedersen RA, Vallier L. (October 2012). The role of novel SMAD2/3-binding epigenetic regulators during early cell fate decisions in human embryonic stem cells. *Abcam conference "Epigenetic and Stem Cells", Cambridge, UK.*

AWARDS

Work from this dissertation has resulted into the following awards:

- June 2015. Travel award from the St Catharine's College Graduate Travel and Research Fund to support the participation to the International Society for Stem Cell Research annual meeting.
- May 2014. Travel award from the Isabelle Bouhon Fund (University of Cambridge) to support the participation to the Keystone Symposia meeting on "Stem Cells and Reprogramming".
- March 2014. Travel award from the St Catharine's College Graduate Travel and Research Fund to support the participation to the Keystone Symposia meeting on "Stem Cells and Reprogramming".
- October 2013. Travel award from the St Catharine's College Graduate Travel and Research Fund to support the participation to the "Biological Interpretation of Next Generation Sequencing Data" course at the European Bioinformatics Institute.
- October 2012. Travel award from the International Society of Differentiation to support the participation to the conference "Stem Cells, Development and Regulation".

SUMMARY

Human pluripotent stem cells (hPSCs) are an invaluable model for cellular and developmental biology, and hold great potential for translational applications. While great progress has been made in elucidating the signalling pathways regulating pluripotency and differentiation, our mechanistic understanding of the downstream regulations is still incomplete. Moreover, studies aimed at clarifying these aspects are severely impeded by the lack of efficient methods to conditionally modulate gene expression in hPSCs and hPSC-derived cells.

In this dissertation I provide new insights into the molecular mechanisms controlled by the Activin/Nodal-SMAD2/3 signalling pathway, whose activity dictates the balance between hPSC pluripotency and differentiation. First, I show that SMAD2/3 modulates the chromatin epigenetic landscape of hPSCs by cooperating with the pluripotency factor NANOG to recruit the DPY30-COMPASS complex and promote histone 3 lysine 4 trimethylation (H3K4me3). This regulation promotes expression of pluripotency genes, while poising developmental regulators for activation during differentiation. Secondly, I describe a novel efficient approach for inducible gene knockdown in hPSCs and hPSC-derived cells. By taking advantage of this technology, I demonstrate that DPY30 is required for early differentiation of hPSCs into certain mesoderm and endoderm derivatives. Finally, I report the first large-scale proteomic identification of SMAD2/3 interacting proteins in both undifferentiated and differentiating hPSCs. This analysis not only confirms that SMAD2/3 interacts with multiple epigenetic modifiers involved in hPSC fate choices, but also implicates SMAD2/3 in several functions other than transcriptional regulation. In particular, I describe how SMAD2/3 physically and functionally interacts with the METTL3-METTL14-WTAP complex to promote the formation of N⁶-methyladenosine (m⁶A). This epitranscriptional modification antagonizes the expression of selected mRNAs, including pluripotency factors whose transcription is promoted by SMAD2/3. Therefore, this provides a negative feedback that facilitates rapid exit from pluripotency upon inhibition of Activin/Nodal signalling.

Overall, the work presented in this dissertation advances the stem cell field in two ways. First, it demonstrates that the Activin/Nodal-SMAD2/3 pathway finely orchestrates the balance between pluripotency and differentiation by shaping both the epigenome and the epitranscriptome of hPSCs. Secondly, it provides a novel powerful technology to facilitate further studies of the mechanisms that regulate cell fate decisions.

CONTENTS

1 INTRODUCTION	1
1.1 EARLY STAGES OF HUMAN DEVELOPMENT.....	2
1.1.1 Pre-implantation: from zygote to blastocyst	2
1.1.2 Post-implantation: epiblast and gastrulation	3
1.1.3 Mechanisms controlling early development	4
1.2 HUMAN PLURIPOTENT STEM CELLS	8
1.2.1 Human embryonic stem cells.....	8
1.2.2 Human induced pluripotent stem cells	9
1.2.3 Naïve and primed hPSCs	11
1.2.4 Applications of hPSCs.....	13
1.3 CONTROL OF PLURIPOTENCY AND DIFFERENTIATION BY EXTRACELLULAR SIGNALLING	14
1.3.1 Signalling pathways regulating pluripotency.....	14
1.3.2 Signalling pathways regulating differentiation.....	17
1.3.2.1 Signalling during mesendoderm differentiation	17
1.3.2.2 Signalling during endoderm and mesoderm differentiation	19
1.3.2.3 Signalling during neuroectoderm differentiation	20
1.4 TRANSCRIPTIONAL CONTROL OF PLURIPOTENCY AND DIFFERENTIATION.....	21
1.4.1 Pluripotency factors.....	21
1.4.1.1 OCT4	21
1.4.1.2 NANOG.....	22
1.4.1.3 SOX2.....	22
1.4.2 Germ layer specification factors	24
1.4.2.1 Mesendoderm transcriptional regulators	24
1.4.2.2 Endoderm transcriptional regulators	25
1.4.2.3 Mesoderm transcriptional regulators	26
1.4.2.4 Neuroectoderm transcriptional regulators	26
1.5 CHROMATIN EPIGENETIC CONTROL OF PLURIPOTENCY AND DIFFERENTIATION .	27
1.5.1 Chromatin organization in pluripotent stem cells.....	27
1.5.2 Histone modifications.....	29
1.5.2.1 H3K4me3.....	29
1.5.2.2 H3K27me3.....	30
1.5.2.3 Bivalent domains	31

1.5.2.4 Additional notable chromatin modifications	33
1.6 POST-TRANSCRIPTIONAL CONTROL OF PLURIPOTENCY AND DIFFERENTIATION ..	35
1.6.1 Control of the transcriptome and of the proteome	35
1.6.2 mRNA modifications	36
1.6.2.1 m6A	36
1.6.2.2 Developmental function of m6A.....	39
1.7 ACTIVIN/NODAL SIGNALLING IN HPSCs	42
1.7.1 Activin/Nodal signalling	42
1.7.2 Contextual determinants of the response to Activin/Nodal signalling	44
1.7.3 SMAD2/3 cofactors.....	46
1.7.4 Control of hPSC fate choices by Activin/Nodal signalling	48
1.7.4.1 Transcriptional regulation by Activin/Nodal signalling in pluripotency and differentiation	48
1.7.4.2 Molecular mechanisms regulating SMAD2/3 function in pluripotency ...	49
1.7.4.3 Molecular mechanisms regulating SMAD2/3 function during differentiation	50
1.7.4.4 Open questions.....	53
1.8 METHODS TO STUDY GENE FUNCTION IN HPSCs	54
1.8.1 Genetic engineering	54
1.8.2 Conditional manipulation of gene expression	56
1.9 OBJECTIVES	59
<u>2 MATERIAL AND METHODS</u>	<u>61</u>
2.1 STATEMENT OF SOURCE	61
2.2 GENERAL METHODS FOR ALL CHAPTERS	62
2.2.1 hPSC culture	62
2.2.2 hPSC differentiation.....	63
2.2.2.1 Germ layer specification.....	63
2.2.2.2 Neural cell types.....	63
2.2.2.3 Neural crest cells and osteocytes	64
2.2.2.4 Chondrocytes	65
2.2.2.5 Smooth muscle cells.....	65
2.2.2.6 Epicardial cells and cardiac fibroblasts.....	65
2.2.2.7 Cardiomyocytes.....	66
2.2.2.8 Intestinal cells	66

2.2.2.9 Pancreatic cells	67
2.2.2.10 Hepatocytes	67
2.2.2.11 Cholangiocytes.....	68
2.2.2.12 Lung cells.....	68
2.2.3 Genetic manipulation of hPSCs	69
2.2.3.1 Lipofection.....	69
2.2.3.2 Nucleofection.....	69
2.2.4 Quantitative real-time polymerase chain reaction (qPCR).....	70
2.2.5 Western blot (WB)	75
2.2.6 Immunofluorescence (IF)	78
2.2.7 Flow cytometry.....	78
2.2.8 Immunoprecipitation (IP).....	78
2.2.9 Chromatin immunoprecipitation (ChIP)	79
2.2.10 Statistical analyses	84
2.3 METHODS SPECIFIC TO CHAPTER 3.....	85
2.3.1 Generation of stable knockdown hPSCs	85
2.3.2 Alkaline phosphatase assay	85
2.3.3 Proliferation curve.....	85
2.3.4 Apoptosis assay	86
2.3.5 Teratoma assay.....	86
2.3.6 Sequential ChIP.....	86
2.3.7 Microarray.....	87
2.3.8 Microarray data analysis	87
2.3.9 ChIP followed by deep sequencing (ChIP-seq)	89
2.3.10 ChIP-seq data analysis.....	89
2.3.11 Integration of ChIP-seq datasets.....	91
2.3.12 Integration of ChIP-seq and microarray data.....	93
2.3.13 Generation of knockout mice	93
2.3.14 Phenotyping of knockout mice	94
2.4 METHODS SPECIFIC TO CHAPTER 4.....	95
2.4.1 Molecular cloning.....	95
2.4.1.1 pTP6Neo_tetR.....	95
2.4.1.2 ROSA26 CRISPR/Cas9n vectors.....	96
2.4.1.3 pR26_ENDO-EGFP	96
2.4.1.4 pR26_EF1 α -EGFP	97
2.4.1.5 pR26_CAG-EGFP	98

2.4.1.6 pR26_CAG-STDtetR	98
2.4.1.7 pR26_CAG-OPTtetR	98
2.4.1.8 AAVS1 ZFNs	99
2.4.1.9 pAAV_iKD and related shRNA-containing plasmids	99
2.4.1.10 pAAV_iKD-2TO and pAAV_iKD-2TO-EGFP	101
2.4.1.11 pAAV_CAG-EGFP	101
2.4.1.12 pAAV_CAG-EGFP_H1-TOEGFPsh	101
2.4.1.13 pAAV_EGFPiKD-TO and pAAV_EGFPiKD-2TO	102
2.4.1.14 pAAV_EGFPsiKD-TO-STD and pAAV_EGFPsiKD-2TO-STD	102
2.4.1.15 pAAV_EGFPsiKD-TO-OPT and pAAV_EGFPsiKD-2TO-OPT	102
2.4.1.16 pAAV_siKD and related shRNA-containing plasmids	102
2.4.2 Gene targeting	103
2.4.3 Inducible gene knockdown	107
2.4.4 EGFP flow cytometry	107
2.5 METHODS SPECIFIC TO CHAPTER 5	108
2.5.1 SMAD2/3 IP for mass spectrometry (co-IP2 protocol)	108
2.5.2 Preparation of samples for mass spectrometry	109
2.5.3 Mass spectrometry	110
2.5.4 Biological interpretation of mass spectrometry data.	111
2.5.5 Generation of inducible knockdown hPSCs.....	111
2.5.6 RNA sequencing (RNA-seq)	113
2.5.7 RNA-seq data analysis.....	114
2.5.8 m6A dot blot.....	115
2.5.9 m6A methylated RNA immunoprecipitation (MeRIP)	116
2.5.10 m6A MeRIP followed by deep sequencing (MeRIP-seq).....	118
2.5.11 m6A MeRIP-seq data analysis	119
2.5.12 RNA co-immunoprecipitation (RNA co-IP)	120

3 ACTIVIN/NODAL CONTROLS THE H3K4ME3

EPIGENETIC LANDSCAPE OF HPSCS..... 123

3.1 STATEMENT OF SOURCE	123
3.2 HYPOTHESIS.....	124
3.3 RESULTS	125
3.3.1 Activin/Nodal signalling maintains H3K4me3 on specific regulators of pluripotency and differentiation	125

3.3.2 Activin/Nodal-dependent H3K4me3 is associated to gene expression changes	130
3.3.3 SMAD2/3 functionally interacts with DPY30 to maintain the pluripotent state	135
3.3.4 Mesendoderm differentiation of hPSCs relies on DPY30-dependent H3K4me3	141
3.3.5 SMAD2/3 cooperates with NANOG to recruit H3K4 methyltransferases on Activin/Nodal-responsive genes	145
3.3.6 Cooperation between DPY30, NANOG, and SMAD2/3 is necessary for H3K4me3 deposition on a core transcriptional network.....	151
3.3.7 DPY30, NANOG and SMAD2/3 colocalize at the genome-wide level	155
3.3.8 Dpy30 is necessary to maintain the pluripotent state of the post-implantation epiblast.....	157
3.4 DISCUSSION	160

4 DEVELOPMENT OF AN OPTIMIZED INDUCIBLE

KNOCKDOWN METHOD 165

4.1 STATEMENT OF SOURCE	165
4.2 HYPOTHESIS	166
4.3 RESULTS.....	167
4.3.1 Expression of inducible shRNAs from the human AAVS1 GSH	167
4.3.2 Validation of the ROSA26 and AAVS1 loci as <i>bona fide</i> GSHs during hPSC differentiation	170
4.3.3 Development of an optimized inducible knockdown method based on dual GSH targeting.....	176
4.3.4 Single-step generation of optimized inducible knockdown hPSCs.....	184
4.3.5 Validation of the optimized inducible knockdown platforms in hPSC-derived differentiated cells.....	187
4.3.6 Inducible DPY30 knockdown at various stages of hPSC differentiation reveals stage- and lineage-specific functions.	191
4.4 DISCUSSION	197

5 IDENTIFICATION OF NOVEL MOLECULAR MECHANISMS

REGULATED BY ACTIVIN/NODAL 203

5.1 HYPOTHESIS	203
-----------------------------	------------

5.2 RESULTS	204
5.2.1 Optimization of the SMAD2/3 co-immunoprecipitation protocol	204
5.2.2 Identification of the SMAD2/3 interactome	206
5.2.3 SMAD2/3 transcriptional and epigenetic cofactors regulate hPSC pluripotency and differentiation	211
5.2.4 The m6A methyltransferase complex inhibits the response to Activin/Nodal signalling	219
5.2.5 Activin/Nodal signalling promotes m6A deposition on specific regulators of pluripotency and differentiation	228
5.3 DISCUSSION	236
<u>6 FUTURE DIRECTIONS AND CONCLUSIONS</u>	<u>243</u>
6.1 FUTURE DIRECTIONS	243
6.1.1 Relationship between epigenome and epitranscriptome controls by SMAD2/3	243
6.1.2 Additional potential functions of SMAD2/3	244
6.1.3 Molecular mechanisms of SMAD2/3 activity in other models	244
6.1.4 Repurposing the OPTiKD methods for other applications	245
6.2 CONCLUSIONS	246
<u>7 APPENDICES</u>	<u>249</u>
APPENDIX I: H3K4ME3 PEAKS DECREASED AFTER INHIBITION OF ACTIVIN/NODAL SIGNALLING FOR 2 HOURS WITH SB	250
APPENDIX II: PLASMID MAPS FOR THE OPTiKD AND SOPTiKD SYSTEMS	258
APPENDIX III: SMAD2/3 INTERACTING PROTEINS IDENTIFIED BY QUANTITATIVE LABEL-FREE MASS SPECTROMETRY	265
APPENDIX IV: M6A PEAKS DECREASED AFTER INHIBITION OF ACTIVIN/NODAL SIGNALLING FOR 2 HOURS WITH SB	267
<u>8 REFERENCES</u>	<u>273</u>

LIST OF TABLES

Table 1. SMAD2/3 interactors.	47
Table 2. Media for hPSC culture and passaging.	62
Table 3. Primers for gene expression qPCR.	71
Table 4. Primary antibodies	76
Table 5. Secondary antibodies	77
Table 6. Buffers for IP and ChIP experiments.....	81
Table 7. Primers for ChIP-qPCR.	82
Table 8. Oligonucleotides for cloning of shRNAs (Chapter 3).....	101
Table 9. Summary of gene targeting experiments.....	104
Table 10. Genotyping strategies for gene targeting of the ROSA26 locus	105
Table 11. Genotyping strategies for gene targeting of the AAVS1 locus.....	106
Table 12. Buffers for SMAD2/3 IP and mass spectrometry	108
Table 13. Oligonucleotides for cloning of shRNAs (Chapter 5).....	113
Table 14. Buffers for m6A dot blot.....	115
Table 15. Buffers for m6A MeRIP.....	117
Table 16. Primers for m6A MeRIP-qPCR and RNA co-IP.	118

LIST OF FIGURES

Figure 1.1. Pre-implantation human development.	2
Figure 1.2. Human gastrulation.....	3
Figure 1.3. Regulation of pluripotency in the pre- and post-implantation mouse embryo.	6
Figure 1.4. Crosstalk between Nodal and BMP4 during mouse gastrulation.	7
Figure 1.5. Origin and developmental potential of hESCs.	9
Figure 1.6. Human induced pluripotent stem cells.	10
Figure 1.7. Different pluripotent states in murine cells.	12
Figure 1.8. Main signalling pathways regulating hPSC pluripotency.	17
Figure 1.9. Differentiation of hPSCs into mesendoderm derivatives.....	19
Figure 1.10. Neuroectoderm differentiation of hPSCs.	20
Figure 1.11. Core pluripotency transcriptional networks.....	23
Figure 1.12. Predominant model for the role of bivalent histone marks in pluripotency and differentiation.	31
Figure 1.13. Regulation of m6A deposition and function.	39
Figure 1.14. Current model for the role of m6A in murine pluripotency.	40
Figure 1.15. Activin/Nodal signalling.	43
Figure 1.16. Regulation of SMAD2/3 transcriptional activity in pluripotency and differentiation.	51
Figure 1.17. Nuclease-facilitated gene editing.....	55
Figure 1.18. Inducible gene knockdown using the TET-OFF system.....	57
Figure 3.1. Effect of SB treatment on SMAD2/3 genomic binding.....	125
Figure 3.2. ChIP-seq for H3K4me3 and H3K27me3 following inhibition of Activin/Nodal signalling.....	126
Figure 3.3. Functional enrichment analysis of differential H3K4me3 peaks.	126
Figure 3.4. Genomic features of Activin/Nodal-dependent H3K4me3.	127
Figure 3.5. Activin/Nodal-dependent H3K4me3 on active and poised promoters.	127
Figure 3.6. Activin/Nodal-dependent H3K4me3 on active enhancers.	128
Figure 3.7. Specificity of Activin/Nodal-dependent H3K4me3.	128
Figure 3.8. Global gene expression response to the inhibition of Activin/Nodal signalling.....	131
Figure 3.9. Expression profiles of genes regulated by Activin/Nodal signalling.....	131
Figure 3.10. qPCR validation of microarray experiments.	132

Figure 3.11. Relationship between gene expression and epigenetic changes following inhibition of Activin/Nodal signalling.....	132
Figure 3.12. Relationship between early epigenetic changes and gene expression dynamics following inhibition of Activin/Nodal signalling.	133
Figure 3.13. H3K4me3 and H3K27me3 dynamics following inhibition of Activin/Nodal signalling.	134
Figure 3.14. SMAD2/3 interacts with H3K4 methyltransferases and DPY30.....	135
Figure 3.15. Generation of DPY30 knockdown hPSCs.....	137
Figure 3.16. Effect of DPY30 knockdown on hPSC self-renewal.	137
Figure 3.17. Effect of DPY30 knockdown on the expression of pluripotency and differentiation markers.	138
Figure 3.18. Global similarities between the gene expression changes induced by knockdown of DPY30 and by the inhibition of Activin/Nodal signalling.	139
Figure 3.19. Effect of DPY30 knockdown on the H3K4me3 levels onto Activin/Nodal target genes.	140
Figure 3.20. Gene expression dynamics following transient DPY30 knockdown.....	140
Figure 3.21. Teratoma assays of DPY30 knockdown hPSCs.	141
Figure 3.22. Transcriptional analysis of DPY30 knockdown cells differentiated into the germ layers.	142
Figure 3.23. Transcription of Activin/Nodal target genes during endoderm differentiation of DPY30 knockdown cells.	142
Figure 3.24. Expression of endoderm marker proteins during differentiation of DPY30 knockdown cells.	143
Figure 3.25. Differentiation of DPY30 knockdown cells into mature endodermal cell types.	143
Figure 3.26. H3K4me3 deposition during endoderm differentiation of DPY30 knockdown cells.	144
Figure 3.27. Effect of NANOG knockdown on the expression of pluripotency and differentiation markers.	145
Figure 3.28. Global similarities between the transcriptional responses induced by the knockdown of NANOG, the knockdown of DPY30, and the inhibition of Activin/Nodal signalling.	146
Figure 3.29. SMAD2/3, NANOG and DPY30 form a chromatin-interacting complex. .	146
Figure 3.30. Effect of NANOG knockdown on the levels of H3K4me3 onto Activin/Nodal target genes.	147
Figure 3.31. SMAD2/3 and NANOG recruit DPY30 onto their genomic targets.	148

Figure 3.32. SMAD2/3 recruits the MLL2 complex onto its genomic targets.....	149
Figure 3.33. Binding of the MLL2 complex to SMAD2/3 target genes does not require DPY30.	150
Figure 3.34. ChIP-seq for H3K4me3 and H3K27me3 following knockdown of DPY30 or NANOG.	151
Figure 3.35. Similarities between the H3K4me3 changes induced by the knockdown of NANOG, the knockdown of DPY30, and the inhibition of Activin/Nodal signalling.....	152
Figure 3.36. Correlation between H3K4me3 decrease after knockdown of NANOG and DPY30 and gene expression changes following inhibition of Activin/Nodal signalling.....	153
Figure 3.37. Progressive loss of H3K4me3 during chronic Activin/Nodal inhibition. ..	153
Figure 3.38. H3K27me3 is unaffected by acute Activin/Nodal inhibition.	154
Figure 3.39. Validation of DPY30 ChIP-seq.	155
Figure 3.40. Global overlap between SMAD2/3, NANOG, and DPY30.	156
Figure 3.41. Binding of SMAD2/3, NANOG, and DPY30 on loci showing Activin/Nodal- dependent H3K4me3.....	156
Figure 3.42. Genotyping of Dpy30 knockout mice.....	157
Figure 3.43. Role of Dpy30 during mouse post-implantation development.....	158
Figure 3.44. Expression of pluripotency and differentiation markers in DPY30 knockout embryos.....	159
Figure 3.45. Proposed model for the functional interaction between SMAD2/3, NANOG, and DPY30-containing COMPASS complexes.	159
Figure 4.1. Generation of tetR expressing hPSCs.....	167
Figure 4.2. Expression of an inducible shRNA against OCT4 from the AAVS1 locus. ..	168
Figure 4.3. Leakiness of the OCT4 inducible shRNA in different culture media.	169
Figure 4.4. Expression of an inducible shRNA against DPY30 from the AAVS1 locus. .	169
Figure 4.5. Generation of ROSA26 reporter hPSCs using different promoters.....	171
Figure 4.6. Comparison of ROSA26 reporter hPSCs.	171
Figure 4.7. Experimental strategy to validate the AAVS1 and ROSA26 loci as “ <i>bona fide</i> ” genomic safe harbours.	172
Figure 4.8. Generation of ROSA26 and AAVS1 CAG-EGFP reporter hPSCs.	172
Figure 4.9. Expression of the ROSA26 and AAVS1 reporters in undifferentiated cells and in the three germ layers.	173
Figure 4.10. Differentiation potential of ROSA26 and AAVS1 reporter hPSCs.	174

Figure 4.11. Expression quantification of the ROSA26 and AAVS1 reporters in differentiated cells.	174
Figure 4.12. Expression of the ROSA26 and AAVS1 reporters in differentiated cells. .	175
Figure 4.13. Experimental strategy to generate inducible knockdown hPSCs following dual GSH targeting.	176
Figure 4.14. Generation of inducible EGFP knockdown hPSCs.	177
Figure 4.15. Inducible EGFP knockdown in hPSCs.....	177
Figure 4.16. Codon optimization of the tetR.	178
Figure 4.17. Expression of the codon-optimized tetR.....	179
Figure 4.18. Comparison of H1 inducible promoters.	179
Figure 4.19. Optimization of EGFP inducible knockdown.....	180
Figure 4.20. Dose responsiveness of EGFP OPTiKD.....	180
Figure 4.21. Dynamics of inducible EGFP knockdown and rescue.	181
Figure 4.22. OPTimized inducible KnockDown.....	181
Figure 4.23. Generation of OCT4 and B2M OPTiKD hPSCs.....	182
Figure 4.24. Efficiency of OCT4 OPTiKD hPSC generation.....	182
Figure 4.25. Effect of OCT4 inducible knockdown in hPSCs.	183
Figure 4.26. Optimization of single-step EGFP inducible knockdown.	184
Figure 4.27. Dose responsiveness and reversibility of EGFP sOPTiKD.....	185
Figure 4.28. Single-step OPTimized inducible KnockDown.....	185
Figure 4.29. Generation of OCT4 and B2M sOPTiKD hESCs and hiPSCs.	186
Figure 4.30. Effect of OCT4 inducible knockdown in hESCs and hiPSCs.	186
Figure 4.31. Performance of OPTiKD and sOPTiKD in differentiated cells.	187
Figure 4.32. Validation of EGFP OPTiKD in differentiated cells.	188
Figure 4.33. Validation of EGFP sOPTiKD in differentiated cells.....	189
Figure 4.34. Prolonged EGFP inducible knockdown in differentiated cells.	190
Figure 4.35. Generation of DPY30 OPTiKD hPSCs.	191
Figure 4.36. Inducible knockdown of DPY30 in hPSCs.	191
Figure 4.37. Inducible knockdown of DPY30 at various stages of endoderm differentiation.	192
Figure 4.38. Inducible DPY30 knockdown during neuroectoderm and lateral plate mesoderm differentiation.	193
Figure 4.39. Inducible DPY30 knockdown at specific stages of hPSC differentiation into mature cell types.....	193
Figure 4.40. Validation of DPY30 knockdown during hPSC differentiation.	194
Figure 4.41. Effect of DPY30 inducible knockdown during hPSC differentiation.	195

Figure 4.42. Knockdown of DPY30 specifically during the early induction of pancreatic endocrine cells.	196
Figure 4.43. Lineage- and stage-specific effects following DPY30 knockdown during hPSC differentiation.	196
Figure 5.1. Comparison of SMAD2/3 co-immunoprecipitation protocols.	205
Figure 5.2. Identification of the SMAD2/3 interactome by label-free quantitative mass spectrometry.	206
Figure 5.3. The SMAD2/3 interactome in hPSCs and early endoderm cells.	207
Figure 5.4. Functional role of SMAD2/3 interacting proteins.	208
Figure 5.5. Known interactions between the SMAD2/3 interacting proteins.	209
Figure 5.6. Interaction dynamics of SMAD2/3 transcriptional and epigenetic cofactors in pluripotent and early endoderm cells.	211
Figure 5.7. Functional characterization of the SMAD2/3 interactome.	212
Figure 5.8. Validation of SMAD2/3 cofactors inducible gene knockdown.	213
Figure 5.9. Effect of SMAD2/3 cofactors knockdown on hPSC morphology.	214
Figure 5.10. Effect of SMAD2/3 cofactors knockdown on hPSC pluripotency.	215
Figure 5.11. Transcriptional effects following knockdown of SMAD2/3 cofactors.	215
Figure 5.12. Validation of SMAD2/3 cofactors gene knockdown during endoderm differentiation.	216
Figure 5.13. Effect of SMAD2/3 cofactors knockdown during endoderm differentiation of hPSCs.	217
Figure 5.14. Transcriptional effects following knockdown of SMAD2/3 cofactors during endoderm differentiation of hPSCs.	217
Figure 5.15. Function of SMAD2/3 cofactors in hPSC pluripotency and endoderm differentiation.	218
Figure 5.16. Interaction of SMAD2/3 with the m6A methyltransferase complex in hPSCs.	219
Figure 5.17. Inducible knockdown of the m6A methyltransferase complex subunits in hPSCs.	220
Figure 5.18. Effect of m6A methyltransferase complex knockdown on hPSC pluripotency.	221
Figure 5.19. Effect of m6A methyltransferase complex knockdown on the deposition of m6A.	221
Figure 5.20. Effect of m6A methyltransferase complex knockdown on endoderm differentiation of hPSCs.	222

Figure 5.21. Effect of m6A methyltransferase complex knockdown on neuroectoderm differentiation of hPSCs.	223
Figure 5.22. Effect of m6A methyltransferase complex knockdown on the downregulation of NANOG following Activin/Nodal inhibition.	224
Figure 5.23. Effect of WTAP knockdown on the expression of Activin/Nodal target genes.	225
Figure 5.24. Genome-wide gene expression changes following Activin/Nodal inhibition in WTAP knockdown hPSCs.	226
Figure 5.25. Transcriptional profiles of genes differentially expressed following Activin/Nodal inhibition in WTAP knockdown hPSCs.	226
Figure 5.26. Global effects of WTAP knockdown on the response to Activin/Nodal signalling in hPSCs.	227
Figure 5.27. Deposition of m6A on transcripts regulated by WTAP or by Activin/Nodal signalling.	227
Figure 5.28. Overview of m6A MeRIP experiments.	228
Figure 5.29. m6A MeRIP-qPCR in hPSCs following Activin/Nodal inhibition.	228
Figure 5.30. m6A MeRIP-seq in hPSCs following Activin/Nodal inhibition.	229
Figure 5.31. Features of m6A deposition in hPSCs.	230
Figure 5.32. Analysis of differential m6A deposition following Activin/Nodal inhibition.	231
Figure 5.33. Correlation between differential m6A deposition and gene expression changes following Activin/Nodal inhibition.	231
Figure 5.34. Functional enrichment analysis of differentially methylated transcripts.	232
Figure 5.35. Differential m6A deposition following Activin/Nodal inhibition.	232
Figure 5.36. Sensitivity of the interaction between SMAD2/3 and the m6A methyltransferase complex to changes in Activin/Nodal signalling.	233
Figure 5.37. Genome-wide correlation between Activin/Nodal-dependent m6A deposition and SMAD2/3 binding.	234
Figure 5.38. DNA binding of the m6A methyltransferase complex.	234
Figure 5.39. RNA binding of SMAD2/3.	235
Figure 5.40. Proposed model for the functional interaction between SMAD2/3 and the m6A methyltransferase complex in hPSCs.	235

LIST OF ABBREVIATIONS AND ACRONYMS

Abbreviation	Meaning
2i	Dual MEK1/2 and GSK3 β inhibition
3' UTR	(mRNA) 3' untranslated region
3'-BB (PCR)	PCR of vector backbone 3'-end genomic integration
3'-INT (PCR)	PCR of transgene 3'-end genomic integration
5' UTR	(mRNA) 5' untranslated region
5'-BB (PCR)	PCR of vector backbone 5'-end genomic integration
5'-INT (PCR)	PCR of transgene 5'-end genomic integration
AAVS1/AAV	AAVS1 locus: PPR1R12C gene on human chromosome 19
abs.FC	Absolute fold-change
adj.p	Adjusted p-value
ALK	(Type I) Activin receptor-like kinases
AP	Alkaline phosphatase
AVE	Anterior visceral endoderm
BB	Binding buffer
BCIP	5-Bromo-4-chloro-3-indolyl phosphate
BMP	Bone morphogenetic protein (signalling)
bp	(Nucleotide/nucleoside) base pairs
BRA	Brachyury
BPTF	Bromodomain PHD finger transcription factor
BS	Binding site
BSA	Bovine serum albumin
CAG	CMV early enhancer, chicken β -actin and rabbit β -globin hybrid promoter
Cas9n	<i>S. Pyogenes</i> Cas9 D10A nickase mutant
CBP/p300	CREB-binding protein/p300 (acetyltransferases)
CBX	Chromobox homolog (proteins)
CDB	Cell dissociation buffer
CDM	Chemically defined medium
cDNA	Complementary DNA
CFIm	Cleavage factor Im complex
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation followed by deep sequencing
ChIP-XL	Chromatin immunoprecipitation performed using additional protein-protein cross-linkers (for transcription factors such as SMAD2/3)
CIP	Calf intestinal alkaline phosphatase
CLB	Cell lysis buffer

co-IP	Co-immunoprecipitation
co-IP1	Co-immunoprecipitation protocol 1
co-IP2	Co-immunoprecipitation protocol 2
co-SMAD	Common-mediator SMAD
COMPASS	Complex proteins associated with Set1
CpG	Cytosine-phosphate-guanine dinucleotide
CRISPR	Clustered regularly interspaced short palindromic repeat
Ct	Cycle threshold
CTR	Control
DAPI	4',6-Diamidine-2'-phenylindole dihydrochloride
DB	Dialysis buffer
DE	Definitive endoderm
DFG	Dorsal foregut
DMSO	Dimethyl sulfoxide
dNTP(s)	Deoxynucleotide(s)
DPY30 KD hESCs	hESC sublines knockdown for DPY30
DSE	Distal sequence element
DTT	Dithiothreitol
E(6.5)	Embryonic day (6.5, as an example)
EB	Elution buffer
EDTA	Ethylenediaminetetraacetic acid
EF1 α	Elongation factor 1 α (promoter)
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescent protein
ELD	Q486E, I499L, N496D (FokI domain mutations)
En2SA	Engrailed-2 exon-2 splice acceptor
ENDO	Endoderm
ERK	Extracellular signal-regulated kinase (pathway)
ExE	Extraembryonic ectoderm
FB	Fragmentation buffer
FBS	Foetal bovine serum
FC	Fold-change
FDR	False discovery ratio
FGF	Fibroblast growth factor (signalling)
FLyAB	Culture conditions containing 12ng/ml FGF2, 10 μ M LY-294002, 100ng/ml Activin, and 10ng/ml BMP4.
FRT	Flippase recognition target site
G1	Gap 1 (cell cycle phase)
GAT	Genomic association test

GB	Gene body
GCN5	Gcn5-related N-acetyltransferases
GO	Gene ontology
GREAT	Genomic regions enrichment of annotations
gRNA(s)	(CRISPR/Cas9) guide RNA(s)
GSEA	Gene set enrichment analysis
GSH(s)	Genomic safe harbour(s)
GTF	Gene transfer format
GTPases	Guanosine triphosphate hydrolases
H1-2TO	Tetracycline-inducible H1 Pol III promoter carrying two tet operons, one before and one after the TATA box
H1-TO	Tetracycline-inducible H1 Pol III promoter carrying one tet operon after the TATA box
H2AK119ub	Histone 2A lysine 119 ubiquitination
H3K18ac	Histone 3 lysine 18 acetylation
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 lysine 27 trimethylation
H3K36me3	Histone 3 lysine 36 trimethylation
H3K4me1	Histone 3 lysine 4 (mono)methylation
H3K4me2	Histone 3 lysine 4 dimethylation
H3K4me3	Histone 3 lysine 4 trimethylation
H3K9me3	Histone 3 lysine 9 trimethylation
H4K16ac	Histone 4 lysine 16 acetylation
H9	Human embryonic stem cell line 9
HAR(s)	Homology arm(s)
HAT(s)	Histone acetyltransferase(s)
hBactP	Human β -actin promoter
HCD	Higher energy collision induced dissociation
HDAC(s)	Histone deacetylase(s)
HDR	Homologous directed repair
hESC(s)	Human embryonic stem cell(s)
HET	Heterozygous
HET+	Heterozygous with additional off-target integration of the targeting vector
HGF	Hepatocyte growth factor
hiPSC(s)	Human induced pluripotent stem cell(s)
HLB	Hypotonic lysis buffer
HOM	Homozygous
HOM+	Homozygous with additional off-target integration of the targeting vector
HP1	Heterochromatin protein 1

hPAN	PARBP-dependent poly-A nuclease complex
hPSC(s)	Human pluripotent stem cell(s)
HRP	Horseradish peroxidase
HSB	High-salt buffer
HSNLB	High-salt nuclear lysis buffer
I-SMAD	Inhibitory SMAD
IC50	Half-maximal inhibitory concentration
iCLIP	Individual-nucleotide resolution cross-linking and immunoprecipitation
ICM	Inner cell mass
IDB	Immunoprecipitation dilution buffer
IF	Immunofluorescence
IGF	Insulin-like growth factor (pathway)
IgG	Class G immunoglobulins
iKD	Inducible knockdown
ILB	Isotonic lysis buffer
IP	Immunoprecipitation
IRES	Intronic ribosomal entry site
IWB1	Immunoprecipitation wash buffer 1
IWB2	Immunoprecipitation wash buffer 2
JAK	Janus activated kinase (signalling)
JmjC	Jumonji C-domain-containing (proteins)
Kb	Kilo base pairs
K_d	Dissociation constant
KD	Knockdown
KKR	E490K, I538K, H537R (FokI domain mutations)
KSR	Knockout serum replacer
LacZ	β -galactosidase (cDNA)
LB	Luria Bertani medium
LFQ	Label free quantification (intensity)
LIF	Leukemia inhibitory factor (pathway)
lncRNA(s)	Long non-coding RNA(s)
Locus (PCR)	PCR of wild-type locus (AAVS1/ROSA26)
loxP	Bacteriophage P1 Cre recombinase recognition site
LSB	Low-salt buffer
m6A	N ⁶ -methyladenosine
m6A-seq	N ⁶ -methyladenosine profiling by deep sequencing
MAPK	Mitogen-activated protein kinase (pathway)
MAPQ	Mapping quality
MEF(s)	Mouse embryonic fibroblast(s)

mEpiSC(s)	Mouse epiblast stem cell(s)
MeRIP	Methylated RNA immunoprecipitation
MeRIP-seq	Methylated RNA immunoprecipitation followed by deep sequencing
mESC(s)	Mouse embryonic stem cell(s)
MESO	Mesoderm
MFI	Median fluorescence intensity
MGI	Mouse Genomic Informatics
MH1	Mad homology 1 (domain)
MH2	Mad homology 2 (domain)
miRNA(s)	Micro RNA(s)
MLL	Mixed-lineage leukemia (enzymes)
MOPS	3-(N-morpholino)propanesulfonic acid (buffer)
mRNA(s)	Messenger RNA(s)
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MYST	MOZ-Ybf2-Sas2-Tip60 (acetyltransferases)
NBT	Nitro blue tetrazolium
Neo	Neomycin resistance gene (Neomycin phosphotransferase II)
NEURO	Neuroectoderm
NHEJ	Non-homologous end joining
NLB	Nuclear lysis buffer
NuRF	Nucleosome remodelling factor
OPTiKD	OPTimized inducible KnockDown
OPTtetR	Codon-optimized tetR
p	P-value
P-bodies	(RNA) processing bodies
P(14)	Postnatal day (14, as an example)
pA	Polyadenylation signal
PAM	Protospacer adjacent motif
PAR-CLIP	Photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation
PBS	Phosphate-buffered saline
PBST	PBS Tween buffer
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PCR	Polymerase chain reaction
PDGF-AA	Platelet-derived growth factor type "AA"
PDGF-BB	Platelet-derived growth factor type "BB"
PFA	Paraformaldehyde

PHD	Plant homeodomain
PI3K	Phosphatidylinositide 3-kinase (pathway)
PKA	Protein kinase A
PKB/AKT	Protein kinase B (pathway)
PLURI	Pluripotent
PMSF	Phenylmethane sulfonyl fluoride
Pol III	RNA polymerase III
poly-A	Polyadenylated (RNA)
PRC1	Polycomb repressive complex 1
PRC2	Polycomb repressive complex 2
PrE	Primitive endoderm
PSC(s)	Pluripotent stem cell(s)
PSE	Proximal sequence element
Puro	Puromycin resistance gene (Puromycin N-acetyltransferase)
PVA	Polyvinyl alcohol
PVDF	Polyvinylidene fluoride
QC	Quality control
qPCR	Quantitative real-time polymerase chain reaction
R-SMAD	Receptor-regulated SMAD
RA	Retinoic acid
RBP(s)	RNA binding protein(s)
RLB	RNA loading buffer
RNA-seq	RNA sequencing
ROSA26/R26	ROSA26 locus: THUMPDS3-AS1 gene on human chromosome 3
RP-HPLC MS/MS	Reverse phase high-performance liquid chromatography tandem mass spectrometry
RPKM	Reads per kilobase of transcript per million mapped reads
RPM	Reads per million mapped reads
RRHO	Rank-rank hypergeometric overlap analysis
RT	Room temperature
rtTA	Reverse tetracycline-controlled transactivator
S/N/D	SMAD2/3, NANOG and DPY30 complex
SA	Splice acceptor
SAGA	Spt-Ada-Gcn5 acetyltransferase
SB	SB431542
SBE(s)	SMAD binding element(s)
SCID	Severe combined immunodeficiency
SCR	Scramble (shRNA)
SDS	Sodium dodecyl sulphate

SEM	Standard error of the mean
shRNA(s)/sh	Short hairpin RNA(s)
SMAD	Sma- and Mad-related (protein)
SMAD2/3	SMAD2 and SMAD3
SNLB	SDS-nuclear lysis buffer
sOPTiKD	Single-step OPTimized inducible KnockDown
SSC	Saline-sodium citrate buffer
SSH	Sonic hedgehog (signalling)
SSPE	Saline-sodium phosphate-EDTA buffer
STDtetR	Standard tetR
Suv39	Su(Var)3-9 (methyltransferase)
T2A	Self-cleaving T2A peptide
TALEN(s)	Transcription activator-like effector nuclease(s)
TBST	TBS Tween buffer
TE	Trophoectoderm
TET	Tetracycline (treatment)
TET-OFF	Tetracycline derepressible
TET-ON	Tetracycline activatable
tetR	Tetracycline-sensitive repressor protein
tetR-nls	Bacterial tetR sequence containing a 5'-terminus SV40 nuclear localization sequence
TFIID	Transcription factor II D
TGF β	Transforming growth factor β (signalling/superfamily)
TO/TETO2	Tet operon
TRE	Tetracycline responsive element
tRNA(s)	Transfer RNA(s)
TSS	Transcription start site
TTB	TE-Tween buffer
UV	Ultraviolet light
VEGF-B	Vascular endothelial growth factor B
vs	Versus
WB	Western blot
WD40	Tryptophan-aspartic acid repeat (domain)
WNT	Wingless-related integration site (signalling)
WRAD	WDR5, RBBP5, ASHL2 and DPY30 (module)
YTH	YT521-B homolog (domain)
Z-score(s)	Standard score(s)
ZFN(s)	Zinc-finger nuclease(s)
$\Delta\Delta Ct$	Delta-delta cycle threshold

LIST OF APPENDICES

Appendix I: H3K4me3 peaks decreased after inhibition of Activin/Nodal signalling for 2 hours with SB	250
Appendix II: Plasmid maps for the OPTiKD and sOPTiKD systems.....	258
Appendix III: SMAD2/3 interacting proteins identified by quantitative label-free mass spectrometry.....	265
Appendix IV: m6A peaks decreased after inhibition of Activin/Nodal signalling for 2 hours with SB	267

1 INTRODUCTION

This chapter begins by describing the earliest stages of human development, and by illustrating how human pluripotent stem cells (hPSCs) can be used to model and study these events. Then, it presents the major signalling pathways, transcriptional networks, epigenetic controls, and post-transcriptional regulations that dictate hPSC early cell fate decisions. Further, a specific focus is dedicated to the mechanisms by which the Activin/Nodal signalling pathway controls both pluripotency and differentiation of hPSCs. Finally, following an overview of the current technologies to study gene function in hPSCs, the aims of this dissertation are presented.

1.1 Early stages of human development

This sub-chapter provides a general background to the stages of human development most relevant to this dissertation. Moreover, it describes the mechanistic notions about such events that can be inferred from studies in model organisms.

1.1.1 Pre-implantation: from zygote to blastocyst

The developmental journey of a new human being starts when the male and female gametes meet in the fallopian tube of the female uterus. The fusion of these two haploid cells is known as fertilization and generates the zygote, the earliest diploid cell in human development (Figure 1.1; Carlson 2014). Such cell is totipotent, as it can differentiate both into any cell type of the embryo proper and into all of the extraembryonic tissues (Niakan et al. 2012). During its migration towards the uterus, the zygote undergoes at least four rounds of cell division without significant growth, and generates a morula comprised of at least sixteen totipotent cells named blastomeres. Following compaction of the 16-32 cells morula, cellular differentiation begins during blastulation (Deglincerti et al. 2016; Shahbazi et al. 2016). This process involves the formation of the trophoblast, an outer layer of cells that will contribute to the foetal part of the placenta, and of the inner cell mass (ICM), a group of cells that will give rise to the embryo, the amnion, the yolk sack, and the allantois (reviewed in Arnold and Robertson, 2009; Parfitt and Shen, 2014). Inside the trophoblast and to one side of the ICM a liquid-filled cavity gradually enlarges (known as blastocyst cavity or blastocoel). At this stage, the embryo is named blastocyst (Figure 1.1). After six to nine days post-fertilization, the late blastocyst hatches from the zona pellucida, a membranous structure that enclosed the conceptus so far, and implants into the uterine wall.

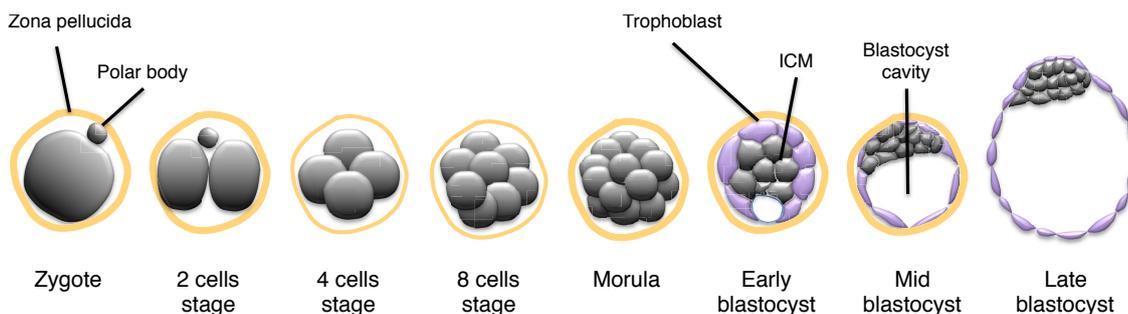


Figure 1.1. Pre-implantation human development.

1.1.2 Post-implantation: epiblast and gastrulation

Following implantation, the ICM further specializes into two cellular layers: the hypoblast, which is closest to the blastocyst cavity and will develop into the extraembryonic endoderm, and the epiblast, which is adjacent to the trophoblast and will differentiate into the three primary germ layers, the amniotic ectoderm, and the extraembryonic mesoderm (Figure 1.2; Deglincerti et al., 2016; Shahbazi et al., 2016). The resulting bilaminar embryo is called embryonic disc, and represents the last stage at which a cell type (the epiblast) has the ability to contribute to all adult cell types. Such competency is defined as pluripotency, and it is lost at around two weeks post fertilization during the process of gastrulation, when the epiblast further specializes into the three primary germ layers: ectoderm, mesoderm and endoderm (reviewed in Rossant and Tam, 2009; Tam and Loebel, 2007).

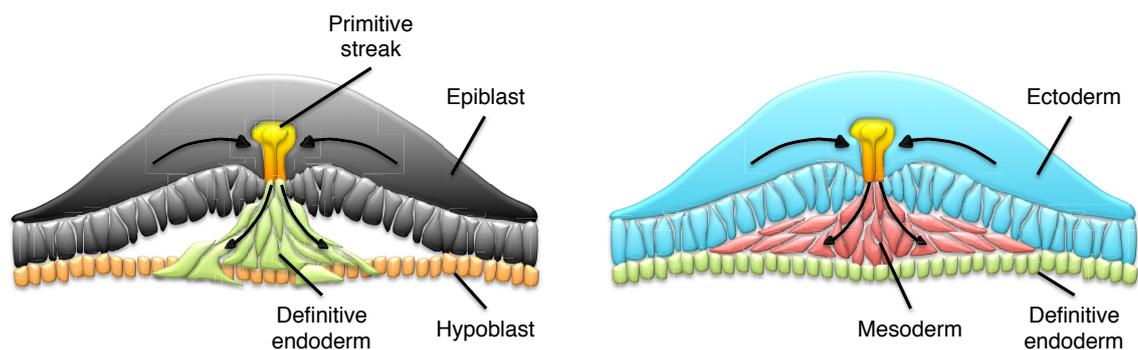


Figure 1.2. Human gastrulation.

The process of gastrulation requires the epiblast to lose its symmetry, with the formation of the proximal-distal and anterior-posterior axes. The start of gastrulation is marked by the appearance of the primitive streak, a linear band of migrating epiblast cells localized at the posterior side of the embryo and juxtaposed to the extraembryonic tissue. Epiblast cells move down into such structure while undergoing an epithelial-mesenchymal transition, a process called ingression. The embryonic endoderm (also known as definitive endoderm) is formed by epiblast cells that while migrating through the anterior side of the primitive streak take the place of the hypoblast, which is moved aside. A middle layer named mesoderm is formed by cells migrating through a more posterior section of the streak, juxtaposing themselves between the epiblast and the definitive endoderm. Among these cells, some migrate outside of the embryo proper to generate the extraembryonic mesoderm. Ectoderm is finally formed by the remaining

more anterior and lateral epiblast cells that do not ingress through the streak. As a result of this, a three-layered embryo is formed, which goes on to develop into all the structures and organs of the body (Figure 1.2; Carlson, 2014). Ectoderm will contribute to the outermost layers of the skin, the eyes, many connective tissues, the inner ear, and both the central and peripheral nervous systems. Accordingly, this germ layer is also often referred to as neuroectoderm, a convention that is maintained throughout this dissertation. Mesoderm will develop into most of the connective tissues including bone, cartilage, muscle, the circulatory system, the heart, blood cells, and kidneys. Finally, definitive endoderm gives rise to the gut tube and its accessory organs, including lungs, liver, pancreas, and bladder.

1.1.3 Mechanisms controlling early development

Studying the earliest stages of human development poses complex ethical and practical limitations. Indeed, while *in vitro* fertilization techniques potentially allow the analysis of pre-implantation development (Niakan et al. 2012), the use of such embryos for research applications is ethically controversial, is subjected to tight regulatory frameworks that prevent certain types of experimentation, and is even banned in certain countries (Daley et al. 2007; Hyun 2010). Moreover, while *in vitro* culture of human embryos up to the 13th day post-fertilization has been recently reported (Deglincerti et al. 2016; Shahbazi et al. 2016), analyses after the 14th day (which conventionally marks the onset of gastrulation) are currently prohibited in most countries. Finally, alternative biological samples can only be obtained in very limited quantities from abortive material, and not during the early time window in which gastrulation occurs. Overall, certain initial stages of post-implantation development of the human embryo are virtually a “black box”.

As a result, most of our understanding about the mechanisms that control early human development is actually inferred from studies performed in model organisms such as chicks, fishes, frogs, and mice. However, it is unclear to what extent such notions can be directly applied to the human model (Chapter 1.2 will describe how hPSCs have been used for *in vitro* experiments aiming to bridge this gap in our knowledge). Due to the extensiveness of the topic, this dissertation will focus on the mouse model when describing *in vivo* developmental mechanisms, as this is evolutionary closest to humans.

In the mouse embryo, the first cell fate decision that occurs in the developing morula is mainly instructed by the mutually exclusive distribution of two transcription factors: Oct4 (also known as Pou5f1), whose expression becomes gradually restricted to the ICM, and Cdx2, which is on the other hand secluded to the trophoectoderm (TE, a structure corresponding to the human trophoblast; Niwa et al., 2005). This specific expression pattern is achieved following asymmetric cell divisions that are reinforced both by a reciprocal antagonistic regulation between Oct4 and Cdx2, and by the cell position-specific activity of the Hippo pathway, which promotes expression of Cdx2 in TE cells (Dietrich & Hiiragi 2007; Nishioka et al. 2009; Ralston & Rossant 2008).

Similarly, subsequent specification of the ICM is primarily dictated by selective expression of the transcription factors Nanog in the epiblast, while Gata6 is restricted to the primitive endoderm (PrE; corresponding to the human hypoblast; Chambers et al., 2003; Fujikura et al., 2002; Mitsui et al., 2003). This process is dictated both by cell position and by fibroblast growth factor (Fgf) signalling (Meilhac et al. 2009). In particular, it is largely dependent on the epiblast-specific expression of Fgf4, which is directly activated by Nanog and promotes Gata6 expression in the PrE (Chazaud et al. 2006; Frankenberg et al. 2011; Morrisey et al. 1998).

Of note, the maintenance of pluripotency in rodents is differentially regulated in the pre- and post-implantation embryo (Figure 1.3). In the early ICM, pluripotency is antagonized by Fgf4, and promoted by signalling *via* the Janus activated kinase (Jak)-Stat3 axis through leukemia inhibitory factor (Lif) and other yet unknown ligands (Do et al., 2013; Nichols et al., 2001; Stewart et al., 1992). On the other hand, pluripotency in the epiblast becomes independent from Jak-Stat3 activation (Posfai et al., 2014). Moreover, bone morphogenetic protein (Bmp) secreted from the TE tissues (the extraembryonic ectoderm, ExE) signals to the underlying epiblast to promote proliferation (Mishina et al. 1995; Lawson et al. 1999). Finally, Nodal signalling in the epiblast is the key factor that preserves its pluripotent state, as absence of Nodal leads to loss of pluripotency markers and to premature and ectopic neuroectoderm differentiation (Camus et al. 2006; Mesnard et al. 2006). Nodal is expressed throughout the early post-implantation epiblast and the overlaying PrE, and is later restricted to the posterior proximal epiblast (Varlet et al. 1997).

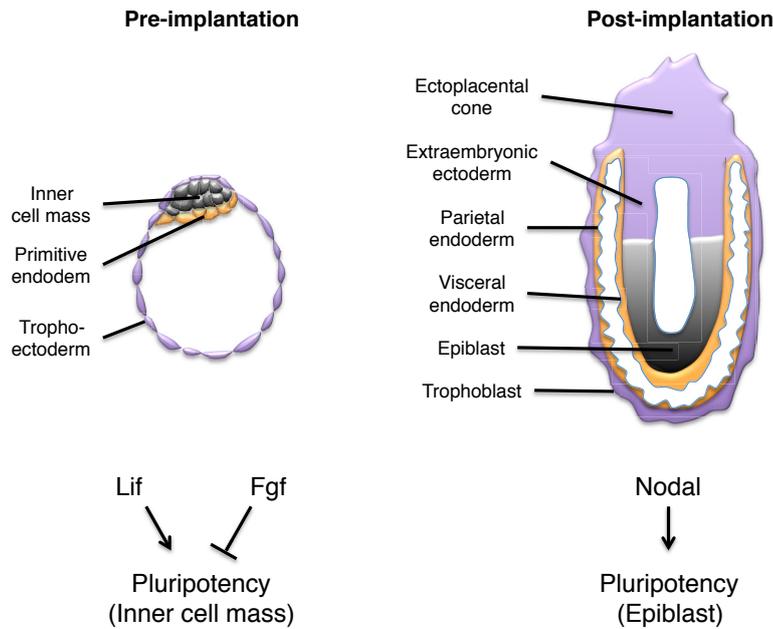


Figure 1.3. Regulation of pluripotency in the pre- and post-implantation mouse embryo.

Simplified schematic of the main pathways involved. Arrows: positive interaction; broken lines: negative regulation.

Gastrulation is a complex process dictated by the interplay of several signalling pathways. Importantly, the mouse epiblast significantly differs from its human counterpart because it is cup-shaped instead of a flat disc, an adaptation thought to facilitate accommodation of multiple embryos in the limited uterine space. In the mouse, the proximal-distal axis is established first by the interplay of ExE and the epiblast (reviewed in Arnold and Robertson, 2009). Then, two main signalling centres determine the anterior-posterior axis, thus driving gastrulation. First, the anterior visceral endoderm (AVE) is localized on the anterior side of the visceral PrE, and produces antagonists of both the Nodal and Wnt pathways (such as Lefty1, Cer1 and Dkk; Mukhopadhyay et al., 2001; Perea-Gomez et al., 2001; Yamamoto et al., 2004). Secondly, the node marks the most anterior part of the primitive streak, and produces the Bmp antagonists Chordin and Noggin (Bachiller et al. 2000). Collectively, these inhibitors counteract signals from the extraembryonic tissues (Bmp; Hogan, 1996) and the epiblast (Nodal and Wnt3; Schier, 2003; Yamaguchi, 2001), thus forming various posterior-anterior gradients for such signalling molecules. Fgf8 forms another such gradient since its rapid messenger RNA (mRNA) decay results in higher expression into the rapidly-proliferating posterior epiblast (Dubrulle & Pourquié 2004). Overall, the interplay between these gradients not only induces the formation of the primitive streak on the posterior side of the embryo, but also finely sub-patterns this dynamic structure to assign specific cell fates to cells

ingressing through the streak in distinct locations and/or at different times (Lawson 1999; Tam et al. 1997; Lawson et al. 1991; Kinder et al. 1999).

Of particular interest for this dissertation, epiblast cells moving through the primitive streak are subjected to opposite signalling gradients of Nodal and Bmp4, with Nodal being higher in the anterior part of the streak, and Bmp4 being strongest posteriorly (Figure 1.4; reviewed in Wu and Hill, 2009). The earliest and most posterior subpopulation delaminates from the streak at the embryonic-extraembryonic junction, and will give rise to the extraembryonic mesoderm and hematopoietic precursors. The lateral plate, paraxial and precardiac mesoderm emerge later from the intermediate and anterior streak. Finally, the midline axial mesoderm tissues and the definitive endoderm are patterned in the extreme anterior tip of the streak. Throughout this process, the anterior side of the epiblast is exposed to very limited Nodal and Bmp signals, and is specified into various both neural and non-neural ectoderm derivatives (reviewed in Ozair et al., 2013). The crucial role of Nodal signalling during gastrulation was established by genetic studies that impaired expression of Nodal (Brennan et al. 2001; Lowe et al. 2001; Conlon et al. 1994), of its co-receptor Cripto (Ding et al. 1998), of its receptor Alk4 (Gu et al. 1998), and of its intracellular effectors Smad2, Smad3, and Smad4 (Heyer et al. 1999; Tremblay et al. 2000; Chu et al. 2004; see Chapter 1.7.1 for a detailed description of the Nodal signalling pathway). In all cases, albeit with some differences in term of penetrance and magnitude of effects, the specification of various mesendoderm lineages was impaired. Moreover, inactivation of Nodal antagonists Lefty1, Lefty2, and Cer1 increases mesendoderm formation (Perea-Gomez et al. 2002; Meno et al. 1999).

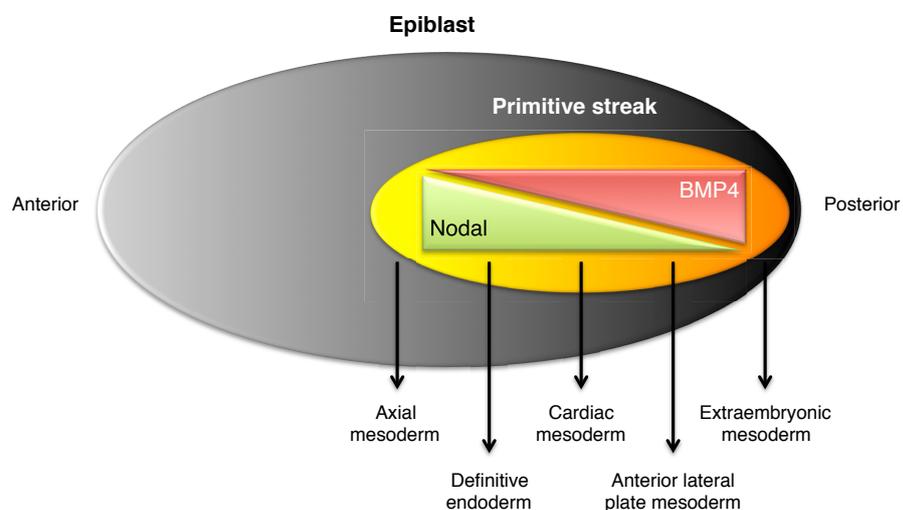


Figure 1.4. Crosstalk between Nodal and BMP4 during mouse gastrulation. For simplicity of representation, the cup-shaped mouse epiblast is shown as if it was flattened.

1.2 Human pluripotent stem cells

As described in the Chapter 1.1.3, while model organisms are powerful tools to investigate early developmental mechanisms, the inability to efficiently test whether these regulations are conserved in humans is a crucial limitation. This sub-chapter introduces hPSCs and describes how this model provides an invaluable complementary approach to understand the mechanism of human development *in vitro*, as well as being a potential source of differentiated cells for various applications.

1.2.1 Human embryonic stem cells.

A stem cell is defined by two key properties linked to its proliferative capacity: it is able to indefinitely produce new copies of itself, a process defined as self-renewal, but it is also capable of generating specialized cell types, namely differentiate. The range of cells that can be obtained from a stem cell defines its potency, which can range between two extremes. On the one hand, the zygote and blastomeres have the ability to generate every cell in the embryo and all the extra-embryonic tissues, and are therefore totipotent. On the other hand, some few stem cells (for example spermatogonial stem cells) have the more limited capacity to specialize into a single cell type, and are called unipotent. Pluripotent stem cells (PSCs) are located between these two extremes, as they can differentiate into every somatic cell type as well as into some (but not all) extra-embryonic tissues.

The initial steps in the history of PSCs date back to the end of the nineteenth century, when researchers initially attempted to maintain animal embryos outside of the womb. These pioneering experiments proved that certain embryonic cells could differentiate into all kinds of different tissues even when explanted from their natural environment. However, attempts to culture such cells for extended periods systematically failed. It was not until the 1970s that pluripotent cell lines were first obtained by culturing *in vitro* mice teratocarcinomas, malignant germ cell tumours (Kahan & Ephrussi 1970; Rosenthal et al. 1970). Nevertheless, these cells presented several chromosomal abnormalities, and were therefore poorly suited to generate useful cell types, and to understand normal developmental mechanisms. A major breakthrough was the work of Martin Evans, Matthew Kaufman and Gail Martin, who in the 1980s managed to culture

in vitro cells from the ICM of a mouse blastocyst, thus obtaining the first karyotypically normal pluripotent stem cell line (Evans & Kaufman 1981; Martin 1981). According to their origin, such cells were named mouse embryonic stem cells, or mESCs. These findings fuelled many years of research towards obtaining similar cells from human embryos, which culminated in the generation of the first human pluripotent stem cell line by James Thomson in 1998 (Thomson 1998). For this, cells were cultured from the ICM of pre-implantation blastocysts obtained from surplus embryos from *in vitro* fertilization procedures, and as such were named human embryonic stem cells, or hESCs (Figure 1.5).

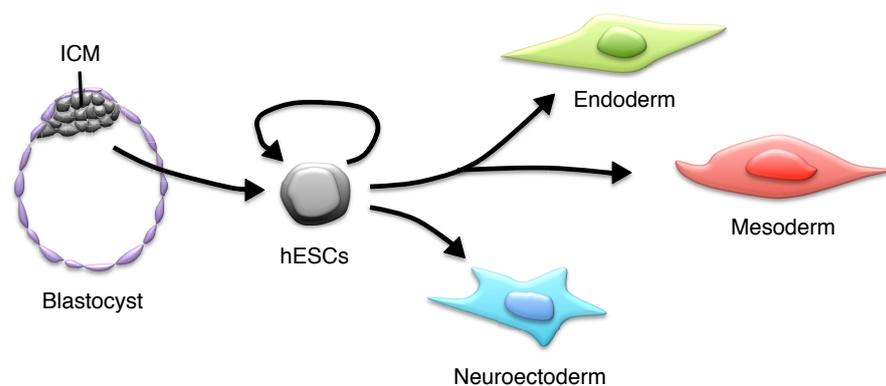


Figure 1.5. Origin and developmental potential of hESCs.

1.2.2 Human induced pluripotent stem cells

The derivation of hESCs was met with equal scientific excitement and ethical concerns about the use of human embryos for research purposes (McLaren 2001). In the United States, the resulting debate led to an almost complete ban of federal funding for embryonic stem cell research, which was initiated in 2001 during the George W. Bush administration (Vogel 2001), and was partially lifted only recently during the Obama presidency (Holden 2009). Of note, derivation of hESCs is still illegal in some countries, for example Italy (Palazzani 2011). Overall, almost 20 years after their discovery, the ethical debate on hESCs is still far from settled.

Moral concerns aside, hESCs present several practical limitations. First, the scarce availability of embryos donated for research purposes complicates large-scale generation of hESC lines, which is required to study the effect of genetic variability during human development. Similarly, generation of hESCs from individuals carrying rare genetic

disorders is extremely challenging. Finally, the use of hESC-derived cells for therapeutic applications of regenerative medicine is limited by the difficulty of obtaining lines with the correct immunological match for each individual patient. Moreover, even when this is possible some level of immunosuppression is still required (Bradley et al. 2002).

Remarkably, both the ethical and practical limitations of hESCs were swept away when Shinya Yamanaka and colleagues showed that both mouse and human adult cells could be reprogrammed to an embryonic-like state by introducing specific transcription factors (Figure 1.6; Takahashi and Yamanaka, 2006; Takahashi et al., 2007). These studies extended the pioneering nuclear reprogramming experiments of Sir John Gurdon, who had shown that the nucleus of a differentiated *Xenopus laevis* cell was developmentally plastic as it could be successfully transplanted into an enucleated egg to generate a viable animal (Gurdon 1962; Gurdon et al. 1958). In an attempt to reprogram human somatic cells to a pluripotent-like state, Yamanaka and colleagues showed that viral-mediated overexpression of four transgenes (OCT4, SOX2, KLF4 and c-MYC) in dermal fibroblasts cultured under appropriate conditions allowed efficient generation of cells that shared most properties of hESCs. These cells were therefore named human induced pluripotent stem cells, or hiPSCs.

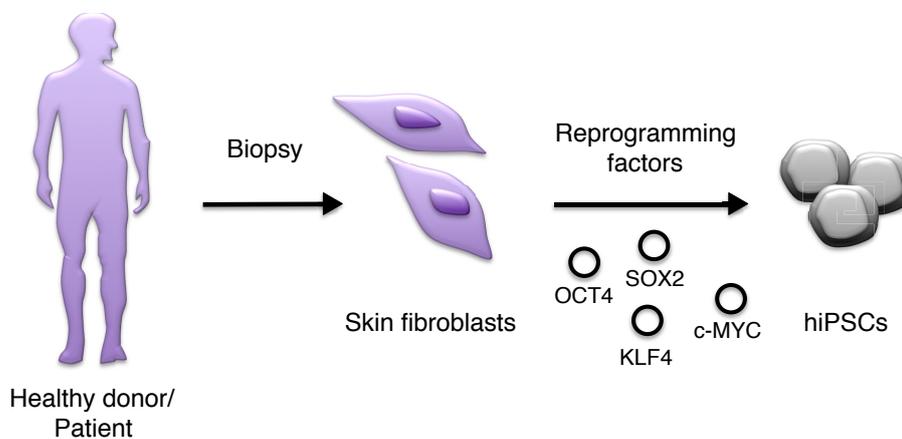


Figure 1.6. Human induced pluripotent stem cells.

Over the last ten years, methods for hiPSC generation have been extensively refined to avoid transgene integration, allow usage of more accessible adult cell types as starting material, and increase reprogramming efficiency and robustness (reviewed in Hochedlinger and Jaenisch, 2015). Most importantly, hiPSCs stood the test of rigorous experimentation with regards to their ability to virtually replace hESCs for most applications (reviewed in Takahashi and Yamanaka, 2016). Therefore, in this

dissertation hESCs and hiPSCs will be collectively referred to as hPSCs when describing their general properties.

1.2.3 Naïve and primed hPSCs

Since the pioneering experiment that isolated mESCs and hESCs from the ICM of the blastocyst, PSCs have been obtained also from several other sources and developmental stages (reviewed in De Los Angeles et al., 2015). Most notably, pluripotent cells can be obtained from the mouse post-implantation epiblast (Tesar et al. 2007; Brons et al. 2007), and from the mouse germ line (Kanatsu-Shinohara et al., 2004). Analysis of cells derived from these different sources revealed that there are multiple and dynamic states of pluripotency that can be captured and controlled *in vitro* (reviewed in Weinberger et al., 2016). A commonly held model is that there are at least two main pluripotent states: naïve pluripotency, in which cells resemble the ICM of the blastocyst, and primed pluripotency, in which cells approximate the post-implantation epiblast and are therefore more poised for differentiation (Figure 1.7; Nichols and Smith, 2009).

In the mouse, these two states are respectively captured by mESCs and mouse epiblast stem cells (mEpiSCs), in line with their developmental origin. mESCs are the only ones capable of colonizing the ICM of the blastocyst and producing chimeric offspring (Nagy et al. 1993; Nagy et al. 1990), while the opposite is true for mEpiSCs and the post-implantation epiblast (Huang et al. 2012; Brons et al. 2007; Tesar et al. 2007). Accordingly, *in vitro*-cultured naïve cells must first transition to the primed state in order to further differentiate into the three germ layers. Naïve and primed cells have also profound differences in their signalling requirements (described in detail in Chapter 1.3). Finally, there are several molecular hallmarks of naïve pluripotency, including the presence of two active X chromosome in female lines, low levels of DNA methylation, preferential usage of the Oct4 distal enhancer, and expression of naïve transcription factors (such as Klf4, Esrrb, Tfcp2l1, and Tbx3; Weinberger et al., 2016).

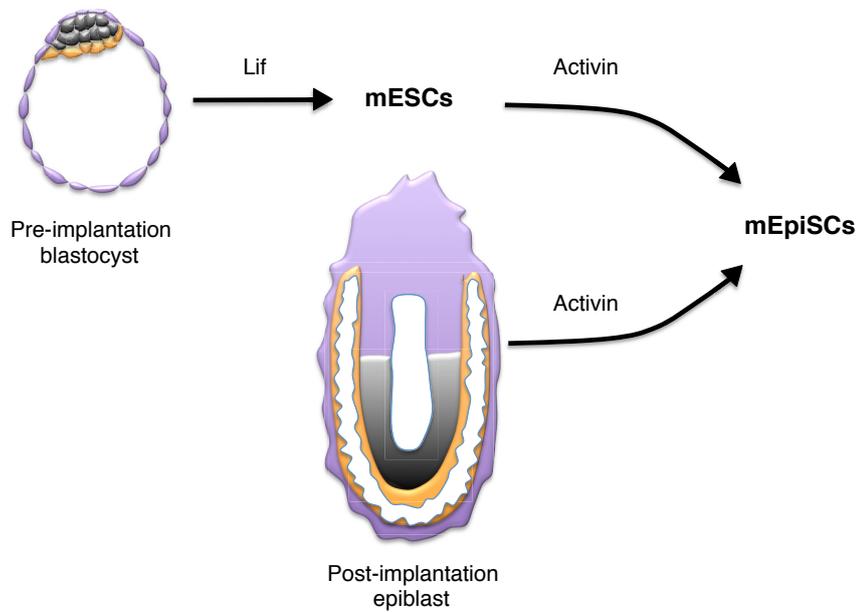


Figure 1.7. Different pluripotent states in murine cells.

In human, conventional hESCs and hiPSCs are in a primed pluripotent state, and are indeed more similar to mEpiSCs than to mESCs (Pauklin et al. 2011). Indeed, the generation of naïve hPSCs proved challenging, since the same culture conditions that enabled mESC isolation proved incompatible with the human system. However, there are now multiple methods to generate hPSCs with certain functional and molecular signatures of naïve pluripotency. Initially, continuous overexpression of pluripotency transgenes proved required to maintain this state (Hanna et al. 2010; Takashima et al. 2014). More recently, various culture conditions that are transgene-independent have emerged (Gafni et al. 2013; Theunissen et al. 2014; Guo et al. 2016; Chan et al. 2013). Of note, these various naïve hPSCs appear to differ from mESCs in some of their signalling requirements (Chapter 6.1.3; Weinberger et al., 2016), thus suggesting that there might be species-specific differences. Alternatively, naïve hPSCs could represent yet another pluripotent state.

It is currently unclear whether naïve hPSCs will provide any substantial practical advantage over conventional hPSCs. Indeed, differentiation of naïve hPSCs requires an initial transition to the primed state, which increases experimental costs, timing, and introduces another potential variable in hPSC differentiation. Moreover, while it has been hypothesized that lack of certain epigenetic marks in naïve hPSCs might result in a broader and less biased differentiation potential (Takashima et al. 2014), this assumption has never been put to the test. On the other side, most current evidence suggests that genetic variability in conventional hPSCs plays a more important role than

epigenetic differences (Kyttälä et al. 2016). Moreover, there are concerns that naïve culture conditions might negatively affect the epigenetic state of hPSCs, for instance by inducing loss of gene imprinting (Pastor et al. 2016). Overall, while naïve hPSCs represent a valuable tool to explore and understand the mechanisms behind different pluripotent states, conventional hPSCs will likely hold a prominent place with regards of downstream applications. As such, the focus of this dissertation is on conventional hPSCs, which will be referred to simply as hPSCs unless otherwise specifically stated.

1.2.4 Applications of hPSCs

hPSCs offer unprecedented opportunities to study human development, investigate human diseases, and develop novel cell therapies to cure pathological conditions (reviewed in Avior et al., 2016; Pourquié et al., 2015; Trounson and DeWitt, 2016; Zhu and Huangfu, 2013). Indeed, hPSCs can be used to test whether developmental notions learned from model organisms are conserved in the human context. Moreover, hiPSCs generated from patients carrying genetic disorders can be used to study the molecular mechanisms behind the disease, and to test potential therapies. Finally, hPSCs can serve as an inexhaustible source to generate virtually any human cell type to regenerate damaged tissues or even whole organs. In this context, transplantation of autologous cells derived from patient-specific hiPSCs prospects the opportunity of preventing immune rejection while avoiding immunosuppressant therapy.

However, all these applications crucially depend on our ability to robustly and reproducibly control both the self-renewal and differentiation of hPSCs. Indeed, genetic and epigenetic variability between different lines can result into poor quality of hPSCs and/or lead to inefficient lineage specification (Cahan & Daley 2013). In this context, the key to develop more robust protocols for hPSC maintenance and differentiation lies in a deeper understanding of the molecular mechanisms that regulate hPSC early cell fate decisions.

1.3 Control of pluripotency and differentiation by extracellular signalling

The early days of hPSC research were characterized by the labour-intensive need of maintaining undifferentiated cells in co-culture on a feeder layer of mitotically inactivated mouse embryonic fibroblasts (MEFs), which empirically demonstrated to provide the necessary pluripotency-maintaining factors (Martin & Evans 1975). hPSCs could then be detached from the feeder layer and induced to differentiate as clumps onto low-attachment plates in the presence of serum. Under these conditions, hPSCs aggregated and formed structures called embryonic bodies, as they roughly resembled early human developing embryos by showing differentiating cells from all of the germ layers (Itskovitz-Eldor et al. 2000).

While feeder culture and embryonic bodies differentiation still represent valuable resources for specific applications, over the last years the field evolved dramatically by identifying robust feeder-free and chemically defined conditions for hPSC culture (Vallier, Reynolds, et al. 2004; Chen et al. 2011; Wiles & Johansson 1999). These advances allowed to define the minimal culture conditions required to maintain hPSC self-renewal, or to drive their differentiation. In turn, this knowledge paved the way to detailed analyses of the mechanisms regulating pluripotency. This sub-chapter describes the signalling pathways involved in hPSC early cell fate decisions.

1.3.1 Signalling pathways regulating pluripotency

As mentioned in Chapter 1.2.3, hPSCs have alternative signalling requirements from mESCs, which reflect their different pluripotent states. The most commonly used conditions for mESC culture primarily rely on Lif stimulation to promote signalling through the Jak-Stat3 pathway (Niwa et al. 1998), which mirrors the requirement the pre-implantation mouse ICM for Stat3 activation (Chapter 1.1.3; Do et al., 2013; Nichols et al., 2001; Stewart et al., 1992). Initially, Lif stimulation was combined with foetal bovine serum (FBS/Lif; Smith et al., 1988; Williams et al., 1988), which provides low doses of Bmp4 and can be substituted accordingly (Ying et al. 2003). However, these conditions result into substantial heterogeneity in mESCs, with only a sub-population of cells presenting a truly naïve pluripotent state (Niakan et al. 2010; Singh et al. 2007; Toyooka et al. 2008). Subsequently, it was demonstrated that the

combination Lif with two small molecule inhibitors for Mek1/2 and Gsk3 β promotes more homogeneous mESC culture in a naïve state (2i/Lif conditions; Hanna et al., 2009; Ying et al., 2008). Interestingly, Mek1 and Mek2 are protein kinases involved in the Erk1/2 mitogen-activated protein kinase (MAPK) pathway, which is activated by various growth factors that include Fgfs. Therefore, the improved mESC culture conditions following Mek1/2 inhibition fit with the antagonistic role of Fgf4 and Fgf2 for the pluripotency of the early pre-implantation mouse ICM (Chazaud et al. 2006; Frankenberg et al. 2011; Morrissey et al. 1998). On the other hand, while Wnt signalling appears dispensable for pluripotency of the ICM, it seems to promote its proliferation (Biechele et al. 2013), which might explain its beneficial roles in 2i/Lif mESC cultures.

Despite several efforts, none of the culture conditions described for mESCs could be successfully applied for feeder-independent culture of hESCs. Indeed, hESC growth relies on a signalling pathway other than LIF as it is Stat3-independent (Humphrey et al. 2004; Dahéron et al. 2004). BMP4 is also not required for hESC pluripotency (R. H. Xu et al. 2002), and on the contrary its inhibition promotes self-renewal (Levine 2005; G. Wang et al. 2005). Efforts to identifying the feeder layer-secreted factor(s) that maintain hPSC pluripotency led to the demonstration that signalling molecules of the transforming growth factor β (TGF β) superfamily were responsible (Vallier, Reynolds, et al. 2004). Indeed, treatment with Activin, Nodal, or TGF β promotes hESC pluripotency, while it inhibits the default neuroectoderm differentiation pathway (Beattie et al., 2005; Vallier et al., 2005, 2004; James et al., 2005; Xiao et al., 2006). Of note, these signalling molecules are functionally very similar in hESCs, as they mainly activate the same downstream signalling pathway (described in detail in Chapter 1.7). Moreover, Activin and Nodal share the same receptors. As such, the notation Activin/Nodal will be used as a convention throughout this dissertation when describing the signalling pathway in general, while Activin, Nodal and TGF β will be individually named to indicate each specific growth factor.

Aside from Activin/Nodal signalling, FGF2 also facilitates maintenance of hESCs in an undifferentiated state (Levenstein et al. 2006; Xu et al. 2005; Vallier et al. 2005; G. Wang et al. 2005). However, while inhibition of FGF receptors can be rescued by increasing the dose of exogenous Activin, absence of Activin cannot be bypassed by high dosage of FGF2 (Vallier et al. 2005). This implies that Activin/Nodal signalling is the dominant pathway for hESC pluripotency, while FGF2 synergizes possibly by regulating

a similar transcriptional network (Göke et al. 2013; Greber et al. 2010). As a result, hESCs can be maintained long-term in chemically defined conditions including Activin and FGF2 in the absence of feeder layer, conditioned media, and serum replacers (Vallier 2011).

Remarkably, the signalling requirements of hESCs closely recapitulate the ones of the post-implantation mouse embryo, in which Nodal is necessary to maintain the expression of pluripotency markers in the epiblast (Chapter 1.1.3; Camus et al., 2006; Mesnard et al., 2006), while *Lif* is dispensable (Posfai et al. 2014). Accordingly, culture conditions with Activin and FGF2 were later shown to support derivation also of mEpiSCs, which could not be otherwise isolated in standard mESC culture conditions (Brons et al. 2007; Tesar et al. 2007; Vallier, Touboul, Chng, et al. 2009; note that derivation of hPSCs from the post-implantation human epiblast has never been attempted due to ethical reasons). Overall, this evidence further reinforced the notion that hESCs are in a developmental stage that is distinct from the one of mESCs and is more similar to that of mEpiSCs (Chapter 1.2.3). Importantly, it was later shown that pluripotency of hiPSCs also relies on Activin/Nodal and FGF2 signalling, further establishing that these cells are functionally equivalent to hESCs (Vallier, Touboul, Brown, et al. 2009). As a result of these studies, hPSC growth conditions based on the stimulation of Activin/Nodal and FGF2 signalling have been adopted by most laboratories worldwide (reviewed in Chen et al., 2014).

Whilst Activin/Nodal and FGF are the key factors that sustain hPSC pluripotency, several other pathways can promote this process (Figure 1.8). Notably, insulin or insulin-like growth factors (IGFs) signalling stimulates pluripotency through the phosphatidylinositide 3-kinase (PI3K)/protein kinase B (PKB/AKT) pathway (reviewed in Voskas et al., 2014). In this, insulin and IGFs are partially redundant with FGF signalling, which activates both the ERK1/2 and PI3K/AKT pathways (Eiselleova et al. 2009). Activation of PI3K/AKT both increases hPSC proliferation (Campbell et al. 2012), and facilitates maintenance of the pluripotent state by inhibiting mesendoderm differentiation (McLean et al. 2007; Singh et al. 2012). Another notable pathway involved in hPSC pluripotency is WNT/ β -catenin (Atlasi et al. 2014). *In vivo*, Wnt marks the developing primitive streak in the post-implantation mouse epiblast (Mohamed et al. 2004), and is required for gastrulation (Liu et al. 1999; Kelly et al. 2004). Accordingly, WNT signalling must be tightly balanced in hPSCs, with low levels

promoting the primed pluripotent state, and a strong stimulation driving mesendoderm specification (Sumi et al. 2013; Singh et al. 2012). Finally, the Hippo pathway also modulates the balance between self-renewal and differentiation by preventing mesendoderm gene expression (Beyer et al. 2013; Hsiao et al. 2016; Varelas et al. 2010).

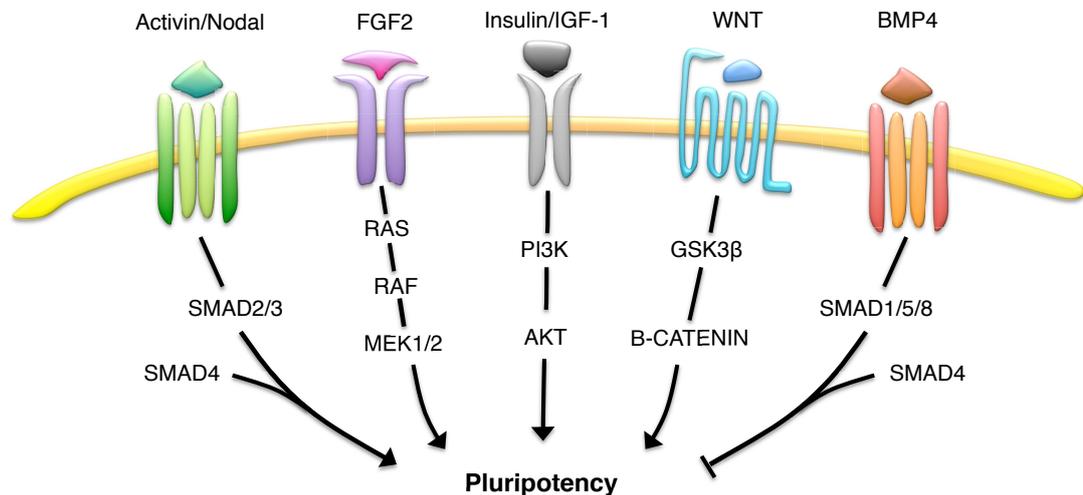


Figure 1.8. Main signalling pathways regulating hPSC pluripotency.

1.3.2 Signalling pathways regulating differentiation

In contrast to the marked differences between the signalling requirements for mESC and hPSC pluripotency, the pathways that drive their differentiation into the primary germ layers are remarkably similar. Strikingly, only a handful of growth factors appears to be required to control this complex and crucial step (reviewed in Cohen and Melton, 2011; Murry and Keller, 2008).

1.3.2.1 Signalling during mesendoderm differentiation

Generation of definitive endoderm and mesoderm cells requires the initial establishment of an intermediate stage that recapitulates the gastrulating primitive streak, and which is commonly described as mesendoderm (Kimelman & Griffin 2000; Rodaway & Patient 2001). Mesendoderm induction relies primarily on the combination of Activin (Kubo et al. 2004; D'Amour et al. 2005), BMP4 (Wiles & Johansson 1999; C. Park et al. 2004; Ng et al. 2005; Nostro et al. 2008), WNT (Lindsley et al. 2006; Naito et al. 2006; Ueno et

al. 2007), and FGF2 (Willems & Leyns 2008; Touboul et al. 2010). By modulating the dosage of these various signalling molecules it is possible to generate different types of mesendoderm cells that have specific potential for further differentiation into individual mesoderm and endoderm subtypes. Remarkably, the *in vitro* patterning of these mesendoderm populations closely mimics the *in vivo* situation in the mouse, in which gradients of Nodal, Bmp, Wnt, and Fgf across the primitive streak dictate specification of epiblast cells into the various mesoderm and endoderm derivatives (Chapter 1.1.3).

This paradigm is exemplified by a series of studies from the Pedersen, Vallier and Sinha groups at the University of Cambridge, which used the H9 hESC line as main model system (Figure 1.9). In this context, differentiation in the presence of high doses of BMP4 and in the absence of FGF2 results into progenitors that can be further specified into extraembryonic mesoderm, thus recapitulating the most posterior part of the primitive streak (Bernardo et al. 2011). On the other side, the combination of BMP4 and FGF2 in the absence of exogenously provided Activin generates mesendoderm cells that can be efficiently differentiated into derivatives of the mid-posterior streak, such as lateral plate mesoderm (Bernardo et al. 2011; Cheung et al. 2012). Further, adding intermediate doses of Activin to BMP4 and FGF2 dictates yet different progenitors that are similar to the middle portion of the primitive streak, and which have the potential of be driven into a cardiogenic mesoderm fate (Mendjan et al. 2014). Finally, cells exposed to high doses of Activin together with FGF2 and BMP4 specify into a mesendoderm subtype analogue to anterior primitive streak, and which readily differentiates into definitive endoderm (Rashid et al. 2010; Touboul et al. 2010; Vallier, Touboul, Brown, et al. 2009). Of note, all these differentiation protocols rely on inhibition of the PI3K/AKT pathway, as such signalling has an inhibitory role in the early specification of mesendoderm progenitors, in particularly definitive endoderm (McLean et al. 2007; Touboul et al. 2010). Interestingly, inhibition of PI3K/AKT also leads to increased β -catenin stabilization (Singh et al. 2012), thus indirectly potentiating WNT signalling. On the other hand, more robust activation of WNT in combination with FGF2 induces yet another type of mesendoderm that recapitulates progenitors observed in the late stages of primitive streak development, and which can be further differentiated into paraxial mesoderm derivatives such as somitic and presomitic mesoderm (Mendjan et al. 2014). Remarkably, similar methods reported by several other research groups allow mesendoderm specification of multiple hPSC lines (Sumi et al. 2008; Kattman et al. 2011; Patsch et al. 2015; Palpant et al. 2015; D'Amour et al. 2005).

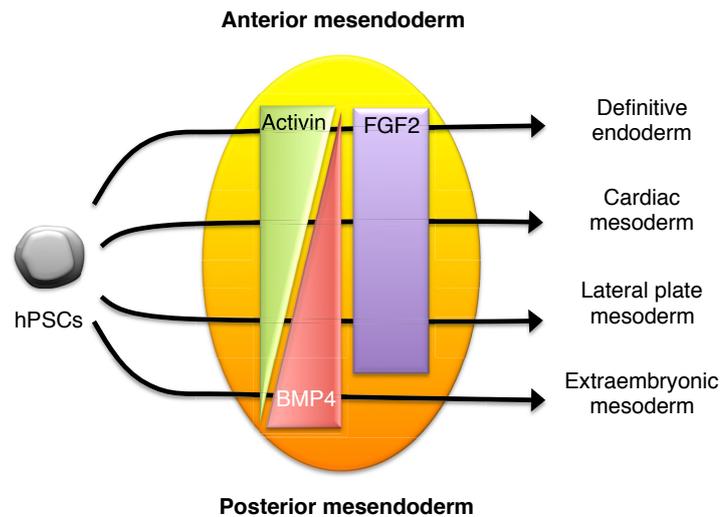


Figure 1.9. Differentiation of hPSCs into mesendoderm derivatives.

1.3.2.2 Signalling during endoderm and mesoderm differentiation

Following anterior mesendoderm induction, definitive endoderm specification is primarily driven by Activin/Nodal signalling, with other accessory factors such as BMP4 and FGF2 being able to reinforce its effects (Touboul et al. 2010; Rashid et al. 2010; Kubo et al. 2004; D'Amour et al. 2005). On the other hand, the differentiation of various mesoderm subtypes from their respective mesendoderm progenitors is largely independent from Activin/Nodal signalling, and relies primarily on BMP4, FGF2 and WNT (Kattman et al. 2011; Patsch et al. 2015; Palpant et al. 2015; Cheung et al. 2012; Mendjan et al. 2014). These findings are in line with elegant experiments in the mouse that demonstrated how progressive genetic ablation of Smad2 and Smad3 (the intracellular transducers of Activin/Nodal signalling, see Chapter 1.7.1) matches a gradual impairment in the specification of mesendoderm lineages arising from the anterior towards the posterior side of the primitive streak (Dunn et al. 2004; Vincent et al. 2003). This demonstrates that these different lineages show different sensitivity to the dose of Activin/Nodal signalling *in vivo* as well as *in vitro*. Of note, the apparently paradoxical feature of Activin/Nodal signalling of being able to both promote pluripotency and drive mesendoderm differentiation has been puzzling developmental biologists, and the mechanism behind this dual function are only partially clear (Chapter 1.7.4 describes this aspect in detail).

1.3.2.3 Signalling during neuroectoderm differentiation

Specification of neuroectoderm from hPSCs is remarkably different from both mesoderm and endoderm differentiation (Figure 1.10). Indeed, neuroectoderm induction is often referred to as the “default” differentiation pathway of pluripotent cells, since it relies on the absence of most factors involved in self-renewal or mesendoderm specification (Ozair et al. 2013). Accordingly, low-density culture of hPSCs in absence of any exogenous signal is sufficient to initiate neuroectoderm specification (S. C. Zhang et al. 2001; Pankratz et al. 2007). This can be further reinforced by using inhibitors of the Activin/Nodal, BMP4 and WNT signalling pathways (Chambers et al. 2009; Itsykson et al. 2005; Yao et al. 2006; Smith, Vallier, et al. 2008; Vallier, Reynolds, et al. 2004; Patani et al. 2009; Eiraku et al. 2008). Among these factors, absence of signalling through Activin/Nodal is the most crucial requirement to drive the exit from pluripotency and allow expression of neuroectoderm regulators (Chng et al. 2010; Vallier, Reynolds, et al. 2004; Smith, Vallier, et al. 2008; Patani et al. 2009). On the other hand, blockage of BMP4 signalling is required to avoid redirection of neuroectoderm into non-neural ectoderm (Kawasaki et al. 2000). Finally, WNT inhibitors prevent both posteriorizing and non-neural signals (Menendez et al. 2011; Fasano et al. 2010). Interestingly, addition of low doses of FGF2 during neuroectoderm differentiation of certain hPSC lines promotes cell growth and improve specification (Pankratz et al. 2007; S. C. Zhang et al. 2001; Smith, Vallier, et al. 2008). Overall, *in vitro* specification of neuroectoderm cells closely recapitulates the *in vivo* scenario in the mouse, in which expression of inhibitors for the Nodal, BMP4 and WNT pathway from the AVE and the node deprives the anterior portion of the epiblast of such signals in order to drive neuralization (Chapter 1.1.3).

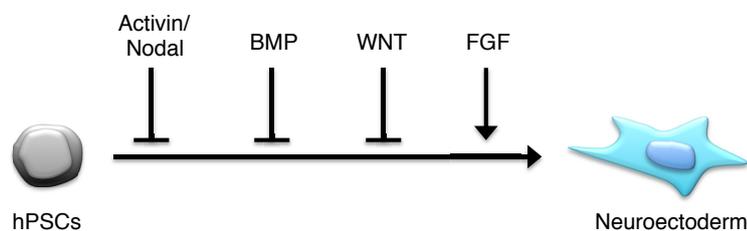


Figure 1.10. Neuroectoderm differentiation of hPSCs.

1.4 Transcriptional control of pluripotency and differentiation

hPSC pluripotency and differentiation are the result of a complex interplay of multiple transcription factors that dictate the pluripotent state or drive cellular specification. This sub-chapter presents the most relevant of such regulators.

1.4.1 Pluripotency factors

Pluripotency is maintained by a complex transcriptional network in which multiple transcription factors, commonly named pluripotency factors, promote each other's expression to collectively stabilize the pluripotent state and stimulate self-renewal (Young 2011; Boyer et al. 2005; Z. Wang et al. 2012). Notably, this network is only partially shared between the naïve and primed pluripotent states (Nichols & Smith 2009; Z. Wang et al. 2012). In naïve mESCs, the pluripotency factors Oct4, Nanog and Sox2 are irreplaceable, and constitute the core components of the network. Additional pluripotency factors are involved (including Klf2, Klf4, Essrb, Sall4, Tfcp2l1, and Tbx3), but these can be individually replaced (Ivanova et al. 2006; Dunn et al. 2014). Most of these “accessory” proteins are strongly downregulated in primed cells such as hPSCs, which primarily rely on the core pluripotency network.

1.4.1.1 OCT4

In the mouse, Oct4 is expressed in the oocyte, morula, ICM, and in both the epiblast and PrE (Palmieri et al. 1994; Rosner et al. 1990). Following implantation, its expression is maintained across the epiblast and the emerging mesendoderm cells. Finally, Oct4 is progressively downregulated from the anterior to the posterior side of the embryo (Schöler et al. 1990). Importantly, Oct4 is required for ICM specification as it counteracts TE differentiation (Chapter 1.1.3; Nichols et al., 1998). Accordingly, mESCs lacking Oct4 cannot be obtained, and conditional loss of Oct4 induces TE (Niwa et al. 2000). Interestingly, Oct4 overexpression in mESCs also results into differentiation, but into mesendoderm (Niwa et al. 2000; Ivanova et al. 2006). Interestingly, OCT4 has a distinct role in hPSCs, as its knockdown can trigger differentiation into PrE, extraembryonic mesoderm, embryonic mesendoderm, or neuroectoderm (Matin et al. 2004; Hay et al. 2004; Z. Wang et al. 2012; Teo et al. 2011; Thomson et al. 2011). In this, the precise outcome depends on the levels of BMP signalling (Z. Wang et al. 2012).

Finally, overexpression of OCT4 in hPSCs has little phenotypic impact (Z. Wang et al. 2012).

1.4.1.2 NANOG

In the early mouse embryo, Nanog is expressed in the compacted morula and in the epiblast (Chambers et al. 2003; Mitsui et al. 2003). Post-implantation, its expression is initially restricted to the posterior epiblast in the region of the presumptive primitive streak, and is then progressively downregulated as epiblast cells ingress through the streak (Hart et al. 2004; Hatano et al. 2005). Nanog is developmentally required, as it dictates the specification of the epiblast by inhibiting PrE differentiation (Chapter 1.1.3; Chazaud et al., 2006). Similarly to Oct4, Nanog knockout mESCs cannot be derived (Chambers et al. 2003; Mitsui et al. 2003). However, conditional loss of Nanog in mESCs is compatible with self-renewal, albeit it results in a marked propensity for differentiation into several cell lineages, especially TE and PrE (Ivanova et al. 2006; Chambers et al. 2007; Hyslop et al. 2005). Remarkably, Nanog overexpression appears sufficient to maintain mESC pluripotency in the absence of Lif (Chambers et al. 2003; Wang et al. 2008). Similarly, NANOG overexpression potentiates hPSC self-renewal (Darr et al. 2006), and makes hPSCs independent from Activin/Nodal signalling (Vallier, Mendjan, et al. 2009). However, knockdown of NANOG in hPSCs has a very different effect than in mESCs, as it induces neuroectoderm while impairing mesendoderm differentiation (Vallier, Mendjan, et al. 2009; Z. Wang et al. 2012; Teo et al. 2011).

1.4.1.3 SOX2

Sox2 is present both in the early morula and in the ICM (Avilion et al. 2003). Oppositely to Nanog, Sox2 is then highly expressed in the anterior part of the post-implantation epiblast (the region of the presumptive neuroectoderm), but is only lowly expressed posteriorly. Unique to the core pluripotency factors, Sox2 is expressed also in multiple other lineages at later developmental stages, notably the brain, neural tube, and gut endoderm (Avilion et al. 2003; Li et al. 1998; Zappone et al. 2000). Knockout of Sox2 in is embryonically lethal post-implantation due to defects in the epiblast (Avilion et al. 2003). Similarly to Oct4, loss of function of Sox2 in mESCs results into TE differentiation (Ivanova et al., 2006; Masui et al., 2007). On the other hand, Sox2 overexpression induces mESC differentiation into a wide range of cell types, in

particular neuroectoderm (Kopp et al. 2008). Interestingly, SOX2 knockdown in hPSCs results in upregulation of mesendoderm markers, an effect which is potentiated by simultaneous knockdown also of the closely related gene SOX3 (Z. Wang et al. 2012; Teo et al. 2011). This suggests that SOX3 can functionally substitute SOX2, and that SOX2/3 are predominantly inhibitors of mesendoderm specification (Chapter 1.4.2.4).

Collectively, the core pluripotency transcriptional network appears to work differently in naïve and primed cells (Figure 1.11). In mESCs, Nanog, Oct4 and Sox2 promote expression of pluripotency factors while predominantly working as pan-repressors of extraembryonic differentiation (Boyer et al. 2005; Young 2011). However, in hPSCs each factor controls specific cell fates, working primarily to repress embryonic lineage specification (Thomson et al. 2011; Z. Wang et al. 2012). Moreover, while Nanog, Oct4 and Sox2 reinforce each other's expression in mESCs, SOX2 appears to be excluded from this network in hPSCs (Z. Wang et al. 2012). Finally, pluripotency factors also have active roles during the induction of hPSC differentiation (Chapters 1.4.2.1 and 1.4.2.4).

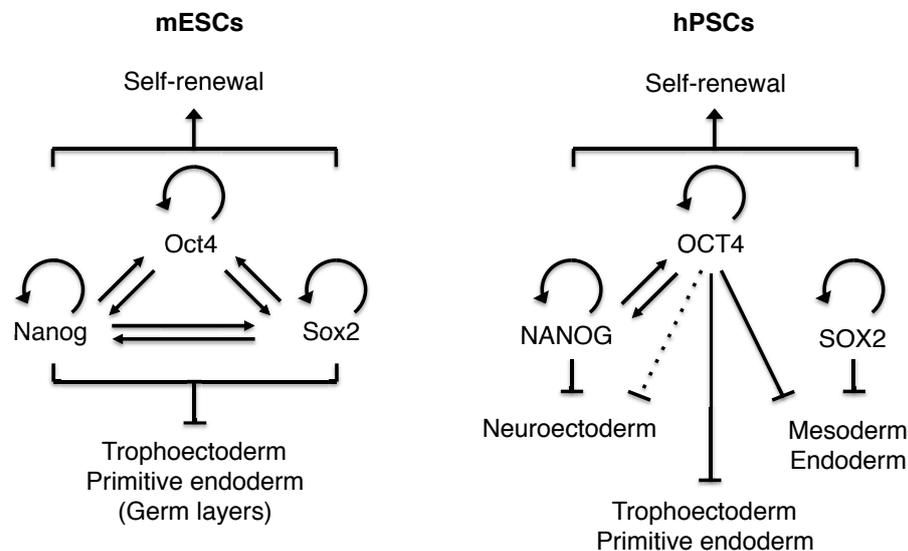


Figure 1.11. Core pluripotency transcriptional networks.

1.4.2 Germ layer specification factors

Germ layer differentiation relies on transcriptional networks that are hierarchical, dynamic, and remarkably intertwined with the core pluripotency factors.

1.4.2.1 Mesendoderm transcriptional regulators

Specification of hPSCs into mesendoderm involves the interplay of transcriptional regulators both of the mesoderm and endoderm germ layers, as well as of pluripotency factors. Indeed, the mechanism of exit from pluripotency initiates different transcriptional networks (Mendjan et al., 2014; Teo et al., 2011).

On the one hand, combination of Activin/Nodal and BMP signalling results in high expression of NANOG during the initial stages of hPSC differentiation, which is necessary to induce EOMES and initiate specification of mesendoderm progenitors recapitulating the mid-anterior portion of the primitive streak (Teo et al. 2011; Mendjan et al. 2014). Eomes is one of the earliest known mesendoderm markers, being expressed in the proximal posterior part of the pre-gastrulating mouse epiblast, in the gastrulating primitive streak, and in the anterior primitive streak progenitors (Ciruna & Rossant 1999; Hancock et al. 1999). Eomes is required for definitive endoderm specification and for the migration of prospective mesoderm cells in the primitive streak (Arnold et al. 2008; Russ et al. 2000). Importantly, Eomes activates other mesendoderm regulators like Mixl1, which cooperate with Eomes for definitive endoderm specification (Chapter 1.4.2.2; Hart et al., 2002; Izumi et al., 2007; Tam et al., 2007).

On the other hand, mesendoderm-inducing conditions that rapidly inhibit NANOG expression during the exit from pluripotency result in induction of different progenitors (Chapter 1.3.2.1). These recapitulate the posterior or late primitive streak, and express high levels of BRA (also known as T) and CDX2 (Mendjan et al. 2014; Bernardo et al. 2011). Bra is another very early mesendoderm marker expressed in the pre-gastrulating mouse epiblast and in the primitive streak (Herrmann 1991; Kispert et al. 1994; Thomas & Beddington 1996). Accordingly, Bra is required for mesoderm specification (Beddington et al. 1992; Wilson et al. 1995). Cdx2 is expressed in the posterior primitive streak, and is also involved in posterior mesoderm development (Savory et al. 2009; van den Akker et al. 2002; Lohnes 2003). While BRA is initially expressed also in

EOMES-positive mesendoderm progenitors, its expression is lower and rapidly downregulated. CDX2 on the other hand is mutually exclusive with EOMES, as it is directly repressed by NANOG. BRA and CDX2 positively regulate each other's expression, and orchestrate further development of posterior mesoderm subtypes (Bernardo et al. 2011; Mendjan et al. 2014; Palpant et al. 2015; Cheung et al. 2012).

1.4.2.2 Endoderm transcriptional regulators

Specification of hPSC-derived mesendoderm progenitors into definitive endoderm is orchestrated by the cooperation of multiple transcription factors, most notably EOMES, MIXL1, GSC, SOX17, FOXA2, and FOXH1 (D'Amour et al. 2005; Touboul et al. 2010; Kubo et al. 2004; Teo et al. 2011). In the mouse, Gsc is expressed both in the anterior primitive streak and in the presumptive definitive endoderm (Blum et al. 1992; Belo et al. 1998). However, loss of Gsc is not sufficient to impair endoderm specification (Yamada et al. 1995; Rivera-Pérez et al. 1995), probably due to compensatory mechanisms by Mixl1 (De Robertis 2004). Sox17 is expressed both in the visceral and definitive endoderm, and is required for definitive endoderm development (Kanai-Azuma et al. 2002; Séguin et al. 2008). Remarkably, SOX17 overexpression is sufficient to convert hESCs into definitive endoderm (Séguin et al. 2008). FOXA2 and FOXH1 are both members of the winged helix/forkhead family of transcription factors (Kaestner et al. 2000). Foxa2 is expressed predominately in the anterior mesendoderm, anterior definitive endoderm, and endoderm derivatives such as the liver, and is essential for foregut and midgut development (Sasaki & Hogan 1993; Monaghan et al. 1993; Ang et al. 1993). In contrast, Foxh1 is expressed throughout the epiblast both prior and during gastrulation (Weisberg et al. 1998), and its knockout induces a spectrum of phenotypes ranging from severe gastrulation defects to milder deficiencies in the specification of anterior mesendoderm derivatives (Hoodless et al. 2001; Yamamoto et al. 2001).

1.4.2.3 Mesoderm transcriptional regulators

Differentiation of hPSC-derived mesendoderm into the different mesodermal subtypes is regulated by partially overlapping yet distinct transcriptional networks that recapitulate *in vivo* events (Mendjan et al. 2014; Cheung et al. 2012; Kattman et al. 2011; Palpant et al. 2015; Patsch et al. 2015). In the mouse, the earliest steps of lateral plate mesoderm and cardiogenic mesoderm specification involve the activation *Mesp1* and *Mesp2* downstream of *Eomes* and *Bra* (Costello et al. 2011; van den Aamele et al. 2012; Saga et al. 1999). Subsequent expression of *Nkx2.5* defines early cardiac progenitors from both heart fields (Lien et al. 2002; Brown et al. 2004; Wu et al. 2006), while *Isl1* is selectively present in lateral plate mesoderm and the secondary heart field (Moretti et al. 2006). Further cardiac maturation is finally initiated by cardiogenic factors such as *Mef2c* and *Tbx5* (Bruneau et al. 2001; Lin et al. 1997). On the other hand, paraxial mesoderm subtypes are specified downstream of *Bra*, *Cdx2* and *Tbx6* (Chapman et al. 1996). Then, ensuing presomitic and somitic mesoderm fates rely on factors such as *Msgn1*, *Pax3*, and *Sox9* (Tam & Trainor 1994; Aulehla & Pourquié 2010; Martin 2015).

1.4.2.4 Neuroectoderm transcriptional regulators

Induction of neuroectoderm in hPSCs depends on the rapid loss of the pluripotency factor *NANOG*, which results in derepression of neuroectoderm genes and is followed by downregulation also of *OCT4* (Vallier, Mendjan, et al. 2009; Chng et al. 2010; Z. Wang et al. 2012). However, *SOX2* expression is not only maintained but also even mildly increased during neuroectoderm differentiation. Indeed, *Sox2* is considered the first general marker of the initiating neural tissue (as described in Chapter 1.4.1.3, *Sox2* is expressed in the prospective neuroectoderm and in its derivatives). *SOX2* promotes neuroectoderm specification mainly by preventing alternative mesendoderm cell fates (Thomson et al. 2011; Z. Wang et al. 2012; Teo et al. 2011). The activity of *Sox2* is reinforced by other neural-initiating factors such as *Zfp521* (Kamiya et al. 2011) and *Sip1* (also known as *Zeb2*, Chng et al., 2010). Subsequent neuroectoderm maturation relies on the sequential expression of early (*Otx2*, *Gbx2*, *Hoxa3*, and *Hoxa1*), and late (*Sox1* and *Pax6*) factors (Lee et al. 2014; Tang et al. 2015; Li & Joyner 2001). In particular, *Sox1* is among the earliest known specific neuronal marker, as its expression begins with the formation of the neural plate (Pevny et al. 1998; Wood & Episkopou 1999).

1.5 Chromatin epigenetic control of pluripotency and differentiation

Extracellular signalling and intracellular transcriptional regulators cooperate to initiate the specific transcriptional programs that drive cell fate decisions. For this, a major mechanism is the imposition of structural changes in the chromatin structure to modulate the transcriptional accessibility of the genome. In turn, these modifications impact on the activity of signalling pathways and transcription factors. Collectively, these so called epigenetic mechanisms drive the acquisition of cellular identity, stabilize such state under physiological conditions, and facilitate its transmission during cellular proliferation (reviewed in Chen and Dent, 2013; Margueron and Reinberg, 2010).

The following sections will present some notable aspects of chromatin epigenetic regulations in PSCs. Due to the extensiveness of this topic, a specific focus will be dedicated to the role of histone epigenetic modifications, since these are particularly relevant to the scope of this dissertation.

1.5.1 Chromatin organization in pluripotent stem cells

In the nucleus, the DNA is wrapped around histone octamers to form nucleosomes, the basic repeating units of the chromatin. A nucleosome is constituted by roughly 2 turns of DNA and core histone proteins formed from a pair of H2A-H2B dimers, and a H3-H4 tetramer (Luger et al. 1997). The linker histone H1 then binds the nucleosome and the linker DNA entering and exiting the core particle, in order to form higher order chromatin structures (Harshman et al. 2013). The association between DNA and histones provides a remarkably flexible template that can be extensively edited to impose epigenetic controls of gene expression. This is achieved in three main ways: first, the DNA can be chemically modified, most notably by methylation (Smith and Meissner, 2013); secondly, the localization and/or composition of histones can be actively modulated (Narlikar et al., 2013); third, histone tails can be targeted by a plethora of covalent changes, including methylation, acetylation, phosphorylation, and ubiquitination (Kouzarides 2007; Tessarz & Kouzarides 2014). Overall, the cooperation between these various regulations shapes the chromatin structure rendering it more or less accessible to transcription factors and the transcriptional machinery.

Compared to other cell types, the chromatin of both mESCs and hPSCs is distinctive because of its open structure: the chromatin is globally decondensed, and only a small fraction of the genome is organized as repressive heterochromatin (Efroni et al., 2008; Park et al., 2004). Remarkably, a similar pattern of chromatin organization appears to be present *in vivo* in the mouse ICM (Ahmed et al. 2010). This peculiar structure has been proposed to allow rapid switching of transcriptional programs during differentiation. Accordingly, the chromatin structure dramatically changes upon PSC differentiation into somatic cells, as large heterochromatin areas are rapidly established (Efroni et al., 2008; Park et al., 2004). However, the open chromatin structure of PSCs results in transcriptional hyperactivity (Efroni et al. 2008), which must be countered to prevent untimely expression of differentiation genes (Guenther et al. 2007). All these features are primarily the result of the concerted activities of several DNA methyltransferases (Meissner et al. 2008; Lister et al. 2009), chromatin remodellers (Schaniel et al. 2009; Meshorer et al. 2006), and histone modifiers (Wen et al. 2009; Bernstein et al. 2006).

Of crucial importance, most of what we know about epigenetic regulation in PSCs comes from studies in mESCs. This is due both to historical reasons (mESCs were available for much longer than hPSCs), and because of practical considerations (until recently, growth and genetic manipulation of mESCs were remarkably simpler, thus facilitating functional studies; Chapter 1.8). However, as explained in previous chapters, mESCs represent a different developmental stage than hPSCs. Overall, it is likely that the epigenetic regulations in mESCs and hPSCs are not identical (Song et al. 2012). However, it is somewhat surprising that the extent of such differences has long been underestimated and poorly explored by the field. Nevertheless, recent data demonstrating that mESCs and hPSCs react differently to the depletion of epigenetic modifiers has brought this issue in the spotlight (reviewed in Weinberger et al., 2016). Overall, it is important to carefully examine the role of individual epigenetic regulations across different pluripotent states.

1.5.2 Histone modifications

Among the many possible covalent modifications of histone tails, trimethylation of histone 3 lysine 4 (H3K4me3) or lysine 27 (H3K27me3) have been proposed to have a prominent role in PSCs (Harikumar & Meshorer 2015; Vastenhouw & Schier 2012; Voigt et al. 2013). Given their relevance to the scope of this dissertation, these marks are the main focus of the following sections, while other prominent histone modifications are described in Chapter 1.5.2.4).

1.5.2.1 H3K4me3

In mammalian cells, H3K4me3 is deposited by SETD1A, SETD1B and mixed lineage leukemia 1 to 4 (MLL1-4) methyltransferases (also known as Tritorax proteins; Shilatifard, 2012). Such enzymes form supramolecular functional complexes named COMPASS (complex proteins associated with Set1) by interacting with common subunits of the WDR5, RBBP5, ASHL2 and DPY30 (WRAD) module, which is essential to promote enzymatic activity and allow efficient methylation (Ernst & Vakoc 2012). Moreover, complex-specific subunits contribute to dictating the functional specificity of various methyltransferases (Ruthenburg et al. 2007; van Nuland et al. 2013). For instance, SETD1A and SETD1B complexes are responsible for the bulk of H3K4me3 during mouse development (Bledau et al. 2014), while MLL enzymes have more selective roles (Yu et al. 1995; Ernst et al. 2004; Yagi et al. 1998; Glaser et al. 2006; Glaser et al. 2009; Andreu-Vieyra et al. 2010; Lee et al. 2013; Lee et al. 2008). Of note, COMPASS complexes are also responsible for mono- and dimethylation of histone 3 lysine 4 (H3K4me1 and H3K4me2), modifications that are prevalent onto distal enhancer elements involved in gene regulation (Calo & Wysocka 2013). Importantly, H3K4me3 can be removed by histone demethylases of the JMJD1 and JARID1 Jumonji C-domain-containing (JmjC) protein families (Cloos et al. 2008), which work as epigenetic “erasers” (Bannister & Kouzarides 2011).

Following histone mark “writing” by the COMPASS complexes, H3K4me3 can be “read” by multiple proteins able to specifically recognize the methylated lysine through domains such as plant homeodomain (PHD), chromo, tryptophan-aspartic acid repeat (WD40), and Tudor (Yun et al. 2011). H3K4me3 readers include the transcription factor II D (TFIID), Spt-Ada-Gcn5 acetyltransferase (SAGA), and nucleosome remodelling

factor/bromodomain PHD finger transcription factor (NuRF/BPTF) complexes (Vermeulen et al. 2007; Vermeulen et al. 2010; Bartke et al. 2010), which are involved in transcriptional activation. Accordingly, deposition of H3K4me3 is generally associated with promoters of actively transcribed genes (Santos-Rosa et al. 2002; Guenther et al. 2007; Mikkelsen et al. 2007), where it is believed to function as an on/off switch (Dong et al. 2012). Moreover, H3K4me3 is deposited in large domains on genes that are key for cellular identity and function, where it buffers transcriptional heterogeneity to ensure consistency of expression and stabilize cell fate (Benayoun et al. 2014). On the other side, H3K4me3 is not fully required for transcription (Hödl & Basler 2012), and can be deposited also on genes that are not transcriptionally active onto so-called bivalently-marked chromatin domains (Chapter 1.5.2.3; Azuara et al., 2006; Bernstein et al., 2006).

1.5.2.2 H3K27me3

H3K27me3 is deposited by the polycomb repressive complex 2 (PRC2) (Schuettengruber et al. 2007), which consists of the H3K27 methyltransferases EZH2 or EZH1 together with EED, SUZ12, RBBP4/RbAp48 and RBBP7/RbAp46 (Müller et al. 2002; Cao et al. 2002). H3K27me3 is read by chromodomain-containing factors such as EED itself, which contributes to spreading the mark in broad domains (Margueron et al. 2009), and the chromobox homolog (CBX) proteins (Kaustov et al. 2011). CBX factors are associated to the polycomb repressive complex 1 (PRC1), which they recruit onto genomic regions marked with H3K27me3 (Morey et al. 2012; Margueron et al. 2009; Morey et al. 2013). As a consequence, the RING1B and BMI1 subunits of PRC1 catalyse ubiquitination of histone 2A lysine 119 (H2AK119ub; Cao et al., 2005; Wang et al., 2004), which represses gene expression by inducing pausing of the elongating RNA polymerase 2 (Min et al. 2011; Brookes et al. 2012). In turn, H2AK119ub induces PRC2 recruitment, thus reinforcing transcriptional repression (Cooper et al. 2014). Overall, H3K27me3 is largely associated with both promoters and enhancers of transcriptionally inactive genes (Rada-Iglesias et al. 2011; Mikkelsen et al. 2007; Boyer et al. 2006; Lee et al. 2006; Squazzo et al. 2006). Finally, the erasing of H3K27me3 is achieved primarily by histone demethylases UTX and JMJD3 (Cloos et al. 2008).

1.5.2.3 Bivalent domains

Despite their apparently opposite function, H3K4me3 and H3K27me3 coexist on selected genomic regions in mESCs (Azuara et al. 2006; Bernstein et al. 2006), hESCs (Pan et al. 2007; Zhao et al. 2007), and hiPSCs (Maherali et al. 2007; Mikkelsen et al. 2007). This situation is defined as “bivalent” histone marking, and is prevalent on developmental genes that show low or no expression in the undifferentiated state. Interestingly, upon differentiation some of these bivalent regions are resolved, thus resulting in the presence of only H3K4me3 or H3K27me3 onto genes that are respectively transcriptionally activated or repressed (Figure 1.12).

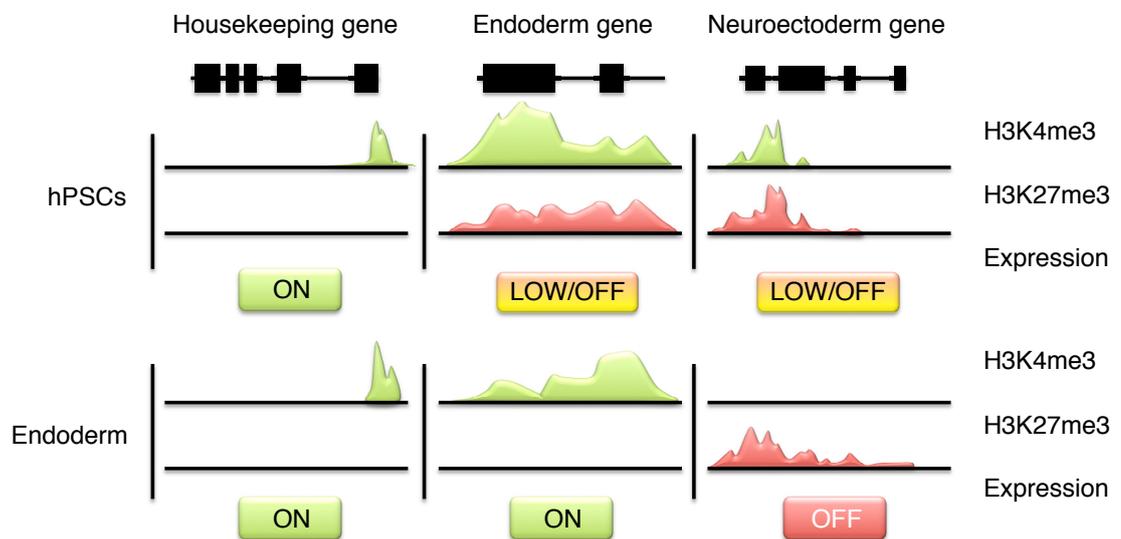


Figure 1.12. Predominant model for the role of bivalent histone marks in pluripotency and differentiation.

Overall, it is commonly hypothesized that bivalent histone marking maintains developmental regulators in an inactive yet poised state, in order to allow rapid and synchronous full activation or repression during differentiation (Vastenhouw & Schier 2012). However, strong experimental evidence of this notion is still unavailable (Voigt et al. 2013). Moreover, it is still not fully established to what extent bivalent marking is the result of true coexistence of H3K4me3 and H3K27me3 on the same nucleosome, rather than being the outcome of cellular heterogeneity and/or of H3K4me3 and H3K27me3 deposition onto separate nearby nucleosomes (Toyooka et al. 2008; Marks et al. 2012; Rotem et al. 2015). Of note, bivalent domains are not unique to PSCs, but are present also in adult stem cells (Cui et al. 2009), fully differentiated cells (Abraham et al. 2013; Roh et al. 2006), and cancers (Rodriguez et al. 2008; Bapat et al. 2010;

McGarvey et al. 2008). Therefore, bivalent histone marks might represent a universal way to regulate developmental plasticity.

The mechanism behind the generation of bivalent histone marks onto specific genomic sites is only partially known. The DNA methylation state likely plays a role, as bivalent domains are predominantly found on cytosine-phosphate-guanine dinucleotide (CpG) islands, CpG-dense regions that are largely resistant to DNA methylation and are prevalent at the transcription start sites of housekeeping and developmental genes (Bernstein et al. 2006). Accordingly, MLL1 and MLL2 contain CXXC motif DNA-binding domains that specifically recognize unmethylated CpG islands (Birke et al. 2002; Bach et al. 2009), while SETD1A and SETD1B complexes contain the CFP1 subunit, which has a similar CXXC domain and is required for H3K4me3 at promoters of strongly expressed genes (Lee et al. 2001; Clouaire et al. 2012). It was proposed that during mouse development Mll2 might work as a H3K4 methyltransferase “pioneering” factor marking bivalent promoters onto which PRC1 and PRC2 somehow prevent recruitment of Setd1A and Setd1B, which are only present when the gene is fully activated following resolution of bivalency (Denissov et al. 2014).

Aside from the methylation state, other factors are likely necessary to dictate the specificity of action of both COMPASS and polycomb complexes (Voigt et al. 2013). For instance, Oct4, physically interact with COMPASS and PRC1 complexes (Ding et al. 2012; Ang et al. 2011), and colocalizes with them onto genomic locations (Bernstein et al. 2006; Ang et al. 2011; Lee et al. 2006). Moreover, knockdown of Oct4 in mESCs impairs H3K4me3 onto selected genes, further suggesting a relationship between pluripotency factors and bivalent domains (Ang et al. 2011). However, whether the decrease of H3K4me3 observed in these experiments is the cause or consequence of transcriptional changes is still unclear, and as such the nature of this potential regulation is not fully established (Henikoff & Shilatifard 2011). Finally, the mechanism by which extracellular signalling might regulate the establishment and resolution of bivalent histone marking in PSCs is very poorly understood.

The function of H3K4me3 and H3K27me3 in mESCs has been extensively probed by genetic studies that impaired the expression of COMPASS and polycomb complexes. With regards to H3K4me3, the knockdown of Wdr5 and Ash2l impairs both self-renewal and differentiation (Wysocka et al. 2005; Ang et al. 2011; Stoller et al. 2010; Wan et al.

2013). On the other side, Rbbp5 and Dpy30 appear to be required only during mESC differentiation (Jiang et al. 2011). Among the core COMPASS enzymes, Mll2 was identified as the H3K4 methyltransferase responsible for the deposition of H3K4me3 at bivalent promoters, while active promoters rely primarily on Setd1a and Setd1b (Denissov et al. 2014; Hu, Garruss, et al. 2013). Interestingly, loss of Mll2 only moderately impairs developmental expression of bivalently-marked genes, but this is sufficient to delay germ layer specification and skew maturation of somatic cells (Lubitz et al. 2007). This is in agreement with the requirement of Mll2 for mouse development (Lubitz et al. 2007; Glaser et al. 2009). On the other hand Mll3 and Mll4 are primarily responsible for H3K4me1 on enhancers (Hu, Gao, et al. 2013). Collectively, these studies suggest that deposition of H3K4me3 in mESCs is largely dispensable for self-renewal, while it contributes to the proper coordination and timing of early cell fate decisions. Of note, whether H3K4me3 plays a similar role in hPSCs is unclear due to the lack of functional studies in this model.

Regarding H3K27me3, depletion of PRC2 subunits in mESCs has a minimal effect on self-renewal, but it strongly impairs expression of lineage specific genes and mESC differentiation (Pasini et al. 2004; Chamberlain et al. 2008; Shen et al. 2008; Leeb et al. 2010). This matches with the post-implantation lethality observed in PRC2 subunits knockout mice (Faust et al. 1995; O'Carroll et al. 2001; Pasini et al. 2004). Moreover, impairment of PRC1 in mESCs also impairs differentiation (Leeb & Wutz 2007; Alkema et al. 1995; van der Lugt et al. 1994). Overall, proper control of H3K27me3 is vital for proper PSC differentiation.

1.5.2.4 Additional notable chromatin modifications

Aside from H3K4me3 and H3K27me3, there are many additional histone modifications (Bannister & Kouzarides 2011; Kouzarides 2007), several of which have been proposed to play some role in the regulation of pluripotency (reviewed in Chen & Dent 2013). While it is beyond the scope of this chapter to cover all of these histone marks, there are a few examples of particular interest for this dissertation.

Histone acetylation has been suggested to regulate pluripotency and differentiation, since inhibitors of histone deacetylases (HDACs) promote reprogramming of somatic cells into iPSCs (Huangfu et al. 2008; Liang et al. 2010; Mali et al. 2010). Moreover,

histone acetylation decreases during mESC differentiation (Gonzales-Cope et al. 2016). Histones can be acetylated on multiple lysine residues by various histone acetyltransferases (HATs), including members of the Gcn5-related N-acetyltransferases (GNAT), MOZ-Ybf2-Sas2-Tip60 (MYST), and CREB-binding protein/p300 (CBP/p300) families (Parthun 2007; Hodawadekar & Marmorstein 2007). In general, lysine acetylation promotes transcriptional activation both by weakening the electrostatic interactions between the DNA and nucleosomes (as a result of neutralizing the positive charge of lysine; Shogren-Knaak et al., 2006), and by triggering the recruitment of bromodomain-containing epigenetic readers (Mujtaba et al. 2007). For instance, histone 4 lysine 16 acetylation (H4K16ac) on gene promoters increases their transcriptional activity (Dion et al. 2005), and histone 3 lysine 27 acetylation (H3K27ac) is associated to active enhancers (Rada-Iglesias et al. 2011).

Histone 3 lysine 9 trimethylation (H3K9me3) is another histone mark that regulates the pluripotent state, as it represents a barrier to somatic cell reprogramming (J. Chen et al. 2013), and is highly enriched following differentiation (Zhu et al. 2013; Gonzales-Cope et al. 2016). H3K9me3 is primarily deposited by methyltransferases of the Su(Var)3-9 (Suv39) family, which include G9a, GLP and SETDB1 (Shinkai & Tachibana 2011; Schultz et al. 2002), and is removed by JMJD2 histone demethylases (Cloos et al. 2008). H3K9me3 is a repressive modification that is associated to heterochromatin and exerts its function by recruiting heterochromatin protein 1 (HP1) factors (Schultz et al. 2002), which in turn promote DNA methylation (Yearim et al. 2015). In mESCs, H3K9me3 is required for silencing of endogenous retroelements (S. Liu et al. 2014), and TE genes (Yuan et al. 2009). During differentiation, deposition of H3K9me3 is essential for silencing of pluripotency genes as well as developmental regulators of alternate lineages (Bilodeau et al. 2009).

1.6 Post-transcriptional control of pluripotency and differentiation

Once transcriptional controls are established, post-transcriptional regulations of the transcriptome and proteome of hPSCs further shape specific gene expression programs. This sub-chapter discusses some notable examples of such regulations, with a particular focus on mRNA modifications given their relevance to the scope of this dissertation.

1.6.1 Control of the transcriptome and of the proteome

Protein levels are controlled at multiple post-transcriptional levels by a wide range of factors such as micro RNAs (miRNAs; Greve et al. 2013), long non-coding RNAs (lncRNAs; Ghosal et al. 2013; Flynn & Chang 2014), and RNA binding proteins (RBPs; Wright & Ciosk 2013; Ye & Blelloch 2014). Collectively, these regulators control every aspect in the life of mRNAs, including capping, splicing, polyadenylation, cleavage, nucleotide modification, nuclear export, subcellular localization, translation, and degradation (Keene 2007). Moreover, proteins are subjected to yet more levels of control through post-translational regulations such as proteolysis and the covalent addition of chemical groups, small peptides, and sugar chains (Y. C. Wang et al. 2014). In turn, these modifications influence protein functions such as enzymatic activity, interactions with other cellular factors, and stability (Deribe et al. 2010).

Overall, it should not be surprising that post-transcriptional regulations might be crucial to modulate the pluripotent state. However, this aspect has been poorly studied, and only gained the spotlight in recent years. As a result, compared to the extensive knowledge of the transcriptional and chromatin epigenetic controls of pluripotency, our understanding of the post-transcriptional mechanisms involved in this process remains relatively limited. Nevertheless, post-transcriptional regulations appear to play an important role in driving cell fate decisions (Lu et al. 2009; Tahmasebi et al. 2014). Indeed at least a few members of each class of factors mentioned above was shown to participate to the regulation of pluripotency (reviewed in Ghosal et al., 2013; Greve et al., 2013; Wang et al., 2014; Ye and Blelloch, 2014).

1.6.2 mRNA modifications

In recent years, it has become apparent that post-transcriptional biochemical modifications of mRNAs are key regulators of gene expression. Indeed, they can modulate a wide range of aspects that include mRNA splicing, localization, stability, and translation (Dominissini et al. 2014; Saletore et al. 2012). By analogy to the term epigenetic, these controls have been named epitranscriptional. Among the many modifications known to be prevalent on eukaryotic mRNAs (Machnicka et al. 2013), methylation of the nitrogen at position 6 of the adenosine base (resulting into N⁶-methyladenosine, or m6A) has been recently suggested to play a key role in the regulation of pluripotency (Liu & Pan 2016; Maity & Das 2015; Yue et al. 2015). On the other hand, the function of other common modifications (such as adenosine to inosine editing, cytosine to uracil deamination, pseudouridylation, and formation of N¹-methyladenosine) has not been extensively explored (Licht & Jantsch 2016; Dominissini et al. 2016). Therefore, the following section will focus on the regulation of m6A, also given its relevance in the context of this dissertation.

1.6.2.1 m6A

m6A was discovered more than 40 years ago (Perry & Kelley 1974; Desrosiers et al. 1974), but gained substantial attention and interest only recently following technological breakthroughs that allowed transcriptome-wide detection of m6A (Meyer et al. 2012; Dominissini et al. 2012; Linder et al. 2015; Ke et al. 2015). For this, the most commonly used methods combine m6A-specific methylated RNA immunoprecipitation with deep sequencing (hence called MeRIP-seq or m6A-seq), and allow identification of thousands of m6A sites with a resolution of ~100 nucleotides (Dominissini et al. 2013).

Early studies showed that m6A is present on average on three to five sites per mammalian RNA (Wei et al. 1975; Dubin & Taylor 1975), and is deposited on a 'RRm6AC[U/A/C]' consensus sequence (R is G or A, with G > A; Csepány et al., 1990; Harper et al., 1990). Transcriptome-wide measurement of m6A in different cells both of human and mouse origin confirmed these early observations, and demonstrated several additional remarkable aspects (Dominissini et al. 2012; Meyer et al. 2012; Schwartz et al. 2014; Ke et al. 2015; Linder et al. 2015; Fustin et al. 2013; Batista et al. 2014). First,

m6A is a widespread and evolutionary conserved modification, as ~12000 peaks are found onto highly conserved sequences of ~7000 mRNAs and ~300 noncoding RNAs. Second, the presence of a consensus m6A site does not always result in methylation. Third, control of methylation is only partially dictated by cis-acting sequences, as a given transcript can be methylated or unmethylated in different cell types, despite being expressed at comparable levels. Fourth, methylation is sub-stoichiometric, with 10-70% (an average of around 25%) of a given methylated transcript harbouring the modification. Fifth, the distribution of m6A along transcripts is non-random, with a marked enrichment in the last exon near the stop codon or the 3' untranslated region (3' UTR). Sixth, methylation can occur early during mRNA biogenesis, as it is detected also on introns. Seventh, m6A deposition can be dynamically modulated by extracellular stimuli or cellular stress. Overall, these features suggest that a remarkable level of control is involved in dictating gene-, location-, and tissue-specific deposition of m6A, as well as in regulating its stoichiometry.

Deposition of m6A is mediated by a ~1 megadalton nuclear-localized multisubunit complex (Bokar et al. 1994). Its core elements are the two methyltransferases METTL3 and METTL14, and the regulatory factor WTAP (Bokar et al., 1997, 1994; Liu et al., 2014; Ping et al., 2014; Schwartz et al., 2014; Wang et al., 2014). WTAP enhances the catalytic competence of the enzymes by altering their localization and by facilitating binding to other factors, including m6A itself. The identity of the additional subunits of the complex is still unclear, although a recent study identified 13 candidate proteins, one of which (KIAA1429) is essential for the activity of the complex (Schwartz et al. 2014). WTAP also interacts with several additional factors, including proteins involved in RNA splicing (Horiuchi et al. 2013), but their functional relevance in the context of m6A biology is unknown. Of note, miRNA have been suggested to regulate binding of METTL3 by a sequence-dependent pairing mechanism, thus potentially conferring some degree of specificity to m6A deposition (T. Chen et al. 2015).

Remarkably, m6A is a reversible modification, as two m6A “erasers” have been described: the demethylases FTO and ALKBH5 (Jia et al. 2011; Fu et al. 2013; Aik et al. 2014; Zheng et al. 2013). These two enzymes have distinct subcellular localization (FTO being both nuclear and cytoplasmic, while ALKBH5 is exclusively nuclear) and tissue distribution, thus suggesting that they might regulate different transcripts and/or

control distinct biological processes (Cheung et al. 2013; Zheng et al. 2013; Gulati et al. 2013; Vujovic et al. 2013; Gerken et al. 2007; Fischer et al. 2009).

Once m6A is deposited, it leads to downstream effects by multiple mechanisms (Figure 1.13). First, the m6A mark can be specifically bound by multiple “readers”, mostly belonging to the YT521-B homolog (YTH) domain protein family (Xu et al. 2014; Dominissini et al. 2012; Schwartz et al. 2014). YTHDF1, YTHDF2, and YTHDF3 are predominantly cytoplasmic, and mediate diverse effects. On the one hand, YTHDF1 binds to m6A on the 3' UTR and stimulates cap-dependent translation by interacting with translation initiation factors that promote ribosome loading (Wang et al. 2015). On the other hand, YTHDF2 triggers mRNA degradation by localizing RNAs to processing bodies (P-bodies; X. Wang et al. 2014; Y. Wang et al. 2014). Overall, the antagonistic action of these two factors results in high expression of proteins whose transcripts have, however, a short half-life. Of note, deposition of m6A in the 5' UTR also facilitates translation, but through a cap-independent mechanism involving the translation factor eIF3 (Zhou et al. 2015; Meyer et al. 2015). YTHDC1 and YTHDC2 are YTH-domain nuclear m6A readers, and YTHDC1 regulates pre-mRNA splicing by promoting exon inclusion (Xiao et al. 2016). HNRNPA2B1 is another m6A nuclear reader that influences alternative splicing and promotes miRNA biogenesis (Alarcón, Lee, et al. 2015; Alarcón, Goodarzi, et al. 2015).

A second way by which m6A can drive downstream responses is by altering the RNA biophysical structure (Figure 1.13). Indeed, m6A can affect base pairing thermodynamics and destabilize RNA duplex structures, thus affecting RNA-protein, RNA-RNA and RNA-DNA interactions (Kierzek 2003; Roost et al. 2015). For instance, so called “m6A-switches” control the accessibility to RNA binding protein HNRNPC (Liu et al. 2015; Zhou et al. 2016), which in turn regulates translation, stability, and alternative splicing (Rajagopalan et al. 1998; König et al. 2010). Moreover, m6A can affect translation dynamics by destabilizing pairing of the mRNA with transfer RNAs (tRNAs) depending on the specific sequence context (Choi et al. 2016).

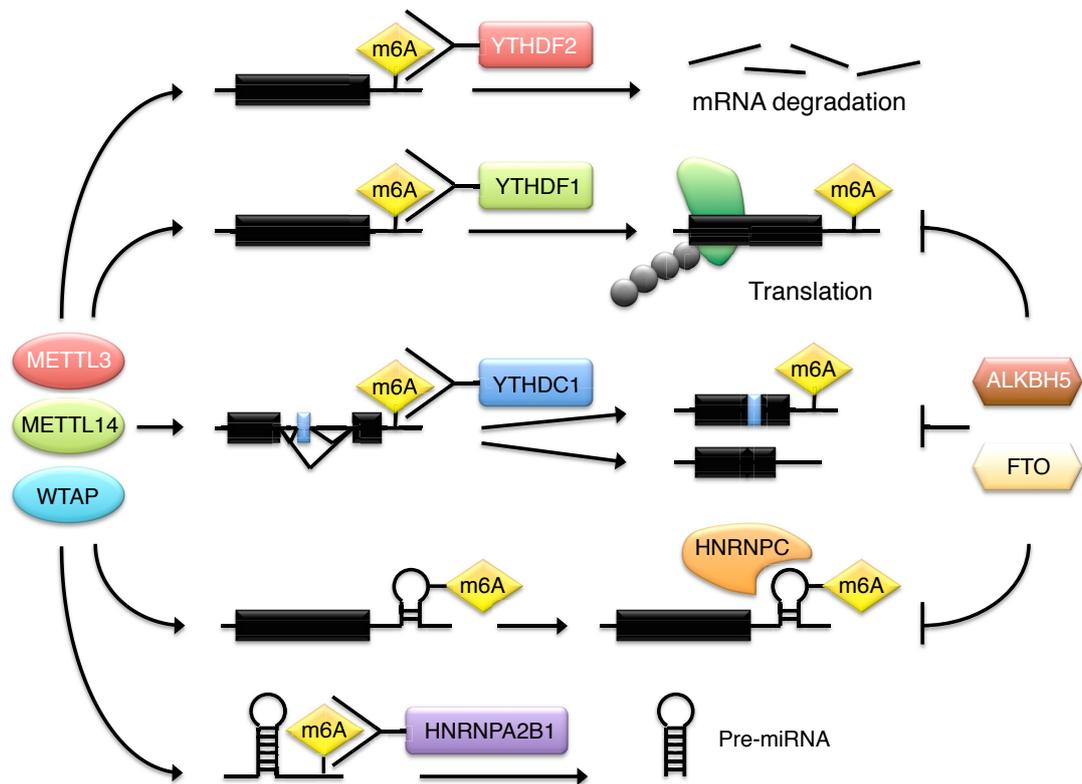


Figure 1.13. Regulation of m6A deposition and function.

Overall, m6A deposition can control several aspects of RNA biogenesis and function. However, aside from some notable examples (Kane & Beemon 1985), the precise role of any particular m6A residue on a given transcript is presently unclear. Indeed, studies reporting site-specific mutagenesis of individual methylation sites are presently scarce, especially due to the difficulty in mapping these sites with single nucleotide resolution. Moreover, the mechanisms that control the specificity of m6A deposition are still largely unknown.

1.6.2.2 Developmental function of m6A

Recent studies explored the function of m6A in pluripotency, leading to somewhat conflicting results (Yue et al. 2015; Maity & Das 2015). It was initially reported that knockdown of either Mettl3 or Mettl14 impairs mESC self-renewal and drives differentiation, suggesting a crucial role for m6A in pluripotency (Y. Wang et al. 2014). However, contrasting findings instead suggested that depletion of Mettl3 in mESCs promotes self-renewal and impairs differentiation (Batista et al. 2014). Finally, a third study proposed a model that reconciliates these observations by advocating that the

function of m6A is different in naïve and primed pluripotent cells (Figure 1.14; Geula et al., 2015). According to this view, m6A destabilizes transcripts of both pluripotency and differentiation factors, and therefore the net result of its activity depends on which of the two transcriptional networks is predominant in a given state. As a consequence, depletion of m6A reinforces the dominant transcriptional network, with the result of locking naïve cells in a “hypernaïve” state, while driving primed cells towards a “hyperprimed” state that results into premature differentiation. This model is mainly supported by *in vitro* experiments in mESCs and mEpiSCs, and is also partially recapitulated *in vivo*. Indeed, Mettl3 knockout mice are developmentally impaired since the ICM does not efficiently progress to the primed epiblast (Geula et al. 2015).

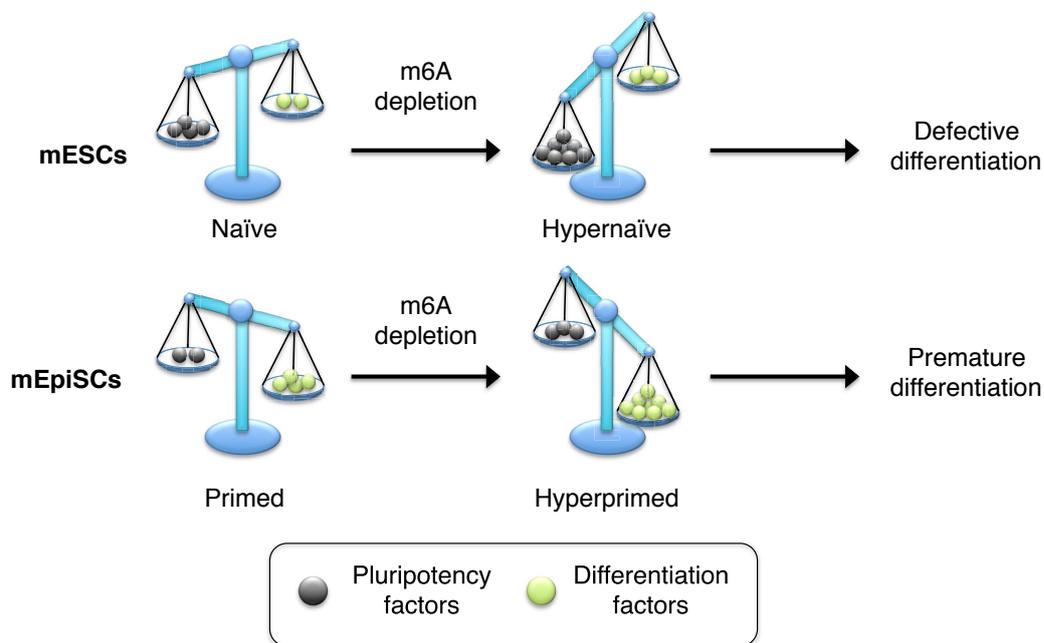


Figure 1.14. Current model for the role of m6A in murine pluripotency.

While this elegant hypothesis might explain the dual function of m6A in mice, it is unclear if this same model is directly applicable to the human context. Indeed, given the fact that conventional hPSCs represent a primed pluripotent state (Chapter 1.2.3), the prediction would be that loss of m6A should induce their differentiation. However, knockdown of METTL3 or METTL14 in hPSCs actually reinforces self-renewal and impairs differentiation (Batista et al. 2014). This observation could be explained by the fact that hPSCs are somewhat “less primed” towards differentiation than mEpiSCs (Chapter 1.3.1), or could be the result of differences between species in the function and regulation of m6A.

With regards to the *in vivo* function of m6A, aside from the mentioned role for Mettl3 in mouse post-implantation development, Wtap knockout also results into early embryonic lethality (Horiuchi et al. 2006; Fukusumi et al. 2008). In particular, Wtap knockout mice fail to gastrulate, and mESCs knockout for Wtap differentiate poorly into endoderm or mesoderm *in vitro* (while neuroectoderm differentiation is unaffected; Fukusumi et al., 2008). However, the interpretation of these results is complicated by the possible role of Wtap in the regulation of the cell cycle (Horiuchi et al. 2006), and by an important function of Wtap also in extraembryonic tissues (Fukusumi et al. 2008). On the other hand, knockout of m6A erasers has little phenotypic effects during early development, and only results in alterations of metabolism or fertility in the adult (Fischer et al. 2009; Zheng et al. 2013). This suggests that the early developmental function of m6A might be mainly regulated by the activity of its writers.

Overall, m6A deposition appears to be crucially implicated in early development, although the mechanistic details of its regulation and function need to be more precisely characterized.

1.7 Activin/Nodal signalling in hPSCs

As introduced in Chapters 1.1 and 1.3, signalling through Activin and/or Nodal is a crucial regulator of PSC early fate choices both *in vivo* and *in vitro*. This sub-chapter will describe in more detail the mechanism by which Activin/Nodal signalling controls hPSC pluripotency and differentiation, as this is the core aspect explored by this dissertation.

1.7.1 Activin/Nodal signalling

Activin and Nodal are members of the TGF β superfamily of signalling molecules, which in humans includes more than 40 members, and is among the most evolutionary conserved regulators of morphogenesis (Oshimori & Fuchs 2012; Pang et al. 2011; Huminiecki et al. 2009). Multiple Activins are generated as homo- and heterodimers of the four inhibin beta genes: INHBA, INHBB, INHBC, and INHBE, which are commonly referred to as inhibin β _a, β _b, β _c, and β _e (Welt et al., 2002). Among these, Activin-A (β _a homodimer) is the most widely used Activin for hPSC culture and differentiation (McLean et al. 2007; Xiao et al. 2006; Vallier et al. 2005; James et al. 2005). On the other hand, the biologically active form of Nodal is obtained following cleavage of a large precursor by the pro-protein convertases SPC1 and SPC4 (also known as Furin and PACE4; Beck et al., 2002). Similarly to Activin, Nodal forms homodimers. Finally, TGF β proper has three isoforms in humans (TGF β ₁, TGF β ₂, and TGF β ₃), and also undergoes cleavage and dimerization (Lawrence 1996).

Activin, Nodal and TGF β exert their function *via* heteromeric transmembrane receptor complexes that consist of two type I and two type II serine/threonine kinase receptors (Figure 1.15; Wrana et al., 1994, 1992). Binding of Activin and Nodal to type II Activin receptors (ActRIIA and ActRIIB) triggers recruitment and activation by phosphorylation of type I Activin receptor-like kinases (ALK), in particular ALK4 (which is known as ActRIB and is preferred by Activin) and ALK7 (Reissmann et al. 2001; Tsuchida et al. 2004). TGF β factors bind to different receptors: the TGFBR1 and TGFBR2 (known as ALK5; Massagué, 1998). Activin/Nodal receptor binding is tightly regulated by the expression of extracellular antagonists, such as the Nodal inhibitors LEFTY1, LEFTY2, and CER1 (Bouwmeester et al., 1996; Meno et al., 1999, 1996; Piccolo et al., 1999), and the Activin inhibitor Follistatin (Harrison et al. 2005). On the other side, the Nodal-

specific co-receptor Cripto facilitates signalling by promoting pro-Nodal processing (Blanchet, Le Good, Oorschot, et al. 2008; Blanchet, Le Good, Mesnard, et al. 2008).

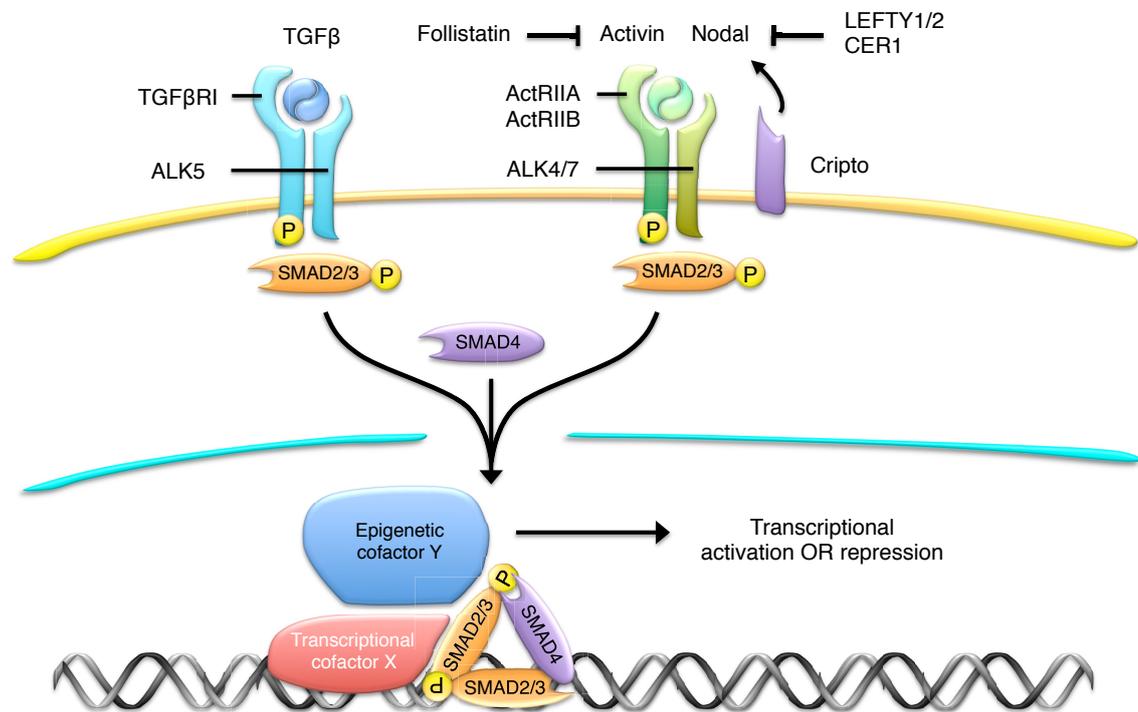


Figure 1.15. Activin/Nodal signalling.

Once activated by ligand binding, the type I receptors rapidly phosphorylate receptor-regulated Sma- and Mad-related proteins (R-SMADs), transcription factors that mediate intracellular signalling (Graff et al. 1996; Zhang et al. 1996). Despite working through different receptors, Activin/Nodal and TGFβ activate the same R-SMADs: SMAD2 and SMAD3 (Nakao, Imamura, et al. 1997; Souchelnytskyi et al. 1997; note that throughout this dissertation the notation SMAD2/3 is used when collectively referring to these two R-SMADs). Following phosphorylation, these form homo- and heteromeric complexes that bind to the common-mediator SMAD (co-SMAD) SMAD4 (Lagna et al. 1996), then accumulate in the nucleus to promote or repress gene expression through interaction with multiple factors (Chapter 1.7.3). Of note, activated SMAD complexes constantly shuttle between the nucleus and the cytoplasm, thus continuously monitoring receptor activity (Inman, Nicolás & Hill 2002; L. Xu et al. 2002). Indeed, SMAD2/3 signalling can be terminated following dephosphorylation by phosphatases such as PP1MA (Lin et al. 2006; Heikkinen et al. 2010; Yu et al. 2010). Moreover, in a negative feedback loop SMAD2/3 promotes expression of the inhibitory SMAD (I-SMAD) SMAD7, which interferes with SMAD2/3 phosphorylation and induces degradation of the receptor

complex (Ebisawa et al. 2001; Hayashi et al. 1997; Nakao, Afrakhte, et al. 1997). Finally, activated SMAD2/3 is targeted for proteasomal degradation by E3 ubiquitin ligases such as SMURF2, NEDD4-2, WWP1 and ROC1 (Lin et al. 2000; Y. Zhang et al. 2001; Seo et al. 2004; Kuratomi et al. 2005; Lo & Massagué 1999; Komuro et al. 2004; Fukuchi et al. 2001).

It is worth mentioning that Activin, Nodal and TGF β can also have SMAD-independent functions by inducing non-canonical TGF β signalling pathways (Mu et al. 2012; Moustakas & Heldin 2005; Zhang 2009; Derynck & Zhang 2003). These are specifically activated by certain TGF β superfamily receptors and coreceptors, and include the p38, c-Jun N-terminal kinase (JNK), and extracellular signal-regulated kinase (ERK) pathways (Gallagher & Schiemann 2007; Yamashita et al. 2008; Sorrentino et al. 2008), the PI3K/AKT axis (Bakin et al. 2000; Yi et al. 2005; Lamouille & Derynck 2007), and the small guanosine triphosphate hydrolases (GTPases) RHO, RAC and CDC42 (Ozdamar et al. 2005; Edlund et al. 2002). However, the function of non-canonical signalling in Activin/Nodal-dependent regulations in hPSCs is unclear, while there is extensive evidence that SMAD2/3 activation plays a major role (Chapter 1.7.4; Fei et al. 2010; Sakaki-Yumoto et al. 2013; Vallier, Mendjan, et al. 2009). Moreover, the predominant function of Smad2/3 in mediating Nodal signalling in the epiblast is well established (Chapter 1.3.2.1; Dunn et al., 2005, 2004; Heyer et al., 1999; Vincent et al., 2003; Waldrip et al., 1998). Overall, the focus of this dissertation is on SMAD2/3-dependent effects downstream of Activin/Nodal signalling.

1.7.2 Contextual determinants of the response to Activin/Nodal signalling

One peculiar characteristic of signalling through the TGF β superfamily is its multifunctional nature. Indeed, differently from the classic principle of endocrinology by which a hormone has one main function and this function only, TGF β molecules can have different and even opposite effects depending on the cellular context in which they act (Massagué 2000; Oshimori & Fuchs 2012; Massagué 2012; Pauklin & Vallier 2015). This results in paradoxical situations in which the same signal can, for example, inhibit cell proliferation in one cell type and stimulate growth in another, or work as a tumour suppressor in pre-malignant cells while encouraging metastasis in cancer cells (Ye et al. 1999; Muñoz et al. 2008; Chaudhury & Howe 2009; Fräter-Schröder et al. 1986;

Stouffer & Owens 1994). Similarly, Activin/Nodal signalling addresses opposite functions in hPSCs by promoting either self-renewal or mesendoderm differentiation depending on the culture conditions (Chapters 1.3.1, 1.3.2.1 and 1.3.2.2). While this dilemma long puzzled cellular and developmental biologists, it is now clear that these apparent contradictions can be largely reconciled and explained by considering how three main types of contextual determinants influence the activity of TGF β superfamily signals in different settings.

First, differences in expression and/or activity of signal transduction regulatory elements can largely modify the response to TGF β . These include both the core and accessory factors in the TGF β pathway itself (such as ligand traps, co-receptors, receptors subtypes, inhibitory SMADs, and E3 ubiquitin ligases), and proteins involved in other signalling pathways that crosstalk with TGF β (Guo & Wang 2009; Wu & Hill 2009). For instance the localization, activity, and degradation of R-SMADs is modulated by a plethora of post-transcriptional modifications controlled by collateral signalling, such as ubiquitination, sumoylation, acetylation and phosphorylation (Xu et al. 2012; Ross & Hill 2008; Massagué et al. 2005).

Secondly, the epigenetic landscape of a given cell dictates which gene regulatory elements are accessible for binding by activated R-SMADs, and therefore susceptible for regulation by the TGF β pathway (Massagué & Xi 2012; Xi et al. 2011). This includes the presence of DNA methylation marks, histone modifications, and nucleosome positioning (Chapter 1.5).

The third and arguably most peculiar type of contextual determinant of the response to TGF β signalling is the cell-specific expression and activity of transcriptional and epigenetic regulators that cooperate with R-SMADs to dictate specialized responses (Mullen et al. 2011; Trompouki et al. 2011; Chen et al. 1997). This is due to the particular structural features of the R-SMADs. First, the affinity of active SMAD complexes for their DNA binding sites (SMAD binding elements, SBEs) is very weak (dissociation constant $K_d \approx 1 \times 10^{-7}$ M), and sufficient binding can be achieved only if multiple SBEs are present (Shi et al. 1998; Zawel et al. 1998). However, natural SMAD target regulatory regions seldom contain such repeated SBEs, and therefore efficient SMAD binding generally relies on the synergy with other DNA-binding factors. Moreover, the most abundant isoform of SMAD2 lacks any DNA binding activity due to

the presence of an additional exon that disrupts the N-terminal DNA-binding domain (Yagi et al. 1999; Dennler et al. 1999). Overall, lineage-specific transcription factors are vital in directing the SMADs to specific loci while stabilizing DNA binding (Table 1). Another aspect of note is that unlike most other known transcription factors, DNA-bound SMAD complexes do not activate transcription by directly recruiting the basal transcription machinery. Indeed, SMADs can only promote gene expression of chromatin templates, but not of naked DNA (Ross et al. 2006). This indicates that SMADs primarily rely on modulation of the chromatin structure for their function. As such, the effect of SMAD binding to the DNA is dictated by the specific epigenetic modifiers expressed in a given cell type and/or recruited by SMAD cofactors.

Collectively, understanding the contextual determinants of the response to TGF β signalling in a given cell type is the first key step towards dissecting the mechanistic regulations involved.

1.7.3 SMAD2/3 cofactors

SMAD2 and SMAD3 consist of a highly conserved N-terminal MH1 (Mad homology 1) domain, a quite divergent linker region, and a relatively conserved C-terminal MH2 (Mad homology 2) domain (Shi & Massagué 2003; Massagué et al. 2005). While the MH1 domain allows DNA binding of SMAD3 (and of the short isoform of SMAD2), the linker and MH2 regions mediate most protein-protein interactions. Table 1 summarizes well-known SMAD2/3 interacting factors in multiple cell types, and the following paragraphs describe some notable general features of SMAD2/3-dependent transcriptional and epigenetic regulations. The specific mechanisms involved in the control of hPSC pluripotency and differentiation are discussed in detail in Chapter 1.7.4.

As described in the previous section, SMAD2/3 relies on transcriptional cofactors to efficiently bind the DNA. FOXH1 was the first known SMAD2/3 cofactor in this class. Originally identified in *Xenopus laevis* embryos (Chen et al. 1996), a Foxh1-Smad2-Smad4 complex binds to an Activin-responsive element on the promoter of Mix2 to promote its expression during mesoderm specification. It was later confirmed that Foxh1 is a functional Smad2/3 partner also during mouse development (Hoodless et al. 2001; Yamamoto et al. 2001). On the other side, SMAD2/3 also interacts with a number

of transcriptional corepressors. Two classic examples are the related proto-oncogenes SKI and SNON, which repress SMAD2/3-dependent gene expression by multiple mechanisms that include competition for SMAD4 and recruitment of histone deacetylases (Liu et al. 2001; Stroschein et al. 1999; Nomura et al. 1999).

Table 1. SMAD2/3 interactors.

Protein	Function	Cell type	Reference
ALK4, ALK7	Activin/Nodal receptor	Multiple	Reissmann et al., 2001
SARA	SMAD adaptor protein for receptor activation	Multiple	Tsukazaki et al., 1998
Nucleoporin	SMAD2/3 nuclear import/export	Multiple	Xu et al., 2000
Importin- β	SMAD3 nuclear import	Multiple	(Xiao et al. 2000)
SMAD4	Co-SMAD	Multiple	Lagna et al., 1996
SMURF2	E3 ubiquitin ligase	Multiple	Lin et al., 2000
NEED4-2	E3 ubiquitin ligase	Multiple	Kuratomi et al., 2005
WWP1	E3 ubiquitin ligase	Multiple	Komuro et al., 2004
ROC1	E3 ubiquitin ligase	Multiple	Fukuchi et al., 2001
NANOG	Transcriptional cofactor	Pluripotent cells	(Vallier, Mendjan, et al. 2009)
OCT4	Transcriptional cofactor	Pluripotent cells	Mullen et al., 2011
TRIM33	Transcriptional cofactor	Pluripotent cells	Xi et al., 2011
FOXH1	Transcriptional cofactor	Mesendoderm	Chen et al., 1996
EOMES	Transcriptional cofactor	Mesendoderm	Teo et al., 2011
β -catenin	Transcriptional cofactor	Mesendoderm	Funa et al., 2014
Mixer	Transcriptional cofactor	Mesoderm (<i>X. laevis</i>)	Germain et al., 2000
MYOD1	Transcriptional cofactor	Myotubes	Mullen et al., 2011
GATA1, GATA2	Transcriptional cofactor	Erythrocyte precursors	Trompouki et al., 2011
PU.1	Transcriptional cofactor	Pro-B cells	Mullen et al., 2011
RUNX1	Transcriptional cofactor	B cells	Hanai et al., 1999
CEBP	Transcriptional cofactor	Adipocytes	Choy and Derynck, 2003
AP-1	Transcriptional cofactor	Epithelial cells	Wong et al., 1999
ATF3	Transcriptional cofactor	Epithelial cells	Kang et al., 2003
FOXO3	Transcriptional cofactor	Epithelial cells	Seoane et al., 2004
NFKB	Transcriptional cofactor	Epithelial cells	López-Rovira et al., 2000
SP1	Transcriptional cofactor	Epithelial cells	Feng et al., 2000
E2F4/5	Transcriptional cofactor	Epithelial cells	Chen et al., 2002
SKI	Transcriptional corepressor	Multiple	Akiyoshi et al., 1999
SNON	Transcriptional corepressor	Multiple	Sun et al., 1999
ZEB-1, ZEB-2	Transcriptional corepressor	Multiple	Postigo et al., 2003
TGIF	Transcriptional corepressor	Multiple	Wotton et al., 1999
EVI-1	Transcriptional corepressor	Multiple	Kurokawa et al., 1998
EP300	Histone acetyltransferase	Multiple	Feng et al., 1998
CREBBP	Histone acetyltransferase	Multiple	Feng et al., 1998
PCAF	Histone acetyltransferase	Multiple	Itoh et al., 2000
GCN5	Histone acetyltransferase	Multiple	Kahata et al., 2004
ARC105/MED15	Mediator subunit	Multiple	Kato et al., 2002
BRG1	Nucleosome remodeller	Multiple	Ross et al., 2006
BAF155, BAF170	Nucleosome remodeller	Multiple	Ross et al., 2006
BPTF	Nucleosome remodeller	Multiple	Landry et al., 2008
HDAC4, HDAC5	Histone deacetylase	Multiple	Kang et al., 2005
SETDB1	H3K9 methyltransferase	Epithelial cells	Wu et al., 2014

Once bound to the DNA, SMAD2/3 recruits several epigenetic modifiers that can either promote or inhibit gene expression (Table 1). Among the former class, several seminal studies proved an important role for the acetyltransferases CREBBP and EP300 (Feng et al. 1998; Pearson et al. 1999; Pouponnot et al. 1998). These factors promote SMAD2/3-dependent gene expression both by directly acetylating SMAD2/3 (an event that enhances its activity) and by increasing histone acetylation (Chapter 1.5.2.4; Ross et al. 2006; Simonsson et al. 2005; Simonsson et al. 2006; Tu & Luo 2007). On the other side, in certain cellular contexts SMAD2/3 can repress gene expression through recruitment of the histone deacetylases HDAC4 and HDAC5 (Kang et al. 2005), or of the H3K9 methyltransferase SETDB1 (Wu et al. 2014).

1.7.4 Control of hPSC fate choices by Activin/Nodal signalling

As introduced in Chapters 1.3.1 and 1.3.2, Activin/Nodal signalling is the key regulator of hPSC pluripotency, while it is also essential for mesendoderm and endoderm specification. Therefore, the molecular mechanisms involved in the dual role of Activin/Nodal signalling in hPSCs have drawn substantial interest in recent years. This section analyses our current knowledge of such aspects, and highlights the open questions in the field.

1.7.4.1 Transcriptional regulation by Activin/Nodal signalling in pluripotency and differentiation

Activin/Nodal signalling maintains the undifferentiated state primarily by directly promoting the expression of pluripotency factors, in particular NANOG and OCT4 (Vallier, Mendjan, et al. 2009; Xu et al. 2008; Brown et al. 2011). Moreover, Activin/Nodal prevents expression of neuroectoderm regulators (Chng et al. 2010; Smith, Vallier, et al. 2008; Vallier, Reynolds, et al. 2004), and inhibits autocrine BMP signalling thus antagonizing its pro-differentiation effects (Galvin et al. 2010; Galvin-Burgess et al. 2013; Sakaki-Yumoto et al. 2013). On the other hand, during mesendoderm differentiation Activin/Nodal synergizes with other signalling pathways (Chapter 1.3.2.1) to promote expression of differentiation regulators such as EOMES, GSC, and SOX17 (Brown et al. 2011; Teo et al. 2011; Kim et al. 2011; Fei et al. 2010).

Considering the known general mechanisms regulating the multifunctional nature of TGF β signalling (Chapter 1.7.2), a straightforward hypothesis to explain the dual role of Activin/Nodal in hPSCs is that SMAD2/3 might bind to two distinct classes of genes following changes in its contextual determinants. However, high-throughput profiling of SMAD2/3-bound genomic regions revealed a more complex picture (Brown et al. 2011; Kim et al. 2011; Beyer et al. 2013). Indeed, these studies demonstrated that genes bound by SMAD2/3 in undifferentiated hPSCs and endoderm cells are largely overlapping, with about half of hPSC-bound genes being maintained during differentiation. Of note, this group of factors includes several master regulators of mesendoderm differentiation that despite strong SMAD2/3 binding are expressed at very low levels in hPSCs. On the other hand, it was observed that the number and/or location of SMAD2/3 binding sites on a large proportion of mesendoderm genes changes during differentiation, thus providing a potential explanation for the change in their expression.

Overall, this evidence suggests the existence of at least three types of contextual determinants that regulate the functional switch of Activin/Nodal signalling in hPSCs: (1) hPSC-specific cofactors that localize SMAD2/3 on pluripotency genes to maintain their expression; (2) factors that inhibit SMAD2/3 transcriptional activity onto mesendodermal loci in undifferentiated cells; (3) factors specific to differentiating cells that relieve such transcriptional repression, and modify the SMAD2/3 binding profile to allow transcriptional activation. Of note, genetic studies in hPSCs showed that the effects of Activin/Nodal signalling in hPSCs are mediated primarily by SMAD2 (Sakaki-Yumoto et al. 2013; Brown et al. 2011). Since SMAD2 does not bind the DNA directly, SMAD2 transcriptional cofactors are likely of prime importance in such regulations.

1.7.4.2 Molecular mechanisms regulating SMAD2/3 function in pluripotency

The pluripotency factors NANOG and OCT4 are key hPSC-specific SMAD2/3 interactors that direct its binding onto genomic loci that promote pluripotency (Mullen et al. 2011; Vallier, Mendjan, et al. 2009). Accordingly, the genome-wide occupancy of such pluripotency factors largely overlaps the one of SMAD2/3 (Mullen et al. 2011; Brown et al. 2011). For instance, SMAD2/3 and NANOG bind to and activate the promoter of *NANOG* itself, thus providing a feed-forward regulatory loop that sustains pluripotency (Xu et al. 2008; Vallier, Mendjan, et al. 2009). Of note, this regulation is conserved *in*

vivo in the mouse epiblast (Sun et al. 2014). Other notable factors that are activated by a similar mechanism are *OCT4*, *FOXD3*, *DPPA4*, *TERT*, *MYC*, *SUZ12*, *LIN28A* and *UTF1* (Brown et al. 2011).

On the other hand, multiple transcriptional corepressors tightly control SMAD2/3 activity in hPSCs. SNON, SIP1, and the Hippo pathway effectors TAZ/YAP/TEAD have all been shown to interact with SMAD2 predominantly on promoters of mesendoderm regulators, thus repressing their expression (Tsuneyoshi et al. 2012; Chng et al. 2010; Beyer et al. 2013). Interestingly, SNON is directly activated by SMAD2/3 and NANOG in undifferentiated cells, hence providing a negative feedback that restricts premature activation of mesendoderm genes (Tsuneyoshi et al. 2012). This repressive activity is lost during mesendoderm differentiation, as SNON is rapidly downregulated due to the loss of NANOG expression and following ubiquitin-induced degradation (Nagano et al. 2007). On the other side, SIP1 is repressed by SMAD2/3 and NANOG, while it is induced by SOX2 (Chng et al. 2010). Collectively, this tightly controls SIP1 expression in undifferentiated cells, while it facilitates SIP1 upregulation during neuroectoderm specification (which is essential for efficient differentiation; Chng et al. 2010). Finally, the repressive activity of the TAZ/YAP/TEAD complex is promoted by OCT4, thus forming yet another negative feedback loop restricting SMAD2/3 activity in hPSCs. However, during mesendoderm differentiation this complex is displaced from SMAD2/3, possibly due to the binding of FOXH1, thus relieving the repression of mesendoderm genes (Beyer et al. 2013).

Overall, the balance between transcriptional coactivators and corepressors finely controls SMAD2/3 activity in hPSCs in order to maintain the pluripotent state (Figure 1.16).

1.7.4.3 Molecular mechanisms regulating SMAD2/3 function during differentiation

FOXH1 is a well-known SMAD2/3 cofactor (Chapter 1.7.3). FOXH1 is expressed both in hPSCs and in mesendoderm cells, but it significantly colocalizes with SMAD2/3 on a genome-wide scale only during differentiation (Kim et al. 2011). Moreover, FOXH1 might displace the TAZ/YAP/TEAD complex during mesendoderm specification in order to allow expression of important lineage regulators (Beyer et al. 2013). As such, it

appears that functional cooperation between FOXH1 and SMAD2/3 is primarily involved in the induction of differentiation, albeit this has never been thoroughly investigated.

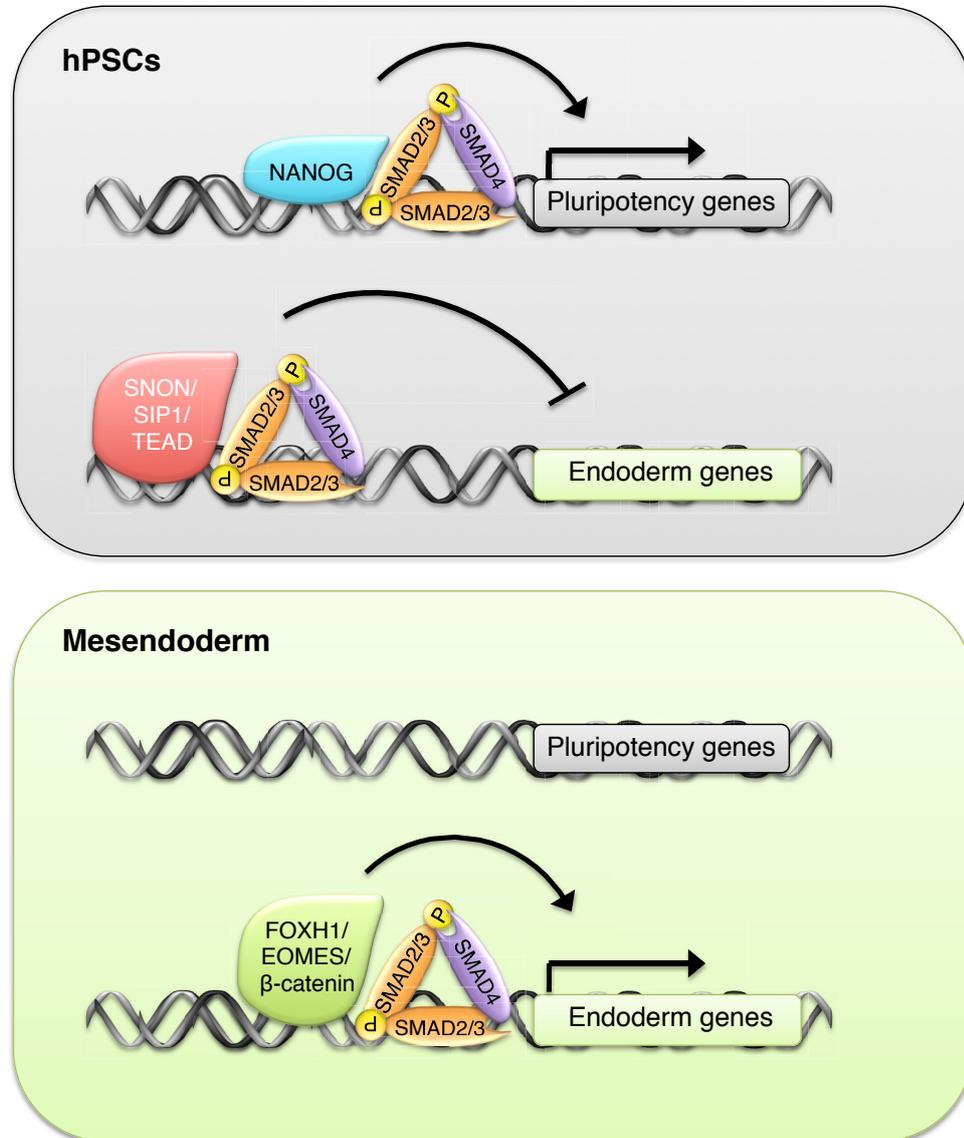


Figure 1.16. Regulation of SMAD2/3 transcriptional activity in pluripotency and differentiation.

EOMES is another important SMAD2/3 cofactor during mesendoderm and endoderm differentiation (Teo et al. 2011). Interestingly, EOMES is among the earliest mesendoderm factors to be expressed during differentiation, and this is directly under the control of SMAD2/3 and NANOG. In turn, EOMES interacts with SMAD2/3 on a genome-wide level to promote expression of a broad network of endoderm genes that include *SOX17*, *GSC*, and *FOXA2*. Therefore, during sequential stages of development

SMAD2/3 promotes the expression of master lineage regulators, and then actively associates with such factors to further drive differentiation.

β -catenin is yet another SMAD2/3 partner, and it mediates the crosstalk between Activin/Nodal and WNT signalling during mesendoderm differentiation (Funa et al. 2014; Mendjan et al. 2014). Indeed, stabilization of β -catenin downstream of WNT facilitates SMAD2/3-dependent activation of mesendoderm genes, while on the other hand SMAD2/3 prevents spurious activation of neural crest genes by β -catenin (Funa et al. 2014). Notably, β -catenin is also involved in the crosstalk between Activin/Nodal and insulin/IGF (Singh et al. 2012). Indeed, inhibition of the PI3K/AKT pathway downstream of insulin/IGF leads to activation of ERK1/2, which represses GSK3 β thus stabilizing β -catenin.

Aside from changes in SMAD2/3 transcriptional cofactors, studies from mESCs suggest that additional mechanisms might facilitate the functional switch of Activin/Nodal signalling during differentiation. Indeed, the histone binding protein Trim33 (also known as Tif1 γ) is a Smad2/3-interacting factor that exposes previously inaccessible Smad2/3 binding sites onto poised mesendoderm genes such as Gsc and Mixl1 (Xi et al. 2011; Massagué & Xi 2012). For this, Trim33 recognizes and binds to genomic regions that are marked by histone marks H3K18ac (histone 3 lysine 18 acetylation) and H3K9me3, then displaces repressing histone marks to reveal previously secluded Smad2/3 binding sites. However, whether this regulation is also important in hPSCs is presently unclear.

A final important regulator of the response to SMAD2/3 during mesendoderm differentiation is the cell cycle (Pauklin & Vallier 2013). Indeed, it was recently reported that the activity of SMAD2/3 is strongly regulated during the various cell cycle phases, and that only cells in the early stages of gap 1 (G1) are sufficiently responsive to Activin/Nodal signalling to efficiently initiate mesendoderm specification. On the other hand, cells in the late G1 phase are so unresponsive to Activin/Nodal that neuroectoderm differentiation is strongly facilitated. Such regulation is achieved by the cell cycle-dependent phosphorylation of SMAD2/3 on its linker region by CDK4/6, which impairs SMAD2/3 nuclear translocation and therefore its transcriptional activity (Pauklin & Vallier 2013; Dalton 2013).

Overall, during the induction of mesendoderm differentiation there are remarkable changes in the transcriptional activity of SMAD2/3 (Figure 1.16). First, mechanisms that repress SMAD2/3-bound mesendoderm genes in hPSCs are relieved. Secondly, SMAD2/3 switches binding partners from pluripotency factors to mesendoderm regulators. Third, chromatin remodelling events expose novel SMAD2/3 binding sites. Fourth, SMAD2/3 transcriptional activity is increased both by augmenting the dose of Activin/Nodal ligands, and by means of cell-cycle dependent intracellular events. Collectively, these regulations allow the functional switch of Activin/Nodal signalling from a pluripotency-maintaining to a pro-differentiation signal.

1.7.4.4 Open questions

Our current knowledge behind the function of SMAD2/3 in hPSCs is fragmentary and comes from few hypothesis-driven studies examining individual regulations. On the other hand, a global picture of the molecular mechanisms regulated by SMAD2/3 is still lacking.

Most importantly, while several transcriptional cofactors regulating SMAD2/3 function during self-renewal and differentiation have been identified, very little is known about the mechanisms by which these interactions activate or repress specific transcriptional responses. Indeed, hPSCs have a distinctive chromatin structure (Chapter 1.5), and it is therefore unclear whether SMAD2/3 controls gene expression in these cells using the same general regulations involved in other lineages, or whether specific mechanisms are at play.

Moreover, it is unknown whether SMAD2/3 might affect mechanisms other than transcriptional and epigenetic controls, as studies so far focused only on such canonical functions. However, post-transcriptional controls of gene expression and function are crucially involved in the control of pluripotency and differentiation (Chapter 1.6). Given the central function of Activin/Nodal signalling in hPSCs, it is foreseeable that it might play a role also in such aspects.

Finally, how the various molecular regulations controlled by Activin/Nodal signalling might contribute to finely regulate the balance between hPSC self-renewal and differentiation is still not fully understood.

1.8 Methods to study gene function in hPSCs

As described so far, hPSCs represent an invaluable model to study human development. However, functional analyses of developmental mechanisms require efficient ways to probe gene function in hPSCs. This sub-chapter describes the main approaches available to achieve this goal, with a focus on the limitations of current technologies.

1.8.1 Genetic engineering

The study of gene function primarily relies on the manipulation of gene expression. Unfortunately, this task proved extremely complex in the early days of hPSC culture, as these cells proved resistant to most traditional methods. First, transfection, electroporation, and nucleofection of hPSCs are relatively inefficient and poorly reproducible (Luo et al. 2016; Costa et al. 2007; Cao et al. 2010; Braam et al. 2008; Green et al. 2008; Beloor et al. 2015). Secondly, while viral transduction is more robust, hPSCs rapidly and efficiently silence most viral transgenes (Ellis 2005; Yao et al. 2004; Herbst et al. 2012). Third, conventional gene editing based on homologous recombination is extremely ineffective in hPSCs, and is further complicated by their very limited clonogenicity (Byrne et al. 2014; Zwaka & Thomson 2003; Urbach et al. 2004). Fourth, while stable integration of plasmids carrying transgenes or short hairpin RNAs (shRNAs) is feasible (Vallier, Rugg-Gunn, et al. 2004; Gerrard et al. 2005; Eiges et al. 2001), their random integration can lead to insertional mutagenesis and is subject to positional effects that can strongly limit transgene expression (Zafarana et al. 2009; Krishnan et al. 2006; Stein et al. 2010; Ott et al. 2006). Moreover, hPSC differentiation can induce silencing of randomly integrated transgenes located in regions where heterochromatin forms following cell fate choices (Herbst et al. 2012; Raya et al. 2009). As a consequence, these practical limitations limited for many years the widespread use of hPSCs for functional genetics studies. Indeed, most researchers opted for models more amenable to genetic manipulation, such as mESCs (Doetschman et al. 1988; Thomas & Capecchi 1987).

This situation was dramatically changed in recent years following the development of more efficient gene editing tools such as zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs) (Figure 1.17; Joung and Sander, 2013; Urnov et al., 2010). These methods are based on customizable DNA binding domains that are

engineered to recognize specific sequences, and are fused to nucleases in order to induce site-specific double-strand DNA breaks. Such events can be repaired by the error-prone non-homologous end joining (NHEJ) pathway, resulting in random small insertions or deletions (indels) that can generate loss-of-function alleles. Alternatively, homologous directed repair (HDR) mechanisms can drive recombination of a donor DNA fragment carrying specific mutations, thus resulting in precise gene editing (Gaj et al. 2013; Kim & Kim 2014).

More recently, this toolbox was further expanded by modified prokaryotic type II clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 systems, whereby a Cas9 endonuclease is selectively targeted to a genomic locus by a guide RNA (gRNA) through Watson-Crick base pairing (Cong et al. 2013; Mali et al. 2013; Jinek et al. 2012). Given its simplicity, efficiency, and flexibility, CRISPR/Cas9-mediated gene editing has quickly become the gold standard for genetic manipulation of multiple cell types, including hPSCs (Wright et al. 2016). Moreover, variation of traditional CRISPR/Cas9 systems that are based on fusion proteins with a catalytically inactive Cas9 allow a wide range of applications, including imaging and purification of genomic loci (Ma et al. 2016; B. Chen et al. 2013; Waldrip et al. 2014; Fujita & Fujii 2013), transcriptional interference or activation (Gilbert et al. 2013; Cheng et al. 2013; Kearns et al. 2014), and targeted epigenetic modification (Hilton et al. 2015; Vojta et al. 2016).

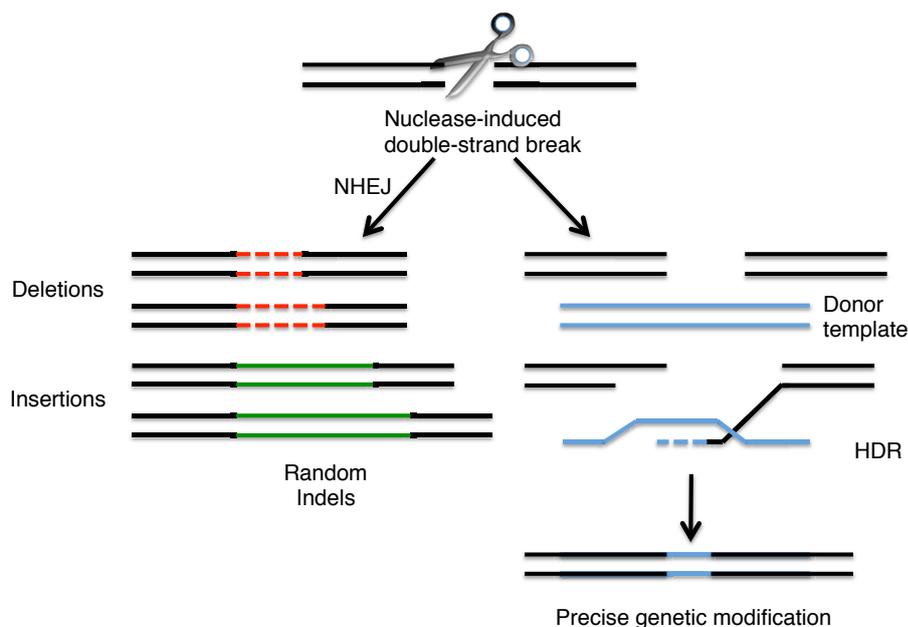


Figure 1.17. Nuclease-facilitated gene editing.

Another major advance in genetic engineering methods was the identification of human genomic safe harbours (GSHs): regions in the human genome that are expressed in virtually any human cell type, are resistant to gene silencing, and can be genetically modified without negatively affecting cellular functionality (Sadellain et al. 2012). Therefore, GSHs are attractive options for robust genetic engineering of hPSCs, and advances in gene targeting methods have fostered the use of GSHs to express transgenes, genetic reporters, and shRNAs (Hockemeyer et al. 2011; Kun et al. 2014; Hockemeyer et al. 2009; DeKolver et al. 2010; Gaj et al. 2013; Smith, Maguire, et al. 2008).

1.8.2 Conditional manipulation of gene expression

Aside from the technical challenges, manipulation of gene expression in hPSCs presents important conceptual problems. First, the study of gene function in hPSCs is complicated by the need of maintaining such cells in a self-renewing undifferentiated state. As such, increased or decreased expression of genes implicated in pluripotency, proliferation, and survival is often incompatible with maintaining healthy hPSCs. Secondly, most methods for efficient genetic manipulation of hPSCs require lengthy clonal selection procedures. Therefore, it is often difficult to distinguish between the immediate effects of any given genetic manipulation and its potential indirect downstream effects. Finally, studying gene function at a specific stage of hPSC differentiation is complicated if such gene is involved at multiple steps during lineage specification. Overall, stable loss- or gain-of-function experiments in hPSCs have only limited value, while conditional manipulation of gene expression is crucial to fully harness the power of hPSCs for functional genetics applications. However, while several robust methods for inducible transgene overexpression in hPSCs have been reported (Vallier et al. 2007; Kim et al. 2016; Kun et al. 2014), inducible loss-of-function approaches are less established.

On the one hand, recently reported technologies for inducible gene knockout in hPSCs still have important drawbacks. First, the most advanced recombination-based inducible knockout method by Chen et al., (2015) is still very complex and time-consuming. Indeed, it requires generation of a different targeting vector for each experiment, and it relies on two steps of gene targeting. Secondly, the inducible CRISPR/Cas9 knockout

approach described by González et al., (2014) relies on the introduction of random indels. Therefore, this generates a mixed cell population carrying different mutations, including some that do not induce loss of gene function. Third, in a recently published inducible CRISPR/Cas9 interference system (Mandegar et al., 2016) the expression of the gRNA is achieved either by transient transfection, which is poorly efficient and not fully reproducible, or by random integration of the gRNA, which can result in mosaic expression. Moreover, all of the current inducible CRISPR/Cas9-based methods involve conditional overexpression of the Cas9 or of Cas9 fusion proteins. However, this is achieved using an inducible promoter (the tetracycline responsive element, or TRE) that is heavily silenced during hPSC differentiation into multiple lineages, even after targeting into GSHs (Mandegar et al. 2016; Haenebalcke et al. 2013; Ordovas et al. 2015). Therefore, these systems are unlikely to work in a diversity of cell types.

Inducible gene knockdown using shRNAs provides an useful alternative to knockout studies (Boettcher & McManus 2015; Lambeth & Smith 2013). The tetracycline derepressible (TET-OFF) system is the most widely used method to inducibly express shRNAs in mammalian cells (Figure 1.18). It relies on a modified RNA polymerase III (Pol III) promoter that is responsive to a tetracycline-sensitive Repressor protein (tetR). Following simple tetracycline treatment, the tetracycline-bound tetR undergoes a conformational change that prevent its binding to the Pol III promoter, which is therefore derepressed. As a result, the shRNA is induced and drives gene knockdown (Lambeth & Smith 2013; Kappel et al. 2007).

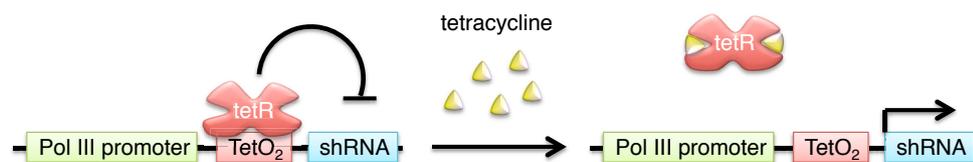


Figure 1.18. Inducible gene knockdown using the TET-OFF system

However, application of the TET-OFF system in hPSCs has proved challenging for two main reasons: (1) tight control of shRNA expression is difficult to achieve; (2) induction of the shRNA rarely works in differentiated derivatives. Indeed, very high and homogenous expression of both the tetR and the inducible shRNA is required to obtain a potent yet controlled knockdown. However, current methods rely either on lentiviruses or randomly integrated plasmids (Zafarana et al. 2009; Tsuneyoshi et al. 2012; Massé et al. 2011; Lian et al. 2012; Laperle et al. 2015), and thus suffer from the important

limitations described above. As a consequence, inducible shRNA expression both in hPSCs and in a wide variety of their differentiated progenies has never been reported.

1.9 Objectives

The overarching goal of this dissertation is to gain insight into the molecular mechanisms involved in the activity of Activin/Nodal signalling in hPSCs (Chapter 1.7.4.4).

Towards this end, this dissertation aims to primarily address the following questions:

- I. What are the chromatin changes induced by Activin/Nodal signalling in hPSCs?
- II. Are mechanisms other than transcriptional and epigenetic regulations controlled by Activin/Nodal signalling in hPSCs?
- III. How does Activin/Nodal signalling regulate both hPSC pluripotency and differentiation?

For this, we employed multiple approaches that will be discussed in the following chapters:

Chapter 3 aims to test the hypothesis that Activin/Nodal signalling might be involved in the regulation of H3K4me3 and H3K27me3 histone marks in hPSCs.

Chapter 4 aims to develop an optimized method to conditionally probe gene function at specific stages of hPSC differentiation, and to apply this approach to investigate the regulations downstream of Activin/Nodal signalling in hPSCs.

Chapter 5 aims to perform proteomic analyses in order to provide a global and unbiased overview of the mechanisms involved in the function of Activin/Nodal signalling during both self-renewal and differentiation of hPSCs.

2 MATERIAL AND METHODS

2.1 Statement of source

Certain methods in this chapter describe experiments that are presented in the following first author manuscripts written by the author of this dissertation. Therefore, some sections are taken *verbatim* or with minor changes from these sources.

Bertero A, Madrigal P, Galli A, Hubner NC, Moreno I, Burks D, Brown S, Pedersen RA, Gaffney D, Mendjan S, et al. 2015. Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev* **29**: 702–17.

Bertero A, Pawlowski M, Ortmann D, Snijders K, Yiangou L, Cardoso de Brito M, Brown S, Bernard WG, Cooper JD, Giacomelli E, et al. Optimized inducible shRNA and CRISPR/Cas9 platforms to study human development. Manuscript in review at *Development*.

2.2 General methods for all chapters

2.2.1 hPSC culture

Feeder- and serum-free hESC (H9 line; WiCell) and hiPSC (A1AT^{R/R} line, Yusa et al., 2011) culture was previously described (Vallier, 2011; refer to Table 2 for media compositions). Cells were plated on gelatin- and MEF medium-coated plates, and cultured in chemically defined medium (CDM) containing either bovine serum albumin (BSA) or polyvinyl alcohol (PVA), respectively for hESCs or hiPSCs. CDM was supplemented with 10ng/ml Activin-A and 12ng/ml FGF2 (both from Dr Marko Hyvonen, Dept. of Biochemistry, University of Cambridge). Cells were passaged every 5-6 days with collagenase or a 50%-50% mixture of collagenase and Dispase (respectively for hESCs or hiPSCs), and plated as clumps of 50-100 cells dispensed at a density of 100-150 clumps/cm². The culture medium was replaced 48h after the split and then every 24h. All assays on undifferentiated hPSCs were performed the third day after the split unless otherwise indicated.

Table 2. Media for hPSC culture and passaging.

Medium	Components	Concentration	Supplier
CDM BSA/PVA	IMDM : F-12 (1:1)	-	Gibco
	Chemically defined concentrated lipids	1%	Gibco
	1-thioglycerol	450mM	Sigma-Aldrich
	Insulin	7µg/ml	Roche
	Transferrin	15µg/ml	Roche
	Penicillin-Streptomycin (optional)	1%	Gibco
	Bovine serum albumin (BSA)	5mg/ml	Europa Bio Product
	OR polyvinyl alcohol (PVA)	1mg/ml	Sigma-Aldrich
KSR medium	Advanced DMEM F12	400mL -	Gibco
	Knockout Serum Replacer (KSR)	20%	Gibco
	L-glutamine	1%	Gibco
	β-mercaptoethanol	100µM	Sigma-Aldrich
	Penicillin-Streptomycin (optional)	1%	Gibco
MEF Medium	Advanced DMEM F12	-	Gibco
	Foetal bovine serum (FBS)	10%	Biosera
	L-glutamine	1%	Gibco
	β-mercaptoethanol	100µM	Sigma-Aldrich
	Penicillin-Streptomycin (optional)	1%	Gibco
Gelatine	Embryo transfer water	-	Gibco
	Gelatine from porcine skin	0.1%	Sigma-Aldrich
Collagenase	Advanced DMEM F12	-	Gibco
	Knockout serum replacer (KSR)	20%	Gibco
	L-glutamine	1%	Gibco
	Collagenase IV	1mg/ml	Gibco
Dispase	DMEM/F12	-	Gibco
	Dispase	1mg/ml	Gibco

2.2.2 hPSC differentiation

Differentiation was initiated in adherent hESC cultures 48h following passaging. Unless otherwise specified, medium changes were performed daily, and volumes were adjusted to cell density.

2.2.2.1 Germ layer specification

Neuroectoderm was induced for 6 days in CDM-BSA with 12ng/ml FGF2 and 10 μ M SB431542 (Activin/Nodal/TGF β signalling inhibitor; Tocris), as previously described (Vallier, Touboul, Chng, et al. 2009). Mesoderm specification was obtained in two steps: (step 1) induction of early mesoderm/primitive streak-like cells (late streak, posterior streak, and anterior streak, respectively for presomitic, lateral plate, and cardiac mesoderm) for 36h, a single medium change; (step 2) mesoderm patterning for three and a half days, two medium changes. Presomitic mesoderm (Mendjan et al. 2014): (step 1) CDM-BSA with 40ng/ml FGF2 and 8 μ M CHIR99021 (WNT signalling activator; Tocris); (step 2) CDM-BSA with 4ng/ml FGF2, 1 μ M all-trans Retinoic Acid (RA; Sigma-Aldrich), 0.1 μ M LDN193189 (BMP signalling inhibitor; Biovision), 10 μ M SB431542, and 1 μ M Purmorphamine (sonic hedgehog/SHH signalling activator; Tocris). Lateral plate mesoderm (Cheung et al. 2012): (step 1) CDM-PVA with 20ng/ml FGF2, 10 μ M LY294002 (PI3K inhibitor; Promega), and 10ng/ml BMP4 (R&D); (step 2) CDM-PVA with 20ng/ml FGF2, and 50ng/ml BMP4 (R&D). Cardiac mesoderm (Mendjan et al. 2014): (step 1) CDM-BSA (without insulin) with 20ng/ml FGF2, 10 μ M LY294002, 50ng/ml Activin-A, and 10ng/ml BMP4; (step 2) CDM-BSA (without insulin) with 8ng/ml FGF2, 10ng/ml BMP4, 1 μ M IWR1 (WNT signalling inhibitor; Sigma-Aldrich), and 0.5 μ M RA. Definitive endoderm was differentiated for 3 days in CDM-PVA (without insulin) with 20ng/ml FGF2, 10 μ M LY294002, 100ng/ml Activin-A, and 10ng/ml BMP4, as previously described (Touboul et al. 2010).

2.2.2.2 Neural cell types

Neurons, oligodendrocytes, and astrocytes were obtained in collaboration with Dr Matthias Pawlowski according to a published protocol (Douvaras et al. 2014), with minor modifications. hESCs were cultured in CDM-BSA with 10 μ M SB431542, 0.1 μ M LDN193189 for 8 days to generate neuroepithelial cells. During this stage, 0.1 μ M RA

and $1\mu\text{M}$ Purmorphamine were respectively added from day 2 and day 4 onwards to promote neuroepithelial patterning towards a caudal and ventral fate. From day 8 to 12, medium was switched to N2 medium (Thermo) with $0.1\mu\text{M}$ RA and $1\mu\text{M}$ Purmorphamine. At day 12, adherent neural progenitors were mechanically lifted and transferred into low-attachment plates to form neurospheres. During this stage, cells were initially cultured in N2B27 medium (Thermo) supplemented with $0.1\mu\text{M}$ RA and $1\mu\text{M}$ Purmorphamine for 8 days. Subsequently, neuronal and glial specification was initiated at day 20 of differentiation by culturing neurospheres in differentiation medium consisting of N2B27 supplemented with 60ng/ml triiodo-L-thyroxine (Sigma-Aldrich), 100ng/ml biotin (Sigma-Aldrich), $1\mu\text{M}$ dibutyryl cAMP (protein kinase A/PKA agonist; Sigma-Aldrich), $10\mu\text{g/ml}$ insulin, 10ng/ml FGF2, 10ng/ml platelet-derived growth factor type "AA" (PDGF-AA; Peprotech), and 10ng/ml Neurotrophin-3 (R&D). From this point onwards, medium changes were performed every other day. At day 30, spheres were plated onto poly-D-lysine- and laminin-coated dishes to promote neuronal and glial outgrowth, and cultured in differentiation medium without FGF2. From day 70 onwards, the concentration of PDGF-AA and Neurotrophin-3 was reduced to half of the previous levels to promote cell maturation. Differentiation concluded at day 95, and resulted into a mixed culture of mostly astrocytes and neurons, and a minority of oligodendrocytes.

2.2.2.3 Neural crest cells and osteocytes

Neural crest and osteocyte differentiation was performed in collaboration with Mr William G. Bernard and Dr Felipe Serrano following modifications of a previously published protocol for neuroectoderm differentiation (Cheung et al. 2014; Felipe Serrano, William G. Bernard, et al., manuscript in preparation). Cells were differentiated into neuroectoderm for 5 days, and neural crest cells were enriched by passaging this population (which resulted in preferential attachment of neural crest cells). Osteocytes were then differentiated using the StemPro Osteogenesis Differentiation Kit (Gibco) according to manufacturer's instructions. Cells were cultured in these conditions for two weeks performing medium changes every other day.

2.2.2.4 Chondrocytes

Chondrocyte differentiation was performed in collaboration with Miss Elisa Giacomelli as previously described (Mendjan et al. 2014), with minor changes. Following presomitic mesoderm differentiation for 5 days, chondrogenic specification was induced for 10 days in CDM-BSA with 8ng/ml FGF2 and 10ng/ml BMP4 and performing medium changes every other day.

2.2.2.5 Smooth muscle cells

Smooth muscle cells were obtained as previously reported (Cheung et al. 2014; Cheung et al. 2012). Following lateral plate mesoderm differentiation for 5 days, cells were dissociated in TrypLE Express (Gibco) for 5' at 37°C, washed with CDM-PVA by centrifuging at 200g for 3' at room temperature (RT), and seeded at a density of 2×10^4 cells/cm² onto gelatin- and MEF medium-coated plates in CDM-PVA supplemented with 10ng/ml platelet-derived growth factor type "BB" (PDRG-BB; Peprotech) and 2ng/ml TGFβ (Peprotech). Cells were cultured in these conditions for 12 days to generate smooth muscle cells. Medium was changed every other day, and cells were split after the first 6 days of culture as just described but seeded at a 1:2 ratio.

2.2.2.6 Epicardial cells and cardiac fibroblasts

Differentiation into epicardium and cardiac fibroblasts was performed in collaboration with Dr Dharini Iyer and Dr Laure Gambardella as previously reported (Iyer et al. 2015). Following lateral plate mesoderm differentiation for 5 days, cells were dissociated in TrypLE Express for 5' at 37°C, washed with CDM-PVA by centrifuging at 200g for 3' at RT, and seeded at a density of 2.5×10^4 cells/cm² onto gelatin- and MEF medium-coated plates in CDM-PVA supplemented with 25ng/ml WNT3A (R&D), 50ng/ml BMP4 and 4μM RA. Cells were cultured in these conditions for 10 days with medium changes every 4 days to generate epicardial cells. Following this, cells were split as just described and seeded at a density of 3×10^4 cells/cm² in CDM-PVA supplemented with 50ng/ml vascular endothelial growth factor B (VEGF-B; Peprotech) and 50ng/ml FGF2. Cells were cultured in these conditions for 12 days with medium changes every other day to generate cardiac fibroblasts.

2.2.2.7 Cardiomyocytes

Cardiac differentiation was performed as previously described (Mendjan et al. 2014), with minor changes. Following cardiac mesoderm differentiation for 5 days, cardiac maturation was initiated by culturing cells in CDM-PVA (without insulin) with 8ng/ml FGF2 and 10ng/ml BMP4 for 2 days (one medium change). Following this, cells were cultured in CDM-PVA (without insulin) until beating clusters appeared (between day 8 and 10 of differentiation). Cardiomyocytes were then matured in CDM-PVA containing insulin for a week. Medium changes were performed every other day.

2.2.2.8 Intestinal cells

Intestinal organoid generation was performed in collaboration with Miss Elisa Giacomelli following a previously published method (Fordham et al. 2013; Hannan, Fordham, et al. 2013), with some minor changes. Following definitive endoderm differentiation for 3 days, cells were cultured in RPMI/B27 medium (consisting of RPMI Medium 1640 with GlutaMAX, 2% B27 supplement, 1% non-essential amino acids, 100U/ml Penicillin, and 100µg/ml Streptomycin; all from Gibco). This was first supplemented with 50ng/ml Activin-A for 1 day, then with 6µM CHIR99021 and 3µM RA for 4 days, in order to pattern definitive endoderm into a posterior fate and obtain hindgut cells. Monolayer cultures were then transferred to three-dimensional conditions. For this, cells were dissociated to small clumps using collagenase for 20' at 37°C, washed twice with basal growth medium (consisting of Advanced DMEM/F12, 10mM HEPES pH 7.4, 2% B27 serum-free supplement, 1% N2 serum-free supplement, and 20mM L-glutamine) by centrifuging at 200g for 3' at RT, and finally resuspended at a density of 1×10^5 cells/ml in a mixture of 70% Matrigel (BD Biosciences) and 30% basal growth medium supplemented with 500ng/ml R-Spondin 1 (R&D), 3µM CHIR99021, 100ng/ml Noggin (BMP signalling inhibitor, R&D), 2.5µM Prostaglandin E2 (Cayman Chemicals), 100ng/ml epidermal growth factor (EGF; R&D), and 0.5µM A83-01 (TGFβ signalling inhibitor, Tocris). The cell-medium-matrigel solution was distributed in 50µl droplets, one for each well of 24-well plates. The droplets were allowed 30' for the gel to solidify, following which 1ml of basal medium with growth factors was added per well. 10µM Y27632 (Rho kinase inhibitor; Tocris) was also added for the first 24h only to promote cell survival. Cells were cultured in these conditions for 10 days with medium changes every other day to generate intestinal organoids.

2.2.2.9 Pancreatic cells

Pancreatic differentiation was done according to what previously reported (Cho et al. 2012), with minor changes. Following definitive endoderm differentiation for 3 days, cells were cultured in Adv-BSA medium (consisting of Advanced DMEM/F12, 5mg/ml BSA, 20mM L-glutamine, 100U/ml Penicillin, and 100µg/ml Streptomycin). This was supplemented with 3µM RA, 50ng/ml FGF10 (Autogen Bioclear), 150ng/ml Noggin, and 10µM SB431542 for 3 days to generate dorsal foregut, then with 3µM RA, 50ng/ml FGF10, 150ng/ml Noggin, and 0.25µM KAAD-Cyclopamine (SHH signalling inhibitor, Toronto Research Chemicals) for 3 days, and finally with 3µM RA, 50ng/ml FGF10, and 0.25µM KAAD-Cyclopamine for 3 days. Pancreatic progenitors were obtained at this stage (day 12 of differentiation). Pancreatic specification was then initiated in Adv-BSA with 3µM RA, 1% B27 supplement, and 1µM DAPT (Notch signalling inhibitor, Sigma-Aldrich) for 3 days. During this stage, 0.1mM 6-Bnz-cAMP (PKA activator, Sigma-Aldrich) was also added for the first 48h only. Cells were then cultured in Adv-BSA with 3µM RA, 1% B27 supplement, and 0.25µM KAAD-Cyclopamine for 3 days (one medium change), in order to obtain immature pancreatic endocrine cells. These were further matured for 6 days (medium changes every 72h) under the same culture conditions to generate mature pancreatic endocrine cells at day 24 of differentiation (predominantly β -cells, some α -cells, and few δ -cells).

2.2.2.10 Hepatocytes

Hepatocytes were generated according to previous reports (Touboul et al. 2010; Hannan, Segeritz, et al. 2013), with minor modifications. Following definitive endoderm differentiation for 3 days, cells were cultured for 5 days in RPMI/B27 medium (Chapter 2.2.2.8) supplemented with 50ng/ml Activin-A to generate anterior foregut cells. Cells were then cultured in Hepatozyme (Gibco) supplemented with 2% non-essential amino-acids, 2% chemically defined concentrated lipids (Gibco), 20mM L-glutamine, 14µg/ml insulin, 15µg/ml transferrin, 100U/ml Penicillin, 100µg/ml Streptomycin, 50ng/ml hepatocyte growth factor (HGF; R&D), and 20ng/ml Oncostatin-M (R&D). Medium changes were performed every other day. After 3 days into these conditions (day 11 of differentiation), hepatic progenitors were obtained. These were further matured under the same culture conditions for 11 more days to generate mature hepatocytes at day 22.

2.2.2.11 Cholangiocytes

Cholangiocyte generation was previously described (Sampaziotis et al. 2015), and performed in collaboration with Dr Fotios Sampaziotis and Miss Mariëlle C.F. Zonneveld. Following definitive endoderm differentiation for 3 days, cells were cultured in RPMI/B27 medium (Chapter 2.2.2.8). This was first supplemented with 50ng/ml Activin-A for 5 days to generate foregut progenitors, then with 10 μ M SB431542 and 50ng/ml BMP4 for 4 days to obtain bipotent hepatoblasts, and finally with 50ng/ml FGF10, 50ng/ml Activin-A, and 3 μ M RA for 4 days to derive cholangiocyte progenitors. Monolayer cultures were then transferred to three-dimensional conditions. For this, cells were incubated with cell dissociation buffer (CDB, Gibco) for 10' at 37°C, mechanically dissociated into small clumps, washed twice with RPMI medium by centrifuging at 200g for 3' at RT, and finally resuspended at a density of 2x10⁵ cells/ml in a mixture of 60% Matrigel and 40% William's E Medium (Gibco) supplemented with 10mM nicotinamide (Sigma-Aldrich), 17mM sodium bicarbonate (Sigma-Aldrich), 0.2mM 2-Phospho-L-ascorbic acid trisodium salt (Sigma-Aldrich), 6.3mM sodium pyruvate (Invitrogen), 14mM glucose (Sigma-Aldrich), 20mM HEPES pH 7.4 (Invitrogen), ITS+ premix (BD Biosciences), 0.1 μ M dexamethasone (R&D Systems), 2mM Glutamax, 100U/ml Penicillin, 100 μ g/ml Streptomycin and 20ng/ml EGF. The cell medium-matrigel solution was distributed in 50 μ l droplets, one for each well of 24-well plates. The droplets were allowed 30' for the gel to solidify, following which 1ml of William's E medium with supplements was added per well. 10 μ M Y27632 was also added for the first 24h only to promote cell survival. Cells were cultured in these conditions for 10 days with medium changes every other day to obtain cholangiocytes.

2.2.2.12 Lung cells

Generation of lung epithelium was recently described (Hannan et al. 2015), and performed in collaboration with Dr Nicholas R.F. Hannan. Following definitive endoderm differentiation for 3 days, cells were cultured in RPMI/B27 medium (Chapter 2.2.2.8) supplemented with 50ng/ml Activin-A for 5 days to generate anterior foregut cells. These cells were cultured in RPMI which was first supplemented with 100ng/ml FGF10 and 1 μ M RA for 5 days to pattern lung endoderm, then with 100ng/ml FGF10 and 50ng/ml HGF for 10 days to generate lung progenitors, and finally with 100ng/ml FGF10 for 15 days to mature lung epithelial cells.

2.2.3 Genetic manipulation of hPSCs

Generation of genetically modified hPSCs was performed following lipofection or nucleofection (Chapters 2.3, 2.4, and 2.5 describe each individual genetic modification in detail).

2.2.3.1 Lipofection

Lipofection was performed as previously described (Vallier, Rugg-Gunn, et al. 2004). hPSCs were seeded feeder-free in 6-well plates and transfected 48h following cell passaging. 4µg of DNA and 10µl of Lipofectamine 2000 were used per well, and lipofection was done in Opti-MEM medium (Gibco) for 20h, all according to manufacturer's instructions. For stable transfections, an appropriate drug was added to the medium from the 5th day post-transfection to select clones that had stably integrated the resistance gene. After 10-14 days of selection, single colonies were picked and clonally expanded. In the case of transient transfections, drug selection was initiated immediately after transfection and maintained for 1-3 days.

2.2.3.2 Nucleofection

hPSCs were pre-treated for 16h with 10µM Y27632 to increase survival (Watanabe et al. 2007). Colonies were dissociated to clumps of 2-8 cells using Accutase (Gibco) for 5' at 37°C, and washed once with KSR medium (Table 2) by centrifuging at 200g for 5'. 2x10⁶ cells were nucleofected with 12µg of DNA in a volume of 100µl by using the Lonza P3 Primary Cell 4D-Nucleofector X Kit and the cycle CA-137 on a Lonza 4D-Nucleofector System, all according to manufacturer's instructions. Nucleofected hPSCs were plated onto a feeder layer of irradiated DR4 MEFs (Puromycin and Neomycin resistant), and cultured in KSR medium supplemented with 4ng/ml FGF2 and 10µM Y27632 (this last for the first 24h only). After 4 days, an appropriate selection drug was added to the medium to select clones that had stably integrated the resistance gene. After 10-14 days of selection, single colonies were picked and clonally expanded in feeder-free conditions.

2.2.4 Quantitative real-time polymerase chain reaction (qPCR)

Cellular RNA was extracted using the GenElute Mammalian Total RNA Miniprep Kit and the On-Column DNase I Digestion Set (both from Sigma-Aldrich) following manufacturer's instructions. 500ng of RNA was used for complementary DNA (cDNA) synthesis in a reaction containing 250ng random primers, 0.5mM deoxynucleotides (dNTPs), 20U RNaseOUT, and 25U of SuperScript II (all from Invitrogen) in a total of 20 μ l, all according to manufacturer's instructions. cDNA was diluted 30-fold, and 5 μ l were used for qPCR using SensiMix SYBR low-ROX (Bioline) and 150nM forward and reverse primers (Sigma-Aldrich; see Table 3 for primer sequences). Samples were run in technical duplicates on 96-well plates on a Stratagene Mx-3005P (Agilent), and results were analysed using the delta-delta cycle threshold ($\Delta\Delta$ Ct) approach using *HMBS/PBGD* or *RPLP0* as housekeeping gene, depending on which proved the most stable across the conditions examined (Livak & Schmittgen 2001). The reference sample used as control to calculate the relative gene expression is indicated in each figure or figure legend. In cases where multiple control samples were used as reference, the average Δ Ct from all controls was used when calculating the $\Delta\Delta$ Ct. Data was expressed as average relative gene expression \pm standard error of the mean (SEM). All primers were designed using PrimerBlast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), and were validated to have a qPCR efficiency >98% and to produce a single PCR product. Perseus software (MaxQuant) was used to generate heatmaps summarizing qPCR data.

Table 3. Primers for gene expression qPCR.

Gene	Species	Primer	Sequence
<i>AAT</i>	Homo Sapiens	FW	CCATTGCTGAAGACCTTAGTGATG
		REV	AGACCCTTTGAAGTCAAGCGACC
<i>ACTA2</i>	Homo Sapiens	FW	GTGTTGCCCTGAAGAGCAT
		REV	GCTGGGACATTGAAAGTCTCA
<i>ACTN2</i>	Homo Sapiens	FW	CAAACCTGACCGGGGAAAAAT
		REV	CTGAATAGCAAAGCGAAGGATGA
<i>AFP</i>	Homo Sapiens	FW	AGAACCTGTCACAAGCTGTG
		REV	GACAGCAAGCTGAGGATGTC
<i>ALBUMIN</i>	Homo Sapiens	FW	CCTTTGGCACAATGAAGTGGGTAACC
		REV	CAGCAGTCAGCCATTTACCATAG
<i>B2M</i>	Homo Sapiens	FW	ATGTCTCGTCCGTGGCCTTAGCT
		REV	CCTGAATCTTTGGAGTACGTGGATAGC
<i>CER1</i>	Homo Sapiens	FW	TTCTCAGGGGGTCATCTTGC
		REV	ATGAACAGACCCGCATTTCC
<i>CITED2</i>	Homo Sapiens	FW	CTGCCGCCAATGTCATAGA
		REV	CTGTTTGACACGAAGTCCG
<i>CDX2</i>	Homo Sapiens	FW	GGGCTCTCGAGAGGCAGGT
		REV	CCTTTGCTCTGCGGTTCTG
<i>CREBBP</i>	Homo Sapiens	FW	ACAAGCGAAACCAACAAACCAT
		REV	ATCTGCTGGTGGGTTGAGG
<i>CYP3A7</i>	Homo Sapiens	FW	ATCCAAGCTATGTCTTCATCAT
		REV	AATCTACTTCCCCAGCACTGA
<i>DPY30</i>	Homo Sapiens	FW	TGCTGGAGGGACAAACGCAGG
		REV	AGGCACGAGTTGCAAGACTGG
<i>DUSP2</i>	Homo Sapiens	FW	GAGGGCCTTTCCGCTACAA
		REV	TTCACCCAGTCAATGAAGCCT
<i>EGFLAM</i>	Homo Sapiens	FW	TGAGACCCAACAGCGACTTC
		REV	CTGACTGGCCATCCCTAACG
<i>EGFP</i>	Synthetic gene	FW	CCCGACAACCACTACCTGAG
		REV	GTCCATGCCGAGAGTGATCC
<i>EGR1</i>	Homo Sapiens	FW	ACGAGCACCTGACCGCAGAGTCTT
		REV	TTGCCACTGTTGGGTGCAGGCT
<i>EOMES</i>	Homo Sapiens	FW	ATCATTACGAAACAGGGCAGGC
		REV	CGGGGTTGGTATTTGTGTAAGG
<i>EP300</i>	Homo Sapiens	FW	GCAGTGTGCCAAACAGATG
		REV	CATAGCCCATAGGCGGGTTG
<i>EPCAM</i>	Homo Sapiens	FW	GTGCTGGTGTGTGAACACTG
		REV	CTCCTTCTGAAGTGCAGTCCG
<i>FOXA2</i>	Homo Sapiens	FW	GGGAGCGGTGAAGATGGA
		REV	TCATGTTGCTCACGGAGGAGTA
<i>FOXH1</i>	Homo Sapiens	FW	GATCGCTTGGTGATTCAG
		REV	TTCCAGCCCTCGTAGTCTTC

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Gene	Species	Primer	Sequence
<i>GATA4</i>	Homo Sapiens	FW	TCCCTCTCCCTCCTCAAAT
		REV	TCAGCGTGTAAGGCATCTG
<i>GBX2</i>	Homo Sapiens	FW	GTTCCACTGCAAAAAGTACCTCT
		REV	GGGACGACGATCTTAGGGTTC
<i>GCG</i>	Homo Sapiens	FW	AAGCATTTACTTTGTGGCTGGATT
		REV	TGATCTGGATTTCTCCTCTGTGTCT
<i>GFAP</i>	Homo Sapiens	FW	ACCTGCAGATTCGAGAAACCA
		REV	CACGGTCTTACCACGATGT
<i>GSC</i>	Homo Sapiens	FW	GAGGAGAAAGTGGAGGCTGGTT
		REV	CTCTGATGAGGACCGCTTCTG
<i>HAND1</i>	Homo Sapiens	FW	GTGCGTCTTTAATCCTCTTC
		REV	GTGAGAGCAAGCGGAAAAG
<i>HEX</i>	Homo Sapiens	FW	CTGCAGCTCAGCGAGAGACA
		REV	CAGGGGAGGGCGAACATGGA
<i>HFN4A</i>	Homo Sapiens	FW	CATGGCCAAGATTGACAACCT
		REV	TTCCCATATGTTCTGCATCAG
<i>HLXB9</i>	Homo Sapiens	FW	CACCGGGGCATGATC
		REV	ACTTCCCCAGGAGGTTCGA
<i>HNF1B</i>	Homo Sapiens	FW	GCACCCCTATGAAGACCCAG
		REV	GGACTGTCTGGTTGAATTGTCTG
<i>HOXA1</i>	Homo Sapiens	FW	CGTGAGAAGGAGGGTCTCTTG
		REV	GTGGGAGGTAGTCAGAGTGTC
<i>ID2</i>	Homo Sapiens	FW	GCCCTGGACTCGCATCCCACTATT
		REV	TCATGAACACCGCTTATTCAGCCACACA
<i>IFITM1</i>	Homo Sapiens	FW	GCAGAAAACCACACTTCTCAAAC
		REV	CCCTAGACTTCACGGAGTAGG
<i>INS</i>	Homo Sapiens	FW	CAGGAGGCGCATCCACA
		REV	AAGAGGCCATCAAGCAGATCA
<i>ISL1</i>	Homo Sapiens	FW	GCAAATGGCAGCGGAGCCCA
		REV	AGCAGGTCCGCAAGGTGTGC
<i>KDR</i>	Homo Sapiens	FW	TTTTTGCCCTTGTCTGTCC
		REV	TCATTGTTCCAGCATTTCA
<i>KLF10</i>	Homo Sapiens	FW	CTGGGTTTTCCCTTCAGCA
		REV	CTTTTCTCCTGTGTGCGTCC
<i>LEFTY1</i>	Homo Sapiens	FW	CGCCAGGAGATGTACATTGA
		REV	CTTGATGCTGACGATCATGG
<i>MAP2</i>	Homo Sapiens	FW	AGACTGCAGCTCTGCCTTTAG
		REV	AGGCTGTAAGTAAATCTTCTCTCC
<i>MEF2C</i>	Homo Sapiens	FW	CAGACATCGTGGAGGCATT
		REV	GGGGTGAGTGCATAAGAGGA
<i>MEIS2</i>	Homo Sapiens	FW	AAGGGGAAGTTGCAGAGC
		REV	AATGCATGGGGGTCCATGT
<i>MESP1</i>	Homo Sapiens	FW	GAAGTGGTTCCTTGGCAGAC
		REV	TCCTGCTTGCCCTCAAAGTGT

Chapter 2: Material and Methods

Gene	Species	Primer	Sequence
<i>METTL14</i>	Homo Sapiens	FW	GGAGGGGTTGGACCTTGGA
		REV	CCCGTCTGTGCTACGCTTCA
<i>METTL3</i>	Homo Sapiens	FW	ATCCAGGCCACAAGAAGCA
		REV	AGGCACTGGGCTGTCACTAC
<i>MHY6</i>	Homo Sapiens	FW	CTCCCGCTTTGGGAAATT
		REV	GGACTTCTCCAGCAGGTAGGT
<i>NAB2</i>	Homo Sapiens	FW	GAGGGCCTTTCCGCTACAA
		REV	TTCACCCAGTCAATGAAGCCT
<i>NANOG</i>	Homo Sapiens	FW	CATGAGTGTGGATCCAGCTTG
		REV	CCTGAATAAGCAGATCCATGG
<i>NEUROG3</i>	Homo Sapiens	FW	GCTCATCGTCTCTATTCTTTGC
		REV	GGTTGAGGCGTCATCCTTTCT
<i>NODAL</i>	Homo Sapiens	FW	TGAGCCAACAAGAGGATCTG
		REV	TGGAAAATCTCAATGGCAAG
<i>OCT4/POU5F1</i>	Homo Sapiens	FW	AGTGAGAGGCAACCTGGAGA
		REV	ACACTCGGACCACATCCTTC
<i>OLIG3</i>	Homo Sapiens	FW	AGCCGTCTCAACTCGGTCT
		REV	CATGGCTAGGTTTCAGTCTG
<i>PAX6</i>	Homo Sapiens	FW	CTTTGCTTGGGAAATCCGAG
		REV	AGCCAGGTTGCGAAGAATC
<i>PBGD/HMBS</i>	Homo Sapiens	FW	GGAGCCATGTCTGGTAACGG
		REV	CCACGCGAATCACTCTCATCT
<i>PCH1</i>	Homo Sapiens	FW	AGGCGAGAAGGCTCTGAGTC
		REV	TACTGAGCTGCATTGGGCTG
<i>PDX1</i>	Homo Sapiens	FW	GATTGGCGTGTGTTGTGGCT
		REV	GCCGGCTTCTCTAAACAGGT
<i>RPLP0</i>	Homo Sapiens	FW	GGCGTCCGTGGAAGTGAC
		REV	GCCTTGCATCATGGTGT
<i>S100B</i>	Homo Sapiens	FW	ATTCTGGAAGGGAGGGAGACA
		REV	GTCCACAACCTCCTGCTCTTT
<i>SETDB1</i>	Homo Sapiens	FW	CATTGCGGCCACTGAAAA
		REV	GCCGTGTAGAGCCTCGATAG
<i>SIP1</i>	Homo Sapiens	FW	CGCTTGACATCACTGAAGGA
		REV	CTTGCCACACTCTGTGCATT
<i>SMAD2</i>	Homo Sapiens	FW	ATGTCGTCCATCTTGCCATTC
		REV	CTCAAGCTCATCTAATCGTCCTG
<i>SOX1</i>	Homo Sapiens	FW + REV	QuantiTect primers (QIAGEN): QT00215299
<i>SOX17</i>	Homo Sapiens	FW	CGCAGGAATTTGAACAGTA
		REV	GGATCAGGACCTGTACAC
<i>SOX18</i>	Homo Sapiens	FW	AGCAGCGGGTCTATTACAG
		REV	CGTATGAGAGAGCAGAGCGG
<i>SOX2</i>	Homo Sapiens	FW	TGGACAGTACGCGCACAT
		REV	CGAGTAGGACATGCTGTAGGT

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Gene	Species	Primer	Sequence
<i>SOX9</i>	Homo Sapiens	FW+ REV	QuantiTect primers (QIAGEN): QT00001498
<i>SST</i>	Homo Sapiens	FW	CCCCAGACTCCGTCAGTTTC
		REV	TCCGCTCTGGTTGGGTTTCAG
<i>SYN1</i>	Homo Sapiens	FW	ACCTCGGGACTCAACACCTCGG
		REV	GAACCACAGGTTGCCGACCCAG
<i>T</i>	Homo Sapiens	FW	TGCTTCCCTGAGACCCAGTT
		REV	GATCACTTCTTTCCTTTCATCAAG
<i>TAGLN</i>	Homo Sapiens	FW	TCTTTGAAGGCAAAGACATGG
		REV	TTATGCTCCTGCGCTTCTT
<i>TBX5</i>	Homo Sapiens	FW	GCTGGAAGGCGGATGTTT
		REV	GATCGTCGGCAGGTACAATG
<i>TBX6</i>	Homo Sapiens	FW	AAGTACCAACCCCGCATACA
		REV	TAGGCTGTACGGAGATGAA
<i>TNNT2</i>	Homo Sapiens	FW	ACAGAGCGAAAAGTGGGAAG
		REV	TCGTTGATCCTGTTTCGGAGA
<i>TRIM2L</i>	Homo Sapiens	FW	AGGAAATGCTACAGAGACTGGG
		REV	TTGAGCAATGCCAAGGTGCC
<i>WTAP</i>	Homo Sapiens	FW	ACCAACGAAGAACCTCTCCCAA
		REV	AAGGATGTTTTCCCTGCGTGC
<i>WNT8A</i>	Homo Sapiens	FW	TCTGCCTGGTCACTGAACAAT
		REV	ACTGGAACCTGCACTCCTCG
<i>Dlx5</i>	Mus Musculus	FW	CTTTCAGCTGGCCGCTTTAC
		REV	ATTTTCACCTGTGTTTGCCTCA
<i>Dpy30</i>	Mus Musculus	FW	ATCCGGTTGATGCGGTTTG
		REV	TGAGGGTTTTCTGCAACCTGT
<i>Eomes</i>	Mus Musculus	FW+REV	QuantiTect primers (QIAGEN): QT01074332
<i>Gapdh</i>	Mus Musculus	FW	CTGCCAGAACATCATCCCT
		REV	ACTTGGCAGGTTTCTCCAGG
<i>Gsc</i>	Mus Musculus	FW	AGAAGGTGGAGGTCTGGTTTA
		REV	TCAGCTGTCCGAGTCCAAT
<i>Hex1</i>	Mus Musculus	FW	CCCAGAACCAGGTCTGAAGTA
		REV	CTTTGCTCGGCGATTTTGAA
<i>Nanog</i>	Mus Musculus	FW	CTGCTCCGCTCCATAACTTC
		REV	GCTTCCAAATTCACCTCAA
<i>Oct4/Pou5f1</i>	Mus Musculus	FW	CACGAGTGAAAGCAACTCA
		REV	TTCATGTCTGGGACTCCTC
<i>Pbgd/Hmbs</i>	Mus Musculus	FW	CATGTATGCTGGGTCCAGG
		REV	AATGCAGCGAAGCAGAGTT
<i>Sox1</i>	Mus Musculus	FW+REV	QuantiTect primers (QIAGEN): QT00289338
<i>Sox17</i>	Mus Musculus	FW	AGATGCTAGGCAAGTCTTGG
		REV	GCTTCATGCGCTTACCT
<i>Sox2</i>	Mus Musculus	FW	ACCAGCTCGCAGACCTACAT
		REV	GACTTGACCACAGAGCCAT

Gene	Species	Primer	Sequence
EGFP	Synthetic gene	FW	CCCGACAACCACTACCTGAG
		REV	GTCCATGCCGAGAGTGATCC
tetR	Synthetic gene	FW	CGACGCCTTAGCCATTGAGA
		REV	TTTCTGTAGGCCGTGTACCT

2.2.5 Western blot (WB)

To prepare total protein lysates, cells were harvested by scraping in phosphate-buffered saline (PBS; Gibco), centrifuged at 250g for 5' at 4°C, and re-suspended in ice-cold CellLytic M buffer (Sigma-Aldrich) supplemented with cOmplete Protease Inhibitor (Roche). After an incubation for 30' at 4°C on a rotating wheel, lysates were clarified by centrifugation at 16,000g for 10' at 4°C. The supernatants were transferred to new tubes, and the protein concentration was assessed using Protein Quantification Kit-Rapid (Sigma-Aldrich). Samples were prepared for Western blot analysis by adding Laemmli buffer (final concentration of 30mM Tris-HCl pH 6.8, 6% glycerol, 2% sodium dodecyl sulphate/SDS, 0.02% bromophenol blue, and 0.25% β -mercaptoethanol), and were denatured at 95°C for 5'. 20 μ g of proteins were loaded and run on 4-12% NuPAGE Bis-Tris Precast Gels (Invitrogen), then transferred to polyvinylidene fluoride (PVDF) membranes by liquid transfer using NuPAGE Transfer buffer (Invitrogen). Membranes were blocked for 1h at RT in PBS 0.05% Tween-20 (PBST) supplemented with 4% non-fat dried milk, and incubated overnight at 4°C with the primary antibody diluted in the same blocking buffer (Table 4). After three washes in PBST, membranes were incubated for 1h at RT with horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in blocking buffer (Table 5), then further washed three times with PBST before being incubated with Pierce ECL2 Western Blotting Substrate (Thermo) and exposed to X-Ray Super RX Films (Fujifilm).

Table 4. Primary antibodies

Protein	Supplier	Product code	WB	IF	Flow cyt.	IP	ChIP	ChIP-seq
ACAN	R&D	AF1220	-	1:50	-	-	-	-
ACTB/ β -actin	Sigma-Aldrich	A22228	1:10000	-	-	-	-	-
ACTN2	Sigma-Aldrich	A7811	-	1:500	-	-	-	-
ALB	Bethyl Lab	A80-229A-3	-	1:100	-	-	-	-
BGLAP	R&D	MAB1419	-	1:25	-	-	-	-
BPTF	Bethyl Lab	A300-973A	-	-	-	-	7ug	-
CREBBP	Cell Signalling	7389	1:1000	-	-	-	-	-
DDR2	Santa Cruz	sc-7555	-	1:50	-	-	-	-
DPY30	Sigma-Aldrich	HPA043761	1:100	1:200	-	5 μ g	6 μ g	-
EOMES	Abcam	ab23345	1:2000	-	-	-	-	-
EP300	Santa Cruz	sc-584	1:1000	-	-	-	-	-
FOXH1 (biotinylated)	R&D	BAF4248	1:250	-	-	-	-	-
GFAP	DAKO	Z0334	-	1:1000	-	-	-	-
GSC	Santa Cruz	sc-133639	1:1000	-	-	-	-	-
H3	Abcam	ab1791	-	-	-	-	3 μ g	-
H3K27me3	Millipore	07-499	-	-	-	-	4 μ g	5 μ g
H3K36me3	Abcam	ab9050	-	-	-	-	5 μ g	-
H3K4me1	Abcam	ab8895	-	-	-	-	3 μ g	-
H3K4me2	Abcam	ab7766	-	-	-	-	3 μ g	-
H3K4me3	Diagenode	C15410003-50	-	-	-	-	3 μ g	4 μ g
HNF4A	Santa Cruz	sc-8987	-	1:100	-	-	-	-
INS	Acris	BM270S	-	1:300	-	-	-	-
KRT19	Abcam	ab7754	-	1:100	-	-	-	-
MAP2	Sigma-Aldrich	M4403	-	1:200	-	-	-	-
MED1	Bethyl Lab	A300-793A	-	-	-	-	7ug	-
MLL1/KMT2A	Bethyl Lab	A300-374A	1:1000	-	-	-	7ug	-
METTL14	Sigma-Aldrich	HPA038002	1:1000	-	-	4ug	-	-
METTL3	Proteintech	15073-1-AP	1:1000	-	-	-	-	-
METTL3	Bethyl Lab	A301-567A	-	-	-	4ug	6ug	-
MLL2/KMT2B	Bethyl Lab	A300-113A	1:1000	-	-	-	7ug	-
NANOG	R&D	AF1997	1:500	1:100	1:100	3 μ g	5 μ g	-
NGFR/p75	Santa Cruz	sc-6188	-	1:100	-	-	-	-
NKX2.5	Santa Cruz	sc-14033	-	1:200	-	-	-	-
O4	Sigma-Aldrich	O7139	-	1:250	-	-	-	-
OCT4/POU5F1	Santa Cruz	sc-5279	1:1000	1:200	-	-	-	-
PAX6	Cambridge Bioscience	PRB-278P-100	1:1000	-	-	-	-	-
RNA Pol II	Abcam	ab817	-	-	-	-	5ug	-
SETD1A	Bethyl Lab	A300-289A	1:2000	-	-	-	7ug	-
SETDB1	Cell Signalling	2196	1:1000	-	-	-	-	-
SFPTC	Santa Cruz	sc-7705	-	1:50	-	-	-	-
SMAD2/3	R&D	AF3797	1:1000	-	-	3 μ g	5 μ g	-

Protein	Supplier	Product code	WB	IF	Flow cyt.	IP	ChIP	ChIP-seq
SMAD2/3 (biotinylated for Co-IP-WB)	Cell Signalling	12470S	1:2000	-	-	-	-	-
Phospho-SMAD2/3 (Ser465/467)	NEB	3108	1:1000	-	-	-	-	-
SNON/SKIL	Santa Cruz	sc-9592x	1:1000	-	-	-	-	-
SOX1	R&D	AF3369	1:1000	1:100	1:100	-	-	-
SOX17	R&D	AF1924	1:1000	1:100	1:100	-	-	-
SOX2	R&D	AF2018	1:1000	-	-	-	-	-
TAGLN	Abcam	ab14106	-	1:1000	-	-	-	-
tetR	Clontech	631131	1:1000	-	-	-	-	-
tetR	Mobitec	TET01	-	1:4000	-	-	-	-
TNTT2	Abcam	ab45932	-	1:500	-	-	-	-
TUBA4A/ α -TUBULIN	Sigma-Aldrich	T6199	1:10000	-	-	-	-	-
TUBB3	Millipore	MAB1637	-	1:1000	-	-	-	-
VIL1/villin	Santa Cruz	sc-58897	-	1:100	-	-	-	-
WDR5	Bethyl Lab	A302-429A	1:500	-	-	-	7ug	-
WT1	Abcam	ab89901	-	1:50	-	-	-	-
WTAP	Bethyl Lab	A301-436A	1:1000	-	-	-	-	-

Table 5. Secondary antibodies

Reagent	Supplier	Product code	Application	Dilution
Anti Goat-HRP	Sigma-Aldrich	A5420	WB	1:10000
Anti Rabbit-HRP	Sigma-Aldrich	A0545	WB	1:10000
Anti Mouse-HRP	Sigma-Aldrich	A2554	WB	1:10000
ImmunoCruz D IP/WB system	Santa Cruz	sc-45041	IP goat; WB goat	1:4000
ImmunoCruz A IP/WB system	Santa Cruz	sc-45038	IP rabbit; WB goat	1:4000
Clean-Blot IP detection reagent	Pierce	21230	IP goat; WB rabbit	1:300
Streptavidin-HRP	Cell Signalling	3999S	IP goat/rabbit; WB SMAD2/3 or FOXH1	1:2000
Alexa Fluor 488 donkey anti-goat	Invitrogen	A11055	IF/flow cytometry	1:1000
Alexa Fluor 488 donkey anti-mouse	Invitrogen	A21202	IF/flow cytometry	1:1000
Alexa Fluor 488 donkey anti-rabbit	Invitrogen	A21206	IF/flow cytometry	1:1000
Alexa Fluor 568 donkey anti-goat	Invitrogen	A11057	IF/flow cytometry	1:1000
Alexa Fluor 568 donkey anti-mouse	Invitrogen	A10037	IF/flow cytometry	1:1000
Alexa Fluor 568 donkey anti-rabbit	Invitrogen	A10042	IF/flow cytometry	1:1000
Alexa Fluor 647 donkey anti-goat	Invitrogen	A21447	IF/flow cytometry	1:1000
Alexa Fluor 647 donkey anti-mouse	Invitrogen	A31571	IF/flow cytometry	1:1000
Alexa Fluor 647 donkey anti-rabbit	Invitrogen	A31573	IF/flow cytometry	1:1000

2.2.6 Immunofluorescence (IF)

Cells were fixed for 20' at 4°C in PBS 4% paraformaldehyde (PFA), rinsed three times with PBS, and blocked and permeabilized for 30' at RT using PBS with 10% donkey serum (Biorad) and 0.1% Triton X-100 (Sigma-Aldrich). Primary antibodies (Table 4) were diluted in PBS 1% donkey serum 0.1% Triton X-100 and incubated overnight at 4°C. This was followed by three washes with PBS and by further incubation with AlexaFluor secondary antibodies (Table 5) for 1h at RT protected from light. Cells were finally washed three times with PBS, and 4',6-Diamidino-2'-phenylindole dihydrochloride (DAPI; Sigma-Aldrich) was added to the first wash to stain nuclei. Images were acquired using a LSM 700 confocal microscope (Leica).

2.2.7 Flow cytometry

Single cell suspensions were prepared by incubation in CDB for 10' at 37° followed by extensive pipetting. Cells were washed twice with PBS and fixed for 20' at 4°C with PBS 4% PFA. After three washes with PBS, cells were first permeabilized for 20' at RT with PBS 0.1% Triton X-100, then blocked for 30' at RT with PBS 10% donkey serum. Primary and secondary antibodies incubations (Table 4 and Table 5) were performed for 1h each at RT in PBS 1% donkey serum 0.1% Triton X-100, and cells were washed three times with this same buffer after each incubation. Flow cytometry was performed using a Cyan ADP flow-cytometer, and at least 10,000 events were recorded.

2.2.8 Immunoprecipitation (IP)

Unless stated otherwise, all steps were performed on ice or at 4°C and ice-cold buffers (Table 6) were used following supplementation with cOmplete Protease Inhibitors (Roche), PhosSTOP Phosphatase Inhibitor Cocktail (Roche), 1mg/ml Leupeptin, 0.2mM dithiothreitol (DTT), 0.2mM phenylmethane sulfonyl fluoride (PMSF), and 10mM sodium butyrate (all from Sigma-Aldrich). Cells were fed with fresh medium for 2h before being washed with PBS, scraped in CDB, and pelleted at 250g for 5'. The cell pellet was incubated in five volumes of ILB for 10' to induce cell swelling. Then, Triton X-100 was added to a final concentration of 0.3% and cells were incubated for 6' to lyse

the plasma membranes. Nuclei were pelleted at 600g for 5', washed once with ten volumes of ILB, and finally re-suspended in two volumes of NLB. The nuclear suspension was transferred to a Dounce homogenizer (Jencons Scientific) and homogenized by performing 70 strokes with a "tight" pestle. The nuclear lysate was first incubated in rotation for 30', after which 125U/ml of benzonase nuclease (Sigma-Aldrich) were added for 45' at RT to digest nucleic acids. The lysate was clarified by ultracentrifugation at 108,000g for 30', and finally pre-cleared before IP. This was done by sequential incubation for 1h with 10 μ g of non-immune class G immunoglobulin (IgG; R&D) and for 2h with 25 μ g of Protein G-Agarose (Roche), both under gentle mixing in rotation. The protein concentration was assessed, and 1mg of protein was used for overnight IP in rotation with the primary antibody (Table 4). IP were further incubated for 1h with 15 μ l of Protein G-Agarose, washed five times with NLB, and resuspended in Laemmli buffer for Western blot analyses.

2.2.9 Chromatin immunoprecipitation (ChIP)

Unless states otherwise, all steps were performed on ice or at 4°C and ice-cold buffers (Table 6) were used following supplementation with 1mg/ml Leupeptin, 0.2mM PMSF, and 10mM sodium butyrate. Cells were fed with fresh medium for 2h before the beginning of the experiment. 2x10⁷ cells were used for transcription factor ChIP (ChIP-XL), and 1.5x10⁶ cells were used for histone ChIP. Cells for ChIP-XL were cross-linked on plates first with protein cross-linkers (10mM dimethyl 3,3'-dithiopropionimidate dihydrochloride and 2.5mM 3,3'-dithiodipropionic acid di-N-hydroxysuccinimide ester; Sigma-Aldrich) for 15' at RT, then with 1% formaldehyde for 15' at RT (Brown et al. 2011). Cells for histone ChIP were cross-linked only with 1% formaldehyde for 15' at RT. Cross-linking was stopped by adding glycine to a final concentration of 0.125M followed by incubation for 10' at RT. Then, the cells were washed with PBS before being scraped off the plates in PBS. The resulting cell suspension was centrifuged at 250g for 5', and the pellet was re-suspended and incubated for 10' in 2ml of CLB to lyse the plasma membranes. Nuclei were pelleted at 600g for 5' and lysed in 1.25ml of SNLB for 10', after which 0.75ml of IDB were added. The chromatin was sonicated using a S-400 Ultrasonic Liquid Processor (Misonix) equipped with a 1-16" dia Ultrasonic Cell Disruptor Microtip Probe (Cole-Palmer). 20 and 12 cycles (15" ON and 45" OFF at 60% amplitude) were performed for ChIP-XL and standard ChIP, respectively. This protocol

generated fragments of 100-400 base pairs (bp). Samples were clarified by centrifugation at 16,000g for 10', and diluted with 3.5ml of IDB.

After pre-clearing with 25µg of non-immune IgG for 1h and 50µl of Protein G-Agarose for 2h, ChIP was performed overnight in rotation using specific antibodies (Table 5) or non-immune IgG as negative control. After incubation for 1h with 30µl of Protein G-Agarose, beads were washed twice with IWB1, once with IWB2, and twice with TE. Precipitated DNA was eluted twice with 150µl of EB for 15' at RT in rotation, and processed as it follows in parallel with 300µl of sonicated chromatin not used for ChIP (input). For ChIP-XL samples, protein-protein cross-linking was first reverted by adding DTT to a concentration of 100µM and by incubating for 30' at 37°C. For all samples, protein-DNA cross-linking was reverted by adding NaCl to a final concentration of 300mM and by incubating the samples at 65°C for 5h. 1µg RNase A (Sigma-Aldrich) was added during this step to digest contaminating RNA. Finally, 60µg of Proteinase K (Sigma-Aldrich) were added overnight at 45°C. DNA was extracted by sequential phenol-chloroform and chloroform extractions using Phase Lock Gel Heavy tubes (5Prime), and precipitated overnight at -80°C in 100mM sodium acetate and 66% ethanol, 50µg of glycogen (Ambion) were added as carrier. After centrifugation at 16,000g for 1h at 4°C, DNA pellets were washed once with ice-cold 70% ethanol, and finally air-dried.

ChIP and input samples were resuspended in 300µl and 700µl, respectively, and 5µl were used for qPCR as described in Chapter 2.2.4 (primer sequences detailed in Table 7). Results of transcription factor ChIP were calculated using the $\Delta\Delta C_t$ approach by using a region in the last exon of *SMAD7* to normalize for background binding, and by normalizing the enrichment to the one of non-immune IgG ChIP. Histone ChIP data were expressed as percentage of input DNA by calculating $(C_{t_{chip}} - C_{t_{input}})^2$ and correcting for the dilutions performed. On top of this, histone ChIP enrichments are presented in certain graphs as normalized to a control condition to facilitate visualization and interpretation of the results.

Table 6. Buffers for IP and ChIP experiments.

Buffer	Protocol	Composition
Isotonic lysis buffer (ILB)	IP	10mM Tris-HCl pH 7.5
		3mM CaCl ₂
		2mM MgCl ₂
		0.32M sucrose
Nuclear lysis buffer (NLB)	IP	50mM Tris-HCl pH 7.5
		100mM NaCl
		50mM KCl
		3mM MgCl ₂
		1mM EDTA
		10% glycerol
		0.1% Tween-20
Cell lysis buffer (CLB)	ChIP	10mM Tris-HCl pH 8
		10mM NaCl
		0.2% NP-40
SDS-nuclear lysis buffer (SNLB)	ChIP	50mM Tris-HCl pH 8
		10mM EDTA
		1% SDS
IP dilution buffer (IDB)	ChIP	20mM Tris-HCl pH 8
		2mM EDTA
		150mM NaCl
		0.01% SDS
		1% Triton X-100
IP wash buffer 1 (IWB1)	ChIP	20mM Tris-HCl pH 8
		2mM EDTA
		50mM NaCl
		0.1% SDS
		1% Triton X-100
IP wash buffer 2 (IWB2)	ChIP	10mM Tris-HCl pH 8
		1mM EDTA
		0.25M LiCl
		1% NP-40
		1% deoxycholic acid
Elution buffer (EB)	ChIP	100mM NaHCO ₃
		1% SDS
Tris-EDTA buffer (TE)	ChIP	10mM Tris-HCl pH 8
		1mM EDTA

Table 7. Primers for ChIP-qPCR.

Gene	Species	Application	Primer	Sequence	Genomic location of amplicon
<i>SMAD7</i>	Homo Sapiens	TF ChIP-qPCR (negative control region)	FW	ACCCTGATAGGAAGAGGGGAAG	Last exon
			REV	TCACACACACTCTTGACAAGTGA	
<i>LEFTY1</i>	Homo Sapiens	TF ChIP-qPCR	FW	AAGCTCACAGCCAGACGAGT	Promoter (SMAD2/3 binding site)
			REV	TCAGGCCTCAACCCACATTAG	
<i>NODAL</i>	Homo Sapiens	TF ChIP-qPCR	FW	CAGGTCGGATCAGATTAGCG	First intron (SMAD2/3 binding site)
			REV	GCCTCAATCCACATTCTGCC	
<i>NANOG</i>	Homo Sapiens	TF ChIP-qPCR	FW	GGTTCTGTTGCTCGGTTTTTC	Promoter (SMAD2/3 binding site)
			REV	ATGAGGCAACCAGCTCAGTC	
<i>OCT4 /POU5F1</i>	Homo Sapiens	TF ChIP-qPCR	FW	AGTCTTTGAGGGGATTGCAGAG	Proximal enhancer (SMAD2/3 binding site)
			REV	GGGCAGCTCTAACCCATAACAA	
<i>EOMES</i>	Homo Sapiens	TF ChIP-qPCR	FW	CTAACCAGGCTGTGCCTTCC	Enhancer (SMAD2/3 binding site)
			REV	TGGTCAGGCGTTTAACCAAG	
<i>GSC</i>	Homo Sapiens	TF ChIP-qPCR	FW	GTGCAGGGCACAGTTCAGAG	Enhancer (SMAD2/3 binding site)
			REV	TATGGGACGCTTTGAATCCC	
<i>SOX17</i>	Homo Sapiens	TF ChIP-qPCR	FW	GCTCCAGGTGATAGCTGTGG	Enhancer (SMAD2/3 binding site)
			REV	ATTTTCACCTGCTCCCCTC	
<i>LEFTY1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	CAGACCAGAGTGGCAGTACG	First intron
			REV	GTCTAAGGGGAGTGAGGTG	
<i>NODAL</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GCATTGTCACTCGCGCTTT	First intron
			REV	AAGTTGCTTATCCCGAGCC	
<i>NANOG</i>	Homo Sapiens	Histone ChIP-qPCR	FW	TCCTCTCCTCTATACTAACATGA	First exon
			REV	CTTTCTTACCAGTCTCCGTGT	
<i>OCT4 /POU5F1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	AGTCTTTGAGGGGATTGCAGAG	Proximal enhancer (SMAD2/3 binding site)
			REV	GGGCAGCTCTAACCCATAACAA	
<i>EOMES</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GTAACCTCCCCTCCTCCCG	First exon
			REV	AAAAGGAGAGGGCTTCTGGG	
<i>GSC</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GAGCTACAGGCAGAGGAAATCGCA	Promoter
			REV	CTGGGCGGGCGCCTAATTG	
<i>SOX17</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GCTTTTCGAGTCTCCCTAACC	Promoter
			REV	CGAGTCCCACGTCCCAGTCCA	
<i>GAPDH</i>	Homo Sapiens	Histone ChIP-qPCR	FW	CCCAGCTACTAGCGGTTTTAC	First exon
			REV	AAAAGAAGATGCGGCTGACTGT	
<i>IFITM1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	AGTGCAGGTCTAGGAGGAGG	First intron
			REV	CACTTTCTGGGAGAGCCTGG	

Gene	Species	Application	Primer	Sequence	Genomic location of amplicon
<i>TRIM2L</i>	Homo Sapiens	Histone ChIP-qPCR	FW	TCCCGTGTCTTAGGTCCCTG	Enhancer
			REV	CTGCTGCTGAGAGCTCAAGG	
<i>CDX2</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GTTTGCAGCAGCCAGAAG	First exon
			REV	GGAGGACTGGAATGGCTACG	
<i>WNT8A</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GAGGCCCCATTGTCTCATC	First exon
			REV	TAGGCCCTCAGTGGGTAAA	
<i>EGFLAM</i>	Homo Sapiens	Histone ChIP-qPCR	FW	AGTGTCTGAAGGGTGCCTG	First intron
			REV	AAATGGCCCAGAGTGAAG	
<i>MYOD1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	CCCTCGCGCCAAAAGATTGA	Last exon
			REV	TTCAGTTCTCCCGCTCTCC	
<i>SOX2</i>	Homo Sapiens	Histone ChIP-qPCR	FW	CCAAAAGCCACCTCCATAC	Promoter
			REV	GGCTGGGGAAAGACACAGAC	
<i>CDH1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GCTCTGAGGAGTGGTGCATT	First intron
			REV	CAAGACCTAGCCCACCGTTC	
<i>PHC1</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GGCATCTGTGGATCAGCTT	First intron
			REV	AACTGCCACAAAACCAGGA	
<i>EPCAM</i>	Homo Sapiens	Histone ChIP-qPCR	FW	CCCCGAAACGGGCATAATA	First intron
			REV	ATTTCCCTACCAAGGCAGC	
<i>SMPDL3B</i>	Homo Sapiens	Histone ChIP-qPCR	FW	GGAAGCCTGTGGTGAACA	First exon
			REV	AGGAAAATCAGCCAGGCGAG	
<i>EOMES</i>	Homo Sapiens	Pol II-MED1 ChIP-qPCR	FW	CAGGTTCACTGAGCTACCAAGA	Transcription start site
			REV	ATAATAAGGCCACCGGAGGAGA	
<i>GSC</i>	Homo Sapiens	Pol II-MED1 ChIP-qPCR	FW	GCGTTCGCTGAACTCAACCC	Transcription start site
			REV	ACCAGGAACTAAATCCAGGATGAC	
<i>SOX17</i>	Homo Sapiens	Pol II-MED1 ChIP-qPCR	FW	GGCCACATCTGTGCAGAAAA	Transcription start site
			REV	CTGTGCGGCTGTGGTCTACA	
<i>GAPDH</i>	Homo Sapiens	Pol II-MED1 ChIP-qPCR	FW	ACGTAGCTCAGGCTCAAGA	Transcription start site
			REV	AGGCTGCGGGCTCAATTTAT	
<i>EOMES</i>	Homo Sapiens	H3K36me3 ChIP-qPCR	FW	AAACTGCGGGAGCCCATCTA	Gene body (first intron)
			REV	ATGTCCACAAGCTGGATTCGT	
<i>GSC</i>	Homo Sapiens	H3K36me3 ChIP-qPCR	FW	CTGTCCGAGTCAAATCGCT	Gene body (last exon)
			REV	TGCGCAACAGGCTGGTTTA	
<i>SOX17</i>	Homo Sapiens	H3K36me3 ChIP-qPCR	FW	GGACCGCACGGAATTTGAAC	Gene body (last exon)
			REV	GGATCAGGGACCTGTCACAC	
<i>GAPDH</i>	Homo Sapiens	H3K36me3 ChIP-qPCR	FW	CCACCACACTGAATCTCCCC	Gene body (last exon)
			REV	CTGTGCCCCAGACCCTAGAA	

2.2.10 Statistical analyses

Statistical analyses were performed using GraphPad Prism 6. The type and number of replicates and the statistical test used are described in the figure legends. The level of significance in all graphs is represented as it follows (p denotes the p-value): $*=p<0.05$, $**=p<0.01$, and $***=p<0.001$. All statistical tests employed were two-tailed. Unless stated otherwise in the figure legends, all data is shown as mean \pm SEM. When representative results are presented, the experiments were reproduced in at least two independent settings.

2.3 Methods specific to chapter 3

2.3.1 Generation of stable knockdown hPSCs

DPY30 knockdown hESCs (DPY30 KD hESCs) were generated by lipofection (Chapter 2.2.3.1) with pLKO.1 vectors expressing validated shRNA (sh; Sigma-Aldrich; clone numbers TRCN0000129317 and TRCN0000131112 for DPY30-sh1 and DPY30-sh2 respectively). A pLKO.1 vector containing a scramble (SCR) shRNA that does not target any known human gene was used as negative control (Sigma-Aldrich; catalogue number SHC002). 1mg/ml Puromycin (Gibco) was used for selection of stably transfected cells. NANOG KD hESCs and matched controls had been obtained using this same protocol as previously described (Teo et al. 2011).

2.3.2 Alkaline phosphatase assay

hESCs were fixed with PBS 4% PFA for 20' at 4°C, washed three times with PBS, and rinsed with AP solution (100mM Tris-HCl pH 9.5 and 100mM NaCl). The AP reaction was performed by incubating cells for 20' at RT with AP solution supplemented with nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl phosphate (BCIP) at a ratio of 33µl of NBT and 16µl BCIP per 5ml AP solution (NBT/BCIP colour development substrate, Promega). After staining, cells were washed three times with PBS and imaged using an Axiovert 200M inverted microscope (Zeiss).

2.3.3 Proliferation curve

hESCs were plated in four replicates for each time point and fixed in PBS 4% PFA for 20' at 4°C at intervals of 24h for four days starting from the day after the split (day 0). After being washed three times with PBS, cells were stained for 10' at RT with a solution of 0.1% crystal violet (Sigma-Aldrich) in deionized water. Plates were then washed three times with deionized water for 10' at RT, air-dried, and stored protected from light. The dye was dissolved in 0.5ml of 10% acetic acid (Sigma-Aldrich), and the absorbance at 590nm was measured using an Envision Multilabel Reader (Perkin

Elmer). The data, which is proportional to the number of cells, was normalized to the average absorbance at day 0.

2.3.4 Apoptosis assay

Apoptosis was measured 12h after the last medium change by using the ApoTox-Glo Triplex Assay from Promega according to manufacturer's instructions. Briefly, cells were first incubated for 30' at 37°C with the viability/cytotoxicity reagent containing the GF-AFC and bis-AAF-R110 substrate, and fluorescence was measured at 400_{ex}/505_{em} (viability) and 485_{ex}/520_{em} (cytotoxicity) using an Envision Multilabel Reader. Then, cells were further incubated for 30' at RT with the Caspase-Glo 3/7 reagent, and Caspase 3 activation was measured by luminescence using a Glomax 96-Microplate luminometer (Promega). Measurements were performed on four replicates and expressed as average Caspase activation/viability.

2.3.5 Teratoma assay

Animal procedures were performed in collaboration with Dr Inmaculada Moreno in accordance with the local committee on animal experimentation at the Centro de Investigación Príncipe Felipe. For each assay, 1x10⁶ hESCs were injected in the right testicle of a 6 to 8 weeks-old severe combined immunodeficiency (SCID) mouse. Three animals were injected in each group. Mice were sacrificed after 12 weeks, and the testicles and tumours were dissected and fixed for 48h in Bouins solution (Sigma-Aldrich). The fixed tissues were then paraffin-embedded and processed according to standard procedures. 5µm sections were stained with haematoxylin/eosin and examined under bright-field using an Axiovert 200M inverted microscope.

2.3.6 Sequential ChIP

Sequential ChIP was performed as previously described (Truax & Greer 2012), with minor modifications. A first round of ChIP-XL was performed as described in Chapter 2.2.9, but elution of the immunoprecipitated material was done for 30' at 37 °C using 75ul of TE buffer supplemented with 1% SDS, fresh 15mM DTT, and cComplete Protease

Inhibitors (Roche). The beads were then washed once with 75ul of IDB and this second eluate was pooled with the first. The total eluate was diluted 10 times in IDB, incubated at 4 °C for 5h, and then subjected to a second round of ChIP. Purified sequential ChIP material was resuspended in 100µl, and 5µl were used for qPCR as described in Chapter 2.2.4 with the exception that KAPA Sybr Fast Low Rox (KAPA Biosystems) was used (as this proved to provide higher sensitivity and reliability in detecting lower amounts of DNA obtained from sequential ChIP). Sequential ChIP enrichment was normalized to that of a control sequential ChIP in which the first ChIP using the antibody of interest (anti-DPY30) was followed by a second round of ChIP using non-immune IgG.

2.3.7 Microarray

Microarrays were performed in collaboration with the Wellcome Trust Sanger Institute microarray facility. 500ng of cellular RNA were amplified and purified using the Illumina TotalPrep-96 RNA Amplification kit (Life Technologies) according to manufacturer's instructions. Three biological replicates for each condition were analysed. Biotin-labelled cDNA was then normalized to a concentration of 150ng/µl, and 750ng were hybridized to Illumina Human-12 v4 BeadChips for 16h (overnight) at 58°C. Following hybridisation, BeadChips were washed and stained with streptavidin-cyanine 3 (GE Healthcare). BeadChips were then scanned using the BeadArray reader, and image data was then processed using Genome Studio software (Illumina). The raw and processed microarray data is available on ArrayExpress (Accession number: E-MTAB-2749).

2.3.8 Microarray data analysis

The primary analysis of microarray results described in this paragraph was performed in collaboration with the Wellcome Trust Sanger Institute microarray facility. Probe summaries for all arrays were obtained from the raw data using the method "Making Probe Summary" in Genome Studio. These values were transformed (variance stabilized) and quantile normalized using the R/Bioconductor package *lumi* (Du et al. 2008). Standard lumi quality control (QC) procedure was applied and no outliers were identified. Differential expression between two conditions was evaluated using the R/Bioconductor package *limma* (Smyth 2004). A linear model fit was applied, and the

top differentially expressed genes were tabulated for each contrast using the method of Benjamini and Hochberg to correct the p-values (Benjamini & Hochberg 1995). Probes that failed to fluoresce above background in both conditions were removed. Differentially expressed probes were selected using a combined cut-off of adjusted p-value (adj.p) <0.05 and absolute fold-change (abs.FC) >1.35 (the recommended threshold for confident measurement of differential expression according to the microarray chip manufacturer).

The analysis of differentially expressed probes during the time-course of Activin/Nodal inhibition was performed using the R/Bioconductor *timecourse* package (Smyth 2004). The Hotelling T^2 score for each probe that fluoresced above background in at least one condition was calculated using the *mb.long* function with all parameters set to their default value. Such score ranked probes according to their differential expression across the time-course. The top 10% probes were selected for complete Euclidean hierarchical clustering (k-means preprocessing; max of 300 clusters) using Perseus software. Standard scores (Z-scores) of the \log_2 normalized expression values across the time-course were calculated and used for this analysis. 19 probe clusters having an Euclidean distance >3.68 were defined, and gene enrichment analysis of genes in selected clusters was performed using Enrichr (E. Y. Chen et al. 2013). The significance of the overlap between genes associated to SMAD2/3 binding sites (Brown et al. 2011) and genes in selected clusters was measured with hypergeometric tests considering the sampling population as the number of expressed genes.

Rank-rank hypergeometric overlap analysis (RRHO, Plaisier et al., 2010) was used to measure the significance of the overlap between genes up- or downregulated in two treatment-control experiments. For each experiment, all probes that significantly fluoresced above background in at least one condition were ranked by the $\log_2\text{FC}$ of normalized expression. This ranking was used as input for the online version of RRHO to calculate Benjamini-corrected hypergeometric probabilities (step size = 150).

Heatmaps were obtained by plotting Z-scores of \log_2 normalized expression values calculated separately within each experiments. The data was subjected to Euclidean hierarchical clustering as described above, but only for rows.

2.3.9 ChIP followed by deep sequencing (ChIP-seq)

For histone mark ChIP-seq, ChIP was performed on three biological replicates per condition as described in Chapter 2.2.9. 7.5µg of fragmented chromatin (corresponding to roughly 1.5×10^6 hESCs) were used per ChIP with anti-H3K4me3 or anti-H3K27me3 antibodies (Table 4). For DPY30 ChIP-sequencing, ChIP-XL was performed on one biological replicate as described in Chapter 2.2.9. 2×10^7 hESCs were used with 5µg of anti-DPY30 antibody (Table 4).

Preparation of sequencing libraries and deep sequencing were performed in collaboration with the Wellcome Trust Sanger Institute next-generation sequencing facility. At the end of the ChIP protocol, fragments between 100bp and 400bp (average length of 200bp) were used to prepare barcoded sequencing libraries using NEBNext Sample Prep Kit1 (NEB) following manufacturer's instructions. 10ng of input material for each condition were also used to prepare libraries to be used as a control during peak calls. Equimolar amounts of each library were pooled, and this multiplexed library was diluted to 8pM before sequencing using an Illumina HiSeq 2000 performing 75bp paired-end reads. The raw and processed ChIP-seq data is publicly available on ArrayExpress (Accession numbers: E-ERAD-191 and E-ERAD-365, respectively for histone mark and DPY30 ChIP-seq).

2.3.10 ChIP-seq data analysis

Analyses described in this section were performed in collaboration with Dr Pedro Madrigal. Reads were mapped to 1000 Genomes Phase II (hs37d5) reference assembly using BWA (Li & Durbin 2009). We kept only reads mapping to autosomal chromosomes or chromosome X (H9 hESCs are female), and with mapping quality (MAPQ) values higher or equal than 20. Peak calling for H3K4me3 ChIP-seq was done versus input controls using PeakRanger 1.16 (Feng et al. 2011). In order to follow the recommendation of control samples having substantially deeper sequencing depth than ChIP-seq samples (Bailey et al. 2013), the inputs were pooled. No substantial differences were found when repeating the analysis without pooling (>99% peaks shared with respect to the pooled control). Reads were extended to 250bp, and the smoothing bandwidth was increased to 200 to accommodate broad histone marks (instead of default mode for narrow peaks produced by transcription factors;

parameters -l 250 -b 200 -q 0.05). Chromosomal locations of significant read-enrichment at false discovery ratio (FDR)<0.05 were kept for further analysis. Peak calling for DPY30 ChIP-seq was done also with PeakRanger (-l 250 -b 150 -q 0.05 -p 0.0000001) using as a control a combined file of the inputs from the SCR shRNA and the DPY30 shRNA conditions. Peak calling for H3K27me3 was done using the CCAT algorithm (Xu et al., 2010) included in the PeakRanger software suite with a sliding window of 1000bp and FDR<0.05 (-win_size 1000 --win_step 100 --min_count 70 --min_score 7 -l 250 -q 0.05).

For visualization purposes, biological replicates of histone mark ChIP-seq were combined. Normalized bedGraph format files were produced for each sample using BEDTools 2.17.0 (Quinlan & Hall 2010). The reads mapped at both DNA strands from 5' to 3' direction were extended to a length of 250bp, and the read-enrichments was normalized by million mapped reads and size of the library. bedGraph files were converted to bigWig using UCSC tool bedGraphToBigWig, and visualized on the Biodalliance genome viewer (Down et al. 2011) to generate figures.

To assess the reproducibility of the histone mark ChIP-seq peaks, we computed the Pearson correlation coefficient (PCC) between pairs of sequenced samples using the tool bigWigCorrelate (UCSC) over an aggregated list of histone mark peaks. All replicates used for further analysis have PCC>0.98 and PCC>0.95, respectively for H3K4me3 and H3K27me3. Three replicates were used for each condition except for H3K27me3 ChIP-seq in NANOG KD and its matched SCR control, where only two replicates were considered because the third showed lower levels of reproducibility (PCC<0.74). The PCC values were computed for the complete genome at base-pair resolution and similar results were obtained. Finally, only peaks reproducible in biological replicates passing PCC analysis were kept. Peaks located equal to or closer than 500bp (H3K4me3) or 5Kb (H3K27me3) were merged.

ChIP-seq peaks were annotated using the R/Bioconductor package *ChIPpeakAnno* (Zhu et al. 2010) using the dataset "hsapiens_gene_ensembl" from Ensembl. The proximal promoter and immediate downstream region were considered respectively 1Kb upstream or downstream of the transcription start site (TSS). Genomic regions enrichment of annotations (GREAT) analysis was performed as previously described (McLean et al. 2010) with all parameters as standard.

Negative Binomial tests implemented in diffReps (Shen et al. 2013) were used to detect regions with differential histone modification. For this, we used a sliding window of 600bp (H3K4me3) or 5,000bp (H3K27me3), $p < 1E-6$, sharp peaks mode for H3K4me3 (--nsb 20) and broad peak mode for H3K27me3 (--nsb 2), hg19 as reference genome, and an average fragment size of 250bp (rest of parameters default). Regions with differential histone modification that did not overlap with previously detected chromatin marks peaks were removed. Genomic regions were ranked by their chromatin mark FC, and reported as differential only if both the $abs.FC > 1.5$ and Benjamini-Hochberg corrected $p < 1E-3$.

2.3.11 Integration of ChIP-seq datasets

The analyses described in this section were performed in collaboration with Dr Pedro Madrigal. The significance of the overlap between H3K4me3 peaks decreased after 2h of SB431542 (SB) treatment and SMAD2/3 binding sites was calculated using the genomic association test (GAT; Heger et al. 2013). For this, the annotations were the SMAD2/3 binding sites (Brown et al. 2011), the segments were the H3K4me3 regions, and the workspace was the hg19 genome after subtraction of the hg19/GRCh37 ENCODE Excludable Mappability Regions (Bernstein et al., 2012). To enable the comparison, SMAD2/3 binding site coordinates were translated from hg18 to hg19 using UCSC liftOver under default parameters. 10,000 iteration of the test were performed, as recommended (Heger et al. 2013).

The overlap between different histone marks was evaluated using the function intersect (with parameter -u) from the BEDTools suite (Quinlan & Hall 2010). Bed format files for H3K4me3 and H3K27me3 peaks were obtained from the reproducible peak calls described above, while data for H3K4me1 and H3K27me3 in H1 hESCs from the ENCODE project (Bernstein et al., 2012) was downloaded from the UCSC browser.

The significance of the multiple overlaps between H3K4me3 peaks decreased after 2h of SB treatment, DPY30 KD, or NANOG KD was calculated using MULTOVL (Aszódi 2012). We pooled the total set of H3K4me3 regions for all conditions considered as reported by the diffReps analysis described above, and considered them as the free space in the tool

multovlprob included in MULTOVL. Then, we randomly reshuffled 10,000 times downregulated H3K4me3 peaks to estimate a null distribution of overlap length. P-values were estimated from the Z-scores as previously described (Aszódi 2012).

Quantitative changes of H3K4me3 and H3K27me3 levels on selected genomic regions were obtained by computing a score from the normalized coverage as if follows. For each region $i=1,\dots,L$, where L is total number of regions considered, we obtained the score S_i from the normalized coverage (bigwig data, as described above) $y_{i,j} \in [x_{i,a},x_{i,b}]$ for a replicate $j=1,2,3$ as:

$$S_{i,j} = \frac{\sum_{w=x_{i,a}}^{x_{i,b}} y_{i,j}(w)}{x_{i,b} - x_{i,a}}$$

where $x_{i,a}$ is the start location of the peak region, and $x_{i,b}$ the end. Scores were calculated for ChIP-seq in the presence of Activin or after 2h of SB treatment, and significant differences in their mean values were assessed by Welch's t-test for two-samples. This analysis was performed on different sets of genomic regions described in the text and figure legends.

Average meta-region plots and coverage heatmaps for NANOG, SMAD2/3 and DPY30 ChIP-seq were done using the genomation toolkit (Akalin et al. 2015). 1,000 bins were used for each window, and a winsorize configuration of (0,99) was set to limit the values to only the 99th percentile for a matrix (everything above the 99th percentile was equalized to the value of the 99th percentile). Signal track for NANOG binding in H1 hESCs was obtained from the ENCODE project (Bernstein et al. 2012). Genome-wide coverage for SMAD2/3 was calculated by combining the alignments in BED format in GEO series GSE19461 (Brown et al. 2011) by using the function 'genomecov' in the BEDTools suite (Quinlan & Hall 2010), and then converting to bigwig format. To allow inter-dataset comparison, 'bwtool lift' (Pohl & Beato 2014) was used to lift hg18 to hg19 coordinates using the liftOver chain file hg18ToHg19.over.chain.gz downloaded from UCSC. The subset of SMAD2/3 binding sites nearby regions where H3K4me3 was Activin/Nodal-dependent was determined by overlapping SMAD2/3 peaks with regions obtained extending by ± 100 Kb the H3K4me3 peaks downregulated after SB treatment.

2.3.12 Integration of ChIP-seq and microarray data

Each H3K4me3 peak was mapped to its closest gene using the `annotatePeaks.pl` function from the HOMER suite (Heinz et al. 2010) using standard parameters and the UCSC hg19 annotation. The 491 H3K4me3 peaks downregulated after 2h of SB mapped to 415 unique genes. The difference in expression of microarray probes mapping to these genes was statistically evaluated by performing Welch's paired t-tests for two-samples. In the case of the SB treatment time-course, Friedman's non-parametric paired test was performed to evaluate the significance of overall changes, and Dunn's corrected multiple comparisons were performed for selected pairs of conditions.

For Gene Set Enrichment Analysis (GSEA; Subramanian et al., 2005), we generated a custom gene set file containing genes significantly downregulated after 48h of SB treatment ($\text{adj.p} < 0.05$ and $\text{FC} < -1.35$, calculated with *limma* as described in Chapter 2.3.8). Then, we tested the enrichment of this gene set in the list of genes closest to H3K4me3 peaks and ranked by the $\log_2\text{FC}$ in Activin versus SB. In cases where multiple H3K4me3 peaks mapped to the same gene, the average $\log_2\text{FC}$ was considered.

2.3.13 Generation of knockout mice

All animal procedures were performed in agreement with the UK Home Office regulations and UK Animals Scientific Procedures Act of 1996. Mice carrying the *Dpy30* "knockout first" *tm1a(KOMP)Wtsi* allele were generated by the Sanger Institute Mouse Genetics Project as previously described (Skarnes et al. 2011; White et al. 2013). Briefly, C57BL/6N strain mESCs were electroporated with a linearized "knockout-first" plasmid carrying an internal ribosome entry site and β -galactosidase cDNA (IRES:LacZ) trapping cassette, and a floxed promoter-driven Neomycin resistance cassette (Neo) inserted between two homology arms mapping to exons 3 and 4 of *Dpy30* (Figure 3.42). Cells were selected for integration using Geneticin (Invitrogen), expanded, and screened with a panel of tests to confirm site-specific homologous recombination (loxP site confirmation qPCR; lacZ confirmation qPCR; loss of wild-type allele copy number qPCR; Valenzuela et al., 2003). After screening, targeted mESCs were microinjected in C57BL/6 blastocysts to obtain chimeras. Germline transmission was detected by Neo gene copy number qPCR, and additional quality control tests were performed to confirm

the structure and targeting of the allele (Ryder et al. 2013). Heterozygotes were crossed with C57BL/6NTac mice, and the colony was expanded before intercrossing.

2.3.14 Phenotyping of knockout mice

The experiments described in this paragraph were performed in collaboration with Dr Antonella Galli. The viability of homozygous *Dpy30* knockout mice was tested by crossing heterozygous animals and genotyping the offspring at postnatal day 14 (P14) and embryonic days encompassing 6.5 to 14.5 (E6.5-14.5). In the latter case, embryos were dissected from the decidua of pregnant females and all extraembryonic tissues were removed. Images were acquired using a Leica MZ16A microscope. Genotype was evaluated both by Neo gene copy number qPCR (TaqMan custom probe, Life Technologies, catalogue number: 4400294-314310o) and loss of wild-type allele copy number qPCR (TaqMan custom probe, Life Technologies, catalogue number: 4400294-210263u). A *tfr* TaqMan probe (Life Technologies, catalogue number: 4458366) was used as endogenous control. For E9.5, E14.5 and P14 embryos, DNA was extracted from earclips using the TaqMan Sample-to-SNP kit (Life Technologies) according to manufacturer's instructions.

For E6.5 and E7.5 embryos, DNA and RNA were simultaneously extracted from whole embryos using the AllPrep DNA/RNA Micro Kit (QIAGEN) according to manufacturer's instructions. The DNA was eluted in 25µl and used for genotyping qPCR. The RNA was eluted in 15µl and used for cDNA synthesis and qPCR to measure gene expression as described in Chapter 2.2.4 (primer sequences are in Table 7). Gene expression was measured with the $\Delta\Delta C_t$ approach using the geometric mean of *HMBS/PBGD* and *GAPDH* Ct values as housekeeping reference.

2.4 Methods specific to chapter 4

2.4.1 Molecular cloning

Unless otherwise indicated, traditional cloning was performed using restriction enzymes from NEB and T4 DNA ligase from Promega, and Gibson cloning was achieved using Gibson Assembly Master Mix (NEB). Polymerase chain reaction (PCR) was performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB), DNA blunting was done with DNA Polymerase I Large (Klenow) Fragment (NEB), and site-directed mutagenesis was achieved using QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Whenever possible, vectors were dephosphorylated before ligation using Calf Intestinal Alkaline Phosphatase (CIP) from NEB. All oligonucleotides were ordered from Sigma-Aldrich as desalted lyophilized products. QIAEX II Gel Extraction Kit and QIAquick PCR Purification Kits (QIAGEN) were used for DNA extraction from agarose gels and for purification of PCR products, respectively. Recombinant plasmids were transformed into DH5 α strain *E. Coli* (Alpha-Select Gold Efficiency, Bioline). QIAGEN plasmid mini, midi, and maxi kits were used for plasmid preparations. All these procedures were performed according to manufacturer's instructions. Additional basic molecular biology procedures (such as DNA electrophoresis and *E. Coli* culture) were performed according to standard protocols (Green & Sambrook 2012). All the plasmids were sequence verified by Sanger sequencing through Beckman Coulter Genomics. The plasmid maps for the OPTiKD and sOPTiKD systems are reported in Appendix II.

2.4.1.1 pTP6Neo_tetR

The wild-type bacterial tetR sequence containing a 5'-terminus SV40 nuclear localization sequence (tetR-nls) was removed from the pCAGTetRnls vector (a gift from Peter Andrews, Addgene plasmid #26599) using XhoI and NotI, blunt-ended, and ligated to a pTP6Neo expression plasmid (Pratt et al. 2000; Vallier, Rugg-Gunn, et al. 2004) that had been cut with EcoRI and blunt ended.

2.4.1.2 ROSA26 CRISPR/Cas9n vectors

These plasmids were generated in collaboration with Dr Matthias Pawlowski. pSpCas9n(BB) (a gift of Feng Zhang, Addgene plasmid #48873) was used to generate two plasmids each expressing a gRNA and the Cas9n (*S. Pyogenes* Cas9 D10A nickase mutant). Left and right gRNAs specific for the human ROSA26 locus (THUMPDS3-AS1 gene on chromosome 3) were designed using the online CRISPR Design Tool from the Zhang lab (MIT, Boston, MA; <http://tools.genome-engineering.org>): gRNA-L 5'-GTCGAGTCGCTTCTCGATTA(TGG)-3'; gRNA-R 5'-GGCGATGACGAGATCACGCG(AGG)-3' (protospacer adjacent motif/PAM sites are reported in parenthesis). The score of the gRNA pair was 97 (high quality), with no predicted off target effects. The gRNAs were cloned exactly as previously described (Ran et al. 2013) using the following single stranded oligonucleotide pairs: gRNA-L-top 5'-CACCGTCGAGTCGCTTCTCGATTA-3'; gRNA-L-bot: 5'-AACTAATCGAGAAGCGACTCGAC-3'; gRNA-R-top 5'-CACCGGCGATGACGAGATCACGCG-3'; gRNA-R-bot 5'-AAACCGCGTGATCTCGTCATCGCC-3'. The combination of the two resulting plasmids (pSpCas9n(BB)_R26-L and pSpCas9n(BB)_R26-R, R26 indicating the ROSA26 locus) is predicted to induce a specific double strand break in the intron between exons 1 and 2 of THUMPDS3-AS1.

2.4.1.3 pR26_ENDO-EGFP

The targeting vector for the human ROSA26 locus was constructed in collaboration with Dr Matthias Pawlowski starting from a pUC plasmid. First, upstream (5', 904bp) and downstream (3', 869bp) homology arms (HARs) were generated by PCR from H9 hESC genomic DNA using primers that inserted restriction sites on the 5' and 3' ends of each amplicon (MfeI and KpnI for the 5'-HAR; Sall and HindIII for the 3'-HAR): 5'-HAR_fw 5'-GACTCAATTGGCTCGAAACCGGACGGAGCCATTGCTC-3'; 5-HAR_rev 5'-GCATGGTACCGATCACGCGAGGAGGAAAGGAGGGAGG-3'; 3'-HAR_fw 5'-GACTGTGCGACGCTTCGATTATGGGCGGGATTCTTTTGC-3'; 3'HAR_rev 5'-GCATAAGCTTGGAAGCTATCACACAGGCATCTGAGATCAG-3'. These amplicons were sequentially inserted into the multiple cloning site of pUC19 by restriction digestion. First, the 5'-HAR was ligated into the EcoRI and KpnI sites. Secondly, the 3'-HAR was ligated into the Sall and HindIII sites. The resulting vector was termed pUC_R26. Next, a promoterless gene-trap vector was constructed by inserting a bicistronic cassette consisting of splice acceptor-Puromycin resistance (Puromycin N-acetyltransferase)-self-cleaving-T2A peptide-

enhanced green fluorescent protein-polyadenylation sequence (SA-Puro-T2A-EGFP-pA). This transgene is placed under control of the endogenous ROSA26 promoter (THUMP3-AS1 gene) following correct genomic integration. For this cloning step, a Gibson assembly was performed in which three inserts were cloned into the KpnI and Sall sites of pUC_R26. The first insert encoded for the adenoviral SA, which was first synthesized *de novo* and cloned into an in house shuttle vector. The Gibson assembly fragment was produced from this vector by PCR amplification using primer fw 5'-TTTCCTCCTCGCGTGATCGGTACCTAGGGCGCAGTAGTCCAGG-3' and primer rev 5'-CTCGGTCATGGTGGCCGGTCCGGGATTCTCCTC-3'. The second insert encoded for the Puro, which was PCR amplified from pTRE-TIGHT-EGFP (a gift from Rudolf Jaenisch, Addgene plasmid #22074) using primer fw 5'-TCCCGGACCGGCCACCATGACCGAGT-ACAAGCCCACGGTG-3' and primer rev 5'-CTCCACTGCCCTTAAGGGCACCGGGCTTGGGGT-3'. The third insert encoded the T2A-EGFP-pA cassette, which was PCR amplified from pSpCas9n(BB)-2A-GFP (a gift from Feng Zhang, Addgene plasmid #48140) using primer fw 5'-GCCCGGTGCC-TTAAGGGCAGTGGAGAGGGCAGA-3' and primer rev 5'-CCGCCATAATCGAGAAGC-GTCGACCCCCAGCATGCCTGCTAT-3'. The resulting vector was termed pR26_ENDO-EGFP (ENDO indicates the use of the ROSA26 endogenous promoter).

2.4.1.4 pR26_EF1 α -EGFP

This plasmid was generated in collaboration with Dr Matthias Pawlowski. A gene-trap vector was constructed in which expression of EGFP was under the control of an exogenous elongation factor 1 α (EF1 α) promoter. For this, a Gibson assembly was performed in which three inserts were cloned into the BglII and SacI sites of pR26_ENDO-EGFP (thus substituting the Puro-T2A-EGFP sequence). The first insert encoded for the Neomycin resistance gene (Neomycin phosphotransferase II, Neo) and was PCR amplified using primer fw 5'-CTTTCCAGTTTCGAACGGGAGATCTGCCA-CCATGGGATCGGCCATTGAA-3' and primer rev 5'-CGGAGCCAATCCATAGAGCCCACCGCAT-3'. The second insert encoded the EF1 α promoter and was PCR amplified from pLVX-EtO (a gift from Oliver Brüstle; Ladewig et al. 2012) using primer fw 5'-GGCTCTATGGATTGGCTCCGGTGCCCGT-3' and primer rev 5'-ATGGTGGCGGCGGATCCGGGTCGAAATTCC-3'. The third insert encoded the EGFP and was PCR amplified from pTRE-TIGHT-EGFP using primer fw 5'-ACCCGGATCCGCCGCCACCATGGTGAGCA-3' and

primer rev 5'-TCGAGGCTGATCAGCGAGCTACGCGTTTATCTAGACTTGTACAGCTCGTC-CATGCCG-3'. The resulting vector was termed pR26_EF1 α -EGFP.

2.4.1.5 pR26_CAG-EGFP

This plasmid was generated in collaboration with Dr Matthias Pawlowski. A gene-trap vector was constructed in which expression of EGFP was under control of an exogenous CMV early enhancer, chicken β -actin, and rabbit β -globin hybrid promoter (CAG). For this, the EF1 α promoter in pR26_EF1 α -EGFP was removed with SacI and BamHI and exchanged with the CAG promoter from an in house vector (pAAV-Neo_CAG) cut by SacI digestion. Both the backbone and insert were blunt-ended. The resulting vector was termed pR26_CAG-EGFP.

2.4.1.6 pR26_CAG-STDtetR

The tetR-nls sequence (which we named standard tetR, or STDtetR) was amplified by PCR from pTP6Neo_tetR using primers fw 5'-CATTTTGGCAAAGAATTAATTCGGATCC-ACCATGCCAAAAAGAAGAGGAAGGTATC-3' and rev 5'-CGAGGCTGATCAGCGAGCTACGCGTTTACCGCGGAGACCCACTTTCAC-3'. The resulting product was cloned by Gibson assembly into pR26_CAG-EGFP following removal of the EGFP by BamHI and MluI digestion.

2.4.1.7 pR26_CAG-OPTtetR

The STDtetR cDNA sequence was used as template for gene synthesis following multi-parameter codon and RNA optimization for expression in human cells (Fath et al. 2011). For this, we used the GeneArt GeneOptimizer service from Invitrogen. The resulting synthetic gene (which we named codon-optimized tetR, or OPTtetR; Figure 4.16) was amplified by PCR using primer fw 5'-CATTTTGGCAAAGAATTAATTCGGATCCACCATGCCCAAAGAAAAAGCGG-3' and rev 5'-CGAGGCTGATCAGCGAGCTACGCGTTCATCTGGGGGAGCCGC-3'). The resulting product was cloned by Gibson assembly into pR26_CAG-EGFP following removal of the EGFP by BamHI and MluI digestion.

2.4.1.8 AAVS1 ZFNs

pZFN_AAVS1-L-ELD and pZFN_AAVS1-R-KKR were a generous gift of Dr Kosuke Yusa (Wellcome Trust Sanger Institute, Hinxton, UK). Previously described left and right ZFN amino acid sequences for the human AAVS1 locus (PPR1R12C gene on chromosome 19; Hockemeyer et al. 2009) were used as basis for artificial gene synthesis following codon optimization for expression in human cells (with insertion of a 5' EcoRI site and a 3' XhoI site). Following digestion with these two enzymes, the ZFNs were cloned into pVAX1 (Invitrogen) using EcoRI and XhoI, thus generating pZFN_AAVS1-L and pZFN_AAVS1-R. These plasmids were then modified to generate obligate heterodimer ZFNs by introducing mutated FokI nuclease domains (Doyon et al. 2011). The amino acid sequences of FokI-ELD (Q486E, I499L, and N496D) and FokI-KKR (E490K, I538K, and H537R) domains were used as basis for artificial gene synthesis following codon optimization for mammalian expression (with insertion of a 5' BamHI site and a 3' XhoI site). Following digestion with these two enzymes, the mutated FokI domains were cloned in ZFN plasmids in which the wild-type FokI domains had been removed by BamHI and XhoI digestion. FokI-ELD was used for AAVS1 left ZFN (pZFN_AAVS1-L-ELD), and FokI-KKR was used for AAVS1 right ZFN (pZFN_AAVS1-R-KKR). The combination of these two plasmids is predicted to induce a specific double strand break between exons 1 and 2 of PPR1R12C.

2.4.1.9 pAAV_iKD and related shRNA-containing plasmids

AAVS1 SA-2A-Puro-pA donor (a gift from Rudolf Jaenisch, Addgene plasmid # 22075; Hockemeyer et al. 2009) was used to construct these plasmids. First, the only BglII restriction site was removed by site-directed mutagenesis using top primer 5'-CACAGGGCCTCGAGAGTTCTGGCAGCGGAGAGG-3' and bottom primer 5'-CCTCTCCGCTGCCAGAACTCTCGAGGCCCTGTG-3'. This step was required as BglII had to be subsequently used to clone the shRNAs. Secondly, the H1-TO promoter (tetracycline-inducible H1 Pol III promoter carrying a tet operon after the TATA box) from pSUPERIOR_Neo (Oligoengine) was inserted in the HincII site following EcoRI-KpnI digestion and blunt-ending. The resulting plasmid was named pAAV_iKD (for inducible knockdown). Finally, shRNAs were cloned to generate the final targeting vectors. For this last step, pAAV_iKD was digested with 25U/ μ g of both BglII and SalI-HF in 1x NEB3.1 at 37°C for 3h, gel extracted following DNA electrophoresis, and

reconstituted at 50ng/ μ l for subsequent ligations. Validated shRNA sequences were obtained either from published reports (*OCT4/POU5F1* and *B2M* shRNAs; Zafarana et al., 2009), from the RNAi Consortium TRC library (*DPY30* shRNA: clone TRCN0000131112; Moffat et al., 2006), or from Sigma-Aldrich (SCR and EGFP shRNA: vectors SHC002 and SHC005). A single guanine was added at the 5' end of the EGFP shRNA, as transcription from the H1 promoter is more effective for sequences starting with guanine or adenosine (the original shRNA started with a thymine). All other sequences were used without any modification. Complementary single-stand oligonucleotides containing the shRNA, the Pol III terminator sequence, and appropriate overhangs (Table 8) were mixed at a 50 μ M concentration of each oligo in a 20 μ l reaction containing 10mM Tris-HCl pH 8, 1mM ethylenediaminetetraacetic acid (EDTA), and 100mM NaCl (annealing buffer). Oligos were annealed in a thermocycler following the protocol: (1) 94°C for 5'; (2) 93°C for 20", reduce by 1°C/cycle and repeat for 12 cycles; (3) 80°C for 4'; (4) 79°C for 20", reduce by 1°C/cycle and repeat for 3 cycles; (5) 75°C for 4'; (6) 74°C for 20", reduce by 1°C/cycle and repeat for 3 cycles; (7) 70°C for 4'; (8) 69°C for 20", reduce by 1°C/cycle and repeat for 60 cycles; (9) hold at 10°C. Annealed oligos were diluted 1:500 in annealing buffer, and 4 μ l were used for ligation with 50ng of pAAV_iKD (prepared by restriction digestion as described above) for 2h at RT. Following transformation, bacterial colonies were screened by colony PCR using primers fw 5'-GTGTCATTCTATTCTGGGGGGTG-3' and rev 5'-GTGGGGGTTAGAC-CCAATATCAG-3' in a 12.5 μ l mix containing 400nM of each primer, 400 μ M dNTPs, 3mM MgCl₂, 1x NH₄ Reaction buffer, and 0.625U of BIOTAQ DNA Polymerase (Bioline). For this, the following PCR program was used: (1) 95°C for 5'; (2) 95°C for 30"; (3) 60°C for 30"; (4) 72°C for 1'; (5) repeat steps 2 to 4 for 34 times; (6) hold at 10°C. Positive clones carried a band of approximately 500bp (depending on the size of the shRNA) instead of the 464bp fragment from the parental pAAV_iKD vector. The shRNA sequence was confirmed by Sanger sequencing using the same primers employed for colony PCR.

Table 8. Oligonucleotides for cloning of shRNAs (Chapter 3).

Gene	Top oligo ^a	Bottom Oligo ^a
<i>OCT4</i>	GATCCCGGATGTGGTCCGAGTGTGGT <u>TTCAAGAGA</u> CCACACTCGGACCACATCC <u>TTTTTTG</u>	TCGACAAAAAAGGATGTGGTCCGAGTGTGG CTCTTGA ACCACACTCGGACCACATCCGG
<i>B2M</i>	GATCCCGGACTGGTCTTTCTATCTCT <u>TTCAAGAGA</u> GAGATAGAAAGACCAGTCC <u>TTTTTTG</u>	TCGACAAAAAAGGACTGGTCTTTCTATCTC CTCTTGA AGATAGAAAGACCAGTCCGG
EGFP	GATCCCGTACAACAGCCACAACGCTATCT <u>CTCGAG</u> ATAGACGTGTGGCTGTGTAT <u>TTTTTTG</u>	TCGACAAAAAATACAACAGCCACAACGCTAT CTCGAG ATAGACGTGTGGCTGTGTACGG
<i>DPY30</i>	GATCCCGTCTCACAGACAACGTTGAGACT <u>CTCGAG</u> TCTCAACGTTGTCTGTGAGACT <u>TTTTTTG</u>	TCGACAAAAAAGTCTCACAGACAACGTTGAGA CTCGAG TCTCAACGTTGTCTGTGAGACGG

^a Sense and anti-sense shRNA strands are underlined; the hairpin loop is in bold; the Pol III terminator is in italic; and the remaining sequences provided the necessary overhangs for cloning into pAAV_iKD or pAAV_siKD. Note that the 5' BglII site is purposely lost after cloning of the shRNA to facilitate screening of recombinant bacteria by restriction digestion.

2.4.1.10 pAAV_iKD-2TO and pAAV_iKD-2TO-EGFP

The H1-2TO promoter was designed by replacing the 19bp sequence before the TATA box of the H1-TO promoter with an additional tet operon (see Figure 4.18), as previously described (Henriksen et al. 2007). This sequence was obtained as double stranded synthetic DNA (Integrated DNA Technologies), with the additional introduction of up- and downstream homology regions to allow subsequent Gibson cloning (5'- GTGGGCTCTATGGGTC-3' and 5'-CCTAGTAAAGCTTAGTACTGTC-3', respectively). The resulting product was Gibson cloned in the AAVS1 SA-2A-Puro-pA donor lacking the BglII site (described above for pAAV_iKD) following HincII digestion. This pAAV_iKD-2TO vector was then used to clone the EGFP shRNA exactly as just described for pAAV_iKD, hence generating the plasmid pAAV_iKD-2TO-EGFP.

2.4.1.11 pAAV_CAG-EGFP

AAV-CAGGS-EGFP (a gift from Rudolf Jaenisch, Addgene plasmid # 22212) was renamed pAAV_CAG-EGFP to provide consistent nomenclature with the rest of the plasmids used in the study.

2.4.1.12 pAAV_CAG-EGFP_H1-TOEGFPsh

A fragment containing the CAG-EGFP-pA cassette was cut from pR26_CAG-EGFP with SpeI and HincII, blunt-ended, and inserted in pAAV_iKD-EGFP following digestion with BstBI and blunt ending.

2.4.1.13 pAAV_EGFPiKD-TO and pAAV_EGFPiKD-2TO

A fragment containing the CAG-EGFP-pA cassette was cut from pR26_CAG-EGFP with SpeI and HincII, blunt-ended, and inserted in pAAV_iKD-EGFP or pAAV_iKD-2TO-EGFP following digestion with PspXI and blunt ending. This generated pAAV_EGFPiKD-TO and pAAV_EGFPiKD-2TO, respectively.

2.4.1.14 pAAV_EGFPsiKD-TO-STD and pAAV_EGFPsiKD-2TO-STD

A fragment containing the CAG-STDtetR-pA cassette was cut from pR26_CAG-STDtetR with SpeI and HincII, blunt-ended, and inserted in pAAV_iKD-EGFP or pAAV_iKD-2TO-EGFP following digestion with PspXI and blunt ending. This generated pAAV_EGFPsiKD-TO-STD and pAAV_EGFPsiKD-2TO-STD, respectively.

2.4.1.15 pAAV_EGFPsiKD-TO-OPT and pAAV_EGFPsiKD-2TO-OPT

A fragment containing the CAG-OPTtetR-pA cassette was cut from pR26_CAG-OPTtetR with SpeI and HincII, blunt-ended, and inserted in pAAV_iKD-EGFP or pAAV_iKD-2TO-EGFP following digestion with PspXI and blunt ending. This generated pAAV_EGFPsiKD-TO-OPT and pAAV_EGFPsiKD-2TO-OPT, respectively.

2.4.1.16 pAAV_siKD and related shRNA-containing plasmids

A fragment containing the CAG-OPTtetR-pA cassette was cut from pR26_CAG-OPTtetR with SpeI and HincII, blunt-ended, and inserted in pAAV_iKD following digestion with PspXI and blunt ending. This generated the pAAV_siKD vector. Such plasmid was used to clone shRNAs as described in Chapter 2.4.1.9 for pAAV_iKD. However, colony PCR screening of the recombinant bacteria was performed using primers fw 5'-CGAACGCTGACGTCATCAACC-3' and rev 5'-GGGCTATGAACTAATGACCCCG-3' with a PCR annealing temperature of 60°C. Positive clones generated a band of 350bp (instead of the 295bp fragment in the parental pAAV_siKD vector).

2.4.2 Gene targeting

ROSA26 locus targeting was performed in collaboration with Dr Matthias Pawlowski. hPSCs were nucleofected with a total of 12 μ g of DNA (equally divided between the two CRISPR/Cas9n plasmids and the targeting vector) as described in Chapter 2.2.3.2. Targeted hPSCs carrying Neo were selected with 50 μ g/ml of Geneticin (G418 Sulfate, Gibco).

AAVS1 locus targeting was performed by lipofection with 4 μ g of DNA (equally divided between the two AAVS1 ZFN plasmids and the targeting vector) as described in Chapter 2.2.3.1. Targeted hPSCs carrying Puro were selected with 1mg/ml of Puromycin.

hPSC clones from gene targeting experiments were screened by genomic PCR to verify site-specific targeting, determine the number of alleles targeted, and exclude off-target integrations of the targeting plasmid (Table 9 and Figure 4.2). For this, the following PCR were performed: Locus PCR (PCR of wild-type AAVS1 locus, indicating a non-targeted allele); Locus PCR/loss-of-allele PCR (PCR of targeted allele/PCR that fails if the transgene contains the GC-rich CAG promoter); 5'-INT/3'-INT PCR (PCR of transgene 5'-end/3'-end genomic integration region, indicative of expected transgene targeting); 5'-BB/3'-BB PCR (PCR of vector backbone 5'-end/3'-end genomic integration region, indicative of non-specific off-target plasmid integration). All PCRs were performed using 100ng of genomic DNA as template in a 10 μ l reaction using LongAmp Taq DNA Polymerase (NEB) according to manufacturer's instructions and including 2% dimethyl sulfoxide (DMSO). DNA was extracted using the Wizard SV Genomic DNA purification system (Promega). Only correctly targeted lines with no off-target integrations were used for the experiments presented, unless otherwise specifically indicated. Homologous recombination of the targeting vectors was confirmed by Sanger sequencing of the genomic PCRs of 5' and 3' vector integration sites. Karyotype analyses of targeted lines were performed by standard G banding techniques (Medical Genetics Service, Cambridge University Hospitals).

Table 9. Summary of gene targeting experiments

Donor plasmid	Tot	Incorr. target ^a	Rand Int. ^b	Het + extra ^c	Homo + extra ^c	Het	Hom	Correct target (%) ^d	Total target (%) ^e
pR26_ENDO-EGFP	12	0	0	9	0	3	0	25	100
pR26_EF1α-EGFP	12	0	1	8	0	3	0	25	92
pR26_CAG-EGFP	11 11 ^f	3 1 ^f	0 0 ^f	2 2 ^f	0 1 ^f	6 5 ^f	0 2 ^f	54 64 ^f	73 91 ^f
pR26_CAG-STDtetR	28	6	1	8	3	8	2	36	75
pR26_CAG-OPTtetR	28 27 ^f	3 11 ^f	0 0 ^f	12 9 ^f	3 2 ^f	10 4 ^f	0 1 ^f	36 19 ^f	89 59 ^f
pAAV_CAG-EGFP	8 4 8 ^g	0 0 0 ^g	0 0 1 ^g	0 0 0 ^g	3 1 2 ^g	0 0 0 ^g	5 3 5 ^g	62 75 62 ^g	100 100 87 ^g
pAAV_EGFPiKD-TO	8 4 8 ^g	0 0 0 ^g	0 0 0 ^g	0 0 0 ^g	6 2 5 ^g	0 0 0 ^g	2 2 3 ^g	25 50 37 ^g	100 100 100 ^g
pAAV_EGFPiKD-2TO	8 4 8 ^g	0 0 0 ^g	0 0 0 ^g	0 0 0 ^g	5 1 6 ^g	1 1 0 ^g	2 2 2 ^g	37 75 25 ^g	100 100 100 ^g
pAAV_iKD-OCT4	6	0	0	1	2	0	3	50	100
pAAV_iKD-B2M	6	0	1	0	4	0	1	17	83
pAAV_iKD-DPY30	6	0	1	0	4	0	1	17	83
pAAV_EGFPsiKD-TO-STD	6	0	0	0	3	0	3	50	100
pAAV_EGFPsiKD-TO-OPT	6	0	0	0	3	0	3	50	100
pAAV_EGFPsiKD-2TO-STD	6	0	1	0	2	1	2	50	83
pAAV_EGFPsiKD-2TO-OPT	6	0	0	0	2	0	4	67	100
pAAV_siKD-OCT4	6 6 ^h	0 0 ^h	0 0 ^h	1 5 ^h	3 0 ^h	1 1 ^h	1 0 ^h	33 17 ^h	100 100 ^h
pAAV_siKD-B2M	6 6 ^h	0 0 ^h	0 0 ^h	4 4 ^h	2 0 ^h	0 2 ^h	0 0 ^h	0 33 ^h	100 100 ^h

^a Evidence of targeting, but incorrect size of 5'- or 3'- genomic integration PCRs.

^b No evidence of targeting (lack of bands in 5'- and 3'- genomic integration PCRs and presence of WT band in locus PCR).

^c Targeting with additional random integration of the plasmid (bands in 5'- and/or 3'-vector backbone PCRs).

^d Heterozygous (Het) + homozygous (homo) targeting only.

^e Including clones with additional random integration of the plasmid (het + extra and homo + extra).

^f The two figures are from two different targeting experiments in hESCs.

^g The three figures are from targeting in three different hESC lines: ROSA26 HOMO STDtetR, ROSA26 HET OPTtetR, and ROSA26 HOMO OPTtetR, respectively.

^h The first figure is from hESC targeting; the second figure is from hiPSC targeting.

Table 10. Genotyping strategies for gene targeting of the ROSA26 locus

Targeting vector(s)	PCR type	Primer location	Primer sequence	Amplicon wild-type	Amplicon transgene	Amplicon plasmid ^a	Temp. ann. ^b	Ext. time ^b
All pR26 vectors	LOCUS	Genomic; 5' to 5'-HAR	GAGAAGAGGCTGTGCTTCGG	2186	Variable ^c (can be loA)	No band	63	>3 ^c
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
All pR26 vectors	5'-INT	Genomic; 5' to 5'-HAR	GAGAAGAGGCTGTGCTTCGG	No band	1264	No band	60	1'30"
		Splice acceptor	AAGACCCGGAAGAGTTGTCC					
pR26_ENDO-EGFP pR26_EF1 α -EGFP pR26_CAG-EGFP	3'-INT	EGFP	CTACCCCGACCACATGAAGC	No band	1746	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
pR26_CAG-STDtetR	3'-INT	STDtetR	CGACGCCTTAGCCATTGAGA	No band	1706	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
pR26_CAG-OPTtetR	3'-INT	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	1548	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
All pR26 vectors	5'-BB	Backbone; 5' to 5'-HAR	CGTTGTAAAACGACGGCCAG	No band	No band	1148	60	1'30"
		Neo	GTGCCAGTCATAGCCGAAT					
pR26_ENDO-EGFP pR26_EF1 α -EGFP pR26_CAG-EGFP	3'-BB	EGFP	ACCACTACCAGCAGAACAC	No band	No band	1315	60	1'30"
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					
pR26_CAG-STDtetR	3'BB	STDtetR	CCCGTAAACTCGCCAGAAG	No band	No band	1681	60	2'
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					
pR26_CAG-OPTtetR	3'BB	OPTtetR	AGAAACTGGGCGTGAACAG	No band	No band	1736	60	2'
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					

^a Result of PCR on targeting vector (positive control for off-target plasmid integration).

^b Variable parameters in PCR program: (1) 94° 5'; (2) 94° 15"; (3) Temp. ann. 30"; (4) 65° Ext. time; (5) Repeat 2 to 4 for 34 cycles; (6) 65° 5'; (7) hold at 10°

^c Size depending on transgene size; the extension time was set accordingly. For CAG promoter-containing transgenes, the PCR fails due to high GC-content and results in loss-of-allele (loA).

Table 11. Genotyping strategies for gene targeting of the AAVS1 locus

Targeting vector(s)	PCR type	Primer location	Primer sequence	Amplicon wild-type	Amplicon transgene	Amplicon plasmid ^a	Temp. ann. ^b	Ext. time ^b
All pAAV vectors	LOCUS	Genomic; 5' to 5'-HAR	CTGTTCCCTTCCAGGACAGGTCC	1692	Variable ^c (can be loA)	No band	65	> 3'
		Genomic; 3' to 3'-HAR	TGCAGGGAAACGGGGCTCAGTCTGA					
All pAAV vectors	5'-INT	Genomic; 5' to 5'-HAR	CTGTTCCCTTCCAGGACAGGTCC	No band	1103	No band	65	1' 30"
		Puro	TCGTCGGGTGGGAGGGCCACCG					
pAAV_EGFP pAAV_EGFPiKD-TO pAAV_EGFPiKD-2TO	3'-INT	EGFP	GGATCACTCTCGGCATGGAC	No band	1656	No band	60	2'
		Genomic; 3' to 3'-HAR	TGCAGGGAAACGGGGCTCAGTCTGA					
All pAAV_iKD vectors	3'-INT	After H1 promoter	GCCGGACAGTACTAAGCTTTACTAGGG	No band	884	No band	60	1'
		Genomic; 3' to 3'-HAR	TGCAGGGAAACGGGGCTCAGTCTGA					
All pAAV_siKD vectors pAAV_EGFPsiKD-TO-OPT pAAV_EGFPsiKD-2TO-OPT	3'-INT	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	1447	No band	60	1' 30"
		Genomic; 3' to 3'-HAR	TGCAGGGAAACGGGGCTCAGTCTGA					
pAAV_EGFPsiKD-TO-STD pAAV_EGFPsiKD-2TO-STD	3'-INT	STDtetR	CGACGCTTAGCCATGAGA	No band	1605	No band	60	2'
		Genomic; 3' to 3'-HAR	TGCAGGGAAACGGGGCTCAGTCTGA					
All pAAV vectors	5'-BB	Backbone; 5' to 5'-HAR	ATGCTTCCGGCTCGTATGTT	No band	No band	1227	60	1' 30"
		Puro	TGAGGAAAGAGTCTTGCAGCTC					
pAAV_EGFP pAAV_EGFPiKD-TO pAAV_EGFPiKD-2TO	3'-BB	EGFP	GGATCACTCTCGGCATGGAC	No band	No band	2011 (EGFP) 1526 (EGFP iKD)	60	2' 30"
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAAGTT					
All pAAV_iKD vectors	3'-BB	After H1 promoter	GCCGGACAGTACTAAGCTTTACTAGGG	No band	No band	1239	60	1' 30"
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAAGTT					
All pAAV_siKD vectors pAAV_EGFPsiKD-TO-OPT pAAV_EGFPsiKD-2TO-OPT	3'-BB	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	No band	1802	60	2'
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAAGTT					
pAAV_EGFPsiKD-TO-STD pAAV_EGFPsiKD-2TO-STD	3'-BB	STDtetR	CGACGCTTAGCCATGAGA	No band	No band	1960	60	2'
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAAGTT					

^{a-c} See previous page

2.4.3 Inducible gene knockdown

Gene knockdown was induced by adding tetracycline hydrochloride (Sigma-Aldrich) to the culture medium. Unless otherwise described in the results or in the figure legend, tetracycline was used at the concentration of 1µg/ml. Medium containing tetracycline was kept light protected, and changed every 24-48h, depending on the cell type and growth conditions. Knockdown rescue was done by omitting tetracycline from the culture medium. Cells were additionally washed three times with PBS for 2' at RT to remove traces of tetracycline.

For DPY30 inducible knockdown during hESC differentiation, knockdown was induced at three stages: (1) in hESCs three days before the start of differentiation, in order to induce DPY30 knockdown specifically from the induction stage; (2) in differentiating cells three days in advance of reaching the state of maturing progenitors, in order to induce knockdown specifically from the specification stage; (3) in maturing cells seven days in advance of the final analysis of mature cells, in order to induce knockdown specifically during cell maturation. Refer to Chapter 2.2.2 for details on the differentiation protocols used.

2.4.4 EGFP flow cytometry

Cells were incubated with TrypLE Select (Gibco) for 5-20' at 37° C to obtain a single cell suspension. Following a wash in PBS, cells were resuspended in ice-cold PBS 1% BSA with 5µg/ml DAPI, incubated for 5' on ice, and filtered through a 100µm cell strainer. Cells were analysed using a Cyan ADP flow-cytometer to determine the EGFP median fluorescence intensity (MFI) of viable cells (DAPI negative). Flow cytometry analysis was performed using FlowJo.

2.5 Methods specific to chapter 5

2.5.1 SMAD2/3 IP for mass spectrometry (co-IP2 protocol)

These experiments were performed in collaboration with Dr Sasha Mendjan. 2×10^7 cells were used for each IP. Unless stated otherwise, all steps were performed on ice or at 4°C and ice-cold buffers (Table 12) were supplemented with cOmplete Protease Inhibitors (Roche), PhosSTOP Phosphatase Inhibitor Cocktail (Roche), 1mg/ml Leupeptin, 0.2mM DTT, 0.2mM PMSF, and 10mM sodium butyrate (all from Sigma-Aldrich).

Table 12. Buffers for SMAD2/3 IP and mass spectrometry

Buffer name	Protocol	Composition
Hypotonic lysis buffer (HLB)	co-IP2	10mM HEPES pH 7.6 10mM KCl 2mM MgCl ₂ 0.2mM EDTA 0.2mM EGTA
High-salt nuclear lysis buffer (HSNLB)	co-IP2	20mM HEPES pH 7.6 420mM NaCl 2mM MgCl ₂ 25% glycerol 0.2mM EDTA 0.2mM EGTA
Dialysis buffer (DB)	co-IP2	20mM HEPES pH 7.6 50mM KCl 100mM NaCl 2mM MgCl ₂ 10% glycerol 0.2mM EDTA 0.2mM EGTA
Solvent A	Mass spectrometry	0.5% acetic Acid
Solvent B	Mass spectrometry	0.5% acetic Acid 80% acetonitrile

Cells were fed with fresh medium for 2h before being washed with PBS, scraped in CDB, and pelleted at 250g for 10'. The cell pellet was then washed once with 10 volumes of PBS and once with 10 volumes of HLB. The pellet was re-suspended in 5 volumes of HLB and incubated for 5' to induce cell swelling. The resulting cell suspension was homogenized using the “loose” pestle of a Dounce homogenizer (Jencons Scientific) for 35-50 strokes until plasma membrane lysis was complete (as judged by microscopic inspection). The nuclei were pelleted at 800g for 5', washed once with 10 volumes of HLB, and resuspended in 1.5 volumes of HSNLB. High-salt nuclear extraction was performed by homogenizing the nuclei using the “tight” pestle of a Dounce homogenizer

for 70 strokes, followed by 45' of incubation in rotation. The resulting lysate was clarified for 30' at 16,000g and transferred to a dialysis cassette using a 19-gauge syringe. Dialysis was performed for 4h in 1l of DB under gentle stirring, and the buffer was changed once after 2h. After the dialysis, the sample was clarified from minor protein precipitates for 10' at 17,000g, and the protein concentration was assessed. Immunoprecipitations were performed by incubating 0.5mg of protein with 5µg of SMAD2/3 antibody (Table 4) or IgG negative control antibody for 3h at 4°C in rotation. This was followed by incubation with 10µl of Protein G-Agarose for 1h. Beads were finally washed three times with DB and processed for mass spectrometry as described below.

Note that the co-IP1 protocol was described in Chapter 2.2.8.

2.5.2 Preparation of samples for mass spectrometry

Label-free quantitative mass spectrometric analysis was performed on triplicate immunoprecipitations for each condition. After immunoprecipitation, samples were prepared as described in (Hubner & Mann 2011), with minor modifications. Proteins were eluted by incubation with 50µl of 2M urea and 10mM DTT for 30' at RT in agitation. Then, 55mM chloroacetamide was added for 20' to alkylate reduced disulphide bonds. Proteins were pre-digested on the beads with 0.4µg of mass spectrometry-quality trypsin (Promega) for 1h at RT in agitation. The suspension was cleared from the beads by centrifugation, the beads were washed with 50ul of 2M Urea, and the merged supernatants were incubated overnight at RT in agitation to complete digestion. 0.1% trifluoroacetic acid was then added to inactivate trypsin, and peptides were loaded on C₁₈ StageTips (Rappsilber et al. 2007). Tips were prepared for binding by sequential equilibration for 2' at 800g with 50µl methanol, 50µl Solvent B, and 50µl Solvent A (Table 12). Subsequently, peptides were loaded, and washed twice with Solvent A. Tips were dry-stored until analysis.

Dimethyl labelling for quantitative mass spectrometric analysis was performed in collaboration with Dr Nina C. Hubner as previously reported (Boersema et al. 2009; Hubner et al. 2015). Protein samples were eluted and digested with trypsin as described above, then incubated with dimethyl labelling reagents (4µl of 0.6M NaBH₃CN together

with 4µl of 4% CH₂O or CD₂O for light or heavy labelling, respectively) for 1h at RT with shaking. The reaction was stopped by adding 16µl of 1% NH₃. Samples were acidified with 0.1% trifluoroacetic acid, and finally loaded on stage-tips as described above. Each immunoprecipitation was performed twice, switching the labels.

2.5.3 Mass spectrometry

The mass spectrometry and primary data analysis described in this paragraph was performed in collaboration with Dr Nina C. Hubner as previously described (Hubner et al. 2010). Peptides were eluted from the StageTips and separated by reversed phase liquid chromatography on a 2.5h long segmented gradient using EASY-nLC 1000 (Thermo Fisher Scientific). Eluting peptides were ionized and injected directly into a Q Exactive mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in a TOP10 sequencing mode, meaning that one full mass spectrometry (MS) scan was followed by higher energy collision induced dissociation (HCD) and subsequent detection of the fragmentation spectra of the 10 most abundant peptide ions (tandem mass spectrometry; MS/MS). The raw data was analysed using the MaxQuant software suite (Cox & Mann 2008). For dimethyl samples, an additional mass of 28.03Da (light) or 32.06Da (heavy) was specified as “labels” at the N-terminus and at lysines. Peptide spectra were searched against the human database (Uniprot) using the integrated Andromeda search engine, and peptides were identified with an FDR<0.01 determined by false matches against a reverse decoy database. Peptides were assembled into protein groups with an FDR<0.01. For label-free quantitative mass spectrometry, protein quantification was performed using the MaxQuant label-free quantification algorithm requiring at least 2 ratio counts, in order to obtain label free quantification (LFQ) intensities.

Statistical analysis of the data was performed using the Perseus software package (MaxQuant). First, common contaminants and reverse hits were removed, and only proteins identified by at least two peptides (one of those being unique to the respective protein group) were considered as high-confidence identifications. Proteins were then filtered for having been identified in all replicates of at least one condition. For label-free samples, LFQ intensities were logarithmized, and missing intensity values were imputed by representing noise values (Hubner et al., 2010). T-tests were then

performed to determine the specific interactors in each condition by comparing the immunoprecipitations with the SMAD2/3 antibody against the IgG negative controls. Statistical significance was set with a $FDR < 0.05$. For dimethyl experiments, the outlier significance was calculated based on protein intensity (Significance B; Cox and Mann, 2008) and was required to be < 0.05 for both the forward and the reverse experiment.

2.5.4 Biological interpretation of mass spectrometry data.

Functional enrichment analysis was performed using Enrichr (E. Y. Chen et al. 2013), and only enriched terms with an $adj.p < 0.05$ were considered. For Gene Ontology (GO) enrichment analysis, the 2015 GO annotation was used. For mouse phenotype enrichment analysis, the level 3 of the Mouse Genomic Informatics (MGI) annotation was used.

The SMAD2/3 protein-protein interaction network was generated using Cytoscape v2.8.3 (Shannon et al. 2003). First, all the annotated interactions involving the SMAD2/3 binding proteins were inferred by interrogating protein-protein interaction databases through the PSIQUIC Universal Web Service Client. IMEx-complying interactions were retained and merged by union. Then, a subnetwork involving only the SMAD2/3 interactors was isolated. Finally, duplicate nodes and self-loops were removed to simplify visualization.

2.5.5 Generation of inducible knockdown hPSCs

Cloning of inducible shRNA vectors was performed essentially as described in Chapter 2.4.1.9, but with the following modifications aimed at facilitating the generation of multiple plasmids in parallel. First, annealing, dilution, and ligation of oligonucleotides (Table 13) were all done in 96-well PCR plates. Bacterial transformations were also performed in 96-well plates, and 10 μ l of DH5 α *E. Coli* (Alpha-Select Gold Efficiency, Bionline) were transformed with 1 μ l of the ligation product according to manufacturer's instructions (with the exception that heat shock was done for 15" at 42°C in a water bath). Bacterial suspensions were then plated onto 25-compartments 100mm petri dishes (Thermo Scientific) containing Luria Bertani (LB) agar medium supplemented with

100µg/ml Ampicillin. For each construct, 80% and 20% of the transformed bacteria were plated in two compartments, and glass beads were used to facilitate spreading. Following overnight incubation at 37°C, 2 to 10 bacterial clones per construct (an average of 4 to 5) could be distinguished in at least one of the wells. Whenever possible, 4 clones were isolated and tested by colony PCR in a 96-well plate format as described earlier (Chapter 2.4.1.9). >50% of the clones demonstrated shRNA integration, and whenever possible 2 clones were sequence-verified. Plasmid minipreps were directly used for subsequent gene targeting. All shRNA sequences were obtained from the RNAi Consortium TRC library (<https://www.broadinstitute.org/rnai/public/>, Moffat et al., 2006). Whenever shRNAs had been validated, the most powerful ones were chosen.

Inducible knockdown hESCs were obtained essentially as described in Chapters 2.2.3.1 and 2.4.2, but with the following modifications aimed at facilitating the generation of multiple lines in parallel. First, lipofection was performed in 12-well plates, with reagent volumes scaled accordingly (2µg of total DNA and 5µl of Lipofectamine 2000 per well). For each shRNA, two wells were transfected in order to generate independent biological replicates. Following drug selection, all the resulting targeted cells in each well were pooled and expanded for further analysis. Given that 20 to 50 clones were obtained for each well, we referred to these lines as “clonal pools”. hESCs expressing a randomly integrated wild-type tetR (clone E, Figure 4.1) were used for the generation of SMAD2, FOXH1, SETDB1, EP300, and CREBBP inducible knockdown hESCs. hESCs expressing the OPTtetR from the ROSA26 locus (OPTiKD platform, Chapter 4.3.3) were used for the generation METTL3, METTL14, and WTAP inducible knockdown hESCs.

Table 13. Oligonucleotides for cloning of shRNAs (Chapter 5).

Gene	sh	Top oligo ^a	Bottom Oligo ^a
<i>SMAD2</i>	1	GATCCCGCGATTAGATGAGCTTGAGAAA CTCGAG TTTCTCAAGCTCATCTAATCGTTTTTTG	TCGACAAAAAACGATTAGATGAGCTTGAGAAA CTCGAG TTTCTCAAGCTCATCTAATCGCGG
	2	GATCCCGCAAGTACTCCTTGCTGGATTG CTCGAG CAATCCAGCAAGGAGTACTGTTTTTTG	TCGACAAAAACAAGTACTCCTTGCTGGATTG CTCGAG CAATCCAGCAAGGAGTACTGCGG
	3	GATCCCGCATGATCCAGTATCACAGTAT CTCGAG ATACTGTGATACTGGATCATGTTTTTTG	TCGACAAAAACATGATCCAGTATCACAGTAT CTCGAG ATACTGTGATACTGGATCATGCGG
<i>FOXH1</i>	1	GATCCCGCACCTCCTACTTGCCATCTACT CTCGAG TAGATAGGCAAGTAGGAGGTGTTTTTTG	TCGACAAAAACACCTCCTACTTGCCATCTA CTCGAG TAGATAGGCAAGTAGGAGGTGCGG
	2	GATCCCGCAGGGCTGGAAGACTCCAT CTCGAG ATGGAGTCTTCCAGCCCTCGTTTTTTG	TCGACAAAAACGAGGGCTGGAAGACTCCAT CTCGAG ATGGAGTCTTCCAGCCCTCGCGG
	3	GATCCCGCCTATCTACACTCCCAATGT CTCGAG ACATTGGGAGTGTAGTAGGCTTTTTTTG	TCGACAAAAAGCCTATCTACACTCCCAATGT CTCGAG ACATTGGGAGTGTAGTAGGCTGCGG
<i>SETDB1</i>	1	GATCCCGCTCAGATGATAACTTCTGTACT CTCGAG TACAGAAGTTATCATCTGAGCTTTTTTTG	TCGACAAAAAGCTCAGATGATAACTTCTGTA CTCGAG TACAGAAGTTATCATCTGAGCGG
	2	GATCCCGATTAGAGACATGGGTAATAC CTCGAG GTATTACCATGTCTCTAACTTTTTTTG	TCGACAAAAAGTTAGAGACATGGGTAATAC CTCGAG GTATTACCATGTCTCTAACTGG
	3	GATCCCGCTGACTTCTAGAGGATAT CTCGAG ATACTCCTCTATGAAGTCACGTTTTTTG	TCGACAAAAAGTACTGACTTCTAGAGGATAT CTCGAG ATACTCCTCTATGAAGTCACGCGG
<i>EP300</i>	1	GATCCCGCCTTCCAAATCCGAGACAT CTCGAG ATGTCGGAATTGTGAAGGCTTTTTTTG	TCGACAAAAAGCCTTCCAAATCCGAGACAT CTCGAG ATGTCGGAATTGTGAAGGCGG
	2	GATCCCATCTACCGCGAGGATATTT CTCGAG AAATATCCTCCGGCTGAGTATTTTTTTG	TCGACAAAAAATACTACCGCGAGGATATTT CTCGAG AAATATCCTCCGGCTGAGTATGG
	3	GATCCCGCCCGGTGAACCTCTCTATAAT CTCGAG ATTATAGGAGAGTTCACCGGTTTTTTG	TCGACAAAAACCCGGTGAACCTCTCTATAAT CTCGAG ATTATAGGAGAGTTCACCGGCGG
<i>CREBBP</i>	1	GATCCCGGATGAATATATCACTAT CTCGAG ATAAGTGATAATATTCATCCCTTTTTTTG	TCGACAAAAAGGATGAATATATCACTAT CTCGAG ATAAGTGATAATATTCATCCCGG
	2	GATCCCGGTTTACATAAACAAGGCAT CTCGAG ATGCCTGTTTATGTAACGCTTTTTTTG	TCGACAAAAAGCGTTTACATAAACAAGGCAT CTCGAG ATGCCTGTTTATGTAACGCGG
	3	GATCCCGCGTTTACCATGAGATCCTTACT CTCGAG TAAGGATCTCATGGTAAACGGTTTTTTG	TCGACAAAAACCGTTTACCATGAGATCCTTA CTCGAG TAAGGATCTCATGGTAAACGCGG
<i>METTL3</i>	1	GATCCCGCTCAGTATCTTGGGCAAGTT CTCGAG AACTTGCCCAAGATACTGACGTTTTTTG	TCGACAAAAACGTGAGTATCTTGGGCAAGTT CTCGAG AACTTGCCCAAGATACTGACGCGG
	2	GATCCCGCCTTAACATTGCCACTGAT CTCGAG ATCAGTGGCAATGTTAAGGCTTTTTTTG	TCGACAAAAAGCCTTAACATTGCCACTGAT CTCGAG ATCAGTGGCAATGTTAAGGCGG
<i>METTL14</i>	1	GATCCCGAAGACGCTTCATCTATTTG CTCGAG CAAATAGATGAAGGCGTCTTCTTTTTTTG	TCGACAAAAAGAAGACGCTTCATCTATTTG CTCGAG CAAATAGATGAAGGCGTCTTCGG
	2	GATCCCGAGGATGAGTTAATAGCTAAAT CTCGAG ATTTAGCTATTAACCTCATCCTTTTTTTG	TCGACAAAAAAGGATGAGTTAATAGCTAAAT CTCGAG ATTTAGCTATTAACCTCATCCTCGG
<i>WTAP</i>	1	GATCCCGCAAGAGTGTACTACTCAAAT CTCGAG ATTTGAGTAGTACACTCTTGCTTTTTTTG	TCGACAAAAACGAAGAGTGTACTACTCAAAT CTCGAG ATTTGAGTAGTACACTCTTGCGG
	2	GATCCCATGGCAAGAGATGAGTTAAT CTCGAG AATTAACCTCATCTCTTGCCATTTTTTTG	TCGACAAAAAATGGCAAGAGATGAGTTAAT CTCGAG AATTAACCTCATCTCTTGCCATGG

^a Sense and anti-sense shRNA strands are underlined; the hairpin loop is in bold; Pol III terminator is in italic; the remaining sequences provided the necessary overhangs for cloning into pAAV_iKD.

2.5.6 RNA sequencing (RNA-seq)

RNA-seq libraries were prepared in collaboration with Ms Stephanie Brown. Polyadenylated (poly-A) purified opposing strand-specific mRNA library libraries were prepared from 200ng of total RNA using the TruSeq Stranded mRNA HT sample preparation kit (Illumina). Samples were individually indexed for pooling using a dual-index strategy. Libraries were quantified both with a Qubit (ThermoFisher Scientific) and by qPCR using the NGS Library Quantification Kit (KAPA Biosystems). Libraries were then normalized and pooled.

Deep sequencing was performed in collaboration with Cambridge Genomics Services. Pooled libraries were diluted and denatured for sequencing on the NextSeq 500 (Illumina) according to the manufacturer's instructions. Samples were pooled so as to obtain >25M unique clusters per sample (18 samples were split across 2 lanes). The PhiX control library (Illumina) was spiked into the main library pool at 1% vol/vol for quality control purposes. Sequencing was performed using a high output flow cell with 2x75 cycles of sequencing, which provided 800M paired end reads from 400M unique clusters from each lane. Overall, a total of ~80M paired end reads per sample were obtained.

2.5.7 RNA-seq data analysis

Primary RNA-seq data analysis described in this paragraph was performed in collaboration with Miss Anna Osnato. Reads were trimmed using Sickle (Joshi & Fass 2011) with q (quality threshold)=20 and l (length threshold)=30. To prepare for reads alignment, the human transcriptome was built with TopHat2 v2.1.0 (Kim et al. 2013) based on Bowtie v2.2.6 (Langmead & Salzberg 2012) by using the human GRCh38.p6 as reference genome, and the Ensembl gene transfer format (GTF) as annotation (http://ftp.ensembl.org/pub/release-83/gtf/homo_sapiens/). All analyses were performed using this transcriptome assembly. Alignment was performed using TopHat2 with standard parameters. Using Samtools view (Li et al. 2009), reads with MAPQ>10 were kept for further analyses. Subsequent quantitative data analysis was performed using SeqMonk (Andrews 2014). The RNA-seq pipeline was used to calculate transcript abundance as reads per million mapped reads (RPM), and differential expression analysis for binary comparisons was performed using the R package *DESeq2* (Love et al. 2014). A combined cut-off of $p < 0.05$ and $\text{abs.FC} > 2$ was chosen.

Analysis of differentially expressed transcripts across all samples was done using the R/Bioconductor *timecourse* package (Smyth 2004). The Hotelling T^2 score for each transcript was calculated using the MB.2D function with all parameters set to their default value. T^2 scores were used to rank probes according to differential expression across the time-course, and the top 5% differentially expressed transcripts were selected for complete Euclidean hierarchical clustering (k-means preprocessing; max of 300 clusters) using Perseus software. Z-scores of \log_2 normalized expression values across

the time-course were calculated and used for this analysis. 8 probe clusters were defined, and gene enrichment analysis of selected clusters was performed using Enrichr (E. Y. Chen et al. 2013). Principal component analysis (PCA) was performed on the same list of top 5% differentially expressed transcripts using Perseus.

2.5.8 m6A dot blot

m6A dot blot was performed in collaboration with Ms Stephanie Brown with minor modifications to what previously described (Batista et al. 2014). poly-A RNA was purified from total cellular RNA using the Dynabeads mRNA Purification Kit (ThermoFisher), diluted in 50 μ l of RLB (Table 14), incubated at 55°C for 15', and snap cooled on ice. Amersham Hybond-XL membrane was rehydrated in water for 3', then in 10x SSC for 10', and finally "sandwiched" in a 96-well dot blot hybridization manifold (ThermoFisher Scientific). Following two washes of the wells with 150 μ l of 10x SSC, the RNA was spotted on the membrane. After ultraviolet light (UV) cross-linking for 2' at 254nm using a Stratalinker 1800 (Stratagene), the membrane was washed once with TBST buffer, and blocked for 1h at RT with TBST supplemented with 4% non-fat dry milk. Incubations with the anti-m6A primary antibody (Synaptic System, catalogue number: 202-111; used at 1 μ g/ml) and the mouse-HRP secondary antibody (Table 5) were each performed in TBST 4% milk for 1h at RT, and were followed by three 10' washes at RT in TBST. Finally, the membrane was incubated with Pierce ECL2 Western Blotting Substrate, and exposed to X-Ray Super RX Films.

Table 14. Buffers for m6A dot blot.

Buffer	Composition
10x saline-sodium citrate buffer (SSC)	1.5M NaCl 150mM Na ₃ C ₆ H ₅ O ₇ @ pH 7.0
3-(N-morpholino)propanesulfonic acid buffer (MOPS)	20mM MOPS 12.5mM CH ₃ COONa 1.25mM EDTA @ pH 7.0
RNA loading buffer (RLB)	2.2M formaldehyde 50% formamide 0.5x MOPS buffer
TBS Tween buffer (TBST)	20mM Tris-HCl pH 7.5 150mM NaCl 0.1% Tween-20

2.5.9 m6A methylated RNA immunoprecipitation (MeRIP)

m6A MeRIP experiments were performed in collaboration with Ms Stephanie Brown following minor modification of previously described methods (Dominissini et al. 2013; Batista et al. 2014). Table 15 details the composition of the buffers used. m6A MeRIP on nuclear or cytoplasmic RNA were performed starting from 2.5×10^7 cells per condition, and cells were fed with fresh culture medium 2h before collection. Subcellular fractionation was performed as described in Chapter 2.2.8, with the exception that 1,000U/ml of RNasin ribonuclease inhibitor (Promega) and 1mM DTT were added to the ILB.

RNA was extracted from the cytoplasmic lysate or from the nuclear pellet as described in Chapter 2.2.4. Residual contaminating DNA was digested in solution using the RNase-free DNase Set from QIAGEN, and RNA was re-purified by sequential acid phenol-chloroform and chloroform extractions followed by ethanol precipitation (Chapter 2.2.9). At this stage, complete removal of DNA contamination was confirmed by qPCR of the resulting RNA without a retrotranscription step (Chapter 2.2.4). RNA was then chemically fragmented in 20 μ l reactions each containing 20 μ g of RNA in FB. Such reactions were incubated at 95°C for 5', followed by inactivation with 50mM EDTA and storage on ice. The fragmented RNA was then cleaned up by ethanol precipitation (Chapter 2.2.9).

In preparation to the MeRIP, 2.5 μ g of anti m6A-antibody (Synaptic Systems, catalogue number: 202-003) or equivalent amounts of rabbit non-immune IgG were cross-linked to 0.5mg of magnetic beads by using the Dynabeads Antibody Coupling Kit (ThermoFisher Scientific) according to manufacturer's instructions. Following equilibration of the magnetic beads by washing with 500 μ l of BB, MeRIP reactions were assembled with 50 μ g of the fragmented RNA in 500 μ l of BB supplemented with 500U of RNasin ribonuclease inhibitor. Samples were incubated at 7rpm for 1h at RT. 5 μ g of fragmented RNA (10% of the amount used for MeRIP) were set aside as pre-MeRIP input control. MeRIP reactions were washed twice with BB, once with LSB, once with HSB, and twice with TTB (each wash performed by incubating the beads with 500 μ l of buffer at 7rpm for 3' at RT). Finally, RNA was eluted from the beads by four successive incubations with 75 μ l of EB at 42°C. Both the RNA from pooled MeRIP eluates and the pre-MeRIP input were purified and concentrated by sequential acid phenol-chloroform

and chloroform extractions followed by ethanol precipitation (Chapter 2.2.9). 30 μ g of glycogen were added as carrier during ethanol precipitation. RNA was resuspended in 15 μ l of ultrapure RNase-free water.

Table 15. Buffers for m6A MeRIP.

Buffer	Composition
Fragmentation buffer (FB)	10mM ZnCl ₂ 10mM Tris-HCl pH 7.0
Binding buffer (BB)	50mM Tris-HCl pH 7.5 150mM NaCl ₂ 1% NP-40 1mM EDTA
Saline-sodium phosphate-EDTA buffer (SSPE)	150mM NaCl 10mM NaHPO ₄ -H ₂ O 10mM Na ₂ -EDTA @ pH 7.4
Low-salt buffer (LSB)	0.25x SSPE 37.5mM NaCl ₂ 1mM EDTA 0.05% Tween-20
High-salt buffer (HSB)	0.25x SSPE 137.5mM NaCl ₂ 1mM EDTA 0.05% Tween-20
TE-Tween buffer (TTB)	10mM Tris-HCl pH 7.4 1mM EDTA 0.05% Tween-20
Elution buffer (EB)	50mM Tris-HCl pH 7.5 150mM NaCl ₂ 20mM DTT 0.1% SDS 1mM EDTA

cDNA synthesis was performed using all of the MeRIP material in a 30 μ l reaction containing 500ng random primers, 0.5mM dNTPs, 20U RNaseOUT, and 200U of SuperScript II, all according to manufacturer's instructions. cDNA was diluted 10-fold, and 5 μ l were used for qPCR using KAPA Sybr Fast Low Rox (KAPA Biosystems) as described in Chapter 2.2.4. For each gene of interest, two primer pairs were designed either against the region containing the m6A peak (Batista et al. 2014), or against a negative region (portion of the same transcript lacking the m6A peak; see Table 16 for primer sequences). Results of MeRIP-qPCR for each gene were then calculated using the $\Delta\Delta$ Ct approach by using the negative region to normalize both for the expression level of the transcript of interest and for background binding.

For m6A MeRIP on total RNA, the protocol just described was followed exactly, with the exception that the subcellular fractionation step was bypassed, and that total RNA was extracted from 5x10⁶ cells. For m6A MeRIP on mRNA, poly-A RNA was purified from

75µg of total RNA using the Dynabeads mRNA Purification Kit, and 2.5µg of the resulting mRNA were used for chemical fragmentation and subsequent MeRIP with 1µg of anti-m6A antibody.

Table 16. Primers for m6A MeRIP-qPCR and RNA co-IP.

Transcript	Region	Forward primer	Reverse primer
<i>DPPA4</i>	m6A peak	CCATGCAGACTGGTGGGGT	TCCACTCACAAAGCAACAGGC
	Negative region	TGTGTGCCTTTGCCTACCCA	TGATTTCCGGATTTTGGCCTCT
<i>LEFTY1</i>	m6A peak	GCTGGCGATGACTGAACTGC	GTCAGAGGAAGCAAATTCAGGGC
	Negative region	AGCCTTCTCAAGGGACAGCC	CAGGGGCAACACCAGAGT
<i>NANOG</i>	m6A peak	ACTCCATGAACATGCAACCTGA	AGGAAGGATTCAGCCAGTGTCC
	Negative region	CCAGAACTGTGTCTCTTCCACC	CTTGCATCTGCTGGAGGCTGA
<i>OCT4/POU5F1</i>	Negative region	AGGCTATGGGAGCCCTCACT	CCCAGAGTGGTGACGGAGAC
<i>SOX2</i>	m6A peak	GTGAACCAGCGCATGGACAGT	CCTGCATCATGCTGTAGCTGCC
	Negative region	GGCGGCAACCAGAAAAACAGC	CGCGGGACCACACCATGAAG

2.5.10 m6A MeRIP followed by deep sequencing (MeRIP-seq)

Samples for MeRIP-seq were prepared in collaboration with Ms Stephanie Brown as described in Chapter 2.5.9, but with the following modifications: (1) 7.5×10^7 cells were used for each replicate; (2) nuclear RNA extraction was performed using the RNeasy Midi Kit (QIAGEN); (3) MeRIP was done in 3ml of binding buffer using 300µg of fragmented RNA and 15µg of anti-m6A antibody; (4) following elution and precipitation, m6A MeRIP and pre-MeRIP (input) samples were used for preparation of sequencing libraries. Three biological replicates for each condition were generated.

Deep sequencing was performed in collaboration with Dr Edward Farnell at Cambridge Genomics Services. Library preparation was performed using the Standard TruSeq total RNA kit (Illumina) according to manufacturer's instructions with the following exceptions: (1) Ribo-Zero treatment was performed only for pre-MeRIP samples, as ribosomal RNA contamination in m6A MeRIP samples was minimal; (1) since samples were pre-fragmented, the fragmentation step was bypassed and 30ng of RNA for each sample were used directly for library prep; (3) due to the small size of the library, a 2-fold excess of Ampure XP beads was used during all purification steps in order to retain small fragments; (4) due to the presence of contaminating adapter dimers, the library was gel extracted using gel safe stain and a dark reader in order to remove fragments smaller than ~120bp. Sequencing of the libraries was performed as described in

Chapter 2.5.6, and an average of ~80M and ~200M paired-end reads were generated for m6A MeRIP and pre-MeRIP samples, respectively.

2.5.11 m6A MeRIP-seq data analysis

These analyses were performed in collaboration with Dr Pedro Madrigal. QC of raw sequencing data was done using Trimmomatic v0.35 (Bolger et al. 2014), with parameters 'LEADING:3 TRAILING:3 SLIDINGWINDOW:5:10 MINLEN:40'. Reads were aligned to GRCh38 human genome assembly using TopHat 2.0.13 (Kim et al. 2013), with parameters '--library-type fr-firststrand' and '-transcriptome-index' and with the Ensembl GTF GRCh38.83. Identification of novel splice junctions was allowed. Forward and reverse paired-end reads were concatenated and mapped in 'single-end' mode in order to be used with MetDiff (Cui et al. 2015), which only supports single-end reads. Reads with MAPQ<20 were filtered out.

m6A peak calling and differential RNA methylation in the exome was assessed using MetDiff (Cui et al. 2015) with pooled inputs for each conditions, GENE_anno_gtf=GRCh38.83, MINIMAL_MAPQ=20, and rest of parameters as default (PEAK_CUTOFF_FDR=0.05; DIFF_PEAK_CUTOFF_FDR=0.05). An additional cut-off of abs.FC>2 was applied for certain analyses as specified in the results. Normalized read coverage files were generated using the function 'normalise_bigwig' in RSeQC-2.6 (L. Wang et al. 2012) with default parameters. The distribution of normalized reads across gene bodies in autosomal chromosomes was plotted using genomation (Akalin et al. 2015), and visualized for 5kb around stop codons using ggplot2. Motif finding on all m6A peaks identified by MetDiff was performed using DREME (Bailey 2011). For visualization purposes, biological replicates were combined. The Biodalliance genome viewer (Down et al. 2011) was used to generate figures.

Expression of transcripts in this experiment was estimated from the pre-MeRIP input samples (which represent an RNA-seq sample on nuclear-enriched RNA species) using FeatureCounts (Liao et al. 2014). Normalized gene expression values (reads per kilobase of transcript per million mapped reads, RPKM) were obtained with *edgeR* (Robinson et al. 2010).

2.5.12 RNA co-immunoprecipitation (RNA co-IP)

RNA co-IP was performed in collaboration with Ms Stephanie Brown. Cells were fed with fresh culture medium 2h before being washed once with PBS and UV cross-linked using a Stratalinker 1800 (irradiation of 400mJ/cm²). Cells were then processed for subcellular fractionation and immunoprecipitation as described in Chapter 2.2.8 with the following exceptions: (1) 800U/ml of RNAsin ribonuclease inhibitor and 1mM DTT were added to the ILB and NLB; (2) for the NLB, 0.3% Triton X-100 was used instead of 0.1% Tween-20; (3) DNase (Sigma-Aldrich) was used instead of benzonase nuclease during nuclear extraction (12.5µg per ml of lysate); (4) immunoprecipitation was performed for 3h at 4°C; (5) beads were washed twice with 1ml of LiCl wash buffer (50mM Tris-HCl pH 7.5, 250mM LiCl, 0.1% Triton X-100, and 1mM DTT) followed by two washes with 1ml of NLB; (6) following the last wash, beads were resuspended in 500µl of TRI Reagent (Sigma-Aldrich), then processed for RNA extraction and ethanol precipitation according to manufacturer's instructions. Finally, the RNA was analysed by qPCR as just described for m6A MeRIP-qPCR.

3 ACTIVIN/NODAL CONTROLS THE H3K4ME3 EPIGENETIC LANDSCAPE OF HPSCs

3.1 Statement of source

The data and text described in this chapter are largely based upon the following first author manuscript written by the author of this dissertation. Therefore, some segments have been taken *verbatim* or with minor changes from this source.

Bertero A, Madrigal P, Galli A, Hubner NC, Moreno I, Burks D, Brown S, Pedersen RA, Gaffney D, Mendjan S, et al. 2015. Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev* **29**: 702–17.

3.2 Hypothesis

As introduced in Chapters 1.3 and 1.7, Activin/Nodal signalling has a central function in controlling hPSC fate decisions. However, the mechanism by which SMAD2/3 controls the transcription of its key target genes in hPSCs remains largely unknown, and it is especially unclear whether this transcriptional control involves epigenetic mechanisms.

We decided to investigate whether and how Activin/Nodal signalling might regulate the epigenetic landscape of hPSCs. In particular, given the proposed central role of H3K4me3 and H3K27me3 in the regulation of pluripotency and differentiation (Chapter 1.5.2), we hypothesized that Activin/Nodal signalling might dictate cell fate decisions by controlling the deposition of such histone marks.

Indeed, there are several important unanswered questions regarding the regulation and function of H3K4me3 and H3K27me3 in hPSCs. First, whether and how extracellular signalling cues might control the H3K4me3 and H3K27me3 epigenetic landscape remains to be uncovered. Secondly, H3K4me3 and H3K27me3 marks could have a direct functional importance in the transcriptional regulation of genes directing cell fate decisions, or could be just a consequence of such regulations. Finally, the function of H3K4me3 and H3K27me3 in pluripotency has been studied primarily in mESCs, which represent a different pluripotent state than hPSCs (Chapter 1.2.3).

3.3 Results

3.3.1 Activin/Nodal signalling maintains H3K4me3 on specific regulators of pluripotency and differentiation

In order to define the function of Activin/Nodal signalling in the control of the epigenetic state of hPSCs, we investigated the effect of SB431542 (SB), an antagonist of the ALK4 and ALK7 type I Activin/Nodal receptors (Inman, Nicolás, Callahan, et al. 2002), on the deposition of H3K4me3 and H3K27me3 histone marks. Preliminary experiments showed that 2h of SB treatment in hESCs was the shortest time required to fully block SMAD2/3 phosphorylation and to abolish binding to its genomic targets (Figure 3.1). Therefore, we performed ChIP-seq for both H3K4me3 and H3K27me3 after 2h of SB treatment, in order to capture the more immediate and hence more likely direct effects of Activin/Nodal inhibition (Figure 3.2). We found that 491 out of 27922 H3K4me3 peaks showed a significant decrease after 2h of SB ($FC < -1.5$; Benjamini-Hochberg $adj.p < 1E-6$), while only 14 were upregulated (Figure 3.2, Appendix I). Interestingly, regions with decreased H3K4me3 were significantly associated with genes involved in Activin/Nodal signalling and with factors expressed in the epiblast and in endoderm cells (GREAT analysis; Figure 3.3). In contrast, we observed almost no significant differences for the 11347 H3K27me3 peaks identified in the analysis, with only one region being increased and none showing decrease (Figure 3.2). Thus, Activin/Nodal signalling is necessary to maintain the positive histone mark H3K4me3 on a subset of genes in hESCs, while the deposition of the negative histone mark H3K27me3 appears to be independent of Activin/Nodal.

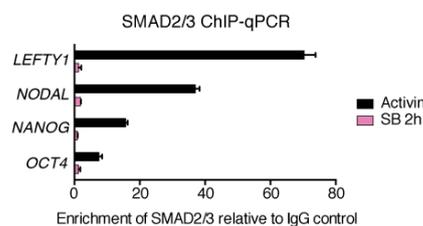


Figure 3.1. Effect of SB treatment on SMAD2/3 genomic binding. ChIP-qPCR for SMAD2/3 on its binding sites associated to the indicated genes.

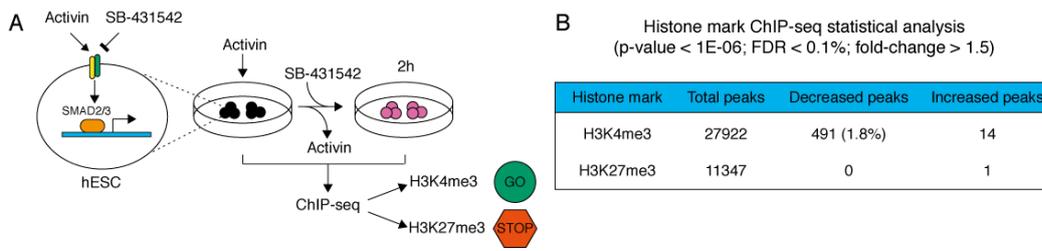


Figure 3.2. ChIP-seq for H3K4me3 and H3K27me3 following inhibition of Activin/Nodal signalling.

(A) Schematics of the experimental approach. ChIP-seq were performed in collaboration with the Wellcome Trust Sanger Institute next-generation sequencing facility. (B) Results of the statistical analysis of ChIP-seq data. ChIP-seq analysis was done in collaboration with Dr Pedro Madrigal.

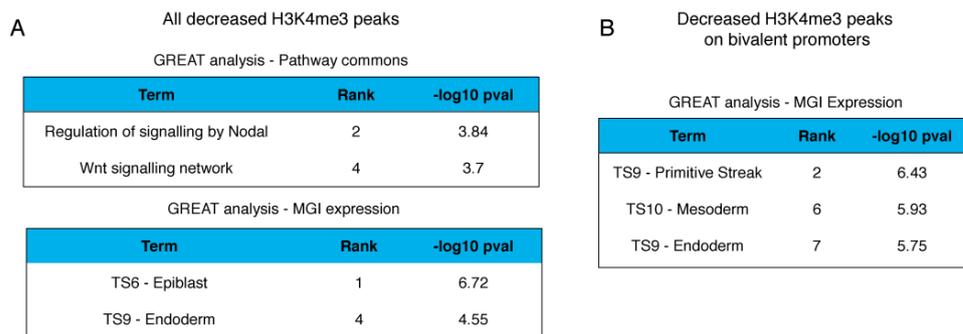


Figure 3.3. Functional enrichment analysis of differential H3K4me3 peaks.

(A) Selected results of GREAT analysis (McLean et al. 2010) on the H3K4me3 peaks downregulated after inhibition of Activin/Nodal signalling for 2h with SB. (B) As in A, but the analysis was performed only on H3K4me3 peaks in a range of \pm 5Kb from the closest TSS and colocalizing with H3K27me3 marks (bivalent promoters).

Accordingly to the known typical localization of H3K4me3 (Calo & Wysocka 2013), several of the H3K4me3 peaks decreased after SB treatment marked proximal promoters and transcription start sites (Figure 3.4). These included several well-known Activin/Nodal transcriptional targets such as *NANOG*, *POU5F1* (more commonly known as *OCT4*), *LEFTY1* and *NODAL* (Figure 3.5). Interestingly, H3K4me3 was decreased on key pluripotency regulators that are highly expressed and marked by H3K4me3 but not H3K27me3, for instance *NANOG*, *OCT4*, *DPPA4*, *GDF3* and *PRDM14* (Young, 2011). However, we observed that inhibition of Activin/Nodal signalling also resulted in impaired H3K4me3 on many genes that show low expression in hESCs (see Figure 3.10 for gene expression data) and are marked both by H3K4me3 and H3K27me3. Examples from this class of genes included *LEFTY1*, *NODAL*, *LEFTY2*, *CER1*, *WNT3*, and *FGF8*. Indeed, 25% of the promoters where H3K4me3 was decreased after 2h of SB displayed this bivalent histone marking (Figure 3.4), representing a similar proportion to the overall abundance of these elements (32%). Remarkably, the bivalent promoters where H3K4me3 was decreased after SB were associated with genes expressed in the primitive

streak, mesoderm, and endoderm (Figure 3.3), thereby supporting the proposed role of bivalent marks in the regulation of developmental gene expression (Chapter 1.5.2.3).

Aside from promoter-associated regions, we observed that almost 60% of H3K4me3 peaks downregulated after 2h of SB fell outside of gene bodies (Figure 3.4). Inspection of these regions revealed a frequent association with the deposition of H3K4me1 and H3K27ac (Figure 3.6), two well-established markers of active enhancers (Calo & Wysocka 2013). Indeed, 25% of intergenic H3K4me3 peaks decreased after 2h of SB showed this particular feature (Figure 3.4). In contrast, only 0.02% of them colocalized with H3K4me1 and H3K27me3, a chromatin signature that identifies poised enhancers (Rada-Iglesias et al. 2011), while the overall abundance of such regions was much higher (18%; Figure 3.4). As such, Activin/Nodal signalling appears to regulate H3K4me3 specifically on active distal enhancers, but not on poised ones.

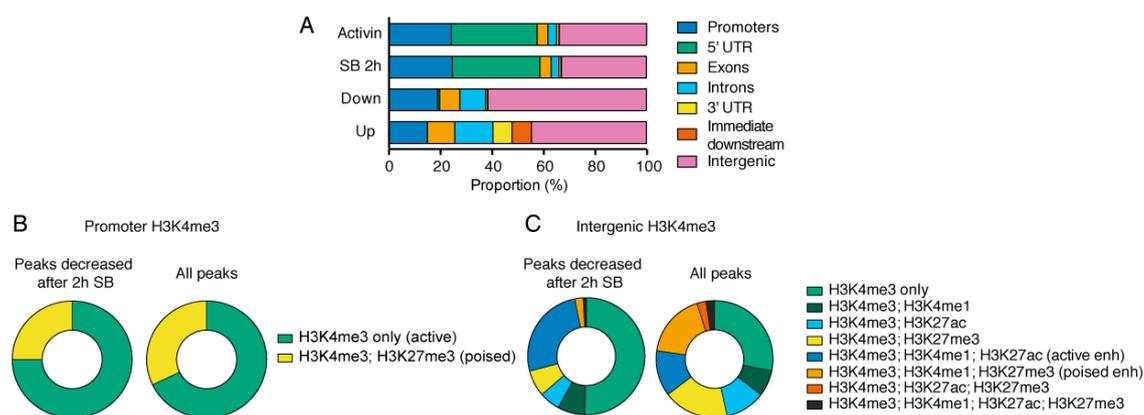


Figure 3.4. Genomic features of Activin/Nodal-dependent H3K4me3.

(A) Annotation of H3K4me3 peaks to genomic features. Down/up: peaks down- or upregulated after 2h of SB. (B) Co-localization of H3K27me3 peaks with H3K4me3 peaks centred in a range of \pm 5Kb from the closest TSS (promoters). (C) Co-localization of H3K4me1 and H3K27ac peaks with H3K4me3 peaks centred outside a range of \pm 5Kb from the closest TSS (intergenic).

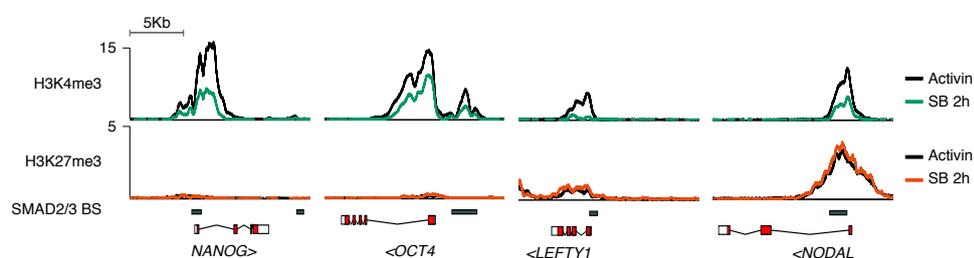


Figure 3.5. Activin/Nodal-dependent H3K4me3 on active and poised promoters.

ChIP-seq results for H3K4me3 and H3K27me3 on selected SMAD2/3 target genes before and after inhibition of Activin/Nodal with SB for 2h. Histograms represent read-enrichments normalized by million mapped reads and size of the library. SMAD2/3 binding sites in hESCs are reported (Brown et al. 2011).

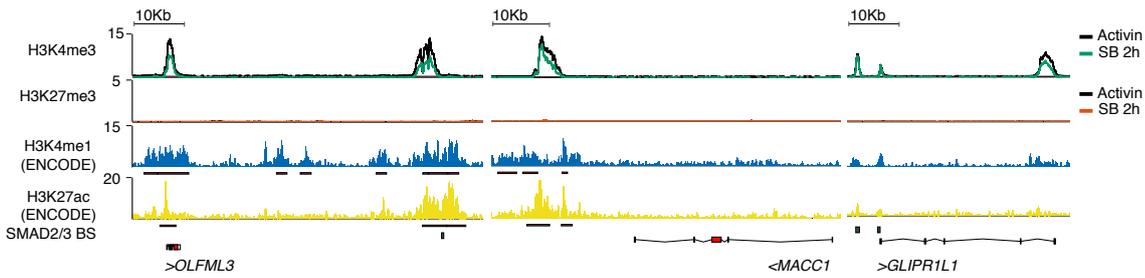


Figure 3.6. Activin/Nodal-dependent H3K4me3 on active enhancers. As in Figure 3.5, but for selected representative intergenic H3K4me3 peaks. H3K4me1 and H3K27ac ChIP-seq data and peak calls for the H1 hESC line from the ENCODE project are also shown.

Importantly, inhibition of Activin/Nodal signalling for 2h specifically impaired H3K4me3 but not H3K4me2 or H3K4me1 both on promoter and enhancers (Figure 3.7). Indeed, the levels of H3K4me2 and H3K4me1 were either unchanged or increased on most regions we analysed. The only exception to this was *LEFTY1*, where both H3K4me3 and H3K4me2 were decreased. Moreover, H3 enrichment was unaffected by inhibition of Activin/Nodal signalling (Figure 3.7), demonstrating that the reduction of H3K4me3 levels was not due to nucleosome repositioning. Overall, these findings suggested that inhibition of Activin/Nodal specifically results in loss of H3K4me3.

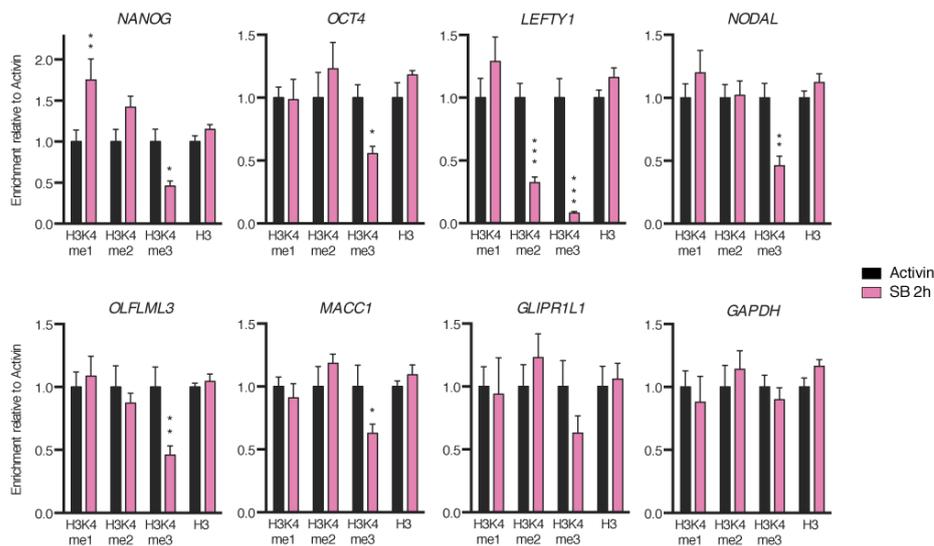


Figure 3.7. Specificity of Activin/Nodal-dependent H3K4me3. ChIP-qPCR for H3, H3K4me1, H3K4me2 and H3K4me3 before and after 2h of SB. qPCR was performed for the genomic regions showing decreased H3K4me3 in Figure 3.5 and Figure 3.6. For each gene, significant differences vs Activin are reported (t-tests; n=3).

Interestingly, 21/491 of the H3K4me3 peaks decreased upon Activin/Nodal signalling inhibition directly overlapped with SMAD2/3 binding sites ($p < 1E-04$, as measured by

GAT analysis; Heger et al. 2013; SMAD2/3 bound genes from Brown et al., 2011). Among others, canonical SMAD2/3 target genes such as *NANOG*, *OCT4*, *NODAL*, and *LEFTY1* showed this association (Figure 3.5). Indeed decrease of H3K4me3 after 2h of SB on such genes correlated with loss of SMAD2/3 binding (Figure 3.1). Moreover, regions with decreased H3K4me3 after Activin/Nodal inhibition were significantly associated with nearby SMAD2/3 binding sites (27% and 100% were 10Kb or 100Kb upstream/downstream from the closest SMAD2/3 binding site; GAT analysis $p < 1E-0.4$ and $p < 0.033$, respectively). This observation is in agreement with previous reports that showed how SMAD2/3 regulates the expression of its target genes mostly by binding to distal enhancers rather than to proximal promoters (Kim et al. 2011; Brown et al. 2011).

Taken together, these findings suggested that Activin/Nodal signalling could control the expression of master regulators of both pluripotency and germ layer specification by maintaining H3K4me3 on both gene promoters and intergenic enhancers.

3.3.2 Activin/Nodal-dependent H3K4me3 is associated to gene expression changes

To test the functional relevance of H3K4me3 loss after SB treatment, we investigated the gene expression dynamics resulting from both acute and chronic Activin/Nodal signalling inhibition. Accordingly, we performed gene expression microarrays of hESCs grown in the presence of Activin or treated with SB for 2h, 4h, 8h, 24h, and 48h (Figure 3.8). Hierarchical clustering of the top 10% differentially expressed probes across the time-course identified three main clusters, two of which contained genes whose expression was decreased after inhibition of Activin/Nodal (Figure 3.8 and Figure 3.9). As expected, these clusters were significantly enriched in genes associated with TGF β signalling, regulation of cell differentiation, and cell cycle (gene enrichment analysis; Figure 3.9). However, these two clusters differed both in their relative size and in the speed of their downregulation: expression of genes in the first smaller cluster started to decrease already after 2h, while the second bigger cluster was significantly affected only after 24h. Importantly, both clusters presented significant overlap with genes bound by SMAD2/3 (70/233 for cluster 1 and 444/1819 for cluster 2, hypergeometric test $p=1.88E-11$ and $p=7.09E-40$), and contained several well-known SMAD2/3 direct targets such as *LEFTY1*, *NODAL*, *NANOG*, *SOX17*, *EOMES*, and *GSC* (cluster 1), and *OCT4*, *DPPA4*, and *EPCAM* (cluster 2). Importantly, these two clusters included not only several pluripotency factors, but also regulators of mesendoderm differentiation (like *SOX17*, *EOMES*, *GSC*, *LEFTY1*, and *NODAL*) that are expressed at background levels in hESCs (Figure 3.10 reports the Ct values for such genes) in agreement with the “primed” pluripotent state of hESCs (Chapter 1.2.3). As such, Activin/Nodal inhibition not only reduced expression of pluripotency genes, but also abolished the primed expression of mesendoderm regulators. Conversely, a third major cluster represented transcripts induced after 24h to 48h of Activin/Nodal inhibition (Figure 3.8 and Figure 3.9). This was significantly enriched in genes involved in neural development, and contained several known SMAD2/3-inhibited factors such as *CDX2*, *WNT8A*, *EGFLAM*, *MEIS2*, and *CITED2* (Figure 3.9). Importantly, qPCR on a subset of genes from each of these three clusters validated the accuracy of the microarray analyses (Figure 3.10). Overall, these observations showed that inhibition of Activin/Nodal signalling leads to both rapid and delayed transcriptional responses that regulate expression of genes involved in pluripotency and cell fate decisions.

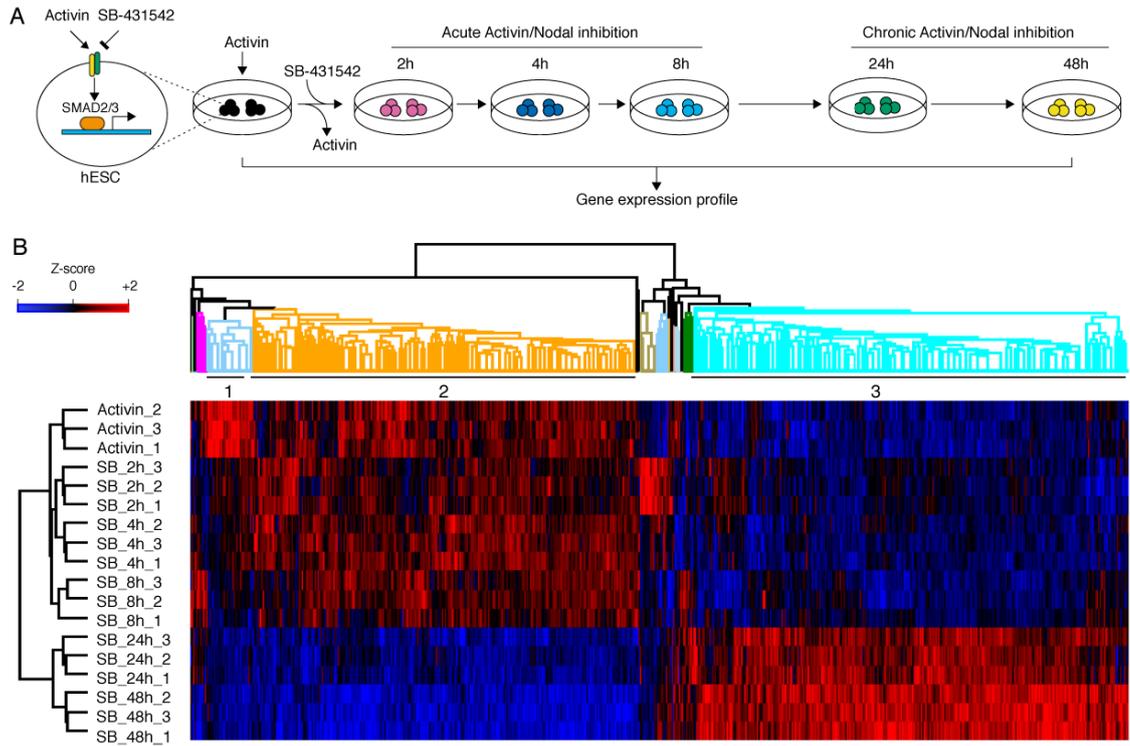


Figure 3.8. Global gene expression response to the inhibition of Activin/Nodal signalling.

(A) Schematics of the experimental approach. Microarrays were performed in collaboration with the Wellcome Trust Sanger Institute microarray facility. (B) Euclidean hierarchical clustering of differentially expressed microarray probes across the time-course of Activin/Nodal inhibition in hESCs (top 10% ranked by Hotelling T^2 statistic). Z-scores indicate the differential expression measured in number of standard deviations from the average level across all the time points. The three major probe clusters are indicated.

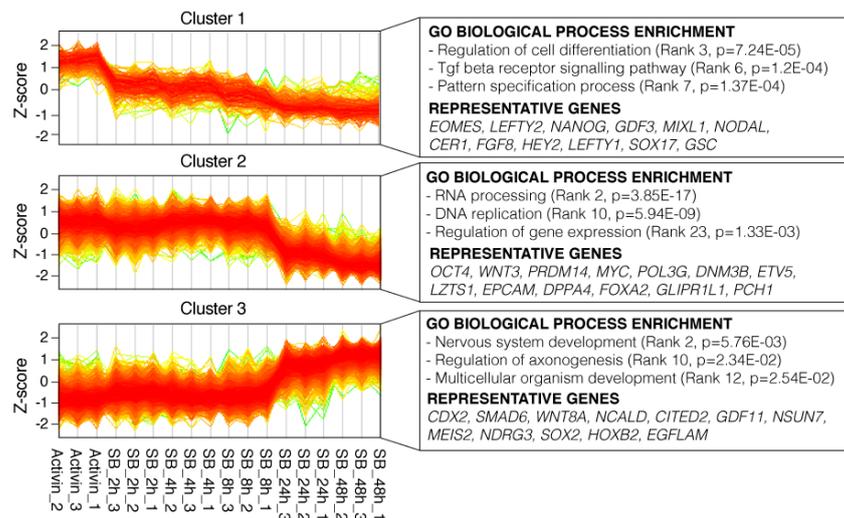


Figure 3.9. Expression profiles of genes regulated by Activin/Nodal signalling.

Expression profiles of probes in the clusters indicated in Figure 3.8. For each cluster, selected results of gene enrichment analysis and representative genes are reported.

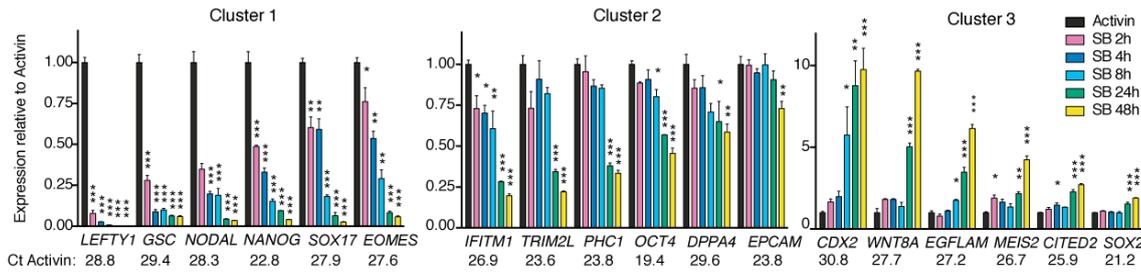


Figure 3.10. qPCR validation of microarray experiments.

Gene expression qPCR before and after inhibition of Activin with SB for the indicated time points. Gene clusters refer to those described in Figure 3.8 and Figure 3.9. The average Ct in hESCs cultured in the presence of Activin is reported (the average Ct for the housekeeping gene *HMBS/PBGD* was 25.02). For each gene, significant differences vs Activin are reported (one-way ANOVA; n=3).

We then investigated the relationship between H3K4me3 reduction and gene expression changes induced by Activin/Nodal inhibition. Interestingly, 108/415 genes associated to decreased H3K4me3 after 2h of SB were among the top 10% differentially expressed genes across the kinetics of Activin/Nodal inhibition. These included several genes important for pluripotency (such as *NANOG*, *OCT4*, *DPPA4*, *GDF3*, and *PRDM14*), as well as regulators of mesendoderm differentiation (such as *CER1*, *WNT3*, *LEFTY1*, *LEFTY2*, *NODAL*, and *FGF8*). These genes belonged to either cluster 1 or 2 described above, and were downregulated upon inhibition of Activin/Nodal signalling (Figure 3.11). This is in agreement with the known role of H3K4me3 as a histone mark that promotes gene expression (Ruthenburg et al. 2007). GSEA analysis confirmed that genes downregulated after inhibition of Activin/Nodal for 48h were significantly associated with regions where H3K4me3 was reduced after 2h of SB (Figure 3.11).

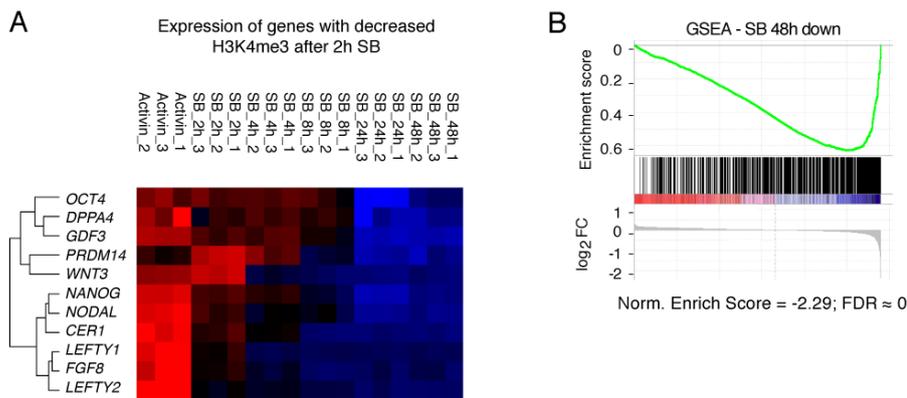


Figure 3.11. Relationship between gene expression and epigenetic changes following inhibition of Activin/Nodal signalling.

(A) As in Figure 3.8, but for selected representative genes that showed decreased H3K4me3 upon 2h of inhibition of Activin/Nodal signalling. (B) GSEA for genes whose expression was decreased after 48h of SB in the list of H3K4me3-associated genes ranked by the differential H3K4me3 enrichment before or after 2h of SB.

Of note, decreased expression in genes with reduced H3K4me3 after 2h of SB occurred progressively and reached the highest significance at 48h (Figure 3.12), showing that changes in the H3K4me3 mark preceded the decrease in gene expression. Thus, Activin/Nodal signalling could maintain the expression of key developmental regulators by controlling the deposition of H3K4me3 on the corresponding genomic regions.

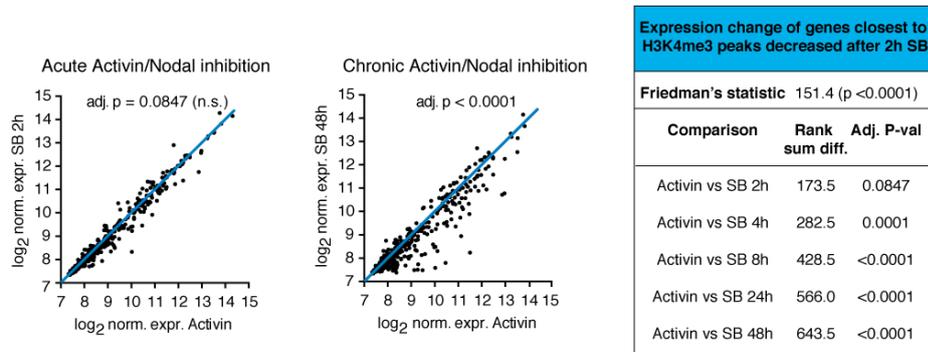


Figure 3.12. Relationship between early epigenetic changes and gene expression dynamics following inhibition of Activin/Nodal signalling.

Gene expression after 2h (acute inhibition) or 48h (chronic inhibition) of SB treatment. The analysis is focused on the genes closest to the H3K4me3 peaks decreased after 2h of SB. The blue lines indicate no expression change. The significance of expression differences across the time-course (Friedman's statistic) and the significance of paired comparisons for each time point vs Activin (Dunn's multiple comparisons test) are reported.

In order to confirm and extend these observations, we monitored the levels of H3K4me3 and H3K27me3 by ChIP-qPCR after 2h, 4h, and 8h of SB treatment on a panel of genes whose expression was fully down- or upregulated after inhibition of Activin/Nodal. Strikingly, H3K4me3 was impaired already after 2h of SB on genes showing both rapid (cluster 1) and slow (cluster 2) downregulation, while upregulated ones (cluster 3) were not affected (Figure 3.13). Interestingly, none of the genes analysed showed a significant change in the levels of H3K27me3 (Figure 3.13).

Collectively, these observations established that loss of H3K4me3 following Activin/Nodal inhibition precedes gene expression impairment. This suggested a functional importance for H3K4me3 regulation downstream of Activin/Nodal signalling to maintain the expression of genes controlling self-renewal and differentiation potency.

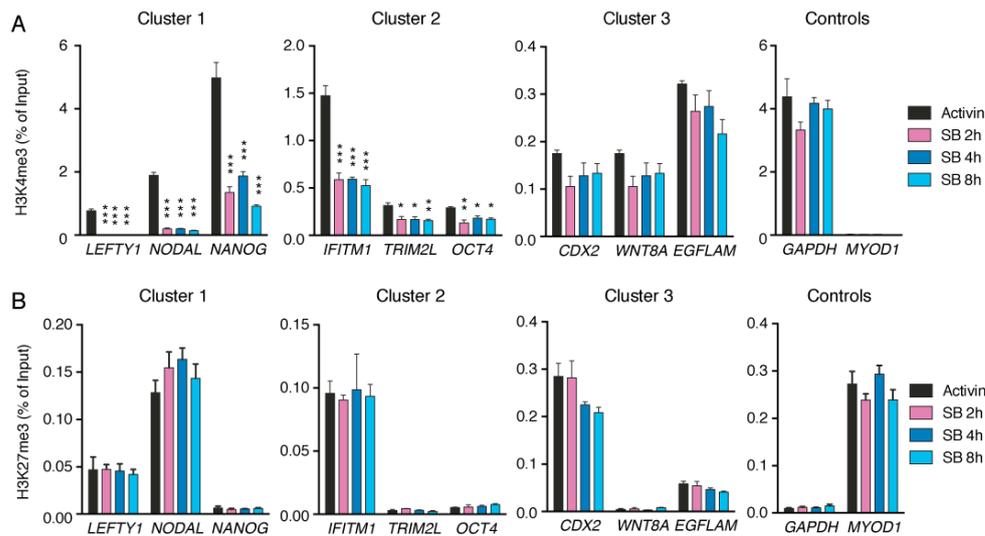


Figure 3.13. H3K4me3 and H3K27me3 dynamics following inhibition of Activin/Nodal signalling.

(A) ChIP-qPCR for H3K4me3 on SMAD2/3 target genes before and after SB for 2h, 4h, and 8h. Gene clusters refer to those from Figure 3.8 and Figure 3.9. *GAPDH* and *MYOD1* are positive and negative controls for this ChIP. For each gene, significant differences vs Activin are reported (one-way ANOVA; n=3). (B) ChIP-qPCR as in A, but for H3K27me3. *GAPDH* and *MYOD1* are negative and positive controls for this ChIP; n=3. qPCR was performed using primer pairs specific for the promoters of the indicated genes based on ChIP-seq data.

3.3.3 SMAD2/3 functionally interacts with DPY30 to maintain the pluripotent state

Having defined the importance of Activin/Nodal signalling to maintain H3K4me3 on key regulatory genes, we sought to understand the underlying molecular mechanism. The association between decreased H3K4me3 and SMAD2/3 binding sites (Chapter 3.3.1) suggested that the Activin/Nodal signalling effector SMAD2/3 could be directly involved in the deposition of H3K4me3. Thus, we decided to test whether SMAD2/3 could interact with the COMPASS complexes, which are responsible for the deposition of this histone modification. As described in detail in Chapter 1.5.2.1, the H3K4 methyltransferases belong to the MLL/SETD1 family. These form six functional complexes classified into three subgroups, all of which include common WRAD (WDR5, RBBP5, ASHL2 and DPY30) cofactors (Ernst & Vakoc 2012; Ruthenburg et al. 2007). Among these enzymes, SETD1A, MLL1/KMT2A, and MLL2/KMT2B have the highest expression in hESCs, and co-immunoprecipitation experiments revealed that SMAD2/3 bound to SETD1A and MLL2/KMT2B, but not MLL1/KMT2A (Figure 3.14).

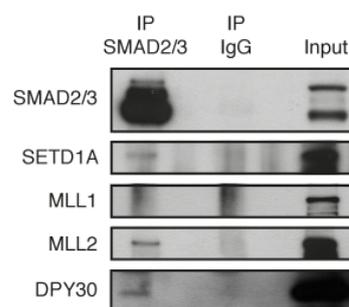


Figure 3.14. SMAD2/3 interacts with H3K4 methyltransferases and DPY30. Western blots of SMAD2/3 or control (IgG) immunoprecipitations from nuclear extracts of hESCs. Input is 5% of the material used for IP.

In order to evaluate the functional relevance of such interactions in the context of Activin/Nodal-dependent H3K4me3 and gene expression, we decided to knockdown the expression of the common COMPASS cofactor DPY30. Indeed, DPY30 is required for efficient COMPASS-dependent H3K4 trimethylation, but not for H3K4 mono- and dimethylation. This is in contrast with the other members of the WRAD module, which are also important for H3K4me2 and H3K4me1 deposition (Ernst & Vakoc 2012). Moreover, being a member of all COMPASS complexes, impairment of DPY30 should prevent compensatory mechanisms between catalytic subunits (Jiang et al. 2011). As such, knockdown of DPY30 has been previously used as a powerful genetic tool to specifically impair H3K4me3 deposition in different cell types (Yang et al. 2014; Jiang

et al. 2011), since specific inhibitors of this histone modification are not currently available.

We stably transfected hESCs with vectors expressing shRNAs directed against DPY30 (Figure 3.15), and isolated individual sublines showing impaired levels of DPY30 at the RNA and protein level (Figure 3.15). Of note, all of the results described below were confirmed with two separate shRNAs, in order to mitigate the possibility that phenotypic effects could be due to off-target activity of the shRNA. Interestingly, we obtained a reduced number of sublines after transfection of DPY30 shRNAs when compared to hESCs transfected with a vector expressing a scramble (SCR) shRNA (Figure 3.15). This suggested that absence of DPY30 expression might interfere with hESC self-renewal. Accordingly, expansion of hESC sublines knockdown for DPY30 (DPY30 KD hESCs) was challenging due to a markedly increased background of differentiation and to a slower proliferation rate (Figure 3.16). Interestingly, apoptosis was not increased in these cells, suggesting that DPY30 was not required for hESC survival (Figure 3.16). Importantly, DPY30 KD hESCs displayed impaired alkaline phosphatase activity (Figure 3.16), which is a hallmark of pluripotency. Finally, the expression of pluripotency and endoderm markers was decreased in DPY30 KD hESCs, while neuroectoderm genes were upregulated (Figure 3.17). These results were confirmed at the protein level by flow cytometry and immunostaining (Figure 3.17). In sum, these observations indicated that DPY30 is important to preserve the pluripotent state of hESCs by maintaining the expression of pluripotency markers and by blocking the expression of neuroectoderm genes.

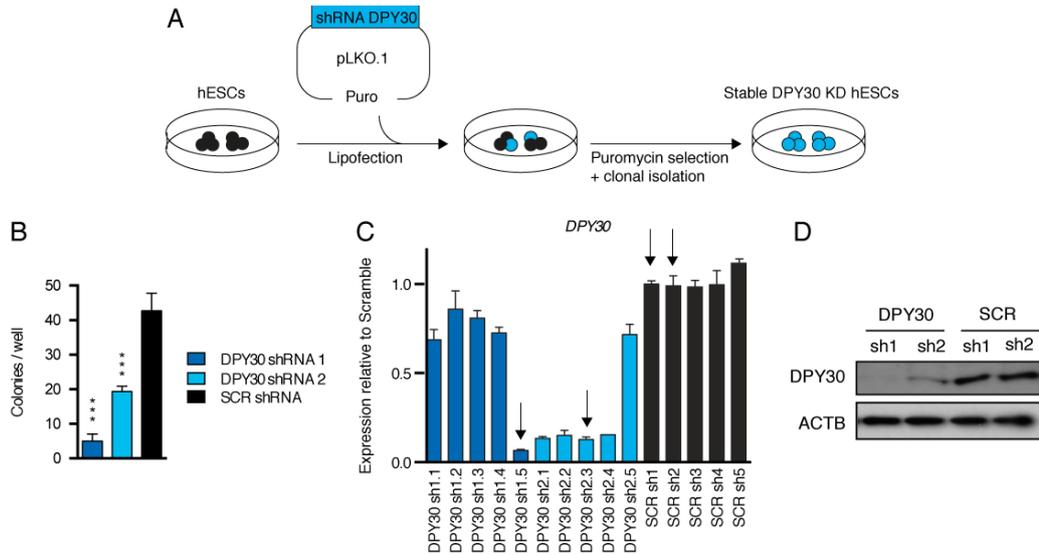


Figure 3.15. Generation of DPY30 knockdown hPSCs.

(A) Schematics of the generation of stable DPY30 knockdown (KD) hESCs. (B) Number of total colonies obtained after selection of stable integrants for two different DPY30 shRNAs or a SCR shRNA. Significant differences vs SCR shRNA are reported (one-way ANOVA; n=3). (C) qPCR for *DPY30* in hESC clonal lines. Arrows indicate the sublines selected for further experiments. (D) Western blots for DPY30 in the selected lines. ACTB (β -actin): loading control.

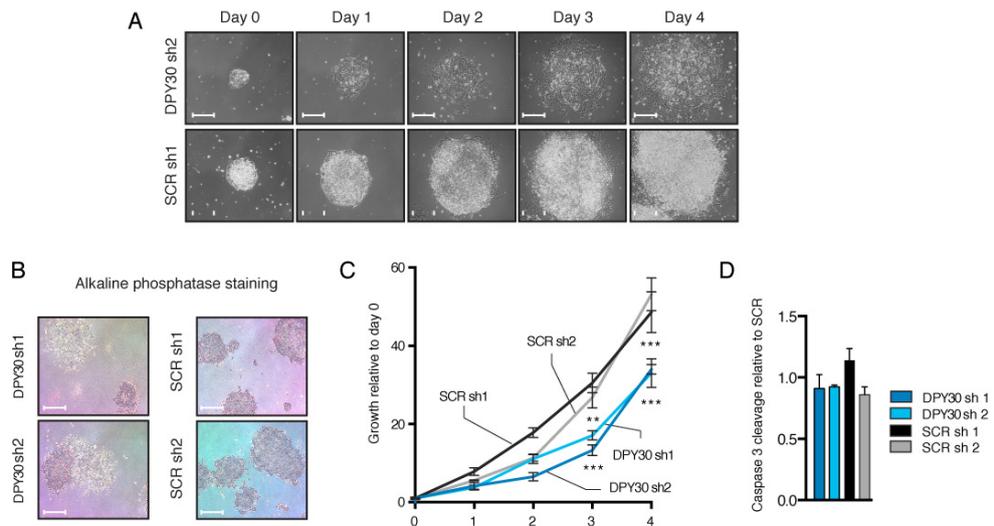


Figure 3.16. Effect of DPY30 knockdown on hPSC self-renewal.

(A) Phase-contrast images of the same DPY30 KD or SCR control hESC colony after the indicated number of days from the last cell split (day 0). Scale bars: 200 μ m. (B) Alkaline phosphatase staining (positivity is indicated by purple/blue colour). Scale bars: 200 μ m. (C) Growth curves. Significant differences vs both SCR sh1 and SCR sh2 are reported (two-way ANOVA, only the highest p is shown; n=4). (D) Apoptotic index (measured by Caspase 3 cleavage assay). No significant differences vs both SCR sh1 and SCR sh2 were identified (one-way ANOVA; n=3).

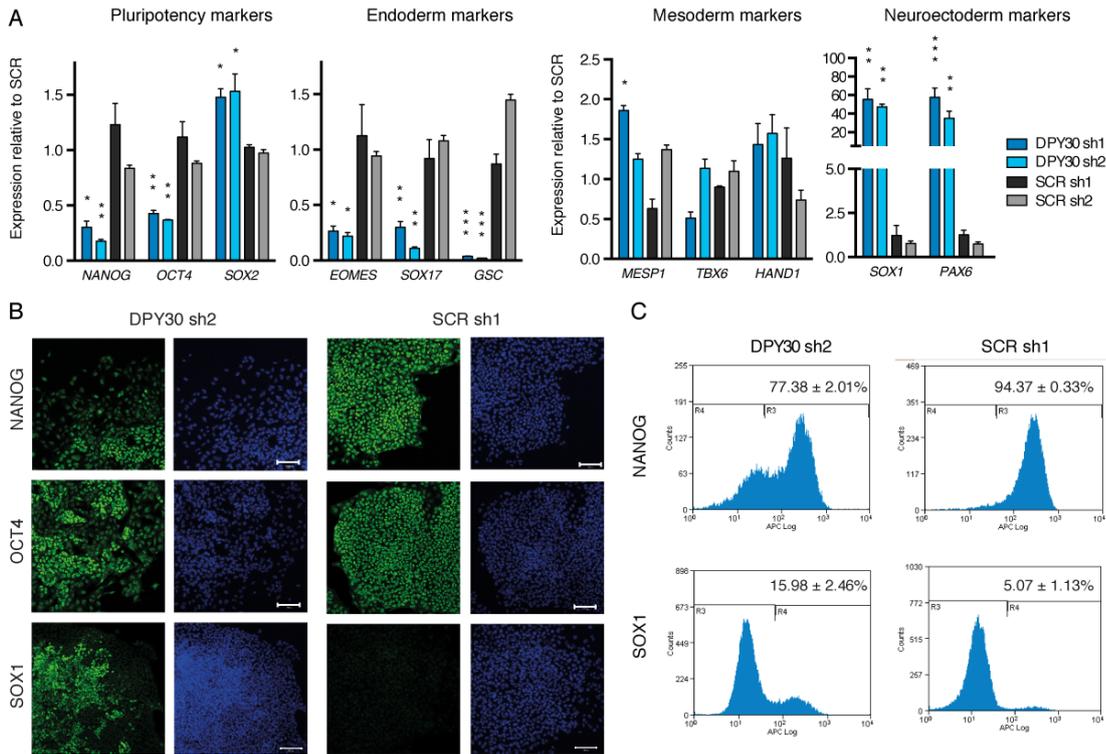


Figure 3.17. Effect of DPY30 knockdown on the expression of pluripotency and differentiation markers.

(A) Gene expression qPCR in DPY30 KD or SCR control hESCs. Note that *SOX2* is both a pluripotency and a neuroectoderm marker. For each gene, significant differences vs both SCR sh1 and SCR sh2 are reported (one-way ANOVA, only the highest p is shown; n=3). (B) Immunofluorescences for the indicated proteins (green) or nuclear staining (DAPI, blue). Scale bars: 100µm. (C) Representative flow cytometry histograms for the indicated proteins. The gates used to define positive cells are shown (based on secondary only staining controls). The average percentages ± SEM of positive cells from three biological replicates are reported.

Interestingly, the phenotype of DPY30 KD hESCs closely resembled the effect of Activin/Nodal signalling inhibition (Chapter 1.3.1). This similarity was confirmed at a global level by microarray analyses, which showed that genes downregulated after DPY30 KD (326 genes; FC<-1.5 and adj.p<0.01) were significantly associated with TGFβ signalling, and had a significant over-representation of factors required for embryonic development (GO enrichment analysis). On the other hand, upregulated transcripts (256 genes; FC>1.5 and adj.p<0.01) had a significant over-representation of genes involved in craniofacial and neural development. Moreover, the transcriptional profile of DPY30 KD hESCs very significantly overlapped with the one resulting from both 2h and 48h of SB treatment (Figure 3.18; highest hypergeometric test p<1E-500, as calculated using RRHO analysis; Plaisier et al., 2010). Indeed, all of the genes that we had validated by qPCR to be increased or decreased during SB treatment (Figure 3.10) followed the same trend in DPY30 KD cells. Notably, these factors included well-known SMAD2/3 targets such as *LEFTY1*, *NODAL*, *NANOG*, *OCT4*, *CDX2*, and *WNT8A*

(Figure 3.18). Altogether, these results demonstrated a functional interaction between DPY30 and the Activin/Nodal-SMAD2/3 signalling pathway.

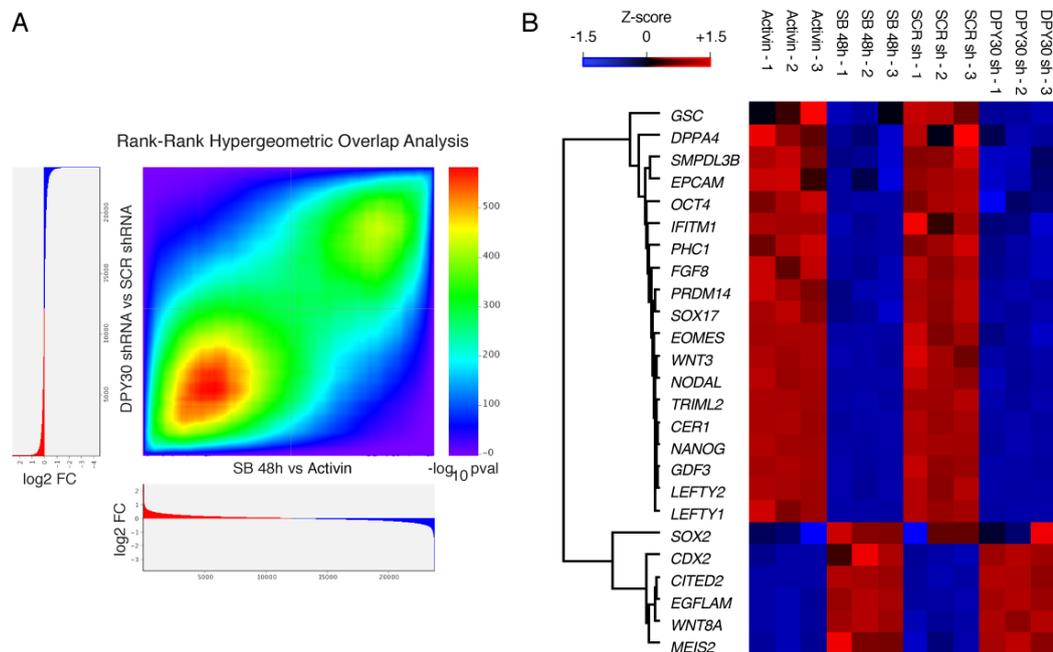


Figure 3.18. Global similarities between the gene expression changes induced by knockdown of DPY30 and by the inhibition of Activin/Nodal signalling.

Microarrays were performed in collaboration with the Wellcome Trust Sanger Institute microarray facility. (A) RRHO for genes ranked by their differential expression after DPY30 KD or inhibition of Activin/Nodal for 48h with SB. Colour-coded \log_{10} p-values indicate the significance of the overlap between genes in the two conditions, as measured by hypergeometric tests. (B) Heatmap showing changes in gene expression of selected SMAD2/3 target genes. Z-scores were separately calculated for each experiment.

We then evaluated if the gene expression changes associated with knockdown of DPY30 correlated with changes in H3K4me3 levels. ChIP-qPCR both on pluripotency and endoderm regulators transcriptionally activated by SMAD2/3 and downregulated after DPY30 KD showed reduced levels of H3K4me3 (Figure 3.19). Interestingly, DPY30 knockdown did not affect H3K4me3 on a diversity of loci, including housekeeping genes such as *GAPDH* (Figure 3.19). This showed that the requirement for high levels of DPY30 in hESCs is gene specific. Interestingly, transient transfection for 48h of a DPY30 shRNA also resulted in a similar gene-specific reduction of H3K4me3 on SMAD2/3 target genes (Figure 3.19), thus indicating that histone marks quickly decreased after DPY30 knockdown. Moreover, this event preceded changes in gene expression (Figure 3.20), further supporting the notion that the epigenetic changes induced transcriptional impairment.

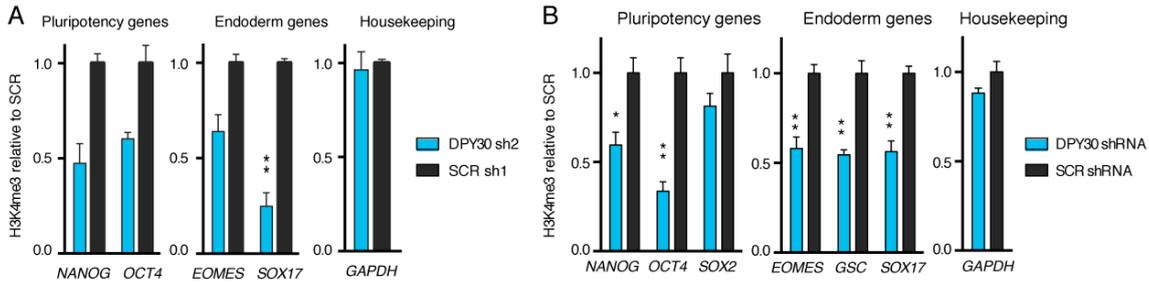


Figure 3.19. Effect of DPY30 knockdown on the H3K4me3 levels onto Activin/Nodal target genes.

ChIP-qPCR for H3K4me3 in hESCs expressing stable (A) or transiently transfected (B; 48h following lipofection, see Figure 3.20) DPY30 or SCR control shRNAs. For each gene, significant differences vs SCR are reported (t-test; n=3). qPCR was performed using primer pairs specific for the promoters of the indicated genes (Figure 3.32).

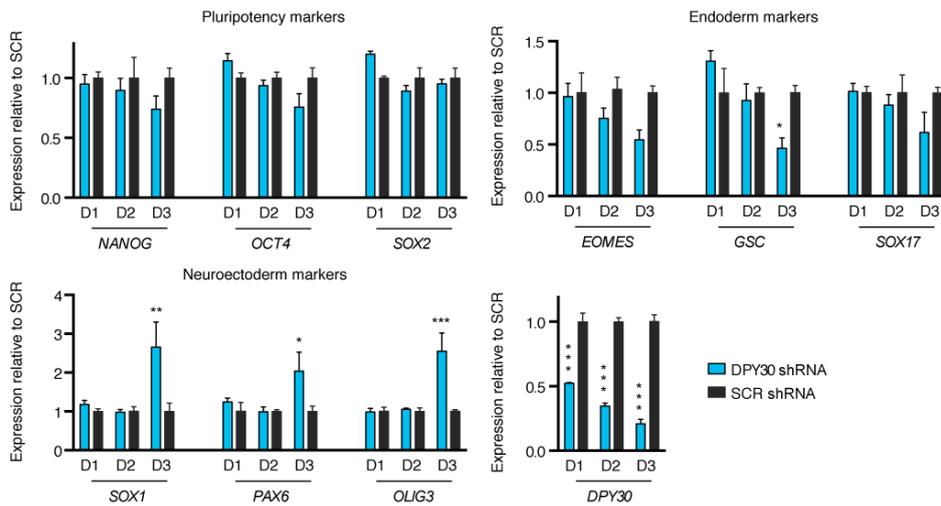


Figure 3.20. Gene expression dynamics following transient DPY30 knockdown.

Gene expression qPCR in hESCs transiently transfected with an shRNA against DPY30 or a SCR control shRNA. Samples were collected after 24h, 48h, and 72h from the transfection (D1, D2, and D3). Note that *SOX2* is both a pluripotency and a neuroectoderm marker. For each gene, significant differences vs SCR are reported (two-ways ANOVA; n=3).

Collectively, these findings suggested that DPY30 and SMAD2/3 might cooperate to preserve the H3K4me3 histone mark on several Activin/Nodal signalling target genes, and that this mechanism is important to maintain the pluripotent state.

3.3.4 Mesendoderm differentiation of hPSCs relies on DPY30-dependent H3K4me3

Considering the effect following DPY30 knockdown in undifferentiated hESCs, we decided to assess the capacity of differentiation of DPY30 KD hESCs by performing *in vivo* teratoma assays. Strikingly, teratomas derived from DPY30 KD hESCs were small and failed to completely invade the testicular capsule of immunodeficient mice, in contrast with those grown from SCR shRNA control hESCs (Figure 3.21). Moreover, the resulting tissue was composed predominantly of neuroectodermal lineages, such as neuroepithelial cells, but lacked mesoderm and endoderm derivatives (Figure 3.21). These results showed that a decrease in DPY30 expression limited the capacity of hESCs to both self-renew and differentiate into all derivatives of the three germ layers.

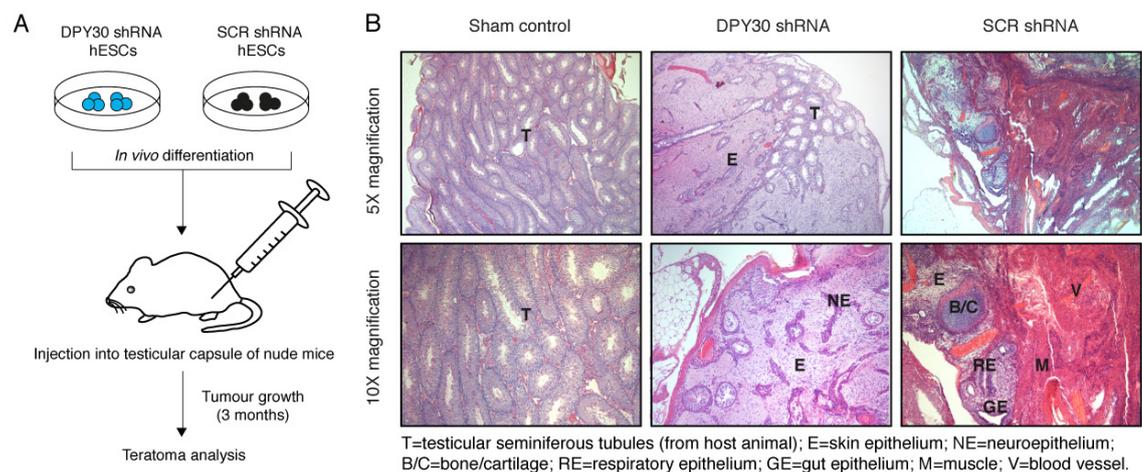


Figure 3.21. Teratoma assays of DPY30 knockdown hPSCs.

(A) Schematic of the experimental approach. (B) Haematoxylin and eosin histological staining of normal testis tissue (sham control) or teratomas. Teratoma assays were performed in collaboration with Dr Deborah Burks.

To more precisely characterize this defect in differentiation potency, DPY30 KD hESCs were grown in chemically defined culture conditions directing differentiation towards endoderm and neuroectoderm. qPCR analyses for the expression of lineage-specific markers proved that DPY30 KD hESCs responded poorly to endoderm differentiation, while neuroectoderm specification was efficient (Figure 3.22). Importantly, reduced mRNA levels for endoderm genes correlated both with impaired RNA polymerase II (Pol II) and Mediator complex recruitment on the transcription start site, and with reduced levels of the elongation marker histone 3 lysine 36 trimethylation (H3K36me3) on the gene body (Figure 3.23). Moreover, SMAD2/3 binding to endoderm genes was also

reduced (Figure 3.23). Overall, DPY30 knockdown caused transcriptional impairment of such regulators.

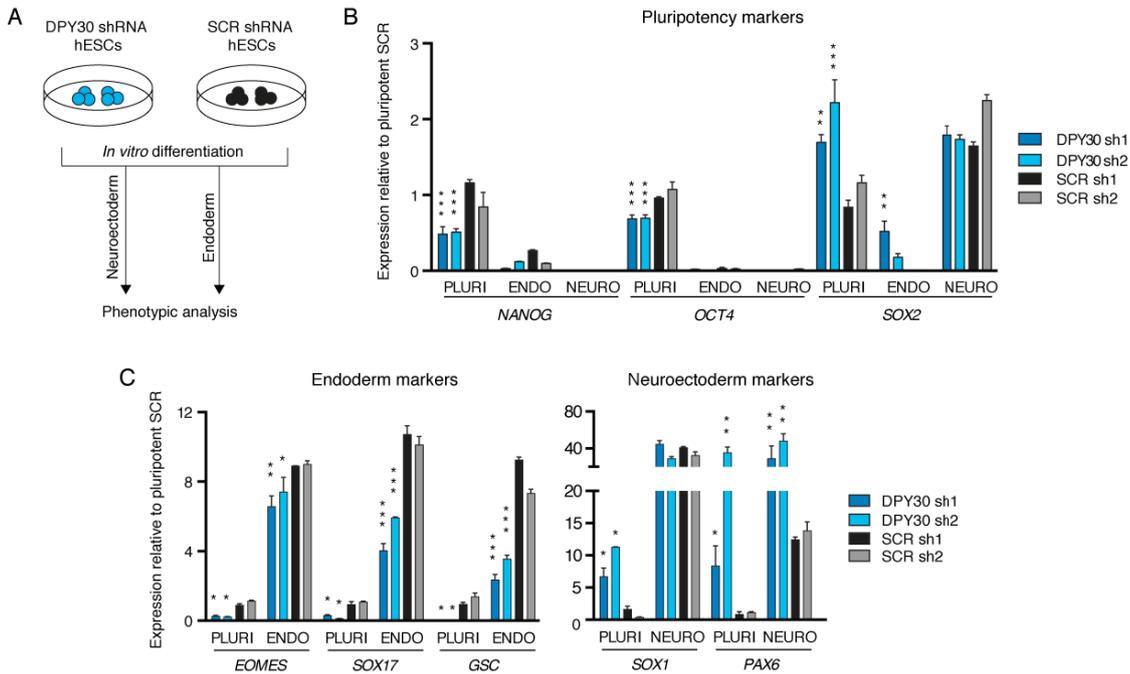


Figure 3.22. Transcriptional analysis of DPY30 knockdown cells differentiated into the germ layers.

(A) Schematic of the experimental approach. (B-C) Gene expression qPCR in DPY30 KD or SCR control hESCs before (PLURI) or after *in vitro* directed differentiation towards endoderm or neuroectoderm. Note that SOX2 is both a pluripotency and a neuroectoderm marker. For each gene, significant differences vs both SCR sh1 and SCR sh2 in the same condition are reported (two-way ANOVA, only the highest p is shown; n=3).

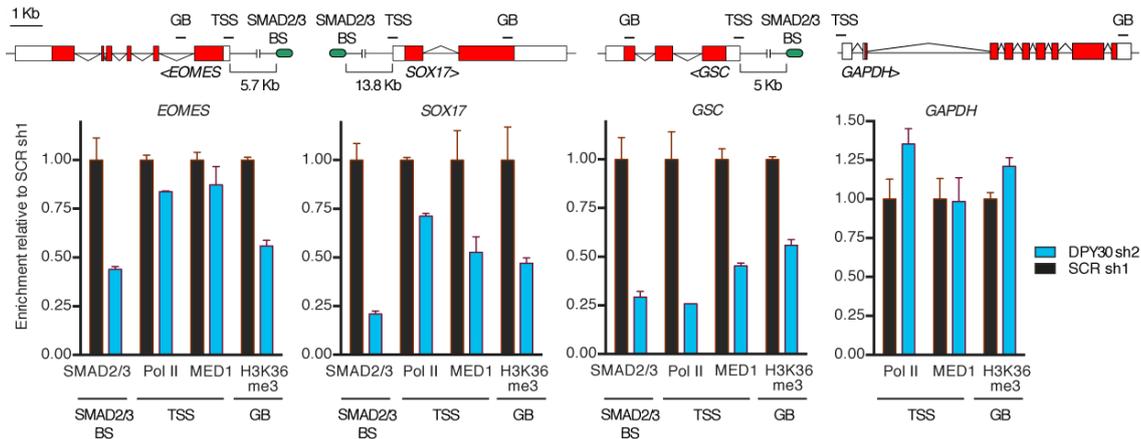


Figure 3.23. Transcription of Activin/Nodal target genes during endoderm differentiation of DPY30 knockdown cells.

ChIP-qPCR for SMAD2/3, RNA polymerase II (Pol II) or Mediator (MED1) in DPY30 KD or SCR control hESCs after endoderm differentiation. The location of the primers used for qPCR is indicated in the schematic above the graphs (TSS: transcription start site; GB: gene body; SMAD2/3 BS: SMAD2/3 binding site). The results are representative of two independent experiments.

Impaired endoderm specification of DPY30 KD hESCs was confirmed at the protein level by Western blot and flow cytometry (Figure 3.24). Of note, this defect resulted into the inability to further differentiate into mature endoderm-derived lineages such as liver and pancreas (Figure 3.25). Finally, the reduced expression of endoderm markers in DPY30 KD hESCs was associated with impaired deposition of H3K4me3 (Figure 3.26), thereby confirming that transcriptional inhibition provoked by the decrease in DPY30 expression is associated with epigenetic deregulation.

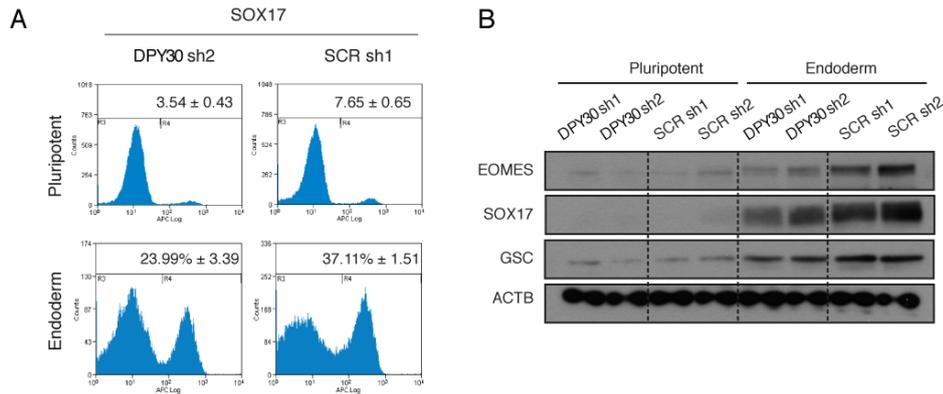


Figure 3.24. Expression of endoderm marker proteins during differentiation of DPY30 knockdown cells.

(A) Representative flow cytometry histograms for SOX17 in DPY30 KD or SCR control hESCs before (pluripotent) or after endoderm differentiation. The gates used to define positive cells are shown (based on secondary only staining controls). The average percentages ± SEM of positive cells from three biological replicates are reported. (B) Western blots for the indicated endoderm markers in cells differentiated as described in A. ACTB (β -actin): loading control.

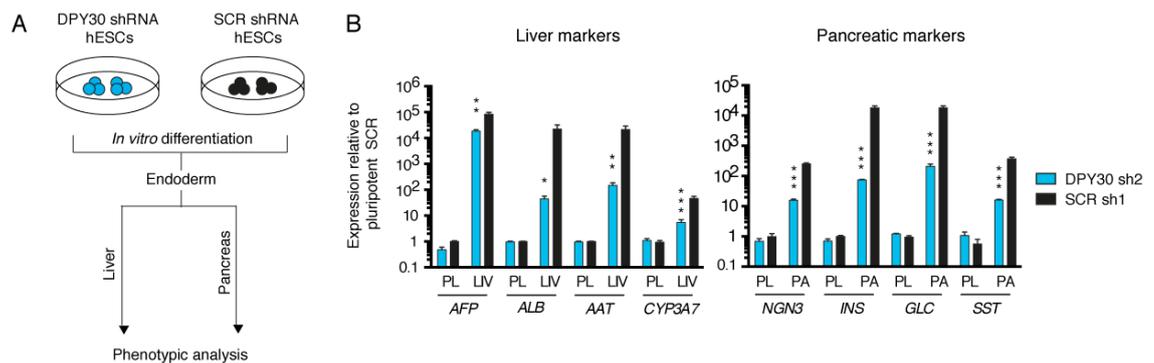


Figure 3.25. Differentiation of DPY30 knockdown cells into mature endodermal cell types.

(A) Schematics of the experimental approach. (B) Gene expression qPCR in DPY30 KD or SCR control hESCs either before (PL, pluripotent) or after liver (LIV) or pancreas (PA) differentiation. For each gene, significant differences vs SCR in the same condition are reported (two-way ANOVA; n=3).

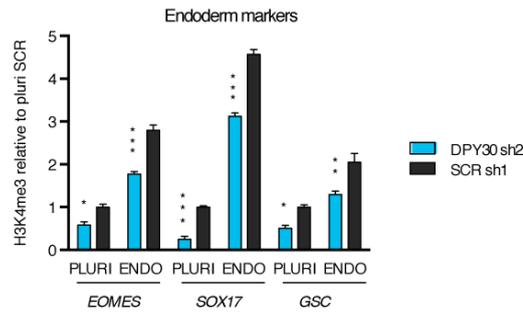


Figure 3.26. H3K4me3 deposition during endoderm differentiation of DPY30 knockdown cells.

ChIP-qPCR for H3K4me3 in DPY30 KD or SCR control hESCs before (PLURI) or after endoderm differentiation. Significant differences vs SCR sh1 in the same condition are reported (two-way ANOVA; n=3). qPCR was performed using primer pairs specific for the promoters of the indicated genes (Figure 3.32).

Overall, these results indicated that loss of H3K4me3 on developmental genes impairs hPSC mesendoderm differentiation potency, thus highlighting the functional importance of the correct deposition of this epigenetic mark for early cell fate decisions.

3.3.5 SMAD2/3 cooperates with NANOG to recruit H3K4 methyltransferases on Activin/Nodal-responsive genes

Decrease in DPY30 expression in hESCs not only recapitulated the effect of Activin/Nodal inhibition, but also closely mimicked the consequence of NANOG knockdown (Vallier, Mendjan, et al. 2009). Indeed, downregulation of NANOG in hESCs impaired expression of pluripotency and mesendoderm markers, while it induced neuroectoderm genes (Figure 3.27). Furthermore, microarray analyses demonstrated a significant similarity between the transcriptional responses to NANOG KD, the inhibition of Activin/Nodal signalling for 48h, and DPY30 KD (Figure 3.28; highest hypergeometric $p < 1E-300$ and $1E-700$ respectively). While these results were in agreement with the known function of NANOG as a SMAD2/3 cofactor (Chapter 1.7.4.2), they also suggested a previously unknown overlap between the DPY30- and NANOG-dependent transcriptional networks. Supporting these notions, co-immunoprecipitation experiments showed that NANOG could be found in protein complexes containing DPY30 and SMAD2/3, and that *vice versa* DPY30 and SMAD2/3 co-immunoprecipitated with NANOG (Figure 3.29 and Figure 3.32). Moreover, NANOG KD hESCs displayed lower levels of H3K4me3 on pluripotency and endoderm genes bound by NANOG and SMAD2/3 (Figure 3.30), further suggesting a functional link between these factors and DPY30. Considered together, these results suggest the existence of a complex between SMAD2/3, NANOG and DPY30, which we will refer to as “S/N/D”.

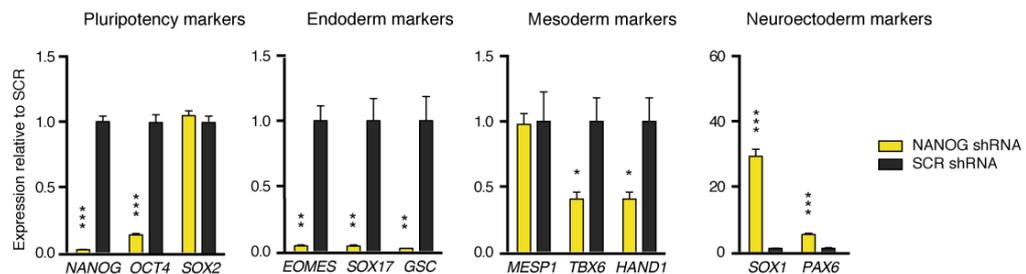


Figure 3.27. Effect of NANOG knockdown on the expression of pluripotency and differentiation markers.

Gene expression qPCR in NANOG KD and SCR control hESCs. Note that *SOX2* is both a pluripotency and a neuroectoderm marker. For each gene, significant differences vs SCR are reported (t-tests; $n=3$).

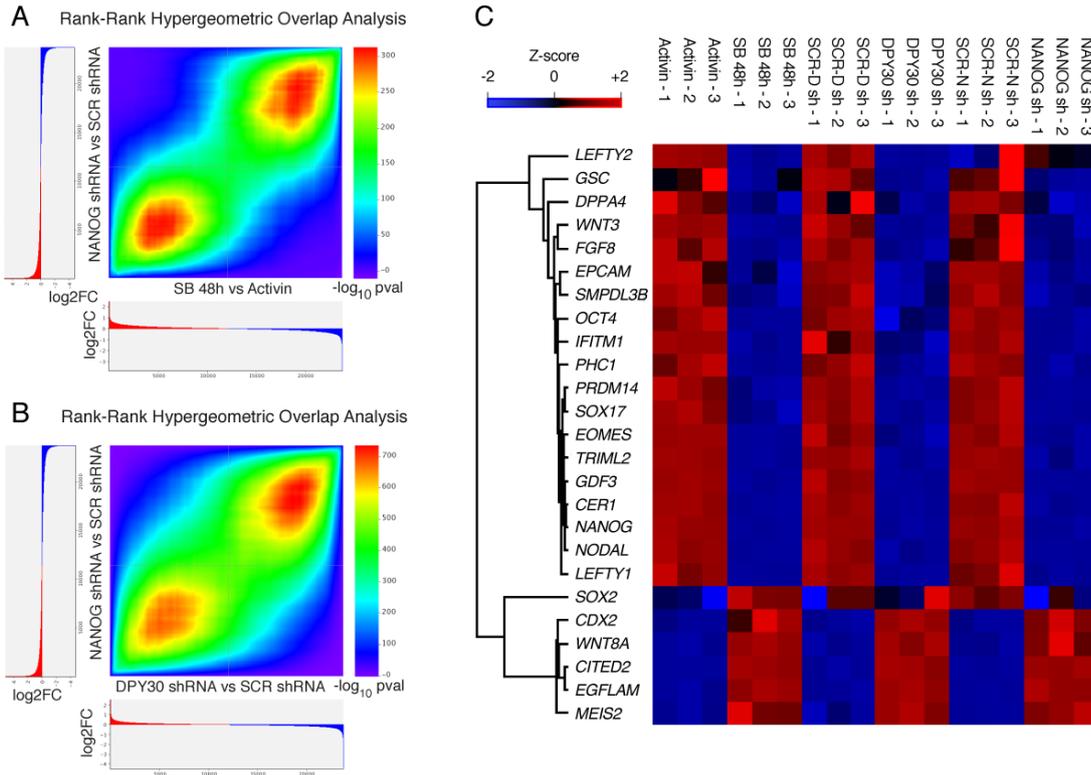


Figure 3.28. Global similarities between the transcriptional responses induced by the knockdown of NANOG, the knockdown of DPY30, and the inhibition of Activin/Nodal signalling.

(A-B) RRHO for genes ranked by their differential expression after NANOG KD, DPY30 KD, or inhibition of Activin/Nodal for 48h with SB. Microarrays were performed in collaboration with the Wellcome Trust Sanger Institute microarray facility. (C) Heatmap reporting changes in gene expression of selected SMAD2/3 target genes (Figure 3.10). Z-scores were separately calculated for each experiment.

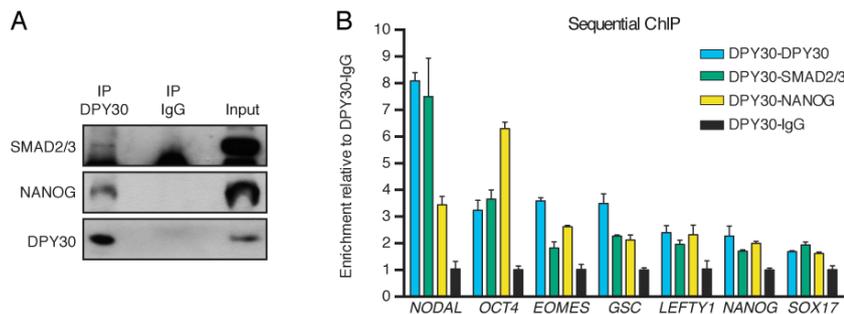


Figure 3.29. SMAD2/3, NANOG and DPY30 form a chromatin-interacting complex.

(A) Western blots of DPY30 or control (IgG) immunoprecipitations from nuclear extracts of hESCs. Input is 5% of the material used for IP. (B) Sequential ChIP-qPCR for DPY30 followed by SMAD2/3, NANOG, or control (IgG) ChIP. qPCR was performed using primer pairs specific for SMAD2/3 binding sites onto the indicated genes (Figure 3.32).

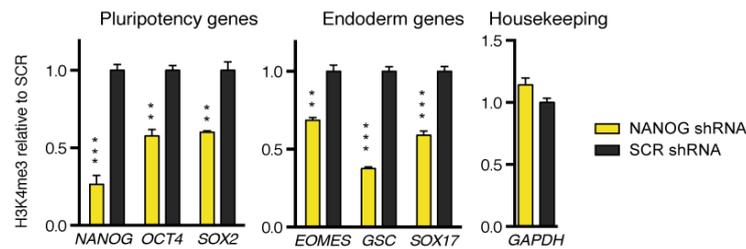


Figure 3.30. Effect of NANOG knockdown on the levels of H3K4me3 onto Activin/Nodal target genes.

ChIP-qPCR for H3K4me3 in NANOG KD or SCR control hESCs. For each gene, significant differences vs SCR are reported (t-test; $n=3$). qPCR was performed using primer pairs specific for the promoters of the indicated genes (Figure 3.32).

To further validate this hypothesis, we performed sequential ChIP experiments demonstrating that DPY30 co-binds both with SMAD2/3 and NANOG onto chromatin regions associated to key regulators of hESC pluripotency and differentiation (Figure 3.29). We then investigated whether genomic binding of DPY30 and SMAD2/3 requires the presence of NANOG in the complex. Strikingly, NANOG knockdown impaired binding of both DPY30 and SMAD2/3 on pluripotency and endoderm genes (Figure 3.31), suggesting that NANOG could be recruiting such factors. On the other hand, inhibition of SMAD2/3 binding with 2h of SB treatment also resulted in loss of both NANOG and DPY30 binding (Figure 3.31), thus indicating that SMAD2/3 was also necessary for their recruitment. Finally, we evaluated the formation of the S/N/D complex by co-immunoprecipitation in the presence or absence of NANOG expression. Interestingly, SMAD2/3 still bound to DPY30 even in the absence of NANOG (Figure 3.31), suggesting that the expression of NANOG is dispensable for the association of SMAD2/3 and DPY30. Taken together, these results implied that SMAD2/3 and NANOG depend on each other to efficiently bind Activin/Nodal-responsive genes and to recruit DPY30 on genes characterizing the pluripotent state of hESCs.

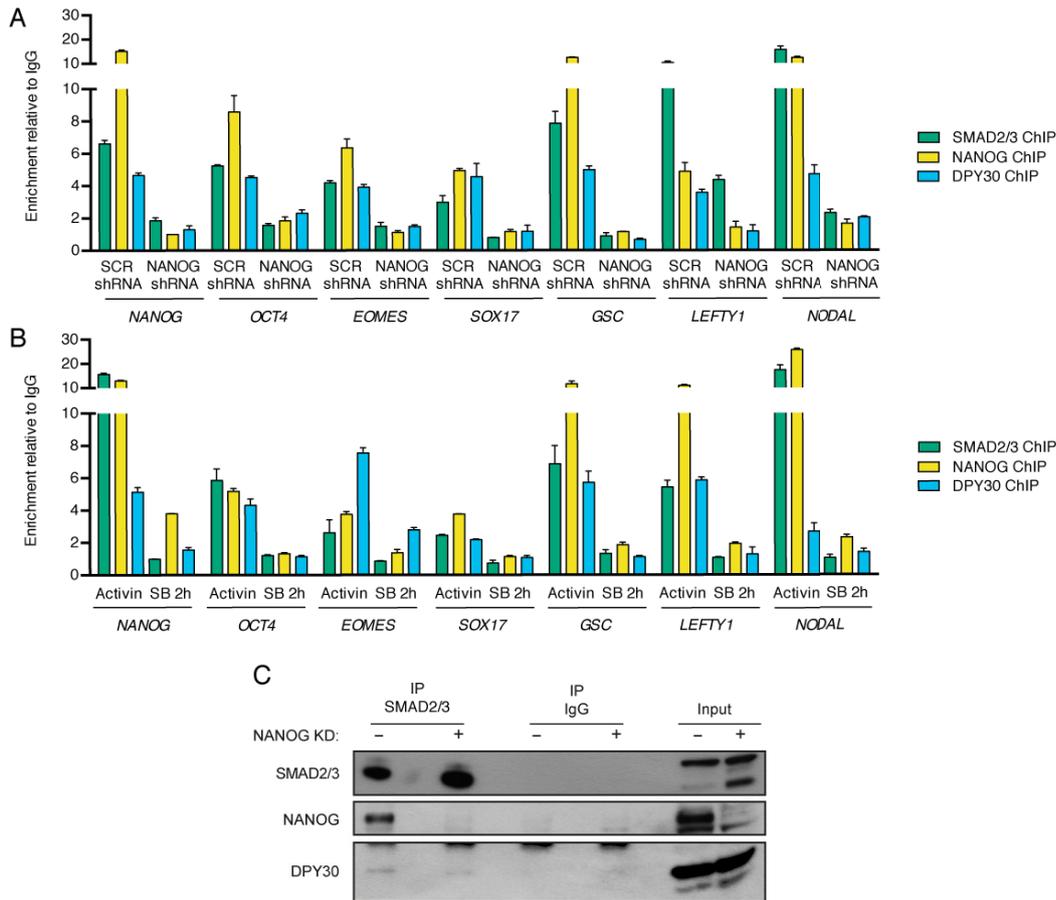


Figure 3.31. SMAD2/3 and NANOG recruit DPY30 onto their genomic targets.

(A-B) ChIP-qPCR for the indicated proteins in NANOG KD or SCR control hESCs (A), or in hESCs before and after inhibition of Activin/Nodal with SB for 2h (B). Results are representative of three independent experiments, and the location of primers used is shown in (Figure 3.32). (C) Western blots of SMAD2/3 or control (IgG) immunoprecipitations from nuclear extracts of NANOG KD or SCR control hESCs. Input is 5% of the material used for IP.

We then investigated whether S/N/D binding correlated with histone methyltransferase recruitment. First, we confirmed that COMPASS complexes interact in an Activin/Nodal-dependent manner with chromatin regions bound by S/N/D, as measured by ChIP for WDR5 before and after inhibition of Activin/Nodal signalling for 2h (Figure 3.32). Furthermore, we observed that the COMPASS catalytic subunit MLL2/KMT2B bound to most of the S/N/D target loci we tested in an Activin/Nodal-dependent manner, while we could not detect any binding for SETD1A or MLL1/KMT2A (Figure 3.32). Accordingly, WDR5 and MLL2/KMT2B appeared to be part of the S/N/D complex, as shown by co-immunoprecipitation (Figure 3.32). Interestingly, knockdown of DPY30 did not result in impaired recruitment of WDR5 or MLL2/KMT2B onto S/N/D target regions (Figure 3.33). This observation suggested that the phenotypic effects following DPY30 knockdown were primarily the results of an impaired H3K4 trimethylation activity of COMPASS complexes, rather than being due to impaired recruitment of such

factors onto the chromatin. This is in agreement with what reported in previous studies (Jiang et al. 2011; Yang et al. 2014; Ernst & Vakoc 2012). It was recently shown that DPY30 can also be found in the NURF nucleosome remodelling complex (van Nuland et al. 2013). We therefore decided to investigate the binding profile of the NURF catalytic subunit BPTF on Activin/Nodal target genes. In contrast to COMPASS complexes, BPTF enrichment on S/N/D bound regions was only limited and independent from the presence of Activin/Nodal signalling (Figure 3.32). This argued against an important role for the NURF complex in the regulation of S/N/D targets.

Overall, these experiments supported a model by which SMAD2/3 and NANOG binding onto Activin/Nodal target genes induces recruitment of histone methyltransferases that regulate H3K4me3 deposition.

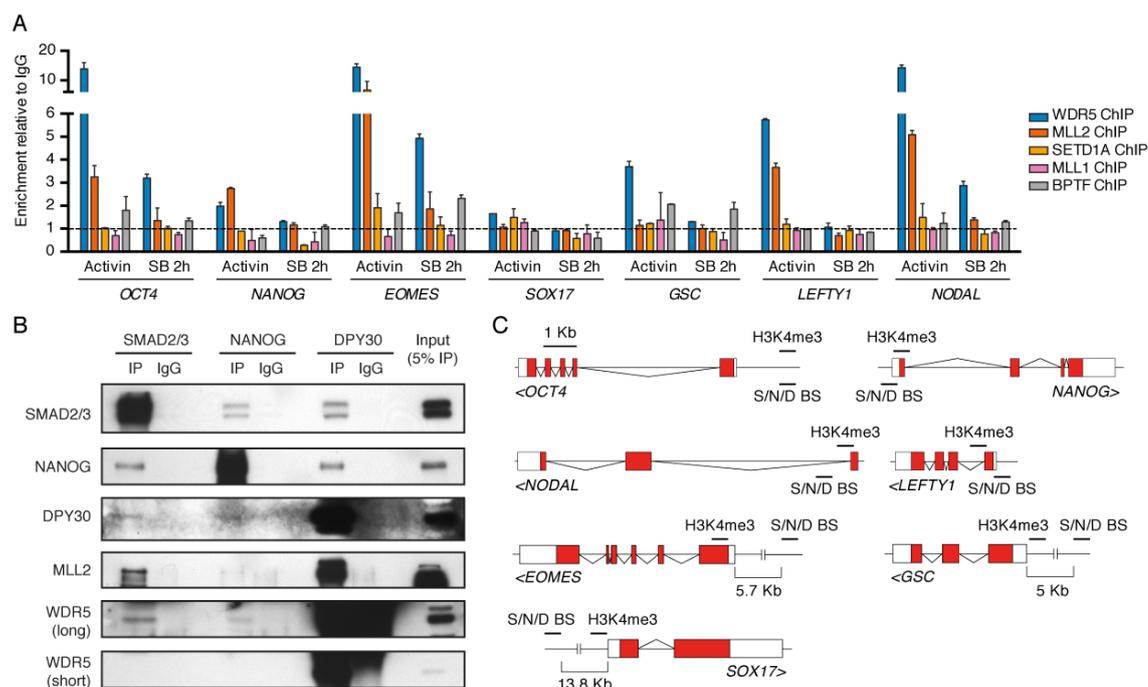


Figure 3.32. SMAD2/3 recruits the MLL2 complex onto its genomic targets.

(A) ChIP-qPCR for the indicated proteins in hESCs before and after inhibition of Activin/Nodal with SB for 2h. The location of primers used for qPCR is shown in panel C. The results are representative of three independent experiments. The dotted line shows the level of enrichment in the control IgG ChIP (no enrichment). (B) Western blots of SMAD2/3, NANOG, DPY30 or control (IgG) immunoprecipitations from nuclear extracts of hESCs. Input is 5% of the material used for IP. Note that the strong background signal for the WDR5 blot in the DPY30 IP and matched rabbit IgG control is due to the reactivity of the anti-rabbit secondary antibody against the rabbit immunoglobulins used for the IP. A short exposure that shows WDR5 binding to DPY30 is provided. (C) Schematic representation of the location of ChIP-qPCR primers used for histone marks ChIP (“H3K4me3”; also used for H3K4me2, H3K4me1, and H3 ChIP) or for transcription factors and chromatin modifiers ChIP (“S/N/D BS”: SMAD2/3, NANOG, and DPY30 binding site; also used for WDR5, MLL2/KMT2B, MLL1/KMT2A, SETD1A, and BPTF ChIP).

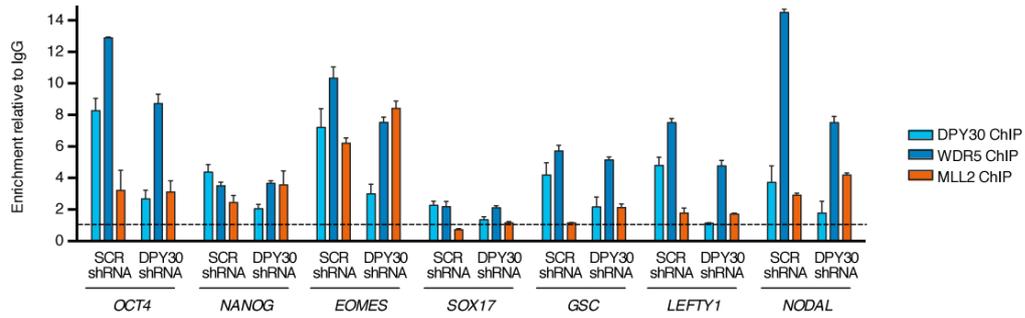


Figure 3.33. Binding of the MLL2 complex to SMAD2/3 target genes does not require DPY30.

ChIP-qPCR for the indicated proteins in DPY30 KD or SCR control hESCs. The results are representative of three independent experiments. The dotted line shows the level of enrichment in the control IgG ChIP (no enrichment).

3.3.6 Cooperation between DPY30, NANOG, and SMAD2/3 is necessary for H3K4me3 deposition on a core transcriptional network

Having established a model for the functional interaction between SMAD2/3, NANOG, and DPY30 on a subset of Activin/Nodal responsive genes, we decided to extend our analysis to a genome-wide scale by performing H3K4me3 ChIP-seq in DPY30 KD and NANOG KD hESCs (Figure 3.34). While both conditions induced a large number of significant differences, deregulation of H3K4me3 was still specific to certain loci: out of 31923 peaks in DPY30 KD hESCs 6482 (20.3%) were decreased and 1394 (4.4%) were increased, while 4028 (12.1%) and 1510 (4.6%) of the 32642 peaks in NANOG KD hESCs were respectively decreased and increased (Figure 3.34). Interestingly, regions with decreased H3K4me3 after DPY30 or NANOG knockdown were associated with genes involved in development of tissues from all germ layers (GREAT analysis). This is in line with an important role for DPY30 and NANOG in the regulation of developmental processes. Importantly, we observed a large overlap between the H3K4me3 peaks downregulated after DPY30 or NANOG knockdown (2062 peaks), implying that these two factors control a large common subset of genes (Figure 3.35). Nevertheless, there were a number of regions specifically affected by either protein. This was particularly striking for DPY30 (4336 peaks), in agreement with its general role in promoting the activity of COMPASS complexes (Jiang et al. 2011). These observations demonstrated that NANOG and DPY30 are required to maintain H3K4me3 on a large but specific subset of genomic regions involved in early cell fate decisions.

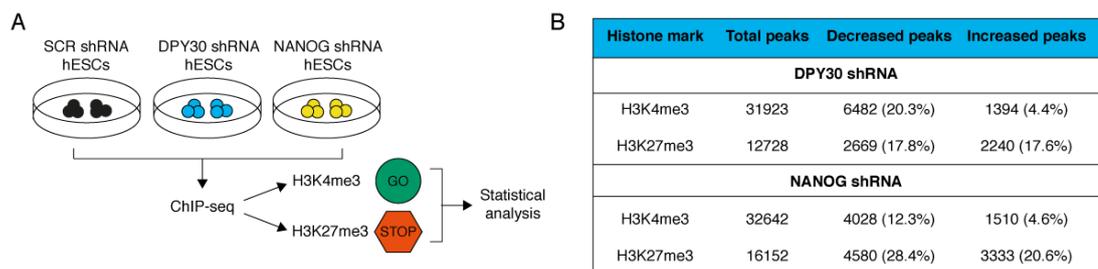


Figure 3.34. ChIP-seq for H3K4me3 and H3K27me3 following knockdown of DPY30 or NANOG.

(A) Schematics of the experimental approach. ChIP-seq were performed in collaboration with the Wellcome Trust Sanger Institute next-generation sequencing facility. (B) Results of the statistical analysis of ChIP-seq data. ChIP-seq analysis was done in collaboration with Dr Pedro Madrigal.

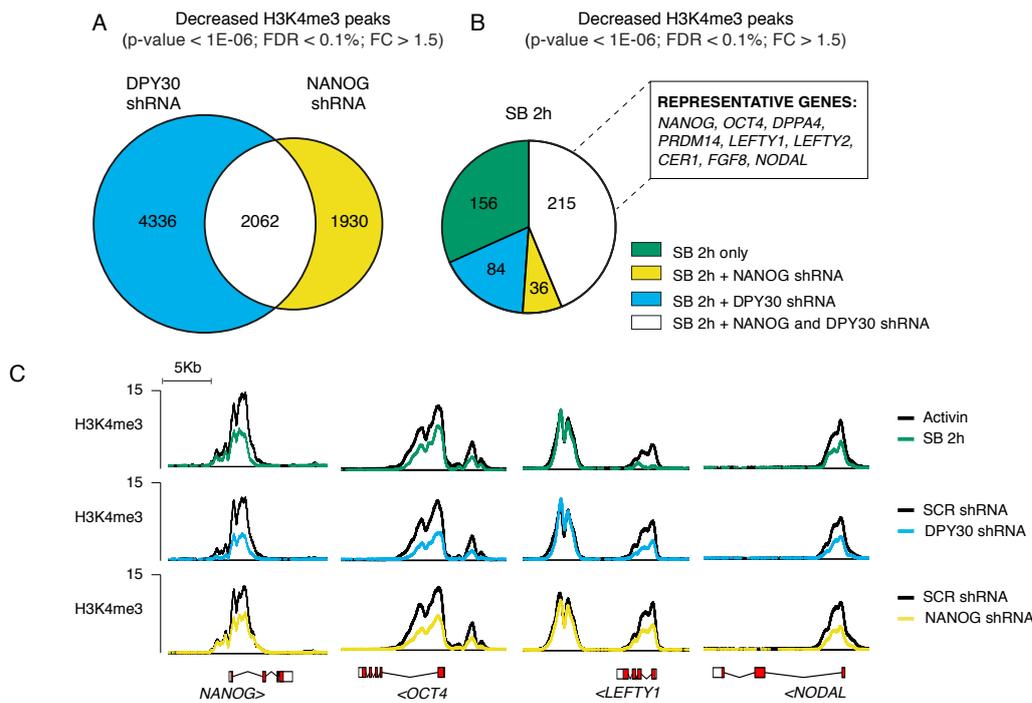


Figure 3.35. Similarities between the H3K4me3 changes induced by the knockdown of NANOG, the knockdown of DPY30, and the inhibition of Activin/Nodal signalling.

(A) Overlap between H3K4me3 peaks significantly downregulated after DPY30 KD and NANOG KD. (B) Proportion of H3K4me3 peaks significantly downregulated after 2h of Activin/Nodal inhibition with SB that were also similarly affected by DPY30 KD, NANOG KD, or both treatments. Representative genes associated to peaks downregulated in all conditions are reported. (C) Examples of ChIP-seq results for H3K4me3 on selected SMAD2/3 target genes. Histograms represent read-enrichments normalized by million mapped reads and size of the library.

Interestingly, a large proportion (44%) of the H3K4me3 peaks decreased after 2h of SB significantly overlapped with those downregulated after both NANOG and DPY30 knockdown (Figure 3.35; $p=4.11E-206$ as calculated with MULTOVL; Aszódi 2012). These regions were associated to genes that included the pluripotency markers *NANOG*, *OCT4*, *DPPA4*, and *PRDM14*, and the mesendoderm regulators *LEFTY1*, *WNT3*, *CER1*, *FGF8* and *NODAL* (Figure 3.35). Therefore, while the long-term knockdown of NANOG and DPY30 induced a larger deregulation of H3K4me3 levels compared to the one resulting from 2h of inhibition of Activin/Nodal signalling, there appears to be a core set of important genes that are directly controlled by SMAD2/3, NANOG and DPY30.

On the other hand, the large number of H3K4me3 peaks decreased by more than 50% after both DPY30 KD and NANOG KD but not after 2h of SB (1847 peaks) could be controlled independently of SMAD2/3. However, the expression of several genes associated to these peaks was significantly impaired after 48h of inhibition of Activin/Nodal signalling (Figure 3.36), arguing against a SMAD2/3-independent

regulation. Furthermore, H3K4me3 onto these genomic regions was modestly yet significantly decreased after 2h of SB treatment (Figure 3.37; note that this mild effect was not uncovered by our previous analysis that focused only on stronger changes of more than 50%). These observations suggested that certain genomic regions might have a slower kinetics of H3K4me3 loss upon Activin/Nodal inhibition. This was supported by ChIP-qPCR showing that the downregulation of H3K4me3 on several loci reached its maximum only after 48h of Activin/Nodal inhibition (Figure 3.37). Therefore, SMAD2/3, DPY30 and NANOG appear to regulate H3K4me3 on several genes in which strong decrease of H3K4me3 levels occurs only after chronic inhibition of Activin/Nodal signalling.

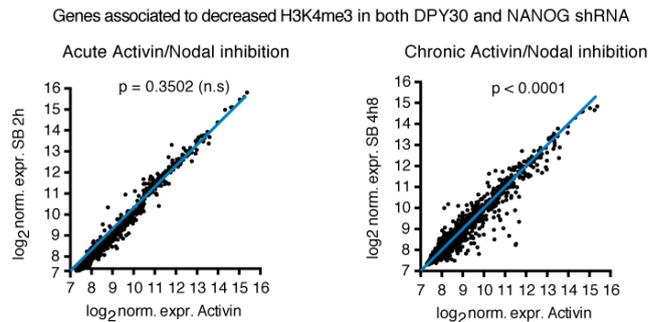


Figure 3.36. Correlation between H3K4me3 decrease after knockdown of NANOG and DPY30 and gene expression changes following inhibition of Activin/Nodal signalling.

Expression of the genes closest to H3K4me3 peaks significantly downregulated by more than 50% both after DPY30 KD and NANOG KD, but not decreased to the same extent after 2h of SB. The significance of differential expression is reported (Welch's t-test). Gene expression data is from the microarrays described in Chapter 3.3.2.

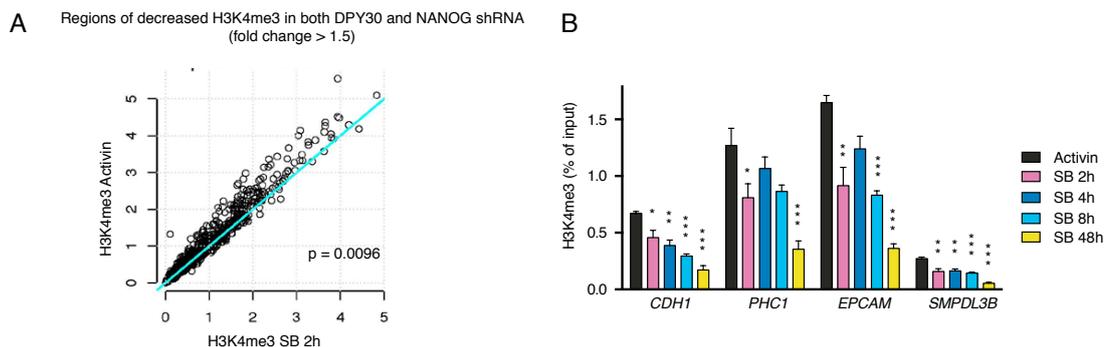


Figure 3.37. Progressive loss of H3K4me3 during chronic Activin/Nodal inhibition.

(A) Average normalized H3K4me3 read enrichment in three biological replicates before or after 2h of SB (Figure 3.2). Data refers only to H3K4me3 peaks described in Figure 3.36. The level of significant change vs Activin is reported (Welch's t-test). This analysis was performed in collaboration with Dr Pedro Madrigal. (B) ChIP-qPCR for H3K4me3 before and after inhibition of Activin/Nodal with SB for 2h, 4h, 8h, or 48h. The genes analysed belong to the group described in Figure 3.36. For each gene, significant changes vs Activin is reported (one-way ANOVA; $n=3$).

In addition to H3K4me3, we also performed ChIP-seq for H3K27me3 after DPY30 KD or NANOG KD (Figure 3.34). In sharp contrast with the results obtained following SB treatment for 2h (Chapter 3.3.1), we identified many regions of differential H3K27me3 in these conditions (2269 decreased and 2240 increased peaks out of 12728 for DPY30 KD; 4580 decreased and 3333 increased peaks out of 16512 for NANOG KD). Moreover, the regions where H3K27me3 changed after DPY30 or NANOG knockdown were not even mildly affected after 2h of SB (Figure 3.38). This further suggests that Activin/Nodal signalling might not be directly involved in the regulation of H3K27me3 in hESCs, while NANOG and DPY30 could have a more important role in this process.

In summary, these genome-wide results supported the model according to which SMAD2/3, NANOG, and DPY30 collaborate to maintain H3K4me3 on a large number of loci.

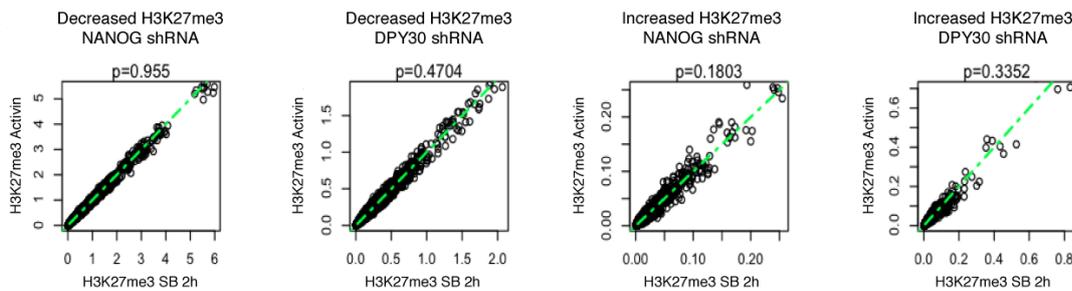


Figure 3.38. H3K27me3 is unaffected by acute Activin/Nodal inhibition. Average normalized H3K27me3 read enrichment in three biological replicates before and after inhibition of Activin with SB for 2h. Data refers to H3K27me3 peaks significantly downregulated or upregulated by more than 50% after NANOG KD or DPY30 KD. The level of significant change vs Activin is reported (Welch’s t-test). This analysis was performed in collaboration with Dr Pedro Madrigal.

3.3.7 DPY30, NANOG, and SMAD2/3 colocalize at the genome-wide level

We then investigated if the functional interactions described so far could imply collaborative mechanisms between SMAD2/3, NANOG, and DPY30 at a genome-wide level. For that, we performed ChIP-seq for DPY30 in control (SCR shRNA) hESCs and DPY30 KD hESCs. These analyses revealed DPY30 binding on 26387 genomic regions in control hESCs, while 6893 peaks were identified following DPY30 knockdown. Moreover, DPY30 ChIP-seq coverage was strongly impaired in DPY30 KD hESCs compared with control hESCs (Figure 3.39). Finally, DPY30 binding correlated with H3K4me3 deposition, as it was previously shown for mESCs (Figure 3.39; Jiang et al. 2011). Overall, these observations validated the specificity of the DPY30 antibody and confirmed the quality of the DPY30 ChIP-seq data.

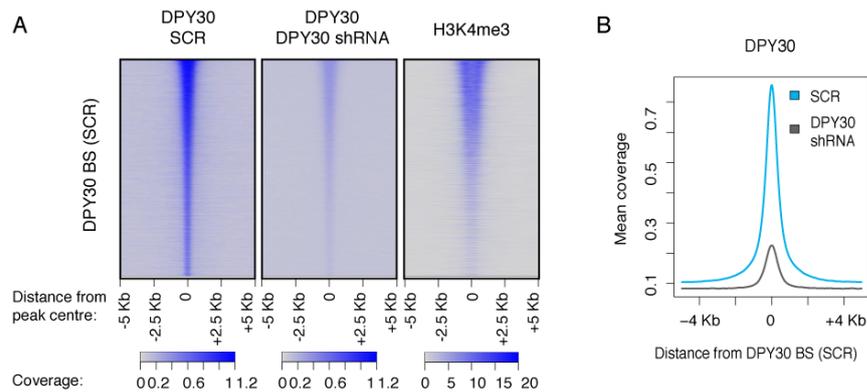


Figure 3.39. Validation of DPY30 ChIP-seq.

(A) Heatmaps of coverage for DPY30 ChIP-seq in control hESCs (SCR) and DPY30 KD hESCs (DPY30 shRNA), and for H3K4me3 in control hESCs. Coverage was calculated relative to the centre of DPY30 binding sites (BS) in control hESCs. ChIP-seq were performed in collaboration with the Wellcome Trust Sanger Institute next-generation sequencing facility, and ChIP-seq analyses were performed in collaboration with Dr Pedro Madrigal. (B) Mean ChIP-seq coverage plot for DPY30 in control and DPY30 KD hESCs.

We then compared the genome-wide binding of DPY30 with that of SMAD2/3 (Brown et al. 2011) and NANOG (ENCODE data, Bernstein et al. 2012). Interestingly, a large proportion of SMAD2/3 binding sites were co-occupied both by NANOG and DPY30 (Figure 3.40). While the colocalization of SMAD2/3 and NANOG is in agreement with previous reports (Mullen et al. 2011), these findings suggests that DPY30 is also present in this complex, in agreement with our earlier observations (Chapter 3.3.5). Moreover, NANOG, and DPY30 globally co-occupied SMAD2/3 binding sites associated with H3K4me3 regions responsive to Activin/Nodal signalling, including several examples of key developmental regulators such as *OCT4* and *NODAL* (Figure 3.41).

Overall, these experiments demonstrated that SMAD2/3, NANOG and DPY30 are found in close proximity at a genome-wide level, thereby reinforcing our hypothesis that these factors form a complex controlling H3K4me3 deposition on target genes regulated by Activin/Nodal signalling.

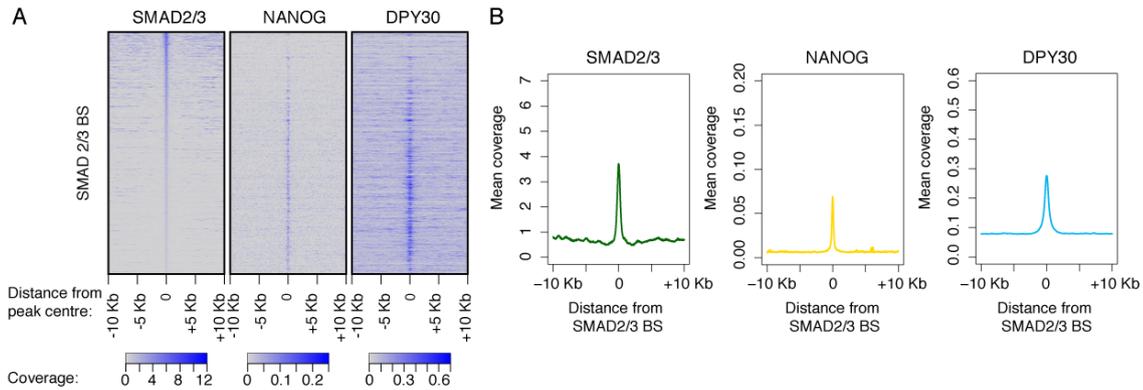


Figure 3.40. Global overlap between SMAD2/3, NANOG, and DPY30. (A) Heatmaps of coverage for SMAD2/3, NANOG, and DPY30 ChIP-seq relative to all SMAD2/3 peaks (14085 peaks, Brown et al. 2011). (B) Mean ChIP-seq coverage for all peaks represented in A. These analyses were performed in collaboration with Dr Pedro Madrigal.

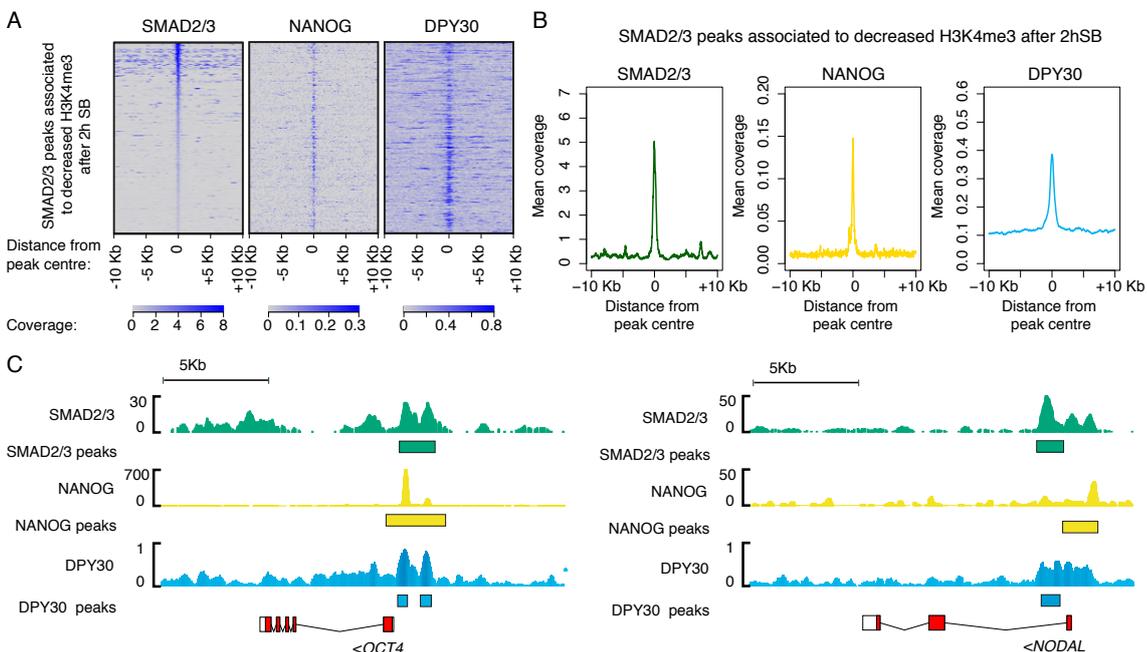


Figure 3.41. Binding of SMAD2/3, NANOG, and DPY30 on loci showing Activin/Nodal-dependent H3K4me3. (A) Heatmaps of coverage for SMAD2/3, NANOG, and DPY30 ChIP-seq relative to the SMAD2/3 peaks associated to H3K4me3 regions decreased after 2h of SB (530 peaks). (B) Mean coverage plots for the peaks considered in A. (C) Examples of ChIP-seq coverage (top) and peaks (bottom). Analyses in A and B were performed in collaboration with Dr Pedro Madrigal.

3.3.8 Dpy30 is necessary to maintain the pluripotent state of the post-implantation epiblast

To confirm the relevance of the mechanisms uncovered by our *in vitro* experiments, we decided to evaluate the function of Dpy30 during mouse embryonic development. For that, we took advantage of Dpy30 knockout mice generated by the Mouse Genetics Project of the Wellcome Trust Sanger Institute (Dpy30^{<tm1a(KOMP)Wtsi>} strain; Figure 3.42). Mice carrying heterozygous mutations were healthy, and we did not observe any obvious phenotype. However, viable homozygous mutants could not be recovered, and further analysis revealed that the absence of Dpy30 was embryonic lethal between E7.5 and E9.5 (Figure 3.43). Indeed, mutant embryos recovered at E6.5 displayed gross morphological abnormalities such as reduced size, impaired anterior-posterior patterning, and lack of a clearly distinguishable primitive streak (Figure 3.43). Accordingly, Dpy30 knockout embryos at E7.5 were severely developmentally delayed and undergoing resorption (Figure 3.43).

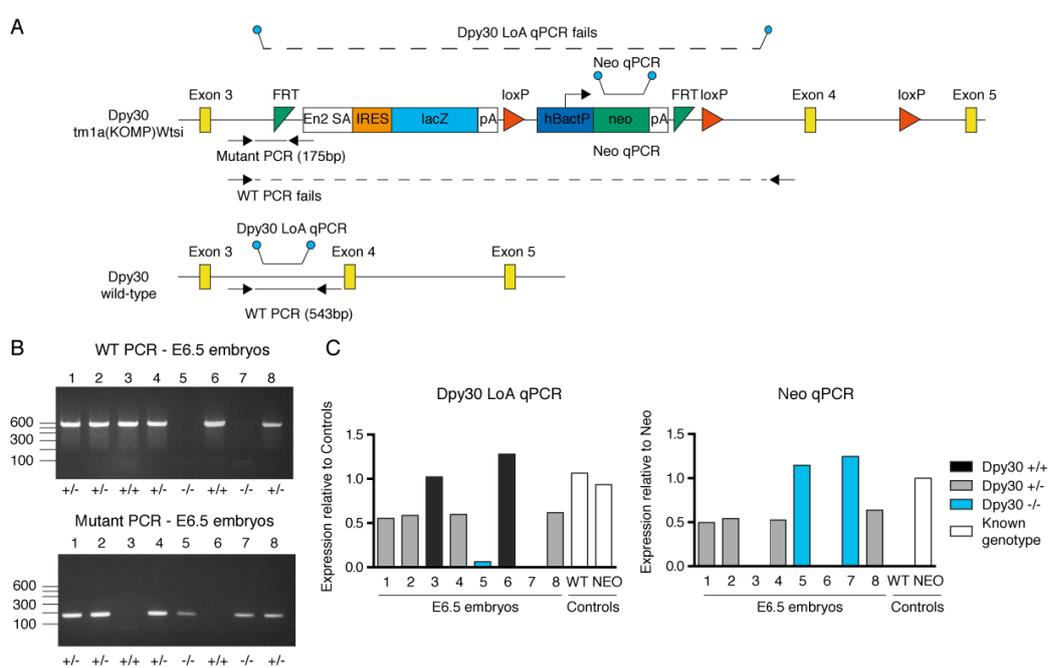


Figure 3.42. Genotyping of Dpy30 knockout mice.

(A) Schematics of the Dpy30 mutant and wild-type alleles. The location of PCR primers used for genotyping is shown. Arrows depict standard PCR, while ball-and-sticks indicate hydrolysis probes-based qPCR. FRT=flippase recognition target site; En2SA=engrailed-2 exon-2 splice acceptor; IRES=internal ribosome entry site; lacZ= β -galactosidase cDNA; pA=polyadenylation site; loxP=bacteriophage P1 Cre recombinase recognition site; hBactP=human β -actin promoter; Neo=Neomycin phosphotransferase II (Neomycin resistance). (B) Representative PCR genotyping of a litter of E6.5 embryos from a Dpy30 mutant heterozygous cross. The inferred genotypes are reported (+/- = heterozygous; +/+ = homozygous wild-type; -/- = homozygous mutant). (C) Representative qPCR-based genotyping on the same embryos as in B.

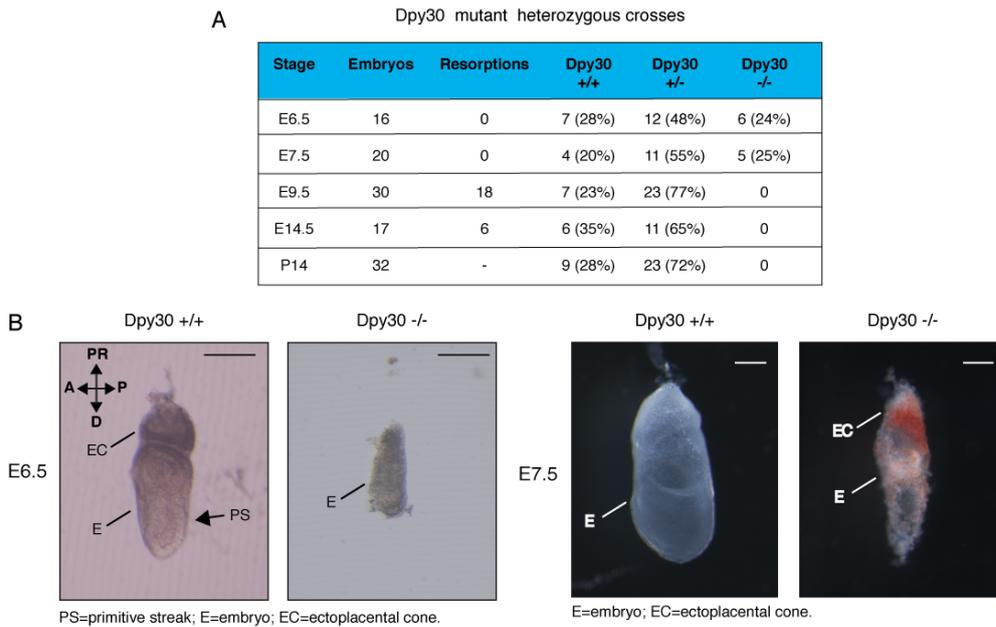


Figure 3.43. Role of Dpy30 during mouse post-implantation development.

(A) Genotyping results from Dpy30 mutant heterozygous crosses. Stages are embryonic (E) or postnatal (P) days. Genotyping of E9.5, E14.5, and P14 embryos was done in collaboration with the Mouse Genetics Project of the Wellcome Trust Sanger Institute. (B) Bright field images of wild-type (+/+) or Dpy30 knockout (-/-) embryos at E6.5 or E7.5. The anterior-posterior (A/P) and proximal-distal (PR/D) axes are shown. Scale bars: 100µm. Embryo dissections and morphological assessments were performed in collaboration with Dr Antonella Galli.

Importantly, absence of Dpy30 expression in homozygous mutant embryos was confirmed by qPCR (Figure 3.44), thus validating that the gene targeting strategy adopted for the generation of these mice resulted in Dpy30 loss-of-function. Interestingly, Dpy30 knockout embryos dissected at E6.5 showed prematurely reduced expression of epiblast markers (*Nanog* and *Oct4*) and failed to properly upregulate mesendoderm genes (*Eomes*, *Gsc* and *Brachyury*). On the other side, neuroectoderm markers were either unaffected (*Sox2*, *Sox1*) or even increased (*Dlx1* and *Hesx*; Figure 3.44). Therefore, we concluded that Dpy30 is necessary both to maintain the pluripotent state of the post-implantation epiblast, and to enable proper specification of the three germ layers *in vivo*. These results strikingly recapitulated the phenotype induced by the knockdown of DPY30 in hESCs, indicating that the molecular regulations we uncovered *in vitro* could also occur at the corresponding developmental stage during early embryogenesis (Figure 3.45).

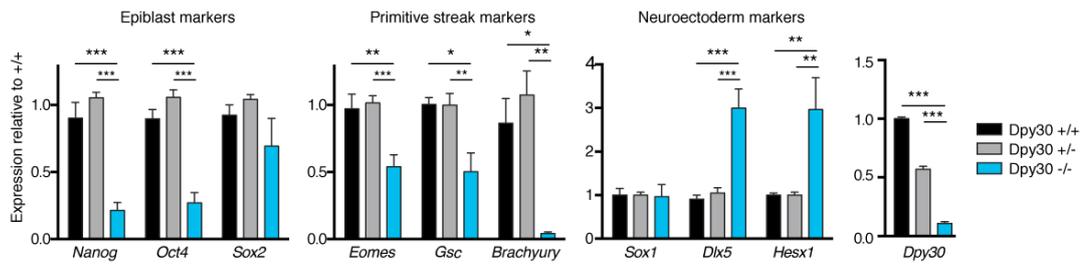


Figure 3.44. Expression of pluripotency and differentiation markers in DPY30 knockout embryos.

Gene expression qPCR in E6.5 embryos from *Dpy30* mutant heterozygous crosses. Note that *Sox2* is both an epiblast and a neuroectoderm marker. Significant differences vs *Dpy30* +/+ are reported (one-way ANOVA; n=6 for *Dpy30* +/+ and *Dpy30* -/-; n=11 for *Dpy30* +/-).

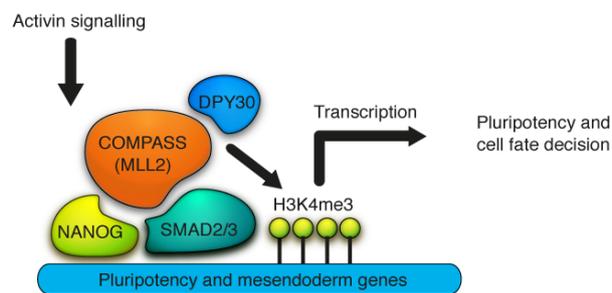


Figure 3.45. Proposed model for the functional interaction between SMAD2/3, NANOG, and DPY30-containing COMPASS complexes.

Schematics of the model we propose for Activin/Nodal-dependent epigenetic and transcriptional regulation of pluripotency and mesendoderm genes in hESCs and in the post-implantation mouse epiblast. Note that the interactions depicted in the model must be interpreted as functional ones rather than direct protein-protein interactions, as this aspect was not addressed in the experiments presented in this chapter.

3.4 Discussion

The results described in this chapter clarify the mechanism by which extracellular signals through the Activin/Nodal signalling pathway are converted into epigenetic and transcriptional regulations necessary for hPSC self-renewal and differentiation. For this, SMAD2/3 and NANOG recruit DPY30-containing COMPASS complexes onto specific genomic regions, where they sustain H3K4me3 deposition (Figure 3.45). This event is primarily required to maintain the expression of several pluripotency factors. Accordingly, inhibition of Activin/Nodal signalling results in a rapid decrease of both H3K4me3 and expression of important regulators, such as NANOG itself. NANOG is involved in a feed-forward regulatory network that maintains both its own expression and that of other pluripotency genes (such as OCT4), all of which are part of a broader SMAD2/3-dependent transcriptional network characterising hPSCs (Chapter 1.4.1; Mullen et al., 2011; Teo et al., 2011). As such, the prompt loss of H3K4me3 upon Activin/Nodal inhibition provides hPSCs with an efficient way to switch off the core pluripotency network, thus initiating the cascade of events necessary for timely exit from the pluripotent state. Importantly, endogenous inhibitors of Activin/Nodal signalling such as Lefty1, Lefty2 and Cer1 are known to promote neuroectoderm specification during mouse embryonic development (Meno et al. 1999; Perea-Gomez et al. 2002; Chapter 1.1.3). This suggests that the regulations we uncovered in hPSCs might explain physiological events that take place also *in vivo*.

Interestingly, the swift decrease of H3K4me3 onto selected loci upon Activin/Nodal inhibition argues against a mechanism of passive loss by gradual dilution following cell division, and suggests that an active process might be responsible. Likely candidates for this are the H3K4-specific histone demethylases of the JARID1 family. Indeed, the profile of histone methylation loss after inhibition of Activin/Nodal signalling is consistent with sequential H3K4me3 and H3K4me2 demethylation that is mediated by this class of enzymes (Cloos et al. 2008). In this light, it is of note that knockdown in Jarid1B in mESCs impairs silencing of pluripotency genes during mESC differentiation, thus interfering with neuroectoderm specification (Schmitz et al. 2011). Therefore, the epigenetic state of the core pluripotency network appears to be tightly controlled by extracellular signals that intervene on the dynamic competition of histone methylation writers and erasers. On the other side, we also observed a large number of genomic regions where loss of H3K4me3 after inhibition of Activin/Nodal follows a slower

kinetics. It is possible that alternative and less efficient mechanisms of H3K4me3 removal are in place onto such loci and/or that some of these effects might represent indirect secondary targets. Overall, multiple mechanisms could contribute to generate gene subsets that respond to the inhibition of Activin/Nodal signalling with different kinetics of epigenetic and transcriptional changes.

Of note, our results show that Activin/Nodal signalling controls H3K4me3 also on “bivalent genes” known to be master regulators of mesendoderm specification. Inhibition of Activin/Nodal, NANOG knockdown, and DPY30 knockdown are all associated with decreased H3K4me3 on these loci, which correlates with impaired mesendoderm differentiation potency. These findings show that H3K4me3 is actively maintained onto certain bivalent genes. Moreover, they suggest that the deposition of these positive histone marks could be indispensable for proper induction of differentiation, in agreement with the proposed function of bivalency (Chapter 1.5.2.3). Therefore, Activin/Nodal-dependent chromatin modifications in undifferentiated hPSCs might not only maintain self-renewal, but also prepare the conditions for mesendoderm differentiation to take place. However, this notion was not directly demonstrated by the results presented in this chapter. Indeed, since DPY30 knockdown was achieved in a stable and long-term fashion, it was not possible to understand whether the impaired differentiation of DPY30 knockdown hESCs was due to an essential function of DPY30 in the undifferentiated state, during endoderm specification, or both. Experiments that rely, for instance, on inducible knockdown of DPY30 would be required to clarify this aspect. Accordingly, Chapter 4 will explore this problem in more detail.

H3K4me3 can be deposited after the assembly of the basal transcription machinery *via* Pol II-dependent induction of H2B monoubiquitination and the subsequent recruitment of COMPASS complexes (Lee et al. 2007). Therefore, H3K4me3 is often regarded as a marker of active transcription. Nevertheless, it is established that in certain contexts deposition of H3K4me3 can drive gene expression, as shown for instance by the activation of Hox genes by Menin-Mll1/2 during mouse development (Wang et al. 2009). In this light, our results suggest that presence of H3K4me3 on Activin/Nodal targets is not only a consequence of the level of transcription, but also a causal event that directly influences gene expression. Indeed, the quick decrease of H3K4me3 on Activin/Nodal target genes after acute SB treatment precedes gene expression downregulation. Moreover, knockdown of the epigenetic remodeller DPY30 is sufficient

to mimic the transcriptional effects of Activin/Nodal inhibition. Several epigenetic readers are known to link H3K4me3 with various factors that can promote gene expression, including the TFIID, SAGA, and NuRF/BPTF complexes (Vermeulen et al. 2010). Thus, we speculate that H3K4me3 deposition could control transcription of Activin/Nodal targets through these interactions. However, more in depth functional analyses will be necessary to establish such a mechanistic link.

The role of H3K4me3 in pluripotency and differentiation has been a matter of debate (Chapter 1.5.2.3). Indeed, reports in mESCs showed that while knockdown of Dpy30 or Rbbp5 only impairs differentiation without affecting the expression of pluripotency genes, knockdown of Wdr5 results in a dramatic loss of self-renewal (Jiang et al. 2011; Ang et al. 2011). Additionally, these studies were limited to *in vitro* experiments that did not test the relevance of H3K4me3 during *in vivo* development, nor provided evidence for an evolutionary conserved role of H3K4me3 in mice and humans. This chapter addressed both these aspects: by combining both experiments in hESCs and *in vivo* analyses in the mouse, we suggest a fundamental role of H3K4me3 in pluripotency and cell fate decisions. Interestingly, our experiments showed that DPY30 has different functions in human and mouse ESCs, which might be explained by the distinct pluripotent states of these cell types (Chapter 1.2.3). Indeed, hESCs are in a primed pluripotent state phenotypically related to the post-implantation epiblast, while mESCs resemble naïve pluripotent cells of the ICM. Thus, deposition of H3K4me3 could be of prominent importance specifically in the primed pluripotent state, when Activin/Nodal signalling orchestrates the core pluripotency network. Accordingly, our genetic study in the mouse shows that Dpy30 is essential only for post-implantation development. Further exploration of this aspect will benefit from the study of conditional Dpy30 knockout mice at various embryonic stages, from *in vitro* analyses of the function of Dpy30 in mESCs versus mEpiSCs, and from similar experiments in conventional versus naïve hPSCs.

It is worth highlighting that the phenotype of Dpy30 knockout mice closely recapitulates the one of embryos mutant for Nodal or Smad2/3 (Brennan et al. 2001; Camus et al. 2006; Dunn et al. 2004). Indeed, absence of Nodal signalling results in premature loss of pluripotency markers in the epiblast, impaired gastrulation, and ectopic expression of neuroectoderm regulators. Therefore, we propose that the epigenetic controls imposed by Activin/Nodal signalling in hPSCs might also take place in the post-implantation

epiblast. Importantly, a more formal demonstration of this notion will require monitoring the levels of H3K4me3 onto Nodal target genes in the epiblast of Dpy30 knockout embryos. Indeed, it is possible that the lethal phenotype induced by loss of Dpy30 might be at least partially Nodal-independent.

Of note, a recent report showed that Setd1a is required for pluripotency specifically in the epiblast and not in the ICM (Bledau et al. 2014). Moreover, knockout of Setd1a results in reduced expression of Oct4 in the epiblast at E6.5, growth retardation, and embryonic lethality at around E7.5, thus mimicking the crucial features of Dpy30 knockout mice. Considered together with the fact that SETD1A interacts with SMAD2/3 in hPSCs, these findings suggest that SETD1A could be at least partially responsible for Activin/Nodal-dependent H3K4me3 in primed pluripotent cells. While we could not demonstrate localization of SETD1A onto the SMAD2/3 binding sites we tested in the current study, this is an aspect that should be further explored in the future. Of the other members of the COMPASS family that have been knocked-out in mice, Setd1b and Mll1 have late developmental phenotypes (Bledau et al. 2014; Yagi et al. 1998) and are therefore unlikely to be involved in the regulations we uncovered. On the other hand, Mll2 knockout embryos die shortly after gastrulation (Glaser et al. 2009; Bledau et al. 2014), and Mll2 regulates H3K4me3 on bivalent promoters in mESCs (Denissov et al. 2014; Hu, Garruss, et al. 2013). This suggests that Mll2 might be “priming” germ layer differentiation during early development. Interestingly, our data showed that MLL2/KMT2B functionally interacts with SMAD2/3. Therefore, we speculate that this methyltransferase might regulate the deposition of H3K4me3 onto bivalent regions controlled by Activin/Nodal signalling, while SETD1A might be important to maintain expression of active genes, such as pluripotency markers.

As a final general remark, we believe that the interconnection between Activin/Nodal signalling and SMAD2/3-dependent deposition of H3K4me3 on master regulators of cell fate could be relevant for many contexts other than the control of pluripotency (Chapter 6.1.3). Indeed, extracellular signalling converging on SMAD2/3 is involved at several stages of development, physiology, and disease (Massagué 2012). Similarly, H3K4me3 deposition has been demonstrated in a large number of cell types (Ruthenburg et al. 2007). Based on this hypothesis, Chapter 4.3.6 will explore the function of DPY30 during the specification of multiple lineages from hPSCs.

4 DEVELOPMENT OF AN OPTIMIZED INDUCIBLE KNOCKDOWN METHOD

4.1 Statement of source

The data and text described in this chapter are largely based upon sections of the following first author manuscript written by the author of this dissertation. Therefore, some segments have been taken *verbatim* or with minor changes from this source.

Bertero A, Pawlowski M, Ortmann D, Snijders K, Yiangou L, Cardoso de Brito M, Brown S, Bernard WG, Cooper JD, Giacomelli E, et al. Optimized inducible shRNA and CRISPR/Cas9 platforms to study human development. Manuscript in review at *Development*

4.2 Hypothesis

As introduced in Chapter 1.8, while hPSCs are an invaluable tool to model human development, functional analyses of complex molecular mechanisms require efficient conditional manipulation of gene expression during hPSC differentiation. For instance, as discussed in Chapter 3.4, the use of stable DPY30 knockdown hESCs did not allow us to tease apart the function of DPY30 in the undifferentiated state or during differentiation.

To address this limitation, we decided to generate hESCs carrying an inducible shRNA against DPY30. However, previous experience from our laboratory had shown that the generation of inducible knockdown hESCs using available techniques based on random integration of the TET-OFF system is poorly efficient and reproducible (Chapter 1.8). As a result, this goal had never been successfully accomplished in our group. This exemplifies how the use of hPSCs to model human development is severely limited by the lack of simple, scalable, rapid, and robust methods to conditionally manipulate gene expression. Therefore, we decided to embark on the optimization of a method for inducible gene knockdown in hPSCs and hPSC-derived cells.

We envisioned that we could develop an improved inducible knockdown platform by ensuring robust, homogeneous, and stable expression of the TET-OFF system, in order to allow tightly regulated transcriptional control both in hPSCs and in differentiated cells. We hypothesized that this could be achieved by targeting genomic safe harbours (GSHs; Chapter 1.8.1) with both an inducible shRNA and a strongly expressed tetR.

4.3 Results

4.3.1 Expression of inducible shRNAs from the human AAVS1 GSH

The AAVS1 GSH has been previously used to express various transgenes, including constitutive shRNAs (DeKolver et al. 2010). Therefore, we tested if this GSH could also be used for expression of inducible shRNAs under the control of the TET-OFF system (Figure 1.18). To this aim, we used a previously described ZFN-based gene trap targeting method (Hockemeyer et al., 2009) to introduce an shRNA against OCT4 whose expression is under the control of a tetracycline-inducible H1 Pol III promoter carrying a single tet operon after the TATA box (H1-TO; Figure 4.2). As a proof-of-principle, this transgene was initially introduced in H9 hESCs that had been previously targeted with a randomly integrated tetR and extensively screened for its high expression in undifferentiated cells as well as in the three germ layers (Figure 4.1). Remarkably, 100% (30/30) of the clonal lines we tested showed robust inducible knockdown of OCT4 after tetracycline treatment (Figure 4.2).

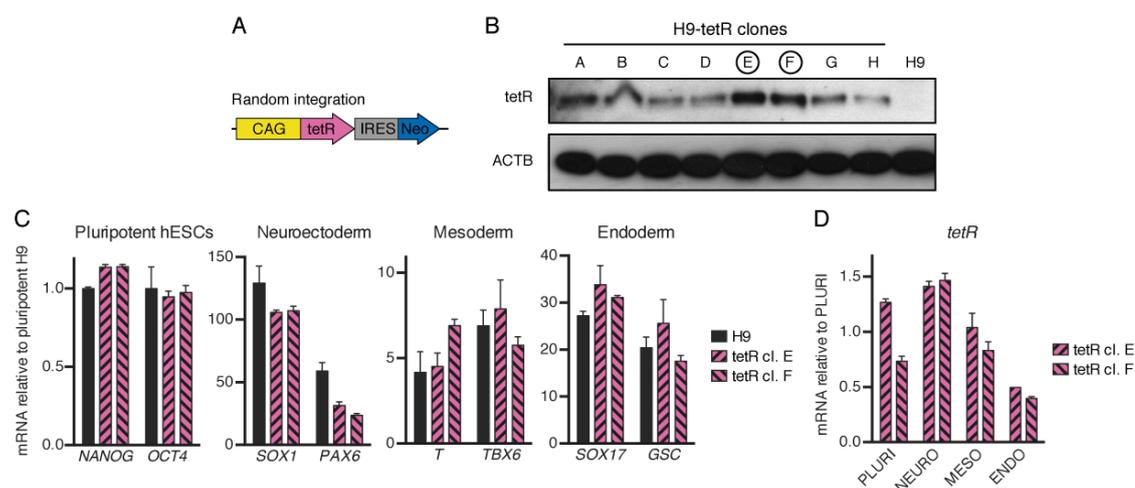


Figure 4.1. Generation of tetR expressing hPSCs.

(A) Schematic of the transgene used to generate hESCs carrying a randomly integrated tetR (pTP6Neo_tetR). (B) Screening of tetR protein expression by Western blot in H9 hESC clonal lines. Clones E and F were selected for further characterization. ACTB (β -actin): loading control. (C) Characterization of the differentiation potential of tetR hESCs by qPCR for lineage-specific markers in undifferentiated cells and in the three germ layers. (D) tetR expression in undifferentiated cells (PLURI) and in the three germ layers (NEURO: neuroectoderm; MESO: mesoderm; ENDO: endoderm). tetR clone E was selected for generation of inducible knockdown lines.

However, we noted that expression of the OCT4 shRNA was substantially leaky in control conditions (up to 60% knockdown even in the absence of tetracycline; Figure

4.2). We initially wondered whether this could be due to tetracycline contamination of the culture medium because of the presence of BSA (which could potentially carry traces of tetracycline if the originating animal was treated with such antibiotic). However, similar levels of shRNA leakiness were observed also when using BSA-free medium (Figure 4.3), thus ruling out this possibility. Moreover, the leakiness observed in these experiments was not specific to the OCT4 shRNA, as generation of DPY30 inducible knockdown hESCs resulted in a similar level of unwanted knockdown in control conditions (Figure 4.4). Therefore, this issue was intrinsic to the TET-OFF system, in agreement with previous reports (Henriksen et al. 2007).

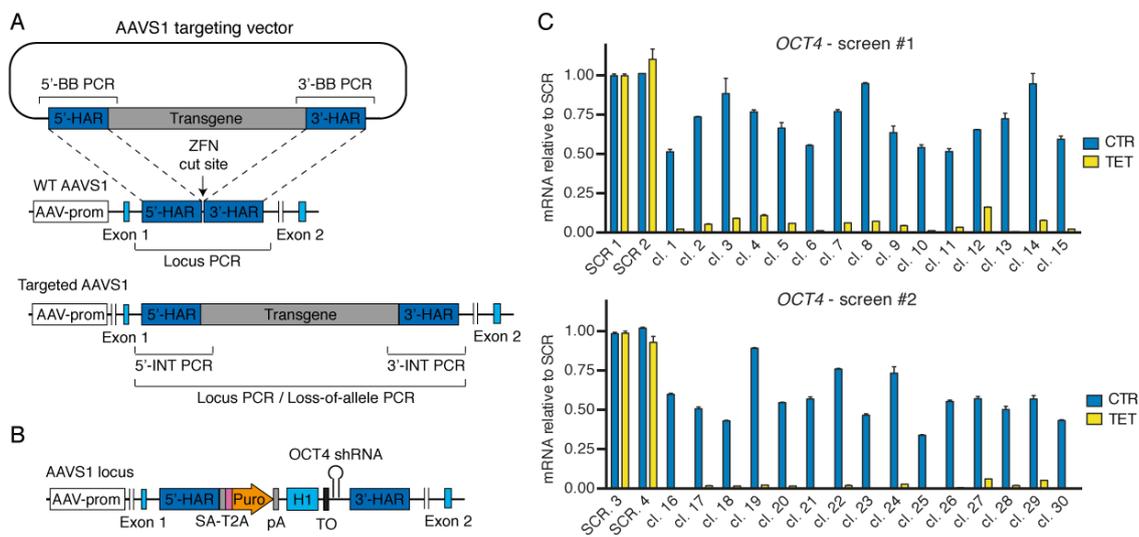


Figure 4.2. Expression of an inducible shRNA against OCT4 from the AAVS1 locus.

(A) Schematic of the AAVS1 targeting method and of the genotyping strategies used to identify correctly targeted lines. ZFN: zinc-finger nucleases; AAV-prom: AAVS1 locus promoter (PPP1R12C gene); 5'-HAR/3'-HAR: upstream/downstream homology arm; Transgene: region integrated following targeting; Locus PCR: PCR of wild-type AAVS1 locus (indicating a non-targeted allele); Locus PCR/loss-of-allele: PCR of targeted allele/PCR that fails if the transgene contains the GC-rich CAG promoter; 5'-INT/3'-INT PCR: PCR of transgene 5'-end/3'-end integration region (indicative of expected transgene targeting); 5'-BB/3'-BB PCR: PCR of vector backbone 5'-end/3'-end integration region (indicative of non-specific off-target plasmid integration). These genotyping strategies were used for all AAVS1 gene targeting experiments presented in this chapter. Note that analogous targeting and genotyping strategies were applied also for all the ROSA26 gene targeting experiments (Figure 4.5 and related figures), but CRISPR/Cas9n was used instead of ZFN for such experiments. (B) Schematic of the locus generated to drive an OCT4 inducible shRNA from the AAVS1 GSH (targeting vector: pAAV_iKD-OCT4). CAG: CMV early enhancer, chicken β -actin, and rabbit β -globin hybrid promoter; tetR: tetracycline-sensitive repressor protein; IRES: internal ribosome entry site; Neo: Neomycin resistance (Neomycin phosphotransferase II); SA: splice acceptor; T2A: self-cleaving T2A peptide; Puro: Puromycin resistance (Puromycin *N*-acetyltransferase); pA: polyadenylation signal; H1: H1 Pol III promoter; TO: tet operon. (C) Results from two independent qPCR screenings of OCT4 inducible knockdown hESC clones. SCR: control hESC clones targeted with an inducible scramble shRNA (targeting vector: pAAV_iKD-SCR). Cells were analysed in the absence (CTR) or presence of tetracycline for 5 days (TET). n=2. qPCR was performed in collaboration with Ms Stephanie Brown.

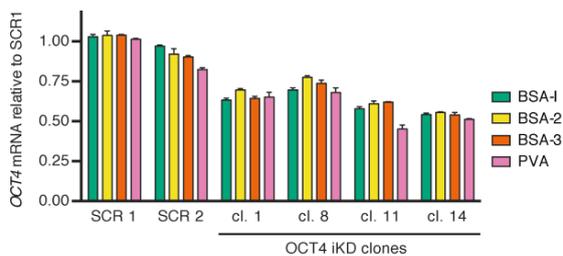


Figure 4.3. Leakiness of the OCT4 inducible shRNA in different culture media.

Expression of OCT4 in inducible knockdown lines or controls (SCR) cultured in absence of tetracycline in medium containing different batches of BSA or PVA. n=2. qPCR was performed in collaboration with Ms Stephanie Brown.

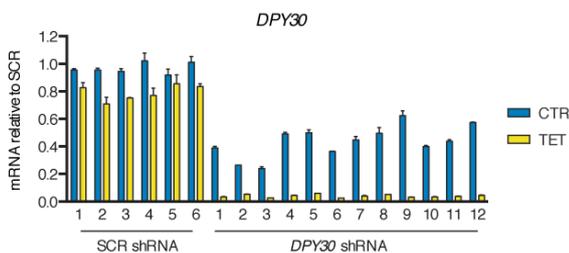


Figure 4.4. Expression of an inducible shRNA against DPY30 from the AAVS1 locus.

As in Figure 4.2, but hESCs were targeted with an shRNA against DPY30 (targeting vector: pAAV_iKD-DPY30) or a SCR control shRNA.

Collectively, these experiments demonstrated the feasibility of using the AAVS1 GSH to efficiently and reproducibly express an inducible shRNA under the control of the TET-OFF system. However, such findings also highlighted the need for optimization to reduce shRNA leakiness and allow tightly controlled inducible knockdown.

4.3.2 Validation of the ROSA26 and AAVS1 loci as *bona fide* GSHs during hPSC differentiation

We then sought to improve the inducible knockdown method by targeting also the tetR to a GSH in order to: (1) provide a reproducible approach that could be easily transferred to other hPSC lines; (2) prevent the possibility of tetR silencing following hPSC differentiation into mature cell types; (3) aim for stronger expression of the tetR to minimize the shRNA leakiness. For this, we reasoned that targeting the inducible shRNA and tetR into two different GSHs would prevent potential promoter interference issues that could arise by placing both elements on a single vector (Shearwin et al. 2005). Moreover, compared to targeting each component onto one allele of the AAVS1 locus, homozygous expression of both transgenes would more strongly support their expression. Overall, we decided to follow this strategy to facilitate individual optimization of each part of the TET-OFF system before attempting to engineer an all-in-one targeting vector. To this aim, we embarked on developing a targeting strategy for another GSH.

The Rosa26 locus is one of the most popular GSH for mouse transgenesis, and a homolog region exists in the human genome (Irion et al. 2007). Therefore, we developed a CRISPR/Cas9n-based strategy to target the human ROSA26 locus by homologous recombination of a gene trap transgene. Then, we tested different targeting vector designs aimed to express a constitutive EGFP transgene in hPSCs (Figure 4.5 and Figure 4.6). Use of the CAG promoter proved to be the best solution, as it increased EGFP levels by an order of magnitude compared to the endogenous ROSA26 promoter, a strategy that was previously described (Irion et al., 2007). Interestingly, the EF1 α promoter was heavily silenced and led to mosaic expression of the transgene (Figure 4.6). This is in agreement with other reports (Luo et al. 2014; Norrman et al. 2010), and indicates that the EF1 α is not generally suited for hPSC transgenesis. On the other hand, the CAG promoter supported stable and homogeneous EGFP expression in both the ROSA26 and AAVS1 loci for more than 30 passages (Figure 4.8). Targeting of the ROSA26 locus was very efficient, with 73% to 100% of integration in at least one allele, and 25% to 64% of clones harbouring no additional copies of the transgene outside of the locus (Table 9). Finally, neither ROSA26 nor AAVS1 targeting resulted in overt chromosomal abnormalities (Figure 4.8).

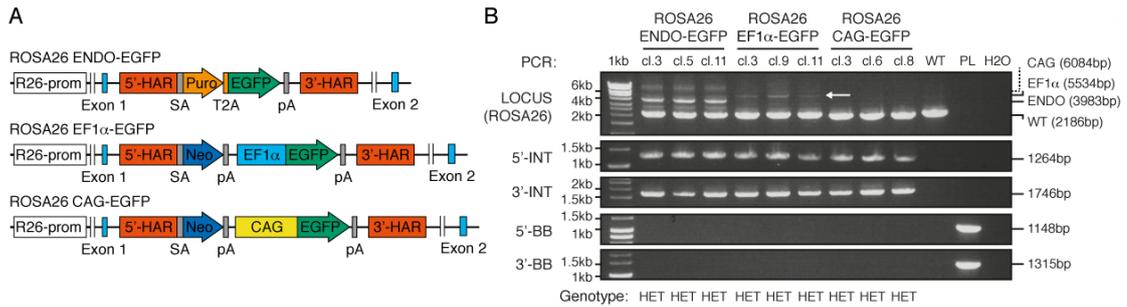


Figure 4.5. Generation of ROSA26 reporter hPSCs using different promoters.

(A) Schematic of the ROSA26 transgenic alleles generated to test the best strategy for constitutive EGFP expression. ENDO-EGFP: EGFP driven by the endogenous ROSA26 promoter (targeting vector pR26_ENDO-EGFP); EF1 α -EGFP: EGFP driven by the elongation factor 1 α promoter (targeting vector pR26_EF1 α -EGFP); CAG-EGFP: EGFP driven by the CAG promoter (targeting vector pR26_CAG-EGFP). See Figure 4.2 for abbreviations. Gene targeting was performed in collaboration with Dr Matthias Pawlowski. (B) Genotyping of selected heterozygous (HET) ROSA26-targeted hESCs. Next to the LOCUS PCR, black lines show the size of amplicons from the different transgenes used, with the dashed line indicating the predicted size for a PCR that fails due to the high GC-content of the CAG promoter (loss-of-allele). 1kb: 1kb DNA ladder; WT: H9 hESCs; PL: ROSA26 targeting plasmid; H2O: water control. See Figure 4.2 for the genotyping strategies.

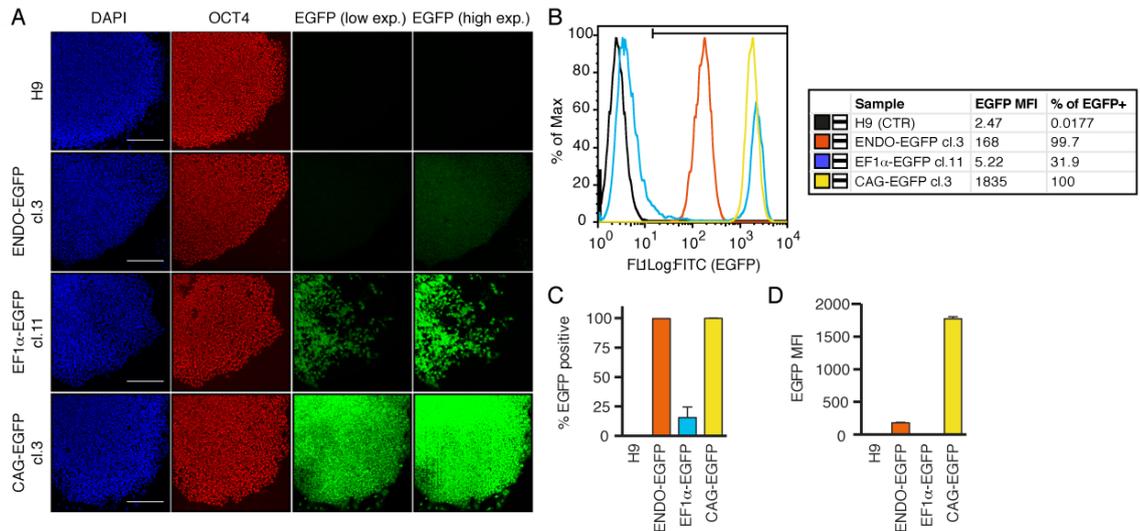


Figure 4.6. Comparison of ROSA26 reporter hPSCs.

(A) EGFP fluorescence in representative clones from the various ROSA26 EGFP reporter hESCs, or wild-type H9 hESCs. Cells were fluorescently immunostained for the pluripotency marker OCT4 (red). DAPI (blue) shows nuclear staining. Scale bars: 200 μ m. (B) Flow cytometry quantification of the percentage of EGFP positive cells (EGFP+; the gate is shown), and of the EGFP median fluorescence intensity (MFI) for the same clones shown in panel A. (C-D) Summary of EGFP quantifications for 3 clones per condition (see Figure 4.5).

We then evaluated the effectiveness of the ROSA26 and AAVS1 targeting strategies for expression of transgenes in hPSCs and hPSC-derived cells. Indeed, while it has been suggested that both these loci are resistant to silencing, an extensive *in vitro* characterization of their expression has never been performed. For this, we generated hESCs with heterozygous or homozygous targeting of a CAG-EGFP transgene in the

ROSA26 or AAVS1 locus, and monitored the expression of the reporter in pluripotent cells and during differentiation into a large variety of lineages (Figure 4.7 and Figure 4.8). In undifferentiated cells and in the three germ layers, EGFP was homogeneously expressed at high levels from either locus, with heterozygous clones showing roughly half the expression of homozygous ones (Figure 4.9). Importantly, targeting did not affect the expression of pluripotency or differentiation markers (Figure 4.10).

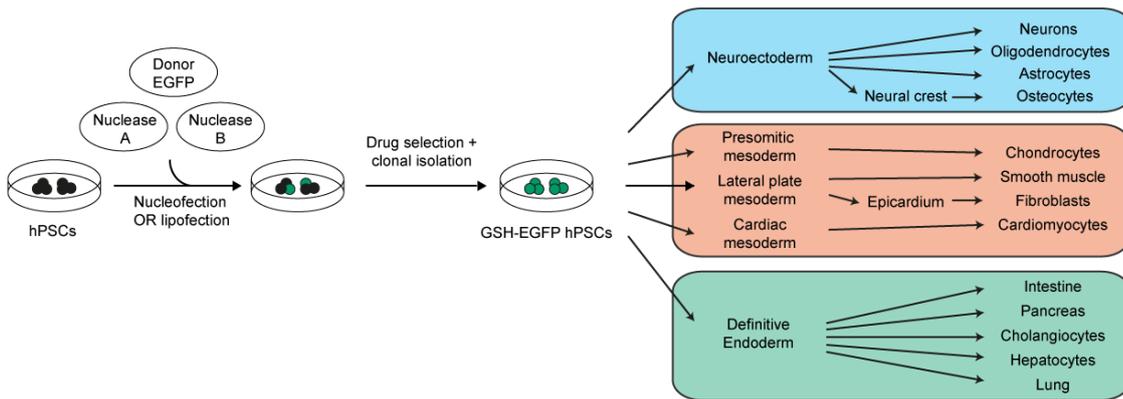


Figure 4.7. Experimental strategy to validate the AAVS1 and ROSA26 loci as “bona fide” genomic safe harbours.

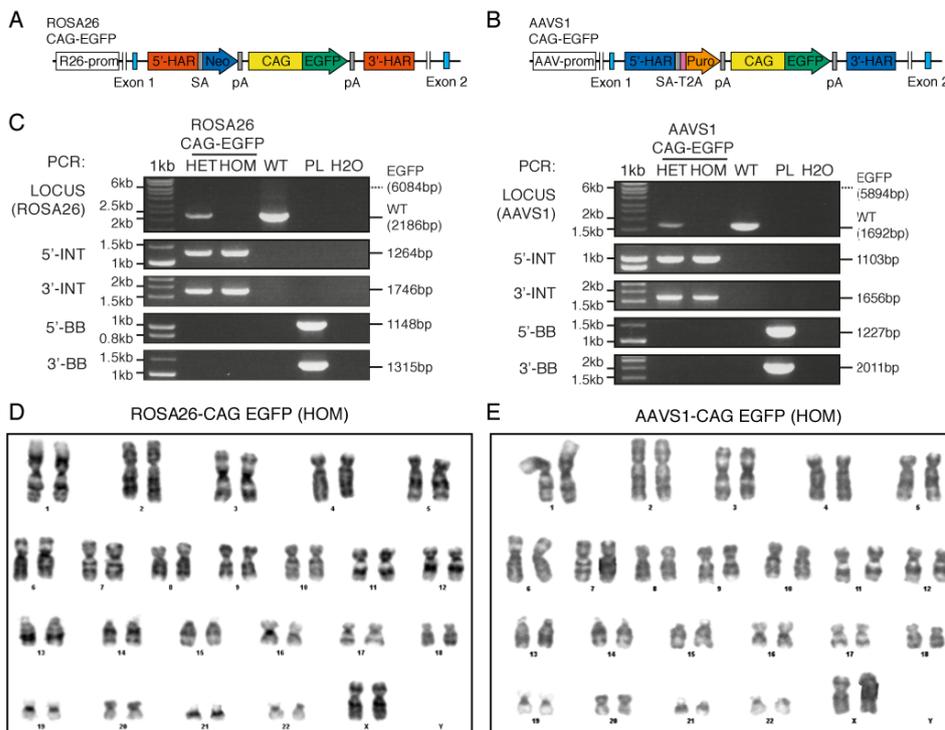


Figure 4.8. Generation of ROSA26 and AAVS1 CAG-EGFP reporter hPSCs. (A-B) Schematic of the transgenic alleles generated. See Figure 4.2 for abbreviations. ROSA26 gene targeting was performed in collaboration with Dr Matthias Pawlowski. (C-D) Genotyping results for selected heterozygous (HET) and homozygous (HOM) reporter hESCs (targeting vectors: pR26_CAG-EGFP; pAAV_CAG-EGFP). See Figure 4.2 and Figure 4.5 for the genotyping strategies. (D-E) Karyotyping of the selected reporter lines, performed in collaboration with the Medical Genetics Service, Cambridge University Hospitals.

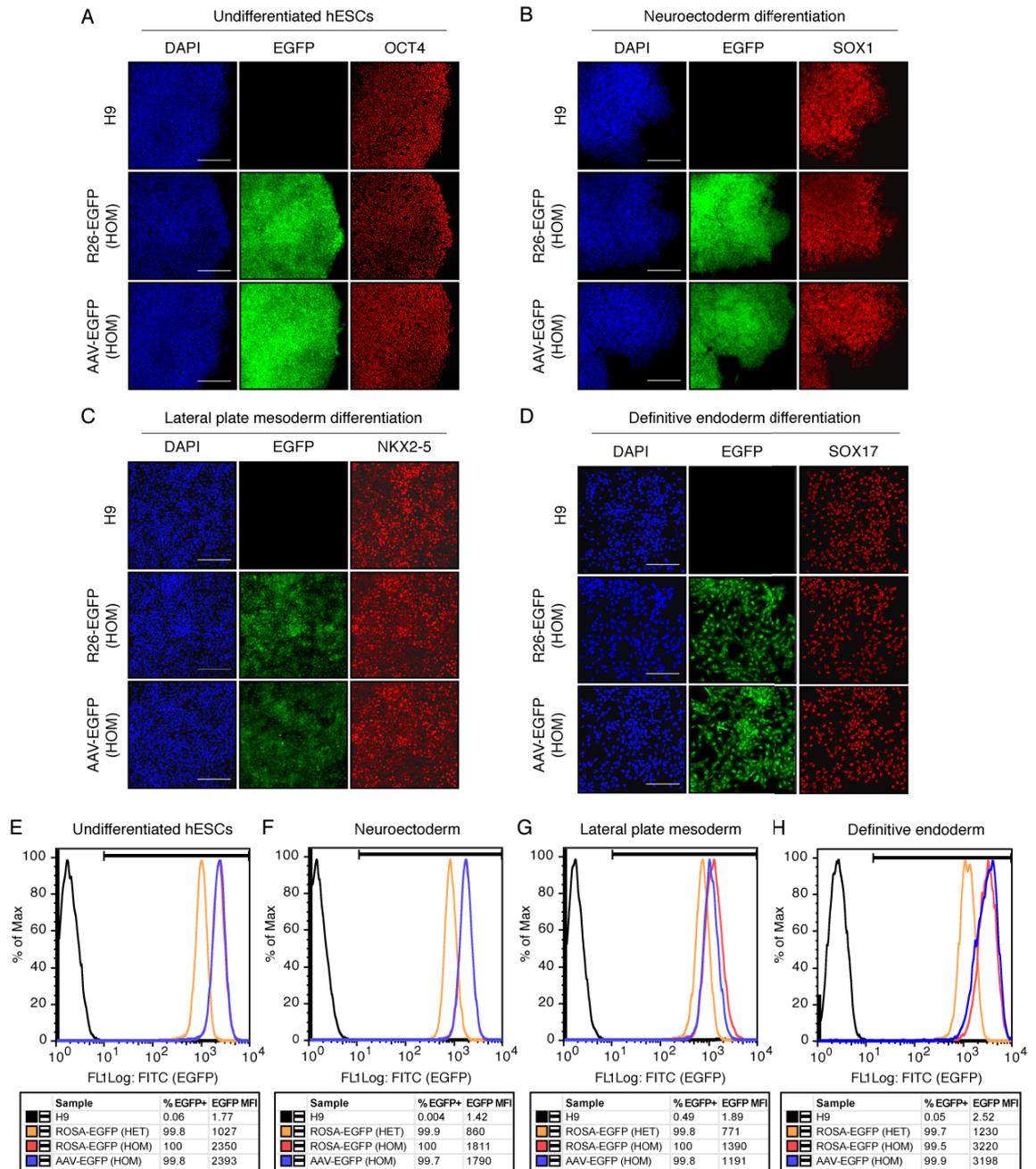


Figure 4.9. Expression of the ROSA26 and AAVS1 reporters in undifferentiated cells and in the three germ layers.

(A-D) Representative immunofluorescent stainings for the indicated lineage-specific markers (red) in undifferentiated hESCs and in the three germ layers obtained from wild-type H9 hESCs and homozygous (HOM) ROSA26 and AAVS1 CAG-EGFP reporter lines. EGFP fluorescence is in green, and DAPI (blue) shows nuclear staining. Scale bars: 200µm. (E-H) Representative flow cytometry quantifications of EGFP levels. HET: heterozygous targeting.

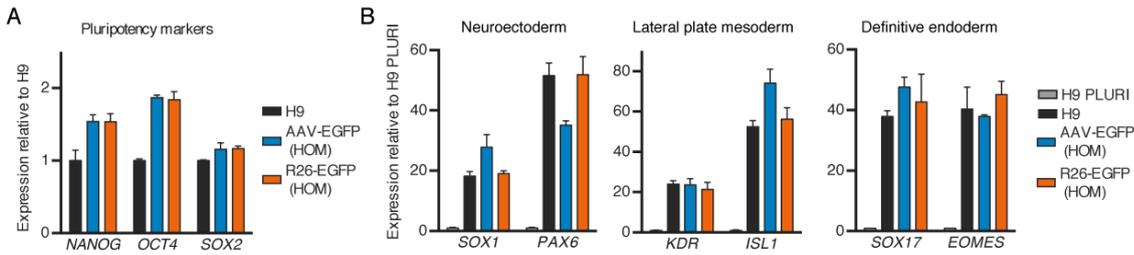


Figure 4.10. Differentiation potential of ROSA26 and AAVS1 reporter hPSCs.

qPCR for lineage-specific markers in undifferentiated cells (A) and in the three germ layers (B). H9 PLURI: undifferentiated H9 hESCs. AAV: AAVS1 targeting; R26: ROSA26 targeting. n=2.

Then, we further differentiated ROSA26 and AAVS1 reporter hPSCs into fifteen different cell types by using protocols that recapitulate the key stages of human development (Figure 4.7). Remarkably, both loci supported strong and homogeneous EGFP expression in all of these cell types, as shown both by flow cytometry and colocalization with lineage-specific markers (Figure 4.11 and Figure 4.12).

Overall, these results validated the ROSA26 and AAVS1 loci as *bona fide* GSHs suitable for robust transgenesis of hPSCs and hPSC-derived cells.

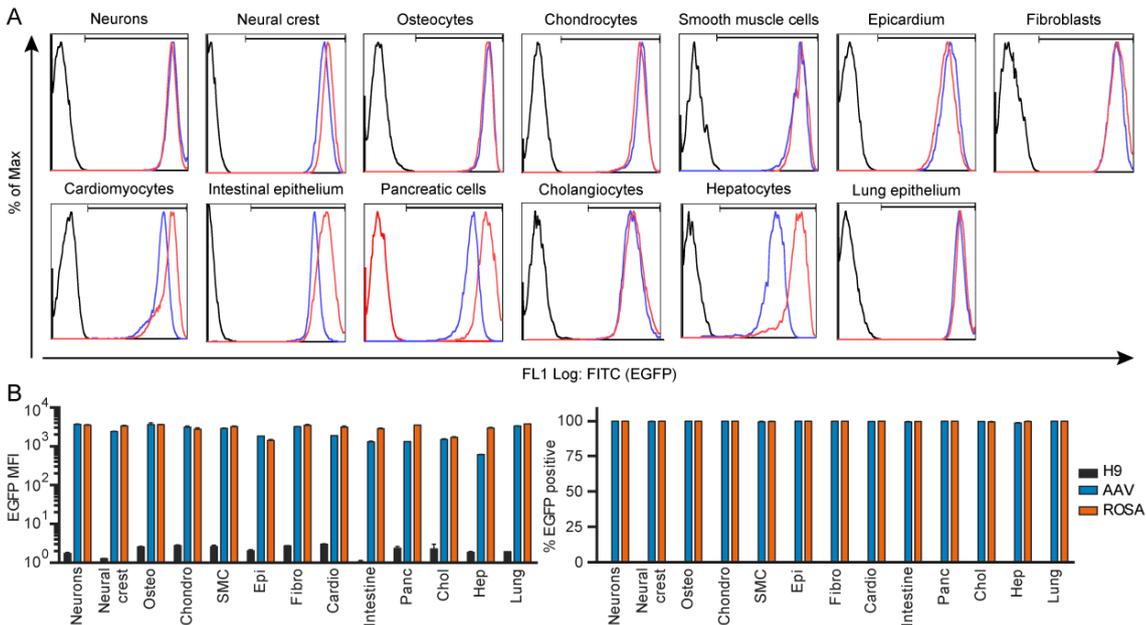


Figure 4.11. Expression quantification of the ROSA26 and AAVS1 reporters in differentiated cells.

(A) Representative flow cytometry quantifications of EGFP levels in cell types generated from homozygous ROSA26 (red lines) and AAVS1 (blue lines) CAG-EGFP reporter hESCs, or wild-type hESCs (black lines). Also refer to Figure 4.7. (A) Summary of the EGFP quantification experiments; n=2. Certain differentiations were performed in collaboration with Dr Matthias Pawlowski (neural cells); Mr William Bernard and Dr Felipe Serrano (neural crest and osteocytes); Ms Elisa Giacomelli (chondrocytes and intestinal epithelium); Dr Dharini Iyer and Dr Laure Gambardella (epicardium and fibroblasts); Dr Fotios Sampaziotis and Ms Mariëlle Zonneveld (cholangiocytes); Mr Miguel Cardoso de Brito (hepatocytes); and Dr Nick Hannan (lung epithelium).

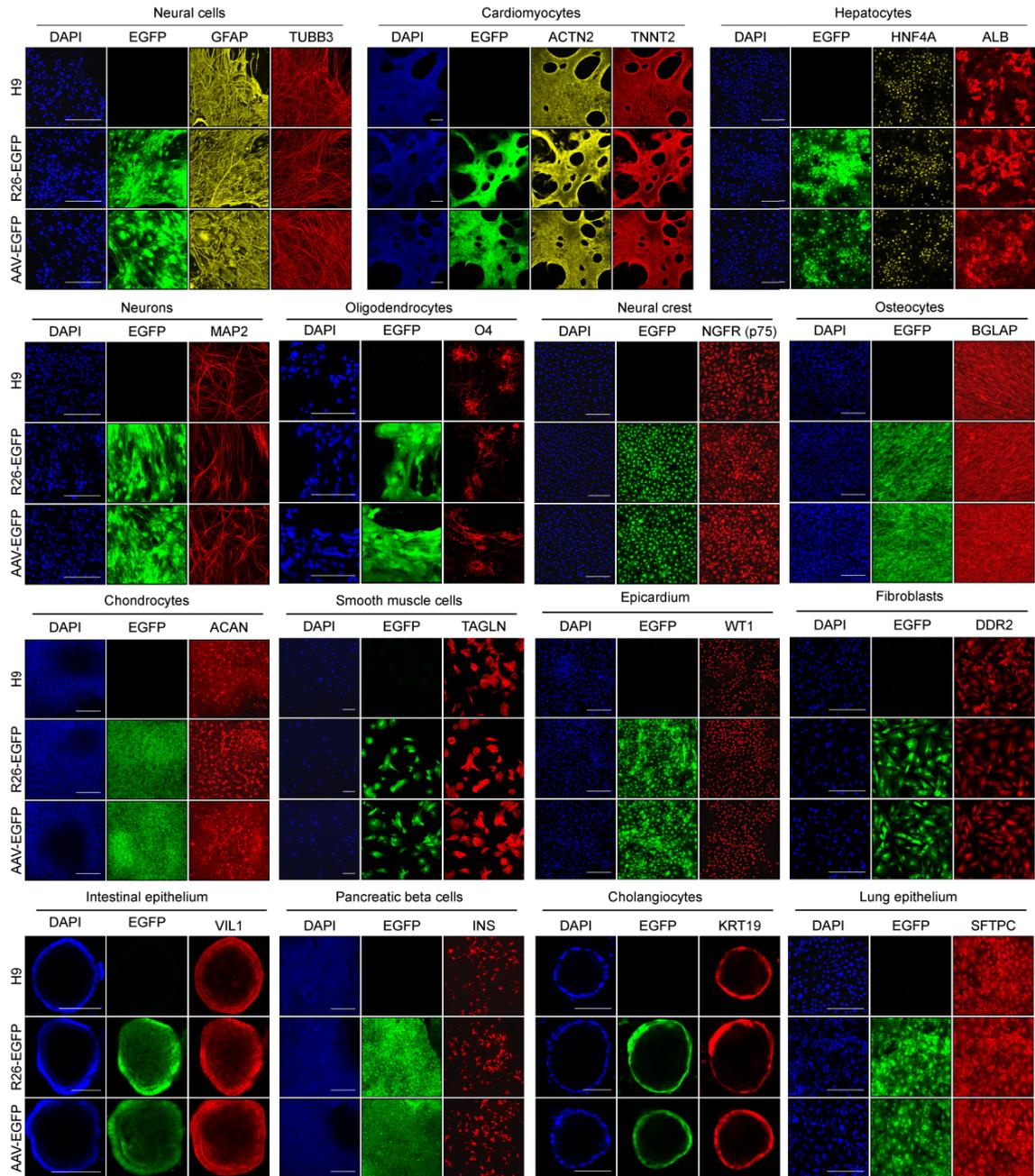


Figure 4.12. Expression of the ROSA26 and AAVS1 reporters in differentiated cells.

Representative immunofluorescence of the indicated lineage-specific markers (yellow or red) in cell types generated from homozygous ROSA26 (R26) and AAVS1 (AAV) CAG-EGFP reporter hESCs, or in wild-type hESCs. EGFP fluorescence from the reporter lines is in green, and DAPI (blue) shows nuclear staining. Scale bars: 100 μ m for oligodendrocytes, intestinal epithelium, and cholangiocytes; 200 μ m for all other lineages. Some of the differentiations were performed in collaboration as described in Figure 4.11.

4.3.3 Development of an optimized inducible knockdown method based on dual GSH targeting.

Having identified two suitable GSHs, we used them to develop an inducible knockdown (iKD) system based on dual GSH targeting (Figure 4.13). In order to simplify the evaluation of gene knockdown and facilitate the optimization of this method, we decided to generate hPSC lines in which a constitutively expressed EGFP transgene could be silenced in an inducible fashion (Figure 4.14). First, we validated that AAVS1-driven expression of a constitutively active EGFP shRNA cassette could potently and homogeneously suppress ROSA26-driven EGFP expression (Figure 4.15). Then, we sequentially targeted hESCs with: (1) a CAG-tetR transgene in the ROSA26 locus; (2) a CAG-EGFP transgene plus an anti-EGFP inducible shRNA cassette into the AAVS1 locus (Figure 4.14). EGFP knockdown was then induced by culturing the resulting sublines in the presence of tetracycline for 5 days.

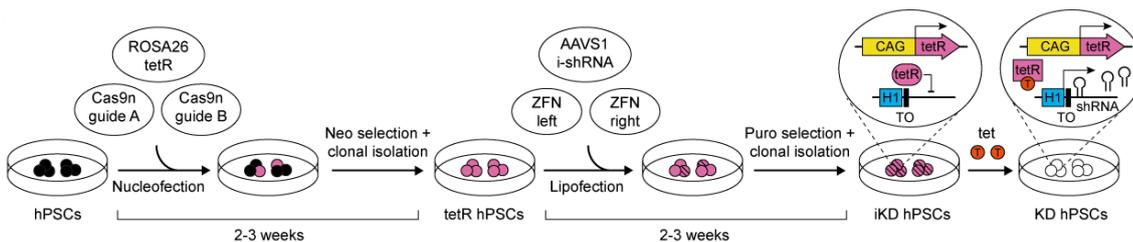


Figure 4.13. Experimental strategy to generate inducible knockdown hPSCs following dual GSH targeting.

Initially, we observed that positioning the CAG promoter in the AAVS1 locus before of the shRNA cassette resulted in poor shRNA induction (Figure 4.15), probably because of promoter interference (Shearwin et al. 2005). However, the opposite arrangement allowed very efficient induction of EGFP knockdown, as shown by an homogeneous decrease of EGFP fluorescence by more than 95% (Figure 4.15). Nevertheless, we observed a significant decrease of EGFP expression even in the absence of tetracycline, which unfortunately suggested a substantial leakiness in the expression of the shRNA. We therefore explored different approaches to overcome this limitation.

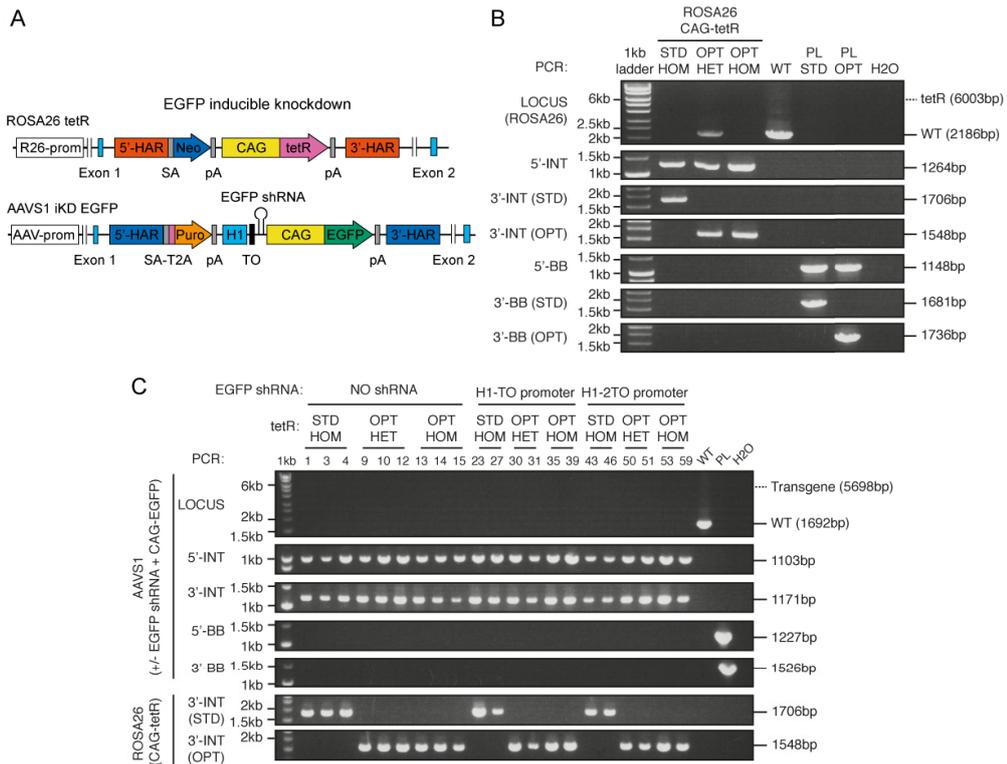


Figure 4.14. Generation of inducible EGFP knockdown hPSCs.

(A) Targeting strategy to generate EGFP inducible knockdown hESCs. See Figure 4.2 for abbreviations. ROSA26 gene targeting was performed in collaboration with Dr Matthias Pawlowski. (B) Genotyping of ROSA26-targeted hESCs carrying the standard tetR (STDtetR; pR26_CAG-STDtetR) or the codon-optimized tetR (OPTtetR; pR26_CAG-OPTtetR). HET: heterozygous targeting; HOM: homozygous targeting. (C) Genotyping of hESCs carrying the indicated combinations of ROSA26-targeted CAG-tetR and of homozygous AAVS1-targeted CAG-EGFP (with or without an inducible EGFP shRNA; see Figure 4.15 and Figure 4.19). AAVS1 targeting vectors used: pAAV_CAG-EGFP (NO shRNA); pAAV_EGFPiKD-TO (H1-TO promoter); pAAV_EGFPiKD-2TO (H1-2TO promoter). H1-TO/H1-2TO: inducible H1 promoter containing one or two tet operons. See Figure 4.2 regarding genotyping strategies used in panels B-C.

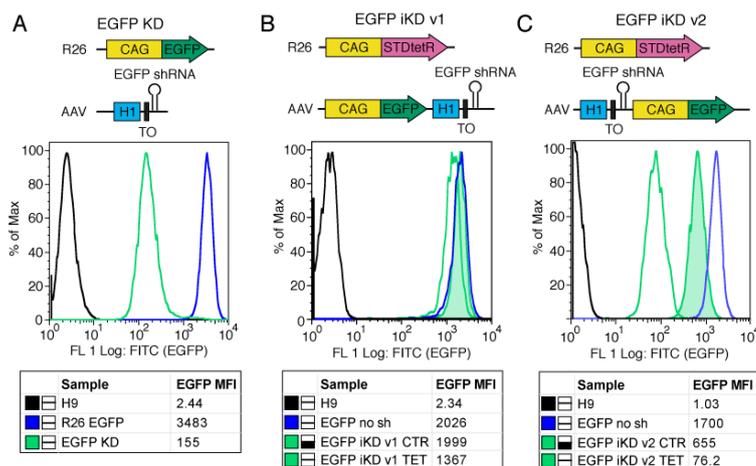


Figure 4.15. Inducible EGFP knockdown in hPSCs.

Representative flow cytometry quantifications of EGFP levels in hESCs carrying various EGFP knockdown systems. The schematics describe the targeting strategies used (see Figure 4.14). For B and C, cells were cultured in the absence (CTR) or presence of tetracycline for 5 days (TET). EGFP levels were compared to untargeted hESCs (H9; negative control) and hESCs targeted with the same strategy apart from the lack of EGFP shRNA (R26 EGFP or EGFP no sh; positive controls).

First, we hypothesized that the leakiness of the system could be reduced by expressing higher levels of the tetR protein, in order to promote stronger repression of the shRNA in the absence of tetracycline. For this reason, we performed multi-parameter RNA and codon optimization of the bacterial tetR sequence to facilitate its expression in human cells (Figure 4.16; Fath et al., 2011; Gray et al., 2007). Compared to the standard tetR (STDtetR), the codon-optimized tetR (OPTtetR), was roughly ten times more expressed when targeted into the ROSA26 locus (Figure 4.17). As a second strategy, we developed another inducible H1 promoter carrying a second tet operon before the TATA box (H1-2TO; Figure 4.18), which should facilitate binding of the tetR in the absence of tetracycline (Henriksen et al. 2007).

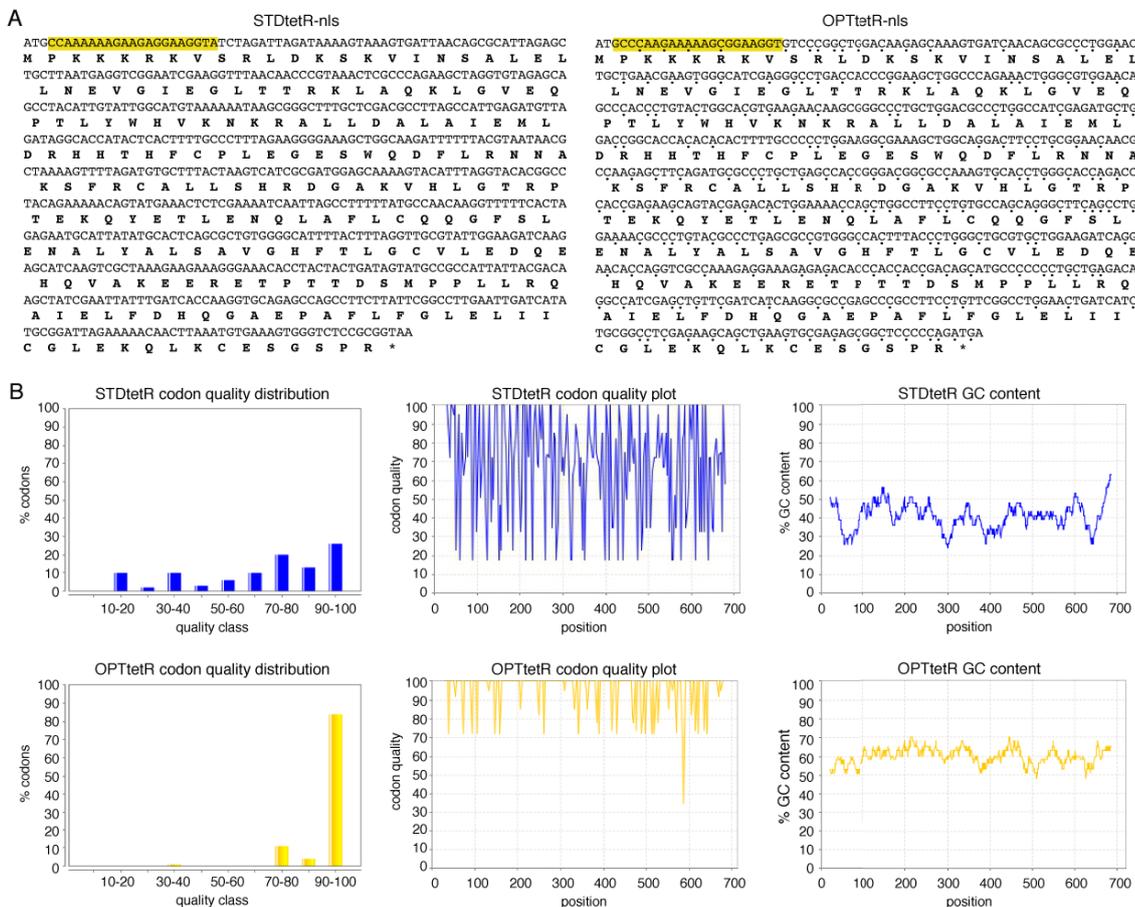


Figure 4.16. Codon optimization of the tetR.

(A) Nucleotide and amino acid sequences of the tetracycline-sensitive repressor protein (tetR) containing an N-terminal SV40 nuclear localization signal (nls, highlighted in yellow). Sequences are reported either before or after codon optimization (STDtetR and OPTtetR, respectively). Dots indicate the synonymous mutations introduced in the OPTtetR. (B) Analysis of the codon quality and GC-content of the tetR-nls sequence before and after optimization.

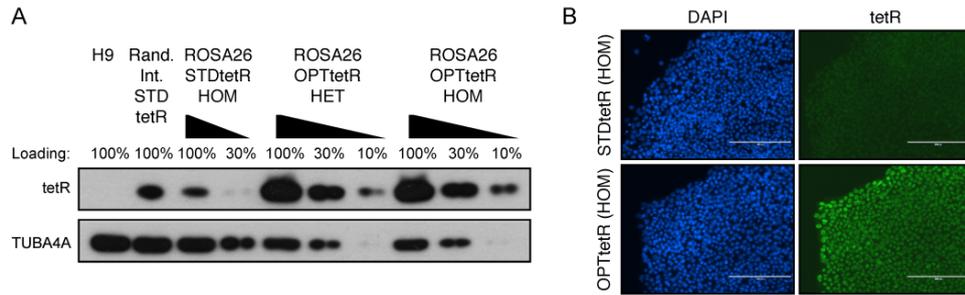


Figure 4.17. Expression of the codon-optimized tetR.

(A) Representative Western blot in ROSA26-targeted hESCs expressing the STDtetR or OPTtetR. HET: heterozygous targeting; HOM: homozygous targeting. hESCs with STDtetR random integration (Rand. Int., clone E described in Figure 4.1) are also shown as a positive reference. Wild-type H9 hESCs are negative controls. As indicated, various protein amounts were loaded to facilitate semi-quantitative comparison. TUBA4A (α -tubulin): loading control. (B) Immunofluorescent staining for the tetR (green). DAPI (blue) shows nuclear staining. Scale bars: 100 μ m.

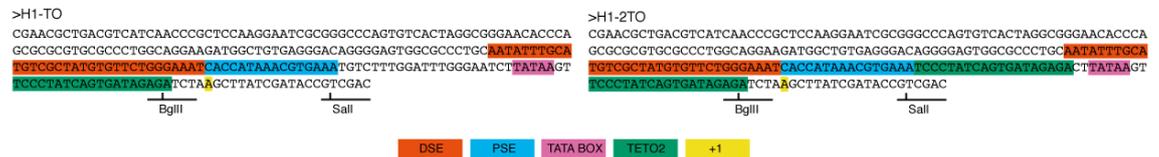


Figure 4.18. Comparison of H1 inducible promoters.

Nucleotide sequences for two inducible H1 Pol III promoters containing one or two tet operons (H1-TO and H1-2TO, respectively). Key sequence features are highlighted with various colours. The cut sites of the two restriction enzymes used for shRNA cloning are shown. DSE: distal sequence element; PSE: proximal sequence element; TETO2: tet operon; +1: start position of RNA transcription.

We then tested the efficiency the four possible combinations of tetR proteins and inducible promoters in the context of the EGFP inducible knockdown system (Figure 4.14 and Figure 4.19). Interestingly, while able to prevent shRNA leakiness in the absence of tetracycline, the H1-2TO promoter showed strongly impaired EGFP knockdown upon induction. This suggested that the addition of a second tet operon interferes with the functionality of the H1 promoter. On the other hand, homozygous expression of the OPTtetR in the context of the H1-TO promoter was sufficient to prevent significant shRNA leakiness while fully preserving maximal knockdown induction (Figure 4.19). Further analyses demonstrated that in these conditions the inducible knockdown was rapid, reversible, and responsive to the dose of tetracycline (Figure 4.20 and Figure 4.21). Overall, we named the method resulting from the combination of these transgenes OPTiKD, for OPTimized inducible KnockDown (Figure 4.22 and Appendix II).

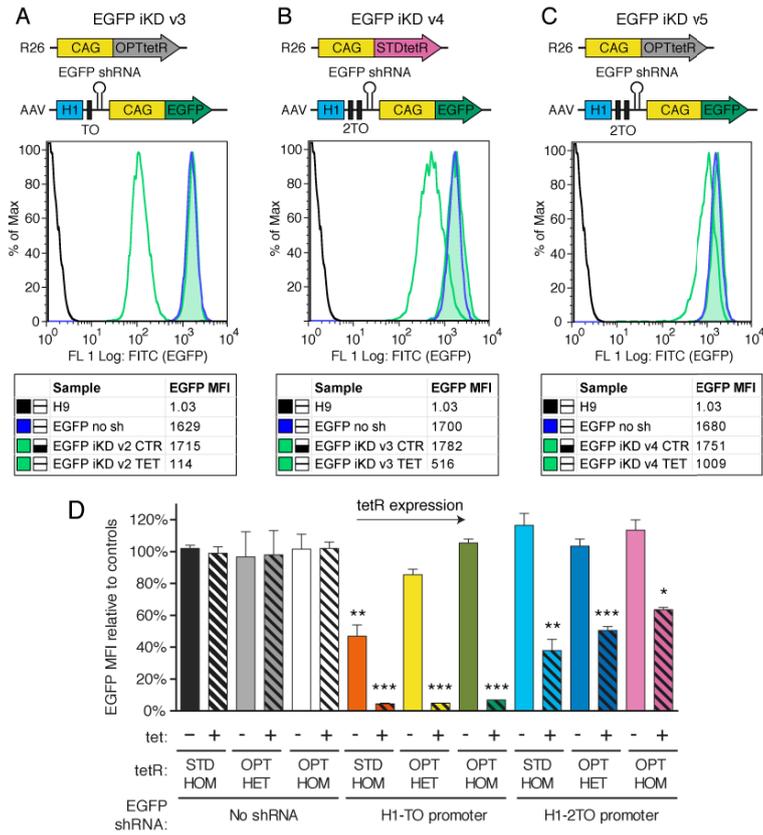


Figure 4.19. Optimization of EGFP inducible knockdown.

(A-C) Representative flow cytometry quantifications of EGFP levels in hESCs carrying various EGFP inducible knockdown systems and analysed as described for Figure 4.15. (D) Quantification of EGFP expression in the various inducible knockdown systems (panel A-C and Figure 4.15). Results are from 2-3 individual lines per condition (Figure 4.14). Significant differences vs same tetR line no tet and no shRNA are reported (ANOVA with post-hoc Holm-Sidak comparisons). EGFP iKD version 3 was chosen for further experiments and named OPTiKD.

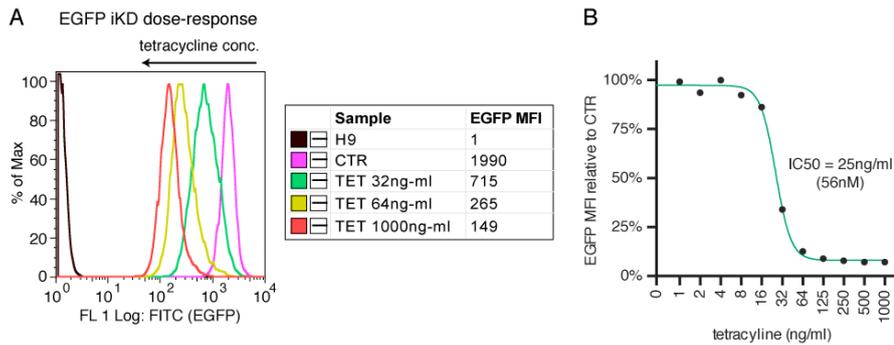


Figure 4.20. Dose responsiveness of EGFP OPTiKD.

(A) Representative flow cytometry quantifications of EGFP knockdown in EGFP OPTiKD hESCs cultured for 5 days with the indicated tetracycline doses. CTR: no tetracycline. (B) Tetracycline dose-response curve for EGFP knockdown. The half-maximal inhibitory concentration (IC50) is reported. Results are from 2 biological replicates per dose, and the mean is shown.

Finally, the generation of OCT4 OPTiKD hESCs with strong and tightly regulated knockdown was so efficient that phenotypic analyses could be performed immediately after antibiotic selection on a mixed population of cells, thereby bypassing the need of picking individual colonies (Figure 4.24).

Overall, these results established that dual targeting of GSHs with an optimized inducible knockdown system is a powerful method to control gene expression in hPSCs.

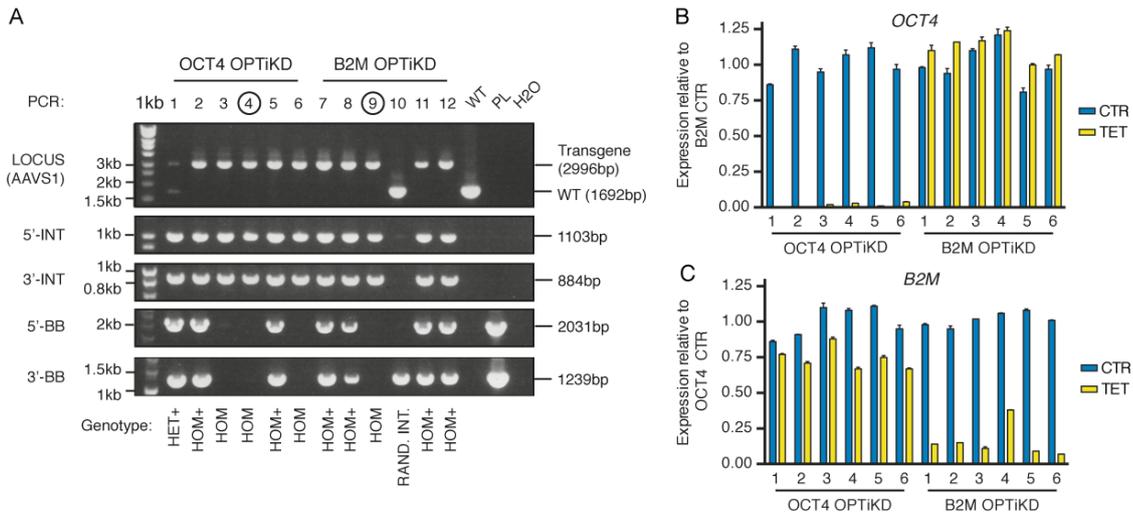


Figure 4.23. Generation of OCT4 and B2M OPTiKD hPSCs.

(A) Genotyping of OCT4 and B2M (negative control gene) OPTiKD hESCs (targeting vectors pAAV_iKD-OCT4 and pAAV_iKD-B2M). HET/HOM: heterozygous or homozygous AAVS1 targeting without additional plasmid integration events; HET+/HOM+: heterozygous or homozygous AAVS1 targeting with additional off-target plasmid integration. See Figure 4.2 regarding genotyping strategies. Clones selected for further experiments are circled. (B-C) qPCR validation of all the OPTiKD lines from panel A cultured in absence (CTR) or presence of tetracycline for 5 days (TET). Individual clones were analysed in duplicate. The expression is shown as normalized on the average level in B2M or OCT4 OPTiKD hESCs in control conditions, as indicated.

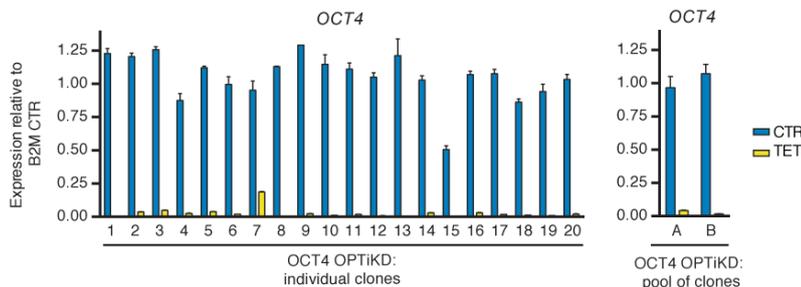


Figure 4.24. Efficiency of OCT4 OPTiKD hPSC generation.

qPCR screening of OCT4 OPTiKD in clonally isolated, or non-clonally isolated (pool of clones) targeted hESCs. Individual clones or pools were analysed in duplicate. The expression is normalized on the average level in B2M OPTiKD hESCs maintained in control conditions. qPCR was performed in collaboration with Ms Stephanie Brown.

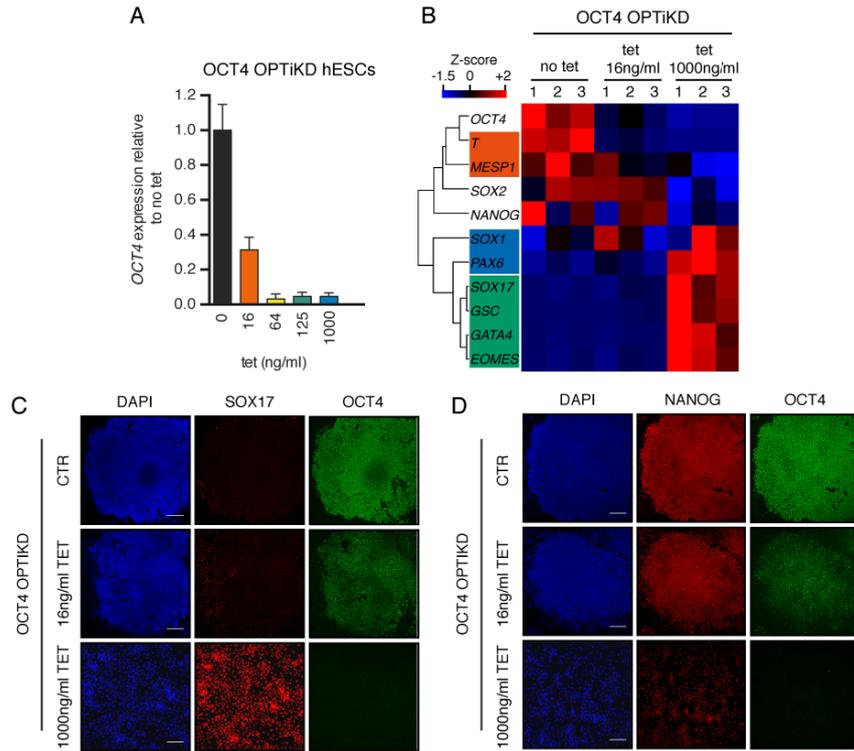


Figure 4.25. Effect of OCT4 inducible knockdown in hPSCs.

(A-B) qPCR analysis of OCT4 OPTiKD hESCs in absence of tetracycline, or following treatment with different doses of tetracycline for 5 days. In the heatmap in B, Z-scores indicate differential expression measured in number of standard deviations from the average level. Markers for each germ layer are highlighted in a coloured box (green: endoderm; red: mesoderm; blue: neuroectoderm). Results are from 3 biological replicates. (C-D) Immunofluorescence of the pluripotency genes OCT4 and NANOG, and of the endoderm marker SOX17 in OCT4 OPTiKD hPSCs maintained in absence of tetracycline (CTR) or cultured for 5 days with different doses of tetracycline that induced intermediate (16ng/ml) or full (1000ng/ml) OCT4 knockdown. DAPI (blue) shows nuclear staining. Scale bars: 200µm.

4.3.4 Single-step generation of optimized inducible knockdown hPSCs

Having taken advantage of a dual GSH targeting strategy to individually optimize the components of the OPTiKD system, we then sought to develop an all-in-one targeting approach that would facilitate the application of this method across multiple hPSC lines. Our previous results proved that positioning a CAG-driven transgene after the inducible H1 promoter does not result in transcriptional interference (Figure 4.15). Hence, we developed various AAVS1 targeting vectors based on this design and carrying both an inducible EGFP shRNA cassette and a CAG-tetR transgene. Then, we tested them for their ability to drive EGFP inducible knockdown in ROSA26-EGFP hESCs (Figure 4.26).

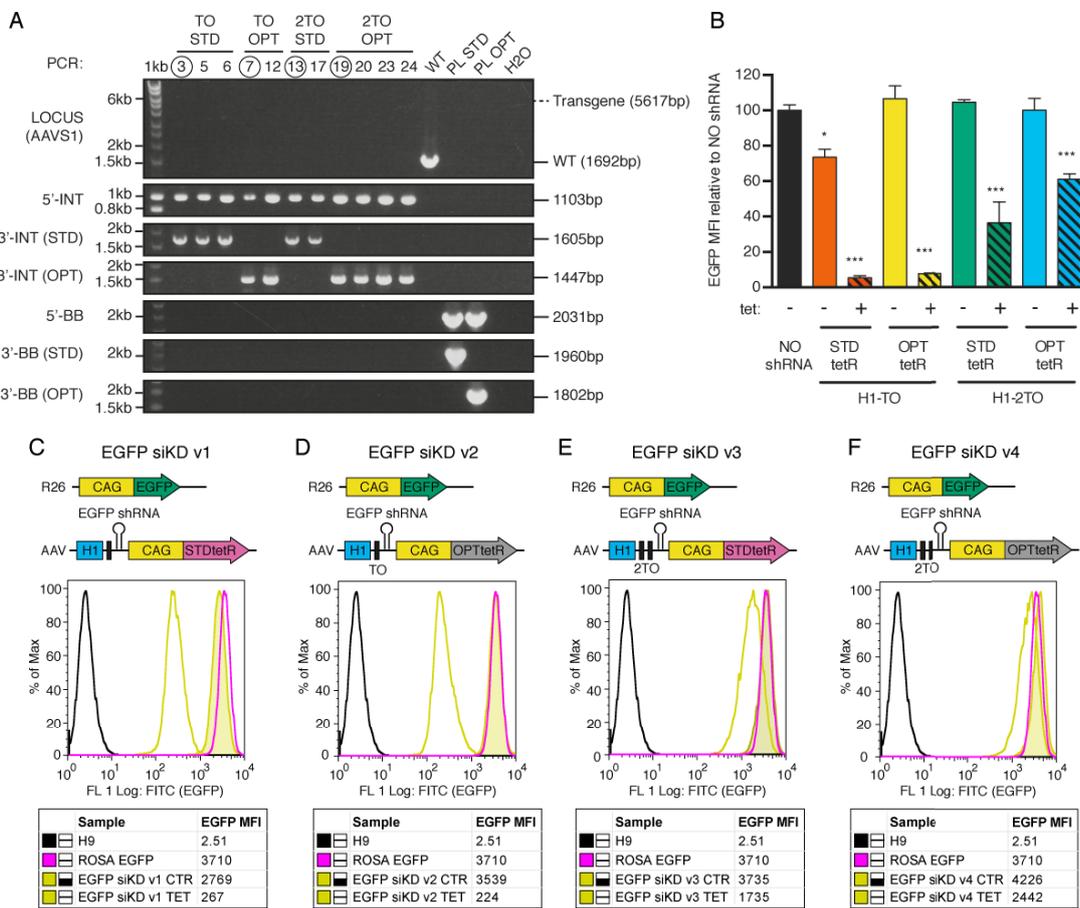


Figure 4.26. Optimization of single-step EGFP inducible knockdown. (A) Genotyping of homozygous ROSA26 CAG-EGFP hESCs carrying the indicated combinations of homozygous AAVS1-targeted inducible EGFP shRNA and CAG-tetR (see schematics in panels C-F). AAVS1 targeting vectors used: pAAV_EGFPsiKD-TO-STD/OPT (TO-STD/TO-OPT); pAAV_EGFPsiKD-2TO-STD/OPT (2TO-STD/2TO-OPT). See Figure 4.2 regarding genotyping strategies. (B) Flow cytometry quantification of EGFP expression in absence or presence of tetracycline for 5 days. EGFP levels were compared to those in homozygous ROSA26 CAG-EGFP hESCs (NO shRNA, positive control) and wild-type H9 (negative control). Results are from 2-4 individual lines per condition (panel A). Significant differences vs NO shRNA are reported (ANOVA with post-hoc Holm-Sidak comparisons). (C-F) Representative examples of the data summarized in panel B. The schematics on top describe the different targeting strategies.

In agreement with our earlier observations, the combination of H1-TO promoter and OPTtetR maximized inducible knockdown while preventing shRNA leakiness (Figure 4.26). We named this strategy sOPTiKD, for single-step OPTimized inducible KnockDown (Figure 4.28). Remarkably, sOPTiKD shared key properties with OPTiKD, such as rapid, reversible, and dose-responsive inducible knockdown (Figure 4.27). Thus, this all-in-one strategy appeared to be as efficient as the original dual targeting approach.

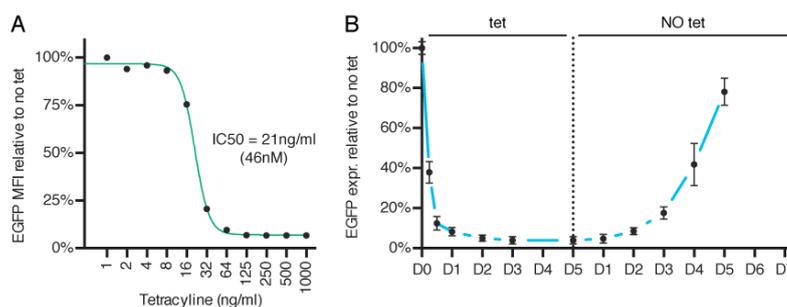


Figure 4.27. Dose responsiveness and reversibility of EGFP sOPTiKD.

(A) Tetracycline dose-response curve for EGFP knockdown in EGFP sOPTiKD hESCs (see Figure 4.20). The IC₅₀ is reported. Results are from 2 biological replicates per dose, and the mean is shown. (B) EGFP mRNA knockdown and rescue kinetics in EGFP sOPTiKD hESCs (see Figure 4.21). Results are from 2 biological replicates per time point. qPCR was performed in collaboration with Ms Stephanie Brown.

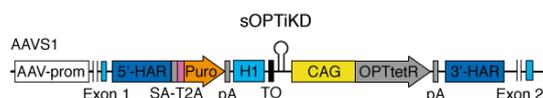


Figure 4.28. Single-step OPTimized inducible KnockDown.

Schematic of the targeting strategy to generate sOPTiKD hPSCs. See Figure 4.2 for abbreviations.

To further demonstrate the versatility of sOPTiKD, we generated both hESC and hiPSC lines carrying an inducible shRNA against OCT4 or B2M (Figure 4.29). Remarkably, all the sublines tested showed robust inducible knockdown with no significant decrease in the absence of tetracycline (Figure 4.29). qPCR analyses confirmed that inducible knockdown of OCT4 using sOPTiKD induced differentiation of both hESCs and hiPSCs, while knockdown of the control gene B2M had no effect (Figure 4.30).

Overall, these experiments showed that sOPTiKD provides an efficient system to knockdown gene expression that can be easily applied to a diversity of hPSC lines (Appendix II).

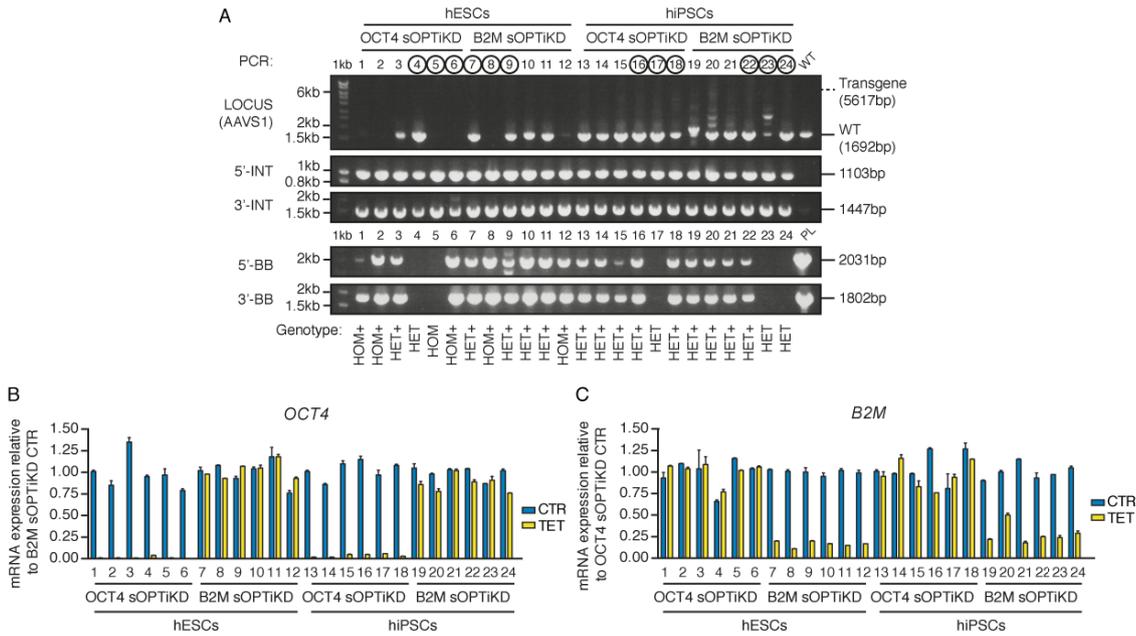


Figure 4.29. Generation of OCT4 and B2M sOPTiKD hESCs and hiPSCs. (A) Genotyping of OCT4 and B2M sOPTiKD hESCs and hiPSCs (targeting vectors pAAV_siKD-OCT4 and pAAV_siKD-B2M). See Figure 4.2 and Figure 4.23 regarding genotyping strategies. Clones selected for further experiments are circled. (B-C) qPCR validation of OCT4 and B2M sOPTiKD hESCs and hiPSCs in absence (CTR) or presence of tetracycline for 5 days (TET). The expression is reported relative to the average level in B2M or OCT4 sOPTiKD hESCs maintained in control conditions, as indicated. Individual clones were analysed in duplicate.

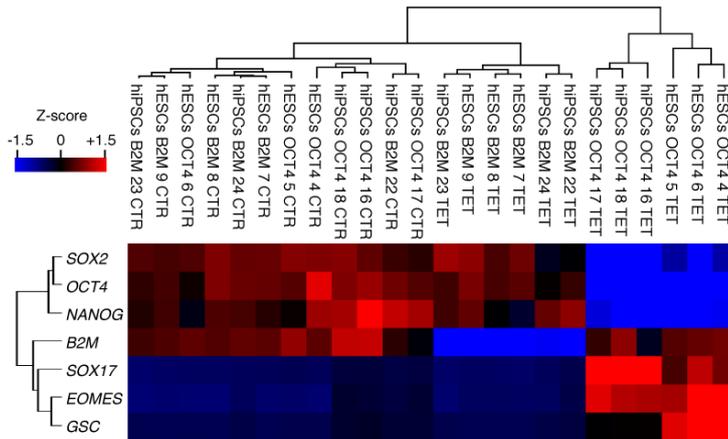


Figure 4.30. Effect of OCT4 inducible knockdown in hESCs and hiPSCs. Heatmap summarizing qPCR data for selected sOPTiKD lines (clone numbers are indicated, see Figure 4.29) before (CTR) and after 5 days of knockdown (TET). Samples and genes were clustered by complete Euclidean distance, and Z-scores indicate differential expression measured in number of standard deviations from the average level.

4.3.5 Validation of the optimized inducible knockdown platforms in hPSC-derived differentiated cells.

The results presented so far demonstrated several key improvements of the methods we developed over previous systems for inducible gene knockdown in hPSCs. Nonetheless, the capacity to knockdown genes in a variety of differentiated cells would represent another major advance in the field, as this has never been previously achieved. We therefore analysed the performance of the OPTiKD and sOPTiKD platforms to knockdown a constitutively expressed EGFP transgene in hPSCs differentiated into the three germ layers, as well as in a panel of thirteen terminally differentiated cell types (Figure 4.7). For both systems, qPCR analyses demonstrated strong and inducible knockdown of EGFP at the transcript level in all lineages tested (Figure 4.31). Further, immunofluorescence confirmed the decrease in EGFP protein expression (Figure 4.32 and Figure 4.33), and flow cytometry showed a decrease of EGFP fluorescence by more than 70% for most lineages (Figure 4.31).

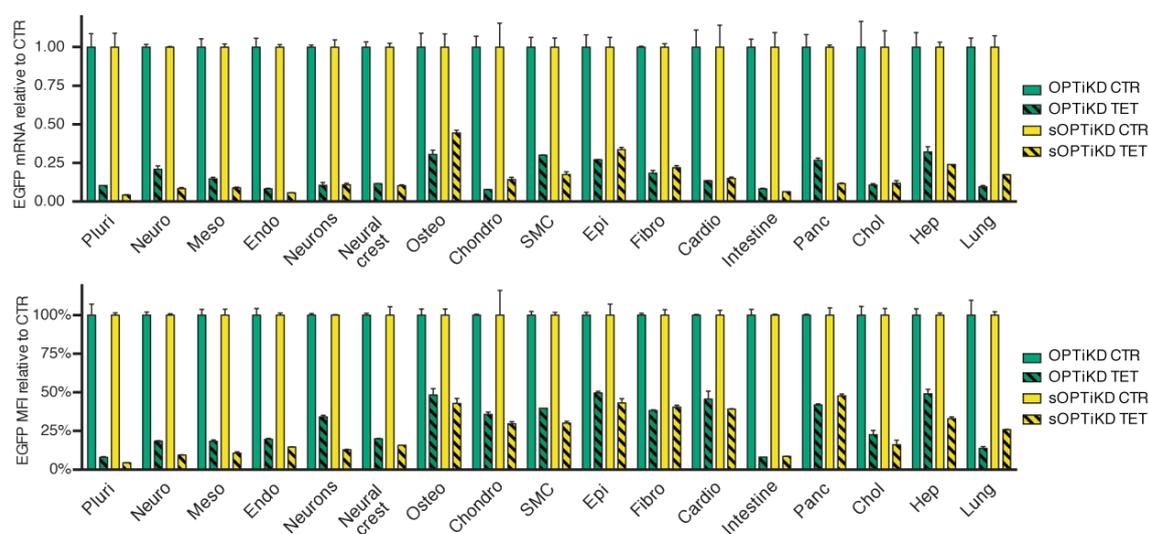


Figure 4.31. Performance of OPTiKD and sOPTiKD in differentiated cells.

EGFP expression measured by qPCR (mRNA, top) or flow cytometry (MFI, bottom) in the indicated cell types derived from EGFP OPTiKD and sOPTiKD hESCs (see Figure 4.7) and cultured in absence (CTR) or presence of tetracycline for 5 days (TET). EGFP levels are reported relative to control conditions in the same line for each individual lineage. Results are from two biological replicates per condition. qPCR was done in collaboration with Ms Loukia Yiangou. Some of the differentiations were performed in collaboration as described in Figure 4.11.

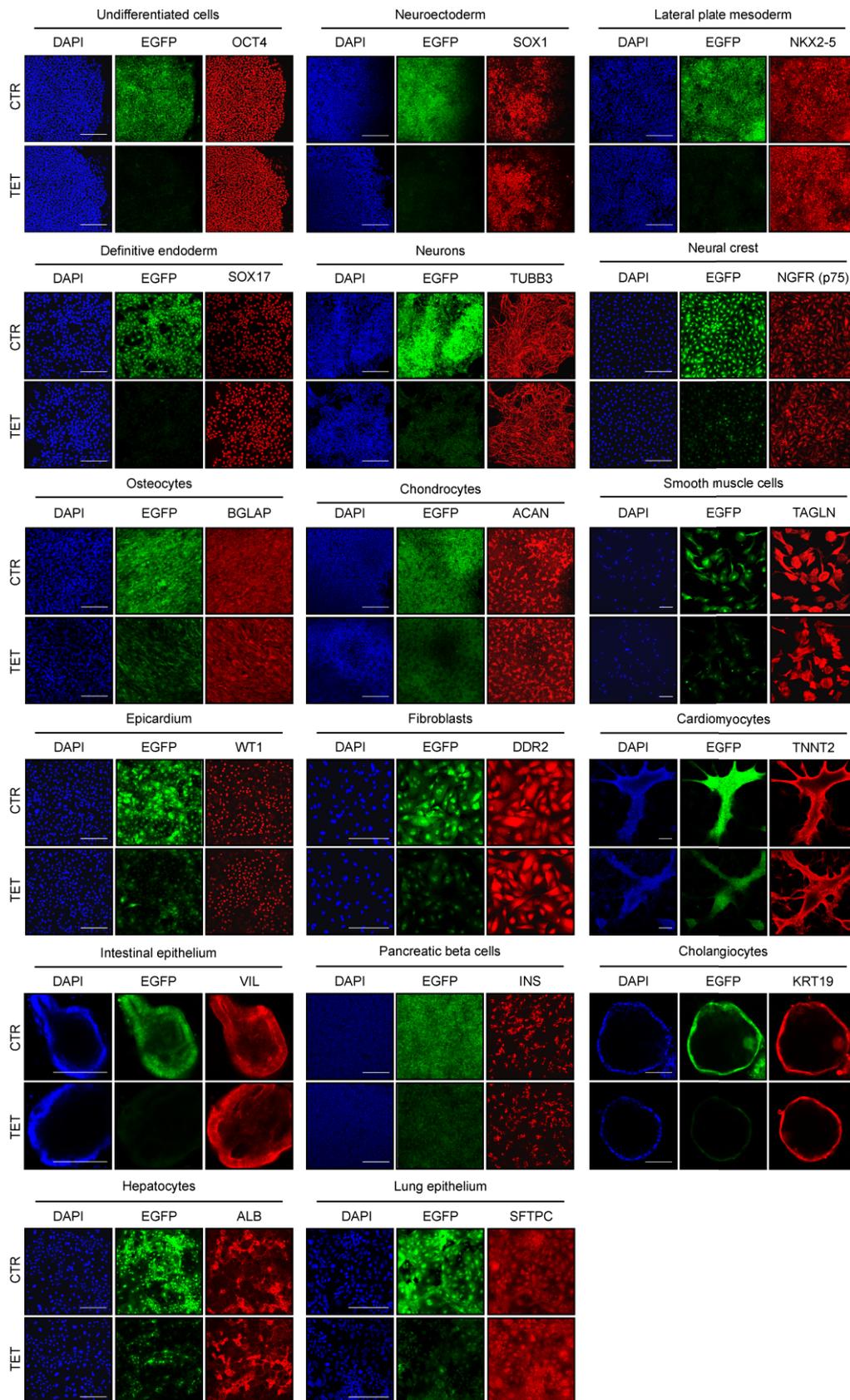


Figure 4.32. Validation of EGFP OPTiKD in differentiated cells.

Representative immunofluorescence of the indicated lineage-specific markers (red) in cells differentiated from EGFP OPTiKD hESCs. EGFP fluorescence in control conditions (CTR) or after 5 days of knockdown (TET) is in green, and DAPI (blue) shows nuclear staining. Scale bars: 100 μ m for intestinal epithelium and cholangiocytes; 200 μ m for all other lineages. Some of the differentiations were performed in collaboration as described in Figure 4.11.

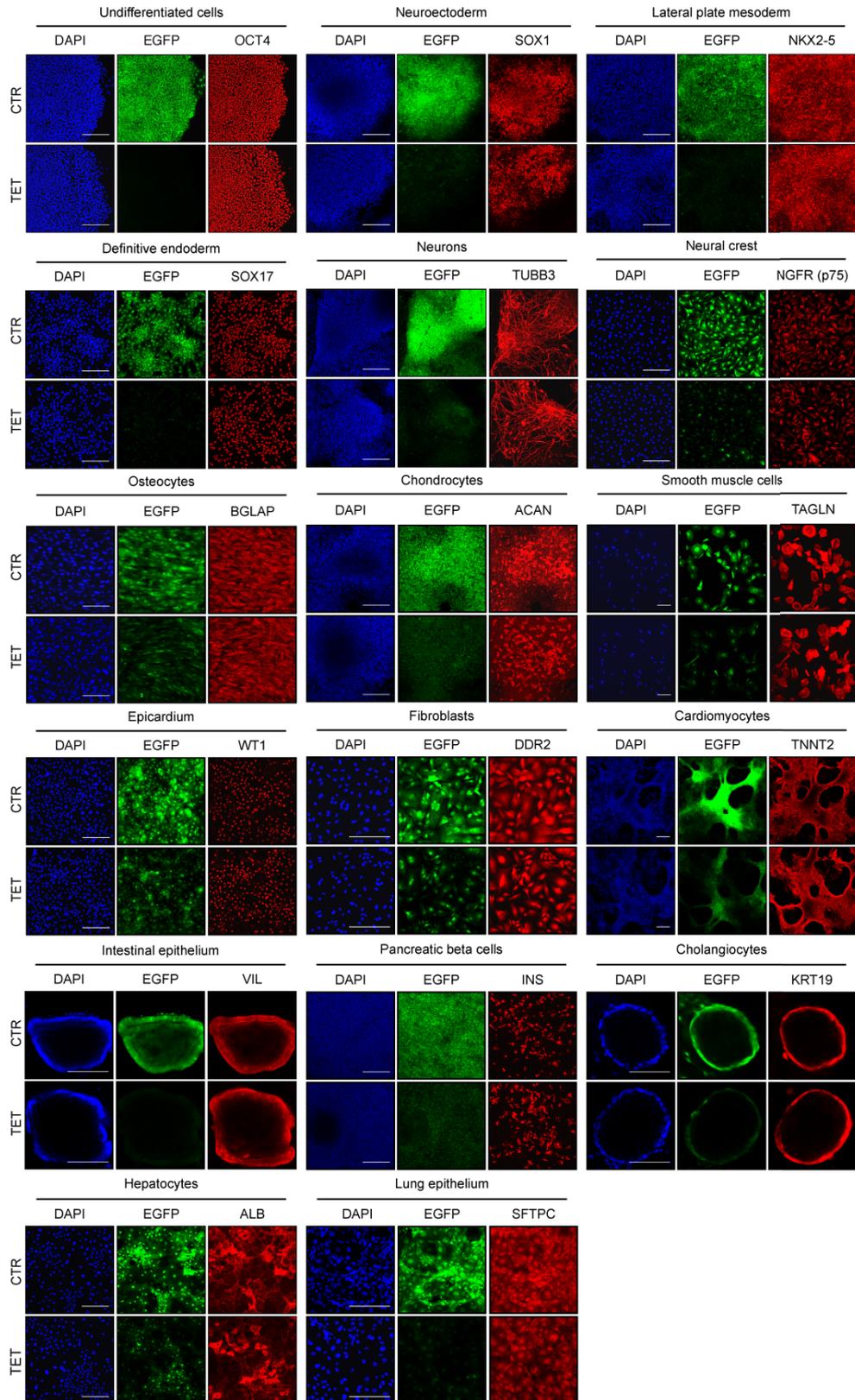


Figure 4.33. Validation of EGFP sOPTiKD in differentiated cells.
As in Figure 4.32, but for EGFP sOPTiKD hESCs.

Of note, EGFP was less reduced in cell types with slower proliferation rates (Figure 4.31). Since EGFP has a long half-life (more than 24h), protein loss upon transcriptional or post-transcriptional inhibition relies heavily upon its dilution following cell division (Li 1998). Considering the strong decrease of EGFP mRNA following induction of gene knockdown, we concluded that the residual protein observed in certain lineages was likely a consequence of the relatively short tetracycline treatment performed in these experiments. To test this hypothesis, we induced prolonged EGFP knockdown in cardiomyocytes, which are non-proliferative cells. Indeed, we observed a slow but constant decrease in EGFP protein expression for up to 20 days, at which point the levels were decreased by more than 75% (Figure 4.34).

Considered together, these results established that both OPTiKD and sOPTiKD allow efficient manipulation of gene expression even after hPSC differentiation.

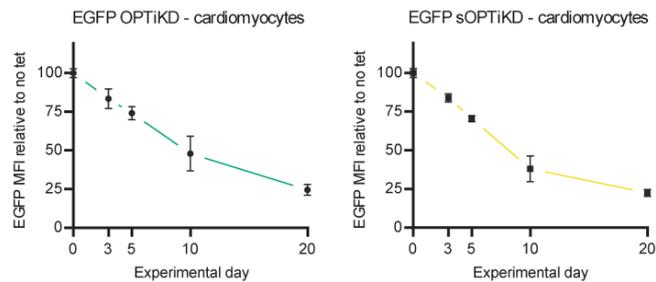


Figure 4.34. Prolonged EGFP inducible knockdown in differentiated cells. Kinetics of EGFP loss following prolonged treatment with tetracycline of cardiomyocytes derived from EGFP OPTiKD or sOPTiKD hESCs. Results are from two biological replicates per condition.

4.3.6 Inducible DPY30 knockdown at various stages of hPSC differentiation reveals stage- and lineage-specific functions.

Having developed a powerful platform for inducible gene knockdown, we applied this tool to study the function of DPY30 during hPSC differentiation. First, we generated DPY30 OPTiKD hESC sublines (Figure 4.35), and demonstrated that inducible DPY30 knockdown in hESCs impaired expression of pluripotency genes and increased neuroectoderm markers (Figure 4.36), in agreement with results obtained by constitutive shRNA expression (Chapter 3).

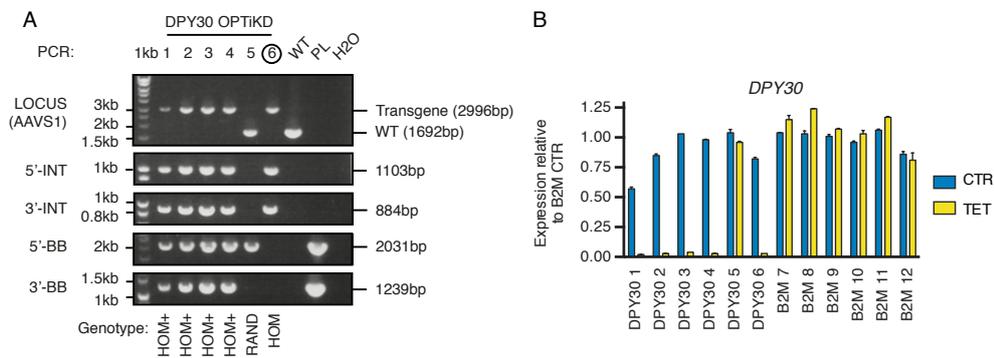


Figure 4.35. Generation of DPY30 OPTiKD hPSCs.

(A) Genotyping of DPY30 OPTiKD hESCs (targeting vector pAAV_iKD-DPY30). See Figure 4.2 and Figure 4.23 regarding the genotyping strategies. RAND: random integration of the targeting vector in the absence of AAVS1 targeting. Clone 6 was selected for further experiments. (B) qPCR validation of the lines from panel A cultured in absence (CTR) or presence of tetracycline for 5 days (TET). B2M OPTiKD hESCs (Figure 4.23) were analysed as negative controls. Note that clone 5 was the result of random integration of the targeting vector (panel A), thus explaining the lack of knockdown following induction. Individual clones were analysed in duplicate.

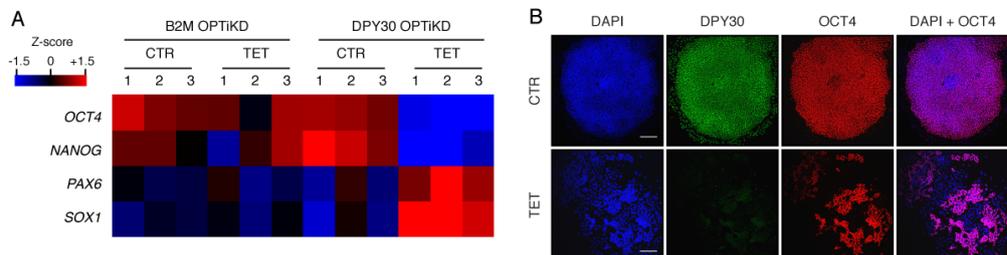


Figure 4.36. Inducible knockdown of DPY30 in hPSCs.

(A) Heatmap summarizing qPCR results from DPY30 and B2M OPTiKD hESCs maintained in control conditions (CTR) or following 10 days of knockdown (TET). Z-scores indicate differential expression measured in number of standard deviations from the average level. Three biological replicates per condition were analysed, as indicated. (B) Immunofluorescence for DPY30 and OCT4 (green and red) in DPY30 OPTiKD hESCs cultured as described in panel A. DAPI (blue) shows nuclear staining. Scale bars: 200µm.

Our previous results showed that DPY30 is required for efficient endoderm specification of hESCs. However, those experiments relied on chronic DPY30 knockdown, and as such did not allow us to dissect if DPY30 is important for the epigenetic priming of hESCs, for the induction of differentiation, or for both aspects. We therefore took advantage of the inducibility and reversibility of the OPTiKD platform to impair DPY30 expression at specific stages by carefully timing the tetracycline treatment of DPY30 OPTiKD hESCs (Figure 4.37). In particular, we induced DPY30 knockdown (1) both during self-renewal and endoderm specification; (2) only in the pluripotent state; (3) only during endoderm induction (Figure 4.37). These experiments showed that DPY30 was primarily required for the induction of endoderm differentiation, while DPY30-dependent hESC epigenetic priming was required only for a subset of markers. Finally, DPY30 knockdown during neuroectoderm or lateral plate mesoderm specification did not have any significant effect (Figure 4.38), confirming that DPY30 is specifically involved in endoderm induction.

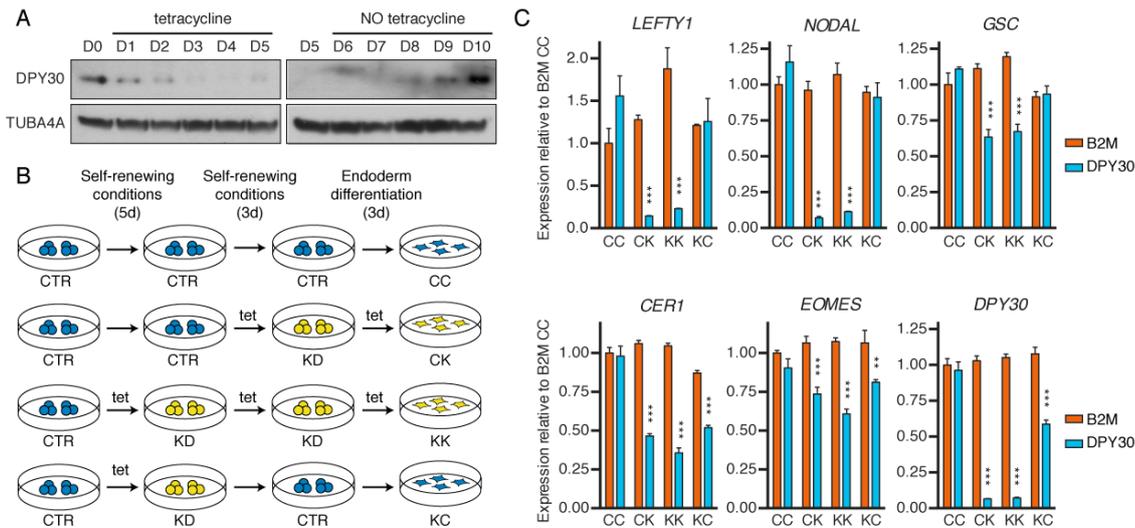


Figure 4.37. Inducible knockdown of DPY30 at various stages of endoderm differentiation.

(A) Kinetics of DPY30 protein knockdown and rescue in DPY30 OPTiKD hESCs. Cells were exposed to tetracycline for 5 days, followed by tetracycline withdrawal for 5 days. Samples were collected at the indicated days (D1-D10). D0: cells maintained in the absence of tetracycline and collected at the start of the experiment. TUBA4A (α -tubulin): loading control. (B) Schematic of the experimental approach to investigate the role of DPY30 during endoderm specification. DPY30 and B2M OPTiKD hESCs were cultured in absence or presence of tetracycline (tet) for the indicated days to induce (KD) or rescue (CTR) gene knockdown. Two-letters abbreviations summarize the expression of DPY30 (or B2M) during hESC self-renewal and endoderm differentiation (C: control; K: knockdown). (C) qPCR for the indicated genes in endoderm cells from the experiment described in panel B, and analysed at the end of the differentiation protocol. Results are from 3 biological replicates per condition. Significant differences vs B2M in the same condition are reported (2-way ANOVA with post-hoc Sidak comparisons).

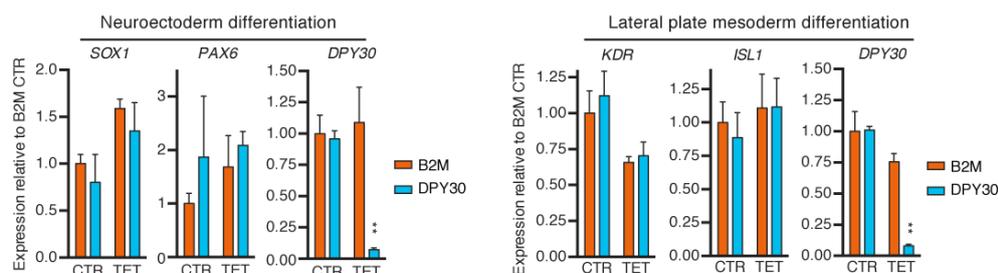


Figure 4.38. Inducible DPY30 knockdown during neuroectoderm and lateral plate mesoderm differentiation.

qPCR for the indicated genes in neuroectoderm and lateral plate mesoderm cells derived from DPY30 and B2M OPTiKD hESCs. Cells were cultured in absence or presence of tetracycline from the third day before differentiation until the end of the induction (knockdown only during differentiation, analogue to the “CK” condition described in Figure 4.37). Results are from 3 biological replicates per condition. Significant differences vs B2M in the same condition are reported (2way ANOVA with post-hoc Sidak comparisons).

We then analysed the function of DPY30 during differentiation of progenitors from the primary germ layers into fully differentiated cells. For this, we differentiated DPY30 OPTiKD hESCs into six different cell types, and induced DPY30 knockdown from the induction, specification, or maturation stages of each lineage (Figure 4.39). Importantly, qPCR analyses confirmed successful knockdown of DPY30 in all of these stages (Figure 4.40), thus validating this experimental approach.

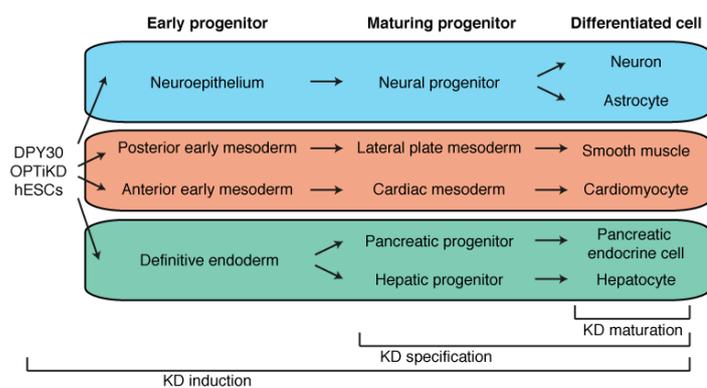


Figure 4.39. Inducible DPY30 knockdown at specific stages of hPSC differentiation into mature cell types.

Schematic summarizing the experimental approach. Refer to the methods (Chapters 2.2.2 and 2.4.3) and Figure 4.40 for details about the timing of tetracycline treatment during differentiation of each lineage.

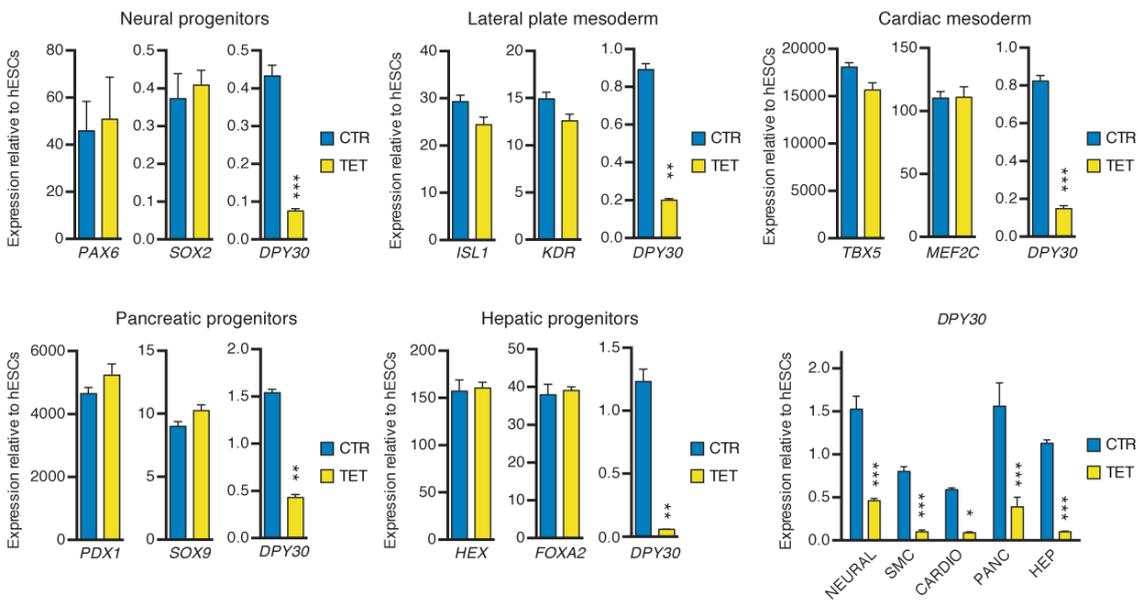


Figure 4.40. Validation of DPY30 knockdown during hPSC differentiation. qPCR for DPY30 and the indicated lineage-specific markers in maturing progenitors or differentiated cells derived from DPY30 OPTiKD hESCs (Figure 4.39). Before the analysis, cells were cultured in the absence (CTR) or presence of tetracycline (TET) for 3 or 7 days, respectively for maturing progenitor and differentiated cells. Results are from 3 biological replicates per condition. Significant differences vs CTR are reported (t-test).

Interestingly, phenotypic analyses demonstrated that DPY30 knockdown from the early induction of cardiac differentiation impaired cardiomyocyte specification, as shown by a decrease in contractile markers in fully differentiated cells (Figure 4.41). However, knockdown at later stages had no significant effects. Interestingly, a similar pattern was observed during hepatocyte differentiation, in which decrease of DPY30 expression from the induction of definitive endoderm progenitors led to extensive cell death at the anterior foregut stage, thus preventing the generation of mature hepatocytes (Figure 4.41). Similarly, specification of pancreatic endocrine cells was also impaired by knockdown of DPY30 from the induction of differentiation (Figure 4.41). Additionally, DPY30 knockdown only during the early differentiation of hPSCs into dorsal foregut cells was sufficient to impair further specification of pancreatic endocrine cells (Figure 4.42), suggesting the existence of DPY30-dependent epigenetic priming events during the early stages of pancreatic induction. However, neither hepatocytes nor pancreatic endocrine cell specification was significantly affected by knockdown of DPY30 in maturing progenitors or differentiated cells (Figure 4.41). Finally, DPY30 knockdown at any analysed stage of smooth muscle, neuronal, or astrocyte differentiation had no effects on the expression of key lineage markers (Figure 4.41).

Considered together, these findings suggest that the requirement for DPY30 expression during hPSC differentiation is stage- and lineage-specific (Figure 4.43). Moreover, these experiments illustrate how the OPTiKD platform can be applied to perform powerful functional studies at different steps of hPSC differentiation.

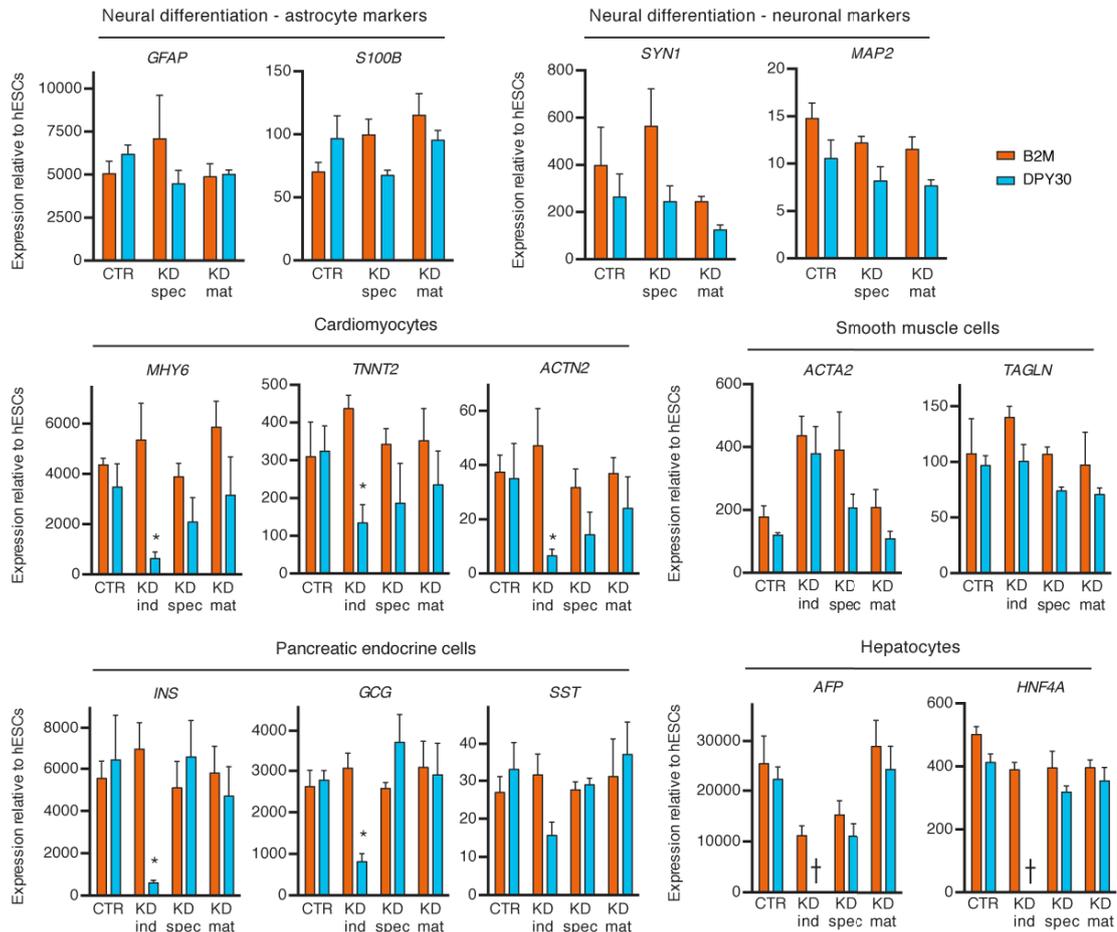


Figure 4.41. Effect of DPY30 inducible knockdown during hPSC differentiation.

qPCR-based phenotypic analyses of DPY30 and B2M OPTiKD hESCs after differentiation into the indicated mature cell types. Gene expression is reported relative to the average level in undifferentiated hESCs. CTR: no knockdown; KD ind/spec/mat: knockdown from induction, specification, or maturation (Figure 4.39). Results are from 3 biological replicates per condition. Significant differences vs B2M in the same condition are reported (2-way ANOVA with post-hoc Sidak comparisons). Cross symbols indicate that cells died during the differentiation and could not be analysed by qPCR. Neural differentiation was performed in collaboration with Dr Matthias Pawlowski.

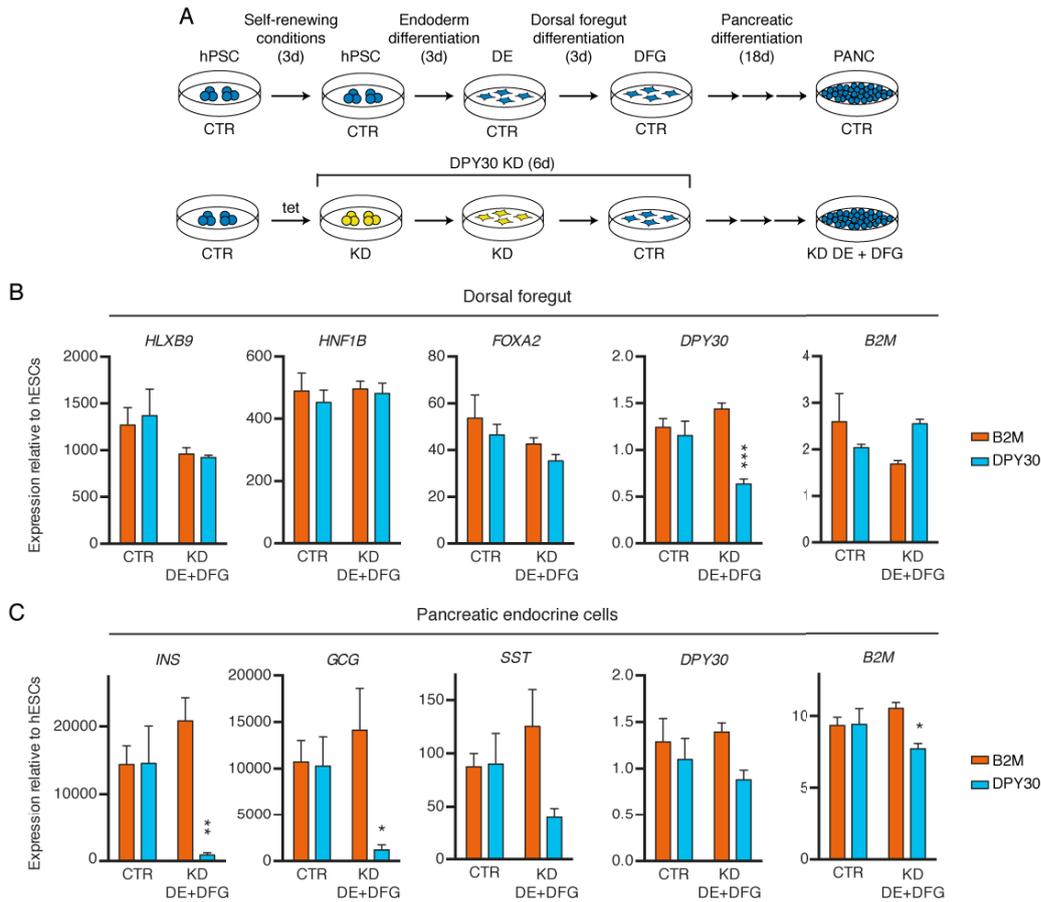


Figure 4.42. Knockdown of DPY30 specifically during the early induction of pancreatic endocrine cells.

(A) Schematic of the experimental approach: DPY30 OPTiKD hESCs were cultured with tetracycline only for three days before induction of endoderm differentiation to induce knockdown (KD) selectively in definitive endoderm (DE) and dorsal foregut cells (DFG). (B-C) Expression of DPY30, B2M, and the indicated lineage specific markers in cells from the experiment described in panel A. Cells were analysed at the dorsal foregut stage (B) or at the end of the differentiation (C). Results are from 3 biological replicates per condition. Significant differences vs B2M in the same condition are reported (2way ANOVA with post-hoc Sidak comparisons).

DPY30 KD:	Induction	Specification	Maturation
Neurons	-	No effect	No effect
Astrocytes	-	No effect	No effect
Smooth muscle	No effect	No effect	No effect
Cardiomyocytes	Impaired differentiation	No effect	No effect
Endocrine pancreas	Impaired differentiation	No effect	No effect
Hepatocytes	Differentiation failure	No effect	No effect

Figure 4.43. Lineage- and stage-specific effects following DPY30 knockdown during hPSC differentiation.

Schematic summary of the results presented in Figure 4.41 and Figure 4.42. Note that the function of DPY30 during the induction of neural differentiation was not addressed in the current study.

4.4 Discussion

This chapter described OPTiKD and sOPTiKD: two new platforms for inducible knockdown of gene expression that address the limitations of previous methods. Of particular interest, we took advantage of GSHs to avoid silencing and mosaic expression of transgenes not only in hPSCs, but also in hPSC-derived cells. Indeed, we performed the largest *in vitro* characterization to date of both the ROSA26 and AAVS1 loci during hPSC differentiation, and demonstrated that the methods we developed can be used to study gene function in a broad number of cell types of considerable interest. Moreover, we engineered a codon-optimized version of the tetR gene that strongly increases its expression levels in human cells. By using such improved transgene, we prevented the common problem of shRNA leakiness in the TET-OFF system, while achieving robust and reproducible knockdown upon induction. Finally, we designed an all-in-one targeting vector that allows straightforward application of this technology in different hPSC lines. Compared with alternative approaches that rely on viral transduction or random integration of inducible shRNAs (Chapter 1.8.2; Lambeth and Smith, 2013; Zafarana et al., 2009), the OPTiKD and sOPTiKD platforms are simpler to use (plasmid-based), quicker (2 weeks to generate stable lines following lipofection), more efficient (>95% of clones showing inducible knockdown), more scalable (isolation of clonal sublines can be entirely bypassed), and more robust (due to the use of GSHs and to the lack of shRNA leakiness). Collectively, these features will greatly facilitate functional studies in a diversity of cell types.

Alternatively to gene knockdown experiments, loss-of-function studies can be performed by generating knockouts. However, as described in Chapter 1.8.2, current methods for inducible knockout in hPSCs suffer from extensive limitations. Aside from these practical considerations, there are important conceptual differences between inducible knockdown and inducible knockout approaches. First, knockdown can be reverted, thus enabling the investigation of gene function only at a specific stage of development. Moreover, the ability to control the level of knockdown allows the study of genes for which complete loss-of-function induces cell death. Finally, this same property facilitates examining gene dosage mechanisms, whose role in human disease is prominent. On the other hand, phenotypic studies following full gene knockout are more powerful in the case of genes that are still functional even when lowly expressed. Moreover, knockout approaches are applicable not only to genes, but also to non-coding genomic regulatory

regions, which represent a large portion of disease-associated genetic traits (Cooper & Shendure 2011). Overall, depending on the experimental approach and biological question to be investigated, gene knockdown and knockout approaches presents distinct advantages and disadvantages and should be chosen accordingly (Boettcher & McManus 2015).

Of particular interest to the scope of this dissertation, we took advantage of the OPTiKD platform to investigate the function of DPY30 during hPSC differentiation. First, we assessed the role of DPY30 during germ layer specification, and confirmed our previous results by demonstrating that DPY30 is required for efficient activation of Activin/Nodal target genes and induction of the definitive endoderm lineage. However, our results showed that only a subset of genes require expression of DPY30 in the undifferentiated state for proper activation during differentiation, while several key regulators of endoderm specification can be efficiently induced even if DPY30 is expressed only during germ layer induction. Given the established link between DPY30 and H3K4me3 deposition onto endoderm regulators (Chapter 3), this suggests that the formation of bivalent domains onto certain endoderm genes might be dispensable for their proper activation during differentiation. However, these results must be interpreted with caution, and further experiments will be required to more formally test this hypothesis. Indeed, our analysis focused only on a subset of genes that might not be fully representative of the genome-wide regulations. Moreover, it will be important to monitor the changes of H3K4me3 deposition following the knockdown and rescue of DPY30 expression. Indeed, it is possible that fluctuations of the levels of DPY30 might differentially impact the epigenetic state of various endodermal regulators. Accordingly, while knockdown of DPY30 is a useful genetic tool to probe the function of H3K4me3, it does not completely abolish H3K4me3 (Chapter 3.3.6). Therefore, it will be important to assess the function of H3K4me3 during germ layer specification also with other complementary approaches, including complete loss-of-function studies for DPY30 as well as for other COMPASS complex proteins. Finally, it will be pivotal to address the relevance of these mechanisms during *in vivo* development, for instance by using conditional Dpy30 knockout mice.

We also investigated the function of DPY30 at different developmental stages of several mature lineages arising from all of the germ layers. These results showed that DPY30 expression is required for the differentiation of certain cell types (hepatocytes,

pancreatic endocrine cells, and cardiomyocytes), but only at an early stage of their specification. These findings further support a crucial role for DPY30 during endoderm differentiation, and show that expression of DPY30 during this early stage is necessary for further specification of mature lineages (in agreement with our previous experiments; Chapter 3.3.4). Moreover, these results suggest that DPY30 might also be involved in the specification of cardiac mesoderm. Interestingly, these two early lineages arise from the mid-anterior portion of the primitive streak in the gastrulating embryo, and rely on Nodal for their early specification (Chapters 1.1.3 and 1.3.2.2). Given our previous finding that DPY30 mediates Activin/Nodal-dependent H3K4me3, we speculate that the lineage-specific function of DPY30 during early cell fate decisions might reflect different requirements for Activin/Nodal-dependent H3K4me3 deposition. However, future detailed studies will be necessary to rigorously test this hypothesis, as it is possible that some of the phenotypic changes observed following DPY30 knockdown are independent from its function downstream of Activin/Nodal signalling.

One additional point of interest of our work is the lack of overt phenotypic effects following knockdown of DPY30 during late stages of hPSC differentiation. These unexpected results suggest that the deposition of H3K4me3 might be required only during the germ layer specification, but might be largely dispensable once this early cell identity is established. However, these results must again be interpreted with care, and will need to be further explored. Indeed, the experimental limitations mentioned above in the context of germ layer specification also apply also to these experiments.

Aside from developmental studies in hPSCs, we envision several other potential applications of the OPTiKD technologies (also see Chapter 6.1.4). The high targeting efficiency and scalability of OPTiKD could allow high-throughput RNA interference screenings by targeting inducible shRNA pools. Compared to viral-based approaches (Chen et al. 2012; Chia et al. 2010), the isogenic integration of inducible shRNAs would reduce heterogeneity in the targeted population, hence increasing the screening sensitivity and specificity. On the other hand, sOPTiKD could allow the simultaneous targeting of several hiPSC lines to probe gene function in different genetic backgrounds. Such an approach could facilitate the identification of genetic disease modifiers, and facilitate the discovery of novel potential drug targets in the context of personalized medicine. Finally, these platforms could be transferred to other cell types amenable to

genetic manipulation, including established cell lines and adult stem cells (Drost et al. 2015; Mandal et al. 2014), thus allowing functional studies in a multitude of systems.

Finally, by developing an improved method for transgene expression in the human ROSA26 locus, we widened the range of human loci available for robust hPSC genome engineering. Accordingly, we have recently used the dual GSHs targeting approach described here to optimize inducible overexpression of transgenes in hPSCs using the tetracycline activatable (TET-ON) system (Matthias Pawlowski, Daniel Ortmann, Alessandro Bertero et al., manuscript in review at *Stem Cell Reports*). For this, we inserted two copies of a reverse tetracycline-controlled transactivator (rtTA) expression cassette into the ROSA26 locus, and two copies of an inducible transgene expression cassette into the AAVS1 locus. By avoiding promoter interference and by maximizing the number of transgenes copies, this approach enabled strong and homogeneous inducible transgene overexpression in hPSCs upon addition of the tetracycline analogue doxycycline, thereby supplying a powerful gain-of-function system. Overall, the work presented in this chapter described general principles that could be applicable to several types of genetic manipulation in human cells.

5 IDENTIFICATION OF NOVEL MOLECULAR MECHANISMS REGULATED BY ACTIVIN/NODAL

5.1 Hypothesis

As introduced in Chapter 1.7, the Activin/Nodal signalling pathway is known to modulate cellular responses primarily through transcriptional and epigenetic regulations controlled by the intracellular effector SMAD2/3. However, while previous work (as well as the results presented in Chapters 3 and 4) partially clarified some of the mechanistic aspects of SMAD2/3 activity in hPSCs, a complete picture is still severely lacking. Moreover, whether mechanisms other than transcriptional and epigenetic controls might be regulated by Activin/Nodal signalling is unknown.

We hypothesized that we could obtain a more complete and unbiased understanding of the molecular functions regulated by Activin/Nodal signalling in hPSCs by interrogating the whole spectrum of factors interacting with SMAD2/3 through mass spectrometry analysis of SMAD2/3 co-immunoprecipitated proteins.

5.2 Results

5.2.1 Optimization of the SMAD2/3 co-immunoprecipitation protocol

The first step towards defining the SMAD2/3 interactome was to identify a co-immunoprecipitation protocol allowing specific identification of the greatest number of interactors following mass spectrometry analysis. Indeed, the biochemical conditions used for protein extraction and immunoprecipitation have a profound effect on the stability of various protein-protein interactions (Sambrook & Russell 2006). Of note, SMAD2 and SMAD3 interact with several of their known binding partners through a set of contiguous hydrophobic patches, referred to as the “hydrophobic corridor”, which is located on the surface of the MH2 domain (Massagué et al. 2005). Therefore, biochemical conditions that stabilize hydrophobic interactions might be preferable. To test this, we compared two SMAD2/3 co-immunoprecipitation (co-IP) methods that are expected to preferentially preserve different types of protein-protein interactions.

First, we tested the protocol that we had previously used to demonstrate the interaction of SMAD2/3 with COMPASS complexes (Chapter 3). This method, which we named co-IP1, relies on an isotonic buffer with low concentration of mild detergent (0.1% Tween-20) both to solubilize the nuclear proteins and to minimize background binding during immunoprecipitation. As such, these biochemical conditions are likely to largely preserve salt-sensitive hydrophilic bonds, while the presence of detergent might partially interfere with hydrophobic interactions. Secondly, we developed a different protocol that avoids the use of detergent at any step, and in which nuclear proteins are extracted using a high-salt buffer followed by dialysis of the lysate to re-adjust the salt content to physiological levels (co-IP2). In this case, hydrophobic interactions should be better preserved due to the lack of detergent. However, the high-salt conditions used for the nuclear extraction could disrupt certain hydrophilic bonds, which might only be partially re-established following dialysis.

Initial comparison in hESCs demonstrated that both conditions allowed to detect well-characterized SMAD2/3 binding factors by Western blot, with co-IP2 being slightly more efficient than co-IP1 (Figure 5.1). We therefore performed small-scale pilot mass spectrometry analyses of SMAD2/3 co-IPs from hESCs using both methods (Figure 5.1).

For these tests, the SMAD2/3 and IgG negative control co-IPs were differentially labelled post-IP using the dimethyl method (Hubner & Mann 2011), followed by a combined run of the two samples in order to compare the abundance of specific peptides and identify enriched ones. Remarkably, co-IP2 allowed the identification of a larger number of SMAD2/3 interacting proteins following statistical analysis (23, compared to 12 for co-IP1). Interestingly, roughly half of the proteins significantly enriched in co-IP1 experiments were also identified as specific binders in co-IP2 ones. On the other side, co-IP2-specific interactors included many notable transcription factors (such as SOX13, ETV6, and SMAD4), epigenetic regulators (like SETDB1 and ATF7IP), and RNA-binding proteins (for instance WTAP and CPSF6).

Overall, these results showed that the co-IP2 protocol is more suitable for the large-scale analysis of SMAD2/3 interacting proteins. We therefore chose such method for our following experiments.

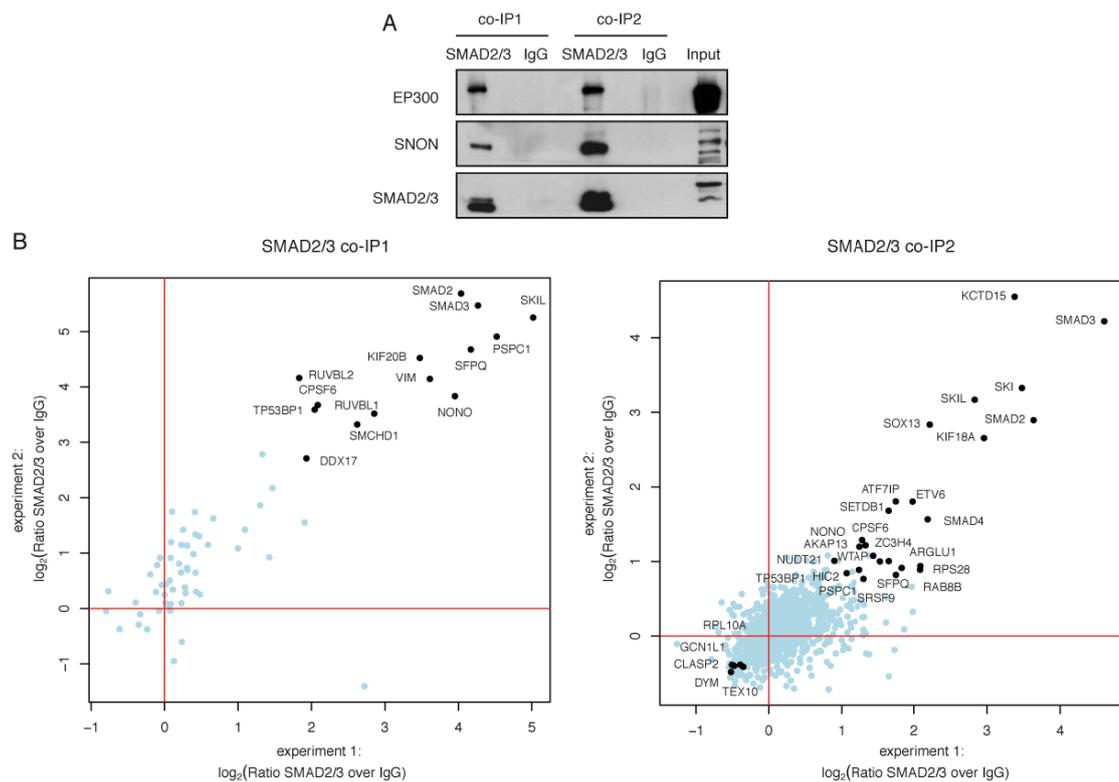


Figure 5.1. Comparison of SMAD2/3 co-immunoprecipitation protocols.

A) Western blots of SMAD2/3 or control (IgG) immunoprecipitations from nuclear extracts of hESCs following the co-IP1 or co-IP2 protocols. Input is 5% of the material used for IP. Co-IP2 experiments were performed in collaboration with Dr Sasha Mendjan. B) Scatter plots of the \log_2 ratios of label-free quantification (LFQ) intensities for proteins identified by quantitative mass spectrometry in SMAD2/3 co-IPs compared with IgG negative control co-IPs. The values for technical dye-swap duplicates are plotted on different axis, and protein whose enrichment was significant (significance $B < 0.01$; see Chapter 2.5.2) are shown in black and named. Mass spectrometry analysis was performed in collaboration with Dr Nina C. Hubner.

5.2.2 Identification of the SMAD2/3 interactome

Having identified a suitable SMAD2/3 co-immunoprecipitation protocol, we performed large-scale proteomic experiments to determine the SMAD2/3 interactome in hESCs. For this, we focused on two conditions: self-renewing hESCs and hESCs induced to differentiate towards endoderm for 36h. Indeed, we hypothesized that this would allow the identification of SMAD2/3 interactors specifically involved in hESC pluripotency or differentiation. SMAD2/3 and IgG negative control IPs were performed in triplicate for each condition, and the samples were analysed by label-free mass spectrometry to allow quantitative comparisons between all samples (Figure 5.2; Hubner and Mann, 2011). Collectively, this generated ~160000 isotope patterns resulting from ~6000 mass spectrometry (MS) runs. Consequently, ~33000 tandem mass spectrometry (MS/MS) spectra were measured and matched to ~20000 known peptides. Overall, this led to the identification of 3635 proteins in at least one of the conditions analysed.

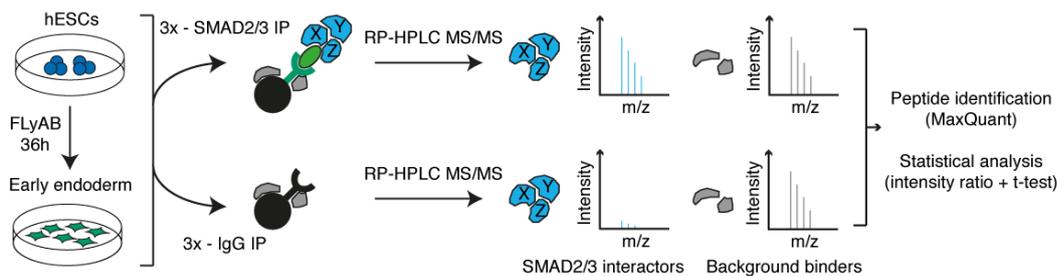


Figure 5.2. Identification of the SMAD2/3 interactome by label-free quantitative mass spectrometry.

Schematic of the experimental approach. FLYAB: cells cultured in 12ng/ml FGF2, 10 μ M LY-294002 (PI3K inhibitor), 100ng/ml Activin, and 10ng/ml BMP4 (Chapter 2.2.2.1). RP-HPLC MS/MS: reverse phase high-performance liquid chromatography tandem mass spectrometry. m/z: mass to charge ratio.

SMAD2/3-specific interactors were then determined by comparing the SMAD2/3 immunoprecipitations in each cell type to the matched IgG control by means of t-tests. This identified 65 and 63 putative SMAD2/3 interacting proteins in pluripotent hESCs and early endoderm cells, respectively (Figure 5.3 and Appendix III). Of note, 39 factors were significantly enriched in both conditions, indicating a substantial overlap in the two datasets (Figure 5.3). Collectively, we found 89 candidate SMAD2/3 partners. Of note, these included 12 proteins whose potential interaction with SMAD2/3 was already annotated in protein-protein interaction databases (Appendix III and Figure 5.5). Among these, there were several known *bona fide* SMAD2/3 interactors (Chapter 1.7), such as the co-SMAD SMAD4, the transcriptional corepressors SKI and SNON, the

transcriptional cofactor FOXH1, the histone acetyltransferases EP300 and CREBBP, and the E3 ubiquitin ligases WWP1 and WWP2 (Figure 5.3). Moreover, the collective list of potential SMAD2/3 interactors contained 16/23 (70%) of the factors identified in the preliminary small-scale test of the co-IP2 protocol (Figure 5.1), thus confirming the reproducibility of our results across independent experiments. Overall, we concluded that the extensive list we generated represents a large snapshot of the SMAD2/3 interactome in self-renewing and differentiating hESCs.

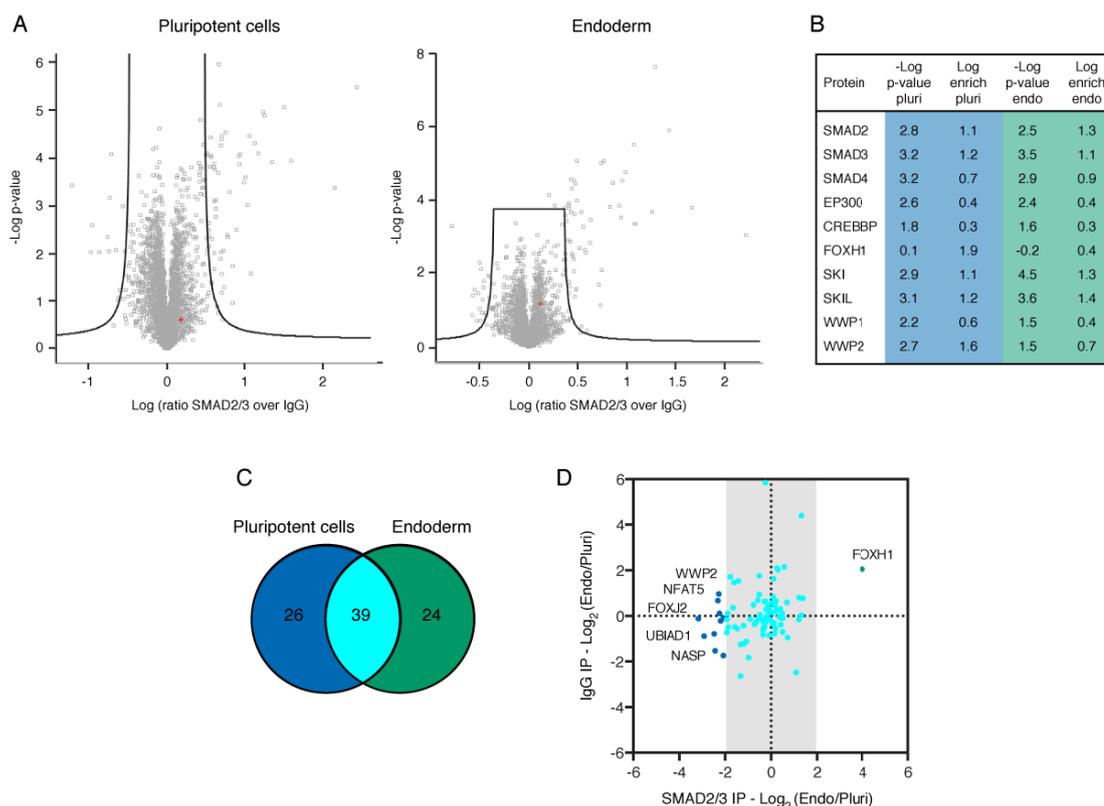


Figure 5.3. The SMAD2/3 interactome in hPSCs and early endoderm cells.

(A) Volcano plots of statistical significance against fold-change for proteins identified by mass spectrometry in SMAD2/3 or IgG negative control IPs in pluripotent cells or early endoderm. The black lines indicate the threshold used to determine specific SMAD2/3 interactors (FDR<0.05; n=3), which are located to the right. Mass spectrometry analysis was performed in collaboration with Dr Nina C. Hubner. (B) Selected results of the analysis described in panel A for SMAD2, SMAD3, and selected known *bona fide* SMAD2/3 binding partners. (C) Venn diagrams reporting the number of putative SMAD2/3 interactors in the two conditions examined. (D) Scatter plot of LFQ intensity log₂ ratios in endoderm (endo) and pluripotency (pluri) for all putative SMAD2/3 interactors. Values for the SMAD2/3 IPs or IgG negative control IPs are plotted against each other to facilitate the estimation of differential enrichment due to changes in SMAD2/3 binding and/or in the expression of the interactor (which can be estimated by changes of background binding to the IgG negative control). The shaded area depicts the cut-off chosen to determine proteins showing enrichment in pluripotent cells (blue dots) or early endoderm (green dot), and selected protein names are reported.

We then examined the changes in the SMAD2/3 interactome between self-renewal and differentiation (Figure 5.3). Remarkably, only a handful of factors showed a substantially different enrichment between SMAD2/3 co-IPs performed in these conditions ($\text{abs.FC} > 2$, as label-free mass spectrometry is at present not sensitive enough to detect smaller changes with confidence; Hubner and Mann, 2011). These included 10 pluripotency-enriched proteins (such as the E3 ubiquitin ligase WWP2 and the transcription factors FOXJ2 and NFAT5) and a single endoderm-enriched protein (the transcription factor FOXH1). Overall, aside from a few notable exceptions the SMAD2/3 interactome appears largely conserved between self-renewing and endoderm-differentiating hESCs,

Functional enrichment analysis showed that, as expected, nuclear transcriptional and cotranscriptional regulators are highly represented among SMAD2/3 interacting proteins (Figure 5.4). However, several other classes of factors also appeared to be enriched in the list. These included multiple molecular functions that have never been previously described to be functionally related to SMAD2/3, such as mRNA processing, modification, and degradation, as well as DNA repair (Figure 5.4). Of note, SMAD2/3 interactors were enriched in genes whose knockout results in developmental defects and embryonic lethality, in agreement with the crucial developmental role of SMAD2/3 (Figure 5.4).

GO molecular function term	Rank	Adj. p-value	Genes	GO biological process term	Rank	Adj. p-value	Genes
RNA Pol II transcription factor binding	1	6.72E-06	8	mRNA processing	1	1.53E-03	12
Poly(A)-specific ribonuclease activity	3	1.97E-03	3	mRNA modification	2	2.60E-02	3
SMAD binding	4	1.97E-03	5	Gene expression	4	4.25E-02	12
Transcription corepressor activity	8	2.57E-03	7	mRNA deadenylation-dependent decay	5	4.25E-02	4
Activating transcription factor binding	9	4.42E-03	4	Embryonic organ development	11	4.78E-02	5

GO cellular compartment term	Rank	Adj. p-value	Genes	MGI mouse phenotype term	Rank	Adj. p-value	Genes
Nuclear body	1	7.33E-07	11	Abnormal embryogenesis/development	1	1.92E-08	19
Transcription factor complex	2	4.06E-06	11	Abnormal survival	2	7.16E-06	24
Nucleotide-excision repair complex	7	2.84E-04	3	Abnormal craniofacial morphology	5	2.82E-02	8
Methyltransferase complex	11	6.52E-03	4	Abnormal respiratory system	6	2.82E-02	6
mRNA cleavage factor complex	16	2.19E-02	3	Abnormal digestive system	7	2.82E-02	7

Figure 5.4. Functional role of SMAD2/3 interacting proteins.

Selected results from GO enrichment analysis, and enrichment analysis for mouse phenotypes annotated in the Mouse Genomics Informatics (MGI) database. For each term, its rank in the analysis, the adj.p, and the number of associated genes are reported.

In order to further explore the molecular function of SMAD2/3-associated proteins, we investigated the known interactions between such factors by interrogating public protein-protein interaction databases (Figure 5.5). This analysis revealed that a large proportion of SMAD2/3 partners are involved in intermolecular interactions with each

other, and that some of them form well-characterized protein complexes. These factors include chromatin epigenetic modifiers already known to interact with SMAD2/3, such as the transcriptional activators EP300 and CREBBP and the transcriptional repressor SETDB1 (Chapter 1.7.3). These proteins interact either directly with SMAD2/3, or through known SMAD2/3 transcriptional coactivators (FOXH1 and SMAD4) or corepressors (SKI and SNON). As such, this sub-network likely represents the central hub that regulates the transcriptional activity of SMAD2/3 in hESCs.

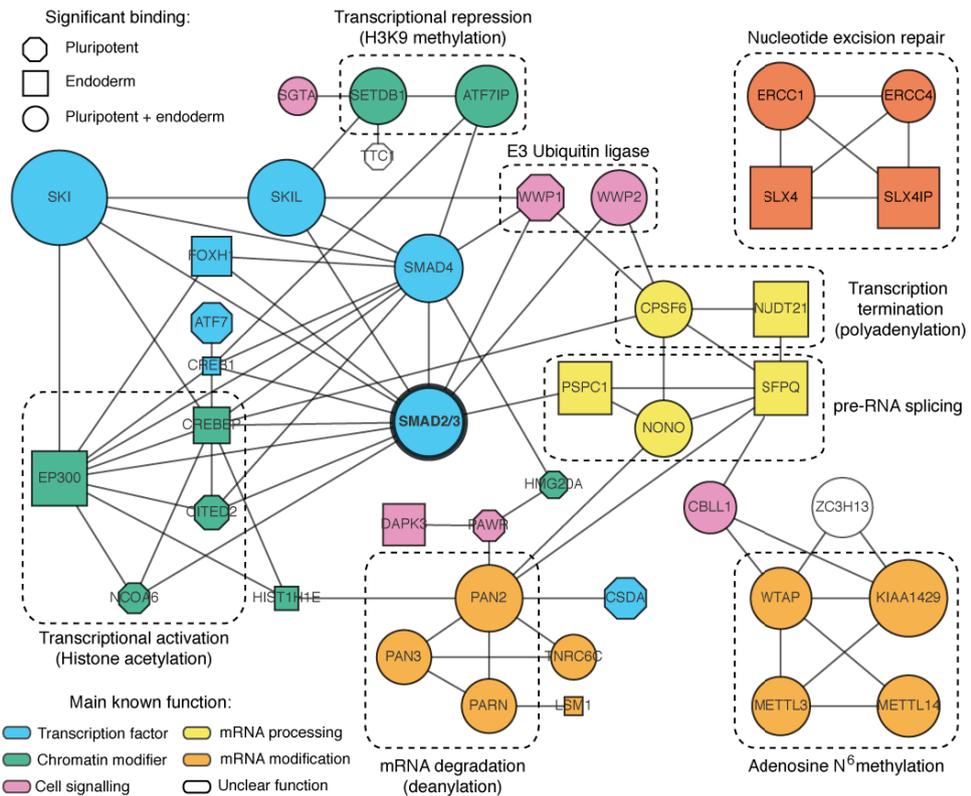


Figure 5.5. Known interactions between the SMAD2/3 interacting proteins.

Interaction network generated from all known protein-protein interactions between SMAD2/3 partners identified in pluripotent and endoderm cells. Note that based on our results all the proteins shown would be connected to SMAD2/3, but such links were omitted to simplify visualization and highlight those interactions with SMAD2/3 that were already known. Proteins lacking any link and small complexes of less than three factors are not shown to improve presentation clarity. The graphical representation of nodes conveys the following information: (1) which lineage the proteins were significantly enriched in (node shape, see legend on the figure); (2) how significant the enrichment was (the node size is proportional to the maximum -log p-value, hence the larger the more significant); (3) the main known function of the factors, as manually annotated from the available literature (node fill colour, see legend on the figure). Selected well-characterized protein complexes of interest are marked. Note that since the nodes representing SMAD2 and SMAD3 shared the very same links, they were fused into a single node (SMAD2/3, in the centre and with a thicker border).

Interestingly, our analyses showed that SMAD2/3 also interacts with multiple complexes involved in RNA regulation, such as the METTL3-METTL14-WTAP complex

(which mediates adenosine N⁶-methylation; Chapter 1.6.2.1), the PARBP-dependent poly(A) nuclease complex (hPAN, which regulates mRNA decay; Uchida et al., 2004), the cleavage factor Im complex (CFIm, involved in pre-mRNA 3' end processing; Rügsegger et al., 1996), and the NONO-SFPQ-PSPC1 complex (which has several functions, including RNA splicing and nuclear retention of defective RNAs; Knott et al., 2016). Finally, additional notable complexes found in the SMAD2/3 interactome are involved in DNA repair (the ERCC1-XPF nucleotide excision repair complex; Manandhar et al., 2015), and in the regulation of apoptosis (DAPK3-PAWR; Boosen et al., 2009)

Overall, these results suggested that SMAD2/3 might be involved in a large number of biological processes in hESCs, which include not only established mechanisms of transcriptional and epigenetic regulations, but also novel “non-canonical” functions, most notably RNA regulation.

5.2.3 SMAD2/3 transcriptional and epigenetic cofactors regulate hPSC pluripotency and differentiation

Having identified the SMAD2/3 interactome in hESCs, we began its validation and functional characterization. First, we focused our attention on the transcriptional and epigenetic cofactors constituting the central hub around SMAD2/3.

Western blots of SMAD2/3 co-IPs validated the interaction with a number of molecular partners identified in our proteomic screening (Figure 5.6). Of note, the epigenetic cofactors CREBBP and SETDB1 were bound at similar levels by SMAD2/3 in hESCs and cells differentiating into endoderm, while the interaction with transcriptional cofactors proved more dynamic. Indeed, SMAD2/3 specifically bound to NANOG in hESCs (Figure 5.6), in agreement with previous reports and our own earlier results (Chapter 3; Vallier, Mendjan, et al. 2009; note that NANOG was, however, not identified in our proteomic screening, likely due to the known technical limitations in its detection by mass spectrometry, see Chapter 5.3). On the other hand, SMAD2/3 binding to FOXH1 was remarkably stronger in endoderm cells (Figure 5.6). Finally, the interaction with the corepressor SNON was diminished during differentiation, in agreement with previous studies (Tsuneyoshi et al. 2012).

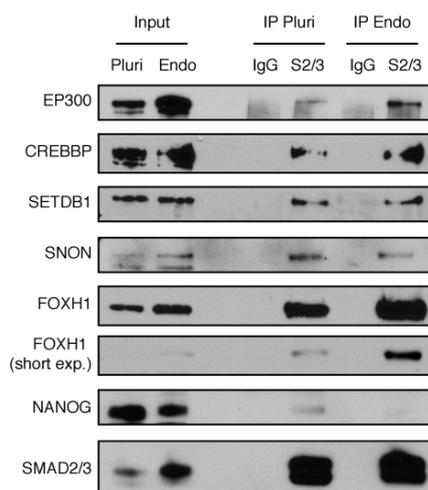


Figure 5.6. Interaction dynamics of SMAD2/3 transcriptional and epigenetic cofactors in pluripotent and early endoderm cells.

Western blots of SMAD2/3 or control (IgG) immunoprecipitations (IP) from nuclear extracts of pluripotent hESCs (Pluri) or hESCs differentiated into endoderm for 36h (Endo). Input is 5% of the material used for IP. For FOXH1, an additional short exposure is shown to facilitate semi-quantitative assessment of protein abundance. Immunoprecipitations were performed in collaboration with Ms Stephanie Brown.

These results support the view that dynamic changes in SMAD2/3 transcriptional cofactors might regulate its transcriptional activity during self-renewal and differentiation (Chapter 1.7.4). In this context, the crucial role of NANOG and SNON in this context has been previously demonstrated (Vallier, Mendjan, et al. 2009; Teo et al. 2011; Tsuneyoshi et al. 2012). On the other hand, despite the abundant evidence that FOXH1 serves as a SMAD2/3 cofactor (Chapter 1.7.3), its role in hPSCs has surprisingly not been directly addressed to date. Moreover, while the cooperation of SMAD2/3 with the coactivators EP300 and CREBBP is well established in multiple cell types (Chapter 1.7.3), the function of such factors in hPSCs has not been determined. Finally, SETDB1 was only recently reported to interact with SMAD2/3 in cancer cells (Wu et al. 2014), and its relevance in hPSCs is poorly understood. Overall, we decided to investigate the role of these four SMAD2/3 cofactors in the context of hPSC pluripotency and differentiation.

To this aim, we took advantage of our recently established inducible gene knockdown platform (Chapter 4; note that for these experiments we used the method described in Chapter 4.3.1 and based on random integration of the tetR, as the OPTiKD and sOPTiKD platforms were still being developed). Further, we developed a pipeline for the rapid functional characterization of multiple genes in parallel (Figure 5.7).

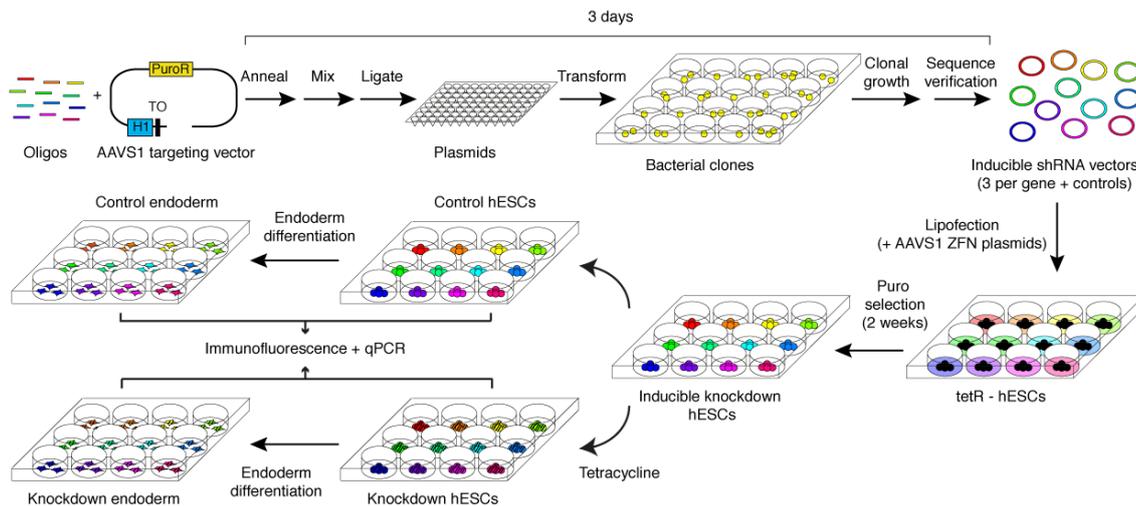


Figure 5.7. Functional characterization of the SMAD2/3 interactome. Schematic of the experimental approach. Following multiplexed generation of inducible shRNA vectors and targeting of tetR-hESCs, the resulting clonal pools were directly experimentally analysed. Refer to Chapters 4 and 2.5.5 for further details on the methodology.

Using this method, we generated 3 inducible shRNAs for each SMAD2/3 cofactor and for SMAD2 itself (in order to have a positive control), and two negative control shRNAs (SCR and B2M shRNAs). We then obtained inducible knockdown hESCs by gene targeting of the AAVS1 locus followed by drug selection (note that clonal isolation of individual sublines was bypassed given the high efficiency of the method). Finally, hESCs showing robust (>70%) inducible knockdown were selected for further analyses (Figure 5.8). Of note, we validated that knockdown could be efficiently induced also during endoderm differentiation (Figure 5.12).

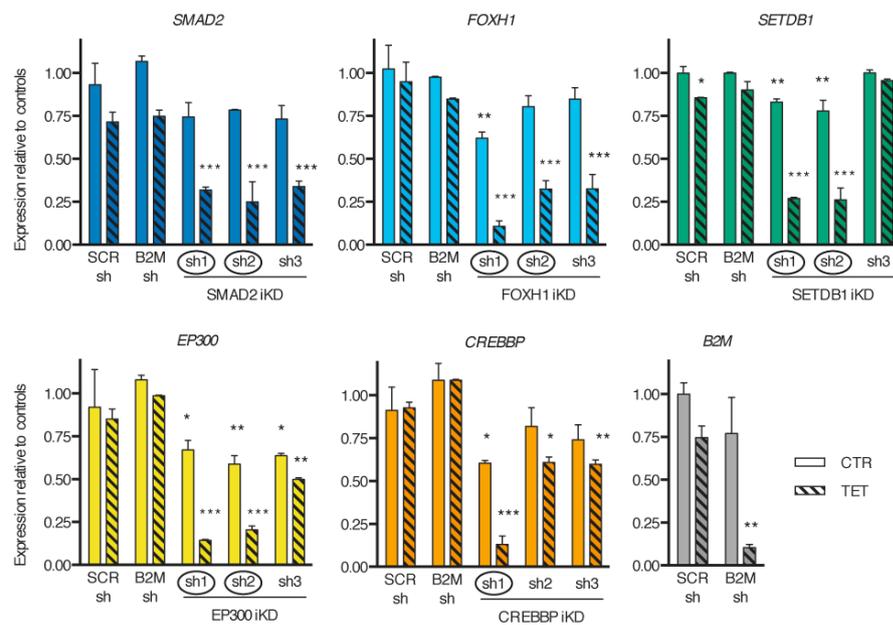


Figure 5.8. Validation of SMAD2/3 cofactors inducible gene knockdown.

qPCR validation of inducible knockdown hESCs cultured in absence (CTR) or presence of tetracycline for 3 days (TET). Three distinct shRNAs were tested for each gene, and two independent clonal pools were analysed for each condition. Expression is shown as normalized on the average level in hESCs carrying negative control shRNAs and cultured in absence of tetracycline. Note that for the B2M shRNA only the SCR shRNA was used as negative control. Significant differences vs SCR CTR are reported (2-way ANOVA with post-hoc Sidak comparisons). shRNAs selected for further experiments are circled.

Initially, we assessed the effect of inducible knockdown in undifferentiated hESCs (Figure 5.9, Figure 5.10, and Figure 5.11). As expected, knockdown of SMAD2 resulted in clear changes of hESC morphology and concomitant loss of NANOG expression, indicative of hESC differentiation. Accordingly, qPCR analyses demonstrated an induction of both mesoderm and neuroectoderm markers, while background expression of endoderm genes was reduced, all in agreement with what recently reported (Sakaki-Yumoto et al. 2013). Moreover, expression of the well-known Activin/Nodal target genes *NODAL* and *LEFTY1* was strongly downregulated. Collectively, inducible knockdown of SMAD2 confirmed the crucial role of Activin/Nodal signalling in hPSCs.

Interestingly, FOXH1 knockdown did not result in overt hESC differentiation, nor it affected expression of pluripotency markers. Nevertheless, the morphology of FOXH1 knockdown hESC colonies was less flat and “dome shaped”, suggesting that FOXH1 might have an important function in hESCs distinct from maintaining the expression of pluripotency factors. Indeed, qPCR analysis showed that FOXH1 knockdown resulted in upregulation of neuroectoderm genes and downregulation of endoderm markers, thus partially mimicking the phenotype of SMAD2 knockdown hESCs. Of note, however, mesoderm makers were unaffected.

With regards to the epigenetic cofactors, knockdown of SETDB1 had only a minor effect on the expression of pluripotency markers, while it resulted in upregulation of neuroectoderm genes. On the other side, knockdown of EP300 or CREBBP strongly impaired expression of pluripotency markers and induced a mild upregulation of endoderm genes. Interestingly, the effect on the protein level of NANOG following CREBBP knockdown was not explained by transcriptional changes, thus suggesting a potential post-transcriptional role.

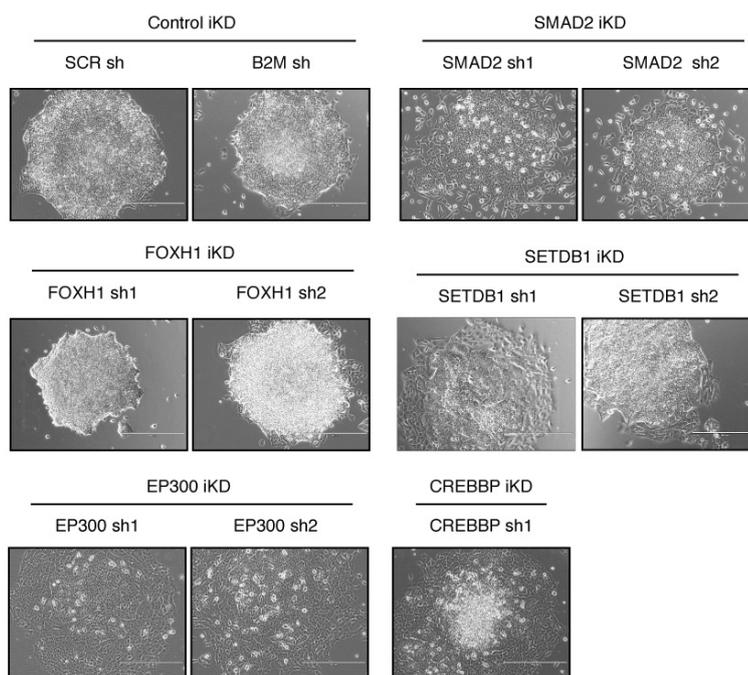


Figure 5.9. Effect of SMAD2/3 cofactors knockdown on hPSC morphology. Representative phase contrast images of inducible knockdown (iKD) hESCs expressing the indicated shRNAs (sh) and cultured in presence of tetracycline for 6 days to induce knockdown. Scale bars: 400µm.

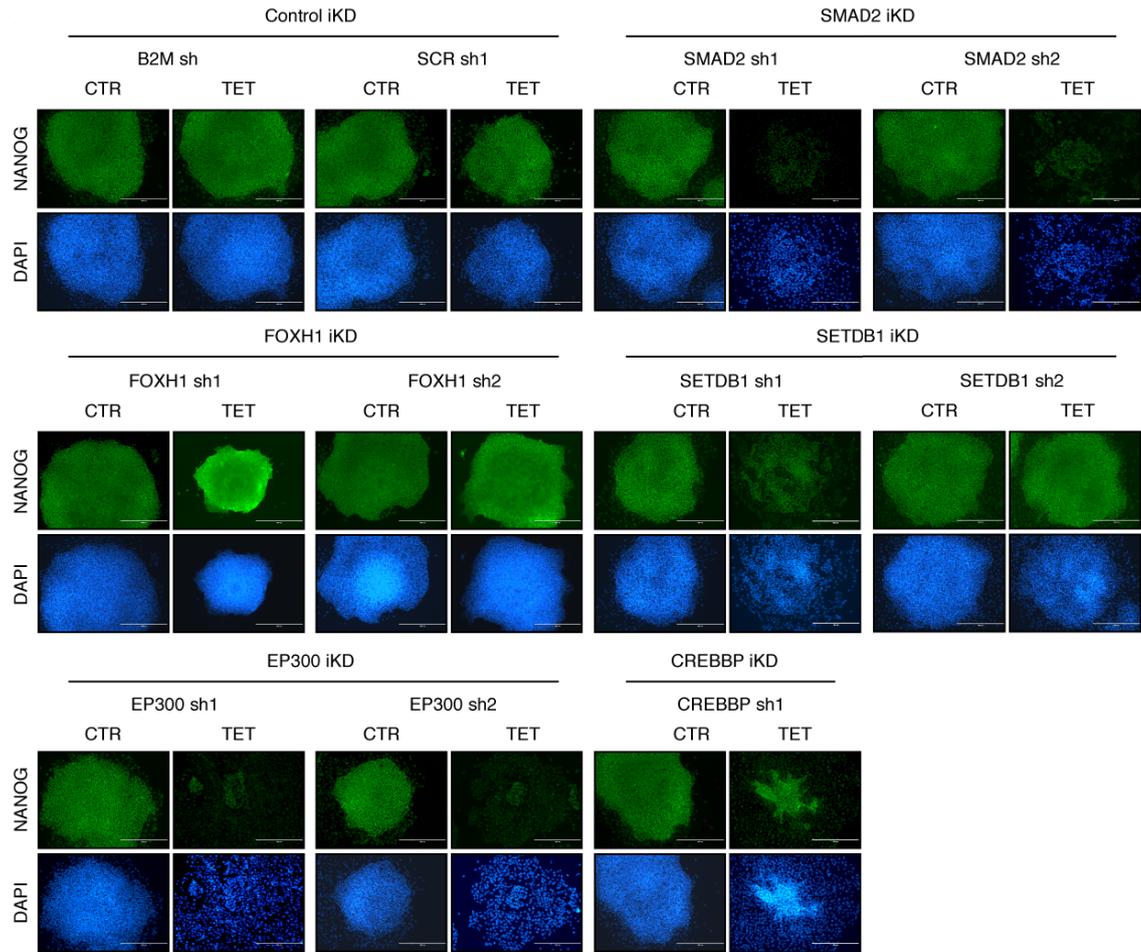


Figure 5.10. Effect of SMAD2/3 cofactors knockdown on hPSC pluripotency.

Representative immunofluorescent stainings for the pluripotency factor NANOG in inducible knockdown (iKD) hESCs expressing the indicated shRNAs (sh) and cultured in absence (CTR) or presence of tetracycline (TET) for 6 days to induce knockdown. DAPI shows nuclear staining. Scale bars: 400 μm.

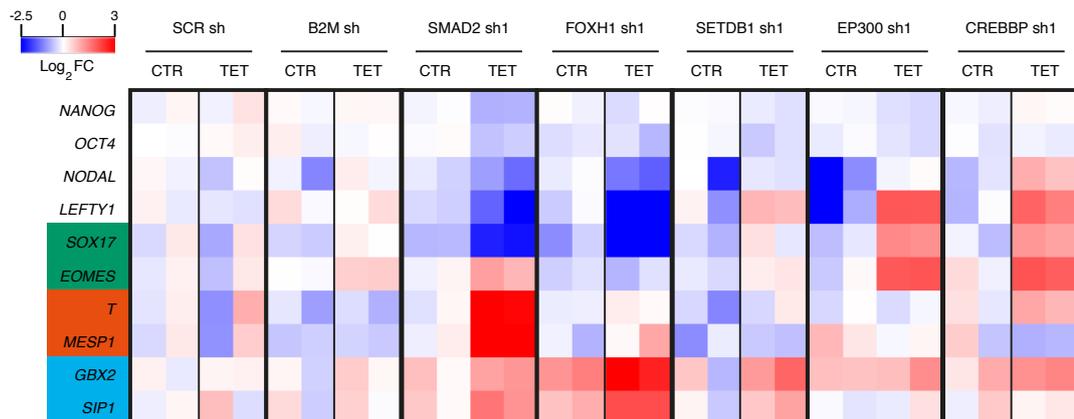


Figure 5.11. Transcriptional effects following knockdown of SMAD2/3 cofactors.

Heatmap summarizing qPCR analyses of inducible knockdown hESCs expressing the indicated shRNAs (sh) and cultured as described in Figure 5.10. Results for two independent clonal pools for each condition are shown. For each gene, the colour code represents log₂ fold-changes (FC) compared to the average expression in hESCs carrying an inducible SCR control shRNA and

cultured in the absence of tetracycline. Markers for each germ layer are highlighted in a coloured box (green: endoderm; red: mesoderm; blue: neuroectoderm).

We then examined the function of the SMAD2/3 cofactors during endoderm differentiation of hESCs (Figure 5.12, Figure 5.13, and Figure 5.14). Remarkably, SMAD2 knockdown almost completely abolished endoderm specification, in agreement with the known crucial role of Activin/Nodal signalling in the induction of this germ layer both *in vitro* and *in vivo* (Chapter 1.3.2). Of note, this effect was nearly phenocopied in hESCs with the strongest level of FOXH1 knockdown, a result that is reminiscent of the *in vivo* function of FOXH1 (Chapters 1.4.2.2 and 1.7.3). On the other hand, knockdown of epigenetic modifiers had only moderate (SETDB1), little (EP300), or no effect (CREBBP) on endoderm specification, suggesting that these factors have only a limited individual involvement in this process.

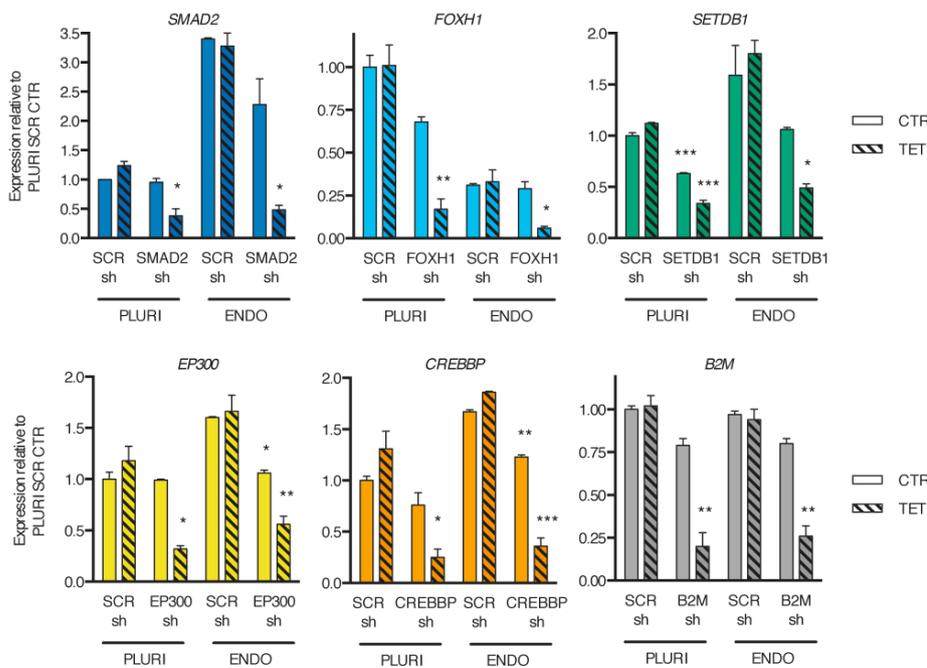


Figure 5.12. Validation of SMAD2/3 cofactors gene knockdown during endoderm differentiation.

qPCR validation of inducible knockdown hESCs in pluripotency (PLURI) and following endoderm differentiation. Pluripotent cells were cultured in absence (CTR) or presence of tetracycline (TET) for 6 days. For endoderm differentiation, tetracycline treatment was initiated in undifferentiated hESCs for 3 days in order to ensure gene knockdown at the start of endoderm specification, and was then maintained during differentiation. For each gene, the shRNA resulting in the strongest level of knockdown was selected, and two independent clonal pools were analysed for each condition. Expression is shown as normalized to the average level in pluripotent hESCs carrying a SCR control shRNAs and cultured in absence of tetracycline. Significant differences vs SCR CTR in the same cell type are reported (2-way ANOVA with post-hoc Sidak comparisons).

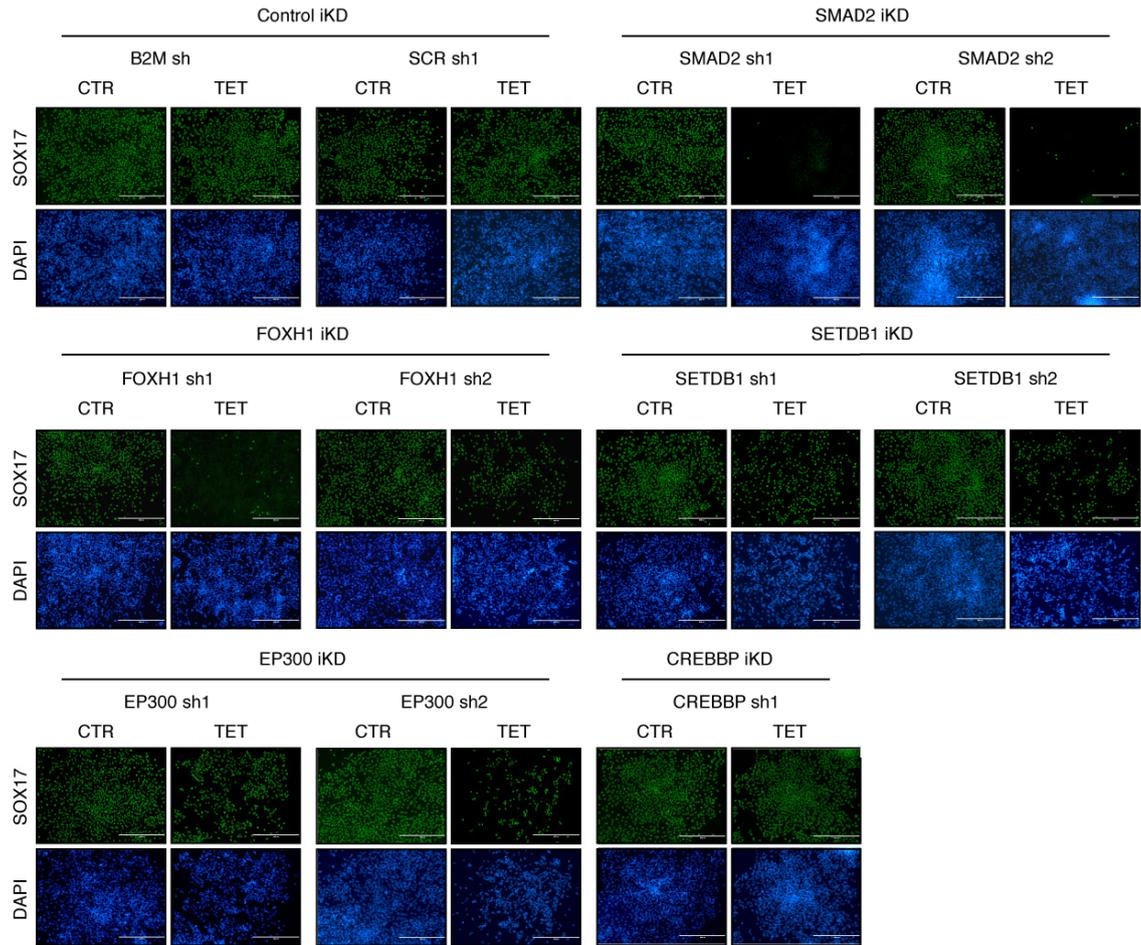


Figure 5.13. Effect of SMAD2/3 cofactors knockdown during endoderm differentiation of hPSCs.

Representative immunofluorescent stainings for the endoderm marker SOX17 following endoderm differentiation of inducible knockdown (iKD) hESCs expressing the indicated shRNAs (sh) and cultured as described in Figure 5.12. DAPI shows nuclear staining. Scale bars: 400µm.

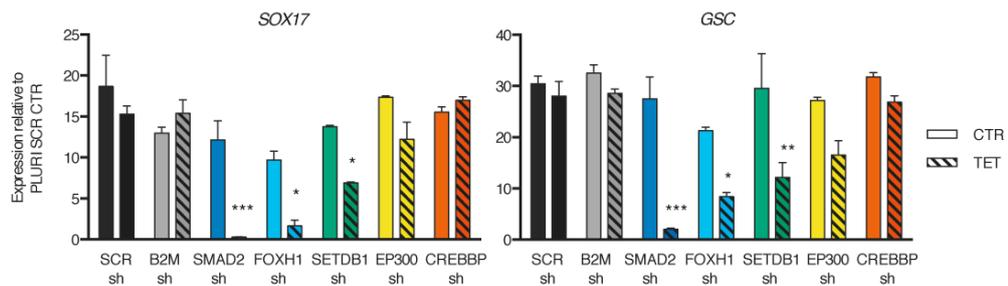


Figure 5.14. Transcriptional effects following knockdown of SMAD2/3 cofactors during endoderm differentiation of hPSCs.

qPCR analysis following endoderm differentiation of inducible knockdown hESCs expressing the indicated shRNAs (sh) and cultured in absence (CTR) or presence of tetracycline (TET) to induce knockdown (Figure 5.12). Two independent clonal pools were analysed for each condition. Significant differences vs same shRNA in control conditions are reported (2-way ANOVA with post-hoc Sidak comparisons).

Collectively, the results of this functional screening demonstrated that several cofactors that interact with SMAD2/3 have important and distinct roles in the control of hPSC pluripotency and differentiation (Figure 5.15). This suggests that SMAD2/3 is a central hub for transcriptional and epigenetic regulations of hPSC fate decisions.

	Pluripotency regulator	Effect of iKD (hPSCs)	Endoderm regulator
SMAD2	+++	↓E ↑N ↑M	+++
FOXH1	± (morphology?)	↓E ↑N	++
SETDB1	+	↑N	++
EP300	++	↑E	+
CREBBP	++ (post-transcr?)	↑E	-

Figure 5.15. Function of SMAD2/3 cofactors in hPSC pluripotency and endoderm differentiation.

Table summarizing the results presented in this sub-chapter. E: endoderm; N: neuroectoderm; M: mesoderm.

5.2.4 The m6A methyltransferase complex inhibits the response to Activin/Nodal signalling

Having investigated the function of several SMAD2/3 epigenetic and transcriptional cofactors, we turned our attention to other “non-canonical” SMAD2/3 interactors identified in our proteomic screening. In particular, we focused on proteins involved in RNA regulations, since a functional role for SMAD2/3 in such mechanisms has never been previously demonstrated.

The interaction between SMAD2/3 and the METTL3-METTL14-WTAP m6A methyltransferase complex appeared particularly interesting (Figure 5.16). Indeed, m6A is the most common epitranscriptome modification, and can regulate multiple aspects of RNA biology such as RNA decay, translation, and splicing (Chapter 1.6.2.1). Moreover, m6A deposition was recently suggested to play a crucial role in murine pluripotency (Chapter 1.6.2.2). However, the role of m6A in hPSCs has been poorly explored to date, and whether m6A deposition is dynamically modulated by extracellular signalling is currently unknown. Overall, we decided to explore the function of the m6A methyltransferase complex in hPSCs and its possible relationship with Activin/Nodal signalling.

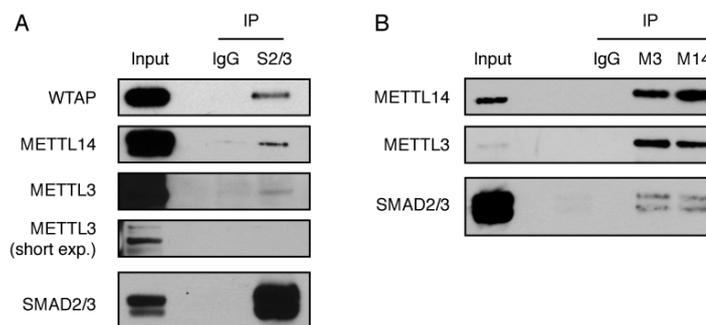


Figure 5.16. Interaction of SMAD2/3 with the m6A methyltransferase complex in hPSCs.

(A-B) Western blots of SMAD2/3 (S2/3), METTL3 (M3), METTL14 (M14), or control (IgG) immunoprecipitations (IP) from nuclear extracts of hESCs. Input is 5% of the material used for IP. Immunoprecipitations were performed in collaboration with Ms Stephanie Brown.

For this, we generated hESCs carrying inducible shRNAs against the various subunits of the m6A methyltransferase complex (note that for these experiments we took advantage of our recently established sOPTiKD platform; Chapter 4.3.4), and validated that these allowed strong inducible gene knockdown (Figure 5.17).

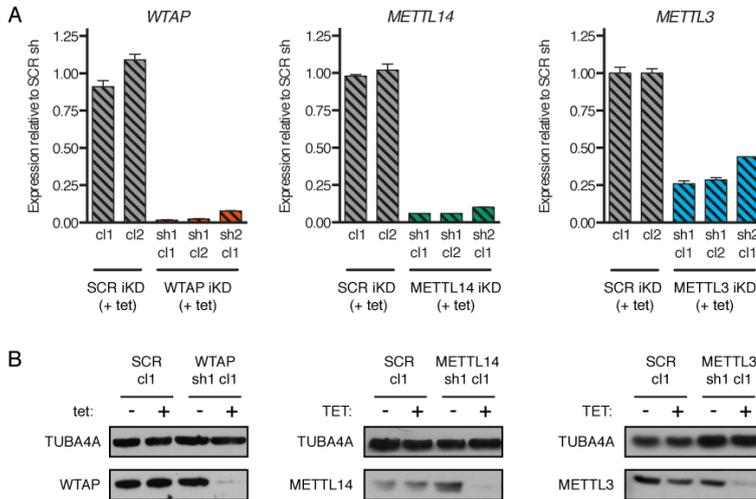


Figure 5.17. Inducible knockdown of the m6A methyltransferase complex subunits in hPSCs.

(A) qPCR validation of inducible knockdown (iKD) hESCs cultured in presence of tetracycline (tet) for 5 days to drive gene knockdown. Two distinct shRNAs (sh) were tested for each gene, and individual clonal sublines (cl) were analysed in duplicate. Expression is shown as normalized on the average level in hESCs carrying a negative control SCR shRNA. (B) Representative Western blots for selected hESCs clones described in panel A. Cells were maintained in the absence or presence of tetracycline for 5 days. TUBA4A (α -tubulin): loading control.

Interestingly, prolonged knockdown of WTAP, METTL14, or METTL3 was compatible with hESC self-renewal and pluripotency, as cells could be cultured in the presence of tetracycline for several passages without any obvious change in the proliferation rates, morphology of hESC colonies, and expression of pluripotency markers (Figure 5.18). On the other hand, the deposition of m6A was reduced following knockdown of these factors, demonstrating that our inducible knockdown strategy successfully impaired the function of the m6A methyltransferase complex (Figure 5.19). In this regards, knockdown of WTAP had the strongest effect, in agreement with previous reports (Schwartz et al. 2014; Ping et al. 2014; J. Liu et al. 2014). On the other hand, knockdown of METTL14 and METTL3 resulted in either moderate or minor reduction of m6A deposition, respectively (possibly due to a less effective knockdown for METTL3; Figure 5.17). Overall, these results showed that robust m6A methylation is not required to maintain the pluripotent state.

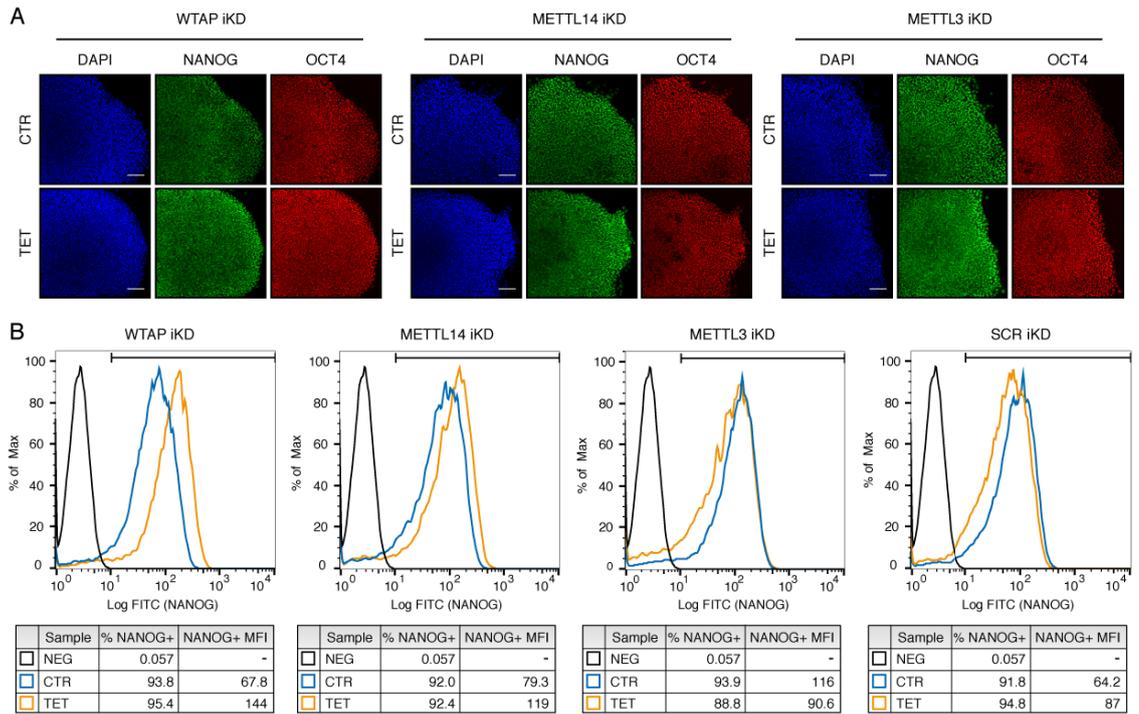


Figure 5.18. Effect of m6A methyltransferase complex knockdown on hPSC pluripotency.

(A) Representative immunofluorescent stainings for the pluripotency markers NANOG and OCT4 in inducible knockdown (iKD) hESCs cultured for three passages (15 days) in absence (CTR) or presence of tetracycline (TET). DAPI shows nuclear staining. Scale bars: 100µm. (B) Representative flow cytometry quantifications for NANOG in cells treated as described for panel A. The percentage and median fluorescence intensity (MFI) of NANOG positive cells (NANOG+) are reported. The gates used for the analysis are shown, and were determined based on a secondary antibody-only negative staining (NEG).

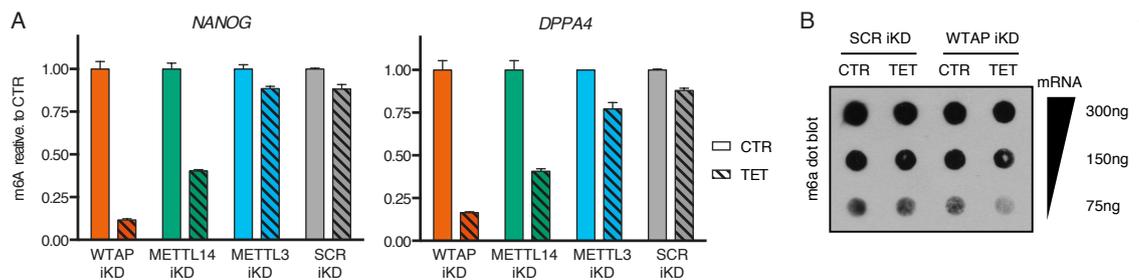


Figure 5.19. Effect of m6A methyltransferase complex knockdown on the deposition of m6A.

(A) Representative m6A MeRIP-qPCR in inducible knockdown (iKD) hESCs cultured for 10 days in absence (CTR) or presence of tetracycline (TET). m6A abundance is reported relative to control conditions in the same hESC line. The error bars depict technical variability in the qPCR analysis. (B) Representative m6A dot blot in WTAP or SCR iKD hESCs treated as described in panel A. Decreasing amounts of mRNA were spotted to facilitate semi-quantitative comparisons, as indicated. The experiments in this figure were performed in collaboration with Ms Stephanie Brown.

We then investigated the consequences of impaired m6A deposition on the differentiation potential of hESCs. For this, we first induced knockdown of WTAP, METTL14, or METTL3 in hESCs, then differentiated such cells into either definitive endoderm or neuroectoderm (Figure 5.20 and Figure 5.21). Endoderm specification was not consistently affected by knockdown of such factors (Figure 5.20). On the other hand, knockdown of any of these genes remarkably impaired neuroectoderm induction, as the expression of several lineage-specific markers was significantly reduced both at the transcript and protein level (Figure 5.21).

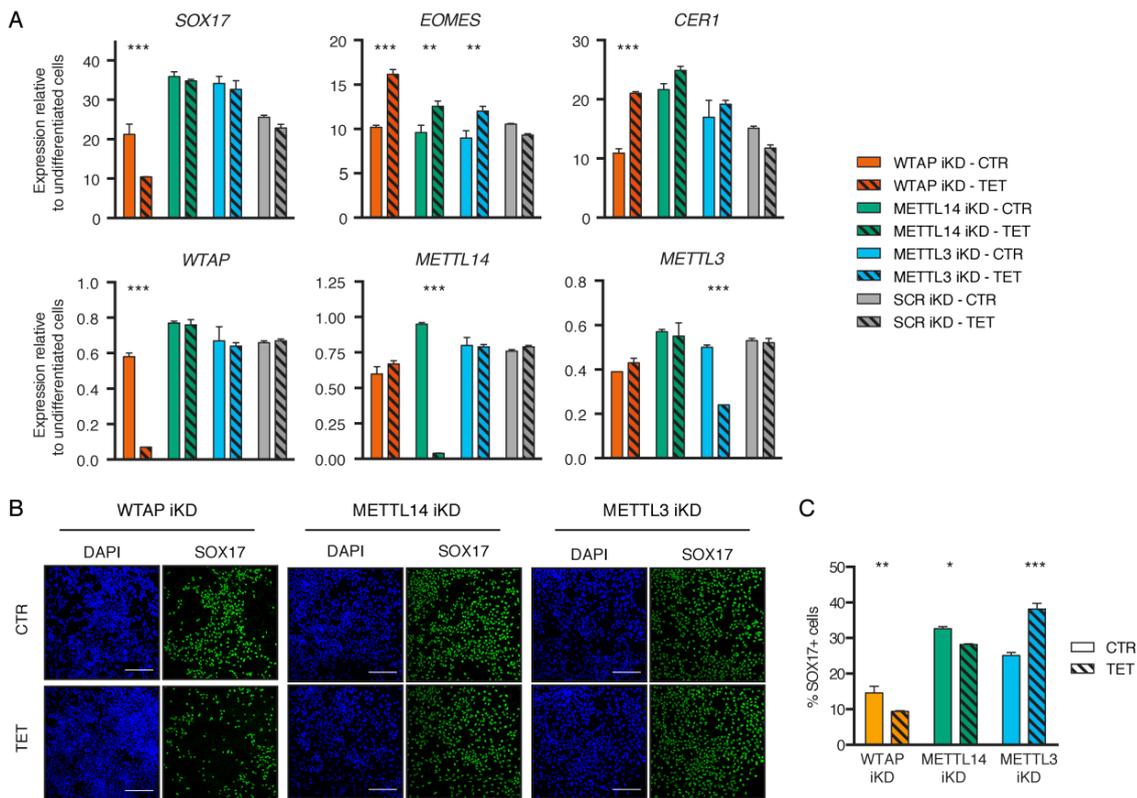


Figure 5.20. Effect of m6A methyltransferase complex knockdown on endoderm differentiation of hPSCs.

(A) qPCR analysis following endoderm differentiation of inducible knockdown (iKD) hESCs cultured in absence (CTR) or presence of tetracycline (TET). Tetracycline treatment was initiated in undifferentiated hESCs for 10 days and was maintained during differentiation (3 days). Expression is shown as normalized on the average level in undifferentiated hESCs. (B) Representative immunofluorescent stainings for the endoderm marker SOX17 in cells treated as described in panel A. DAPI shows nuclear staining. Scale bars: 100 μ m. (C) Flow cytometry quantification of the percentage of SOX17 positive cells (SOX17+) in the experiment described in panel A. For panels A and C, results are from 3 biological replicates per condition, and significant differences vs same iKD line in control conditions are reported (2-way ANOVA with post-hoc Sidak comparisons).

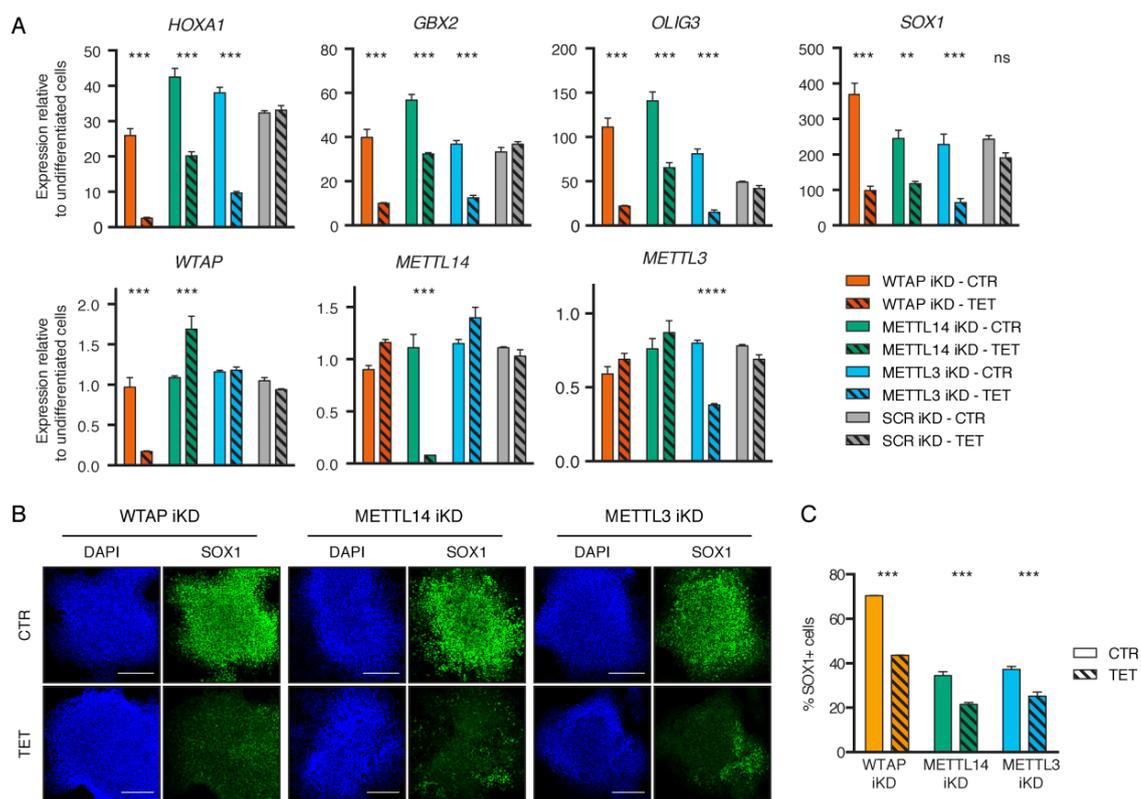


Figure 5.21. Effect of m6A methyltransferase complex knockdown on neuroectoderm differentiation of hPSCs.

As in Figure 5.20, but cells were differentiated into neuroectoderm and analysed for the expression of the neuroectoderm marker SOX1.

Neuroectoderm specification of hESCs relies primarily on the inhibition of Activin/Nodal signalling, which leads to loss of the pluripotency factor NANOG thus driving the exit from the pluripotent state (Chapter 1.3.2.3 and Chapter 3). Interestingly, knockdown of WTAP, METTL14, or METTL3 impaired this mechanism, as the downregulation of NANOG was delayed in such conditions (Figure 5.22). The expression of NANOG was particularly affected by knockdown of WTAP, which resulted in a modest yet significant upregulation already in the pluripotent state, and prevented full downregulation even following 32h of Activin/Nodal signalling inhibition. NANOG is known to repress the expression of neuroectoderm genes (Vallier, Mendjan, et al. 2009), which could explain the defective neuroectoderm specification of WTAP knockdown hESCs. Of interest, these results are in agreement with what previously reported for mESCs, in which impairment of the m6A methyltransferase complex and the subsequent reduction of m6A deposition result in an increased stability of the *Nanog* transcript, thus preventing timely exit of mESCs from the pluripotent state (Batista et al. 2014; Geula et al. 2015).

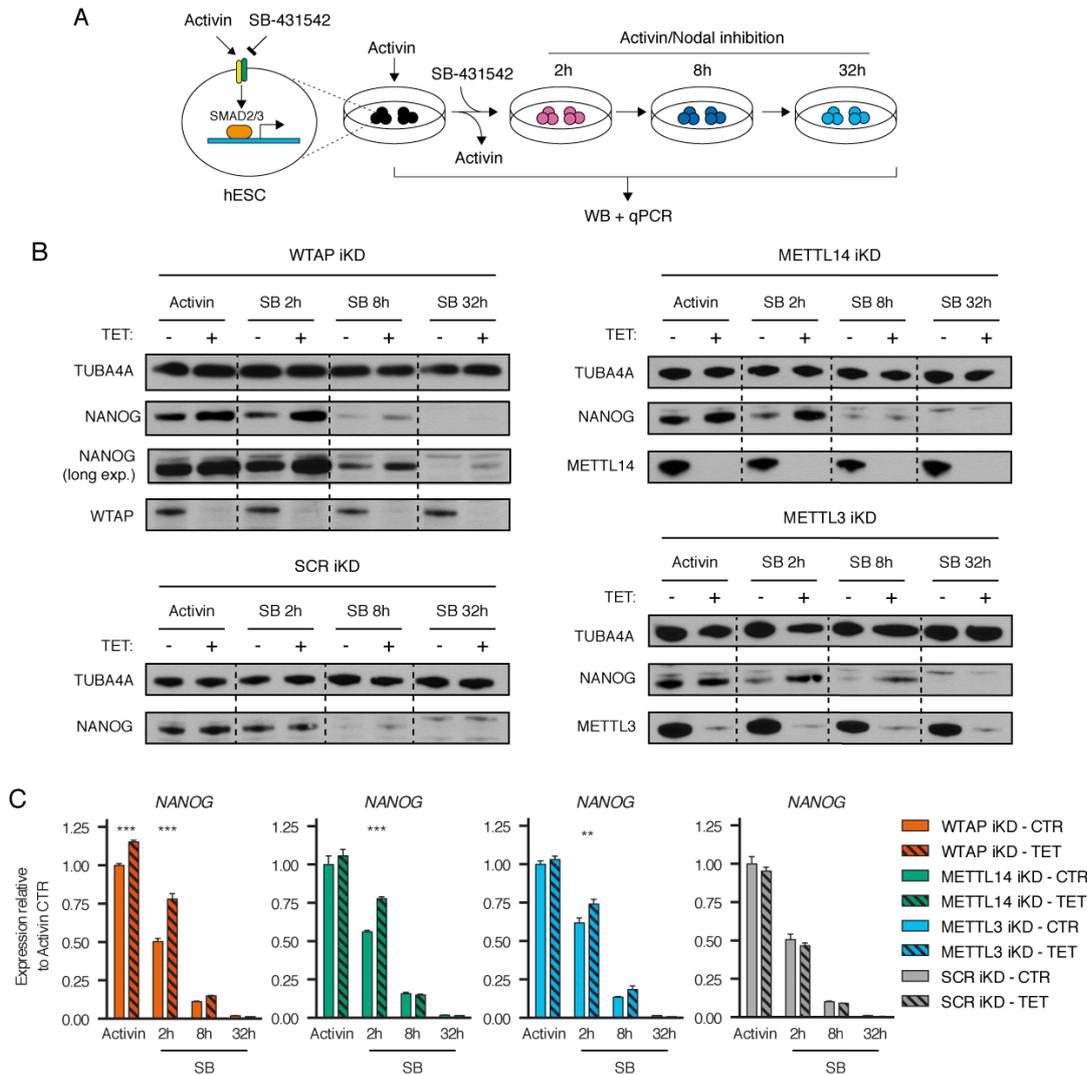


Figure 5.22. Effect of m6A methyltransferase complex knockdown on the downregulation of NANOG following Activin/Nodal inhibition.

(A) Schematic of the experimental approach. (B) Representative Western blots in inducible knockdown (iKD) hESCs cultured in absence or presence of tetracycline (TET) for 10 days, then subjected to Activin/Nodal signalling inhibition with SB431542 (SB) for the indicated time. Activin: cells maintained in standard pluripotency-promoting culture conditions containing Activin and collected at the beginning of the experiment. TUBA4A (α -tubulin): loading control. (C) qPCR analyses of cells treated as described in panel A. Results are from 3 biological replicates per condition, and significant differences vs same iKD line in control conditions at the same time point are reported (2-way ANOVA with post-hoc Sidak comparisons).

In hESCs, the expression of NANOG is directly activated by SMAD2/3 downstream of the Activin/Nodal signalling pathway (Chapter 3; Vallier, Mendjan, et al. 2009; Xu et al. 2008). We therefore investigated whether other SMAD2/3 target genes were deregulated following knockdown of the m6A methyltransferase complex. For this, we focused on WTAP inducible knockdown hESCs, since these had demonstrated to represent the strongest genetic tool to probe the function of m6A (Figure 5.19). Remarkably, qPCR demonstrated that WTAP knockdown resulted in upregulation of several well-known SMAD2/3 target genes activated by Activin/Nodal signalling, including *LEFTY1*, *NODAL*, and *CER1* (Figure 5.23). Moreover, WTAP knockdown

impaired the upregulation of several genes induced following inhibition of Activin/Nodal signalling, such as *CDX2*, *HOXA1*, and *GBX2* (Figure 5.23).

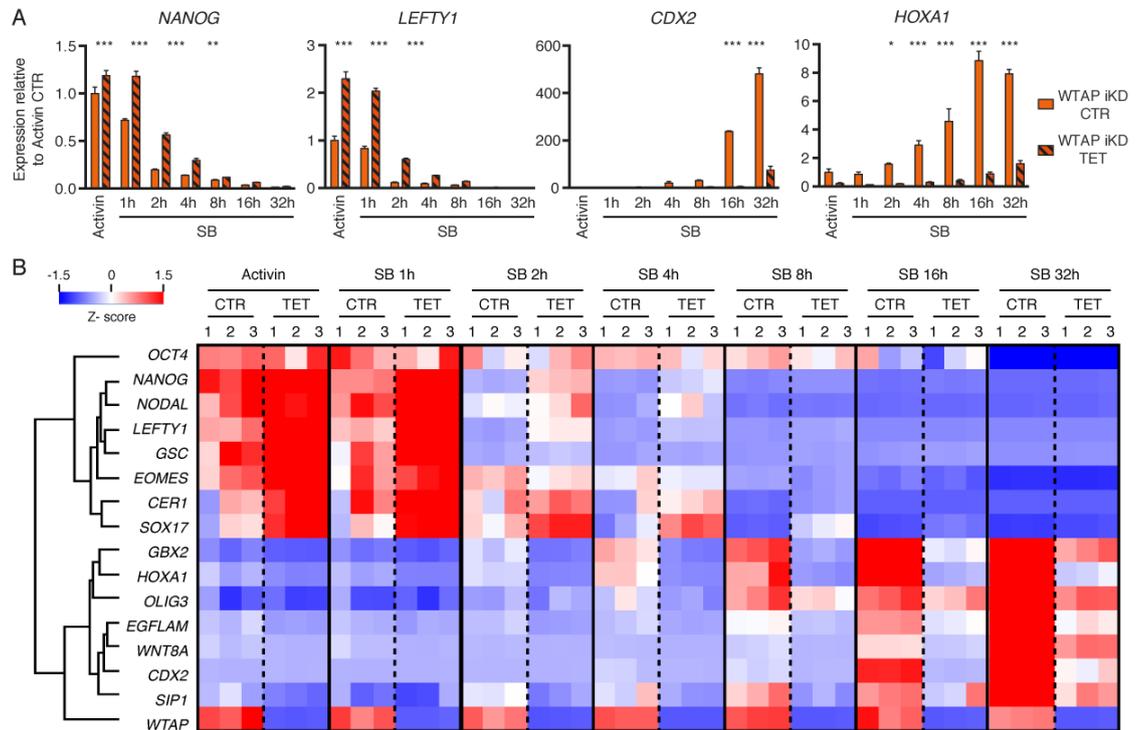


Figure 5.23. Effect of WTAP knockdown on the expression of Activin/Nodal target genes.

As in Figure 5.22 panel C, but WTAP inducible knockdown (iKD) cells were analysed at multiple time points following Activin/Nodal inhibition with SB. In B, the heatmap summarizes qPCR results, and Z-scores indicate the differential expression measured in number of standard deviations from the average level across all the time points. Genes were clustered based on their Euclidean distance.

Given these interesting findings, we extended our analyses to a genome-wide level by performing RNA sequencing (RNA-seq) experiments (Figure 5.24, Figure 5.25, and Figure 5.26). These results confirmed that knockdown of WTAP globally alters the response to Activin/Nodal signalling by: (1) upregulating a large cohort of developmental regulators whose expression is maintained by Activin/Nodal in the pluripotent state, and by delaying the downregulation of such genes upon Activin/Nodal inhibition (cluster 2); (2) impairing the upregulation of neuroectoderm genes induced following inhibition of Activin/Nodal (cluster 3). Additionally, WTAP knockdown resulted in up- and downregulation of additional factors whose expression is largely independent from Activin/Nodal signalling, and which are not associated to developmental regulations (cluster 1 and 4, respectively). This showed that WTAP has also functions other than modulating the response to Activin/Nodal signalling, in agreement with its role as a general regulator of the epitranscriptome.

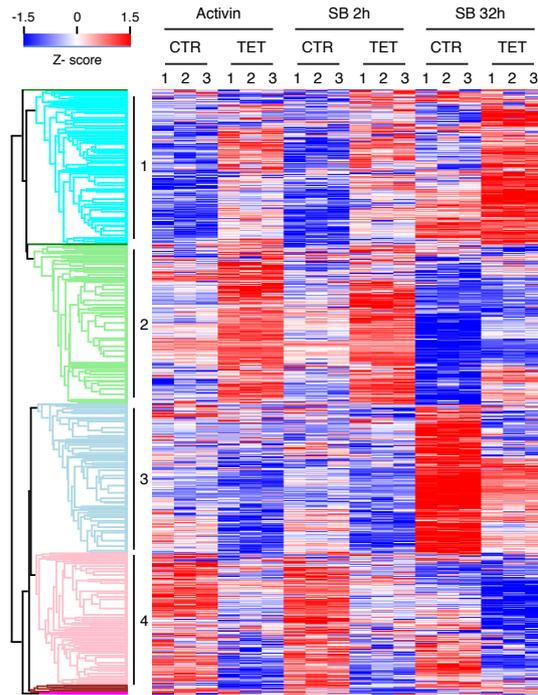


Figure 5.24. Genome-wide gene expression changes following Activin/Nodal inhibition in WTAP knockdown hPSCs.

As in Figure 5.23, but gene expression in WTAP inducible knockdown hESCs was measured by RNA-seq. RNA-seq and primary data analysis was performed in collaboration with Ms Stephanie Brown, the Cambridge Genomic Services facility, and Miss Anna Osnato. The heatmap depicts Z-scores for the top 5% differentially expressed genes (1789 genes, as ranked by the Hotelling T² statistic). Genes and samples were clustered based on their Euclidean distance, and the four major gene clusters are indicated.

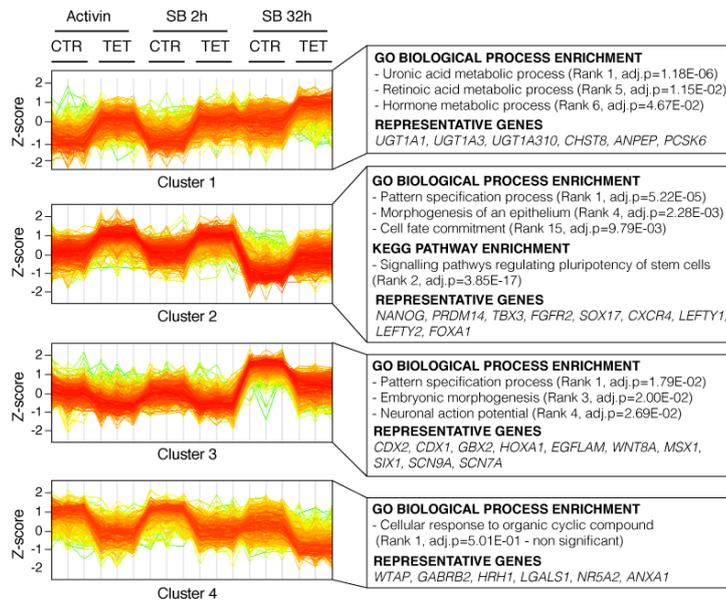


Figure 5.25. Transcriptional profiles of genes differentially expressed following Activin/Nodal inhibition in WTAP knockdown hPSCs.

Expression profiles of genes belonging to the clusters indicated in Figure 5.24. Selected results of gene enrichment analysis and representative genes for each cluster are reported.

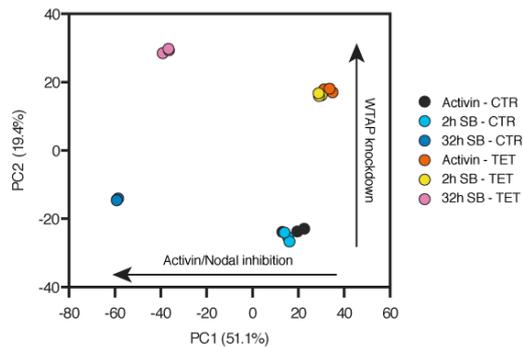


Figure 5.26. Global effects of WTAP knockdown on the response to Activin/Nodal signalling in hPSCs.

Principal component analysis (PCA) of RNA-seq results described in Figure 5.24. The top 5% differentially expressed genes were considered for this analysis. For each of the two main principal components (PC1 and PC2), the fraction of inter-sample variance that they explain and their proposed biological meaning are reported.

Overall, the knockdown of WTAP in hESCs induced a transcriptional signature indicative of an increased response to Activin/Nodal. Interestingly, mRNAs upregulated following WTAP knockdown were significantly associated to m6A deposition (Figure 5.27), in agreement with the notion that presence of m6A generally reduces mRNA stability (Chapter 1.6.2.1). Moreover, transcripts rapidly decreased after Activin/Nodal inhibition were enriched in m6A-marked mRNAs (Figure 5.27), suggesting that deposition of m6A could be at least partially responsible for their rapid downregulation.

Collectively, these results showed that the m6A methyltransferase complex antagonizes the transcriptional responses initiated downstream of the Activin/Nodal signalling pathway. Moreover, they suggested that this effect could be mediated by m6A-dependent destabilization of Activin/Nodal transcriptional targets.

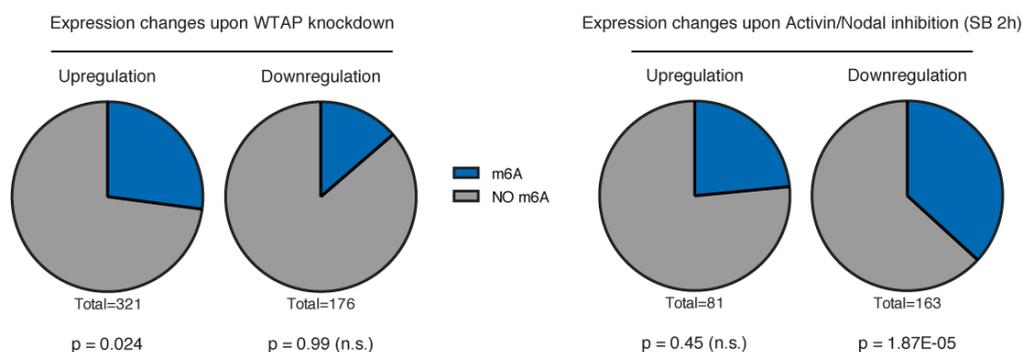


Figure 5.27. Deposition of m6A on transcripts regulated by WTAP or by Activin/Nodal signalling.

Proportion of transcripts marked by at least one high-confidence m6A (data from Batista et al., 2014) in transcripts significantly up- or downregulated ($p < 0.05$ and $\text{abs.FC} > 2$) following WTAP inducible knockdown in hESCs or Activin/Nodal inhibition for 2h with SB (Figure 5.24). The hypergeometric probability of the observed overlap is reported (n.s.: non-significant).

5.2.5 Activin/Nodal signalling promotes m6A deposition on specific regulators of pluripotency and differentiation

Since SMAD2/3 physically interacts with the m6A methyltransferase complex, we hypothesized that the functional interaction between Activin/Nodal signalling and m6A deposition might be the result of a direct crosstalk between these two pathways. To test this notion, we monitored the level of m6A onto Activin/Nodal-regulated transcripts following short-term signalling inhibition (Figure 5.28 and Figure 5.29). Interestingly, treatment of hESCs for 2h with the Activin/Nodal inhibitor SB decreased m6A levels of specific nuclear transcripts (such as *NANOG* and *LEFTY1*, but not *DPPA4*), while cytoplasmic mRNAs were unaffected (Figure 5.29). This suggested that SMAD2/3 might promote m6A deposition onto certain transcripts at the nuclear level.

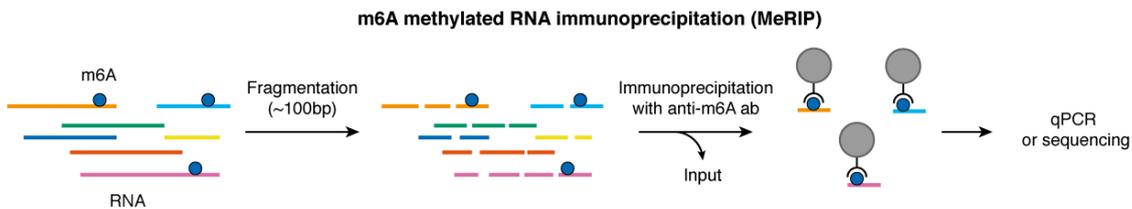


Figure 5.28. Overview of m6A MeRIP experiments.

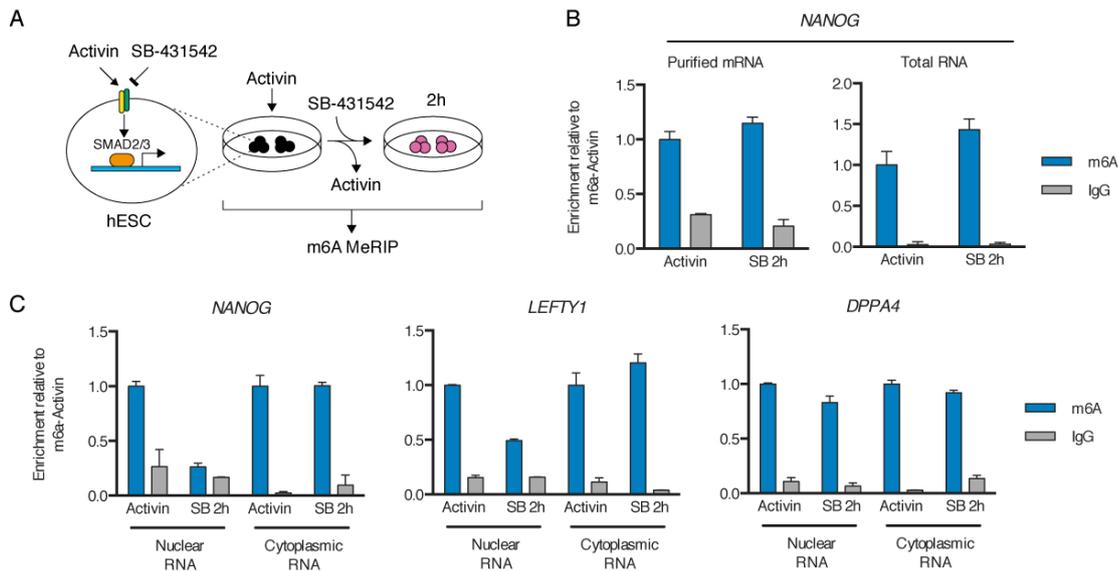


Figure 5.29. m6A MeRIP-qPCR in hPSCs following Activin/Nodal inhibition.

(A) Schematic of the experimental approach. (B) Representative m6A MeRIP-qPCR results from purified mRNA or total cellular RNA. IgG MeRIP experiments were performed as negative controls, and samples were analysed in technical duplicate. (C) As in B, but cellular RNA species were extracted following nuclear/cytoplasmic subcellular fractionation. The experiments in this figure were performed in collaboration with Ms Stephanie Brown.

Encouraged by these results, we assessed the transcriptome-wide effects of Activin/Nodal inhibition on the deposition of m6A onto nuclear transcripts. Deep sequencing of m6A methylated RNA immunoprecipitations (MeRIP) confirmed that several Activin/Nodal-regulated transcripts are heavily methylated in an Activin/Nodal-dependent fashion (Figure 5.30). In agreement with what previously reported (Chapter 1.6.2.1), deposition of m6A was specific only to certain transcripts, was enriched around stop codons, and occurred on a motif closely reminiscent of the m6A consensus sequence (Figure 5.31).

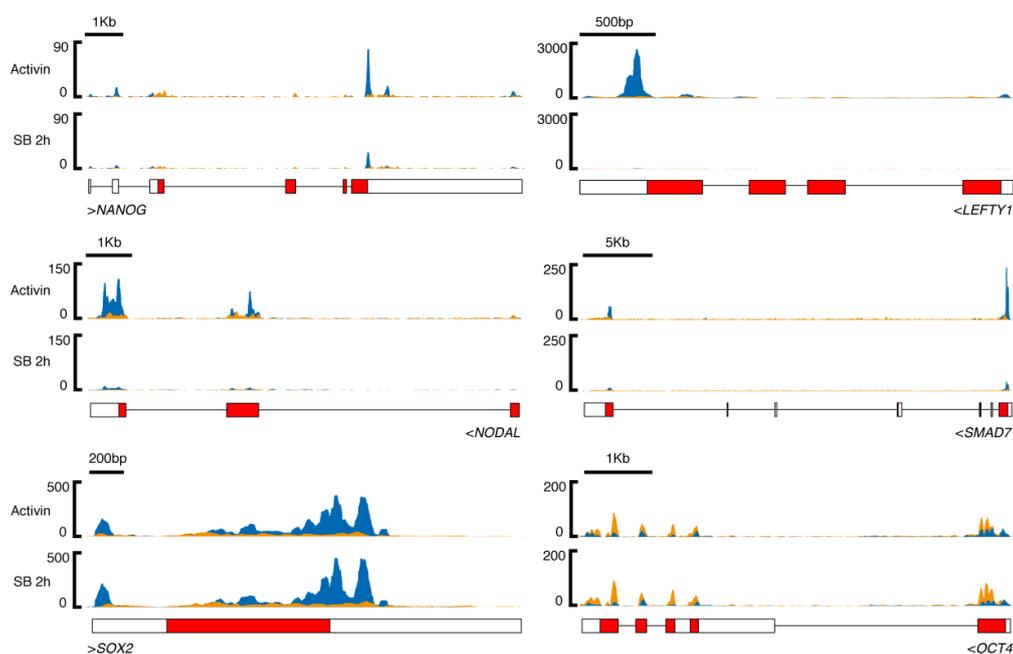


Figure 5.30. m6A MeRIP-seq in hPSCs following Activin/Nodal inhibition.

m6A MeRIP-seq results for selected transcripts before and after inhibition of Activin/Nodal with SB for 2h. Three biological replicates per condition were analysed, and these were combined for visualization purposes. Histograms represent read-enrichments normalized by million mapped reads and size of the library. Blue: sequencing results of m6A MeRIP. Orange: sequencing results of pre-MeRIP input RNA (negative control). GENECODE gene annotations are shown (red: protein coding exons; white: untranslated exons; note that all potential exons are shown and overlaid). Compared to the other genes shown, the m6A levels on *SOX2* were unaffected by Activin/Nodal inhibition, suggesting a specificity of action. *OCT4/POU5F1* is reported as negative control since it is known not to have any m6A site (Batista et al., 2014), as confirmed by the lack of m6A enrichment compared to the input. These experiments were performed in collaboration with Ms Stephanie Brown, the Cambridge Genomic Services facility, and Dr Pedro Madrigal.

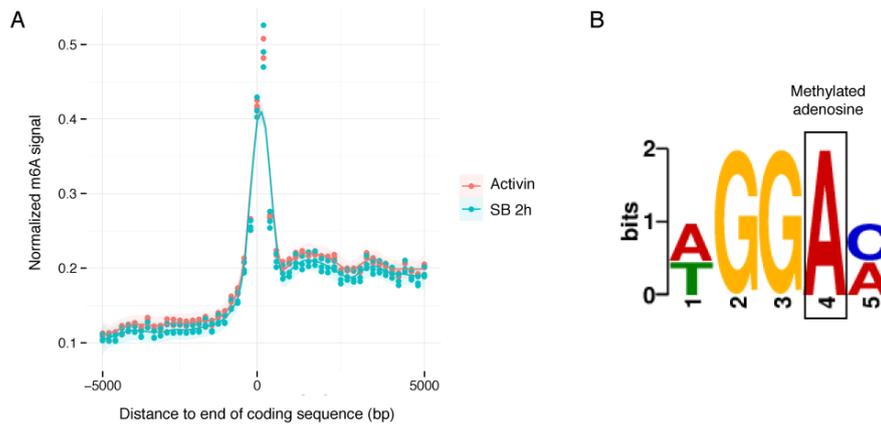


Figure 5.31. Features of m6A deposition in hPSCs.

(A) Normalized m6A coverage relative to stop codons in MeRIP-seq data from the experiment described in Figure 5.30. (B) Top sequence motif identified *de novo* using all m6A peaks across the two conditions (E-value: 8.4E-138). The position of the methylated adenosine is shown. These analyses were performed in collaboration with Dr Pedro Madrigal.

Global assessment of differential m6A deposition revealed that Activin/Nodal inhibition predominantly resulted in reduced m6A levels onto selected transcripts (Figure 5.32, Appendix IV). Interestingly, downregulation of m6A largely correlated with rapid decrease of gene expression following Activin/Nodal inhibition (Figure 5.33; note that the bioinformatics method used to measure differential methylation takes into account such changes, as it considers the m6A abundance relative to the transcript level). Transcripts behaving in this fashion were enriched in pluripotency regulators and in factors involved in the Activin/Nodal signalling pathway (Figure 5.34). Moreover, they included several well-known SMAD2/3 transcriptional targets (such as *NANOG*, *NODAL*, and *SMAD7*). This suggests that Activin/Nodal-dependent transcriptional and epitranscriptional regulations might be tightly interconnected. Moreover, the rapid decrease in expression of these mRNAs is in agreement with the notion that m6A deposition might decrease their stability.

On the other hand, the expression of a large number of transcripts associated to Activin/Nodal-sensitive m6A deposition was unchanged following 2h of Activin/Nodal inhibition (Figure 5.33 and Figure 5.35). This class of genes was more heterogeneous, but was nonetheless enriched in developmental regulators (Figure 5.35). These findings established that Activin/Nodal signalling regulates m6A deposition of numerous transcripts also independently of transcriptional changes.

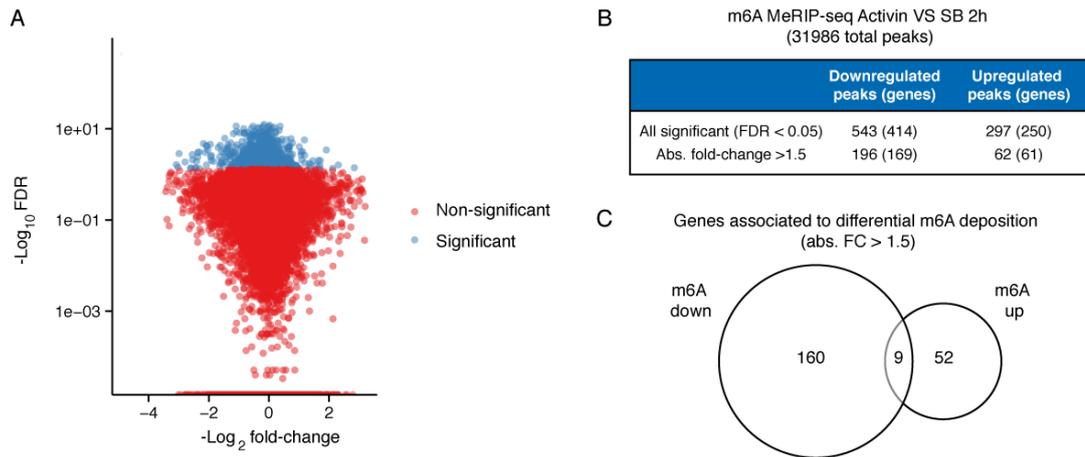


Figure 5.32. Analysis of differential m6A deposition following Activin/Nodal inhibition.

(A) Plot of statistical significance against fold-change of m6A peaks in the experiment described in Figure 5.30. Data is colour-coded according to the selected significance cut-off (FDR < 0.05). (B) Quantification of differential m6A deposition. (C) Overlap between genes showing decreased (down) or increased (up) m6A deposition. These analyses were performed in collaboration with Dr Pedro Madrigal.

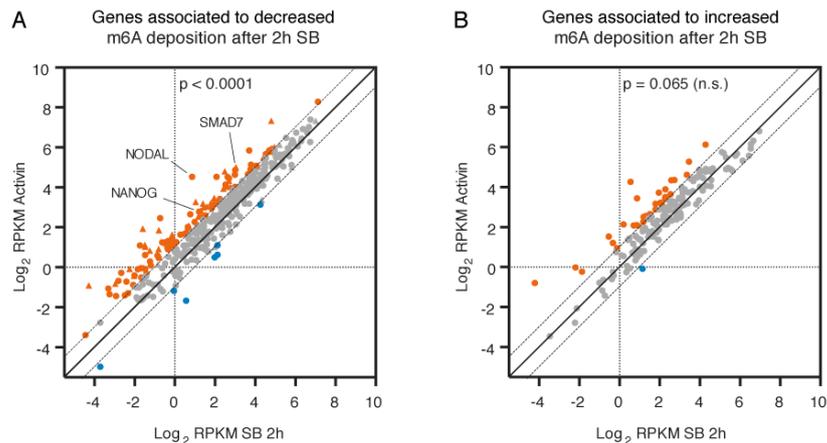


Figure 5.33. Correlation between differential m6A deposition and gene expression changes following Activin/Nodal inhibition.

Average expression after 2h of SB treatment of the genes associated to decreased (A) or increased (B) m6A abundance (all significant differences, Figure 5.32). Gene expression was quantified from the pre-MeRIP input RNA (Figure 5.30). The dotted lines indicate the selected cut-off for differential expression (abs.FC > 2), and dots are colour-coded accordingly (grey: no difference; red: upregulation in SB; blue: downregulation in SB). In A, triangles indicate genes that were associated both to decreased and increased m6A peaks. The significance of global expression differences compared to the null model (no changes) is reported (F test comparison of fit following non-linear regression).

Genes associated to Activin/Nodal signalling-sensitive m6A deposition

Decreased expression after 2h SB	Stable expression after 2h SB
<p>KEGG PATHWAY ENRICHMENT</p> <ul style="list-style-type: none"> - TGF-beta signaling pathway (Rank 1, adj.p=9.87E-07) - Signaling pathways regulating pluripotency of stem cells (Rank 2, adj.p=2.87E-06) <p>GO BIOLOGICAL PROCESS ENRICHMENT</p> <ul style="list-style-type: none"> - Epithelium development (Rank 2, adj.p=1.37E-05) - Cellular response to transforming growth factor beta stimulus (Rank 4, adj.p=1.42E-56) <p>REPRESENTATIVE GENES</p> <p><i>NANOG, WNT3, LIF, NODAL, SMAD7, SKIL, SMAD3, SMAD6, CDKN2B, BMP2, RUNX1, JUN</i></p>	<p>GO BIOLOGICAL PROCESS ENRICHMENT</p> <ul style="list-style-type: none"> - Tissue morphogenesis (Rank 1, adj.p=5.52E-03) - Morphogenesis of an epithelium (Rank 2, adj.p=5.53E-03) - Embryo development (Rank , adj.p=1.92E-02) <p>GO CELLULAR COMPONENT ENRICHMENT</p> <ul style="list-style-type: none"> - Extracellular matrix (Rank 1, adj.p=1.77E-02) - Transcriptional repressor complex (Rank 3, adj.p=4.55E-02) - Stress fiber (Rank 5, adj.p=4.99E-02) <p>REPRESENTATIVE GENES</p> <p><i>TGFB1, NOTCH1, WNT3A, GDF7, HEY2, TEAD1, COL18A1, COL16A1, MMP17, NCOR2, SATB2, MYL9</i></p>

Figure 5.34. Functional enrichment analysis of differentially methylated transcripts.

Selected results from gene GO or KEGG pathway enrichment analysis on genes associated to differential m6A deposition following 2h of SB. Genes were further subdivided based on the stability of their expression during the experiment (Figure 5.33). For each term, its rank in the analysis and the significance of its enrichment are reported.

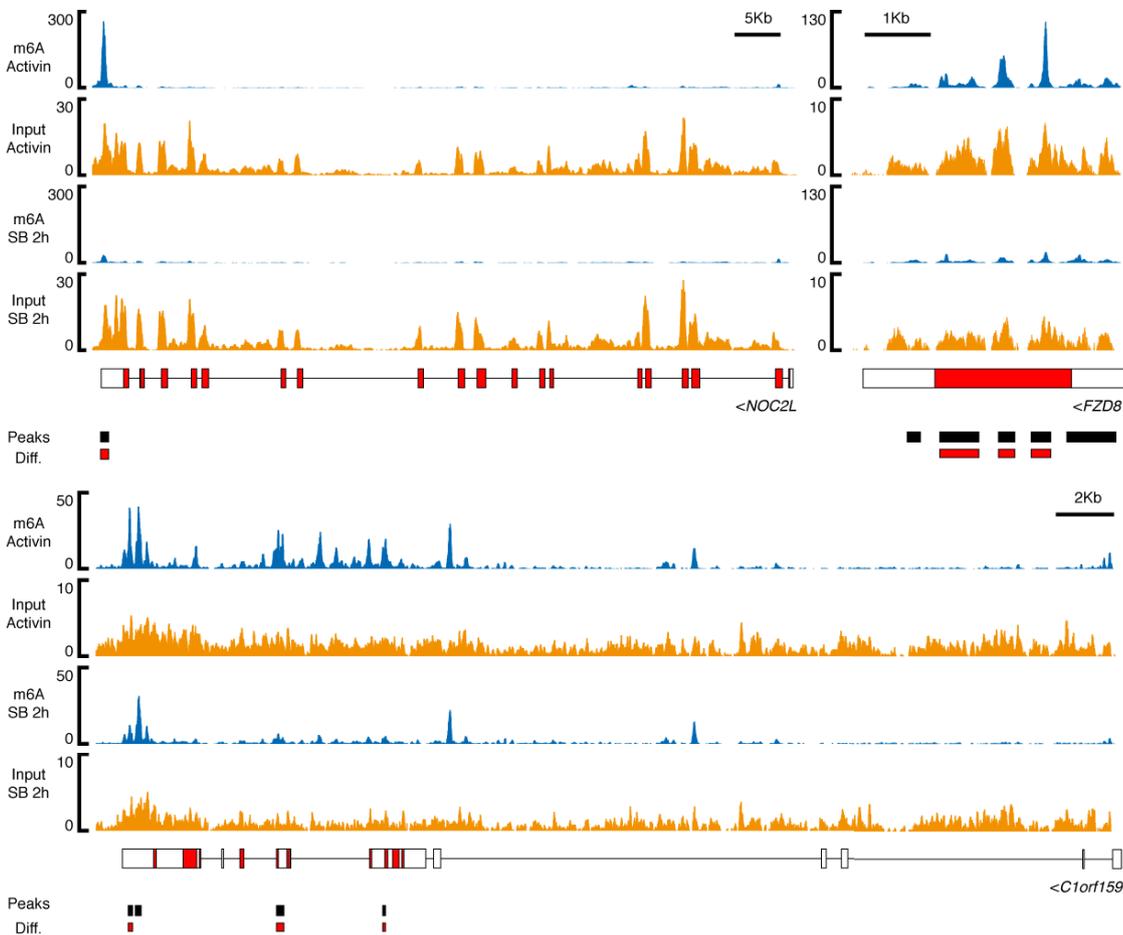


Figure 5.35. Differential m6A deposition following Activin/Nodal inhibition.

As in Figure 5.30, but for representative transcripts whose expression is stable following Activin/Nodal inhibition for 2h. The m6A MeRIP and input tracks were separated and have a different scale in order to facilitate visual comparison between the conditions. The m6A peaks and those differentially abundant (diff.) are indicated. Note that the bioinformatics analysis focused only on exons, and as such the intronic peaks in *C1orf159* were not considered.

Having shown that Activin/Nodal can influence m6A deposition, we then sought to clarify the underlying mechanism. Interestingly, we observed that SMAD2/3 interacts with the m6A methyltransferase complex only in the presence of Activin/Nodal signalling (Figure 5.36). This suggests that phosphorylation of SMAD2/3 can affect binding between these factors, which could allow signalling-controlled regulation of m6A. In principle, this interaction could affect m6A deposition in at least two ways: (1) SMAD2/3 could recruit the m6A methyltransferase complex onto specific genomic regions, thus facilitating m6A marking of nascent transcripts; (2) SMAD2/3 could facilitate the recruitment and/or activity of the m6A methyltransferase complex directly onto nuclear RNAs.

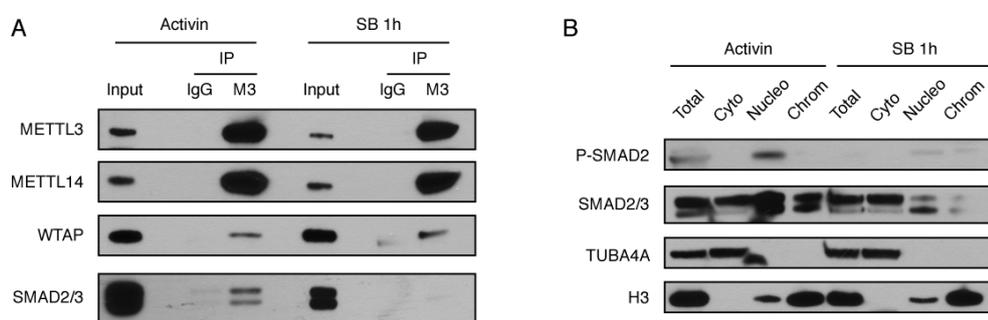


Figure 5.36. Sensitivity of the interaction between SMAD2/3 and the m6A methyltransferase complex to changes in Activin/Nodal signalling.

(A) Representative Western blots of METTL3 (M3) or control (IgG) immunoprecipitations (IP) from nuclear extracts of hESCs maintained in presence of Activin or treated for 1h with SB. Input is 5% of the material used for IP. (B) Western blots of subcellular fractions from the experiment described in A. Total: non-fractionated cell lysate; cyto: cytoplasmic extract; nucleo: nuclear extract; chrom: residual chromatin fraction following extraction of nuclear proteins. Total SMAD2/3 and phosphorylated SMAD2 (P-SMAD2) were probed to validate the efficacy of Activin/Nodal inhibition. Fraction-specific markers confirmed successful subcellular fractionation (TUBA4A/ α -tubulin: cytoplasm; Histone H3: nucleus-chromatin).

The first possibility is somewhat supported by the fact that SMAD2/3 binds closely to a large number of genes whose transcripts are epitranscriptionally regulated by Activin/Nodal signalling (Figure 5.37; of note, however, SMAD2/3 binding sites do not significantly overlap directly with genomic regions corresponding to m6A sites). Nevertheless, ChIP-qPCR experiments for METTL3 failed to identify a significant Activin/Nodal-dependent enrichment for this protein onto any of the well-known SMAD2/3 binding sites that we tested (Figure 5.38). Therefore, we could not find evidence that this type of mechanism is at play.

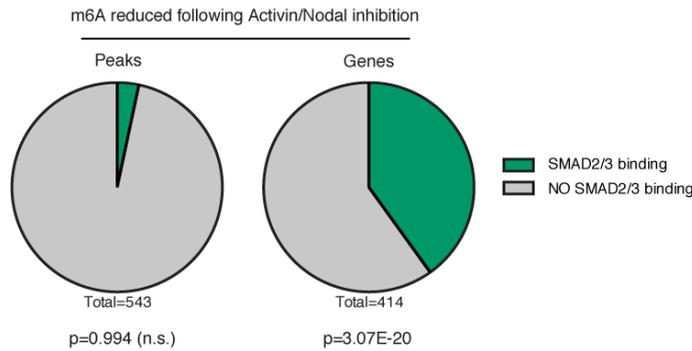


Figure 5.37. Genome-wide correlation between Activin/Nodal-dependent m6A deposition and SMAD2/3 binding.

On the left, proportion of m6A peaks downregulated after 2h of Activin/Nodal inhibition (all significant changes; Figure 5.32) that contain a SMAD2/3 binding site on the corresponding genomic regions. The significance of the observed overlap was calculated using GAT (Heger et al. 2013). n.s: non-significant. On the right, proportion of genes showing Activin/Nodal-dependent m6A deposition that are associated to SMAD2/3 genomic binding sites. The hypergeometric probability of the observed overlap is reported. SMAD2/3 ChIP-seq data for both analysis is from Brown et al., 2011.

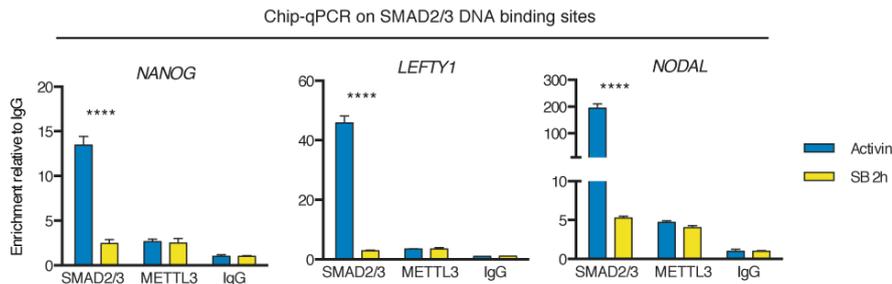


Figure 5.38. DNA binding of the m6A methyltransferase complex.

Results of SMAD2/3, METTL3, or negative control IgG ChIP-qPCR for genomic SMAD2/3 binding sites on the indicated genes. hESCs were cultured in presence of Activin or treated for 2h with SB. Three biological replicates were analysed per condition. Significant differences for Activin vs SB 2h are reported (two-way ANOVA with post-hoc Sidak comparisons).

To explore the second possibility, we performed RNA co-immunoprecipitations for SMAD2/3 to test whether it could be directly binding to nuclear RNAs (Figure 5.39). However, we could not identify any reproducible binding of SMAD2/3 to any RNA we tested. On the other side, we observed that binding of METTL3 to *NANOG* and *LEFTY1* is impaired following inhibition of Activin/Nodal signalling (Figure 5.39). Therefore, while we could not identify the precise molecular mechanism behind this, Activin/Nodal signalling might control m6A deposition by modulating the recruitment of the m6A methyltransferase complex onto target RNAs.

Concluding, the results presented in this sub-chapter suggested that Activin/Nodal signalling modulates the epitranscriptional m6A landscape of hPSCs in order to finely modulate the expression of several genes, including factors transcriptionally regulated

by SMAD2/3 (Figure 5.40). Moreover, such regulations appear important to allow rapid changes of transcriptional programs during the exit from pluripotency of hPSCs.

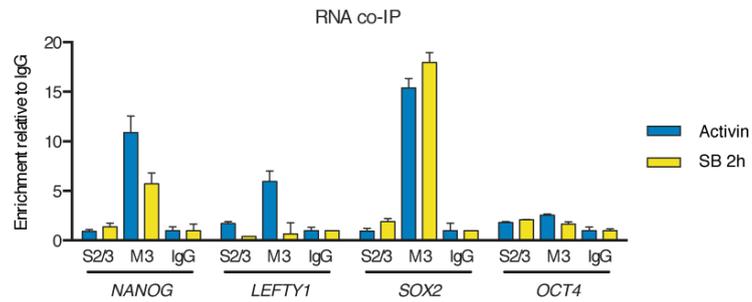


Figure 5.39. RNA binding of SMAD2/3.

Representative results of SMAD2/3 (S2/3), METTL3 (M3), or negative control IgG RNA co-immunoprecipitations from nuclear extracts of hESCs. The samples were analysed by qPCR for regions containing m6A sites. hESCs were cultured in presence of Activin or treated for 2h with SB. *OCT4* is a negative control for the experiment as it does not contain any m6A site. This experiment was performed in collaboration with Ms Stephanie Brown.

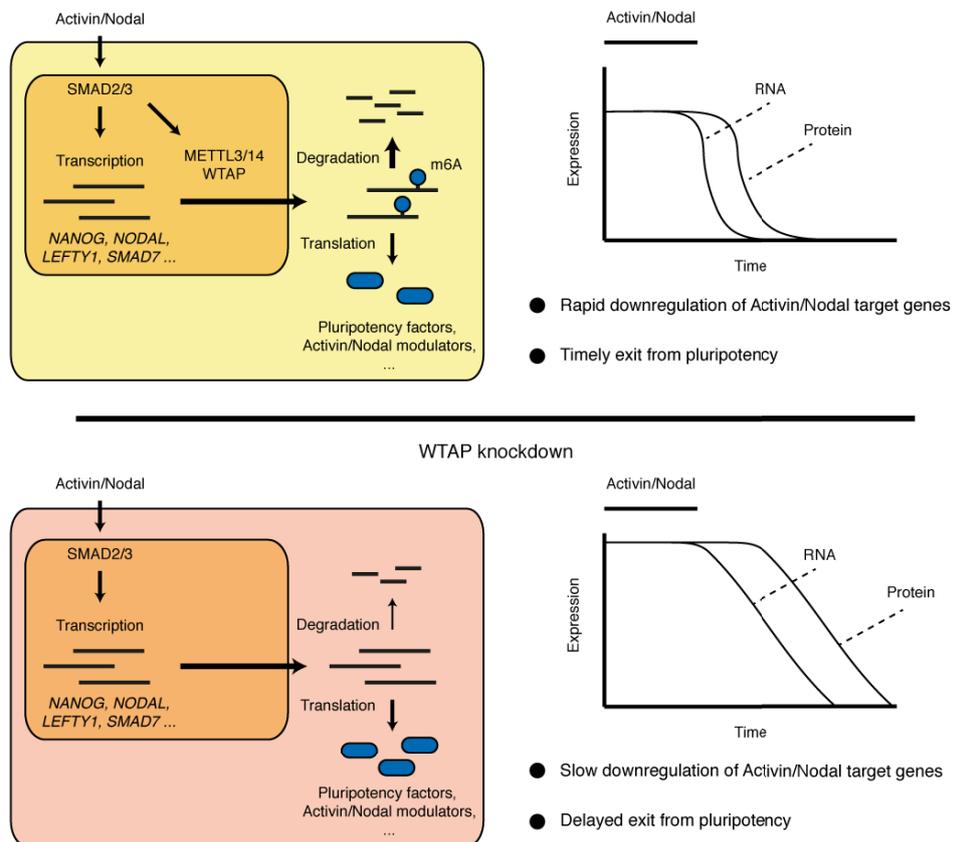


Figure 5.40. Proposed model for the functional interaction between SMAD2/3 and the m6A methyltransferase complex in hPSCs.

SMAD2/3 promotes expression of Activin/Nodal target genes while promoting deposition of m6A on the resulting transcripts, thus decreasing their stability. This represents a negative feedback that allows rapid downregulation of Activin/Nodal target genes and rapid exit from pluripotency upon withdrawal of Activin/Nodal. Such regulation is impaired when the m6A methyltransferase complex is dysfunctional, for instance upon knockdown of WTAP. The detailed mechanism by which SMAD2/3 promotes m6A deposition and its precise effect on the marked transcripts will need to be further elucidated.

5.3 Discussion

This chapter described the identification and initial functional characterization of the SMAD2/3 interactome in both self-renewing and differentiating hPSCs. While similar experiments have been previously performed to clarify the interactome of the core pluripotency factors SOX2, NANOG, and OCT4 (van den Berg et al. 2010; Pardo et al. 2010; Ding et al. 2012; Costa et al. 2013; Wang et al. 2006; Gao et al. 2012), this is the first time that such a study is performed for the Activin/Nodal-regulated transcription factor SMAD2/3, whose role in hPSC early cell-fate decisions is pivotal. More in general, to the best of our knowledge this represents the first mass spectrometry-based unbiased and large-scale analysis of SMAD2/3 interacting proteins in any cell type. By identifying 89 putative SMAD2/3 partners in hPSCs, we provide an invaluable resource to explain the role of Activin/Nodal signalling both in the maintenance of pluripotency and during the induction of differentiation. Indeed, our results not only clarified the transcriptional and epigenetic cofactors of SMAD2/3 in hPSCs, but also led to the discovery that Activin/Nodal signalling is involved in epitranscriptional regulations.

On the other side, it is important to mention that our experiments did not identify certain previously reported SMAD2/3-interacting proteins, most notably NANOG and the COMPASS complexes (Chapter 3). This might be the result of experimental differences such as the specific biochemical conditions used during the preparation of protein extracts, which might have negatively affected the stability of certain protein-protein interactions (Chapter 5.2.1). Alternatively, technical limitations intrinsic to the mass spectrometry analysis could be to blame. In particular, it has been reported that NANOG is particularly difficult to identify by mass spectrometry both because it is resistant to tryptic digestion, and since only a few of its derived peptides can be measured by the mass spectrometer due to poor ionization and suboptimal length (van den Berg et al. 2010; Ding et al. 2012). Nevertheless, despite these limitations our analysis greatly expanded the repertoire of information regarding SMAD2/3-centered protein-protein interactions in hPSCs.

Of particular interest, our data suggests that SMAD2/3 interacts with more than 15 different transcription factors in hPSCs. Aside from well-known SMAD2/3 cofactors such as SMAD4, SKI, and SNON, most of these proteins have never been previously reported to interact with SMAD2/3. These include multiple transcription factors (for

instance ETV6, NFAT5, and SOX13) whose role in hPSCs is unknown despite being crucial for other developmental processes (Tsuzuki & Seto 2013; Ford et al. 2009; Melichar et al. 2007; Y. Wang et al. 2005; Berga-Bolaños et al. 2013). Additionally, SMAD2/3 binds to a large number of proteins involved in chromatin epigenetic modifications. Among these, we showed that the histone acetyltransferases EP300 and CREBBP are particularly important to maintain the pluripotent state. Given their well-known functional cooperation with SMAD2/3 in other cellular contexts (Chapter 1.7.3), it is likely that a large proportion of the effects induced by knockdown of these factors is due to impairment of Activin/Nodal-dependent transcriptional responses. Nevertheless, our experiments did not aim to directly address this point, which will need to be further clarified in future studies.

To our surprise, the interactome of SMAD2/3 is remarkably similar in undifferentiated hPSCs and hPSCs differentiating into endoderm. Nevertheless, a few factors appear to differentially bind to SMAD2/3 in each condition. The most remarkable example is FOXH1, which strongly interacts with SMAD2/3 only during endoderm differentiation. This is in agreement with previous ChIP-seq results, which showed that FOXH1 and SMAD2/3 only weakly colocalize in pluripotent cells, while their genomic binding largely overlaps during endoderm differentiation (Kim et al. 2011). Interestingly, our experiments showed that the knockdown of FOXH1 specifically impairs endoderm specification, while it has only a limited effect on the pluripotent state. Finally, loss of *Foxh1* in the mouse embryo does not cause overt defects in the post-implantation epiblast, while it specifically impairs patterning of the anterior primitive streak and formation of the node, prechordal mesoderm, notochord, and definitive endoderm (von Both et al. 2004; Yamamoto et al. 2001; Hoodless et al. 2001). Altogether, these observations suggest that FOXH1 is a key differentiation-specific SMAD2/3 cofactor that regulates SMAD2/3 genomic binding to drive differentiation of anterior mesendoderm lineages, in particular definitive endoderm (Chapter 1.7.4.3). Overall, only limited differences in the SMAD2/3 interactome could be sufficient to substantially modify the outcome of Activin/Nodal signalling over the timescale analysed. On the other side, it is possible that at later stages of hPSC differentiation the changes in the SMAD2/3 interactome might become more significant, with novel partners such as EOMES driving yet other SMAD2/3-dependent transcriptional responses (Teo et al. 2011; Chapter 1.7.4.3).

Interestingly, our results showed that knockdown of the histone acetyltransferases EP300 and CREBBP has only a minimal effect on endoderm specification of hPSCs. This is in agreement with the fact that gastrulation is not affected in mice knockout for Ep300 or Crebbp, which only show later embryonic defects such as heart malformations, defective neurulation, and impaired haematopoiesis and vasculogenesis (Oike et al. 1999; Kung et al. 2000; Yao et al. 1998; Shikama et al. 2003). This suggests that EP300 and CREBBP might be redundant during Activin/Nodal-induced endoderm specification, either because they compensate for each other or because other epigenetic regulators play a more significant role (for instance the COMPASS complexes, Chapters 3 and 4.3.6).

Perhaps the most interesting finding of our mass spectrometry experiments was that SMAD2/3 interacts not only with transcription factors and epigenetic regulators, but also with complexes involved in several other biological processes. To date, the only other established function of SMAD2/3 aside from modulation of transcriptional responses is the control of miRNA processing (Davis et al. 2008; Blahna & Hata 2012). Our results argue that SMAD2/3 might be involved in multiple other regulations of nuclear RNAs, including splicing, modification, export and degradation. Additionally, SMAD2/3 could also be implicated in the control of DNA repair and apoptosis. While each of these aspects will need to be individually assessed in detail in future studies (Chapter 6.1.2), it appears likely that SMAD2/3 has many “non-canonical” roles in hPSCs.

In particular, we showed that Activin/Nodal signalling controls m6A deposition, as inhibition of Activin/Nodal rapidly results into reduced m6A levels onto multiple transcripts. Further, our results suggest that m6A marking might represent a negative feedback by which Activin/Nodal signalling “poises” several of its transcriptional targets for rapid degradation upon signalling withdrawal. This model is largely extrapolated from the phenotype we observed following genetic impairment of the m6A methyltransferase complex in hESCs, in which the transcriptional response to Activin/Nodal signalling is potentiated. However, the precise effects of Activin/Nodal-dependent m6A deposition onto individual mRNAs will need to be more precisely explored. Indeed, while it has been established that m6A deposition is generally associated with decreased stability of mRNAs, there are several other possible functions

linked to m6A (Chapter 1.6.2.1). For instance, Activin/Nodal-dependent m6A deposition could regulate mRNA splicing, an aspect that was not explored in our study.

Of note, the results presented in this chapter did not provide a detailed mechanistic explanation of how Activin/Nodal signalling controls the levels of m6A on selected transcripts. Indeed, while SMAD2/3 interacts with the m6A methyltransferase complex in an Activin/Nodal-dependent fashion, the precise consequences of this interaction are unclear. Nevertheless, our findings suggest that inhibition of Activin/Nodal signalling impairs binding of the m6A methyltransferase complex to certain transcripts such as *NANOG* and *LEFTY1*. We hypothesize that this could explain the decreased deposition of m6A on such mRNAs in the absence of Activin/Nodal signalling. Of note, while we could not demonstrate binding of SMAD2/3 to these transcripts, it is possible that this was due to the technical limitations of the UV cross-linking RNA co-immunoprecipitation method used, which is known to only capture the most direct and stable RNA-protein interactions. The use of more sophisticated and sensitive approaches such as photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) or individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP) will be pivotal in further exploring this aspect (Hafner et al. 2010; König et al. 2010; Li et al. 2014). On the other side, while ChIP experiments could not demonstrate genomic binding for METTL3, this technique is also known to efficiently capture only stable and direct protein-DNA interactions. Moreover, ChIP for SMAD2/3 is already very challenging and results in poor enrichments due to the low affinity of SMAD2/3 for the DNA (Brown et al. 2011). Therefore, it is possible that traditional ChIP experiments might not be sensitive enough to identify potential genomic recruitment of m6A methyltransferase, which would probably be transient and not relying on direct binding to the DNA. Overall, multiple mechanisms open for future investigation might explain the functional cooperation between SMAD2/3 and m6A methyltransferases.

The role of m6A in pluripotency has been the subject of several contradictory reports, and eventually it was proposed that the function of m6A differs in the naïve and primed pluripotent states (described in detail in Chapter 1.6.2.2). This model was mainly developed following studies in the mouse, and based on these predictions depletion of m6A in a primed pluripotent state should reinforce the predominant transcriptional network, thus driving differentiation. However, our results demonstrated that decreased

m6A in hESCs does not interfere with self-renewal and pluripotency. On the contrary, hESCs defective for the m6A methyltransferase complexes are resistant to differentiation into the neuroectodermal lineage, as the pluripotency factor *NANOG* is not rapidly downregulated. Our findings are in agreement with the only other published study that looked at the function of m6A in hESCs (Batista et al. 2014), and suggest that correct m6A deposition in hESCs is important to facilitate the exit from pluripotency and initiation of neuroectoderm differentiation. Of note, however, we did not observe any significant defects during endoderm differentiation of hESCs depleted for the subunits of the m6A methyltransferase complex. This apparent paradox could be explained by the fact that expression of *NANOG* is actually necessary for the early stages of endoderm specification (Teo et al. 2011). As such, delayed downregulation of *NANOG* might not have a strong detrimental effect on the specification of this germ layer.

Concluding, the results presented in this chapter establish that SMAD2/3 is not only involved in transcriptional and epigenetic regulations, but it is implicated also in post-transcriptional controls of gene expression. Aside from their immediate relevance to the hPSC field, these findings open the possibility that similar mechanisms might be important also in other cellular contexts. Therefore, our results provide an invaluable resource to explore the molecular mechanisms involved in TGF β superfamily signalling during human development, physiology, and disease (Chapter 6.1.3).

6 FUTURE DIRECTIONS AND CONCLUSIONS

6.1 Future directions

The discussions in Chapters 3, 4, and 5 outlined several important experimental avenues to further validate, explore, and expand the findings presented in each individual chapter. In this section, potential future directions of the collective work presented in this dissertation are discussed.

6.1.1 Relationship between epigenome and epitranscriptome controls by SMAD2/3

The work presented in Chapters 3 and 5 showed that SMAD2/3 regulates both the epigenome and epitranscriptome of hPSCs. It is intriguing to speculate that these two aspects might be coordinated. On one level, these regulations could be functionally correlated: histone marks might dictate the propensity to epitranscriptome modifications (for instance by affecting enzyme recruitment on nascent transcripts), and deposition of post-translational marks might in turn influence chromatin dynamics. On the other hand, these mechanisms could even be physically coordinated by the formation of “oligoenzymatic” complexes linked by SMAD2/3. Global mapping of H3K4me3 and m6A marks upon genetic impairment of the m6A methyltransferase

complex and of COMPASS complexes, respectively, will be instructive in testing this hypothesis.

6.1.2 Additional potential functions of SMAD2/3

The proteomic screening of SMAD2/3 interactors presented in Chapter 5 highlighted that SMAD2/3 interacts with several complexes involved in diverse nuclear functions. Aside from those explored in this dissertation, there are several other potentially interesting aspects that might be controlled by Activin/Nodal signalling in hPSCs, such as various RNA processing events and DNA repair mechanisms. Each of these aspects should be investigated by combining functional studies (for instance by applying the OPTiKD platforms described in Chapter 4), and other specific methods. Accordingly, this work is already ongoing in our laboratory.

6.1.3 Molecular mechanisms of SMAD2/3 activity in other models

Moving forward from conventional hPSCs, it will be of important to test whether the molecular regulations uncovered in this work are relevant also in other cellular contexts.

On such model of prime interest is human naïve pluripotency. Indeed, in contrast to naïve mESCs, all reported naïve hPSCs either explicitly rely on exogenous Activin or TGF β , or require this signalling by autocrine or paracrine mechanisms (Gafni et al. 2013; Theunissen et al. 2014; Chan et al. 2013). Interestingly, recent sequencing analyses of the human pre-implantation epiblast showed that these cells express key components of the Activin/Nodal signalling pathway (Blakeley et al. 2015; Petropoulos et al. 2016). On the contrary, mouse embryos only start expressing Nodal in the peri-implantation epiblast (Varlet et al. 1997; Papanayotou et al. 2014). Moreover, inhibition of Activin/Nodal in human pre-implantation blastocysts impairs expression of pluripotency markers, thus demonstrating that this pathway is involved in maintaining the human naïve pluripotent state (Blakeley et al. 2015). Remarkably, similar observations were made in non-human primates (Boroviak et al. 2015), showing that this mechanism is evolutionary conserved. Overall, these findings suggest that Activin/Nodal is not only required for primed human pluripotency, but also for the

naïve state. Therefore, it will be of interest to address whether the mechanisms uncovered in this dissertation are also involved in the regulation of cell fate choices in naïve hPSCs. This could be readily assessed by performing experiments similar to those described in Chapters 3 and 5, but in hPSCs cultured in naïve conditions.

Finally, as discussed in Chapters 3.4 and 5.3, TGF β signalling is involved in numerous aspects of human development, physiology, and disease. Of particular interest, TGF β signalling regulates the balance between self-renewal and differentiation of several adult stem cells as well as of many cancer stem cells (Caja et al. 2012). Therefore, the regulations uncovered in this dissertation could have a broad implication in multiple fields of investigation.

6.1.4 Repurposing the OPTiKD methods for other applications

We envision that the inducible knockdown platforms developed in Chapter 4 could be repurposed to drive other types of inducible small RNAs, such as gRNAs for CRISPR/Cas9 applications (Chapter 1.8.1) or antagomir and miRNA sponges in order to study miRNA function (Ebert & Sharp 2010). Collectively, these methods could largely expand the toolbox available to the field to probe the function of both coding and non-coding genetic elements.

In this light, we have been recently exploring the development of a conditional knockout system based on CRISPR/Cas9 technology. For that, we replaced the inducible shRNA used in the sOPTiKD method with an inducible gRNA. In the presence of a constitutively expressed Cas9, the resulting system allows efficient generation of knockout cells following tetracycline-mediated induction (>95%) without leakiness (less than 1% gene editing following more than 10 passages in the absence of tetracycline). In the future, we plan to further refine this method and expand it also to allow other types of inducible CRISPR/Cas9-based applications, such as transcriptional interference or activation, and targeted epigenetic modifications (Chapter 1.8.1).

6.2 Conclusions

According to the University of Cambridge guidelines for the award of a PhD Degree, the candidate must demonstrate that his or her dissertation “represents a significant contribution to learning, for example through the discovery of new knowledge, the connection of previously unrelated facts, the development of new theory, or the revision of older views”. To this end, this dissertation has accomplished such criteria through the following:

- Uncovering that the Activin/Nodal signalling pathway modulates the H3K4me3 epigenetic landscape of hPSCs in order to regulate early cell fate decisions.
- Demonstrating that SMAD2/3 functionally cooperates with NANOG and DPY30 to promote H3K4me3 of Activin/Nodal target genes.
- Elucidating the function of DPY30 during early embryonic development.
- Engineering a robust method to conditionally probe gene function in hPSCs and during their differentiation into mature lineages.
- Revealing that DPY30 has stage- and lineage-specific functions during hPSC differentiation.
- Identifying the SMAD2/3 interactome in undifferentiated and endoderm-differentiating hPSCs.
- Characterizing the function of multiple SMAD2/3 transcriptional and epigenetic cofactors in hPSCs.
- Discovering that Activin/Nodal signalling modulates the epitranscriptome of hPSCs by affecting the deposition of m6A.
- Clarifying the function of the m6A methyltransferase complex in hPSCs.

Collectively, the results presented in this dissertation advanced both the knowledge and the methodologies available to the field, and will significantly contribute to several exciting future developments in the years to come.

7 APPENDICES

Appendix I: H3K4me3 peaks decreased after inhibition of Activin/Nodal signalling for 2 hours with SB.....	250
Appendix II: Plasmid maps for the OPTiKD and sOPTiKD systems	258
Appendix III: SMAD2/3 interacting proteins identified by quantitative label-free mass spectrometry	265
Appendix IV: m6A peaks decreased after inhibition of Activin/Nodal signalling for 2 hours with SB.....	267

APPENDIX I: H3K4ME3 PEAKS DECREASED AFTER INHIBITION OF ACTIVIN/NODAL SIGNALLING FOR 2 HOURS WITH SB

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr1	226074361	226077600	-2.34	2.77E-08	865	<i>LEFTY1</i>	protein-coding
chr10	104965141	104967120	-1.77	1.65E-08	-13068	<i>NT5C2</i>	protein-coding
chr9	14720941	14722740	-1.69	6.89E-09	874	<i>CER1</i>	protein-coding
chr11	68401501	68403300	-1.68	3.28E-09	-49582	<i>GAL</i>	protein-coding
chr1	204682561	204684240	-1.66	3.28E-09	-28804	<i>LRRN2</i>	protein-coding
chr20	23729881	23731620	-1.66	3.28E-09	823	<i>CST1</i>	protein-coding
chr11	81736801	81738300	-1.63	3.28E-09	-135673	<i>MIR4300</i>	ncRNA
chr4	3928921	3932340	-1.63	3.28E-09	26517	<i>FAM86EP</i>	pseudo
chr6	160711561	160712820	-1.63	1.16E-08	-32228	<i>SLC22A2</i>	protein-coding
chr4	89453701	89454540	-1.6	3.28E-09	-9169	<i>PIGY</i>	protein-coding
chr3	191617201	191618580	-1.57	3.28E-09	438939	<i>PYDC2</i>	protein-coding
chr3	106640641	106641780	-1.56	3.28E-09	318274	<i>LINC00882</i>	ncRNA
chr11	87776821	87779040	-1.52	3.28E-09	130704	<i>RAB38</i>	protein-coding
chr3	129777841	129780720	-1.49	3.28E-09	-21393	<i>ALG1L2</i>	protein-coding
chr1	225032281	225033360	-1.45	7.99E-09	-84535	<i>DNAH14</i>	protein-coding
chr18	32796121	32797620	-1.45	3.45E-09	-24123	<i>ZNF397</i>	protein-coding
chr5	53175841	53176740	-1.45	1.41E-08	71138	<i>MIR581</i>	ncRNA
chr5	116106601	116108340	-1.44	3.28E-09	-9566	<i>LOC102467223</i>	ncRNA
chr8	82814761	82815660	-1.43	1.42E-07	-60690	<i>SNX16</i>	protein-coding
chr9	114291121	114291960	-1.41	9.65E-09	4102	<i>ZNF483</i>	protein-coding
chr13	107270161	107271060	-1.39	3.28E-09	453	<i>LINC00551</i>	ncRNA
chr4	120920041	120921360	-1.39	3.28E-09	67312	<i>MAD2L1</i>	protein-coding
chr6	98230681	98231640	-1.39	1.72E-08	-241246	<i>MIR2113</i>	ncRNA
chr4	134190061	134191140	-1.38	3.28E-09	120131	<i>PCDH10</i>	protein-coding
chrX	12961561	12962460	-1.38	9.41E-07	-592	<i>TLR8-AS1</i>	ncRNA
chr14	51360601	51361980	-1.37	7.01E-09	22413	<i>ABHD12B</i>	protein-coding
chr5	166728121	166730580	-1.37	3.28E-09	17508	<i>TENM2</i>	protein-coding
chr18	46466461	46467180	-1.36	6.02E-09	2356	<i>SMAD7</i>	protein-coding
chr15	92666461	92667480	-1.35	3.28E-09	270033	<i>SLCO3A1</i>	protein-coding
chr3	137314621	137315820	-1.35	3.28E-09	-167913	<i>SOX14</i>	protein-coding
chr3	143520241	143522460	-1.34	3.28E-09	46022	<i>SLC9A9</i>	protein-coding
chr5	135903961	135905040	-1.34	5.23E-09	-203337	<i>TRPC7</i>	protein-coding
chr2	7569901	7570500	-1.31	4.04E-08	8809	<i>LOC100506274</i>	ncRNA
chr8	33484201	33484800	-1.31	7.43E-08	-27062	<i>DUSP26</i>	protein-coding
chrX	31395841	31399140	-1.31	3.28E-09	-112467	<i>DMD</i>	protein-coding
chr2	209051221	209052060	-1.3	3.95E-07	3132	<i>C2orf80</i>	protein-coding
chr5	18669421	18671100	-1.3	3.28E-09	-1282842	<i>LOC401177</i>	ncRNA
chr1	53897701	53901300	-1.29	4.05E-09	6192	<i>SLC25A3P1</i>	pseudo
chr3	186717361	186718320	-1.28	5.58E-08	-21824	<i>ST6GAL1</i>	protein-coding
chr4	87401641	87403200	-1.28	3.28E-09	-28138	<i>MAPK10</i>	protein-coding
chr6	47623861	47625420	-1.28	5.04E-07	315	<i>GPR111</i>	protein-coding
chr1	209623861	209626020	-1.27	3.28E-09	19463	<i>MIR205</i>	ncRNA
chr20	60010561	60011160	-1.27	1.93E-07	-63616	<i>CDH4</i>	protein-coding
chr11	63226921	63228060	-1.25	3.28E-09	31189	<i>HRASLS5</i>	protein-coding
chr6	16263421	16264920	-1.25	4.06E-09	25360	<i>GMPR</i>	protein-coding
chr1	79958161	79959360	-1.24	1.24E-06	-486266	<i>ELTD1</i>	protein-coding
chr12	12211681	12213360	-1.24	3.28E-09	-11357	<i>BCL2L14</i>	protein-coding
chr4	134182741	134184600	-1.24	3.28E-09	113201	<i>PCDH10</i>	protein-coding
chr1	55490281	55491060	-1.23	1.78E-08	-14478	<i>PCSK9</i>	protein-coding
chr2	200049901	200051100	-1.23	3.28E-09	272318	<i>SATB2</i>	protein-coding
chr10	27754441	27755820	-1.22	1.92E-08	-37972	<i>RAB18</i>	protein-coding
chr10	91044841	91045980	-1.22	7.89E-08	-16295	<i>IFIT2</i>	protein-coding
chr11	128452981	128453880	-1.22	6.60E-09	4022	<i>ETS1</i>	protein-coding
chr15	74642881	74646120	-1.22	3.28E-09	-9392	<i>LOC729739</i>	pseudo
chr22	19867081	19868460	-1.22	3.45E-09	-25309	<i>GNB1L</i>	protein-coding
chr12	8958181	8958960	-1.21	7.33E-09	-16579	<i>A2ML1</i>	protein-coding
chr14	71591101	71593200	-1.21	7.55E-08	218029	<i>PCNX</i>	protein-coding
chr3	179242801	179244060	-1.21	3.28E-09	-37277	<i>ACTL6A</i>	protein-coding

Chapter 7: Appendices

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr8	128444161	128444880	-1.21	8.13E-08	16664	<i>POU5F1B</i>	protein-coding
chr9	117942361	117944580	-1.21	3.28E-09	39374	<i>DEC1</i>	protein-coding
chr18	56308921	56309700	-1.2	3.28E-09	-13122	<i>ALPK2</i>	protein-coding
chr2	238515661	238519440	-1.19	3.28E-09	-17782	<i>RAB17</i>	protein-coding
chr4	98060701	98061780	-1.19	3.52E-08	-226836	<i>STPG2-AS1</i>	ncRNA
chr5	135905101	135906420	-1.19	7.15E-07	-204597	<i>TRPC7</i>	protein-coding
chr6	16886521	16890540	-1.18	3.28E-09	-126810	<i>ATXN1</i>	protein-coding
chr9	6776521	6777120	-1.18	3.41E-06	19180	<i>KDM4C</i>	protein-coding
chr9	102905641	102908820	-1.18	4.37E-09	45729	<i>INVS</i>	protein-coding
chr16	21233641	21234240	-1.17	1.12E-06	-11073	<i>ZP2</i>	protein-coding
chr6	142337161	142338720	-1.16	8.46E-09	71995	<i>NMBR</i>	protein-coding
chr9	95726401	95727360	-1.16	1.96E-08	638	<i>FGD3</i>	protein-coding
chrX	28522801	28525980	-1.16	3.28E-09	-10611	<i>MIR6134</i>	ncRNA
chr17	44850901	44851500	-1.15	1.86E-06	44925	<i>WNT3</i>	protein-coding
chr18	71949241	71949840	-1.15	3.43E-06	9710	<i>CYB5A</i>	protein-coding
chr21	27453061	27453780	-1.15	8.42E-06	59287	<i>APP</i>	protein-coding
chr5	59004301	59004900	-1.15	1.61E-06	59837	<i>PDE4D</i>	protein-coding
chr7	124897441	124898640	-1.15	9.34E-09	6304	<i>RP11-3B12.3</i>	ncRNA
chrX	51199201	51200880	-1.15	3.28E-09	39418	<i>NUDT11</i>	protein-coding
chr10	53673721	53674800	-1.14	3.28E-09	-214906	<i>CSTF2T</i>	protein-coding
chr11	121511281	121514160	-1.14	3.28E-09	189809	<i>SORL1</i>	protein-coding
chr17	72888601	72889200	-1.14	3.47E-06	804	<i>FADS6</i>	protein-coding
chr19	782581	783840	-1.14	3.28E-09	-14181	<i>PTBP1</i>	protein-coding
chr4	140079541	140080200	-1.14	8.99E-08	-19241	<i>ELF2</i>	protein-coding
chr4	157026541	157027620	-1.14	5.71E-09	-152033	<i>CTSO</i>	protein-coding
chrX	129411241	129414960	-1.14	3.28E-09	-10179	<i>ZNF280C</i>	protein-coding
chr17	46810201	46812780	-1.13	3.28E-09	-5380	<i>HOXB13</i>	protein-coding
chr2	82239541	82241280	-1.13	6.95E-09	-546341	<i>ACO12075.2</i>	ncRNA
chr6	14301601	14302200	-1.13	2.93E-07	-16216	<i>LINC01108</i>	ncRNA
chrX	124530421	124531200	-1.13	3.76E-06	76842	<i>LOC100129520</i>	protein-coding
chr16	521101	521820	-1.12	2.39E-06	-3389	<i>RAB11FIP3</i>	protein-coding
chr2	8394961	8396100	-1.12	3.28E-09	73018	<i>LINC00299</i>	ncRNA
chr2	238498801	238499580	-1.12	9.18E-09	578	<i>RAB17</i>	protein-coding
chr6	37094641	37095480	-1.12	3.28E-09	-42861	<i>PIM1</i>	protein-coding
chr7	131338321	131339160	-1.12	2.87E-06	-97365	<i>PODXL</i>	protein-coding
chr8	20133241	20133840	-1.12	3.07E-06	257	<i>LZTS1-AS1</i>	ncRNA
chr1	147038761	147039360	-1.11	1.49E-06	25790	<i>BCL9</i>	protein-coding
chr13	98296261	98298360	-1.11	3.28E-09	210836	<i>RAP2A</i>	protein-coding
chr21	44225761	44226360	-1.11	1.01E-06	73324	<i>WDR4</i>	protein-coding
chr5	109077421	109078140	-1.11	1.13E-06	52714	<i>MAN2A1</i>	protein-coding
chr5	177579181	177579780	-1.11	5.34E-06	1480	<i>NHP2</i>	protein-coding
chr7	116140981	116141760	-1.11	1.29E-06	1273	<i>CAV2</i>	protein-coding
chrX	124521001	124523220	-1.11	3.28E-09	68142	<i>LOC100129520</i>	protein-coding
chrX	137346541	137351880	-1.11	3.45E-09	350588	<i>LINC00889</i>	ncRNA
chr1	27384781	27385740	-1.1	2.05E-07	-45928	<i>FAM46B</i>	protein-coding
chr11	83434501	83435280	-1.1	2.37E-06	-41423	<i>DLG2</i>	protein-coding
chr12	7847281	7848600	-1.1	1.51E-08	419	<i>GDF3</i>	protein-coding
chr9	139465741	139467000	-1.1	3.71E-07	-25660	<i>MIR4674</i>	ncRNA
chr1	224686621	224687220	-1.09	4.70E-06	-64920	<i>WDR26</i>	protein-coding
chr2	157578481	157582200	-1.09	3.28E-09	287441	<i>GPD2</i>	protein-coding
chr21	32532421	32533200	-1.09	6.30E-08	-122016	<i>KRTAP19-8</i>	protein-coding
chr3	162431401	162434160	-1.09	3.28E-09	588308	<i>LINC01192</i>	ncRNA
chr8	8918221	8919000	-1.09	6.05E-08	12656	<i>MIR4660</i>	ncRNA
chr8	68656561	68657280	-1.09	2.04E-07	1699	<i>CPA6</i>	protein-coding
chrX	102684721	102690360	-1.09	4.06E-09	55432	<i>NGFRAP1</i>	protein-coding
chr13	91503181	91504440	-1.08	9.20E-07	75040	<i>LINC00410</i>	ncRNA
chrX	23677381	23678040	-1.08	1.85E-08	-7934	<i>PRDX4</i>	protein-coding
chr10	91845361	91846380	-1.07	3.28E-09	256621	<i>LINC00865</i>	ncRNA
chr22	25090861	25092420	-1.06	3.28E-09	-36527	<i>POM121L10P</i>	pseudo
chr6	33590341	33590940	-1.06	2.76E-06	1485	<i>ITPR3</i>	protein-coding
chr2	35663221	35665500	-1.05	3.28E-09	918352	<i>LOC100288911</i>	ncRNA
chr3	97910281	97912560	-1.05	3.28E-09	23877	<i>OR5H15</i>	protein-coding
chr8	103821841	103822440	-1.05	6.24E-07	54256	<i>AZIN1</i>	protein-coding
chr11	117069841	117072900	-1.04	3.66E-09	1259	<i>LOC100652768</i>	ncRNA
chr12	7940641	7946040	-1.04	1.04E-08	1346	<i>NANOG</i>	protein-coding
chr12	120667141	120668160	-1.04	1.40E-07	20313	<i>PXN</i>	protein-coding

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr2	201065941	201066540	-1.04	1.57E-06	-104363	<i>SPATS2L</i>	protein-coding
chr22	29427901	29428500	-1.04	2.15E-06	-737	<i>ZNRF3-AS1</i>	ncRNA
chr5	135069421	135073080	-1.04	3.45E-09	-81627	<i>LOC340074</i>	ncRNA
chr7	20158921	20159940	-1.04	3.28E-09	20618	<i>AC005062.2</i>	ncRNA
chr8	133093201	133096500	-1.04	2.37E-07	22661	<i>HHLA1</i>	protein-coding
chr9	2281801	2282460	-1.04	1.42E-07	123675	<i>SMARCA2</i>	protein-coding
chrX	13180741	13181400	-1.04	4.03E-06	-118154	<i>FAM9C</i>	protein-coding
chr1	56840221	56840880	-1.03	7.53E-06	204706	<i>PPAP2B</i>	protein-coding
chr21	27600481	27605580	-1.03	3.45E-09	-59585	<i>APP</i>	protein-coding
chr7	130596121	130596960	-1.03	2.92E-08	1528	<i>AC058791.1</i>	ncRNA
chrX	152991121	152992080	-1.03	8.57E-07	1278	<i>ABCD1</i>	protein-coding
chr7	22601701	22602300	-1.02	7.37E-07	-955	<i>AC002480.3</i>	ncRNA
chr8	8919421	8920140	-1.02	8.53E-08	13826	<i>MIR4660</i>	ncRNA
chrX	73028401	73029180	-1.02	1.22E-06	16751	<i>TSIX</i>	ncRNA
chr1	45784741	45787320	-1.01	5.85E-07	-6514	<i>HPDL</i>	protein-coding
chr17	27371281	27371880	-1.01	4.86E-06	1663	<i>PIPOX</i>	protein-coding
chr3	32836501	32837580	-1.01	6.73E-09	-22469	<i>TRIM71</i>	protein-coding
chr1	183782641	183783420	-1	2.09E-06	-160583	<i>APOBEC4</i>	protein-coding
chr11	69588361	69589020	-1	4.01E-07	1480	<i>FGF4</i>	protein-coding
chr18	56245801	56247000	-1	3.89E-08	49788	<i>ALPK2</i>	protein-coding
chr5	106319701	106321020	-1	3.05E-08	26354	<i>LOC102467213</i>	ncRNA
chr9	116357461	116359080	-1	8.45E-07	1837	<i>RGS3</i>	protein-coding
chr14	48730141	48737040	-0.99	6.30E-09	-469374	<i>LINC00648</i>	ncRNA
chr19	781681	782340	-0.99	1.06E-07	-15381	<i>PTBP1</i>	protein-coding
chr22	25029361	25030020	-0.99	2.01E-08	809	<i>BCRP3</i>	pseudo
chr7	93269401	93274680	-0.99	1.28E-08	-67999	<i>CALCR</i>	protein-coding
chr8	70985221	70986240	-0.99	3.28E-09	-2169	<i>PRDM14</i>	protein-coding
chr1	178253101	178254060	-0.98	3.28E-09	-57025	<i>RASAL2</i>	protein-coding
chr11	116313901	116316600	-0.98	4.06E-09	328463	<i>BUD13</i>	protein-coding
chr18	35157661	35158260	-0.98	1.26E-06	-11961	<i>CELFA</i>	protein-coding
chr2	84020041	84020940	-0.98	1.04E-07	-497315	<i>FUNDC2P2</i>	pseudo
chr2	188426881	188429400	-0.98	3.45E-09	-8922	<i>TFPI</i>	protein-coding
chr20	37230301	37232700	-0.98	3.28E-09	924	<i>ARHGAP40</i>	protein-coding
chr20	56281801	56284620	-0.98	8.17E-09	1820	<i>PMEP1</i>	protein-coding
chr21	38960761	38964960	-0.98	4.43E-09	170259	<i>DYRK1A</i>	protein-coding
chr1	175892041	175892940	-0.97	3.45E-09	45185	<i>SCARNA3</i>	ncRNA
chr13	66156841	66157860	-0.97	4.35E-07	-1241950	<i>PCDH9-AS2</i>	ncRNA
chr5	173171101	173173260	-0.97	1.32E-08	1031	<i>LOC101928136</i>	ncRNA
chr7	13055401	13056480	-0.97	3.38E-08	329030	<i>ARL4A</i>	protein-coding
chr22	50730841	50731500	-0.96	3.67E-07	14830	<i>PLXNB2</i>	protein-coding
chrX	122319781	122320380	-0.96	2.74E-07	1985	<i>GRIA3</i>	protein-coding
chr1	119836621	119837580	-0.95	4.83E-07	-74301	<i>HAO2</i>	protein-coding
chr15	24932341	24936000	-0.95	4.07E-09	13630	<i>NPAP1</i>	protein-coding
chr15	93196741	93198540	-0.95	1.76E-08	1390	<i>FAM174B</i>	protein-coding
chr1	167134621	167135340	-0.94	3.28E-09	-55085	<i>POU2F1</i>	protein-coding
chr1	216182761	216186180	-0.94	3.75E-09	412267	<i>USH2A</i>	protein-coding
chr10	6837061	6840720	-0.94	3.29E-09	17331	<i>LINC00707</i>	ncRNA
chr12	46714261	46714980	-0.94	3.28E-09	-51413	<i>SLC38A1</i>	protein-coding
chr3	106630321	106631460	-0.94	3.28E-09	328594	<i>LINC00882</i>	ncRNA
chr4	171038821	171043920	-0.94	4.91E-09	-29833	<i>AADAT</i>	protein-coding
chr1	116191861	116192520	-0.93	2.29E-07	6941	<i>VANGL1</i>	protein-coding
chr10	53801221	53801820	-0.93	5.28E-06	272367	<i>PRKG1-AS1</i>	ncRNA
chr2	202835701	202836300	-0.93	5.14E-06	-63309	<i>FZD7</i>	protein-coding
chr3	99593161	99594060	-0.93	2.08E-04	1436	<i>FILIP1L</i>	protein-coding
chr5	141615781	141616440	-0.93	9.72E-07	88509	<i>SPRY4</i>	protein-coding
chr7	139166761	139167360	-0.93	9.96E-07	1396	<i>KLRG2</i>	protein-coding
chrX	97778761	97779540	-0.93	1.90E-06	1415690	<i>XRCC6P5</i>	pseudo
chr1	211573201	211574700	-0.92	3.28E-09	17854	<i>LINC00467</i>	ncRNA
chr10	72199201	72202560	-0.92	4.96E-09	584	<i>NODAL</i>	protein-coding
chr14	77414341	77415300	-0.92	7.56E-06	80221	<i>IRF2BPL</i>	protein-coding
chr3	192713581	192716400	-0.92	3.79E-09	-79041	<i>MB21D2</i>	protein-coding
chr4	114897961	114898980	-0.92	3.28E-09	2407	<i>ARSJ</i>	protein-coding
chr4	146619061	146619840	-0.92	3.42E-07	18095	<i>C4orf51</i>	protein-coding
chr4	183003841	183006540	-0.92	3.28E-09	60477	<i>MGCA5800</i>	ncRNA
chr5	109076041	109077060	-0.92	3.28E-09	51484	<i>MAN2A1</i>	protein-coding
chr6	63507661	63509100	-0.92	3.17E-07	-51281	<i>KHDRBS2</i>	protein-coding

Chapter 7: Appendices

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr1	53883121	53883960	-0.91	1.99E-07	22152	<i>SLC25A3P1</i>	pseudo
chr15	93054781	93059760	-0.91	4.33E-09	42364	<i>C15orf32</i>	protein-coding
chr3	8218981	8219940	-0.91	1.51E-08	-161467	<i>AC087859.1</i>	ncRNA
chr3	191387041	191389500	-0.91	1.91E-05	209319	<i>PYDC2</i>	protein-coding
chr10	75760501	75761220	-0.9	5.23E-06	2989	<i>VCL</i>	protein-coding
chr12	4134841	4136760	-0.9	2.20E-08	-153187	<i>PARP11</i>	protein-coding
chr13	93881701	93882780	-0.9	2.54E-05	3163	<i>GPC6</i>	protein-coding
chr5	133763581	133766340	-0.9	3.45E-09	5777	<i>LOC102546229</i>	ncRNA
chr6	16259401	16260480	-0.9	3.28E-09	21130	<i>GMPR</i>	protein-coding
chr9	132097081	132097800	-0.9	1.68E-06	14146	<i>C9orf106</i>	protein-coding
chrX	90027961	90028560	-0.9	7.11E-07	-661336	<i>PABPC5</i>	protein-coding
chrX	93960001	93961020	-0.9	1.27E-05	1031499	<i>FAM133A</i>	protein-coding
chr1	213089161	213091020	-0.89	3.28E-09	-33796	<i>VASH2</i>	protein-coding
chr11	9598321	9600120	-0.89	3.28E-09	2987	<i>WEE1</i>	protein-coding
chr11	128455141	128457780	-0.89	3.45E-09	992	<i>ETS1</i>	protein-coding
chr15	57678421	57679140	-0.89	9.78E-06	10078	<i>CNGL1</i>	protein-coding
chr3	54668281	54668880	-0.89	1.50E-06	5303	<i>ESRG</i>	ncRNA
chr4	134187121	134189760	-0.89	3.28E-09	117971	<i>PCDH10</i>	protein-coding
chr5	3102841	3103440	-0.89	4.80E-06	-74808	<i>LOC102467074</i>	ncRNA
chrX	68440261	68441220	-0.89	3.28E-09	41341	<i>LINC00269</i>	ncRNA
chrX	134080201	134081400	-0.89	4.19E-09	-31504	<i>MOSPD1</i>	protein-coding
chr1	223197961	223200660	-0.88	3.45E-09	117313	<i>TLR5</i>	protein-coding
chr1	247540501	247541100	-0.88	2.56E-07	-38657	<i>NLRP3</i>	protein-coding
chr13	54700501	54707040	-0.88	4.16E-08	3235	<i>LINC00458</i>	ncRNA
chr13	74601541	74602380	-0.88	1.71E-05	106105	<i>KLF12</i>	protein-coding
chr16	85344241	85345440	-0.88	2.69E-08	-4910	<i>MIR5093</i>	ncRNA
chr20	12736501	12738660	-0.88	2.62E-07	180030	<i>LOC102606466</i>	ncRNA
chr4	80479201	80480640	-0.88	3.28E-09	66174	<i>LINC00989</i>	ncRNA
chr4	176383021	176388120	-0.88	4.25E-09	323256	<i>GPM6A</i>	protein-coding
chr5	116274721	116275800	-0.88	1.28E-08	-177356	<i>LOC102467223</i>	ncRNA
chr7	148034041	148035360	-0.88	3.28E-09	-252956	<i>C7orf33</i>	protein-coding
chr8	101319181	101320200	-0.88	1.37E-05	2703	<i>RNF19A</i>	protein-coding
chrX	124748461	124749180	-0.88	7.85E-07	294852	<i>LOC100129520</i>	protein-coding
chr1	36352081	36353700	-0.87	4.06E-09	4081	<i>AGO1</i>	protein-coding
chr20	59267341	59268540	-0.87	3.28E-09	214772	<i>MIR4533</i>	ncRNA
chr4	131689201	131690040	-0.87	1.87E-08	1672339	<i>C4orf33</i>	protein-coding
chr8	3098461	3099480	-0.87	3.28E-09	1074302	<i>MIR7160</i>	ncRNA
chr8	116228761	116231400	-0.87	3.28E-09	450158	<i>TRPS1</i>	protein-coding
chrX	113705941	113706540	-0.87	3.02E-06	-112310	<i>HTR2C</i>	protein-coding
chr1	38890741	38895960	-0.86	1.47E-08	-212912	<i>RP11-214L19.1</i>	ncRNA
chr14	47575261	47577180	-0.86	3.28E-09	236228	<i>MDGA2</i>	protein-coding
chr2	76802761	76804560	-0.86	3.28E-09	-409430	<i>AC079117.1</i>	ncRNA
chr22	39637561	39638340	-0.86	5.73E-09	-1037	<i>PDGFB</i>	protein-coding
chr5	116157961	116160360	-0.86	3.28E-09	-61256	<i>LOC102467223</i>	ncRNA
chr7	83045041	83047020	-0.86	7.27E-09	200706	<i>MIR7976</i>	ncRNA
chrX	93954661	93956460	-0.86	3.72E-07	1026549	<i>FAM133A</i>	protein-coding
chrX	113289241	113291340	-0.86	3.28E-09	-528260	<i>HTR2C</i>	protein-coding
chr1	114692941	114693720	-0.85	2.46E-05	1731	<i>SYT6</i>	protein-coding
chr1	147243541	147245700	-0.85	3.28E-09	845	<i>GJA5</i>	protein-coding
chr14	102196861	102197820	-0.85	6.83E-09	567	<i>LINC00239</i>	ncRNA
chr17	1655701	1656480	-0.85	4.50E-05	-9168	<i>SERPINF1</i>	protein-coding
chr17	73342441	73343280	-0.85	1.36E-07	-57331	<i>SLC25A19</i>	protein-coding
chr18	56310721	56313960	-0.85	3.42E-09	-16152	<i>ALPK2</i>	protein-coding
chr22	39417121	39418080	-0.85	1.55E-08	483	<i>APOBEC3D</i>	protein-coding
chr3	112119421	112120980	-0.85	4.06E-09	68285	<i>CD200</i>	protein-coding
chr3	148132801	148133880	-0.85	2.31E-07	-282317	<i>AGTR1</i>	protein-coding
chr4	133045201	133047660	-0.85	4.92E-09	-1024039	<i>PCDH10</i>	protein-coding
chr5	18667561	18669240	-0.85	1.93E-08	-1280982	<i>LOC401177</i>	ncRNA
chr6	11229841	11230920	-0.85	6.39E-08	2534	<i>NEDD9</i>	protein-coding
chr7	75408241	75411420	-0.85	4.69E-09	9233	<i>CCL26</i>	protein-coding
chr8	68657761	68658360	-0.85	2.88E-06	559	<i>CPA6</i>	protein-coding
chr9	2276701	2281260	-0.85	1.22E-08	120525	<i>SMARCA2</i>	protein-coding
chr9	118237321	118241940	-0.85	6.48E-09	266887	<i>LOC101928775</i>	ncRNA
chr17	79138861	79139520	-0.84	3.39E-06	-116	<i>AATK-AS1</i>	ncRNA
chr3	58419781	58420380	-0.84	4.89E-06	-502	<i>PDHB</i>	protein-coding
chr4	111539581	111543360	-0.84	5.24E-09	2783	<i>PITX2</i>	protein-coding

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr5	146941681	146942580	-0.84	6.84E-09	2574	<i>JAKMIP2-AS1</i>	ncRNA
chr1	48677041	48680460	-0.83	4.06E-09	-9606	<i>SLC5A9</i>	protein-coding
chr14	48723241	48728460	-0.83	4.83E-09	-461634	<i>LINC00648</i>	ncRNA
chr4	24502801	24505620	-0.83	4.06E-09	17702	<i>MIR573</i>	ncRNA
chr8	8920381	8921820	-0.83	3.28E-09	15146	<i>MIR4660</i>	ncRNA
chrX	137790541	137791440	-0.83	2.41E-06	2836	<i>FGF13</i>	protein-coding
chr13	69774001	69774780	-0.82	3.15E-06	-314934	<i>LINC00550</i>	ncRNA
chr16	9773881	9775080	-0.82	3.28E-09	-116233	<i>MIR7641-2</i>	ncRNA
chr2	171542041	171542640	-0.82	5.72E-07	28736	<i>LINC01124</i>	ncRNA
chr3	90135481	90137280	-0.82	3.28E-09	979707	<i>EPHA3</i>	protein-coding
chr6	132225901	132228000	-0.82	3.36E-09	45567	<i>CTGF</i>	protein-coding
chr7	25991521	25992120	-0.82	3.28E-09	-2215	<i>MIR148A</i>	ncRNA
chr8	115301401	115302180	-0.82	1.65E-06	-852549	<i>CSMD3</i>	protein-coding
chr1	17914801	17915820	-0.81	2.46E-08	8263	<i>ARHGEF10L</i>	protein-coding
chr18	3624121	3624840	-0.81	6.62E-07	30369	<i>DLGAP1-AS1</i>	ncRNA
chr22	35230741	35235000	-0.81	4.91E-09	-229259	<i>ISX</i>	protein-coding
chr3	21010261	21014820	-0.81	4.06E-09	-434677	<i>VENTXP7</i>	pseudo
chr8	115302481	115303260	-0.81	3.28E-09	-853629	<i>CSMD3</i>	protein-coding
chrX	115949041	115950960	-0.81	3.82E-09	-355807	<i>CT83</i>	protein-coding
chr10	6836221	6836820	-0.8	4.64E-06	14961	<i>LINC00707</i>	ncRNA
chr11	66341221	66344100	-0.8	1.36E-07	-6614	<i>CTSF</i>	protein-coding
chr14	21267301	21271320	-0.8	3.95E-09	1247	<i>RNASE1</i>	protein-coding
chr2	188375281	188376600	-0.8	3.81E-07	43278	<i>TFPI</i>	protein-coding
chr20	48937021	48937620	-0.8	3.26E-06	558	<i>RP11-290F20.2</i>	ncRNA
chr5	126016081	126017100	-0.8	8.69E-09	49177	<i>C5orf48</i>	protein-coding
chr9	88073401	88075620	-0.8	1.59E-08	282433	<i>AGTPBP1</i>	protein-coding
chr1	73444981	73450320	-0.79	6.65E-09	-699374	<i>NEGR1</i>	protein-coding
chr1	229310161	229313820	-0.79	5.01E-09	-94818	<i>RAB4A</i>	protein-coding
chr10	53792401	53793900	-0.79	3.28E-09	280737	<i>PRKG1-AS1</i>	ncRNA
chr10	103533481	103535880	-0.79	6.34E-09	1078	<i>FGF8</i>	protein-coding
chr15	57680101	57680760	-0.79	2.13E-07	11728	<i>CGNL1</i>	protein-coding
chr15	65099941	65100900	-0.79	1.03E-06	17446	<i>PIF1</i>	protein-coding
chr19	47896501	47900100	-0.79	2.46E-08	24484	<i>MEIS3</i>	protein-coding
chr20	12735181	12736260	-0.79	2.99E-06	181890	<i>LOC102606466</i>	ncRNA
chr3	112136401	112138680	-0.79	9.02E-09	80867	<i>BTLA</i>	protein-coding
chr5	150969721	150970680	-0.79	2.20E-07	-21696	<i>FAT2</i>	protein-coding
chr6	37092541	37093320	-0.79	2.03E-08	-44991	<i>PIM1</i>	protein-coding
chr1	185921161	185923140	-0.78	3.25E-08	218468	<i>HMCN1</i>	protein-coding
chr10	82654741	82657320	-0.78	3.42E-07	355456	<i>SH2D4B</i>	protein-coding
chr12	25053961	25054800	-0.78	1.09E-07	941	<i>BCAT1</i>	protein-coding
chr2	216787801	216788460	-0.78	3.28E-09	-79872	<i>LINC00607</i>	ncRNA
chrX	113833501	113835720	-0.78	3.86E-09	16060	<i>HTR2C</i>	protein-coding
chr12	87333241	87334200	-0.77	3.28E-09	-101040	<i>MGAT4C</i>	protein-coding
chr19	54336241	54340020	-0.77	1.33E-06	-10474	<i>NLRP12</i>	protein-coding
chr2	190042261	190042860	-0.77	6.19E-09	2044	<i>COL5A2</i>	protein-coding
chr3	39619501	39621480	-0.77	3.45E-09	76934	<i>MOBP</i>	protein-coding
chr9	97812001	97812840	-0.77	9.39E-08	-15211	<i>MIR6081</i>	ncRNA
chr1	18807541	18808260	-0.76	7.56E-09	477	<i>KLHDC7A</i>	protein-coding
chr1	222139201	222146760	-0.76	8.98E-09	-227465	<i>DUSP10</i>	protein-coding
chr11	94373101	94374300	-0.76	8.67E-09	73227	<i>PIWILA</i>	protein-coding
chr2	28617061	28618020	-0.76	5.01E-08	-2	<i>FLJ31356</i>	ncRNA
chr2	62693461	62694060	-0.76	4.82E-06	39843	<i>TMEM17</i>	protein-coding
chr2	173421061	173424420	-0.76	9.91E-09	2044	<i>PDK1</i>	protein-coding
chr2	183914161	183915180	-0.76	3.28E-09	-11085	<i>NCKAP1</i>	protein-coding
chr2	207672121	207676080	-0.76	1.32E-08	-26069	<i>MIR3130-1</i>	ncRNA
chr3	38115781	38118000	-0.76	3.28E-09	36195	<i>DLEC1</i>	protein-coding
chr4	2813881	2814540	-0.76	3.28E-09	265	<i>SH3BP2</i>	protein-coding
chr7	22766761	22768200	-0.76	3.28E-09	715	<i>IL6</i>	protein-coding
chr7	124899421	124904460	-0.76	1.09E-08	2404	<i>RP11-3B12.3</i>	ncRNA
chr18	54736981	54739680	-0.75	4.10E-09	1019	<i>LINC-ROR</i>	ncRNA
chr2	85804021	85806720	-0.75	4.12E-09	757	<i>VAMP8</i>	protein-coding
chr2	227657041	227658300	-0.75	1.52E-07	5835	<i>IRS1</i>	protein-coding
chr3	143518621	143520120	-0.75	7.49E-06	48002	<i>SLC9A9</i>	protein-coding
chr3	176916121	176916780	-0.75	2.26E-06	-1403	<i>TBL1XR1</i>	protein-coding
chr5	138907201	138908100	-0.75	4.03E-06	-33100	<i>UBE2D2</i>	protein-coding
chr7	123635101	123636180	-0.75	3.28E-09	37882	<i>TMEM229A</i>	protein-coding

Chapter 7: Appendices

Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr1	204680341	204681300	-0.74	3.28E-09	-26224	<i>LRRN2</i>	protein-coding
chr11	67121701	67122300	-0.74	9.69E-07	-934	<i>POLD4</i>	protein-coding
chr2	47795881	47796480	-0.74	4.30E-06	1289	<i>KCNK12</i>	protein-coding
chr20	38902981	38904240	-0.74	3.28E-09	270508	<i>RP11-101E14.2</i>	ncRNA
chr3	189870241	189870960	-0.74	1.30E-05	-30375	<i>LEPREL1</i>	protein-coding
chr6	77272141	77274060	-0.74	9.42E-07	-490706	<i>IMPG1</i>	protein-coding
chr1	209602621	209604120	-0.73	2.09E-05	1203	<i>MIR205HG</i>	protein-coding
chr13	56616661	56617260	-0.73	5.68E-07	868372	<i>MIR5007</i>	ncRNA
chr14	21173041	21174420	-0.73	1.16E-07	16795	<i>RNASE4</i>	protein-coding
chr2	82230421	82232820	-0.73	3.28E-09	-537551	<i>ACO12075.2</i>	ncRNA
chr20	45978241	45979020	-0.73	5.48E-07	7002	<i>ZMYND8</i>	protein-coding
chr3	141087421	141088080	-0.73	2.86E-07	44696	<i>ZBTB38</i>	protein-coding
chr4	120916381	120918240	-0.73	3.58E-05	70702	<i>MAD2L1</i>	protein-coding
chr8	115295821	115300800	-0.73	1.88E-07	-849069	<i>CSMD3</i>	protein-coding
chr9	102434221	102436200	-0.73	3.45E-09	146960	<i>RP11-554F20.1</i>	ncRNA
chrX	16594741	16595400	-0.73	3.28E-09	-73210	<i>S100G</i>	protein-coding
chr1	114572761	114578700	-0.72	1.29E-08	53718	<i>OLFML3</i>	protein-coding
chr1	223196701	223197720	-0.72	3.52E-08	119413	<i>TLR5</i>	protein-coding
chr12	51319141	51320520	-0.72	3.28E-09	1297	<i>METTL7A</i>	protein-coding
chr13	54846601	54847260	-0.72	1.22E-06	39252	<i>MIR1297</i>	ncRNA
chr19	6458341	6458940	-0.72	9.65E-09	1140	<i>SLC25A23</i>	protein-coding
chr2	116934361	116936520	-0.72	3.34E-09	1015757	<i>DPP10</i>	protein-coding
chr3	11679841	11680680	-0.72	5.53E-08	5190	<i>VGLL4</i>	protein-coding
chr4	126484561	126486420	-0.72	3.07E-08	57077	<i>MIR2054</i>	ncRNA
chr6	3164521	3166020	-0.72	3.28E-09	-7488	<i>TUBB2A</i>	protein-coding
chr8	129625921	129629820	-0.72	6.03E-09	465509	<i>MIR1208</i>	ncRNA
chr9	117877441	117880200	-0.72	6.47E-09	1715	<i>TNC</i>	protein-coding
chr9	134165221	134165940	-0.72	3.28E-09	500	<i>PPAPDC3</i>	protein-coding
chrX	84211081	84213120	-0.72	6.27E-09	22944	<i>UBE2DNL</i>	pseudo
chr11	61159141	61159740	-0.71	5.34E-06	-391	<i>TMEM216</i>	protein-coding
chr11	94376161	94385880	-0.71	1.99E-08	80547	<i>PIWIL4</i>	protein-coding
chr11	118588381	118589820	-0.71	3.45E-09	-38720	<i>TREH</i>	protein-coding
chr12	79935901	79940220	-0.71	3.85E-08	87410	<i>MIR5692B</i>	ncRNA
chr19	18415561	18416340	-0.71	3.28E-09	18050	<i>LSM4</i>	protein-coding
chr5	118284301	118288500	-0.71	7.36E-09	-23880	<i>MIR1244-1</i>	ncRNA
chr6	138864481	138865080	-0.71	3.28E-09	28887	<i>NHSL1</i>	protein-coding
chr7	109232221	109235280	-0.71	4.06E-09	366519	<i>EIF3IP1</i>	pseudo
chr9	96110221	96113940	-0.71	5.79E-09	-3385	<i>C9orf129</i>	protein-coding
chr1	113364541	113366880	-0.7	3.45E-09	27554	<i>RP11-426L16.8</i>	ncRNA
chr1	156629941	156631380	-0.7	3.28E-09	16528	<i>NES</i>	protein-coding
chr12	124872601	124873440	-0.7	1.51E-06	-51233	<i>MIR6880</i>	ncRNA
chr2	188417821	188418540	-0.7	3.28E-09	1038	<i>TPPI</i>	protein-coding
chr2	232485841	232486440	-0.7	3.28E-06	28566	<i>C2orf57</i>	protein-coding
chr4	189027781	189029280	-0.7	3.28E-09	-2123	<i>TRIML2</i>	protein-coding
chr1	45274141	45275160	-0.69	3.28E-09	497	<i>BTBD19</i>	protein-coding
chr1	163170361	163170960	-0.69	3.28E-09	2302	<i>RGS5</i>	protein-coding
chr12	47617501	47619060	-0.69	3.28E-09	-8055	<i>PCED1B-AS1</i>	ncRNA
chr12	120700861	120702960	-0.69	3.65E-09	1663	<i>PXN</i>	protein-coding
chr19	1508401	1509120	-0.69	8.35E-07	4427	<i>ADAMTSL5</i>	protein-coding
chr2	52564921	52568460	-0.69	4.06E-09	-1307017	<i>NRXN1</i>	protein-coding
chr2	227658601	227660640	-0.69	3.28E-09	3885	<i>IRS1</i>	protein-coding
chr20	62129341	62129940	-0.69	4.91E-09	1027	<i>EEF1A2</i>	protein-coding
chr4	16921021	16921920	-0.69	9.89E-08	-21047	<i>LDB2</i>	protein-coding
chr4	146620201	146626680	-0.69	1.60E-07	22085	<i>C4orf51</i>	protein-coding
chr5	1009621	1010640	-0.69	3.92E-07	1054	<i>NKD2</i>	protein-coding
chr6	31139701	31141980	-0.69	3.45E-09	-2371	<i>POU5F1</i>	protein-coding
chr6	111925201	111926220	-0.69	3.07E-07	1766	<i>TRAF3IP2</i>	protein-coding
chr6	126024721	126028800	-0.69	6.51E-09	31262	<i>RP1-293L8.2</i>	ncRNA
chr6	144327421	144328500	-0.69	3.28E-09	1580	<i>PLAGL1</i>	protein-coding
chr7	125560261	125562540	-0.69	3.86E-07	-657056	<i>RP11-3B12.3</i>	ncRNA
chr8	30254101	30254820	-0.69	3.28E-09	-11544	<i>RBPMS-AS1</i>	ncRNA
chrX	90029461	90030900	-0.69	1.28E-07	-659416	<i>PABPC5</i>	protein-coding
chr12	14858941	14861400	-0.68	6.76E-08	-10652	<i>GUCY2C</i>	protein-coding
chr14	38664541	38665140	-0.68	4.53E-06	-12363	<i>SSTR1</i>	protein-coding
chr14	41986381	41991360	-0.68	7.94E-09	-87893	<i>LRFN5</i>	protein-coding
chr18	40317121	40322100	-0.68	8.97E-08	376046	<i>RIT2</i>	protein-coding

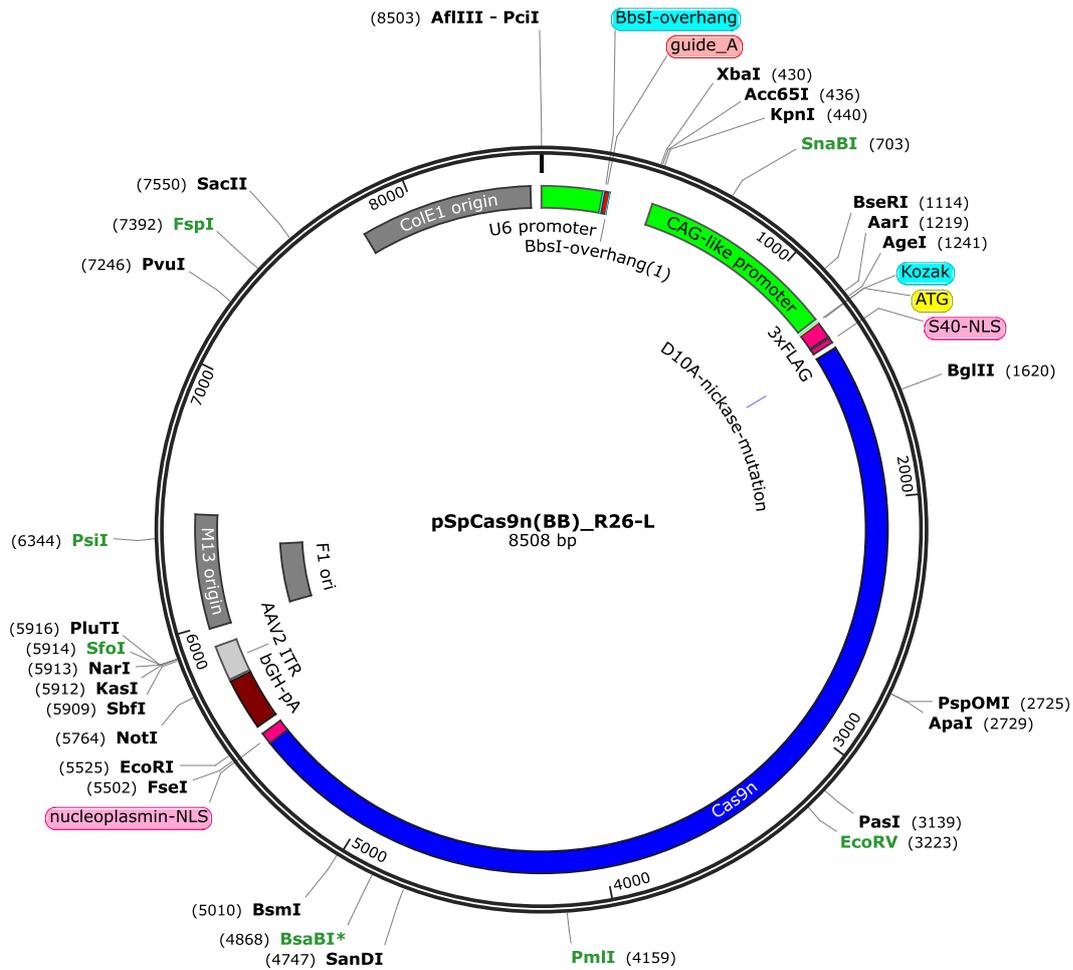
Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

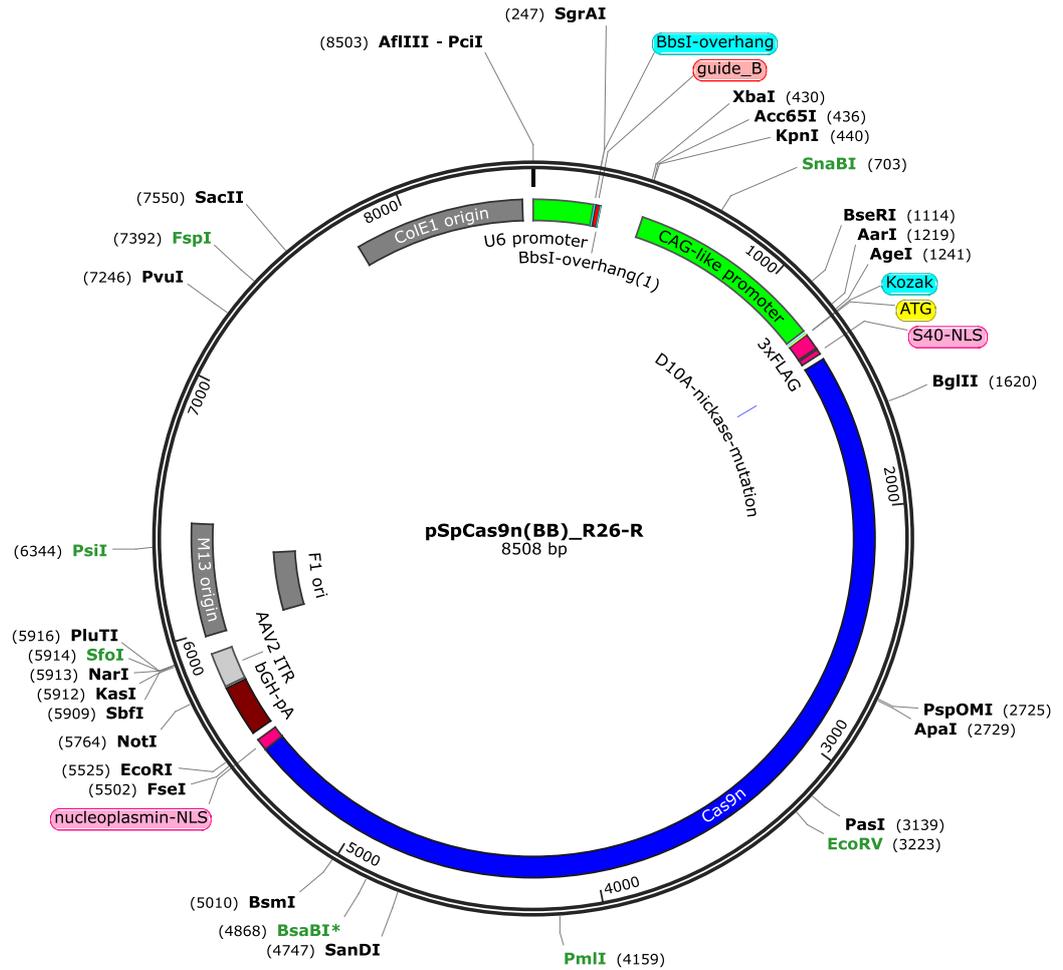
Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr18	40322881	40323900	-0.68	1.56E-07	372266	RIT2	protein-coding
chr19	47924281	47925060	-0.68	7.21E-08	-1886	MEIS3	protein-coding
chr5	27725401	27726240	-0.68	3.28E-09	253422	LINC01021	ncRNA
chr5	126175441	126181080	-0.68	9.21E-09	65416	LMNB1	protein-coding
chr6	110499001	110499600	-0.68	3.71E-06	1906	WASF1	protein-coding
chr6	131618221	131620920	-0.68	5.71E-09	48272	AKAP7	protein-coding
chrX	16194241	16201860	-0.68	1.92E-08	12447	MAGEB17	protein-coding
chr12	75761041	75764640	-0.67	9.45E-09	21861	CAPS2	protein-coding
chr16	70652221	70655820	-0.67	7.43E-09	-3874	IL34	protein-coding
chr18	40856221	40856940	-0.67	8.06E-06	1034	SYT4	protein-coding
chr19	16190041	16190700	-0.67	1.54E-05	3236	TPM4	protein-coding
chr2	188536741	188540820	-0.67	4.60E-09	-119562	TFPI	protein-coding
chr3	152555761	152556900	-0.67	2.02E-07	3595	P2RY1	protein-coding
chr4	17005141	17006460	-0.67	3.28E-09	-105377	LDB2	protein-coding
chr5	12627061	12627840	-0.67	5.96E-09	52482	LINC01194	ncRNA
chrX	68369941	68370960	-0.67	3.60E-05	14914	PJA1	protein-coding
chr12	7996261	7997220	-0.66	1.07E-08	28894	SLC2A14	protein-coding
chr13	54840061	54842520	-0.66	1.45E-08	44892	MIR1297	ncRNA
chr13	78487321	78490620	-0.66	9.68E-09	3995	EDNRB	protein-coding
chr15	72519241	72520920	-0.66	3.28E-09	1982	PKM	protein-coding
chr5	76100761	76101960	-0.66	3.28E-09	-13472	F2RL1	protein-coding
chr1	81716641	81717960	-0.65	3.28E-09	-548781	LPHN2	protein-coding
chr12	65384701	65386560	-0.65	3.45E-09	108077	FLJ41278	ncRNA
chr13	56150401	56152140	-0.65	3.63E-09	402682	MIR5007	ncRNA
chr13	80909941	80910600	-0.65	2.66E-06	4815	SPRY2	protein-coding
chr16	11410201	11411640	-0.65	5.05E-08	-28390	RMI2	protein-coding
chr19	47925601	47926740	-0.65	3.28E-09	-3386	MEIS3	protein-coding
chr2	64484581	64485960	-0.65	1.70E-06	-52652	LINC00309	ncRNA
chr20	38900821	38902620	-0.65	2.90E-08	268618	RP11-101E14.2	ncRNA
chr20	44390641	44393760	-0.65	3.85E-07	28346	WFDC3	protein-coding
chr5	147247141	147247740	-0.65	3.38E-06	-10833	SCGB3A2	protein-coding
chrX	16595521	16598340	-0.65	3.28E-09	-71350	S100G	protein-coding
chr10	57658141	57658860	-0.64	3.28E-09	299751	MTRNR2L5	protein-coding
chr11	311101	320700	-0.64	5.20E-08	1910	IFITM1	protein-coding
chr11	811021	811680	-0.64	3.28E-09	-330	SNORA52	snoRNA
chr12	109530541	109531140	-0.64	3.28E-09	452	ALKBH2	protein-coding
chr13	40174021	40175160	-0.64	3.28E-09	2765	LHFP	protein-coding
chr13	51749101	51749940	-0.64	3.16E-06	-2997	LINC00371	ncRNA
chr17	74607601	74608440	-0.64	1.17E-06	-25876	ST6GALNAC2	protein-coding
chr22	40391281	40392300	-0.64	7.15E-07	838	FAM83F	protein-coding
chr3	46617781	46618740	-0.64	3.28E-09	-815	TDGF1	protein-coding
chr4	87395221	87401520	-0.64	1.15E-08	-24088	MAPK10	protein-coding
chr7	22743421	22749000	-0.64	6.48E-08	-20555	IL6	protein-coding
chr7	100771441	100772340	-0.64	2.09E-06	1521	SERPINE1	protein-coding
chr8	72582361	72584880	-0.64	3.45E-09	-171737	LOC100132891	ncRNA
chr9	90026161	90031080	-0.64	3.01E-08	-83522	DAPK1	protein-coding
chrX	103015141	103021140	-0.64	2.14E-08	-13298	PLP1	protein-coding
chr1	24477961	24479100	-0.63	8.17E-07	-8756	IL22RA1	protein-coding
chr11	118587181	118587780	-0.63	5.39E-08	-37100	TREH	protein-coding
chr16	65265301	65265900	-0.63	5.05E-06	-109682	CDH11	protein-coding
chr17	27369721	27370980	-0.63	4.32E-08	433	PIPOX	protein-coding
chr18	22005301	22006140	-0.63	3.28E-09	-888	IMPACT	protein-coding
chr22	17745241	17747820	-0.63	4.07E-09	1092	CECR3	ncRNA
chr7	137927341	137928180	-0.63	5.35E-06	119257	MIR4468	ncRNA
chr8	137632081	137633580	-0.63	3.45E-09	1163115	KHDRBS3	protein-coding
chr9	76702021	76703520	-0.63	2.12E-08	-409481	RORB	protein-coding
chr9	118234201	118235280	-0.63	3.28E-09	271777	LOC101928775	ncRNA
chrX	2878321	2882820	-0.63	9.35E-09	1923	ARSE	protein-coding
chrX	124523281	124527180	-0.63	8.57E-08	71262	LOC100129520	protein-coding
chr15	83479981	83480700	-0.62	3.28E-09	1961	WHAMM	protein-coding
chr16	85364041	85364700	-0.62	3.28E-09	-24440	MIR5093	ncRNA
chr20	45985501	45986700	-0.62	3.28E-09	-468	ZMYND8	protein-coding
chr22	17741941	17743080	-0.62	3.28E-09	5112	CECR3	ncRNA
chr5	172710301	172710960	-0.62	3.28E-09	45875	STC2	protein-coding
chr8	1772821	1774680	-0.62	3.45E-09	1602	ARHGEF10	protein-coding
chr8	133096621	133097580	-0.62	1.33E-06	20411	HHLA1	protein-coding

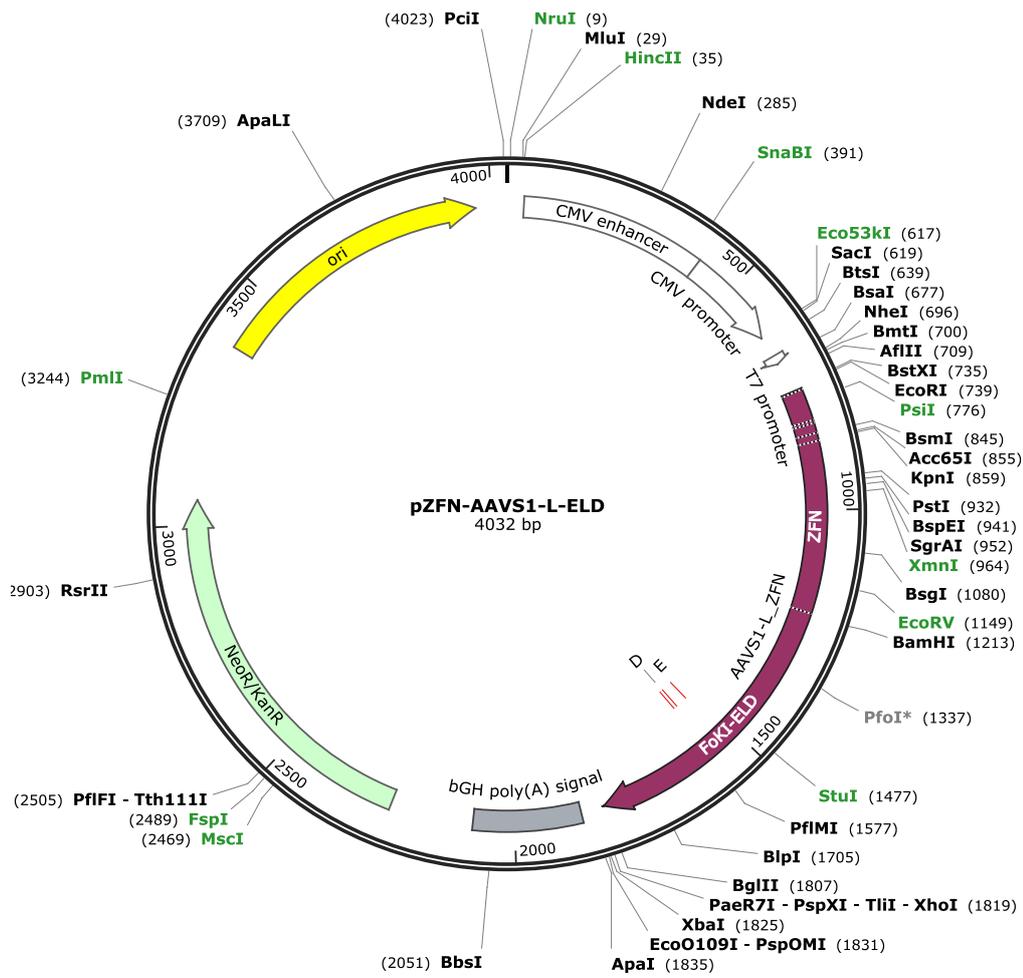
Chapter 7: Appendices

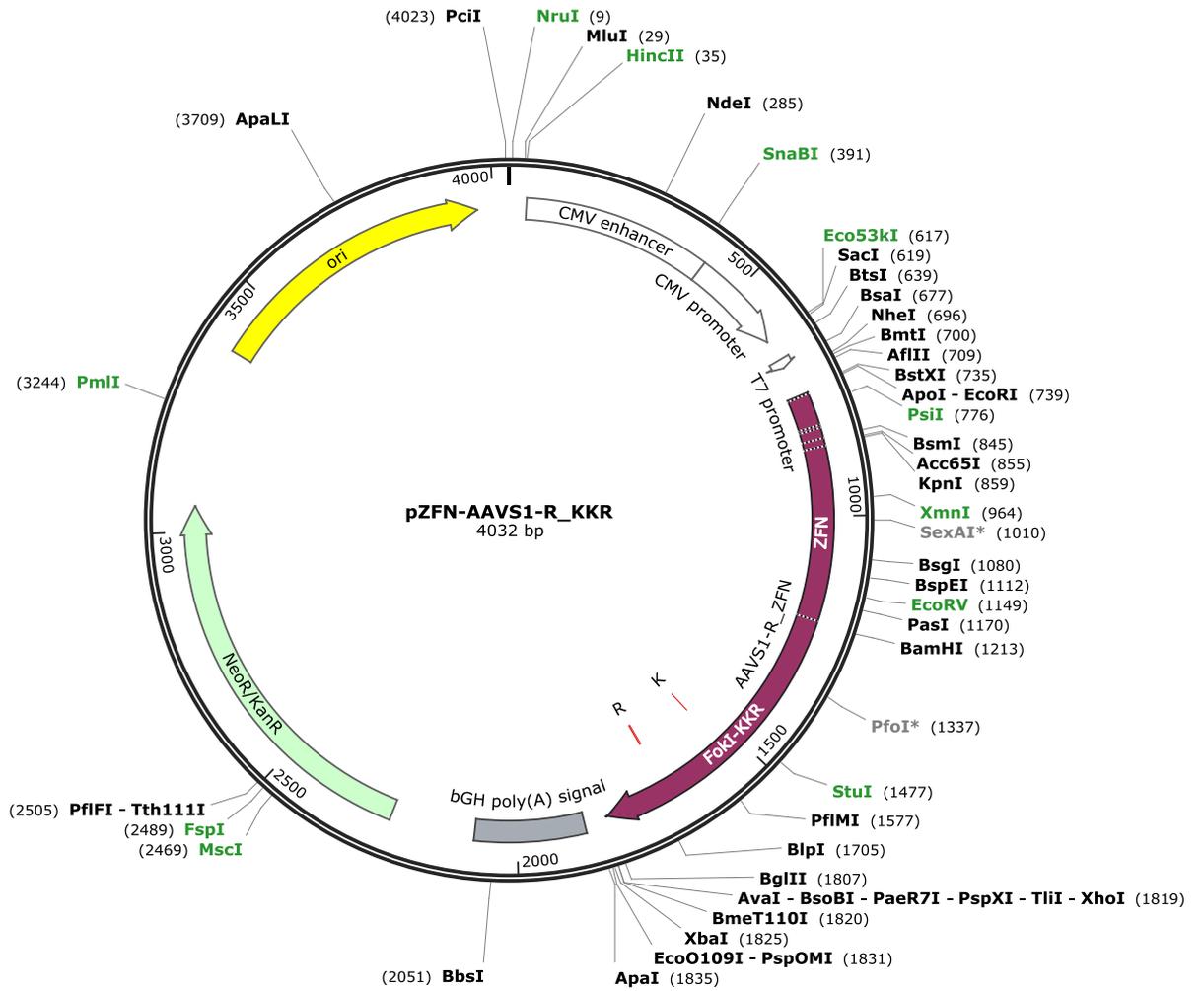
Chr.	Start	End	log ₂ FC	adj.p	Distance to TSS	Gene name	Gene type
chr9	136399501	136400700	-0.62	1.43E-06	126	<i>ADAMTSL2</i>	protein-coding
chrX	113837161	113839560	-0.62	2.89E-07	19810	<i>HTR2C</i>	protein-coding
chr19	17436841	17438220	-0.61	3.42E-07	8107	<i>ANO8</i>	protein-coding
chr2	8012161	8013000	-0.61	2.43E-07	27205	<i>AC007463.2</i>	ncRNA
chr2	38307661	38313360	-0.61	1.04E-08	-7188	<i>CYP1B1</i>	protein-coding
chr2	231732361	231733380	-0.61	2.93E-08	3206	<i>ITM2C</i>	protein-coding
chr20	48600541	48601260	-0.61	1.95E-07	1388	<i>SNAI1</i>	protein-coding
chr3	109051201	109051800	-0.61	2.68E-06	4918	<i>DPPA4</i>	protein-coding
chr3	143517601	143518560	-0.61	7.51E-08	49292	<i>SLC9A9</i>	protein-coding
chr4	4047001	4050480	-0.61	5.97E-09	-91593	<i>FAM86EP</i>	pseudo
chr6	169651681	169654320	-0.61	4.06E-09	1208	<i>THBS2</i>	protein-coding
chr7	20162281	20165760	-0.61	1.86E-08	16028	<i>AC005062.2</i>	ncRNA
chr8	8747281	8748000	-0.61	2.91E-06	3490	<i>MFHAS1</i>	protein-coding
chr8	142402921	142403520	-0.61	8.68E-09	-25856	<i>GPR20</i>	protein-coding
chr1	7844401	7845300	-0.6	6.55E-08	88	<i>PER3</i>	protein-coding
chr2	108995041	108995880	-0.6	2.75E-07	1040	<i>SULT1C4</i>	protein-coding
chr20	50414641	50415480	-0.6	2.84E-08	3987	<i>SALL4</i>	protein-coding
chr3	179244601	179245740	-0.6	3.28E-09	-35537	<i>ACTL6A</i>	protein-coding
chr4	95374441	95375280	-0.6	1.09E-08	1853	<i>PDLIM5</i>	protein-coding
chr4	135121921	135122520	-0.6	1.02E-08	682	<i>PABPC4L</i>	protein-coding
chr5	12662941	12666300	-0.6	5.20E-09	89652	<i>LINC01194</i>	ncRNA
chr6	131615281	131616360	-0.6	7.92E-07	44522	<i>AKAP7</i>	protein-coding
chrX	153602101	153602760	-0.6	7.32E-06	575	<i>FLNA</i>	protein-coding
chr1	8935321	8936520	-0.59	1.88E-08	-2973	<i>ENO1-AS1</i>	ncRNA
chr1	16493341	16493940	-0.59	1.39E-06	-11059	<i>EPHA2</i>	protein-coding
chr1	77334961	77335620	-0.59	3.28E-09	2105	<i>ST6GALNAC5</i>	protein-coding
chr1	150667921	150668640	-0.59	1.69E-08	1391	<i>GOLPH3L</i>	protein-coding
chr11	77183101	77184000	-0.59	3.06E-06	1557	<i>PAK1</i>	protein-coding
chr12	14863621	14864640	-0.59	2.10E-06	-14612	<i>GUCY2C</i>	protein-coding
chr12	86267641	86271060	-0.59	5.81E-09	1278	<i>NTS</i>	protein-coding
chr12	122020981	122021580	-0.59	3.28E-09	-2361	<i>KDM2B</i>	protein-coding
chr19	17516581	17517720	-0.59	3.28E-09	-694	<i>BST2</i>	protein-coding
chr3	69061021	69062280	-0.59	3.42E-09	1394	<i>EOGT</i>	protein-coding
chr5	124075741	124076820	-0.59	3.28E-09	4524	<i>ZNF608</i>	protein-coding
chr7	134463421	134467200	-0.59	5.33E-09	1147	<i>CALD1</i>	protein-coding
chr7	148843261	148844100	-0.59	1.22E-06	-879	<i>ZNF398</i>	protein-coding
chr9	85500421	85503660	-0.59	4.95E-09	176002	<i>RASEF</i>	protein-coding

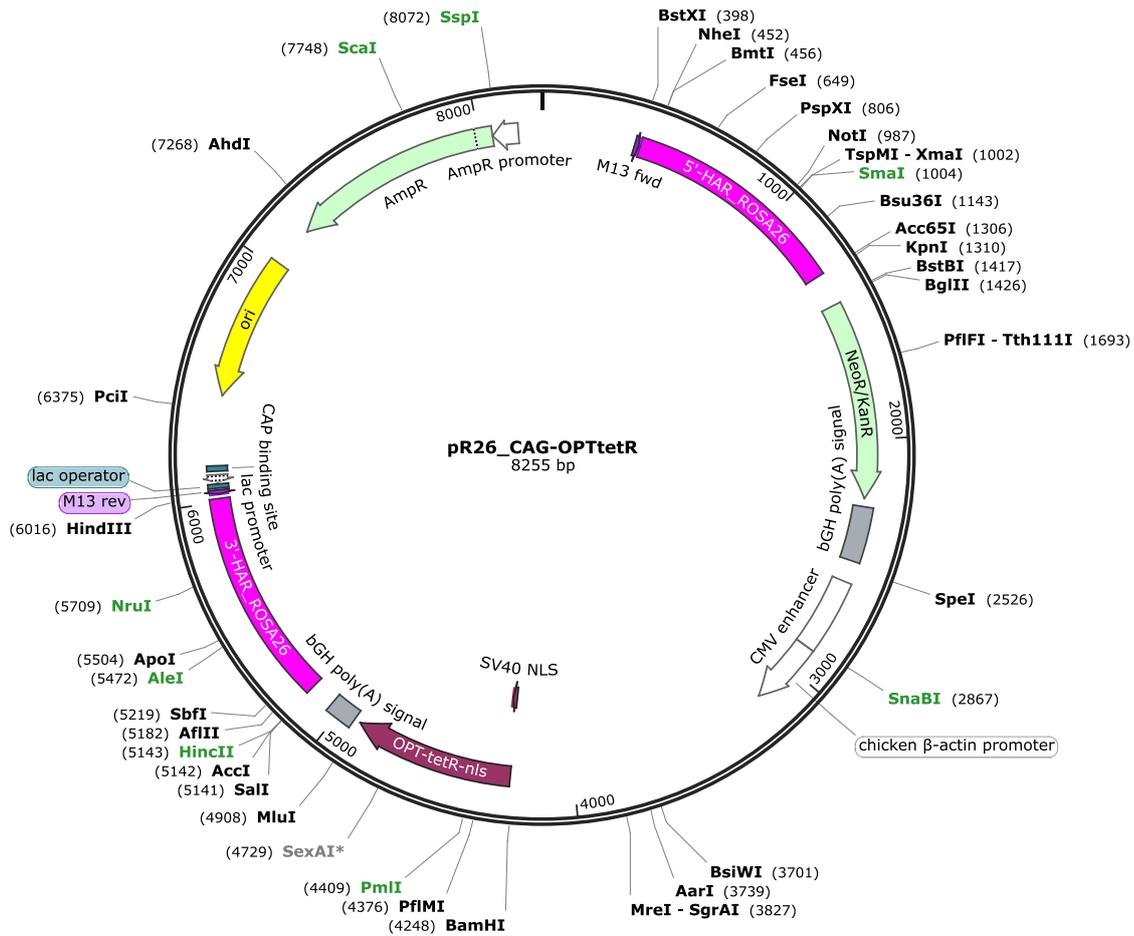
APPENDIX II: PLASMID MAPS FOR THE OPTIKD AND SOPTIKD SYSTEMS

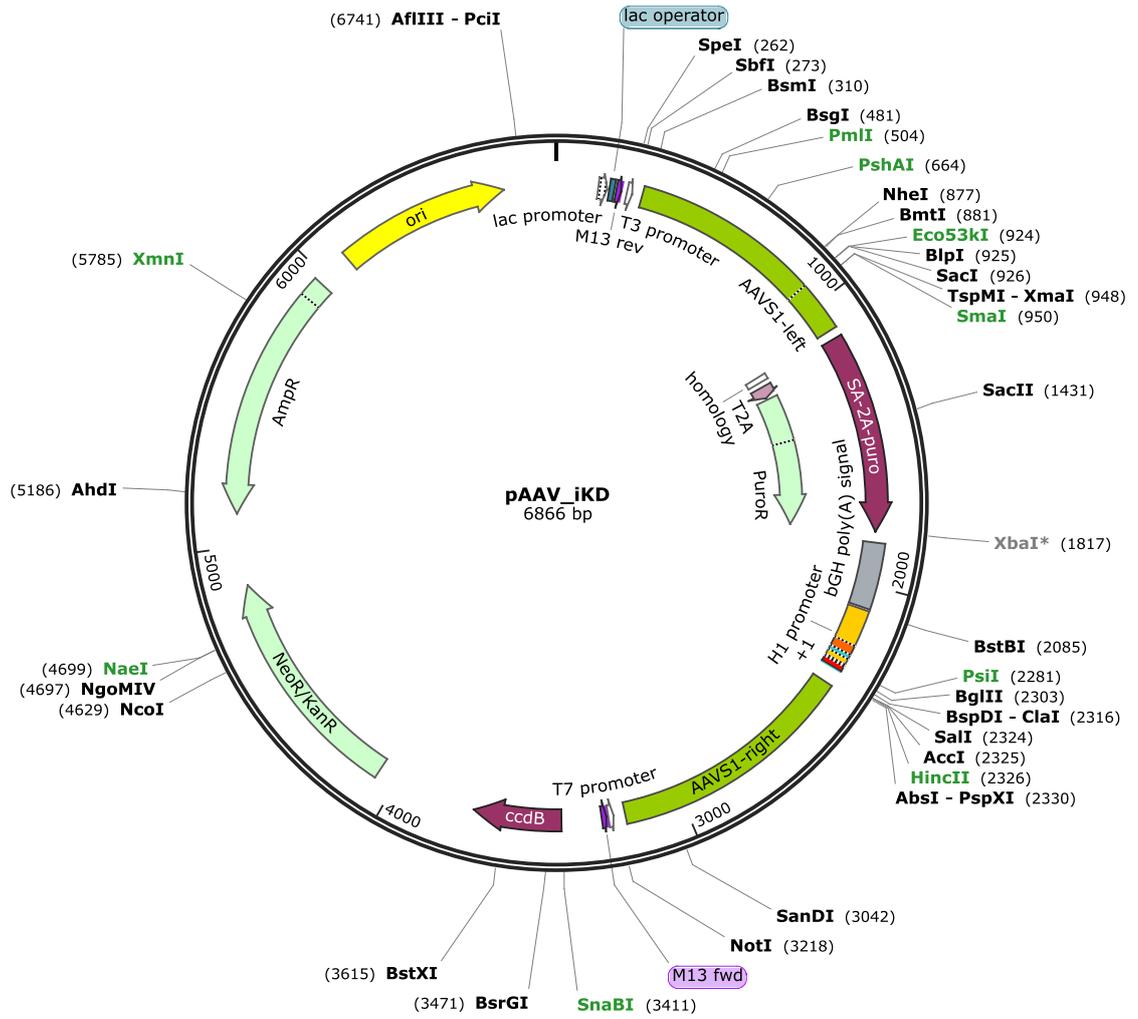


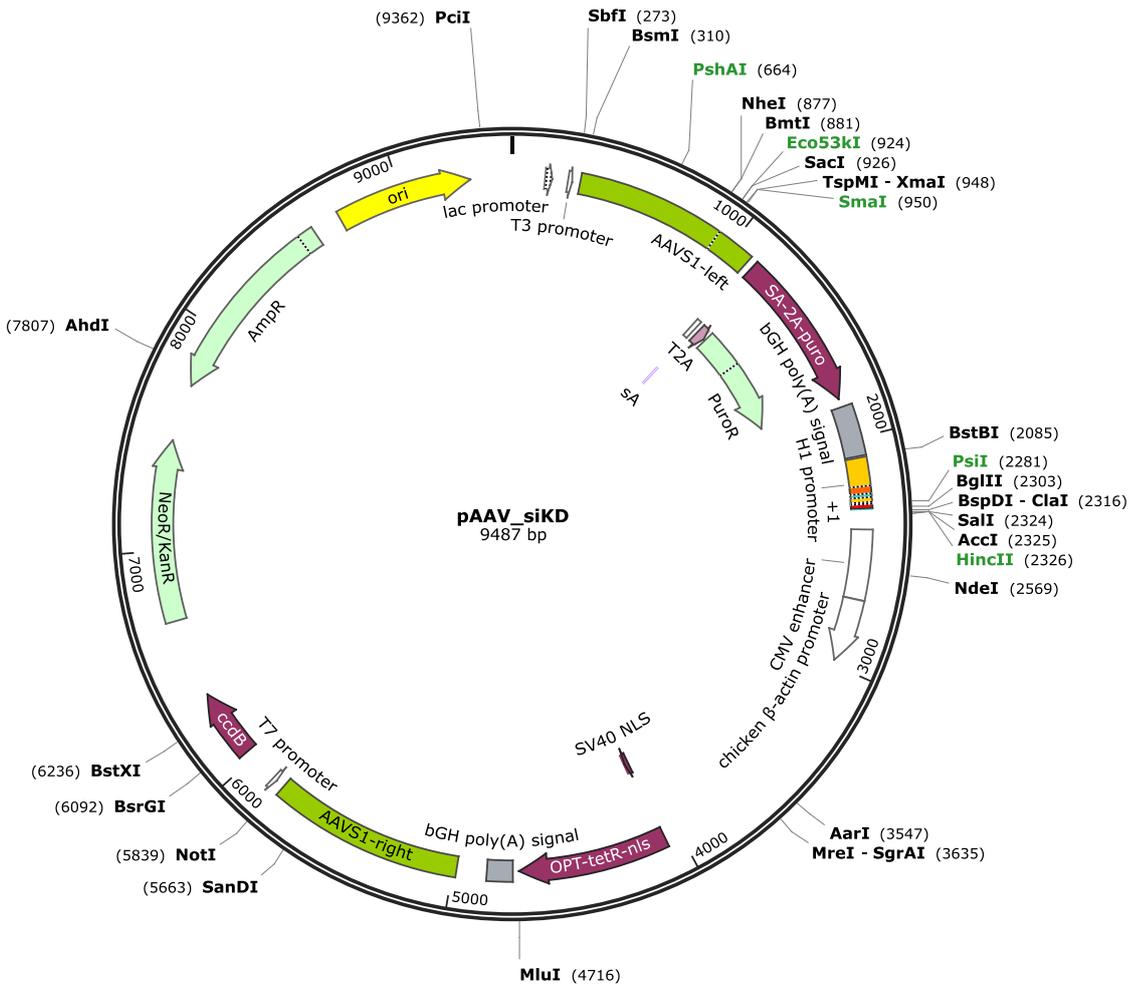












APPENDIX III: SMAD2/3 INTERACTING PROTEINS IDENTIFIED BY QUANTITATIVE LABEL-FREE MASS SPECTROMETRY

Protein	Uniprot ID	Main known function	Sign. pluri	Sign. endo	p-val pluri	p-val endo	log enrich. pluri	log enrich. endo
ATF7	P17544	Transcription	+		1.8	2.0	0.5	0.2
C1D	Q13901	Transcription	+		1.7	0.6	0.7	0.4
CREB1	P16220	Transcription		+	0.8	0.9	0.1	0.3
CSDA	P16989	Transcription	+		2.0	0.9	0.6	0.3
ETV6	P41212	Transcription repression	+	+	2.9	2.0	0.7	0.6
FOXH1	O75593	Transcription activation		+	0.1	1.9	-0.2	0.4
FOXJ2	Q9P0K8	Transcription activation	+		0.9	1.3	0.4	-0.1
GZF1	Q9H116	Transcription repression	+		1.4	0.8	0.7	0.1
IRF2BP2	Q7Z5L9	Transcription repression	+		2.3	1.6	0.8	0.2
KCTD1	Q719H9	Transcription repression		+	1.4	1.4	0.9	0.5
KCTD15	Q96SI1	Transcription	+	+	3.4	2.2	2.3	2.3
NFAT5	O94916	Transcription	+	+	2.6	1.4	1.4	0.5
SATB2	Q9UPW6	Transcription	+	+	2.6	1.9	0.6	0.4
SKI	P12755	Transcription repression	+	+	2.9	4.5	1.1	1.3
SKIL	P12757	Transcription repression	+	+	3.1	3.6	1.2	1.4
SMAD4	Q13485	Transcription	+	+	3.2	2.9	0.7	0.9
SOX13	Q9UN79	Transcription	+	+	2.7	2.4	1.2	1.1
ATF7IP	Q6VMQ6	Epigenetic regulation	+	+	2.7	3.0	0.9	1.0
CITED2	Q99967	Transcriptional coactivator	+		1.7	0.2	0.3	-0.1
CREBBP	Q92793	Histone acetyltransferase		+	1.8	1.6	0.3	0.3
EP300	Q09472	Histone acetyltransferase		+	2.6	2.4	0.4	0.4
HIST1H1E	P10412	Histone variant		+	0.2	1.2	-0.1	0.6
HMG20A	Q9NP66	Chromatin associated protein	+		1.3	0.0	0.6	-0.2
NASP	P49321	Histone 1 chaperone	+		2.5	1.2	0.5	0.2
NCOA6	Q14686	Transcriptional coactivator	+		1.5	0.1	0.7	-0.2
SETDB1	Q15047	H3K9 methyltransferase	+	+	2.7	2.5	0.9	0.9
SMCHD1	A6NHR9	X-inactivation	+	+	2.9	2.4	0.7	0.6
CPSF6	Q16630	mRNA processing	+	+	2.7	2.5	0.6	0.5
DDX19B	Q9UMR2	mRNA export	+		0.8	0.3	0.7	0.0
NONO	Q15233	mRNA splicing	+	+	2.7	2.5	0.5	0.4
NUDT21	O43809	mRNA processing		+	2.4	2.7	0.5	0.4
PSPC1	Q8WXF1	mRNA splicing		+	2.5	2.5	0.4	0.4
SFPQ	P23246	mRNA splicing		+	2.6	2.6	0.4	0.4
SYF2	O95926	mRNA splicing	+		1.7	0.2	0.5	0.0
KIAA1429	Q69YN4	RNA methylation	+	+	3.7	2.7	0.7	0.5
LSM1	O15116	RNA degradation		+	0.0	0.9	-0.7	0.4
METTL14	Q9HCE5	RNA methylation	+	+	3.0	1.9	0.7	0.5
METTL3	Q86U44	RNA methylation	+	+	2.8	2.6	0.7	0.8
PAN2	Q504Q3	RNA degradation	+	+	2.1	3.2	0.6	0.7
PAN3	Q58A45	RNA degradation	+	+	1.9	2.6	0.5	0.4
PARN	O95453	RNA degradation	+	+	2.6	2.6	0.9	0.9
TNRC6C	Q9HCJ0	miRNA-mediated silencing	+	+	2.2	1.5	1.1	0.4

Protein	Uniprot ID	Main known function	Sign. pluri	Sign. endo	p-val pluri	p-val endo	log enrich. pluri	log enrich. endo
TRMT112	Q9UI30	tRNA methylation		+	0.8	1.9	0.1	0.6
WTAP	Q15007	RNA methylation	+	+	2.9	3.0	0.6	0.5
ERCC1	P07992	DNA repair	+	+	2.1	3.2	0.6	0.8
ERCC4	Q92889	DNA repair	+	+	2.5	2.3	0.7	0.6
SLX4	Q8IY92	DNA repair		+	2.3	3.0	0.4	0.4
SLX4IP	Q5VYV7	DNA repair		+	2.3	2.9	0.4	0.5
AKAP13	Q12802	PKA signalling	+	+	2.7	2.1	0.6	0.4
BSG	P35613	Transporter		+	0.6	0.9	0.0	0.4
CBLL1	Q75N03	Ubiquitination	+	+	2.2	2.5	0.6	0.5
CLPTM1	O96005	Unknown	+		1.9	0.7	0.8	0.4
DAPK3	O43293	Protein kinase		+	2.0	1.7	0.4	0.5
DUSP5	Q16690	ERK phosphatase		+	1.1	1.1	-0.3	0.0
OTUD4	Q01804	Deubiquitination	+	+	1.9	3.1	0.8	1.0
PAWR	Q96IZ0	Pro-apoptotic factor	+		1.5	0.2	0.6	-0.1
RASAL2	Q9UJF2	RAS signalling		+	1.7	2.6	0.4	0.4
RNF220	Q5VTB9	E3 Ubiquitin ligase		+	0.9	1.5	0.2	0.4
SGOL2	Q562F6	Cytokinesis	+	+	2.3	3.2	0.5	0.5
SGTA	O43765	Chaperone	+	+	1.8	1.9	0.6	0.5
SRCIN1	Q9C0H9	SRC inhibition	+	+	2.6	1.3	0.8	0.6
UBL5	Q9BZL1	Unknown	+		1.8	0.5	0.5	0.0
WWP1	Q9H0M0	E3 Ubiquitin ligase	+		2.2	1.5	0.6	0.4
WWP2	O00308	E3 Ubiquitin ligase	+	+	2.7	1.5	1.6	0.7
ARL14EP	Q8N8R7	Vesicle transport	+	+	2.4	2.1	1.0	0.6
CHMP1B	Q7LBR1	Vesicle transport	+		2.6	0.7	0.7	0.4
KIAA1279	Q96EK5	Cytoskeletal dynamics	+	+	2.5	2.4	1.0	1.1
KIF18A	Q8NI77	Cytokinesis	+	+	3.2	2.6	1.5	1.7
PLS3	P13797	Cytoskeletal dynamics	+		1.7	0.8	0.3	0.0
SYNE1	Q8NF91	Vesicle transport		+	0.1	0.7	0.0	0.5
ZMYM4	Q5VZL5	Cytoskeletal dynamics	+	+	2.8	2.4	0.5	0.3
ALG13	Q9NP73	Glycosylation	+		1.6	0.3	0.7	-0.3
EXTL3	O43909	Glycosylation	+		1.1	0.3	0.6	-0.1
PNPLA6	Q8IY17	Phospholipase		+	0.1	1.9	0.2	0.3
SLC38A10	Q9HBR0	Ion transport	+		1.7	0.1	0.5	-0.1
SLC7A5	Q01650	Amino acid transport	+		1.3	0.7	1.0	0.0
TM7SF2	O76062	Cholesterol metabolism		+	0.1	2.1	0.0	0.4
UBIAD1	Q9Y5Z9	Vitamin metabolism	+		1.2	0.3	0.8	0.0
NSA2	O95478	Ribosome biogenesis	+	+	2.4	0.6	2.0	0.6
RPL22L1	Q6P5R6	Ribosomal protein		+	0.2	0.9	-0.3	0.1
RPS15A	P662244	Ribosomal protein	+		1.0	0.4	0.7	0.0
BTBD10	Q9BSF8	Unknown	+		1.9	0.8	0.5	0.0
LUZP1	Q86V48	Unknown	+	+	2.7	2.0	0.5	0.6
RNF175	Q8N4F7	Unknown		+	1.2	1.7	0.2	0.3
RNF219	Q5W0B1	Unknown	+	+	2.6	2.0	0.5	0.5
SS18L2	Q9UHA2	Unknown	+	+	1.7	1.4	0.7	0.5
TTC1	Q99614	Unknown	+		1.3	0.1	0.5	-0.3

APPENDIX IV: M6A PEAKS DECREASED AFTER INHIBITION OF ACTIVIN/NODAL SIGNALLING FOR 2 HOURS WITH SB

Chr.	Start	End	log2FC	logFDR	Gene name	Gene type
7	41700562	41703058	-4.81	-4.34	<i>INHBA</i>	protein coding
8	140098487	140098588	-3.08	-1.64	<i>PEG13</i>	sense intronic
2	46512163	46512512	-3.03	-1.43	<i>ATP6V1E2</i>	protein coding
17	77215750	77216151	-2.99	-2.86	<i>SEC14L1</i>	protein coding
21	26923782	26923982	-2.79	-1.87	<i>ADAMTSS5</i>	protein coding
3	39188879	39188980	-2.61	-1.35	<i>XIRP1</i>	protein coding
13	36879515	36879815	-2.59	-1.39	<i>SMAD9</i>	protein coding
5	76132007	76132107	-2.55	-2.72	<i>SV2C</i>	protein coding
1	228060961	228061061	-2.54	-2.1	<i>WNT3A</i>	protein coding
9	12821585	12821784	-2.46	-1.36	<i>LURAP1L</i>	protein coding
1	11766416	11766832	-2.42	-1.4	<i>C1orf167</i>	protein coding
17	70176114	70176513	-2.41	-1.93	<i>KCNJ2</i>	protein coding
20	51523533	51523634	-2.37	-3.02	<i>NFATC2</i>	protein coding
21	34888555	34888656	-2.37	-2.41	<i>RUNX1</i>	protein coding
1	944253	944404	-2.32	-3.1	<i>NOC2L</i>	protein coding
7	151374464	151374564	-2.31	-1.33	<i>NUB1</i>	protein coding
1	944234	944383	-2.18	-4.73	<i>SAMD11</i>	protein coding
4	148435210	148435459	-2.12	-3.04	<i>NR3C2</i>	protein coding
15	79457220	79457321	-2.08	-1.6	<i>KIAA1024</i>	protein coding
7	132129071	132129271	-2.08	-4.29	<i>PLXNA4</i>	protein coding
5	76131757	76131858	-2.06	-4.15	<i>SV2C</i>	protein coding
X	47512797	47512945	-2.03	-3.42	<i>NUS1P1</i>	processed pseudogene
1	6104939	6105239	-1.95	-4.58	<i>CHD5</i>	protein coding
19	52028358	52028558	-1.94	-2.44	<i>ZNF614</i>	protein coding
19	54216725	54217025	-1.92	-2.05	<i>LILRB3</i>	protein coding
19	31278672	31278773	-1.91	-2.59	<i>TSHZ3</i>	protein coding
11	18701790	18702290	-1.85	-1.93	<i>TMEM86A</i>	protein coding
16	88709498	88709647	-1.69	-3.9	<i>RP5-1142A6.10</i>	antisense
21	26924130	26924231	-1.64	-2.39	<i>ADAMTSS5</i>	protein coding
17	48726644	48726744	-1.64	-1.54	<i>HOXB13</i>	protein coding
7	124765428	124765528	-1.62	-1.78	<i>GPR37</i>	protein coding
11	1597613	1597905	-1.62	-4.36	<i>KRTAP5-2</i>	protein coding
11	1597594	1597893	-1.58	-3.18	<i>KRTAP5-AS1</i>	antisense
22	30661731	30661982	-1.57	-1.75	<i>DUSP18</i>	protein coding
21	26965368	26965468	-1.56	-2.03	<i>ADAMTSS5</i>	protein coding
10	46462418	46462516	-1.53	-1.55	<i>NPY4R</i>	protein coding
16	1079801	1080142	-1.5	-2.95	<i>SSTR5</i>	protein coding
X	155776280	155776680	-1.49	-1.89	<i>SPRY3</i>	protein coding
19	50486870	50487171	-1.47	-6.2	<i>EMC10</i>	protein coding
1	1090852	1090953	-1.45	-3.66	<i>C1orf159</i>	protein coding
7	155458877	155459027	-1.44	-2.55	<i>EN2</i>	protein coding
17	34626375	34626574	-1.42	-1.86	<i>TMEM132E</i>	protein coding
1	919681	919781	-1.39	-3.41	<i>RP11-5407.3</i>	lincRNA
10	35640984	35641283	-1.38	-3.71	<i>FZD8</i>	protein coding
12	132554152	132554252	-1.38	-2.5	<i>RP13-554M15.7</i>	TEC
17	77319851	77320052	-1.36	-4.53	<i>SEPT9</i>	protein coding
8	140096940	140097989	-1.35	-3.96	<i>PEG13</i>	sense intronic
15	68199960	68205362	-1.32	-1.36	<i>CALML4</i>	protein coding
8	96145092	96145339	-1.31	-2.76	<i>GDF6</i>	protein coding
19	31279468	31279717	-1.3	-2.5	<i>TSHZ3</i>	protein coding
11	61510771	61510969	-1.28	-2.07	<i>LRRC10B</i>	protein coding
11	89491359	89492165	-1.27	-4.93	<i>NOX4</i>	protein coding
X	100599071	100599170	-1.27	-1.5	<i>TNMD</i>	protein coding
20	63695376	63695477	-1.26	-1.43	<i>RTEL1</i>	protein coding

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Chr.	Start	End	log2FC	logFDR	Gene name	Gene type
19	49070546	49070746	-1.25	-3.26	<i>KCNA7</i>	protein_coding
1	2504681	2504831	-1.25	-2.77	<i>PLCH2</i>	protein_coding
1	16692135	16692235	-1.23	-1.61	<i>ESPNP</i>	transcribed_unprocessed_pseudogene
11	562313	562563	-1.22	-1.4	<i>RASSF7</i>	protein_coding
1	203165309	203165410	-1.21	-1.68	<i>ADORA1</i>	protein_coding
3	50288530	50288630	-1.21	-2.97	<i>IFRD2</i>	protein_coding
6	149972812	149973062	-1.21	-1.35	<i>ULBP1</i>	protein_coding
2	218893382	218893482	-1.2	-1.52	<i>WNT10A</i>	protein_coding
10	35640486	35640736	-1.18	-7.8	<i>FZD8</i>	protein_coding
19	31277927	31278027	-1.17	-2.4	<i>TSHZ3</i>	protein_coding
9	137079023	137079330	-1.17	-3.16	<i>UAP1L1</i>	protein_coding
18	22200880	22201130	-1.14	-1.8	<i>GATA6</i>	protein_coding
11	13010320	13010470	-1.14	-1.89	<i>RASSF10</i>	protein_coding
6	41684627	41684778	-1.13	-1.9	<i>TFEB</i>	protein_coding
19	49070844	49072196	-1.12	-2.99	<i>KCNA7</i>	protein_coding
2	104857643	104857892	-1.12	-3.94	<i>POU3F3</i>	protein_coding
3	48609745	48609846	-1.1	-2.5	<i>UQCRC1</i>	protein_coding
19	1775957	1777153	-1.09	-11.5	<i>ONECUT3</i>	protein_coding
9	22005326	22006119	-1.08	-8.78	<i>CDKN2B</i>	protein_coding
17	18411207	18411848	-1.07	-4.44	<i>FLJ35934</i>	lincRNA
11	64103260	64103653	-1.06	-1.46	<i>RP11-21A7A.2</i>	antisense
1	1082066	1082216	-1.05	-1.34	<i>C1orf159</i>	protein_coding
7	1541408	1541508	-1.02	-1.76	<i>MAFK</i>	protein_coding
7	155458577	155458678	-1.01	-2.13	<i>EN2</i>	protein_coding
17	12763364	12763615	-1.01	-1.83	<i>MYOCD</i>	protein_coding
1	156914044	156914989	-1.01	-2.87	<i>PEAR1</i>	protein_coding
19	46492926	46493473	-1	-1.92	<i>PNMAL2</i>	protein_coding
15	101075403	101075504	-0.994	-2.17	<i>LRRK1</i>	protein_coding
20	38213231	38213382	-0.98	-1.42	<i>KIAA1755</i>	protein_coding
X	40097337	40146502	-0.979	-1.99	<i>BCOR</i>	protein_coding
4	152975999	152976200	-0.979	-1.97	<i>FHDC1</i>	protein_coding
2	109792857	109792958	-0.975	-8.16	<i>RGPD5</i>	protein_coding
19	31277380	31277480	-0.973	-1.52	<i>TSHZ3</i>	protein_coding
6	32195792	32196043	-0.966	-2.27	<i>NOTCH4</i>	protein_coding
16	1348691	1348792	-0.957	-2.2	<i>BAIAP3</i>	protein_coding
18	33744744	33744944	-0.951	-1.79	<i>ASXL3</i>	protein_coding
6	139166969	139167268	-0.938	-1.34	<i>HECA</i>	protein_coding
X	38285695	38287244	-0.935	-3.87	<i>RPGR</i>	protein_coding
X	153795912	153796213	-0.922	-1.44	<i>SSR4</i>	protein_coding
15	77615897	77634290	-0.919	-1.77	<i>LINGO1</i>	protein_coding
11	18615221	18615321	-0.907	-1.33	<i>SPTY2D1</i>	protein_coding
1	10658532	10659758	-0.904	-2.49	<i>CASZ1</i>	protein_coding
X	153982257	153982406	-0.902	-1.64	<i>TMEM187</i>	protein_coding
7	195251	195627	-0.897	-4.09	<i>FAM20C</i>	protein_coding
16	66403740	66404487	-0.895	-2.33	<i>CDH5</i>	protein_coding
17	19021313	19021659	-0.895	-2.16	<i>GRAP</i>	protein_coding
1	41582124	41582324	-0.895	-5.28	<i>HIVEP3</i>	protein_coding
13	48412873	48415712	-0.895	-1.75	<i>LPAR6</i>	protein_coding
16	377403	377554	-0.882	-2.12	<i>TMEM8A</i>	protein_coding
11	64103287	64103738	-0.881	-2.44	<i>FLRT1</i>	protein_coding
3	50275706	50275806	-0.874	-2.79	<i>SEMA3B</i>	protein_coding
14	99174008	99174506	-0.872	-2.75	<i>BCL11B</i>	protein_coding
6	125759724	125760223	-0.866	-1.44	<i>HEY2</i>	protein_coding
11	119308677	119309027	-0.858	-2.86	<i>MCAM</i>	protein_coding
11	14643772	14644271	-0.855	-1.64	<i>PDE3B</i>	protein_coding
9	137164281	137164581	-0.85	-2.38	<i>GRIN1</i>	protein_coding
1	111755161	111755411	-0.847	-3.25	<i>FAM212B</i>	protein_coding
7	2541412	2541746	-0.841	-1.75	<i>BRAT1</i>	protein_coding
1	1087191	1087440	-0.841	-10.1	<i>C1orf159</i>	protein_coding
6	168308138	168308437	-0.839	-3.49	<i>DACT2</i>	protein_coding

Chapter 7: Appendices

Chr.	Start	End	log2FC	logFDR	Gene name	Gene type
10	7562963	7563214	-0.835	-1.39	<i>ITIH5</i>	protein_coding
1	10638401	10638501	-0.834	-1.5	<i>CASZ1</i>	protein_coding
1	29120821	29121120	-0.834	-3.06	<i>TMEM200B</i>	protein_coding
4	110617920	110618469	-0.833	-8.14	<i>PITX2</i>	protein_coding
18	7567581	7567732	-0.82	-1.48	<i>PTPRM</i>	protein_coding
19	616565	616665	-0.818	-1.53	<i>HCN2</i>	protein_coding
2	234834364	234834612	-0.815	-1.58	<i>ACO10148.1</i>	processed_transcript
8	22127569	22127720	-0.813	-1.83	<i>HR</i>	protein_coding
15	79457970	79463092	-0.811	-7.8	<i>KIAA1024</i>	protein_coding
11	61897399	61898237	-0.811	-2.23	<i>RAB3IL1</i>	protein_coding
19	45468341	45468491	-0.807	-8.28	<i>FOSB</i>	protein_coding
5	127457173	127457672	-0.807	-2.96	<i>MEGF10</i>	protein_coding
16	3054122	3054917	-0.804	-1.86	<i>MMP25-AS1</i>	antisense
11	58579160	58579310	-0.795	-1.32	<i>ZFP91</i>	protein_coding
20	58515030	58515131	-0.791	-1.88	<i>APCDD1L</i>	protein_coding
5	59038968	59039069	-0.777	-1.57	<i>PDE4D</i>	protein_coding
5	58974180	58975030	-0.771	-2.33	<i>PDE4D</i>	protein_coding
1	6249983	6250569	-0.767	-1.74	<i>GPR153</i>	protein_coding
19	5679277	5680106	-0.763	-1.38	<i>C19orf70</i>	protein_coding
8	22128626	22128727	-0.763	-1.35	<i>HR</i>	protein_coding
7	1540170	1540319	-0.759	-1.53	<i>MAFK</i>	protein_coding
10	45443853	45444003	-0.755	-2.27	<i>ALOX5</i>	protein_coding
9	22008867	22009116	-0.75	-3.73	<i>CDKN2B</i>	protein_coding
2	111121118	111122700	-0.742	-1.6	<i>BCL2L11</i>	protein_coding
7	50446619	50447267	-0.741	-1.9	<i>FIGNL1</i>	protein_coding
6	167825970	167826709	-0.732	-4.29	<i>MLLT4-AS1</i>	lincRNA
11	67425009	67425358	-0.727	-1.75	<i>CARNS1</i>	protein_coding
3	159764507	159764953	-0.726	-5.7	<i>IQCJ-SCHIP1</i>	protein_coding
19	18435412	18435612	-0.726	-1.46	<i>ISYNA1</i>	protein_coding
5	149004063	149004564	-0.718	-1.93	<i>SH3TC2</i>	protein_coding
7	92833274	92833474	-0.712	-2.71	<i>CDK6</i>	protein_coding
9	133575269	133575519	-0.711	-1.76	<i>ADAMTSL2</i>	protein_coding
21	34887307	34887657	-0.709	-8.34	<i>RUNX1</i>	protein_coding
1	182600355	182600455	-0.708	-2.85	<i>RGS16</i>	protein_coding
12	57225099	57225399	-0.691	-1.77	<i>NXPH4</i>	protein_coding
12	54549399	54549748	-0.688	-1.61	<i>PDE1B</i>	protein_coding
8	69672841	69673142	-0.685	-4.19	<i>SLCO5A1</i>	protein_coding
11	20045075	20045176	-0.684	-3.23	<i>NAV2</i>	protein_coding
21	34887806	34887907	-0.684	-1.87	<i>RUNX1</i>	protein_coding
1	44813750	44813999	-0.681	-3.54	<i>BTBD19</i>	protein_coding
16	733120	733679	-0.678	-2.1	<i>HAGHL</i>	protein_coding
19	31276386	31276983	-0.673	-3.11	<i>TSHZ3</i>	protein_coding
21	46129756	46132051	-0.67	-1.56	<i>COL6A2</i>	protein_coding
14	105050788	105051239	-0.667	-2.84	<i>GPR132</i>	protein_coding
16	70380923	70381074	-0.667	-3	<i>ST3GAL2</i>	protein_coding
12	128414846	128415496	-0.665	-2.78	<i>TMEM132C</i>	protein_coding
19	4449056	4449156	-0.663	-1.4	<i>CTB-50L17.8</i>	TEC
22	42571366	42571515	-0.66	-1.44	<i>SERHL2</i>	protein_coding
19	45142513	45142613	-0.653	-1.75	<i>PPP1R37</i>	protein_coding
3	52487421	52487571	-0.649	-2.05	<i>NISCH</i>	protein_coding
12	4913306	4913457	-0.644	-3.16	<i>KCNA1</i>	protein_coding
7	132508252	132508602	-0.643	-1.76	<i>PLXNA4</i>	protein_coding
1	157124568	157124668	-0.64	-1.63	<i>ETV3</i>	protein_coding
9	137461057	137461257	-0.64	-2.08	<i>PNPLA7</i>	protein_coding
7	27246198	27246741	-0.639	-2.32	<i>EVX1</i>	protein_coding
1	10653485	10653586	-0.639	-1.7	<i>RP4-734G22.3</i>	antisense
19	416831	417031	-0.639	-1.52	<i>SHC2</i>	protein_coding
19	55621574	55621917	-0.638	-1.42	<i>ZNF784</i>	protein_coding
1	31798491	31798592	-0.637	-1.87	<i>SPOCD1</i>	protein_coding
1	203307140	203307535	-0.635	-4.35	<i>BTG2</i>	protein_coding
10	35639591	35640189	-0.635	-7.22	<i>FZD8</i>	protein_coding

Activin/Nodal Signalling Controls the Epigenome and Epitranscriptome of Human Pluripotent Stem Cells

Chr.	Start	End	log2FC	logFDR	Gene name	Gene type
15	30224215	30224365	-0.63	-1.32	<i>RP11-261B23.1</i>	processed_pseudogene
1	208243970	208244270	-0.629	-2.5	<i>PLXNA2</i>	protein_coding
15	66702277	66702676	-0.629	-3.42	<i>SMAD6</i>	protein_coding
X	153862161	153862311	-0.625	-1.63	<i>L1CAM</i>	protein_coding
10	1737276	1737476	-0.623	-6.18	<i>ADARB2</i>	protein_coding
7	73831901	73832099	-0.623	-7.22	<i>CLDN4</i>	protein_coding
7	12369647	12370097	-0.619	-1.88	<i>VWDE</i>	protein_coding
2	104857196	104857545	-0.616	-5.25	<i>POU3F3</i>	protein_coding
16	30399059	30399260	-0.614	-2.27	<i>ZNF48</i>	protein_coding
16	2967642	2967742	-0.6	-3.82	<i>KREMEN2</i>	protein_coding
16	86568637	86569183	-0.599	-2.44	<i>FOXC2</i>	protein_coding
9	136049731	136049981	-0.596	-7.03	<i>NACC2</i>	protein_coding
2	128266226	128266327	-0.595	-2.01	<i>HS6ST1</i>	protein_coding
2	98822835	98823182	-0.595	-2.92	<i>KIAA1211L</i>	protein_coding
20	38240805	38241753	-0.595	-4.33	<i>KIAA1755</i>	protein_coding
1	111755294	111755394	-0.591	-2.09	<i>DDX20</i>	protein_coding
11	61509982	61510427	-0.591	-1.87	<i>LRRC10B</i>	protein_coding
22	43203424	43203874	-0.591	-1.93	<i>SCUBE1</i>	protein_coding
11	17381195	17381593	-0.586	-1.35	<i>RP1-239B22.5</i>	antisense

8 REFERENCES

- Abraham, B.J. et al., 2013. Dynamic regulation of epigenomic landscapes during hematopoiesis. *BMC genomics*, 14, pp.193–208.
- Ahmed, K. et al., 2010. Global chromatin architecture reflects pluripotency and lineage commitment in the early mouse embryo. *PLoS one*, 5(5), p.e10531.
- Aik, W. et al., 2014. Structure of human RNA N⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic acids research*, 42(7), pp.4741–54.
- Akalin, A. et al., 2015. Genomation: a Toolkit To Summarize, Annotate and Visualize Genomic Intervals. *Bioinformatics*, 31(7), pp.1127–9.
- Akiyoshi, S. et al., 1999. c-Ski acts as a transcriptional co-repressor in transforming growth factor-beta signaling through interaction with smads. *The Journal of biological chemistry*, 274(49), pp.35269–77.
- van den Akker, E. et al., 2002. Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development*, 129(9), pp.2181–93.
- Alarcón, C.R., Goodarzi, H., et al., 2015. HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events. *Cell*, 162(6), pp.1299–308.
- Alarcón, C.R., Lee, H., et al., 2015. N6-methyladenosine marks primary microRNAs for processing. *Nature*, 519(7544), pp.482–5.
- Alkema, M.J. et al., 1995. Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature*, 374(6524), pp.724–7.
- van den Aamele, J. et al., 2012. Eomesodermin induces Mesp1 expression and cardiac differentiation from embryonic stem cells in the absence of Activin. *EMBO reports*, 13(4), pp.355–62.
- Andreu-Vieyra, C. V et al., 2010. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS biology*, 8(8), p.e1000453.
- Andrews, S., 2014. SeqMonk: A tool to visualise and analyse high throughput mapped sequence data. <http://www.bioinformatics.bbsrc.ac.uk/projects/seqmonk/>.
- Ang, S.L. et al., 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development*, 119(4), pp.1301–15.
- Ang, Y.-S. et al., 2011. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell*, 145(2), pp.183–97.
- Arnold, S.J. et al., 2008. Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse. *Development*, 135(3), pp.501–11.
- Arnold, S.J. & Robertson, E.J., 2009. Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature reviews. Molecular cell biology*, 10(2), pp.91–103.

- Aszodi, A., 2012. MULTOVL: fast multiple overlaps of genomic regions. *Bioinformatics*, 28(24), pp.3318–9.
- Atlasi, Y., Looijenga, L. & Fodde, R., 2014. Cancer stem cells, pluripotency, and cellular heterogeneity: a WNTer perspective. *Current topics in developmental biology*, 107, pp.373–404.
- Aulehla, A. & Pourquié, O., 2010. Signaling gradients during paraxial mesoderm development. *Cold Spring Harbor perspectives in biology*, 2(2), p.a000869.
- Avilion, A.A. et al., 2003. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development*, 17(1), pp.126–40.
- Avior, Y., Sagi, I. & Benvenisty, N., 2016. Pluripotent stem cells in disease modelling and drug discovery. *Nature reviews. Molecular cell biology*, 17(3), pp.170–182.
- Azuara, V. et al., 2006. Chromatin signatures of pluripotent cell lines. *Nature cell biology*, 8(5), pp.532–8.
- Bach, C. et al., 2009. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene*, 28(6), pp.815–23.
- Bachiller, D. et al., 2000. The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature*, 403(6770), pp.658–61.
- Bailey, T. et al., 2013. Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. *PLoS computational biology*, 9(11), p.e1003326.
- Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics*, 27(12), pp.1653–9.
- Bakin, A. V et al., 2000. Phosphatidylinositol 3-kinase function is required for transforming growth factor beta-mediated epithelial to mesenchymal transition and cell migration. *The Journal of biological chemistry*, 275(47), pp.36803–10.
- Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–95.
- Bapat, S.A. et al., 2010. Multivalent epigenetic marks confer microenvironment-responsive epigenetic plasticity to ovarian cancer cells. *Epigenetics*, 5(8), pp.716–29.
- Bartke, T. et al., 2010. Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, 143(3), pp.470–84.
- Batista, P.J. et al., 2014. m6A RNA Modification Controls Cell Fate Transition in Mammalian Embryonic Stem Cells. *Cell Stem Cell*, 15(6), pp.707–19.
- Beattie, G.M. et al., 2005. Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem cells*, 23(4), pp.489–95.
- Beck, S. et al., 2002. Extraembryonic proteases regulate Nodal signalling during gastrulation. *Nature cell biology*, 4(12), pp.981–5.
- Beddington, R.S., Rashbass, P. & Wilson, V., 1992. Brachyury--a gene affecting mouse gastrulation and early organogenesis. *Development*, pp.157–65.
- Belo, J.A. et al., 1998. The prechordal midline of the chondrocranium is defective in Goosecoid-1 mouse mutants. *Mechanisms of development*, 72(1-2), pp.15–25.
- Beloor, J. et al., 2015. Effective gene delivery into human stem cells with a cell-targeting Peptide-modified bioreducible polymer. *Small*, 11(17), pp.2069–79.
- Benayoun, B.A. et al., 2014. H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency. *Cell*, 158(3), pp.673–688.
- Benjamini, Y. & Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, 57(1), pp.289–300.
- van den Berg, D.L.C. et al., 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell*, 6(4), pp.369–81.
- Berga-Bolaños, R. et al., 2013. NFAT5 induction by the pre-T-cell receptor serves as a selective survival signal in T-lymphocyte development. *PNAS*, 110(40), pp.16091–6.
- Bernardo, A.S. et al., 2011. BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell*, 9(2), pp.144–55.
- Bernstein, B.E. et al., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), pp.315–26.
- Bernstein, B.E. et al., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

- Beyer, T. a et al., 2013. Switch Enhancers Interpret TGF- β and Hippo Signaling to Control Cell Fate in Human Embryonic Stem Cells. *Cell reports*, 5(6), pp.1611–24.
- Biechele, S. et al., 2013. Porcn-dependent Wnt signaling is not required prior to mouse gastrulation. *Development*, 140(14), pp.2961–71.
- Bilodeau, S. et al., 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes & development*, 23(21), pp.2484–9.
- Birke, M. et al., 2002. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. *Nucleic acids research*, 30(4), pp.958–65.
- Blahna, M.T. & Hata, A., 2012. Smad-mediated regulation of microRNA biosynthesis. *FEBS letters*, 586(14), pp.1906–12.
- Blakeley, P. et al., 2015. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development*, 142(18), pp.3151–65.
- Blanchet, M.H., Le Good, J.A., Oorschot, V., et al., 2008. Cripto localizes Nodal at the limiting membrane of early endosomes. *Science signaling*, 1(45), p.ra13.
- Blanchet, M.H., Le Good, J.A., Mesnard, D., et al., 2008. Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation. *The EMBO journal*, 27(19), pp.2580–91.
- Bledau, A.S. et al., 2014. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development*, 141(5), pp.1022–35.
- Blum, M. et al., 1992. Gastrulation in the mouse: the role of the homeobox gene goosecoid. *Cell*, 69(7), pp.1097–106.
- Boersema, P.J. et al., 2009. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature protocols*, 4(4), pp.484–494.
- Boettcher, M. & McManus, M.T., 2015. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular cell*, 58(4), pp.575–585.
- Bokar, J.A. et al., 1994. Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *The Journal of biological chemistry*, 269(26), pp.17697–704.
- Bokar, J.A. et al., 1997. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *RNA*, 3(11), pp.1233–47.
- Bolger, A.M., Lohse, M. & Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), pp.2114–20.
- Boosen, M. et al., 2009. Par-4 is an essential downstream target of DAP-like kinase (Dlk) in Dlk/Par-4-mediated apoptosis. *Molecular biology of the cell*, 20(18), pp.4010–20.
- Boroviak, T. et al., 2015. Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Developmental cell*, 35(3), pp.366–382.
- von Both, I. et al., 2004. Foxh1 is essential for development of the anterior heart field. *Developmental cell*, 7(3), pp.331–45.
- Bouwmeester, T. et al., 1996. Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature*, 382(6592), pp.595–601.
- Boyer, L. a et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–53.
- Boyer, L.A. et al., 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6), pp.947–56.
- Braam, S.R. et al., 2008. Improved genetic manipulation of human embryonic stem cells. *Nature methods*, 5(5), pp.389–92.
- Bradley, J.A., Bolton, E.M. & Pedersen, R.A., 2002. Stem cell medicine encounters the immune system. *Nature reviews. Immunology*, 2(11), pp.859–71.
- Brennan, J. et al., 2001. Nodal signalling in the epiblast patterns the early mouse embryo. *Nature*, 411(6840), pp.965–9.
- Brons, I.G. et al., 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448(7150), pp.191–5.
- Brookes, E. et al., 2012. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell*, 10(2), pp.157–70.
- Brown, C.O. et al., 2004. The cardiac determination factor, Nkx2-5, is activated by mutual cofactors GATA-4 and Smad1/4 via a novel upstream enhancer. *The Journal of biological chemistry*, 279(11), pp.10659–69.

- Brown, S. et al., 2011. Activin/Nodal signaling controls divergent transcriptional networks in human embryonic stem cells and in endoderm progenitors. *Stem cells*, 29(8), pp.1176–85.
- Bruneau, B.G. et al., 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell*, 106(6), pp.709–21.
- Byrne, S.M., Mali, P. & Church, G.M., 2014. Genome editing in human stem cells. *Methods in enzymology*, 546, pp.119–38.
- Cahan, P. & Daley, G.Q., 2013. Origins and implications of pluripotent stem cell variability and heterogeneity. *Nature reviews. Molecular cell biology*, 14(6), pp.357–68.
- Caja, L., Kahata, K. & Moustakas, A., 2012. Context-dependent action of transforming growth factor β family members on normal and cancer stem cells. *Current pharmaceutical design*, 18(27), pp.4072–86.
- Calo, E. & Wysocka, J., 2013. Modification of enhancer chromatin: what, how, and why? *Molecular cell*, 49(5), pp.825–37.
- Campbell, J.M. et al., 2012. Insulin increases epiblast cell number of in vitro cultured mouse embryos via the PI3K/GSK3/p53 pathway. *Stem cells and development*, 21(13), pp.2430–41.
- Camus, A. et al., 2006. Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Developmental biology*, 295(2), pp.743–55.
- Cao, F. et al., 2010. Comparison of gene-transfer efficiency in human embryonic stem cells. *Molecular imaging and biology*, 12(1), pp.15–24.
- Cao, R. et al., 2002. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, 298(5595), pp.1039–43.
- Cao, R., Tsukada, Y.-I. & Zhang, Y., 2005. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Molecular cell*, 20(6), pp.845–54.
- Carlson, B.M., 2014. *Human Embryology and Developmental Biology* 5th ed., Elsevier.
- Chamberlain, S.J., Yee, D. & Magnuson, T., 2008. Polycomb repressive complex 2 is dispensable for maintenance of embryonic stem cell pluripotency. *Stem cells*, 26(6), pp.1496–505.
- Chambers, I. et al., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5), pp.643–55.
- Chambers, I. et al., 2007. Nanog safeguards pluripotency and mediates germline development. *Nature*, 450(7173), pp.1230–4.
- Chambers, S.M. et al., 2009. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology*, 27(3), pp.275–280.
- Chan, Y.-S. et al., 2013. Induction of a Human Pluripotent State with Distinct Regulatory Circuitry that Resembles Preimplantation Epiblast. *Cell Stem Cell*, 13(6), pp.663–675.
- Chapman, D.L. et al., 1996. Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Developmental biology*, 180(2), pp.534–42.
- Chaudhury, A. & Howe, P.H., 2009. The tale of transforming growth factor-beta (TGFbeta) signaling: a soigné enigma. *IUBMB life*, 61(10), pp.929–39.
- Chazaud, C. et al., 2006. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell*, 10(5), pp.615–24.
- Chen, B. et al., 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7), pp.1479–91.
- Chen, C.-R. et al., 2002. E2F4/5 and p107 as Smad cofactors linking the TGFbeta receptor to c-myc repression. *Cell*, 110(1), pp.19–32.
- Chen, E.Y. et al., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14, p.128.
- Chen, G. et al., 2011. Chemically defined conditions for human iPSC derivation and culture. *Nature methods*, 8(5), pp.424–9.
- Chen, J. et al., 2013. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature genetics*, 45(1), pp.34–42.
- Chen, K.G. et al., 2014. Human pluripotent stem cell culture: considerations for maintenance, expansion, and therapeutics. *Cell Stem Cell*, 14(1), pp.13–26.
- Chen, T. et al., 2012. An RNA interference screen uncovers a new molecule in stem cell self-renewal and long-term regeneration. *Nature*, 485(7396), pp.104–8.
- Chen, T. et al., 2015. m6A RNA Methylation Is Regulated by MicroRNAs and Promotes Reprogramming to Pluripotency. *Cell Stem Cell*, 16(3), pp.289–301.
- Chen, T. & Dent, S.Y.R., 2013. Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature Reviews Genetics*, 15(2), pp.93–106.

- Chen, X. et al., 1997. Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature*, 389(6646), pp.85–9.
- Chen, X., Rubock, M.J. & Whitman, M., 1996. A transcriptional partner for MAD proteins in TGF-beta signalling. *Nature*, 383(6602), pp.691–6.
- Chen, Y. et al., 2015. Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. *Cell Stem Cell*, 17(2), pp.233–244.
- Cheng, A.W. et al., 2013. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research*, 23(10), pp.1163–71.
- Cheung, C. et al., 2014. Directed differentiation of embryonic origin-specific vascular smooth muscle subtypes from human pluripotent stem cells. *Nature protocols*, 9(4), pp.929–38.
- Cheung, C. et al., 2012. Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nature biotechnology*, 30(2), pp.165–73.
- Cheung, M.K. et al., 2013. FTO expression is regulated by availability of essential amino acids. *International journal of obesity*, 37(5), pp.744–7.
- Chia, N.-Y. et al., 2010. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, 468(7321), pp.316–20.
- Chng, Z. et al., 2010. SIP1 mediates cell-fate decisions between neuroectoderm and mesendoderm in human pluripotent stem cells. *Cell Stem Cell*, 6(1), pp.59–70.
- Cho, C.H. et al., 2012. Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells. *Diabetologia*, 55(12), pp.3284–95.
- Choi, J. et al., 2016. N(6)-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nature structural & molecular biology*, 23(2), pp.110–5.
- Choy, L. & Derynck, R., 2003. Transforming growth factor-beta inhibits adipocyte differentiation by Smad3 interacting with CCAAT/enhancer-binding protein (C/EBP) and repressing C/EBP transactivation function. *The Journal of biological chemistry*, 278(11), pp.9609–19.
- Chu, G.C. et al., 2004. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development*, 131(15), pp.3501–12.
- Ciruna, B.G. & Rossant, J., 1999. Expression of the T-box gene Eomesodermin during early mouse development. *Mechanisms of development*, 81(1-2), pp.199–203.
- Cloos, P.A.C. et al., 2008. Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease. *Genes & development*, 22(9), pp.1115–1140.
- Clouaire, T. et al., 2012. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes & development*, 26(15), pp.1714–28.
- Cohen, D.E. & Melton, D., 2011. Turning straw into gold: directing cell fate for regenerative medicine. *Nature reviews. Genetics*, 12(4), pp.243–252.
- Cong, L. et al., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), pp.819–23.
- Conlon, F.L. et al., 1994. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development*, 120(7), pp.1919–28.
- Conrad, S. et al., 2008. Generation of pluripotent stem cells from adult human testis. *Nature*, 456(7220), pp.344–9.
- Cooper, G.M. & Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature reviews. Genetics*, 12(9), pp.628–40.
- Cooper, S. et al., 2014. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell reports*, 7(5), pp.1456–70.
- Costa, M. et al., 2007. A method for genetic modification of human embryonic stem cells using electroporation. *Nature protocols*, 2(4), pp.792–6.
- Costa, Y. et al., 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, 495(7441), pp.370–4.
- Costello, I. et al., 2011. The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature cell biology*, 13(9), pp.1084–91.
- Cox, J. & Mann, M., 2008. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12), pp.1367–72.
- Csepány, T. et al., 1990. Sequence specificity of mRNA N6-adenosine methyltransferase. *The Journal of biological chemistry*, 265(33), pp.20117–22.

- Cui, K. et al., 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell*, 4(1), pp.80–93.
- Cui, X. et al., 2015. MeTDiff: a Novel Differential RNA Methylation Analysis for MeRIP-Seq Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, in press.
- D'Amour, K.A. et al., 2005. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature biotechnology*, 23(12), pp.1534–41.
- Dahéron, L. et al., 2004. LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem cells*, 22(5), pp.770–8.
- Daley, G.Q. et al., 2007. Ethics. The ISSCR guidelines for human embryonic stem cell research. *Science*, 315(5812), pp.603–4.
- Dalton, S., 2013. G1 compartmentalization and cell fate coordination. *Cell*, 155(1), pp.13–4.
- Darr, H., Mayshar, Y. & Benvenisty, N., 2006. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development*, 133(6), pp.1193–201.
- Davis, B.N. et al., 2008. SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*, 454(7200), pp.56–61.
- Deglincerti, A. et al., 2016. Self-organization of the in vitro attached human embryo. *Nature*, 533(7602), pp.251–4.
- DeKolver, R.C. et al., 2010. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome research*, 20(8), pp.1133–42.
- Denissov, S. et al., 2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development*, 141(3), pp.526–37.
- Dennler, S., Huet, S. & Gauthier, J.M., 1999. A short amino-acid sequence in MH1 domain is responsible for functional differences between Smad2 and Smad3. *Oncogene*, 18(8), pp.1643–8.
- Deribe, Y.L., Pawson, T. & Dikic, I., 2010. Post-translational modifications in signal integration. *Nature structural & molecular biology*, 17(6), pp.666–72.
- Derynck, R. & Zhang, Y.E., 2003. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature*, 425(6958), pp.577–84.
- Desrosiers, R., Friderici, K. & Rottman, F., 1974. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *PNAS*, 71(10), pp.3971–5.
- Dietrich, J.E. & Hiiragi, T., 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development*, 134(23), pp.4219–31.
- Ding, J. et al., 1998. Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature*, 395(6703), pp.702–7.
- Ding, J. et al., 2012. Oct4 links multiple epigenetic pathways to the pluripotency network. *Cell research*, 22(1), pp.155–67.
- Dion, M.F. et al., 2005. Genomic characterization reveals a simple histone H4 acetylation code. *PNAS*, 102(15), pp.5501–6.
- Do, D.V. et al., 2013. A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo. *Genes & development*, 27(12), pp.1378–90.
- Doetschman, T., Maeda, N. & Smithies, O., 1988. Targeted mutation of the Hprt gene in mouse embryonic stem cells. *PNAS*, 85(22), pp.8583–7.
- Dominissini, D. et al., 2014. Genomics and Proteomics. Roadmap to the epitranscriptome. *Science*, 346(6214), p.1192.
- Dominissini, D. et al., 2016. The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature*, 530(7591), pp.441–446.
- Dominissini, D. et al., 2012. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, 485(7397), pp.201–6.
- Dominissini, D. et al., 2013. Transcriptome-wide mapping of N(6)-methyladenosine by m(6)A-seq based on immunocapturing and massively parallel sequencing. *Nature protocols*, 8(1), pp.176–89.
- Dong, X. et al., 2012. Modeling gene expression using chromatin features in various cellular contexts. *Genome biology*, 13(9), p.R53.

- Douvaras, P. et al., 2014. Efficient generation of myelinating oligodendrocytes from primary progressive multiple sclerosis patients by induced pluripotent stem cells. *Stem Cell Reports*, 3(2), pp.250–9.
- Down, T.A., Piipari, M. & Hubbard, T.J.P., 2011. Dalliace: interactive genome viewing on the web. *Bioinformatics*, 27(6), pp.889–90.
- Doyon, Y. et al., 2011. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature methods*, 8(1), pp.74–9.
- Drost, J. et al., 2015. Sequential cancer mutations in cultured human intestinal stem cells. *Nature*, 521(7550), pp.43–7.
- Du, P., Kibbe, W. a & Lin, S.M., 2008. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24(13), pp.1547–8.
- Dubin, D.T. & Taylor, R.H., 1975. The methylation state of poly A-containing messenger RNA from cultured hamster cells. *Nucleic acids research*, 2(10), pp.1653–68.
- Dubrulle, J. & Pourquié, O., 2004. fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature*, 427(6973), pp.419–22.
- Dunn, N.R. et al., 2004. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development*, 131(8), pp.1717–1728.
- Dunn, N.R. et al., 2005. Mice exclusively expressing the short isoform of Smad2 develop normally and are viable and fertile. *Genes & development*, 19(1), pp.152–163.
- Dunn, S.J. et al., 2014. Defining an essential transcription factor program for naïve pluripotency. *Science*, 344(6188), pp.1156–60.
- Ebert, M.S. & Sharp, P.A., 2010. MicroRNA sponges: Progress and possibilities. *RNA*, 16(11), pp.2043–2050.
- Ebisawa, T. et al., 2001. Smurf1 interacts with transforming growth factor-beta type I receptor through Smad7 and induces receptor degradation. *The Journal of biological chemistry*, 276(16), pp.12477–80.
- Edlund, S. et al., 2002. Transforming growth factor-beta-induced mobilization of actin cytoskeleton requires signaling by small GTPases Cdc42 and RhoA. *Molecular biology of the cell*, 13(3), pp.902–14.
- Efroni, S. et al., 2008. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, 2(5), pp.437–47.
- Eiges, R. et al., 2001. Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. *Current biology*, 11(7), pp.514–8.
- Eiraku, M. et al., 2008. Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell*, 3(5), pp.519–32.
- Eiselleova, L. et al., 2009. A complex role for FGF-2 in self-renewal, survival, and adhesion of human embryonic stem cells. *Stem cells*, 27(8), pp.1847–57.
- Ellis, J., 2005. Silencing and variegation of gammaretrovirus and lentivirus vectors. *Human gene therapy*, 16(11), pp.1241–6.
- Ernst, P. et al., 2004. An Mll-dependent Hox program drives hematopoietic progenitor expansion. *Current biology*, 14(22), pp.2063–9.
- Ernst, P. & Vakoc, C.R., 2012. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Briefings in functional genomics*, 11(3), pp.217–26.
- Evans, M.J. & Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292, pp.154–156.
- Fasano, C.A. et al., 2010. Efficient derivation of functional floor plate tissue from human embryonic stem cells. *Cell Stem Cell*, 6(4), pp.336–47.
- Fath, S. et al., 2011. Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PloS one*, 6(3), p.e17596.
- Faust, C. et al., 1995. The eed mutation disrupts anterior mesoderm production in mice. *Development*, 121(2), pp.273–85.
- Fei, T. et al., 2010. Smad2 mediates Activin/Nodal signaling in mesendoderm differentiation of mouse embryonic stem cells. *Cell research*, 20(12), pp.1306–18.
- Feng, X., Grossman, R. & Stein, L., 2011. PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC bioinformatics*, 12(1), p.139.
- Feng, X.H. et al., 1998. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for smad3 in TGF-beta-induced transcriptional activation. *Genes & development*, 12(14), pp.2153–63.

- Feng, X.H., Lin, X. & Derynck, R., 2000. Smad2, Smad3 and Smad4 cooperate with Sp1 to induce p15(Ink4B) transcription in response to TGF-beta. *The EMBO journal*, 19(19), pp.5178–93.
- Fischer, J. et al., 2009. Inactivation of the Fto gene protects from obesity. *Nature*, 458(7240), pp.894–8.
- Flynn, R.A. & Chang, H.Y., 2014. Long noncoding RNAs in cell-fate programming and reprogramming. *Cell Stem Cell*, 14(6), pp.752–61.
- Fong, H., Hohenstein, K.A. & Donovan, P.J., 2008. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. *Stem cells*, 26(8), pp.1931–8.
- Ford, A.M. et al., 2009. The TEL-AML1 leukemia fusion gene dysregulates the TGF-beta pathway in early B lineage progenitor cells. *The Journal of clinical investigation*, 119(4), pp.826–36.
- Fordham, R.P. et al., 2013. Transplantation of expanded fetal intestinal progenitors contributes to colon regeneration after injury. *Cell Stem Cell*, 13(6), pp.734–44.
- Frankenberg, S. et al., 2011. Primitive endoderm differentiates via a three-step mechanism involving Nanog and RTK signaling. *Developmental cell*, 21(6), pp.1005–13.
- Fräter-Schröder, M. et al., 1986. Transforming growth factor-beta inhibits endothelial cell proliferation. *Biochemical and Biophysical Research Communications*, 137(1), pp.295–302.
- Fu, Y. et al., 2013. FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nature communications*, 4, p.1798.
- Fujikura, J. et al., 2002. Differentiation of embryonic stem cells is induced by GATA factors. *Genes & development*, 16(7), pp.784–9.
- Fujita, T. & Fujii, H., 2013. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and biophysical research communications*, 439(1), pp.132–6.
- Fukuchi, M. et al., 2001. Ligand-dependent degradation of Smad3 by a ubiquitin ligase complex of ROC1 and associated proteins. *Molecular biology of the cell*, 12(5), pp.1431–43.
- Fukusumi, Y., Naruse, C. & Asano, M., 2008. Wtap is required for differentiation of endoderm and mesoderm in the mouse embryo. *Developmental dynamics*, 237(3), pp.618–629.
- Funa, N.S. et al., 2014. B-Catenin Regulates Primitive Streak Induction through Collaborative Interactions with SMAD2/SMAD3 and OCT4. *Cell Stem Cell*, pp.639–652.
- Fustin, J.M. et al., 2013. RNA-Methylation-Dependent RNA Processing Controls the Speed of the Circadian Clock. *Cell*, 155(4), pp.793–806.
- Gafni, O. et al., 2013. Derivation of novel human ground state naive pluripotent stem cells. *Nature*.
- Gaj, T., Gersbach, C.A. & Barbas, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*, 31(7), pp.397–405.
- Galliher, A.J. & Schiemann, W.P., 2007. Src phosphorylates Tyr284 in TGF-beta type II receptor and regulates TGF-beta stimulation of p38 MAPK during breast cancer cell proliferation and invasion. *Cancer research*, 67(8), pp.3752–8.
- Galvin-Burgess, K.E. et al., 2013. TGF- β -superfamily signaling regulates embryonic stem cell heterogeneity: self-renewal as a dynamic and regulated equilibrium. *Stem cells*, 31(1), pp.48–58.
- Galvin, K.E. et al., 2010. Nodal signaling regulates the bone morphogenic protein pluripotency pathway in mouse embryonic stem cells. *The Journal of biological chemistry*, 285(26), pp.19747–56.
- Gao, Z. et al., 2012. Determination of protein interactome of transcription factor Sox2 in embryonic stem cells engineered for inducible expression of four reprogramming factors. *The Journal of biological chemistry*, 287(14), pp.11384–97.
- Gerken, T. et al., 2007. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, 318(5855), pp.1469–72.
- Germain, S. et al., 2000. Homeodomain and winged-helix transcription factors recruit activated Smads to distinct promoter elements via a common Smad interaction motif. *Genes & development*, 14(4), pp.435–51.
- Gerrard, L. et al., 2005. Stably transfected human embryonic stem cell clones express OCT4-specific green fluorescent protein and maintain self-renewal and pluripotency. *Stem cells*, 23(1), pp.124–33.
- Geula, S. et al., 2015. m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science*, 347(6225), pp.1002–6.

- Ghosal, S., Das, S. & Chakrabarti, J., 2013. Long noncoding RNAs: new players in the molecular mechanism for maintenance and differentiation of pluripotent stem cells. *Stem cells and development*, 22(16), pp.2240–53.
- Gilbert, L.A. et al., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), pp.442–51.
- Glaser, S. et al., 2006. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*, 133(8), pp.1423–32.
- Glaser, S. et al., 2009. The histone 3 lysine 4 methyltransferase, Mll2, is only required briefly in development and spermatogenesis. *Epigenetics & chromatin*, 2(1), p.5.
- Göke, J. et al., 2013. Genome-wide kinase-chromatin interactions reveal the regulatory network of ERK signaling in human embryonic stem cells. *Molecular cell*, 50(6), pp.844–55.
- Gonzales-Cope, M. et al., 2016. Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *BMC Genomics*, 17(1), p.95.
- González, F. et al., 2014. An iCRISPR Platform for Rapid, Multiplexable, and Inducible Genome Editing in Human Pluripotent Stem Cells. *Cell Stem Cell*, 15(2), pp.215–26.
- Graff, J.M., Bansal, A. & Melton, D.A., 1996. Xenopus Mad proteins transduce distinct subsets of signals for the TGF beta superfamily. *Cell*, 85(4), pp.479–87.
- Gray, D.C. et al., 2007. pHUSH: a single vector system for conditional gene expression. *BMC biotechnology*, 7, p.61.
- Greber, B. et al., 2010. Conserved and Divergent Roles of FGF Signaling in Mouse Epiblast Stem Cells and Human Embryonic Stem Cells. *Cell Stem Cell*, 6(3), pp.215–226.
- Green, J.J. et al., 2008. Nanoparticles for gene transfer to human embryonic stem cell colonies. *Nano letters*, 8(10), pp.3126–30.
- Green, M.R. & Sambrook, J., 2012. *Molecular Cloning: A Laboratory Manual (4th Edition)*, Cold Spring Harbor Laboratory Press.
- Greve, T.S., Judson, R.L. & Belloch, R., 2013. microRNA control of mouse and human pluripotent stem cell behavior. *Annual review of cell and developmental biology*, 29, pp.213–39.
- Gu, Z. et al., 1998. The type I activin receptor ActRIB is required for egg cylinder organization and gastrulation in the mouse. *Genes & development*, 12(6), pp.844–57.
- Guenther, M.G. et al., 2007. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell*, 130(1), pp.77–88.
- Gulati, P. et al., 2013. Role for the obesity-related FTO gene in the cellular sensing of amino acids. *PNAS*, 110(7), pp.2557–62.
- Guo, G. et al., 2016. Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports*, 6(4), pp.437–46.
- Guo, X. & Wang, X.F., 2009. Signaling cross-talk between TGF-beta/BMP and other pathways. *Cell research*, 19(1), pp.71–88.
- Gurdon, J.B., 1962. The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of embryology and experimental morphology*, 10, pp.622–40.
- Gurdon, J.B., Elsdale, T.R. & Fischberg, M., 1958. Sexually Mature Individuals of *Xenopus laevis* from the Transplantation of Single Somatic Nuclei. *Nature*, 182(4627), pp.64–65.
- Haenebalcke, L. et al., 2013. Efficient ROSA26-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem cell reviews*, 9(6), pp.774–85.
- Hafner, M. et al., 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141(1), pp.129–41.
- Hanai, J. et al., 1999. Interaction and functional cooperation of PEBP2/CBF with Smads. Synergistic induction of the immunoglobulin germline Calpha promoter. *The Journal of biological chemistry*, 274(44), pp.31577–82.
- Hancock, S.N. et al., 1999. Mapping and expression analysis of the mouse ortholog of *Xenopus* Eomesodermin. *Mechanisms of development*, 81(1-2), pp.205–8.
- Hanna, J. et al., 2010. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. *PNAS*, 107(20), pp.9222–7.
- Hanna, J. et al., 2009. Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell*, 4(6), pp.513–24.
- Hannan, N.R.F. et al., 2015. Generation of Distal Airway Epithelium from Multipotent Human Foregut Stem Cells. *Stem cells and development*, 24(14), pp.1680–90.

- Hannan, N.R.F., Fordham, R.P., et al., 2013. Generation of multipotent foregut stem cells from human pluripotent stem cells. *Stem Cell Reports*, 1(4), pp.293–306.
- Hannan, N.R.F., Segeritz, C.-P., et al., 2013. Production of hepatocyte-like cells from human pluripotent stem cells. *Nature protocols*, 8(2), pp.430–7.
- Harikumar, A. & Meshorer, E., 2015. Chromatin remodeling and bivalent histone modifications in embryonic stem cells. *EMBO reports*, 16(12), pp.1609–1619.
- Harper, J.E. et al., 1990. Sequence specificity of the human mRNA N6-adenosine methylase in vitro. *Nucleic acids research*, 18(19), pp.5735–41.
- Harrison, C.A. et al., 2005. Antagonists of activin signaling: mechanisms and potential biological applications. *Trends in endocrinology and metabolism*, 16(2), pp.73–8.
- Harshman, S.W. et al., 2013. H1 histones: current perspectives and challenges. *Nucleic acids research*, 41(21), pp.9593–609.
- Hart, A.H. et al., 2004. Identification, cloning and expression analysis of the pluripotency promoting Nanog genes in mouse and human. *Developmental dynamics*, 230(1), pp.187–98.
- Hart, A.H. et al., 2002. Mixl1 is required for axial mesendoderm morphogenesis and patterning in the murine embryo. *Development*, 129(15), pp.3597–608.
- Hatano, S.Y. et al., 2005. Pluripotential competence of cells associated with Nanog activity. *Mechanisms of development*, 122(1), pp.67–79.
- Hay, D.C. et al., 2004. Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem cells*, 22(2), pp.225–35.
- Hayashi, H. et al., 1997. The MAD-Related Protein Smad7 Associates with the TGF β Receptor and Functions as an Antagonist of TGF β Signaling. *Cell*, 89(7), pp.1165–1173.
- Heger, A. et al., 2013. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics*, 29(16), pp.2046–8.
- Heikkinen, P.T. et al., 2010. Hypoxia-activated Smad3-specific dephosphorylation by PP2A. *The Journal of biological chemistry*, 285(6), pp.3740–9.
- Heinz, S. et al., 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), pp.576–89.
- Henikoff, S. & Shilatifard, A., 2011. Histone modification: cause or cog? *Trends in genetics*, 27(10), pp.389–96.
- Henriksen, J.R. et al., 2007. Comparison of RNAi efficiency mediated by tetracycline-responsive H1 and U6 promoter variants in mammalian cell lines. *Nucleic acids research*, 35(9), p.e67.
- Herbst, F. et al., 2012. Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo. *Molecular therapy*, 20(5), pp.1014–21.
- Herrmann, B.G., 1991. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development*, 113(3), pp.913–7.
- Heyer, J. et al., 1999. Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. *PNAS*, 96(22), pp.12595–600.
- Hilton, I.B. et al., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology*, 33(5), pp.510–7.
- Hochedlinger, K. & Jaenisch, R., 2015. Induced Pluripotency and Epigenetic Reprogramming. *Cold Spring Harbor perspectives in biology*, 7(12), p.a019448.
- Hockemeyer, D. et al., 2009. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature biotechnology*, 27(9), pp.851–7.
- Hockemeyer, D. et al., 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology*, 29(8), pp.731–4.
- Hodawadekar, S.C. & Marmorstein, R., 2007. Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene*, 26(37), pp.5528–40.
- Hödl, M. & Basler, K., 2012. Transcription in the absence of histone H3.2 and H3K4 methylation. *Current biology*, 22(23), pp.2253–7.
- Hogan, B.L., 1996. Bone morphogenetic proteins in development. *Current opinion in genetics & development*, 6(4), pp.432–8.
- Holden, C., 2009. Obama executive order. For Congress and NIH, headaches ahead on stem cells. *Science*, 323(5921), pp.1552–3.

- Hoodless, P. a et al., 2001. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes & development*, 15(10), pp.1257–71.
- Horiuchi, K. et al., 2013. Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. *The Journal of biological chemistry*, 288(46), pp.33292–302.
- Horiuchi, K. et al., 2006. Wilms' tumor 1-associating protein regulates G2/M transition through stabilization of cyclin A2 mRNA. *PNAS*, 103(46), pp.17278–83.
- Hsiao, C. et al., 2016. Human pluripotent stem cell culture density modulates YAP signaling. *Biotechnology journal*, 11(5), pp.662–75.
- Hu, D., Garruss, A.S., et al., 2013. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature structural & molecular biology*, 20(9), pp.1093–7.
- Hu, D., Gao, X., et al., 2013. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Molecular and cellular biology*, 33(23), pp.4745–54.
- Huang, Y. et al., 2012. In Vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation. *Cell reports*, 2(6), pp.1571–8.
- Huangfu, D. et al., 2008. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature biotechnology*, 26(11), pp.1269–75.
- Hubner, N.C. et al., 2015. A quantitative proteomics tool to identify DNA-protein interactions in primary cells or blood. *Journal of proteome research*, 14(2), pp.1315–29.
- Hubner, N.C. et al., 2010. Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of cell biology*, 189(4), pp.739–54.
- Hubner, N.C. & Mann, M., 2011. Extracting gene function from protein-protein interactions using Quantitative BAC InteraCtomics (QUBIC). *Methods*, 53(4), pp.453–9.
- Huminiacki, L. et al., 2009. Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom. *BMC evolutionary biology*, 9, p.28.
- Humphrey, R.K. et al., 2004. Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem cells*, 22(4), pp.522–30.
- Hyslop, L. et al., 2005. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. *Stem cells*, 23(8), pp.1035–43.
- Hyun, I., 2010. The bioethics of stem cell research and therapy. *The Journal of clinical investigation*, 120(1), pp.71–5.
- Inman, G.J., Nicolás, F.J., Callahan, J.F., et al., 2002. SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Molecular pharmacology*, 62(1), pp.65–74.
- Inman, G.J., Nicolás, F.J. & Hill, C.S., 2002. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. *Molecular cell*, 10(2), pp.283–94.
- Irion, S. et al., 2007. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nature biotechnology*, 25(12), pp.1477–82.
- Itoh, S. et al., 2000. The transcriptional co-activator P/CAF potentiates TGF-beta/Smad signaling. *Nucleic acids research*, 28(21), pp.4291–8.
- Itskovitz-Eldor, J. et al., 2000. Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers. *Molecular medicine*, 6(2), pp.88–95.
- Itsykson, P. et al., 2005. Derivation of neural precursors from human embryonic stem cells in the presence of noggin. *Molecular and cellular neurosciences*, 30(1), pp.24–36.
- Ivanova, N. et al., 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature*, 442(7102), pp.533–8.
- Iyer, D. et al., 2015. Robust derivation of epicardium and its differentiated smooth muscle cell progeny from human pluripotent stem cells. *Development*, 142(8), pp.1528–41.
- Izumi, N. et al., 2007. Dissecting the molecular hierarchy for mesendoderm differentiation through a combination of embryonic stem cell culture and RNA interference. *Stem cells*, 25(7), pp.1664–74.
- James, D. et al., 2005. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development*, 132(6), pp.1273–82.
- Jia, G. et al., 2011. N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature chemical biology*, 7(12), pp.885–7.

- Jiang, H. et al., 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell*, 144(4), pp.513–25.
- Jinek, M. et al., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096), pp.816–21.
- Joshi, N. & Fass, J., 2011. Sickie: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle>, p.2011.
- Joung, J.K. & Sander, J.D., 2013. TALENs: a widely applicable technology for targeted genome editing. *Nature reviews. Molecular cell biology*, 14(1), pp.49–55.
- Kaestner, K.H., Knochel, W. & Martinez, D.E., 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes & development*, 14(2), pp.142–146.
- Kahan, B.W. & Ephrussi, B., 1970. Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *Journal of the National Cancer Institute*, 44(5), pp.1015–36.
- Kahata, K. et al., 2004. Regulation of transforming growth factor-beta and bone morphogenetic protein signalling by transcriptional coactivator GCN5. *Genes to cells*, 9(2), pp.143–51.
- Kamiya, D. et al., 2011. Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature*, 470(7335), pp.503–9.
- Kanai-Azuma, M. et al., 2002. Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development*, 129(10), pp.2367–79.
- Kanatsu-Shinohara, M. et al., 2004. Generation of pluripotent stem cells from neonatal mouse testis. *Cell*, 119(7), pp.1001–12.
- Kane, S.E. & Beemon, K., 1985. Precise localization of m6A in Rous sarcoma virus RNA reveals clustering of methylation sites: implications for RNA processing. *Molecular and cellular biology*, 5(9), pp.2298–306.
- Kang, J.S. et al., 2005. Repression of Runx2 function by TGF-beta through recruitment of class II histone deacetylases by Smad3. *The EMBO journal*, 24(14), pp.2543–55.
- Kang, Y., Chen, C.R. & Massagué, J., 2003. A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. *Molecular cell*, 11(4), pp.915–26.
- Kappel, S. et al., 2007. Silencing of mammalian genes by tetracycline-inducible shRNA expression. *Nature protocols*, 2(12), pp.3257–69.
- Kato, Y. et al., 2002. A component of the ARC/Mediator complex required for TGF beta/Nodal signalling. *Nature*, 418(6898), pp.641–6.
- Kattman, S.J. et al., 2011. Stage-specific optimization of activin/nodal and BMP signaling promotes cardiac differentiation of mouse and human pluripotent stem cell lines. *Cell Stem Cell*, 8(2), pp.228–40.
- Kaustov, L. et al., 2011. Recognition and specificity determinants of the human cbx chromodomains. *The Journal of biological chemistry*, 286(1), pp.521–9.
- Kawasaki, H. et al., 2000. Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron*, 28(1), pp.31–40.
- Ke, S. et al., 2015. A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes & Development*, 29(19), pp.2037–53.
- Kearns, N. a et al., 2014. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development*, 141(1), pp.219–23.
- Keene, J.D., 2007. RNA regulons: coordination of post-transcriptional events. *Nature reviews. Genetics*, 8(7), pp.533–43.
- Kelly, O.G., Pinson, K.I. & Skarnes, W.C., 2004. The Wnt co-receptors Lrp5 and Lrp6 are essential for gastrulation in mice. *Development*, 131(12), pp.2803–15.
- Kierzek, E., 2003. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic acids research*, 31(15), pp.4472–4480.
- Kim, D. et al., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 14(4), p.R36.
- Kim, H. & Kim, J.S., 2014. A guide to genome engineering with programmable nucleases. *Nature reviews. Genetics*, 15(5), pp.321–34.
- Kim, S.I. et al., 2016. Inducible Transgene Expression in Human iPS Cells Using Versatile All-in-One piggyBac Transposons. *Methods in Molecular Biology*, 1357, pp.111–31.

- Kim, S.W. et al., 2011. Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Developmental biology*, 357(2), pp.492–504.
- Kimelman, D. & Griffin, K.J., 2000. Vertebrate mesendoderm induction and patterning. *Current opinion in genetics & development*, 10(4), pp.350–6.
- Kinder, S.J. et al., 1999. The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development*, 126(21), pp.4691–701.
- Kispert, A. et al., 1994. Homologs of the mouse Brachyury gene are involved in the specification of posterior terminal structures in *Drosophila*, *Tribolium*, and *Locusta*. *Genes & development*, 8(18), pp.2137–50.
- Knott, G.J., Bond, C.S. & Fox, A.H., 2016. The DBHS proteins SFPQ, NONO and PSPC1: a multipurpose molecular scaffold. *Nucleic acids research*, 44(9), pp.3989–4004.
- Komuro, A. et al., 2004. Negative regulation of transforming growth factor-beta (TGF-beta) signaling by WW domain-containing protein 1 (WWP1). *Oncogene*, 23(41), pp.6914–23.
- König, J. et al., 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature structural & molecular biology*, 17(7), pp.909–15.
- Kopp, J.L. et al., 2008. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem cells*, 26(4), pp.903–11.
- Kouzarides, T., 2007. Chromatin modifications and their function. *Cell*, 128(4), pp.693–705.
- Krishnan, M. et al., 2006. Effects of epigenetic modulation on reporter gene expression: implications for stem cell imaging. *FEBS Journal*, 20(1), pp.106–8.
- Kubo, A. et al., 2004. Development of definitive endoderm from embryonic stem cells in culture. *Development*, 131(7), pp.1651–62.
- Kun, Q. et al., 2014. A Simple and Efficient System for Regulating Gene Expression in Human Pluripotent Stem Cells and Derivatives. *Stem cells*, 32(5), pp.1230–38.
- Kung, A.L. et al., 2000. Gene dose-dependent control of hematopoiesis and hematologic tumor suppression by CBP. *Genes & development*, 14(3), pp.272–7.
- Kuratomi, G. et al., 2005. NEDD4-2 negatively regulates TGF-beta signalling by inducing ubiquitin-mediated degradation of Smad2 and TGF-beta type I receptor. *The Biochemical journal*, 386(Pt 3), pp.461–70.
- Kurokawa, M. et al., 1998. The oncoprotein Evi-1 represses TGF-beta signalling by inhibiting Smad3. *Nature*, 394(6688), pp.92–6.
- Kyttälä, A. et al., 2016. Genetic Variability Overrides the Impact of Parental Cell Type and Determines iPSC Differentiation Potential. *Stem Cell Reports*, 6(2), pp.200–12.
- Ladewig, J. et al., 2012. Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nature methods*, 9(6), pp.575–8.
- Lagna, G. et al., 1996. Partnership between DPC4 and SMAD proteins in TGF-beta signalling pathways. *Nature*, 383(6603), pp.832–6.
- Lambeth, L.S. & Smith, C.A., 2013. Short hairpin RNA-mediated gene silencing. *Methods in Molecular Biology*, 942, pp.205–32.
- Lamouille, S. & Derynck, R., 2007. Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway. *The Journal of cell biology*, 178(3), pp.437–51.
- Landry, J. et al., 2008. Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS genetics*, 4(10), p.e1000241.
- Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4), pp.357–9.
- Laperle, A. et al., 2015. α -5 Laminin Synthesized by Human Pluripotent Stem Cells Promotes Self-Renewal. *Stem Cell Reports*, 5(2), pp.195–206.
- Lawrence, D.A., 1996. Transforming growth factor-beta: a general review. *European cytokine network*, 7(3), pp.363–74.
- Lawson, K.A. et al., 1999. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes & development*, 13(4), pp.424–36.
- Lawson, K.A., 1999. Fate mapping the mouse embryo. *The International journal of developmental biology*, 43(7), pp.773–5.
- Lawson, K.A., Meneses, J.J. & Pedersen, R.A., 1991. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, 113(3), pp.891–911.

- Lee, H.K., Lee, H.S. & Moody, S.A., 2014. Neural transcription factors: from embryos to neural stem cells. *Molecules and cells*, 37(10), pp.705–12.
- Lee, J. et al., 2008. Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *PNAS*, 105(49), pp.19229–34.
- Lee, J.E. et al., 2013. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *eLife*, 2, p.e01503.
- Lee, J.H., Voo, K.S. & Skalnik, D.G., 2001. Identification and characterization of the DNA binding domain of CpG-binding protein. *The Journal of biological chemistry*, 276(48), pp.44669–76.
- Lee, J.S. et al., 2007. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell*, 131(6), pp.1084–96.
- Lee, T.I. et al., 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*, 125(2), pp.301–13.
- Leeb, M. et al., 2010. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes & development*, 24(3), pp.265–76.
- Leeb, M. & Wutz, A., 2007. Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *The Journal of cell biology*, 178(2), pp.219–29.
- Levenstein, M.E. et al., 2006. Basic fibroblast growth factor support of human embryonic stem cell self-renewal. *Stem cells*, 24(3), pp.568–74.
- Levine, A.J., 2005. GDF3, a BMP inhibitor, regulates cell fate in stem cells and early embryos. *Development*, 133(2), pp.209–216.
- Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–9.
- Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), pp.1754–60.
- Li, J.Y. & Joyner, A.L., 2001. Otx2 and Gbx2 are required for refinement and not induction of mid-hindbrain gene expression. *Development*, 128(24), pp.4979–91.
- Li, M. et al., 1998. Generation of purified neural precursors from embryonic stem cells by lineage selection. *Current biology*, 8(17), pp.971–4.
- Li, X., 1998. Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *Journal of Biological Chemistry*, 273(52), pp.34970–34975.
- Li, X., Song, J. & Yi, C., 2014. Genome-wide Mapping of Cellular Protein–RNA Interactions Enabled by Chemical Crosslinking. *Genomics, Proteomics & Bioinformatics*, 12(2), pp.72–78.
- Lian, X. et al., 2012. Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *PNAS*, 109(27), pp.E1848–57.
- Liang, G. et al., 2010. Butyrate promotes induced pluripotent stem cell generation. *The Journal of biological chemistry*, 285(33), pp.25516–21.
- Liao, Y., Smyth, G.K. & Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7), pp.923–30.
- Licht, K. & Jantsch, M.F., 2016. Rapid and dynamic transcriptome regulation by RNA editing and RNA modifications. *The Journal of cell biology*, 213(1), pp.15–22.
- Lien, C.-L. et al., 2002. Cardiac-specific activity of an Nkx2-5 enhancer requires an evolutionarily conserved Smad binding site. *Developmental biology*, 244(2), pp.257–66.
- Lin, Q. et al., 1997. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science*, 276(5317), pp.1404–7.
- Lin, X. et al., 2006. PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. *Cell*, 125(5), pp.915–28.
- Lin, X., Liang, M. & Feng, X.H., 2000. Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor-beta signaling. *The Journal of biological chemistry*, 275(47), pp.36818–22.
- Linder, B. et al., 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nature methods*, 12(8), pp.767–72.
- Lindsley, R.C. et al., 2006. Canonical Wnt signaling is required for development of embryonic stem cell-derived mesoderm. *Development*, 133(19), pp.3787–96.
- Lister, R. et al., 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), pp.315–22.
- Liu, J. et al., 2014. A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. *Nature chemical biology*, 10(2), pp.93–5.

- Liu, N. et al., 2015. N6-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature*, 518(7540), pp.560–564.
- Liu, N. & Pan, T., 2016. N6-methyladenosine–encoded epitranscriptomics. *Nature Structural & Molecular Biology*, 23(2), pp.98–102.
- Liu, P. et al., 1999. Requirement for Wnt3 in vertebrate axis formation. *Nature genetics*, 22(4), pp.361–5.
- Liu, S. et al., 2014. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & development*, 28(18), pp.2041–55.
- Liu, X. et al., 2001. Ski/Sno and TGF-beta signaling. *Cytokine & growth factor reviews*, 12(1), pp.1–8.
- 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(4), pp.402–408.
- Lo, R.S. & Massagué, J., 1999. Ubiquitin-dependent degradation of TGF-beta-activated smad2. *Nature cell biology*, 1(8), pp.472–8.
- Lohnes, D., 2003. The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. *BioEssays*, 25(10), pp.971–80.
- López-Rovira, T. et al., 2000. Interaction and functional cooperation of NF-kappa B with Smads. Transcriptional regulation of the junB promoter. *The Journal of biological chemistry*, 275(37), pp.28937–46.
- De Los Angeles, A. et al., 2015. Hallmarks of pluripotency. *Nature*, 525(7570), pp.469–78.
- Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), p.550.
- Lowe, L.A., Yamada, S. & Kuehn, M.R., 2001. Genetic dissection of nodal function in patterning the mouse embryo. *Development*, 128(10), pp.1831–43.
- Lu, R. et al., 2009. Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature*, 462(7271), pp.358–62.
- Lubitz, S. et al., 2007. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Molecular biology of the cell*, 18(6), pp.2356–66.
- Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389(6648), pp.251–60.
- van der Lugt, N.M. et al., 1994. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes & development*, 8(7), pp.757–69.
- Luo, C. et al., 2016. Improving the Gene Transfection in Human Embryonic Stem Cells: Balancing with Cytotoxicity and Pluripotent Maintenance. *ACS applied materials & interfaces*, 8(13), pp.8367–75.
- Luo, Y. et al., 2014. Stable Enhanced Green Fluorescent Protein Expression After Differentiation and Transplantation of Reporter Human Induced Pluripotent Stem Cells Generated by AAVS1 Transcription Activator-Like Effector Nucleases. *Stem cells translational medicine*, 3(7), pp.821–35.
- Ma, H. et al., 2016. Multiplexed labeling of genomic loci with dCas9 and engineered sgRNAs using CRISPRainbow. *Nature biotechnology*, 34(5), pp.528–530.
- Machnicka, M.A. et al., 2013. MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic acids research*, 41(D1), pp.D262–7.
- Maherali, N. et al., 2007. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell*, 1(1), pp.55–70.
- Maity, A. & Das, B., 2015. N6-methyladenosine modification in mRNA: Machinery, function and implications for health and diseases. *FEBS Journal*, 283(9), pp.1607–30.
- Mali, P. et al., 2010. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. *Stem cells*, 28(4), pp.713–20.
- Mali, P. et al., 2013. RNA-guided human genome engineering via Cas9. *Science*, 339(6121), pp.823–6.
- Manandhar, M., Boulware, K.S. & Wood, R.D., 2015. The ERCC1 and ERCC4 (XPF) genes and gene products. *Gene*, 569(2), pp.153–61.

- Mandal, P.K. et al., 2014. Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. *Cell Stem Cell*, 15(5), pp.643–652.
- Mandegar, M.A. et al., 2016. CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell*, 18(4), pp.541–53.
- Margueron, R. et al., 2009. Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature*, 461(7265), pp.762–7.
- Margueron, R. & Reinberg, D., 2010. Chromatin structure and the inheritance of epigenetic information. *Nature reviews. Genetics*, 11(4), pp.285–96.
- Marks, H. et al., 2012. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, 149(3), pp.590–604.
- Martin, B.L., 2015. Factors that coordinate mesoderm specification from neuromesodermal progenitors with segmentation during vertebrate axial extension. *Seminars in cell & developmental biology*, 49, pp.59–67.
- Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *PNAS*, 78(12), pp.7634–7638.
- Martin, G.R. & Evans, M.J., 1975. Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. *PNAS*, 72(4), pp.1441–5.
- Massagué, J., 2000. How cells read TGF-beta signals. *Nature reviews. Molecular cell biology*, 1(3), pp.169–78.
- Massagué, J., 1998. TGF-beta signal transduction. *Annual review of biochemistry*, 67, pp.753–91.
- Massagué, J., 2012. TGFβ signalling in context. *Nature reviews. Molecular cell biology*, 13(10), pp.616–30.
- Massagué, J., Seoane, J. & Wotton, D., 2005. Smad transcription factors. *Genes & development*, 19(23), pp.2783–810.
- Massagué, J. & Xi, Q., 2012. TGF-β control of stem cell differentiation genes. *FEBS letters*, 586(14), pp.1953–8.
- Massé, J. et al., 2011. ZFP191/Zfp462 is involved in P19 cell pluripotency and in their neuronal fate. *Experimental cell research*, 317(13), pp.1922–34.
- Masui, S. et al., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature cell biology*, 9(6), pp.625–35.
- Matin, M.M. et al., 2004. Specific knockdown of Oct4 and beta2-microglobulin expression by RNA interference in human embryonic stem cells and embryonic carcinoma cells. *Stem cells*, 22(5), pp.659–68.
- McGarvey, K.M. et al., 2008. Defining a chromatin pattern that characterizes DNA-hypermethylated genes in colon cancer cells. *Cancer research*, 68(14), pp.5753–9.
- McLaren, A., 2001. Ethical and social considerations of stem cell research. *Nature*, 414(6859), pp.129–31.
- McLean, A.B. et al., 2007. Activin efficiently specifies definitive endoderm from human embryonic stem cells only when phosphatidylinositol 3-kinase signaling is suppressed. *Stem cells*, 25(1), pp.29–38.
- McLean, C.Y. et al., 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5), pp.495–501.
- Meilhac, S.M. et al., 2009. Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst. *Developmental biology*, 331(2), pp.210–21.
- Meissner, A. et al., 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), pp.766–70.
- Melichar, H.J. et al., 2007. Regulation of gamma delta versus alpha beta T lymphocyte differentiation by the transcription factor SOX13. *Science*, 315(5809), pp.230–3.
- Mendjan, S. et al., 2014. NANOG and CDX2 Pattern Distinct Subtypes of Human Mesoderm during Exit from Pluripotency. *Cell Stem Cell*, 15(3), pp.310–325.
- Menendez, L. et al., 2011. Wnt signaling and a Smad pathway blockade direct the differentiation of human pluripotent stem cells to multipotent neural crest cells. *PNAS*, 108(48), pp.19240–5.
- Meno, C. et al., 1996. Left-right asymmetric expression of the TGF beta-family member lefty in mouse embryos. *Nature*, 381(6578), pp.151–5.
- Meno, C. et al., 1999. Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Molecular cell*, 4(3), pp.287–98.

- Meshorer, E. et al., 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Developmental cell*, 10(1), pp.105–16.
- Mesnard, D., Guzman-Ayala, M. & Constam, D.B., 2006. Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning. *Development*, 133(13), pp.2497–505.
- Meyer, K.D. et al., 2015. 5' UTR m6A Promotes Cap-Independent Translation. *Cell*, 163(4), pp.999–1010.
- Meyer, K.D. et al., 2012. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, 149(7), pp.1635–46.
- Mikkelsen, T.S. et al., 2007. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, 448(7153), pp.553–60.
- Min, I.M. et al., 2011. Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes & development*, 25(7), pp.742–54.
- Mishina, Y. et al., 1995. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes & development*, 9(24), pp.3027–37.
- Mitsui, K. et al., 2003. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*, 113(5), pp.631–42.
- Moffat, J. et al., 2006. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, 124(6), pp.1283–98.
- Mohamed, O.A., Clarke, H.J. & Dufort, D., 2004. Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo. *Developmental dynamics*, 231(2), pp.416–24.
- Monaghan, A.P. et al., 1993. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development*, 119(3), pp.567–78.
- Moretti, A. et al., 2006. Multipotent embryonic isl1+ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell*, 127(6), pp.1151–65.
- Morey, L. et al., 2012. Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell*, 10(1), pp.47–62.
- Morey, L. et al., 2013. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell reports*, 3(1), pp.60–9.
- Morrissey, E.E. et al., 1998. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development*, 12(22), pp.3579–90.
- Moustakas, A. & Heldin, C.H., 2005. Non-Smad TGF-beta signals. *Journal of cell science*, 118, pp.3573–84.
- Mu, Y., Gudey, S.K. & Landström, M., 2012. Non-Smad signaling pathways. *Cell and Tissue Research*, 347(1), pp.11–20.
- Mujtaba, S., Zeng, L. & Zhou, M.M., 2007. Structure and acetyl-lysine recognition of the bromodomain. *Oncogene*, 26(37), pp.5521–7.
- Mukhopadhyay, M. et al., 2001. Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. *Developmental cell*, 1(3), pp.423–34.
- Mullen, A.C. et al., 2011. Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell*, 147(3), pp.565–76.
- Müller, J. et al., 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, 111(2), pp.197–208.
- Muñoz, N.M., Baek, J.Y. & Grady, W.M., 2008. TGF-beta has paradoxical and context dependent effects on proliferation and anoikis in human colorectal cancer cell lines. *Growth factors*, 26(5), pp.254–62.
- Murry, C.E. & Keller, G., 2008. Differentiation of Embryonic Stem Cells to Clinically Relevant Populations: Lessons from Embryonic Development. *Cell*, 132(4), pp.661–680.
- Nagano, Y. et al., 2007. Arkadia induces degradation of SnoN and c-Ski to enhance transforming growth factor-beta signaling. *The Journal of biological chemistry*, 282(28), pp.20492–501.
- Nagy, A. et al., 1993. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *PNAS*, 90(18), pp.8424–8.
- Nagy, A. et al., 1990. Embryonic stem cells alone are able to support fetal development in the mouse. *Development*, 110(3), pp.815–21.

- Naito, A.T. et al., 2006. Developmental stage-specific biphasic roles of Wnt/beta-catenin signaling in cardiomyogenesis and hematopoiesis. *PNAS*, 103(52), pp.19812–7.
- Nakao, A., Afrakhte, M., et al., 1997. Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. *Nature*, 389(6651), pp.631–5.
- Nakao, A., Imamura, T., et al., 1997. TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. *The EMBO journal*, 16(17), pp.5353–62.
- Narlikar, G.J., Sundaramoorthy, R. & Owen-Hughes, T., 2013. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 154(3), pp.490–503.
- Ng, E.S. et al., 2005. Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust, reproducible hematopoietic differentiation. *Blood*, 106(5), pp.1601–3.
- Niakan, K.K. et al., 2012. Human pre-implantation embryo development. *Development*, 139(5), pp.829–841.
- Niakan, K.K. et al., 2010. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes & development*, 24(3), pp.312–26.
- Nichols, J. et al., 1998. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell*, 95(3), pp.379–91.
- Nichols, J. et al., 2001. Physiological rationale for responsiveness of mouse embryonic stem cells to gp130 cytokines. *Development*, 128(12), pp.2333–9.
- Nichols, J. & Smith, A., 2009. Naive and primed pluripotent states. *Cell Stem Cell*, 4(6), pp.487–92.
- Nishioka, N. et al., 2009. The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Developmental cell*, 16(3), pp.398–410.
- Niwa, H. et al., 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*, 123(5), pp.917–29.
- Niwa, H. et al., 1998. Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development*, 12(13), pp.2048–60.
- Niwa, H., Miyazaki, J. & Smith, A.G., 2000. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature genetics*, 24(4), pp.372–6.
- Nomura, T. et al., 1999. Ski is a component of the histone deacetylase complex required for transcriptional repression by Mad and thyroid hormone receptor. *Genes & development*, 13(4), pp.412–23.
- Norrman, K. et al., 2010. Quantitative comparison of constitutive promoters in human ES cells. *PloS one*, 5(8), p.e12413.
- Nostro, M.C. et al., 2008. Wnt, activin, and BMP signaling regulate distinct stages in the developmental pathway from embryonic stem cells to blood. *Cell Stem Cell*, 2(1), pp.60–71.
- van Nuland, R. et al., 2013. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Molecular and cellular biology*, 33(10), pp.2067–77.
- O'Carroll, D. et al., 2001. The polycomb-group gene Ezh2 is required for early mouse development. *Molecular and cellular biology*, 21(13), pp.4330–6.
- Oike, Y. et al., 1999. Mice homozygous for a truncated form of CREB-binding protein exhibit defects in hematopoiesis and vasculo-angiogenesis. *Blood*, 93(9), pp.2771–9.
- Ordovas, L. et al., 2015. Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition. *Stem Cell Reports*, 5(5), pp.918–931.
- Oshimori, N. & Fuchs, E., 2012. The harmonies played by TGF-β in stem cell biology. *Cell Stem Cell*, 11(6), pp.751–64.
- Ott, M.G. et al., 2006. Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nature medicine*, 12(4), pp.401–9.
- Ozair, M.Z., Kintner, C. & Brivanlou, A.H., 2013. Neural induction and early patterning in vertebrates. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(4), pp.479–498.

- Ozdamar, B. et al., 2005. Regulation of the polarity protein Par6 by TGFbeta receptors controls epithelial cell plasticity. *Science*, 307(5715), pp.1603–9.
- Palazzani, L., 2011. Embryo research in Italy: the bioethical and biojuridical debate. *Human reproduction and genetic ethics*, 17(1), pp.28–39.
- Palmieri, S.L. et al., 1994. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Developmental biology*, 166(1), pp.259–67.
- Palpant, N.J. et al., 2015. Inhibition of β -catenin signaling respecifies anterior-like endothelium into beating human cardiomyocytes. *Development*, 142(18), pp.3198–209.
- Pan, G. et al., 2007. Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell*, 1(3), pp.299–312.
- Pang, K. et al., 2011. Evolution of the TGF- β signaling pathway and its potential role in the ctenophore, *Mnemiopsis leidyi*. *PloS one*, 6(9), p.e24152.
- Pankratz, M.T. et al., 2007. Directed Neural Differentiation of Human Embryonic Stem Cells via an Obligated Primitive Anterior Stage. *Stem Cells*, 25(6), pp.1511–1520.
- Papanayotou, C. et al., 2014. A Novel Nodal Enhancer Dependent on Pluripotency Factors and Smad2/3 Signaling Conditions a Regulatory Switch During Epiblast Maturation. *PLoS biology*, 12(6), p.e1001890.
- Pardo, M. et al., 2010. An expanded Oct4 interaction network: implications for stem cell biology, development, and disease. *Cell Stem Cell*, 6(4), pp.382–95.
- Parfitt, D.E. & Shen, M.M., 2014. From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis. *Philosophical transactions of the Royal Society of London*, 369(1657), p.ii: 20130542.
- Park, C. et al., 2004. A hierarchical order of factors in the generation of FLK1- and SCL-expressing hematopoietic and endothelial progenitors from embryonic stem cells. *Development*, 131(11), pp.2749–62.
- Park, S.H. et al., 2004. Ultrastructure of human embryonic stem cells and spontaneous and retinoic acid-induced differentiating cells. *Ultrastructural pathology*, 28(4), pp.229–38.
- Parthun, M.R., 2007. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene*, 26(37), pp.5319–28.
- Pasini, D. et al., 2004. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO journal*, 23(20), pp.4061–71.
- Pastor, W.A. et al., 2016. Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell Stem Cell*, 18(3), pp.323–329.
- Patani, R. et al., 2009. Activin/Nodal inhibition alone accelerates highly efficient neural conversion from human embryonic stem cells and imposes a caudal positional identity. *PloS one*, 4(10), p.e7327.
- Patsch, C. et al., 2015. Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. *Nature cell biology*, 17(8), pp.994–1003.
- Pauklin, S., Pedersen, R.A. & Vallier, L., 2011. Mouse pluripotent stem cells at a glance. *Journal of cell science*, 124, pp.3727–32.
- Pauklin, S. & Vallier, L., 2015. Activin/Nodal signalling in stem cells. *Development*, 142(4), pp.607–619.
- Pauklin, S. & Vallier, L., 2013. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. *Cell*, 155(1), pp.135–147.
- Pearson, K.L., Hunter, T. & Janknecht, R., 1999. Activation of Smad1-mediated transcription by p300/CBP. *Biochimica et biophysica acta*, 1489(2-3), pp.354–64.
- Perea-Gomez, A. et al., 2002. Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Developmental cell*, 3(5), pp.745–56.
- Perea-Gomez, A. et al., 2001. Otx2 is required for visceral endoderm movement and for the restriction of posterior signals in the epiblast of the mouse embryo. *Development*, 128(5), pp.753–65.
- Perry, R.P. & Kelley, D.E., 1974. Existence of methylated messenger RNA in mouse L cells. *Cell*, 1(1), pp.37–42.
- Petropoulos, S. et al., 2016. Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*, 165(4), pp.1012–26.
- Pevny, L.H. et al., 1998. A role for SOX1 in neural determination. *Development*, 125(10), pp.1967–78.

- Piccolo, S. et al., 1999. The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature*, 397(6721), pp.707–10.
- Ping, X.-L. et al., 2014. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell research*, 24(2), pp.177–89.
- Plaisier, S.B. et al., 2010. Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic acids research*, 38(17), p.e169.
- Pohl, A. & Beato, M., 2014. bwtool: a tool for bigWig files. *Bioinformatics*, 30(11), pp.1618–9.
- Posfai, E., Tam, O.H. & Rossant, J., 2014. Mechanisms of Pluripotency In Vivo and In Vitro. *Current topics in developmental biology*, 107, pp.1–37.
- Postigo, A.A. et al., 2003. Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins. *The EMBO journal*, 22(10), pp.2453–62.
- Pouponnot, C., Jayaraman, L. & Massagué, J., 1998. Physical and functional interaction of SMADs and p300/CBP. *The Journal of biological chemistry*, 273(36), pp.22865–8.
- Pourquié, O. et al., 2015. Looking inwards: opening a window onto human development. *Development*, 142(1), pp.1–2.
- Pratt, T. et al., 2000. Embryonic stem cells and transgenic mice ubiquitously expressing a tau-tagged green fluorescent protein. *Developmental biology*, 228(1), pp.19–28.
- Quinlan, A.R. & Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), pp.841–2.
- Rada-Iglesias, A. et al., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), pp.279–83.
- Rajagopalan, L.E. et al., 1998. hnRNP C increases amyloid precursor protein (APP) production by stabilizing APP mRNA. *Nucleic acids research*, 26(14), pp.3418–23.
- Ralston, A. & Rossant, J., 2008. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophoblast fate in the early mouse embryo. *Developmental biology*, 313(2), pp.614–29.
- Ran, F.A. et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nature protocols*, 8(11), pp.2281–308.
- Rappsilber, J., Mann, M. & Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature protocols*, 2(8), pp.1896–906.
- Rashid, S.T. et al., 2010. Technical advance Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *The Journal of clinical investigation*, 120(9), pp.3127–3136.
- Raya, A. et al., 2009. Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature*, 460(7251), pp.53–9.
- Reissmann, E. et al., 2001. The orphan receptor ALK7 and the Activin receptor ALK4 mediate signaling by Nodal proteins during vertebrate development. *Genes & development*, 15(15), pp.2010–22.
- Rivera-Pérez, J.A. et al., 1995. Goosecoid is not an essential component of the mouse gastrula organizer but is required for craniofacial and rib development. *Development*, 121(9), pp.3005–12.
- De Robertis, E., 2004. Goosecoid and gastrulation. In *Gastrulation*. Cold Spring Harbor Laboratory Press, pp. 581–589.
- 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(1), pp.139–40.
- Rodaway, A. & Patient, R., 2001. Mesendoderm. an ancient germ layer? *Cell*, 105(2), pp.169–72.
- Rodriguez, J. et al., 2008. Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. *PNAS*, 105(50), pp.19809–14.
- Roh, T.-Y. et al., 2006. The genomic landscape of histone modifications in human T cells. *PNAS*, 103(43), pp.15782–7.
- Roost, C. et al., 2015. Structure and thermodynamics of N6-methyladenosine in RNA: a spring-loaded base modification. *Journal of the American Chemical Society*, 137(5), pp.2107–15.
- Rosenthal, M.D., Wishnow, R.M. & Sato, G.H., 1970. In vitro growth and differentiation of clonal populations of multipotential mouse cells derived from a transplantable testicular teratocarcinoma. *Journal of the National Cancer Institute*, 44(5), pp.1001–14.

- Rosner, M.H. et al., 1990. A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature*, 345(6277), pp.686–92.
- Ross, S. et al., 2006. Smads orchestrate specific histone modifications and chromatin remodeling to activate transcription. *The EMBO journal*, 25(19), pp.4490–502.
- Ross, S. & Hill, C.S., 2008. How the Smads regulate transcription. *The international journal of biochemistry & cell biology*, 40(3), pp.383–408.
- Rossant, J. & Tam, P.P.L., 2009. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development*, 136(5), pp.701–713.
- Rotem, A. et al., 2015. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature biotechnology*, 33(11), pp.1165–1172.
- Rüegsegger, U., Beyer, K. & Keller, W., 1996. Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *The Journal of biological chemistry*, 271(11), pp.6107–13.
- Russ, A.P. et al., 2000. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature*, 404(6773), pp.95–9.
- Ruthenburg, A.J., Allis, C.D. & Wysocka, J., 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell*, 25(1), pp.15–30.
- Ryder, E. et al., 2013. Molecular characterization of mutant mouse strains generated from the EUCOMM/KOMP-CSD ES cell resource. *Mammalian genome*, 24(7-8), pp.286–94.
- Sadelain, M., Papapetrou, E.P. & Bushman, F.D., 2012. Safe harbours for the integration of new DNA in the human genome. *Nature reviews. Cancer*, 12(1), pp.51–8.
- Saga, Y. et al., 1999. MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*, 126(15), pp.3437–47.
- Sakaki-Yumoto, M. et al., 2013. Smad2 is essential for maintenance of the human and mouse primed pluripotent stem cell state. *Journal of Biological Chemistry*, 288(25), pp.18546–18560.
- Saletore, Y. et al., 2012. The birth of the Epitranscriptome: deciphering the function of RNA modifications. *Genome Biology*, 13(10), p.175.
- Sambrook, J. & Russell, D.W., 2006. Identification of associated proteins by coimmunoprecipitation. *CSH protocols*, 2006(1), p.pdb.prot3898.
- Sampaziotis, F. et al., 2015. Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. *Nature biotechnology*, 33(8), pp.845–852.
- Santos-Rosa, H. et al., 2002. Active genes are tri-methylated at K4 of histone H3. *Nature*, 419(6905), pp.407–11.
- Sasaki, H. & Hogan, B.L., 1993. Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development*, 118(1), pp.47–59.
- Savory, J.G.A. et al., 2009. Cdx2 regulation of posterior development through non-Hox targets. *Development*, 136(24), pp.4099–110.
- Schaniel, C. et al., 2009. Smarcc1/Baf155 couples self-renewal gene repression with changes in chromatin structure in mouse embryonic stem cells. *Stem cells*, 27(12), pp.2979–91.
- Schier, A.F., 2003. Nodal signaling in vertebrate development. *Annual review of cell and developmental biology*, 19, pp.589–621.
- Schmitz, S.U. et al., 2011. Jarid1b targets genes regulating development and is involved in neural differentiation. *The EMBO journal*, 30(22), pp.4586–600.
- Schöler, H.R. et al., 1990. New type of POU domain in germ line-specific protein Oct-4. *Nature*, 344(6265), pp.435–9.
- Schuettengruber, B. et al., 2007. Genome regulation by polycomb and trithorax proteins. *Cell*, 128(4), pp.735–45.
- Schultz, D.C. et al., 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & development*, 16(8), pp.919–32.
- Schwartz, S. et al., 2014. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell reports*, 8(1), pp.284–96.
- Séguin, C.A. et al., 2008. Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells. *Cell Stem Cell*, 3(2), pp.182–95.
- Seo, S.R. et al., 2004. The novel E3 ubiquitin ligase Tiul1 associates with TGIF to target Smad2 for degradation. *The EMBO journal*, 23(19), pp.3780–92.

- Seoane, J. et al., 2004. Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell*, 117(2), pp.211–23.
- Shahbazi, M.N. et al., 2016. Self-organization of the human embryo in the absence of maternal tissues. *Nature cell biology*, 18(6), pp.700–8.
- Shannon, P. et al., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), pp.2498–504.
- Shearwin, K.E., Callen, B.P. & Egan, J.B., 2005. Transcriptional interference--a crash course. *Trends in genetics*, 21(6), pp.339–45.
- Shen, L. et al., 2013. diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates. *PLoS one*, 8(6), p.e65598.
- Shen, X. et al., 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell*, 32(4), pp.491–502.
- Shi, Y. et al., 1998. Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA binding in TGF-beta signaling. *Cell*, 94(5), pp.585–94.
- Shi, Y. & Massagué, J., 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*, 113(6), pp.685–700.
- Shikama, N. et al., 2003. Essential function of p300 acetyltransferase activity in heart, lung and small intestine formation. *The EMBO journal*, 22(19), pp.5175–85.
- Shilatifard, A., 2012. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry*, 81, pp.65–95.
- Shinkai, Y. & Tachibana, M., 2011. H3K9 methyltransferase G9a and the related molecule GLP. *Genes & development*, 25(8), pp.781–8.
- Shogren-Knaak, M. et al., 2006. Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, 311(5762), pp.844–7.
- Simonsson, M. et al., 2005. The balance between acetylation and deacetylation controls Smad7 stability. *The Journal of biological chemistry*, 280(23), pp.21797–803.
- Simonsson, M. et al., 2006. The DNA binding activities of Smad2 and Smad3 are regulated by coactivator-mediated acetylation. *The Journal of biological chemistry*, 281(52), pp.39870–80.
- Singh, A.M. et al., 2007. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells*, 25(10), pp.2534–42.
- Singh, A.M. et al., 2012. Signaling network crosstalk in human pluripotent cells: a Smad2/3-regulated switch that controls the balance between self-renewal and differentiation. *Cell Stem Cell*, 10(3), pp.312–26.
- Skarnes, W.C. et al., 2011. A conditional knockout resource for the genome-wide study of mouse gene function. *Nature*, 474(7351), pp.337–42.
- Smith, A.G. et al., 1988. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, 336(6200), pp.688–90.
- Smith, J.R., Vallier, L., et al., 2008. Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Developmental biology*, 313(1), pp.107–17.
- Smith, J.R., Maguire, S., et al., 2008. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem cells*, 26(2), pp.496–504.
- Smith, Z.D. & Meissner, A., 2013. DNA methylation: roles in mammalian development. *Nature reviews. Genetics*, 14(3), pp.204–20.
- Smyth, G.K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3(1), pp.1–25.
- Song, J. et al., 2012. DNA and chromatin modification networks distinguish stem cell pluripotent ground states. *Molecular & cellular proteomics*, 11(10), pp.1036–47.
- Sorrentino, A. et al., 2008. The type I TGF-beta receptor engages TRAF6 to activate TAK1 in a receptor kinase-independent manner. *Nature cell biology*, 10(10), pp.1199–207.
- Souchelnyskiy, S. et al., 1997. Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. *The Journal of biological chemistry*, 272(44), pp.28107–15.

- Squazzo, S.L. et al., 2006. Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome research*, 16(7), pp.890–900.
- Stavridis, M.P. et al., 2007. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. *Development*, 134(16), pp.2889–94.
- Stein, S. et al., 2010. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nature medicine*, 16(2), pp.198–204.
- Stewart, C.L. et al., 1992. Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature*, 359(6390), pp.76–9.
- Stoller, J.Z. et al., 2010. Ash2l interacts with Tbx1 and is required during early embryogenesis. *Experimental biology and medicine*, 235(5), pp.569–76.
- Stouffer, G.A. & Owens, G.K., 1994. TGF-beta promotes proliferation of cultured SMC via both PDGF-AA-dependent and PDGF-AA-independent mechanisms. *The Journal of clinical investigation*, 93(5), pp.2048–55.
- Stroschein, S.L. et al., 1999. Negative feedback regulation of TGF-beta signaling by the SnoN oncoprotein. *Science*, 286(5440), pp.771–4.
- Subramanian, A. et al., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102(43), pp.15545–15550.
- Sumi, T. et al., 2008. Defining early lineage specification of human embryonic stem cells by the orchestrated balance of canonical Wnt/beta-catenin, Activin/Nodal and BMP signaling. *Development*, 135(17), pp.2969–79.
- Sumi, T. et al., 2013. Epiblast ground state is controlled by canonical Wnt/ β -catenin signaling in the postimplantation mouse embryo and epiblast stem cells. *PloS one*, 8(5), p.e63378.
- Sun, L.T. et al., 2014. Nanog co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast. *Developmental biology*, 392(2), pp.182–92.
- Sun, Y. et al., 1999. SnoN and Ski protooncoproteins are rapidly degraded in response to transforming growth factor beta signaling. *PNAS*, 96(22), pp.12442–7.
- Tahmasebi, S. et al., 2014. Multifaceted regulation of somatic cell reprogramming by mRNA translational control. *Cell Stem Cell*, 14(5), pp.606–16.
- Takahashi, K. et al., 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5), pp.861–72.
- Takahashi, K. & Yamanaka, S., 2016. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology*, 17(3), pp.183–193.
- Takahashi, K. & Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4), pp.663–76.
- Takashima, Y. et al., 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, 158(6), pp.1254–69.
- Tam, P.P. et al., 1997. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development*, 124(9), pp.1631–42.
- Tam, P.P. & Trainor, P.A., 1994. Specification and segmentation of the paraxial mesoderm. *Anatomy and embryology*, 189(4), pp.275–305.
- Tam, P.P.L. et al., 2007. Sequential allocation and global pattern of movement of the definitive endoderm in the mouse embryo during gastrulation. *Development*, 134(2), pp.251–60.
- Tam, P.P.L. & Loebel, D. a F., 2007. Gene function in mouse embryogenesis: get set for gastrulation. *Nature reviews. Genetics*, 8(5), pp.368–81.
- Tang, K. et al., 2015. Intrinsic regulations in neural fate commitment. *Development Growth and Differentiation*, 57(2), pp.109–120.
- Teo, A. et al., 2011. Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes & development*, 25(2), pp.238–250.
- Tesar, P.J. et al., 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150), pp.196–9.
- Tessarz, P. & Kouzarides, T., 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nature reviews. Molecular cell biology*, 15(11), pp.703–8.
- Theunissen, T.W. et al., 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell Stem Cell*, 15(4), pp.471–87.
- Thomas, K.R. & Capecchi, M.R., 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3), pp.503–12.

- Thomas, P. & Beddington, R., 1996. Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Current biology*, 6(11), pp.1487–96.
- Thomson, J. a., 1998. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science*, 282(5391), pp.1145–1147.
- Thomson, M. et al., 2011. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*, 145(6), pp.875–89.
- Touboul, T. et al., 2010. Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology*, 51(5), pp.1754–1765.
- Toyooka, Y. et al., 2008. Identification and characterization of subpopulations in undifferentiated ES cell culture. *Development*, 135(5), pp.909–18.
- Tremblay, K.D. et al., 2000. Formation of the definitive endoderm in mouse is a Smad2-dependent process. *Development*, 127(14), pp.3079–90.
- Trompouki, E. et al., 2011. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell*, 147(3), pp.577–89.
- Trounson, A. & DeWitt, N.D., 2016. Pluripotent stem cells progressing to the clinic. *Nature Reviews Molecular Cell Biology*, 17(3), pp.194–200.
- Truax, A.D. & Greer, S.F., 2012. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods in Molecular Biology*, 809, pp.175–188.
- Tsuchida, K. et al., 2004. Activin isoforms signal through type I receptor serine/threonine kinase ALK7. *Molecular and cellular endocrinology*, 220(1-2), pp.59–65.
- Tsukazaki, T. et al., 1998. SARA, a FYVE domain protein that recruits Smad2 to the TGFbeta receptor. *Cell*, 95(6), pp.779–91.
- Tsuneyoshi, N. et al., 2012. The SMAD2/3 corepressor SNON maintains pluripotency through selective repression of mesendodermal genes in human ES cells. *Genes & development*, 26(22), pp.2471–6.
- Tsuzuki, S. & Seto, M., 2013. TEL (ETV6)-AML1 (RUNX1) initiates self-renewing fetal pro-B cells in association with a transcriptional program shared with embryonic stem cells in mice. *Stem cells*, 31(2), pp.236–47.
- Tu, A.W. & Luo, K., 2007. Acetylation of Smad2 by the co-activator p300 regulates activin and transforming growth factor beta response. *The Journal of biological chemistry*, 282(29), pp.21187–96.
- Uchida, N., Hoshino, S.I. & Katada, T., 2004. Identification of a human cytoplasmic poly(A) nuclease complex stimulated by poly(A)-binding protein. *The Journal of biological chemistry*, 279(2), pp.1383–91.
- Ueno, S. et al., 2007. Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *PNAS*, 104(23), pp.9685–90.
- Urbach, A., Schuldiner, M. & Benvenisty, N., 2004. Modeling for Lesch-Nyhan disease by gene targeting in human embryonic stem cells. *Stem cells*, 22(4), pp.635–41.
- Urnov, F.D. et al., 2010. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*, 11(9), pp.636–46.
- Valenzuela, D.M. et al., 2003. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nature biotechnology*, 21(6), pp.652–9.
- Vallier, L., Mendjan, S., et al., 2009. Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development*, 136(8), pp.1339–49.
- Vallier, L., Touboul, T., Chng, Z., et al., 2009. Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PLoS one*, 4(6), p.e6082.
- Vallier, L., Rugg-Gunn, P.J., et al., 2004. Enhancing and diminishing gene function in human embryonic stem cells. *Stem cells*, 22(1), pp.2–11.
- Vallier, L., 2011. Serum-Free and Feeder-Free Culture Conditions for Human Embryonic Stem Cells. *Springer protocols*, 690(1), pp.57–66.
- Vallier, L., Touboul, T., Brown, S., et al., 2009. Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem cells*, 27(11), pp.2655–66.

- Vallier, L., Alexander, M. & Pedersen, R.A., 2005. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *Journal of cell science*, 118, pp.4495–509.
- Vallier, L., Alexander, M. & Pedersen, R.A., 2007. Conditional gene expression in human embryonic stem cells. *Stem cells*, 25(6), pp.1490–7.
- Vallier, L., Reynolds, D. & Pedersen, R.A., 2004. Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Developmental biology*, 275(2), pp.403–21.
- Varelas, X. et al., 2010. The Crumbs complex couples cell density sensing to Hippo-dependent control of the TGF- β -SMAD pathway. *Developmental cell*, 19(6), pp.831–44.
- Varlet, I., Collignon, J. & Robertson, E.J., 1997. Nodal Expression in the Primitive Endoderm Is Required for Specification of the Anterior Axis During Mouse Gastrulation. *Development*, 124(5), pp.1033–1044.
- Vastenhouw, N.L. & Schier, A.F., 2012. Bivalent histone modifications in early embryogenesis. *Current opinion in cell biology*, 24(3), pp.374–86.
- Vermeulen, M. et al., 2010. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, 142(6), pp.967–80.
- Vermeulen, M. et al., 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, 131(1), pp.58–69.
- Vincent, S.D. et al., 2003. Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes & development*, 17(13), pp.1646–1662.
- Vogel, G., 2001. Biomedical policy. Bush squeezes between the lines on stem cells. *Science*, 293(5533), pp.1242–5.
- Voigt, P., Tee, W.-W. & Reinberg, D., 2013. A double take on bivalent promoters. *Genes & development*, 27(12), pp.1318–38.
- Vojta, A. et al., 2016. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic acids research*, in press.
- Voskas, D., Ling, L.S. & Woodgett, J.R., 2014. Signals controlling un-differentiated states in embryonic stem and cancer cells: role of the phosphatidylinositol 3' kinase pathway. *Journal of cellular physiology*, 229(10), pp.1312–22.
- Vujovic, P. et al., 2013. Fasting induced cytoplasmic Fto expression in some neurons of rat hypothalamus. *PloS one*, 8(5), p.e63694.
- Waldrip, W.R. et al., 1998. Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. *Cell*, 92(6), pp.797–808.
- Waldrip, Z.J. et al., 2014. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics*, 9(9), pp.1207–11.
- Wan, M. et al., 2013. The trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells. *The Journal of biological chemistry*, 288(7), pp.5039–48.
- Wang, G. et al., 2005. Noggin and bFGF cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers. *Biochemical and Biophysical Research Communications*, 330(3), pp.934–942.
- Wang, H. et al., 2004. Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, 431(7010), pp.873–8.
- Wang, J. et al., 2006. A protein interaction network for pluripotency of embryonic stem cells. *Nature*, 444(7117), pp.364–8.
- Wang, J., Levasseur, D.N. & Orkin, S.H., 2008. Requirement of Nanog dimerization for stem cell self-renewal and pluripotency. *PNAS*, 105(17), pp.6326–31.
- Wang, L., Wang, S. & Li, W., 2012. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, 28(16), pp.2184–5.
- Wang, P. et al., 2009. Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Molecular and cellular biology*, 29(22), pp.6074–85.
- Wang, X. et al., 2015. N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*, 161(6), pp.1388–1399.
- Wang, X. et al., 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, 505(7481), pp.117–20.

- Wang, Y. et al., 2014. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nature cell biology*, 16(2), pp.191–8.
- Wang, Y., Bagheri-Fam, S. & Harley, V.R., 2005. SOX13 is up-regulated in the developing mouse neuroepithelium and identifies a sub-population of differentiating neurons. *Developmental brain research*, 157(2), pp.201–8.
- Wang, Y.C., Peterson, S.E. & Loring, J.F., 2014. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell research*, 24(2), pp.143–60.
- Wang, Z. et al., 2012. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*, 10(4), pp.440–54.
- Watanabe, K. et al., 2007. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nature biotechnology*, 25(6), pp.681–6.
- Wei, C.M., Gershowitz, A. & Moss, B., 1975. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell*, 4(4), pp.379–86.
- Weinberger, L. et al., 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nature Reviews Molecular Cell Biology*, 17(3), pp.155–169.
- Weisberg, E. et al., 1998. A mouse homologue of FAST-1 transduces TGF beta superfamily signals and is expressed during early embryogenesis. *Mechanisms of development*, 79(1-2), pp.17–27.
- Welt, C. et al., 2002. Activins, Inhibins, and Follistatins: From Endocrinology to Signaling. A Paradigm for the New Millennium. *Experimental Biology and Medicine*, 227(9), pp.724–752.
- Wen, B. et al., 2009. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature genetics*, 41(2), pp.246–50.
- White, J.K. et al., 2013. Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell*, 154(2), pp.452–64.
- Wiles, M. V & Johansson, B.M., 1999. Embryonic stem cell development in a chemically defined medium. *Experimental cell research*, 247(1), pp.241–8.
- Willems, E. & Leyns, L., 2008. Patterning of mouse embryonic stem cell-derived pan-mesoderm by Activin A/Nodal and Bmp4 signaling requires Fibroblast Growth Factor activity. *Differentiation*, 76(7), pp.745–59.
- Williams, R.L. et al., 1988. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, 336(6200), pp.684–7.
- Wilson, V. et al., 1995. The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development*, 121(3), pp.877–86.
- Wong, C. et al., 1999. Smad3-Smad4 and AP-1 complexes synergize in transcriptional activation of the c-Jun promoter by transforming growth factor beta. *Molecular and cellular biology*, 19(3), pp.1821–30.
- Wood, H.B. & Episkopou, V., 1999. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of development*, 86(1-2), pp.197–201.
- Wotton, D. et al., 1999. Multiple modes of repression by the Smad transcriptional corepressor TGIF. *The Journal of biological chemistry*, 274(52), pp.37105–10.
- Wrana, J.L. et al., 1994. Mechanism of activation of the TGF-beta receptor. *Nature*, 370(6488), pp.341–7.
- Wrana, J.L. et al., 1992. TGFβ signals through a heteromeric protein kinase receptor complex. *Cell*, 71(6), pp.1003–1014.
- Wright, A.V., Nuñez, J.K. & Doudna, J.A., 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, 164(1-2), pp.29–44.
- Wright, J.E. & Ciosk, R., 2013. RNA-based regulation of pluripotency. *Trends in genetics*, 29(2), pp.99–107.
- Wu, M.Y. & Hill, C.S., 2009. Tgf-beta superfamily signaling in embryonic development and homeostasis. *Developmental cell*, 16(3), pp.329–43.
- Wu, P.-C. et al., 2014. H3K9 Histone Methyltransferase, KMT1E/SETDB1, Cooperates with the SMAD2/3 Pathway to Suppress Lung Cancer Metastasis. *Cancer research*, pp.1–12.
- Wu, S.M. et al., 2006. Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell*, 127(6), pp.1137–50.
- Wysocka, J. et al., 2005. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6), pp.859–72.

- Xi, Q. et al., 2011. A poised chromatin platform for TGF- β access to master regulators. *Cell*, 147(7), pp.1511–24.
- Xiao, L., Yuan, X. & Sharkis, S.J., 2006. Activin A maintains self-renewal and regulates fibroblast growth factor, Wnt, and bone morphogenic protein pathways in human embryonic stem cells. *Stem cells*, 24(6), pp.1476–86.
- Xiao, W. et al., 2016. Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Molecular cell*, 61(4), pp.507–19.
- Xiao, Z., Liu, X. & Lodish, H.F., 2000. Importin beta mediates nuclear translocation of Smad 3. *The Journal of biological chemistry*, 275(31), pp.23425–8.
- Xu, C. et al., 2005. Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem cells*, 23(3), pp.315–23.
- Xu, C. et al., 2014. Structural basis for selective binding of m6A RNA by the YTHDC1 YTH domain. *Nature chemical biology*, 10(11), pp.927–9.
- Xu, H. et al., 2010. A signal-noise model for significance analysis of ChIP-seq with negative control. *Bioinformatics*, 26(9), pp.1199–204.
- Xu, L. et al., 2002. Smad2 nucleocytoplasmic shuttling by nucleoporins CAN/Nup214 and Nup153 feeds TGFbeta signaling complexes in the cytoplasm and nucleus. *Molecular cell*, 10(2), pp.271–82.
- Xu, L., Chen, Y.G. & Massagué, J., 2000. The nuclear import function of Smad2 is masked by SARA and unmasked by TGFbeta-dependent phosphorylation. *Nature cell biology*, 2(8), pp.559–62.
- Xu, P., Liu, J. & Derynck, R., 2012. Post-translational regulation of TGF- β receptor and Smad signaling. *FEBS letters*, 586(14), pp.1871–1884.
- Xu, R.H. et al., 2002. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature biotechnology*, 20(12), pp.1261–4.
- Xu, R.H. et al., 2008. NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell Stem Cell*, 3(2), pp.196–206.
- Yagi, H. et al., 1998. Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood*, 92(1), pp.108–17.
- Yagi, K. et al., 1999. Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. *The Journal of biological chemistry*, 274(2), pp.703–9.
- Yamada, G. et al., 1995. Targeted mutation of the murine goosecoid gene results in craniofacial defects and neonatal death. *Development*, 121(9), pp.2917–22.
- Yamaguchi, T.P., 2001. Heads or tails: Wnts and anterior-posterior patterning. *Current biology*, 11(17), pp.R713–24.
- Yamamoto, M. et al., 2004. Nodal antagonists regulate formation of the anteroposterior axis of the mouse embryo. *Nature*, 428(6981), pp.387–92.
- Yamamoto, M. et al., 2001. The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes & development*, 15(10), pp.1242–56.
- Yamashita, M. et al., 2008. TRAF6 mediates Smad-independent activation of JNK and p38 by TGF-beta. *Molecular cell*, 31(6), pp.918–24.
- Yang, Z. et al., 2014. The DPY30 subunit in SET1/MLL complexes regulates the proliferation and differentiation of hematopoietic progenitor cells. *Blood*, 124(13), pp.2025–2.
- Yao, S. et al., 2006. Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. *PNAS*, 103(18), pp.6907–12.
- Yao, S. et al., 2004. Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. *Molecular therapy*, 10(1), pp.27–36.
- Yao, T.P. et al., 1998. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell*, 93(3), pp.361–72.
- Ye, J. & Blelloch, R., 2014. Regulation of Pluripotency by RNA Binding Proteins. *Cell Stem Cell*, 15(3), pp.271–280.
- Ye, S.C. et al., 1999. Contextual effects of transforming growth factor beta on the tumorigenicity of human colon carcinoma cells. *Cancer research*, 59(18), pp.4725–31.
- Yearim, A. et al., 2015. HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell reports*, 10(7), pp.1122–34.

- Yi, J.Y., Shin, I. & Arteaga, C.L., 2005. Type I transforming growth factor beta receptor binds to and activates phosphatidylinositol 3-kinase. *The Journal of biological chemistry*, 280(11), pp.10870–6.
- Ying, Q.L. et al., 2003. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, 115(3), pp.281–92.
- Ying, Q.L. et al., 2008. The ground state of embryonic stem cell self-renewal. *Nature*, 453(7194), pp.519–23.
- Young, R. a, 2011. Control of the embryonic stem cell state. *Cell*, 144(6), pp.940–54.
- Yu, B.D. et al., 1995. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*, 378(6556), pp.505–8.
- Yu, J. et al., 2010. MTMR4 attenuates transforming growth factor beta (TGFbeta) signaling by dephosphorylating R-Smads in endosomes. *The Journal of biological chemistry*, 285(11), pp.8454–62.
- Yuan, P. et al., 2009. Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes & development*, 23(21), pp.2507–20.
- Yue, Y., Liu, J. & He, C., 2015. RNA N⁶-methyladenosine methylation in post-transcriptional gene expression regulation. *Genes & development*, pp.1343–1355.
- Yun, M. et al., 2011. Readers of histone modifications. *Cell research*, 21(4), pp.564–78.
- Yusa, K. et al., 2011. Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, 478(7369), pp.391–4.
- Zafarana, G. et al., 2009. Specific knockdown of OCT4 in human embryonic stem cells by inducible short hairpin RNA interference. *Stem cells*, 27(4), pp.776–82.
- Zappone, M. V. et al., 2000. Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. *Development*, 127(11), pp.2367–82.
- Zawel, L. et al., 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Molecular cell*, 1(4), pp.611–7.
- Zhang, S.C. et al., 2001. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nature biotechnology*, 19(12), pp.1129–33.
- Zhang, Y. et al., 1996. Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. *Nature*, 383(6596), pp.168–72.
- Zhang, Y. et al., 2001. Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase. *PNAS*, 98(3), pp.974–9.
- Zhang, Y.E., 2009. Non-Smad pathways in TGF-beta signaling. *Cell research*, 19(1), pp.128–39.
- Zhao, X.D. et al., 2007. Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell*, 1(3), pp.286–98.
- Zheng, G. et al., 2013. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Molecular cell*, 49(1), pp.18–29.
- Zhou, J. et al., 2015. Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature*, 526(7574), pp.591–4.
- Zhou, K.I. et al., 2016. N(6)-Methyladenosine Modification in a Long Noncoding RNA Hairpin Predisposes Its Conformation to Protein Binding. *Journal of molecular biology*, 428(5 Pt A), pp.822–33.
- Zhu, J. et al., 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell*, 152(3), pp.642–54.
- Zhu, L.J. et al., 2010. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC bioinformatics*, 11(1), p.237.
- Zhu, Z. & Huangfu, D., 2013. Human pluripotent stem cells: an emerging model in developmental biology. *Development*, 140(4), pp.705–17.
- Zwaka, T.P. & Thomson, J.A., 2003. Homologous recombination in human embryonic stem cells. *Nature biotechnology*, 21(3), pp.319–21.