

L-Proline-mediated mechanisms driving neural differentiation from mouse embryonic stem cells

Sally Vuong

430110650



Supervisor: Dr. Michael B. Morris

This thesis is submitted in fulfilment of the requirements for the
degree of

Doctor of Philosophy

School of Medical Sciences
Faculty of Medicine and Health
University of Sydney, Australia

2024

Declaration of Authorship

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Signature

Name

Sally Vuong

Abstract

Development of the nervous system in the mammalian embryo requires inner cell mass cells in the blastocyst to transition to a second pluripotent population, primitive ectoderm, then the multipotent germ layer, definitive ectoderm, and then the initial precursor of the nervous system, neurectoderm, before differentiating into mature neural cells. The steps that produce these cell types depend on complex molecular mechanisms that are tightly controlled in terms of activity, timing and location. Early neural development in mammals, however, remains poorly understood in part due to the relatively inaccessible environment of the embryo.

Amino acids and metabolic processes are now recognised as part of the complex system of mechanisms required to drive cell fate. In particular, L-Proline acts as a growth factor driving progression along the pluripotency spectrum and neural lineage commitment by (i) activating signalling pathways such as mTOR and MAPK/ERK and (ii) through its metabolism.

This thesis investigates various mechanisms driving pluripotent cells to commit to the neural lineage using mouse ESCs as an *in vitro* model of mammalian development. Emphasis is placed on the amino acid L-Proline, its metabolism, and the effect it has on the transcription of various genes and transcriptional circuits. While many studies have looked at neural-cell production following the formation of neurectoderm, few have focused, as we do here, on events prior to this: The formation of pluripotent primitive ectoderm, the multipotent germ layer definitive ectoderm and the first tangible evidence of the nervous system, neurectoderm.

In addition, we also look at the expression of genes associated with motor neurone disease (MND). According to the Developmental Origin of Health and Disease Hypothesis, compromised expression of such genes during early development *in utero* may combine with environmental factors which result in the onset of this neurodegenerative disease later in life.

We used bioinformatics to analyse previously collected RNA sequencing (RNAseq) data sets representing 4 increasingly ‘primed’ cell types across the pluripotency spectrum: Ground-state ESCs, naïve ESCs, early primitive ectoderm-like cells (EPLCs; produced from naïve ESCs

by the addition of L-Proline) and EpiSCs, which are fully primed and ready to lose pluripotency and form the three multipotent germ layers. Our approach consisted of examining changes in gene expression in the 4 pluripotent cell types and transforming the data sets using Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE), which provides candidate genes that are responsible for driving cell fate.

Our results show that many of these candidates are transcription factors, and that some of these are part of gene circuits associated with (i) pluripotency (e.g., *Tbx3* and *Esrrb*, which are part of the extended pluripotency network), (ii) lineage commitment (e.g., members of PcG complex of genes, which promote epigenetic changes), and (iii) L-Proline-mediated effects on transcription (e.g., *Dnmt3b* and *Fgf5*). In addition, some candidates are genes whose products control the cell cycle, are intermediates in cell signalling, or are enzymes in selected metabolic pathways. Changes in cell cycle, cell signalling, and metabolism are all important mechanistic contributors to cells moving across the pluripotency spectrum, including the L-Proline-mediated ES-to-EPLC transition. Further TRIAGE analysis using TRIAGE-Parser defines gene and protein-protein interactions, which help define the biological relevance of circuitry that drives progression along the pluripotency spectrum. These circuits indicate candidates of genes/proteins which act as master regulators of cell fate. They also indicate that EPLCs are likely more fated to become neural cells via the germ layer neuroectoderm, while EpiSCs are more like to lose pluripotency and become the other two germ layers: Mesoderm and endoderm.

In addition, results from RNAseq analysis are consistent with results from Seahorse extracellular flux assays, which show reduced reliance on aerobic metabolism as cells transition across the pluripotency spectrum. However, Seahorse assay results show that the addition of L-Proline naïve ESCs *per se* doesn't affect either aerobic or anaerobic metabolism, at least not over the 1-2 h time period of these experiments.

We also examined the differentiation of ESCs to neural cell types using a modified protocol that employs the conditioned medium, MEDII (which contains L-Proline as the bioactive component), and L-Proline in place of MEDII. Gene expression and TRIAGE analysis of previously collected microarray data of the MEDII-mediated differentiation to neural cell types show that some genes are key mediators of neural fate (e.g., *Neurogenin1*).

Finally, we show that many MND-associated genes and proteins are expressed very early in development, including prior to the first signs of the development of the nervous system. These results provide a baseline for further examination of the timing, location and role of these genes, including in mutated form, in contributing to the onset of MND in later life.

To my dear maternal grandmother, who unfortunately could not see this thesis finished.

I love you and still miss you heaps Ba Ngoai. Although you left sooner than we all had hoped, I pray that you are at peace with Buddha and had the chance to reunite with grandpa (Ong Ngoai).

Acknowledgements

What a roller coaster this journey has been! With disruption after disruption (and COVID-19) it's finally done! So many people I'd like to thank throughout this journey (probably could write another thesis for this!). First and foremost, to my primary supervisor Dr. Michael Morris and co-supervisor A/Prof. Margot Day. Thank you both very much for welcoming me back into your lab and for your guidance and support during these last few odd years. No amount of words is enough to express my gratitude for you both. Your patience, knowledge and kindness has made the Morris-Day Lab a wonderful environment to work in.

Huge thanks to Dr. Rachel Shparberg and Dr. Hannah Glover for your time and mentorship throughout the years. I'll never forget the guidance you both showed me during my baby Honours year that started my research journey. To you Rach, who is willing to lend some time, whether in the form of text rants, chats about random stuff or just looking over my thesis chapters. And Hannah, your valuable knowledge in the bioinformatics world helped me a lot during the beginning stages of the PhD when I didn't know what I was doing. And your confidence is something I look up to as a role model.

To all the wonderful members of the Morris-Day lab, both past and present. Especially to Maddy, Monique, Dani and Kate. Congrats to you Maddy for your completion of your PhD and for riding the journey with me (and back at you for our morning chats). Also, for you for selflessly allowing me to use your coffee machine to the max! To you Monique for always "believing in (me)". To you Lt. Dani. Your resilience, hard work and dedication is something to be praised about for years on end. To you Kate for coming into my "rock pool", your bubbly positivity, cheerleading, supply of dark chocolate and love of pink! Special mention to Chloe Purnell for your time and early discussions on making manuscripts.

Huge thanks as well to Dr. Woo Jun (Chris) Shim and A/Prof. Nathan Palpant from the University of Queensland. Our meetings have been invaluable to this project and your efforts deserve to be recognized. I also thank you for your time and patience for said meetings and for answering many of my questions.

I would also like to thank A/Prof. Anandwardhan Hardikar, Dr. Mugdha Joglekar, Dr. Wilson Wong, Dr. Luke Carroll and Dr. Andrzej Januszewski for your guidance in PCR and borrowing your equipment. I also would like to acknowledge Dr. Kristina Cook from the Centenary Institute and Charles Perkins Centre for allowing me to access and borrow the Seahorse XF96 analyser.

Of course, to all fellow members of the Medical Foundation Building community. Your chats and rants on experiments and writing really are something I take to heart. Despite how isolated it can be, we are all in this together. Big shout out to Dr. Laura Lindsay for your time in teaching me about BSA protein assays and your sass that never stops putting a smile on my face! And to Lara, for our occasional lunches together.

I would also like to acknowledge the Sydney Informatics Hub and the University of Sydney's high performance computing cluster, Artemis, for providing the computing resources that have contributed to the results reported herein. I'd like to especially thank Dr. Tracy Chew

and Dr. Georgina Samaha for your guidance to RNA sequencing. I would also like to thank Dr. Jianzhou Zhao and Khuong Tran from Intersect for your teaching and training in R, Unix and Artemis.

I would also like to acknowledge Dr. Sheng Hua, Dr. Donna Lai and Shirley Nakhla from Bosch, and Padmaja Dhanvate from Faculty of Medicine and Health for managing the equipment.

Big thanks to also my external mentor on the other side of the world, Dr. Shanell T. Smith. Your tough love and no-nonsense attitude are something I will cherish and walk forward with beyond this PhD. And for taking an interest in my work and reminding me of my “why”.

Huge thanks again and again to Australian Rotary Health for your support in not just the financials, but the emotional side of things too. Especially to Cheryl Deguara for your support throughout these years. And thank you again to my sponsors from Cronulla Rotary Club, especially to Kim Ruth, Maria Moran and Carol Denison. Thank you everyone for being so welcoming and for your time listening to my talks!

To Emily, I thank you for all the years you had to put up with me. I appreciated your help in my scholarship application, and I would not have been as resilient as I am today if not for you.

To members of Zany Dodgeball Club, for the reminders that there is life outside of the lab! Especially to you Rob and Lauren. Big shout out as well to Mandy and Morgan, your generosity and hilarious banter is like no other. Thank you for your sense of humour, whether it be sane or just outright dark. To Tajida and Daniel, for your kindness and keen interest in what I am doing. To my sister from another mister Clare, for your pillar of strength, no-nonsense attitude and for a shoulder to cry on when I was overwhelmed. I am and will be your (hard-boiled potato) rock when **** hits the fan.

To my partner Michael H. for your love, support, your boundless patience and honesty with me. Thank you for sticking by me when I seem somewhat aloof to others. Although we met during the later stages of my PhD journey, I do wonder what would have been like if we met at the beginning instead. Regardless, I am grateful that the algorithm caught us both to match (with your sister Stephanie catfishing in the first and only message LOL). I must also thank the rest of your family: Stephanie, Kristy, Olivia, Akbar, James, Liam, Maliwan, Mel, Tony, Lisa, Tomika, Penny and Tajeeb for welcoming me with open arms while I complete this PhD journey. Your family antics are wholesome and hilarious, and I am excited to create more memories with you all.

Of course, to my parents, especially Mum for making me delicious food for my overthinking and overworking brain, and Dad for your resilience and the necessary push and wakeup call (even if painful) on life’s ups and downs. Overbearing you both may be, thank you very much for making me the person I am today to finish this super massive project(s). No amount of words can describe the boundless love I have for the both of you.

Table of Contents

Title	Page
Chapter 1: Literature Review	1
1.1. Introduction	2
1.2. Signalling pathways influencing neural cell fate from pluripotent stem cells	5
1.2.1. LIF and BMP4 signalling are required for mouse ESC self-renewal and maintenance in the naïve pluripotency state	5
1.2.2. Ground-state mouse ESCs are more resistant to pro-differentiation signals than naïve state mouse ESCs	9
1.2.3. Mouse EpiSCs are more responsive to pro-differentiation signals than naïve mouse ESCs and preferentially favour mesendoderm lineage commitment	10
1.2.4. L-Proline-mediated activation of mammalian target of rapamycin is required but not sufficient for differentiation of mouse ESCs	14
1.2.5. L-Proline-mediated activation of mTOR and MAPK/ERK pathway activities are required for neural differentiation	15
1.3. Metabolism in stem-cell fate	16
1.3.1. Metabolism maintaining pluripotency and loss of pluripotency	16
1.3.2. Amino acids and lineage commitment	18
1.3.2.1. Amino acids strongly influence ESC function and differentiation	18
1.3.2.2. L-Proline metabolism mediates anaerobic and aerobic metabolism in mESCs	20
1.4. Transcriptional mechanisms in stem-cell fate	22
1.4.1. Core and extended pluripotency networks maintain ESC self-renewal and pluripotency	22
1.4.2. Multiple molecular networks work interdependently to drive ESC differentiation and lineage commitment	30
1.4.2.1. Transcriptional control works interdependently with epigenetics, translational control, and signalling to drive cellular activity during embryo development	30
1.4.2.2. A better understanding of the transcriptome is required to understand cell-fate specification.	33
1.4.3. Profiling the transcriptome can be used to predict mechanisms, including metabolic mechanisms, driving differentiation of mESCs to neural-cell fate	34
1.4.3.1. Single-cell RNAseq may reveal novel molecular mechanisms on differentiation while accounting for heterogeneity and revealing new cell types and states	34
1.4.3.2. Bioinformatics pipelining is required to process raw RNAseq data to generate hypotheses for further studies	37
1.5. Mechanisms driving mESCs to a neural fate during development may highlight mechanisms responsible for the onset of neurodegenerative diseases later in life	39
1.5.1. MND-associated genes and proteins and their contribution to the mechanism of neural degeneration	39
1.5.1.1. Fused in sarcoma (Fus) and Transactivating response DNA-binding protein of 43 kDa (Tdp-43) are MND-associated proteins involved in RNA metabolism	40

1.5.1.2. Optineurin (Otpn) and its association with trafficking and cell-cycle regulation	40
1.5.1.3. Survival motor neuron 1 (Smn1) and its coordinated role in RNA metabolism	41
1.5.1.4. Superoxide dismutase (Sod1) and its role as an antioxidant	42
1.5.1.5. Vesicle-associated membrane protein-associated protein B and C (Vapb) and its coordinated role in Sod1-linked MND	42
1.5.2. Developmental Origin of Health and Disease (DOHaD): A potential trigger of sporadic MND later in life	43
1.6. Concluding remarks	44
1.7. Overall aims and hypotheses	45
1.8. References	47
Chapter 2: Materials and Methods	87
2.1. Cell Culture	88
2.1.1. Materials	88
2.1.2. Cell culture media	90
2.1.3. Cell lines	92
2.1.4. Thawing mESCs and HepG2 cells	92
2.1.5. Mouse ESCs and HepG2 cell maintenance	93
2.1.6. Conditioned medium from HepG2 cells (MEDII)	93
2.1.7. EPLC formation and ground-state mouse ESC maintenance from mouse ESCs	94
2.1.8. Mouse embryonic fibroblast (MEF) derivation and culture and epiblast stem cell (EpiSC) culture	94
2.1.9. Neural cell formation from mESCs	95
2.2. Seahorse extracellular flux (XF) assay	98
2.2.1. Hydrating the sensor cartridge	98
2.2.2. Culturing and seeding cells	99
2.2.3. Preparation of XF assay medium and injections	99
2.2.4. Running the protocol and analysis	100
2.3. Gene Expression	102
2.3.1. Materials	102
2.3.2. RNA isolation and purification	104
2.3.3. RNA quantification and assessment	104
2.3.4. DNase treatment	105
2.3.5. cDNA production	105
2.3.6. Quantitative real-time polymerase chain reaction (qPCR) using Taqman™ gene expression assay	106
2.4. RNA sequencing (RNAseq) analysis	107
2.4.1. Sample preparation	107
2.4.2. Sequencing	107
2.4.3. RNAseq bioinformatics pipeline	107
2.4.4. Biological interpretation of TRIAGE-transformed genes	109
2.5. Microarray analysis	111
2.6. Western blot analysis	113

2.6.1. Materials	113
2.6.2. Gel preparation	114
2.6.3. Cell lysis for protein quantification and Western blot	115
2.6.4. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE)	116
2.6.5. Protein transfer and quality check	116
2.6.6. Membrane blocking and antibody staining	117
2.7. Microscopy	118
2.7.1. Phase contrast microscopy	118
2.7.2. Time-lapse microscopy with IncuCyte ZOOM	118
2.7.3. Colony morphology scoring	118
2.7.4. Doubling time calculations	121
2.8. Statistical analyses	121
2.9. References	122
Chapter 3: Investigating cell-fate drivers and L-Proline-mediated metabolic processes across the pluripotency spectrum	125
3.1. Introduction	126
3.2. Hypotheses and aims	132
3.3. Results	133
3.3.1. RNAseq analysis of cells at various stages along the pluripotency spectrum from ground state to primed state	133
3.3.1.1. Principle component analysis of RNAseq data	134
3.3.1.2. Differential gene expression analysis across the pluripotency spectrum	136
3.3.1.3. TRIAGE transformation of RNAseq data	139
3.3.1.4. Gene ontology and pathway analysis of top 100 DS genes	153
3.3.2. Relative expression and TRIAGE analysis of L-Proline metabolism-associated genes in cells across the pluripotency spectrum	166
3.3.3. Relative expression and TRIAGE analysis of genes associated with glycolysis in cells across the pluripotency spectrum	175
3.3.4. Relative expression and TRIAGE analysis of genes associated with the pentose phosphate pathway in cells across the pluripotency spectrum	180
3.3.5. Relative expression and TRIAGE analysis of genes associated with the TCA cycle and the electron transport chain in cells across the pluripotency spectrum	184
3.3.6. Metabolic profiles of ground- and naïve-state mESCs and EPLCs in response to L-Proline	192
3.3.7. The role of L-Proline metabolism in the mESC-to-EPLC transition	200
3.3.8. Effect of POX inhibitors on the L-Proline-mediated mESC-to-EPLC transition	205
3.3.9. L-Proline effects on colony morphology, doubling time and gene expression profiles in the presence of DCA	209
3.3.10. L-Proline effects on colony morphology, doubling time and gene expression profiles in the presence of 6AN	213
3.4. Discussion	219
3.4.1. Differential gene expression analysis and TRIAGE transformation provides candidate genes and predict protein-protein interactions driving cell fate across pluripotency	219

3.4.2. Selected genes encoding metabolic enzymes may drive cell fate and identity across the pluripotency spectrum	222
3.4.3. Ground and naïve states and EPLCs have different basal aerobic profiles as measured by the Seahorse extracellular flux assay but do not respond to acute exposure to L-Proline	226
3.4.4. Inhibition of POX-mediated conversion of L-Proline to P5C has different effects on parameters associated with the mESC-to-EPLC transition	227
3.4.5. Anaerobic metabolism and the PPP drive some aspects of the mESC-to-EPLC transition	228
3.5. References	230
Chapter 4: Transcription, epigenetics and L-Proline-mediated metabolism involved in neural lineage commitment from pluripotent mouse ESCs	246
4.1. Introduction	247
4.2. Aims and hypotheses	253
4.3. Results	254
4.3.1. Differential expression analysis of genes associated with pluripotency and lineage restriction in MEDII-treated embryoid bodies	254
4.3.2. Gene expression profile and phenotype changes across neural lineage commitment	259
4.3.3. TRIAGE transformation provides candidate genes controlling differentiation of mESC to neural fate	265
4.3.4. Relative expression and TRIAGE analysis of genes associated with PcG and TrxG proteins across pluripotency and neural lineage commitment	280
4.3.5. Relative expression and TRIAGE analysis of genes associated with L-Proline metabolism across neural lineage commitment	295
4.3.6. Relative expression and TRIAGE analysis of genes associated with glycolysis across neural lineage commitment	300
4.3.7. Relative expression and TRIAGE analysis of genes associated with pentose phosphate pathway across neural lineage commitment	303
4.3.8. Relative expression and TRIAGE analysis of genes associated with the TCA cycle across neural lineage commitment	306
4.3.9. Relative expression and TRIAGE analysis of genes associated with the electron transport chain across neural lineage commitment	309
4.4. Discussion	312
4.4.1. MEDII and L-Proline primes mESCs to commit to neural fate as soon as mESCs leaves the naïve state	312
4.4.2. TRIAGE transformation of genes shows candidate transcription factors directly driving neural fate commitment	317
4.4.3. PcG genes drive cell fate and identity across neural fate commitment, but TrxG genes do not	320
4.4.4. Metabolism mediates other mechanisms driving neural fate commitment	323
4.5. References	326
Chapter 5: Modelling early neural developmental processes to determine contributors to onset of motor neurone disease post-partum	346
5.1. Introduction	347
5.2. Hypotheses and aims	353

5.3. Results	354
5.3.1. Relative expression and TRIAGE analysis of MND-associated genes across the pluripotency spectrum	354
5.3.2. Interrogation of MND-associated genes in relevant cell populations across early gastrulation	360
5.3.3. Relative expression and TRIAGE analysis of MND-associated genes across neural lineage commitment	364
5.3.4. qPCR analysis of the expression of MND-associated genes across the pluripotency spectrum and neural lineage commitment	367
5.3.5. MND-associated protein expression analysis across the pluripotency spectrum	370
5.3.6. MND-associated protein expression analysis across neural lineage commitment	372
5.4. Discussion	374
5.4.1. MND-associated genes are expressed across the pluripotency spectrum and during neural fate commitment but do not appear to contribute directly to cell-fate decisions	375
5.4.2. Expression of MND-associated genes and proteins do not always follow the same trends as bioinformatics interrogation	376
5.4.3. Expression of MND-associated proteins play their roles in housekeeping and cell structure across pluripotency and neural lineage commitment	376
5.4.4. Limitations and future directions	377
5.5. References	381
Chapter 6: Final Discussion	400
6.1. References	407

Presentations and Published Abstracts

***Denotes prizes awarded*

Scholarship presentations at the Rotary Club of Cronulla Meeting, Australian Rotary Health

“Understanding developmental processes that can lead to motor neurone disease”

Oral presentation, Rotary Club of Cronulla, Cronulla, Australia

Nov 2022

“Understanding developmental processes that can lead to motor neurone disease”

Zoom presentation

Nov 2021

“Understanding neurone development for motor neurone disease”

Zoom presentation

Sept 2020

Academic presentations

“Triggers of Motor Neuron Disease during development? Using embryonic stem cells as a model”

Three-minute thesis, Molecular Biomedicine Theme Meeting, University of Sydney

Oct 2022

“Modelling neural development using L-Proline-mediated differentiation of mouse embryonic stem cells”

Zoom presentation, University of Sydney

Jun 2020

“Changes in metabolism may drive differentiation of mouse embryonic stem cells to primitive ectoderm-like cells”

Three-minute thesis, Bosch Annual Scientific Meeting, University of Sydney

Aug 2019

Poster presentations

“Delayed onset of sporadic motor neuron disease may be set up during pluripotent cell commitment to the neural lineage”

Australasian Society for Stem Cell Research Annual Scientific Meeting, Melbourne, Australia

Nov 2022

“Pentose phosphate pathway is not involved in the L-Proline-concentration-dependent mouse embryonic stem cell to early primitive ectoderm-like cell transition”.

Twitter presentation, Australian Society for Medical Research 28th Annual Scientific Meeting**

Nov 2020

“Pentose phosphate pathway is required for L-proline-mediated transition of mouse embryonic stem cells to early primitive ectoderm-like cells”.

Australasian Society for Stem Cell Research Annual Scientific Meeting, Brisbane, Australia
Nov 2019

“Pentose phosphate pathway is required for L-proline-mediated transition of mouse embryonic stem cells to early primitive ectoderm-like cells”.

Australian and New Zealand Stem Cell and Developmental Biology Cell and Developmental Biology Meeting, University of Sydney AND Australian Society for Medical Research 27th Annual Scientific Meeting, Sydney, Australia
Jun 2019

List of Figures

Figure number	Title	Page
1.1	The dynamic pluripotency spectrum recapitulates early embryo development	7
1.2	Cell signalling pathways contributing to the switch between self-renewal and differentiation in mouse embryonic stem cells (ESCs).	8
1.3	Cell signalling pathways contributing to the switch between self-renewal and differentiation in mouse epiblast stem cells and human embryonic stem cells.	12
1.4	L-Proline mediates anaerobic and aerobic metabolism in mouse embryonic stem cells	17
1.5	Recruitment of polycomb and trithorax group proteins to repress and express genes during cell-fate commitment.	24
1.6	Core units of polycomb-repressive complexes 1 (PRC1) and PRC2 and their mammalian homologs.	26
1.7	COMPASS complexes are composed of trithorax proteins.	28
2.1	L-Proline-based protocol to model neural differentiation from mouse ESCs.	97
2.2	RNAseq bioinformatics pipeline.	110
2.3	Hybridisation strategy used for microarray datasets.	112
2.4	Representative reference images for scoring colony morphology.	119
2.5	Representative images of neural features used for phenotypic scoring of EBs in adherent culture at Day 15.	120
3.1	Metabolism in ESCs, including proposed emergent properties from pathways active during ES-to-EPLC transition, and the role of metabolic inhibitors dichloroacetate (DCA), L-tetrahydro-2-furoic acid (THFA), 3,4 dehydro-L-Proline (DHP) and 6-aminonicotinamide (6AN).	128
3.2	Bioinformatics and experimental approaches used to determine L-Proline-mediated metabolic changes across the pluripotency spectrum.	131
3.3	Principle components analysis (PCA) of cells from the D3 and A2 cell lines at various states along the pluripotency spectrum.	135
3.4	Expression of genes associated with pluripotency and lineage restriction across the ESC pluripotency spectrum.	138
3.5	D3 ground-state mESC RNAseq data transformed using Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE).	141
3.6	D3 and A2 ground-state mESC RNAseq data transformed using TRIAGE.	143-144
3.7	D3 and A2 naïve-state mESC RNAseq data transformed using TRIAGE.	146-147
3.8	D3 and A2 EPLC RNAseq data transformed using TRIAGE	149-150

3.9	EpiSC RNAseq data transformed using TRIAGE.	152
3.10	Gene ontology (GO) analysis of the top 100 ranked DS genes using TRIAGE-Parser for the D3 cell line to identify biological processes associated with the A. ground state, B. naïve state and C. EPLCs.	154-158
3.11	Gene ontology (GO) analysis of the top 100 ranked DS genes using TRIAGE-Parser for the A2 cell line to identify biological processes associated with the A. ground state, B. naïve state, C. EPLCs and D. EpiSCs.	160-165
3.12	Relative expression of genes encoding enzymes associated with proline catabolism across the pluripotency spectrum.	168
3.13	TRIAGE transformation of genes encoding enzymes associated with proline catabolism across the pluripotency spectrum.	169-170
3.14	Relative expression of genes encoding enzymes associated with proline biosynthesis across the pluripotency spectrum.	172
3.15	TRIAGE transformation of genes encoding enzymes associated with proline biosynthesis across the pluripotency spectrum.	173-174
3.16	Expression of genes encoding enzymes associated with glycolysis across the pluripotency spectrum.	176-177
3.17	TRIAGE transformation of genes encoding enzymes associated with glycolysis across the pluripotency spectrum.	178-179
3.18	Expression of genes encoding enzymes associated with oxidative and non-oxidative phases of the PPP across the pluripotency spectrum.	181
3.19	TRIAGE transformation of genes encoding enzymes associated with oxidative and non-oxidative phases of the PPP.	182-183
3.20	Expression of genes encoding enzymes associated with TCA cycle across the pluripotency spectrum.	185
3.21	TRIAGE transformation of genes encoding enzymes associated with the TCA cycle across the pluripotency spectrum.	186-187
3.22	Expression of genes encoding complexes and enzymes involved in the ETC across the pluripotency spectrum.	189
3.23	TRIAGE transformation of genes encoding complexes and enzymes associated with the ETC across the pluripotency spectrum.	190-191
3.24	Measurement of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) in ground- and naïve-state mESCs and EPLCs using the Seahorse extracellular flux assay.	194-195
3.25	Basal metabolic rates of ground and naïve state ESCs and EPLCs.	197
3.26	Corrected OCR and ECAR values of cells across the pluripotency spectrum in response to acute injections of L-Proline and mitochondrial uncoupler FCCP.	199
3.27	Colony morphology and doubling time profiles of naïve mESCs cultured in LIF and 0.1% DMSO vehicle control.	201-202
3.28	Gene expression profiles of naïve mESCs cultured in LIF and 0.1% DMSO vehicle control.	204
3.29	Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline with or without POX inhibitors.	206

3.30	Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± DHP ± THFA.	208
3.31	Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline with or without suppression of anaerobic metabolism by DCA.	210
3.32	Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± DCA.	212
3.33	Dose culture of 6AN using doubling time to determine the optimal concentration.	214
3.34	Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline ± 6AN.	216
3.35	Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± 6AN.	218
4.1	MEDII-based (A, B) and L-Proline-based (C) protocols used to generate cell types as mESCs differentiate towards neural fate via embryologically relevant cell types.	249-250
4.2	Pluripotency network and lineage restriction genes in a MEDII-based protocol driving mESCs to neural fate.	255
4.3	Differential expression analysis of genes associated with pluripotency and lineage restriction in MEDII-treated and spontaneously differentiated naïve mESCs.	257-258
4.4	Gene expression profiles of EBs cultured by the hanging drop method in L-Proline for neural lineage commitment or in Basal medium to allow spontaneous differentiation.	260-262
4.5	EBs scored for neural features following seeding down at Day 9 after suspension culture.	264
4.6	TRIAGE transformation of microarray data for mESCs differentiated in MEDII to neural-cell fate (A) or allowed to spontaneously differentiate (B).	267-269
4.7	TRIAGE-Parser analysis of genes with the top 100 DS rank for cells at Day 6 and 9 differentiated in MEDII to neural fate.	271-274
4.8	TRIAGE-Parser analysis of genes with the top 100 DS rank for EPLCs left to spontaneously differentiate to Day 5 and Day 7.	276-279
4.9	Changes in PcG core component gene expression (A) across the pluripotency spectrum and (B) in MEDII-directed neural differentiation and spontaneously differentiated cells.	282-283
4.10	Changes in TrxG gene expression (A) across the pluripotency spectrum and (B) in MEDII-directed neural differentiation and spontaneously differentiated cells.	285-286
4.11	TRIAGE-transformed PcG genes across the pluripotency spectrum.	288-289
4.12	TRIAGE-transformed TrxG genes across the pluripotency spectrum	290-291
4.13	TRIAGE-transformed PcG genes across intermediate stages of neural fate commitment.	293
4.14	TRIAGE-transformed TrxG genes across intermediate stages of neural fate commitment.	294

4.15	Relative expression and TRIAGE transformation of proline catabolism genes across neural lineage commitment.	296-297
4.16	Relative expression and TRIAGE transformation of proline biosynthesis genes across neural lineage commitment.	299
4.17	Relative expression and TRIAGE transformation of glycolysis genes across neural lineage commitment.	301-302
4.18	Relative expression and TRIAGE transformation of PPP genes across neural lineage commitment	304-305
4.19	Relative expression and TRIAGE transformation of TCA cycle genes across neural lineage commitment	307-308
4.20	Relative expression and TRIAGE transformation of ETC genes across neural lineage commitment.	310-311
5.1	Relative expression of MND-associated genes across the pluripotency spectrum.	356
5.2	TRIAGE transformation of MND-associated genes across the pluripotency spectrum.	358-359
5.3	Expression of MND-associated genes in different cell types of the mouse embryo during early gastrulation.	361-363
5.4	Relative expression of MND-associated genes across neural lineage commitment.	365
5.5	TRIAGE transformation of MND-associated genes across neural lineage commitment.	366
5.6	Expression of 5 MND-associated genes (A-E) in i) cells across the pluripotency spectrum and ii) EBs cultured by hanging drop method in L-Proline for neural lineage commitment.	368-369
5.7	Western blot analysis of the expression of 3 MND-associated proteins across the pluripotency spectrum.	371
5.8	Western blot analysis of the expression of 3 MND-associated proteins across neural lineage commitment.	373
6.1	Summary of the proposed L-Proline-mediated mechanisms that drive neural fate commitment from mESCs.	403-404

List of Tables

Table number	Title	Page
2.1	Cell culture media for mESC, HepG2, MEF and EpiSC maintenance, as well as adherent culture and embryoid body culture experiments.	90
2.2	Plate sizes, seeding densities, and media volumes for maintenance of adherent culture.	91
2.3	Taqman™ gene expression assays for qPCR.	103
2.4	Primary antibodies used for Western blot analysis.	114
2.5	Secondary antibodies used for Western Blot analysis.	114
2.6	Separating gel components for different acrylamide/bis-acrylamide % (v/v).	114
2.7	Western blot buffers and solutions and their components.	115

List of Abbreviations

Literature

Abbreviation	Full description
4EBP1	Eukaryotic initiation factor 4E-binding protein 1
6-PG	6-phosphogluconate
AA	Amino acid
AAR	Amino-acid starvation response
ALK	Activin receptor-like kinase
ALS	Amyotrophic lateral sclerosis
AP	Alkaline phosphatase
Atf4	Activating Transcription Factor 4
ATP	Adenosine triphosphate
bHLH	Basic helix–loop–helix
BMP4	Bone morphogenetic protein 4
BMPR	Bone morphogenetic protein receptor
CDK	Cyclin-dependent kinases
CNS	Central nervous system
COMPASS	Complex associated with trithorax-related SET protein
DMSO	Dimethyl sulfoxide
DOHaD	Developmental Origins of Health and Disease
dpc	Days post coitum
DUSP9	Dual-specificity phosphatase 9
E	Embryonic day
ECAR	Extracellular acidification rate
ECM	Extracellular matrix
Eif2e	Eukaryotic initiation factor 4E
Eif2α	Eukaryotic initiation factor 4 α
EMT	Epithelial-to-mesenchymal transition
EPL	Primitive ectoderm-like
EPLCs	Primitive ectoderm-like cells
ERK	Extracellular signal-regulated kinase
ES	Embryonic stem
ESCs	Embryonic stem cells
ETS	Electron transport system
FBS	Fetal bovine serum
Fgf	Fibroblast growth factor
Gcn2	General control non-repressed 2
GF	Growth factor
Gln	L-Glutamine
Glu	L-Glutamate
Glucose-6P	Glucose-6-phosphate
GLUT	Glucose transporter

Gly3P	Glyceraldehyde 3-phosphate
gp130	Glycoprotein 130
Grb2	Growth factor receptor bound protein 2
GSA	Glutamate-5-semialdehyde
H3K36me3	Histone H3 lysine 36 trimethylation
H3K4me3	Histone H3 lysine 4 trimethylation
H3K9me3	Histone H3 lysine 9 trimethylation
HIF1a	hypoxia inducible factor-1a
ICM	Inner cell mass
Ids	Inhibitor of differentiation
IRS1	Insulin receptor substrate 1
JAK	Janus Kinase
lcnRNA	Long non-coding RNAs
LDH	Lactate dehydrogenase
LIF	Leukemia inhibitory factor
LIFRβ	Leukemia inhibitory factor receptor beta
L-Pro	L-Proline
MAPK	Mitogen-Activated Protein Kinase
MCT	Monocarboxylate transporter
MEDII	Conditioned medium derived from the hepatocarcinoma cell line HepG2
mESCs	Mouse embryonic stem cells
MND	Motor neurone disease
mTOR	Mammalian target of rapamycin
mTORC1	Mammalian target of rapamycin complex 1
mTORC2	Mammalian target of rapamycin complex 2
NADPH	Nicotinamide adenine dinucleotide phosphate
OAA	Oxaloacetate
OAT	Ornithine-aminotransferase
OCR	Oxygen consumption rate
Orn	L-Ornithine
OxPhos	Oxidative phosphorylation
P5C	Pyrroline-5-carboxylate
PcG	Polycomb group
PDKs	Pyruvate dehydrogenase kinases
PE	Primitive ectoderm
PEn	Primitive endoderm
PI3K	Phosphoinositide 3-kinase
POX	Proline oxidase
PPAR	Peroxisome proliferator-activated receptor
PPARα	Peroxisome proliferator-activated receptor α
PPARγ	Peroxisome proliferator-activated receptor γ
PPIs	Protein-protein interactions
PPP	Pentose phosphate pathway
PRC1	Polycomb-repressive complex 1
PRC2	Polycomb-repressive complex 2

ROS	Reactive oxygen species
Rosi	Rosiglitazone
Ru5P	Ribulose-5-phosphate
S6K	S6 kinase
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
Shh	Sonic hedgehog
SNAT2	Sodium-coupled neutral amino acid transporter 2
SOS	Son of sevenless guanine nucleotide exchange factor
STAT3	Signal Transducer and Activator of Transcription 3
SucCoA	Succinyl coenzyme A
TCA	Tricarboxylic acid
TGF-β	Transforming growth factor beta
TGF-βR	Transforming growth factor beta receptor
TMRM	Tetramethylrhodamine methyl ester perchlorate
TrxG	Trithorax group
XF	Extracellular flux
α-KG	Alpha-ketoglutarate

Methodology

Abbreviation	Description
6AN	6-aminonicotinamide
APS	Ammonium persulphate
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
DCA	Sodium dichloroacetate
DHP	3,4-dehydro-L-proline
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DS	Discordance score
DTT	Dithiothreitol
EBs	Embryoid bodies
FAM	Fluorescein amidite
FBS	Fetal bovine serum
FCCP	Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FDR	False discovery rate
FPKM	Fragments per kilobase of transcript per million mapped reads
HISAT2	Hierarchical indexing for spliced alignment of transcripts 2
HPC	High Performance Computing
ITSS	Insulin-transferrin-sodium selenite
KO-DMEM	Knockout-Dulbecco's Modified Eagle Medium
KOSR	Knockout serum replacement

MEFs	Mouse embryonic fibroblasts
NaCl	Sodium chloride
NEAA	Non-essential amino acids
PBS	Phosphate buffered saline
PBST	0.1% (v/v) Triton-X in PBS
Pen-strep	Penicillin/streptomycin
PFA	Paraformaldehyde
PMSF	Phenylmethanesulfonyl fluoride
qPCR	Quantitative real-time polymerase chain reaction
RNAseq	RNA sequencing
RT	Reverse transcriptase
SDS	Sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
TBS	Tris buffered saline
TBST	0.1% (v/v) of 10% (v/v) Tween in Tris buffered saline
TEMED	Tetramethylethylenediamine
THFA	L-tetrahydro-2-furoic acid
TRIAGE	Transcriptional Regulatory Inference Analysis from Gene Expression
Tris	Trisaminomethane
VIC	2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein
β-mec	Beta-mercaptoethanol

Gene and Protein Abbreviations

Gene (Protein)	Full description	Category
<i>Aco1</i>	Aconitase 1	TCA
<i>Aco2</i>	Aconitase 2, mitochondrial	TCA
<i>Aldh18a1</i> (P5CS)	Aldehyde dehydrogenase 18 family, member A1 (pyrroline-5-carboxylate synthase)	Pro metabolism
<i>Aldh4a1</i> (P5CDH)	Aldehyde dehydrogenase 4 Family, member A1 (Pyrroline-5-carboxylate dehydrogenase)	Pro metabolism
<i>Aldoa</i>	Fructose-bisphosphate aldolase	Glycolysis
BMP4	Bone morphogenic protein 4	
<i>Brp44</i> (MPC2)	Mitochondrial pyruvate carrier 2	Glycolysis
<i>Brp44l</i> (MPC1)	Mitochondrial pyruvate carrier 1	Glycolysis
<i>C9orf72</i> (C9orf72)	Chromosome 9 open reading frame 72	MND
<i>Cbx2</i>	Chromobox protein homolog 2	Polycomb-repressive complex 1
<i>Cbx4</i>	Chromobox protein homolog 4	Polycomb-repressive complex 1
<i>Cbx6</i>	Chromobox protein homolog 6	Polycomb-repressive complex 1
<i>Cbx7</i>	Chromobox protein homolog 7	Polycomb-repressive complex 1
<i>Cbx8</i>	Chromobox protein homolog 8	Polycomb-repressive complex 1
<i>Cdx2</i>	Caudal type homeobox 2	
<i>Cox10</i>	Heme A:farnesyltransferase cytochrome c oxidase assembly factor 10	
<i>CoxIV/Cox4i1</i> (COX IV)	Cytochrome C oxidase subunit IV	ETC
<i>Cs</i>	Citrate synthase	TCA
<i>Csl</i>	Citrate synthase-like	TCA
<i>Dnmt3a</i> (Dnmt3a)	DNA methyltransferase 3-alpha	
<i>Dnmt3b</i> (Dnmt3b)	DNA methyltransferase 3-beta	
<i>Eed</i>	Embryonic ectoderm development protein	Polycomb-repressive complex 2
<i>Eno1</i>	Enolase 1	Glycolysis
<i>Esrrb</i>	Estrogen related receptor, beta	

Gene (Protein)	Full description	Category
<i>Ezh1</i>	enhancer of zeste 1 polycomb repressive complex 2 subunit	Polycomb-repressive complex 2
<i>Ezh2</i>	enhancer of zeste 2 polycomb repressive complex 2 subunit	Polycomb-repressive complex 2
<i>Fgf4</i>	Fibroblast growth factor 4	
<i>Fgf5</i>	Fibroblast growth factor 5	
<i>Fh1</i>	Fumarate hydratase 1	TCA
<i>Fus (Fus)</i>	Fused in sarcoma	MND
<i>G6pdx (G6PD)</i>	Glucose-6-phosphate dehydrogenase	PPP
<i>Gapdh</i>	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis
<i>Gata3</i>	GATA binding protein 3	
<i>Gata6</i>	GATA binding protein 6	
<i>Gbx2</i>	Gastrulation brain homeobox 2	
<i>Gclc</i>	Glutamate-cysteine ligase catalytic subunit	Pro metabolism
<i>Gls</i>	Glutaminase (mitochondrial)	Pro metabolism
<i>Gls2</i>	Glutaminase 2	Pro metabolism
<i>Glud1 (GLUDH)</i>	Glutamate dehydrogenase 1	Pro metabolism
<i>Glul (GS)</i>	Glutamate-ammonia ligase/Glutamine synthetase	Pro metabolism
<i>Glut1/Slc2a1</i>	Glucose transporter 1	Metabolism
<i>Glut2/Slc1a2</i>	Glucose transporter 2	Metabolism
<i>Gpi1</i>	Glucose-6-phosphate isomerase	Glycolysis
<i>Gsr</i>	Glutathione reductase	Pro metabolism
<i>Gss</i>	Glutathione synthetase	Pro metabolism
<i>Hk</i>	Hexokinase	Metabolism
<i>Hnf-3β</i>	Hepatocyte nuclear factor-3 beta	
HNRNPK	Heterogeneous nuclear ribonucleoprotein K	
<i>Idh1</i>	Isocitrate dehydrogenase 1 (NADP+), soluble	TCA
<i>Idh2</i>	Isocitrate dehydrogenase 2 (NADP+), mitochondrial	TCA
<i>Idh3a</i>	Isocitrate dehydrogenase 3 (NAD+) alpha	TCA
<i>Idh3b</i>	Isocitrate dehydrogenase 3 (NAD+) beta	TCA
<i>Idh3g</i>	Isocitrate dehydrogenase 3 (NAD+), gamma	TCA
<i>Kdm3a</i>	H3K9 demethylase	Pro metabolism
<i>Kdr/Flk1</i>	Kinase insert domain protein receptor	
<i>Ldha</i>	Pyruvate dehydrogenase E1 Subunit Beta	Glycolysis
<i>Let-7</i>	Lethal-7 microRNA precursor	
<i>lin28</i>	Lin-28 homolog	
<i>Lncenc1</i>	Long non-coding RNA, embryonic stem cells expressed 1	
<i>Mdh2</i>	Malate dehydrogenase 2, NAD (mitochondrial)	TCA
<i>Mixl1</i>	Mixl homeobox-like 1	

Gene (Protein)	Full description	Category
<i>Nanog</i>	Nanog homeobox	
<i>Ndufa10</i> (NADH)	NADH dehydrogenase [ubiquinone] 1	ETC
<i>Nek1</i> (Nek1)	NIMA (never in mitosis gene a)-related expressed kinase 1	MND
<i>NeuroG1</i>	Neurogenin 1	
<i>Oat</i> (Oat)	Ornithine aminotransferase	Pro metabolism
<i>Ogdh</i>	Oxoglutarate (alpha-ketoglutarate) dehydrogenase (lipoamide)	TCA
<i>Optn</i> (Optn)	Optineurin	MND
<i>Pard6b</i>	Par-6 family cell polarity regulator beta	
<i>Pcgf1</i>	Polycomb group ring finger 1	Polycomb-repressive complex 1
<i>Pcgf2</i>	Polycomb group ring finger 2	Polycomb-repressive complex 1
<i>Pcgf3</i>	Polycomb group ring finger 3	Polycomb-repressive complex 1
<i>Pcgf4/Bmi1</i>	Polycomb group ring finger 4	Polycomb-repressive complex 1
<i>Pcgf5</i>	Polycomb group ring finger 5	Polycomb-repressive complex 1
<i>Pcgf6</i>	Polycomb group ring finger 6	Polycomb-repressive complex 1
<i>Pcx</i>	Pyruvate carboxylase	TCA
<i>Pdha1</i>	Lactate dehydrogenase A	Glycolysis
<i>Pdhb</i>	Pyruvate dehydrogenase E1 component subunit alpha	Glycolysis
<i>Pdk1</i>	Pyruvate dehydrogenase kinase 1	Glycolysis
<i>Penk1</i>	Preproenkephalin 1	
<i>Pfkm</i>	Phosphofructokinase	Glycolysis
<i>Pgam1</i>	Phosphoglycerate mutase 1	Glycolysis
<i>Pgd</i> (6PGD)	6-phosphogluconate dehydrogenase	PPP
<i>Pgk1</i>	Phosphoglycerate kinase 1	Glycolysis
<i>Pgl3</i> (6PGL)	6-phosphogluconolactonase	PPP
<i>Pgrn</i> (Grn)	Progranulin/granulin	MND
<i>Phc1</i>	Polyhomeotic 1	Polycomb-repressive complex 1
<i>Phc2</i>	Polyhomeotic 2	Polycomb-repressive complex 1

Gene (Protein)	Full description	Category
<i>Phc3</i>	Polyhomeotic 3	Polycomb-repressive complex 1
<i>Pkm</i>	Pyruvate kinase M1/2	Glycolysis
<i>Pkm2</i>	Pyruvate kinase M2	Metabolism
<i>Pou5f1/Oct4</i>	POU domain, class 5, transcription factor 1	
<i>Prodh (Pox)</i>	Proline dehydrogenase 1	
PTBP1	Polypyrimidine tract-binding protein 1	
<i>Pycr1</i>	Pyrroline-5-carboxylate reductase 1	Pro metabolism
<i>Pycr2</i>	Pyrroline-5-carboxylate reductase 2	Pro metabolism
<i>Pycrl</i>	Pyrroline-5-carboxylate reductase-like	Pro metabolism
<i>Rbbp4</i>	Retinoblastoma binding protein 4	Polycomb-repressive complex 2
<i>Rbbp7</i>	Retinoblastoma binding protein 7	Polycomb-repressive complex 2
<i>Ring1</i>	Ring finger protein 1	Polycomb-repressive complex 2
<i>Rnf2</i>	Ring finger protein 2	Polycomb-repressive complex 2
<i>Rpe</i>	Ribulose phosphate 3-epimerase/ phosphopentose epimerase	PPP
<i>Rpia</i>	Ribose-5-phosphate isomerase/ phosphopentose isomerase	PPP
<i>Sall4</i>	Spalt-like transcription factor 4	
<i>SdhA (SDHA)</i>	Succinate dehydrogenase subunit A, flavoprotein (Fp)	TCA, ETC
<i>SdhB (SDHB)</i>	Succinate dehydrogenase subunit B, iron sulfur (Ip)	TCA, ETC
<i>Slc16a1 (MCT1)</i>	Solute Carrier Family 16 member 1/Monocarboxylate transporter 1	Glycolysis
<i>Slc16a3 (MCT4)</i>	Solute Carrier Family 16 member 3/Monocarboxylate transporter 4	Glycolysis
<i>Slc16a7 (MCT2)</i>	Solute Carrier Family 16 member 7/Monocarboxylate transporter 2	Glycolysis
<i>Slc16a8 (MCT3)</i>	Solute Carrier Family 16 member 8/Monocarboxylate transporter 3	Glycolysis
<i>Slc38a2 (SNAT2)</i>	Solute Carrier Family 38 Member 2/Sodium-coupled neutral amino acid transporter 2	Pro metabolism
SMAD1/5/8	SMAD family member 1/5/8	
<i>Smn1 (Smn1)</i>	Survival of motor neuron 1, telomeric	MND
<i>Sod1 (Sod1)</i>	Superoxide desmutase enzyme	MND
<i>Sox17</i>	SRY (sex determining region Y)-box 17	
<i>Sucla2</i>	Succinate-Coenzyme A ligase, ADP-forming, beta subunit	TCA

Gene (Protein)	Full description	Category
<i>Suclg1</i>	Succinate-CoA ligase, GDP-forming, alpha subunit	TCA
<i>Suclg2</i>	Succinyl-CoA synthetase/ligase beta subunit, GDP-forming	TCA
<i>Suz12</i>	Suppressor of zeste 12	Polycomb-repressive complex 2
<i>T</i>	Brachyury	
<i>Taldo1</i>	Transaldolase	PPP
<i>Tardbp</i> (Tdp-43)	Transactive response DNA binding protein	MND
<i>Tbk1</i> (Tbk1)	TANK-binding kinase 1	MND
<i>Tet1</i>	Ten–Eleven-Translocation DNA dioxygenase 1	
<i>Tkt</i>	Transketolase	PPP
<i>Tktl1</i>	Transketolase-like 1	PPP
<i>Tktl2</i>	Transketolase-like 2	PPP
<i>Tpi1</i>	Triosephosphate Isomerase 1	Glycolysis
<i>Trp53</i>	p53 tumour suppressor	ETC
<i>Upcl</i> (UCP1)	Uncoupling protein 1	ETC
<i>Upcl</i> (UCP2)	Uncoupling protein 2	ETC
<i>Upcl</i> (UCP3)	Uncoupling protein 3	ETC
<i>Uqcrrf1</i> (CIII-R)	Complex III Rieske subunit	ETC
<i>Vapb</i> (Vapb)	Vesicle-associated membrane protein-associated protein B and C	MND
<i>Vcp</i> (Vcp)	Valosin-containing protein	MND
<i>Wnt3</i>	Wingless-type MMTV integration site family, member 3	
<i>Zfp42/Rex1</i>	Zinc finger protein 42	
<i>α-tubb</i> (α-tub)	Alpha tubulin	
<i>β-actin</i> (β-actin)	Beta actin	

Chapter 1: Literature Review

1.1. Introduction:

Development of the early nervous system from inner cell mass (ICM) cells requires their lineage-directed differentiation to multiple cell types (Glover, 2018; Shparberg, 2018). In the mouse, cell migration occurs in the ICM after blastocyst implantation into the uterine wall 4.5 days post coitus (dpc) to produce 2 distinct cell lineages: *Gata6*⁺/*Sox17*⁺ cells and *Nanog*⁺/*Rex1*⁺ cells. *Gata6*⁺/*Sox17*⁺ cells move to the blastocoel surface to become the hypoblast (or primitive endoderm) (Artus *et al.*, 2011). These cells lay down a collagen-rich extracellular matrix (ECM) that separates the ICM from the hypoblast (Chazaud *et al.*, 2006; Arnold and Robertson, 2009). *Nanog*⁺/*Rex1*⁺ cells remain as pluripotent epiblast cells, contributing to the development of all cells of the embryo proper and extraembryonic tissue (excluding the placenta) (Nichols and Smith, 2009; Krupinski *et al.*, 2011). In keeping with their pluripotency, epiblast cells maintain expression of *Oct4* and alkaline phosphatase activity (Chazaud *et al.*, 2006; Nichols and Smith, 2009, 2012). By 5.5-6.5 dpc, the remaining epiblast cells reduce doubling time from ~12 h to ~4 h (Snow, 1977) as they form the second pluripotent cell population called primitive ectoderm which prepares the embryo for gastrulation (Tam and Behringer, 1997; Pelton *et al.*, 2002). Pluripotency markers *Oct4* expression and alkaline phosphatase activity are maintained in primitive ectoderm while the expression of ICM markers such as *Nanog* and *Rex1* are downregulated and expression of primitive ectoderm markers *Fgf5* and *Dnmt3b* are upregulated (Pelton *et al.*, 2002; Nichols and Smith, 2009).

Primitive ectoderm undergoes gastrulation to form the 3 multipotent germ layers of the embryo proper: Definitive ectoderm, definitive endoderm and definitive mesoderm by 7.5 dpc. Gastrulation begins at ~6.5 dpc with the formation of the primitive streak which reaches its final length by 7.5 dpc (Tam and Behringer, 1997; Loebel *et al.*, 2003; Arnold and Robertson, 2009). Location-specific and highly regulated secretion of Nodal, Wnt3 and bone morphogenetic protein (BMP)4 is responsible for primitive streak elongation (Tam *et al.*, 2006;

Arnold and Robertson, 2009; Deepak and Sudipto, 2013; Shparberg *et al.*, 2019b). Wnt3 signalling and downregulation of E-cadherin expression results in some primitive ectoderm cells undergoing an epithelial-to-mesenchymal transition (EMT), resulting in their migration through the primitive streak to form mesendoderm, and eventually the definitive mesoderm and definitive endoderm (Tam and Behringer, 1997; Loebel *et al.*, 2003; Arnold and Robertson, 2009). Primitive ectoderm cells that do not pass through primitive streak develop into definitive ectoderm (Tam and Behringer, 1997; Lu *et al.*, 2001; Tam *et al.*, 2006; Tam and Loebel, 2007; Li *et al.*, 2013). Inhibition of Nodal, Wnt3 or BMP4 signalling (for instance, by culturing *ex vivo* 6.5 dpc embryos in the presence of Nodal inhibitor, SB431542) results in a failure to produce mesendoderm. Instead, the resulting primitive ectoderm differentiates into the definitive ectoderm derivatives: surface ectoderm and neurectoderm (Inman *et al.*, 2002; Li *et al.*, 2013).

Commitment to surface ectoderm or neurectoderm from definitive ectoderm depends, in part, on the presence or absence of BMP4-mediated SMAD1/5/8 signalling, respectively (Di-Gregorio *et al.*, 2007). *Penk1* and *Pard6b* are two potential definitive ectoderm markers based on an *in vitro* neural differentiation protocol derived from mouse embryonic stem cells (ESCs) (Harvey *et al.*, 2010). These markers have yet to be confirmed in the 6.5-7.0 dpc embryo. Besides *Penk1* and *Pard6b*, there is a lack of definitive ectoderm markers, unlike mesendoderm lineages for which several markers have been confirmed *in vivo* and *in vitro*; e.g. *Mixl1*, *T*, *Flk1* and *Sox17* (Arnold and Robertson, 2009; Ishitobi *et al.*, 2011). The lack of markers makes it more difficult to derive the molecular mechanisms driving definitive ectoderm formation (Loebel *et al.*, 2003; Tam *et al.*, 2006; Arnold and Robertson, 2009; Harvey *et al.*, 2010; Li *et al.*, 2013).

At 7.0-7.5 dpc, a transient population of ciliated *Hnf-3 β* ⁺ cells is produced at the distal tip of the embryo, called the node. Primitive ectoderm cells that migrate through the node towards

the anterior side of the embryo produce the prechordal plate and notochord, which are required for generating the neural plate and patterning of the neural tube (Ang and Rossant, 1994; Chiang *et al.*, 1996; Ding *et al.*, 1998). The node secretes BMP4 antagonists including chordin, noggin and follistatin at 7.0 dpc (Bachiller *et al.*, 2000; Brazil *et al.*, 2015). The portion of the definitive ectoderm closest to the node (where BMP signalling is low) differentiates into neurectoderm, which forms the neural plate and subsequent central nervous system (CNS). Neurectoderm (neural plate) transiently expresses *Sox1* followed closely by *Pax6* in mice (Rathjen *et al.*, 2002; Suter *et al.*, 2009). The neural plate thickens via proliferation, then begins to elevate and fold, converging to the midline at 8.5 dpc (Ybot-Gonzalez *et al.*, 2002). Lateral neural plate cells come into contact with overlying surface ectoderm. The resulting neural tube is now exposed to a dorsal-ventral gradient of morphogens from the prechordal plate and notochord (McMahon *et al.*, 1998; Placzek and Briscoe, 2005). Sonic hedgehog (Shh) is secreted from the floor plate (ventral side) of the neural tube (Chiang *et al.*, 1996; Ding *et al.*, 1998; Briscoe and Ericson, 1999; Jeong and Epstein, 2003) and a range of BMPs, including BMP4/7, are secreted from the roof plate (dorsal side) of the neural tube and overlying surface ectoderm (Liem *et al.*, 1997; Lee *et al.*, 2000; Wine-Lee *et al.*, 2004). These morphogen gradients help establish the dorsal-ventral patterning of the CNS. Cells closest to Shh activity (ventral side) develop into V3 neurons. Cells with 'moderate' Shh activity develop into motor neurons. As Shh activity decreases, BMP activity from the roof plate increases, where cells develop into V2, V1 then V0 interneurons, and D2 then D1 interneurons (Shparberg *et al.*, 2019b).

In vitro models of early development (including ESCs) and neurogenesis (neural stem cells) have been used to focus on cell signalling mechanisms that are coordinated to regulate the transition through multiple cell types (Bratt-Leal *et al.*, 2009; Martynoga *et al.*, 2012; Shin *et al.*, 2015; Telley *et al.*, 2016). Metabolism, including those of amino acids, are now known to play their part in driving development and cell fate, with L-Proline acting as a growth factor driving neural lineage commitment from pluripotent stem cells (Rathjen *et al.*, 2002;

Washington *et al.*, 2010; D’Aniello *et al.*, 2015; Shparberg, 2018). The first section of the literature review will focus on signalling pathways contributing to neural differentiation from mouse ESCs. The second section will focus on emerging evidence of amino acids and metabolism driving cell fate, and metabolism involvement in neural differentiation. The third section will focus on transcriptional mechanisms contributing their part to drive cell fate and how we can discover novel mechanisms that include, but are not limited to, metabolic pathways. The fourth section will apply the *in vitro* model of early establishment of neural lineage commitment to a potential mechanism of the onset of motor neurone disease (MND), a family of related neurodegenerative diseases, later in life.

1.2. Signalling pathways influencing neural cell fate from pluripotent stem cells

1.2.1. *LIF and BMP4 signalling are required for mouse ESC self-renewal and maintenance in the naïve pluripotency state*

Studying mechanisms driving ICM-to-neural development is difficult *in vivo* because the embryo develops very quickly, involves a small number of cells, and is hidden away in the uterus (Jones *et al.*, 2002; Anderson and Stern, 2016). ICM cells in pre-implanted mouse blastocysts can be extracted and grown in tissue culture to produce mouse ESC lines (Figure 1.1) (Evans and Kaufman, 1981; Martin, 1981; Doetschman *et al.*, 1985). Mouse ESCs express the same pluripotency markers (e.g., *Oct4* and alkaline phosphatase activity) and ICM markers (e.g., *Rex1*, *Nanog* and *Gbx2*) as the cells from which they were derived (Figure 1.1) (Rathjen *et al.*, 1999; Casalino *et al.*, 2011). Mouse ESCs are defined by their ability to (i) self-renew, (ii) maintain a normal karyotype, and (iii) display pluripotency as shown, for example, by their ability to produce germ-line chimeras (Evans and Kaufman, 1981; Bryja *et al.*, 2006). Leukemia inhibitory factor (LIF) is one factor responsible for maintaining pluripotency and self-renewal in mouse ESCs (Figure 1.1). LIF binds to the LIF receptor-glycoprotein 130 (LIFR-gp130)

complex to then activate 3 main intracellular pathways: (i) Janus kinase (JAK)/signal transducer and activator of transcription 3 (STAT3), (ii) phosphoinositide 3-kinase (PI3K)/Akt and (iii) mitogen-activated protein kinase (MAPK)/ERK (Figure 1.2) (Hirai *et al.*, 2011). JAK/STAT3 and PI3K/Akt pathways promote self-renewal and proliferation (Burdon *et al.*, 2002; Ying *et al.*, 2003; Cartwright *et al.*, 2005; Graf *et al.*, 2011), while the MAPK/ERK pathway promotes differentiation (Figure 1.2) (Burdon *et al.*, 2002; Nichols and Smith, 2009; Casanova *et al.*, 2011; Graf *et al.*, 2011). In self-renewing mouse ESCs, the balance of LIF-mediated signalling pathway activities lie in favour of self-renewal but the cells are primed for differentiation. This state is considered to be the naïve state of pluripotency (Nichols and Smith, 2009).

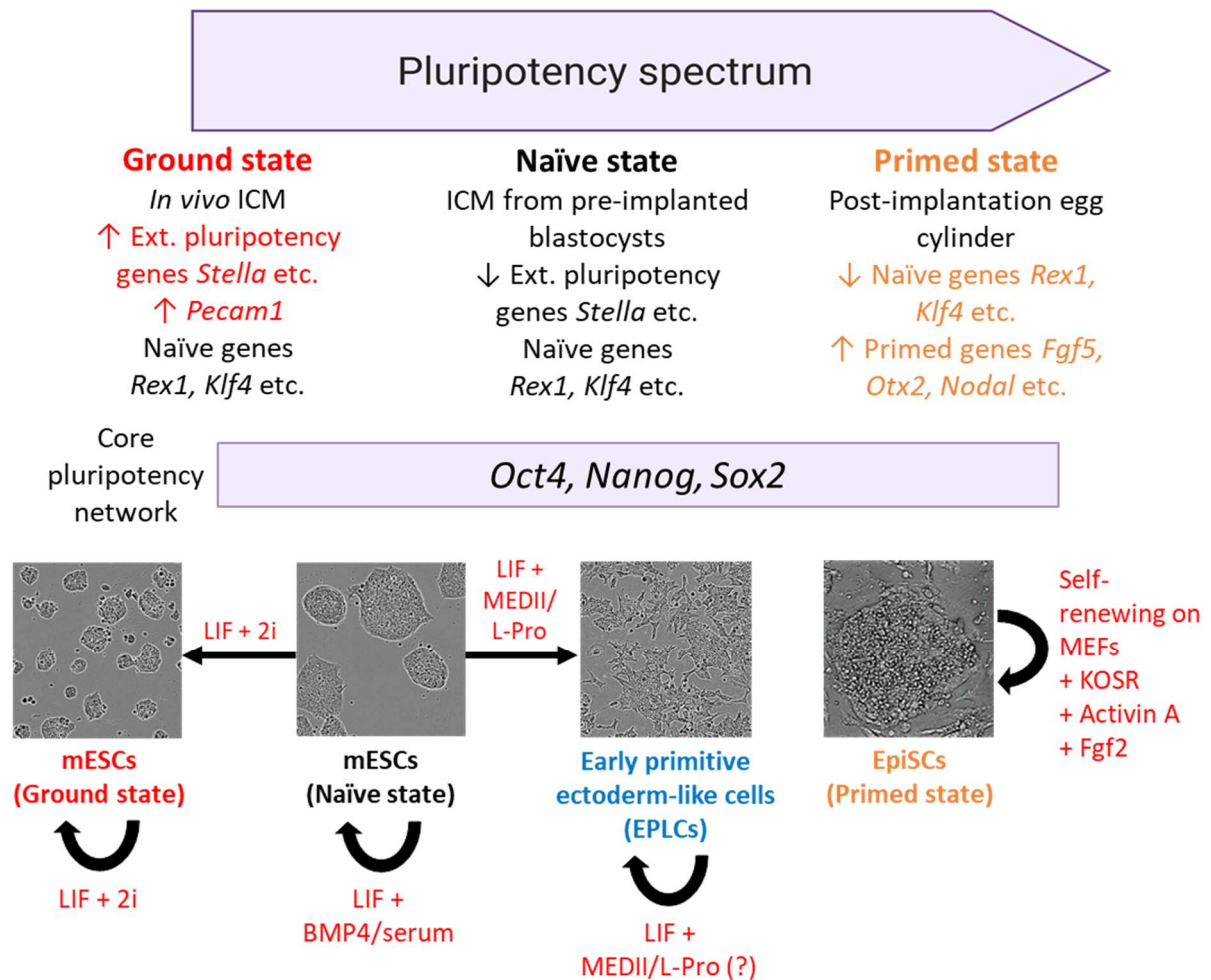


Figure 1.1 The dynamic pluripotency spectrum recapitulates early embryo development. Naïve-state mouse embryonic stem cells (mESCs) can be derived from pre-implanted blastocysts at 3.5-4.5 dpc. They express naïve genes, such as *Rex1* and *Klf4* while maintaining expression of core pluripotency genes *Oct4*, *Nanog* and *Sox2*. Naïve mESCs can self-renew and be maintained in LIF + serum or serum-derived BMP4 *in vitro*. Naïve-state mESCs grow as a mix of round, domed shaped colonies and colonies with irregular borders. The naïve state can be driven back to the ground state, which is derived from naïve mESCs cultured in LIF and the presence of MEK inhibitor PD0325901 and glycogen synthase kinase 3 inhibitor CHIR99021 (2i). Ground-state mESCs represent *in vivo* ICM in the 4.0 dpc mouse embryo, can self-renew and be maintained *in vitro* using LIF + 2i. The 2i system allows these ground-state mESCs to resist pro-differentiation signals like *Fgf4* and remain in a transcriptionally repressed state. Like naïve mESCs, ground-state mESCs demonstrate naïve genes while maintaining core pluripotency gene. However, ground-state mESCs have high expression of extended pluripotency network genes such as *Stella* and ICM markers like *Pecam1*. Ground-state mESCs grow uniformly as round, domed shaped colonies. Addition of MEDII or L-Proline to the LIF + BMP4/serum system (LIF + MEDII/L-Pro) produces early primitive ectoderm-like cells (EPLCs). EPLCs represent an intermediate/formative stage between naïve and primed states. EPLCs grow as flat colonies with irregular borders and can be maintained *in vitro* in LIF + MEDII/L-Pro. However, it is unknown if they can self-renew. The primed state is represented by mouse epiblast stem cells (EpiSCs). They are derived from the egg cylinder in the 5.5-7.5 dpc post-implantation mouse embryo. EpiSCs have high expression of primed genes while maintaining core pluripotency genes. EpiSCs grow as flat, compact colonies, self-renew and be maintained *in vitro* on feeder layers such as mouse embryonic fibroblasts (MEFs) in the presence of Knockout Serum Replacement (KOSR), activin A and *Fgf2*. Figure derived from Glover 2019.

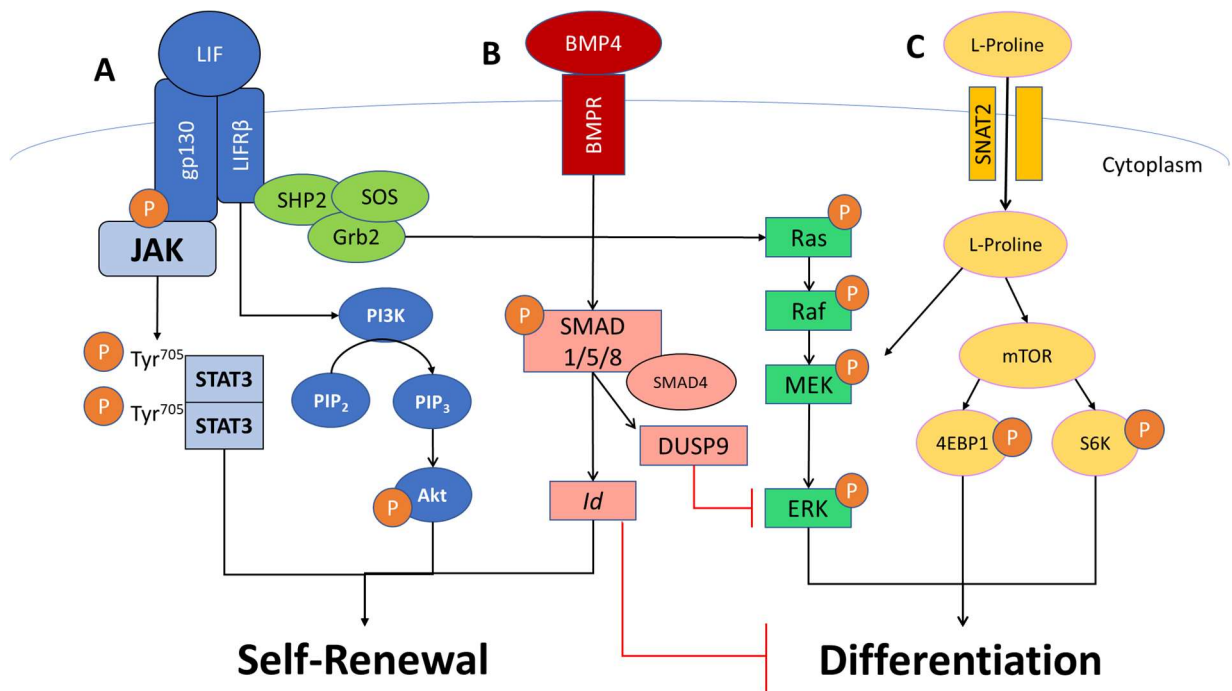


Figure 1.2 Cell signalling pathways contributing to the switch between self-renewal and differentiation in mouse embryonic stem cells (ESCs). **A.** Leukemia inhibitory factor (LIF) binds to LIF beta receptor (LIFRβ) and glycoprotein-130 (gp130), forming a heterodimeric complex. Janus kinase (JAK) gets recruited to the heterodimer and becomes phosphorylated. This results in the heterodimer becoming a docking site to recruit signal transducer and activator of transcription 3 (STAT3). Recruited STAT3 is phosphorylated at Tyr⁷⁰⁵, homodimerizes, then translocates to the nucleus to induce transcription of self-renewal genes. The LIF-gp130-LIFRβ complex also activates the phosphoinositide 3-kinase (PI3K) pathway, resulting in conversion of PIP₂ to PIP₃ and downstream phosphorylation of AKT. This results in cell-cycle progression and proliferation, contributing to self-renewal. The LIF-gp130-LIFRβ complex also contributes to differentiation by recruiting the tyrosine phosphatase SHP2, growth factor receptor bound protein 2 (Grb2) and son of sevenless guanine nucleotide exchange factor (SOS), which activates the mitogen-activated protein kinase (MAPK)/ERK pathway. **B.** LIF-induced self-renewal is complemented by bone morphogenetic protein 4 (BMP4) binding to BMP receptor (BMPR), resulting in phosphorylation of SMAD1/5/8. Phosphorylated SMAD1/5/8 heterodimerises with SMAD4, which translocates into the nucleus to upregulate expression of inhibitor of differentiation (*Id*) genes. *Id* gene products prevent differentiation and induce self-renewal. SMAD signalling also results in upregulation of expression of the phosphatase, DUSP9. DUSP9 inhibits phosphorylation of ERK, thereby preventing differentiation. **C.** LIF-induced activation of the MAPK/ERK differentiation pathway is complemented by the actions of L-Proline. L-Proline enters mESCs via the sodium-coupled neutral amino acid transporter 2 (SNAT2) to activate MAPK/ERK and mammalian target of rapamycin (mTOR) signalling, indicated by phosphorylated downstream proteins, eukaryotic initiation factor 4E-binding protein 1 (4EBP1) and ribosomal protein S6 kinase (S6K). MAPK/ERK and mTOR signalling induces differentiation of mouse ESCs to early primitive ectoderm-like cells. Figure adapted from: (Burdon *et al.* 2002, Ying *et al.* 2003, Cartwright *et al.* 2005, Graf *et al.* 2011, Casanova *et al.* 2011, Hirai *et al.* 2011, Washington *et al.* 2010, Shparberg *et al.* 2019b).

BMP4, either added exogenously or present in serum, is also required to maintain self-renewal in naïve-state mouse ESCs (Figure 1.1). BMP4 stimulates SMAD1/5/8 signalling, promoting expression of inhibitor of differentiation genes to suppress lineage commitment (Figure 1.2) (Ying *et al.*, 2003; Qi *et al.*, 2004). BMP4 signalling also attenuates MAPK/ERK activity by upregulating expression of dual-specificity phosphatase 9 (DUSP9) (Figure 1.2) (Qi *et al.*, 2004; Hirai *et al.*, 2011). DUSP9 dephosphorylates ERK in mouse ESCs, suppressing differentiation-mediated signalling via this pathway (Figure 1.2) (Qi *et al.*, 2004; Li *et al.*, 2012).

1.2.2. Ground-state mouse ESCs are more resistant to pro-differentiation signals than naïve state mouse ESCs

The pluripotency continuum is a dynamic, reversible set of states ranging from the ground state to naïve state to a primed pluripotency state seen in primitive ectoderm and cell lines of epiblast stem cells (EpiSCs) (Figure 1.1) (Hackett *et al.*, 2013). Ground-state mouse ESCs represent cells of the ICM at an earlier stage of blastocyst development (4.0 dpc) and are considered the “blank state” or “zero state” of pluripotency (Figure 1.1) (Wray *et al.*, 2010; Hackett and Surani, 2014; Glover *et al.*, 2022). Ground-state mouse ESCs can be maintained in LIF + 2i medium, where 2i represents the MAPK inhibitor PD0325091, and the glycogen synthase kinase (GSK)-3 β inhibitor CHIR99021 (Figure 1.1) (Ying *et al.*, 2008; Wray *et al.*, 2011; Marks *et al.*, 2012; Tosolini and Jouneau, 2015). GSK-3 β inhibition makes the cells resistant to pro-differentiation signals like Fgf4 (Kunath *et al.*, 2007; Ying *et al.*, 2008; Nichols and Smith, 2009; Chen, C. W. *et al.*, 2010). The colonies are uniformly round and dome shaped (Figure 1.1) (Ying *et al.*, 2008; Hackett *et al.*, 2013; Glover, 2018; Glover *et al.*, 2022). This contrasts with mouse ESCs in LIF + BMP4/serum medium, which are further along the pluripotency system and where colony morphology is a mix of round and dome shaped together with colonies with somewhat irregular borders (Figure 1.1) (Wray *et al.*, 2010; Tosolini and Jouneau, 2015). Ground-state mouse ESCs demonstrate high expression of genes in the extended pluripotency network such as *Stella* and

ICM markers like *Pecam1* in combination with expression of the 3 core pluripotency genes, *Oct4*, *Nanog* and *Sox2* (Figure 1.1) (Hayashi *et al.*, 2008; Zhou *et al.*, 2010). Ground-state mouse ESCs express *Stella* homogeneously, while naïve-state mouse ESCs also express *Stella* but is distributed heterogeneously (Galonska *et al.*, 2015).

1.2.3. Mouse EpiSCs are more responsive to pro-differentiation signals than naïve mouse ESCs and preferentially favour mesendoderm lineage commitment

Mouse EpiSC lines are derived from mouse post-implantation embryos at 5.5–7.5 dpc and can generate chimeras when grafted into post-implantation embryos but not pre-implantation embryos (Figure 1.1) (Brons *et al.*, 2007; Bernemann *et al.*, 2011). They also generally don't integrate when injected into embryos after gastrulation when pluripotency is lost, although this depends on the cell line (Bernemann *et al.*, 2011; Huang *et al.*, 2012; Kojima *et al.*, 2014). They show strong resemblance to cells of the anterior primitive streak of the embryo, indicated by their transcriptional signatures derived from transcriptome analysis such as RNA sequencing (RNAseq) and microarray (Section 1.4.3) (Kojima *et al.*, 2014; Peng *et al.*, 2016). Mouse EpiSCs are also similar to human ESCs: They do not rely on LIF but rely on activin A-mediated Smad transcriptional activity and Fgf2-mediated MAPK/ERK signalling for self-renewal (Figure 1.1 and 1.3) (Brons *et al.*, 2007; Tesar *et al.*, 2007; Vallier, Touboul, *et al.*, 2009; Kim *et al.*, 2013; Kojima *et al.*, 2014; Osteil *et al.*, 2016). EpiSCs maintain expression of the core pluripotency gene network consisting of *Oct4*, *Nanog*, and *Sox2* but also begin to express higher levels of genes associated with lineage specification and differentiation (Figure 1.1) (Brons *et al.*, 2007; Tesar *et al.*, 2007; Bernemann *et al.*, 2011; Kojima *et al.*, 2014). For this reason, these cells are said to be 'primed' and are ready to rapidly exit pluripotency if exposed to the right exogenous signals which initiate differentiation (Vallier, Touboul, *et al.*, 2009; Osteil *et al.*, 2016). EpiSCs can be maintained in culture with Fgf2, activin A and Knockout Serum Replacement (KOSR) and generally require mouse embryonic fibroblasts (MEFs) as a feeder

layer to maintain their primed pluripotent state (Figure 1.1) (Greber *et al.*, 2007; Tesar *et al.*, 2007; Kim *et al.*, 2013; Liu *et al.*, 2019).

In contrast with MAPK/ERK signalling in mouse ESCs (Figure 1.2), MAPK/ERK signalling in mouse EpiSCs and human ESCs, mediated by Fgf2, maintains self-renewal by suppressing differentiation (Figure 1.3A): Fgf receptor or MAPK/ERK inhibitors in self-renewal medium causes mouse EpiSCs and human ESCs to differentiate to the neural lineage (Greber *et al.*, 2010). Fgf signalling also indirectly promotes self-renewal using feeder layers: Fgf2 stimulates MEFs to produce TGF- β ligands that bind to TGF- β receptors, which activates Smad signalling to facilitate self-renewal (Figure 1.3A and 1.3B) (Greber *et al.*, 2007). In mouse EpiSCs, Fgf2-mediated MAPK/ERK signalling also maintains self-renewal by suppressing *Klf2* expression, a marker of naïve pluripotency, to prevent conversion back to the naïve state, and prevents neurectoderm differentiation by repressing *Pax6* expression (Greber *et al.*, 2007, 2010).

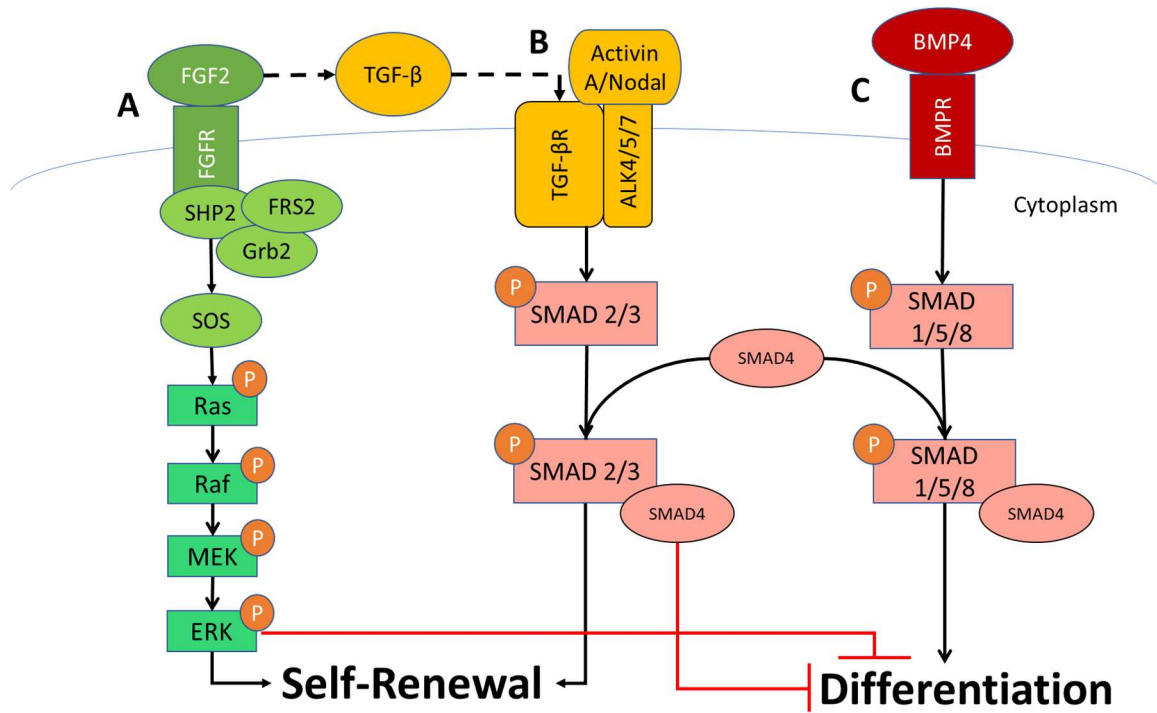


Figure 1.3. Cell signalling pathways contributing to the switch between self-renewal and differentiation in mouse epiblast stem cells and human embryonic stem cells. **A)** Fibroblast growth factors, such as fibroblast growth factor 2 (FGF2) binds to FGF receptors (FGFR), and this recruits scaffolding proteins SHP2, growth factor receptor bound protein 2 (Grb2) and FGF receptor substrate 2 (FRS2), which activates son of sevenless guanine nucleotide exchange factor (SOS). As a result, SOS activates the mitogen-activated protein kinase (MAPK)/ERK pathway, promoting self-renewal and suppressing differentiation. FGF2 also stimulates feeder layers like mouse embryonic fibroblasts to produce TGF-β ligands to bind to TGF-β receptors (TGF-βR). **B)** Activin A or Nodal also binds to TGF-βR and activin receptor-like kinase (ALK)4/5/7 receptors. This forms a heterodimeric complex which results in transphosphorylation of SMAD2/3. SMAD4 dimerises with SMAD2/3, which translocates into the nucleus and binds to Smad-binding elements on DNA to express pluripotency genes such as *Nanog* to promote self-renewal and repress differentiation. **C)** Bone morphogenetic protein 4 (BMP4) binds to BMP receptor (BMPR), resulting in phosphorylation of SMAD1/5/8. Phosphorylated SMAD1/5/8 heterodimerises with SMAD4, which translocates into the nucleus to promote transcription of differentiation genes. Adapted from Glover, 2019.

Activin A contributes to self-renewal and pluripotency in mouse EpiSCs and human ESCs via Nodal/TGF- β /Smad signalling and also prevents neural differentiation in EpiSCs (Figure 1.3B) (Camus *et al.*, 2006; Brons *et al.*, 2007; Tesar *et al.*, 2007; Smith *et al.*, 2008; Vallier, Mendjan, *et al.*, 2009; Greber *et al.*, 2010). Activin A binds to TGF- β and activin receptor-like kinase (ALK)4/5/7 receptors, which results in transphosphorylation of Smad2/3 (Figure 1.3B). Smad2/3 then dimerises with Smad4 (Figure 1.3B) and the resulting complex translocates into the nucleus to bind Smad-binding elements on DNA, which regulate gene expression (Vallier, Mendjan, *et al.*, 2009; Pauklin and Vallier, 2015). In EpiSCs, this includes induction of *Nanog* expression (Greber *et al.*, 2008; Vallier, Mendjan, *et al.*, 2009; Pauklin and Vallier, 2015). In turn, *Nanog* inhibits Fgf-mediated signalling, limiting neurectoderm differentiation from EpiSCs (Vallier, Mendjan, *et al.*, 2009). *Nanog* also interacts directly with Smad2/3 proteins that usually promote mesendoderm differentiation (Pauklin and Vallier, 2013; Dalton, 2015). The end result is a negative feedback system that cancels out both neurectoderm and mesendoderm differentiation, and promotes maintenance of pluripotency.

In contrast to naïve state mouse ESCs, where BMP4 activation promotes self-renewal and pluripotency (Figure 1.1 and 1.2), in mouse EpiSCs and human ESCs, BMP4 promotes differentiation (Figure 1.3C). Thus, KOSR is used in mouse EpiSC and human ESC culture since it lacks the BMP4 found in serum (Figure 1.1) (Xu *et al.*, 2005; Greber *et al.*, 2007; Osteil *et al.*, 2016; Liu *et al.*, 2019). EpiSCs can also be derived from naïve mouse ESCs by producing chimeras. Naïve mouse ESCs are incorporated into the blastocyst, then implanted into pseudo-pregnant mice. After blastocysts are implanted, they develop further, then isolated and maintained as an EpiSC line on an MEF feeder layer with the conventional Fgf2, activin A and KOSR culture medium. This allows production of EpiSCs which have the same genetic background, including any genetic modification, as the naïve mouse ESC line from which they were derived (Osteil *et al.*, 2016).

Wnt/ β -catenin signalling facilitates both self-renewal and differentiation in mouse EpiSCs and human ESCs (Kim *et al.*, 2013). Its dual role in self-renewal and differentiation relies on the location of stabilised β -catenin: Retention of stabilised β -catenin in the cytoplasm maintains self-renewal, while translocation of β -catenin into the nucleus and subsequent interaction with T-cell factors regulate gene expression for differentiation (Kim *et al.*, 2013). In the presence of Fgf2, activin A and KOSR, Wnt/ β -catenin signalling lies in favour towards self-renewal.

EpiSCs have the potential to commit to any of the three germ layers (Brons *et al.*, 2007). However, this is cell line-dependent and most of the available cell lines appear primed for mesendoderm commitment (Bernemann *et al.*, 2011). This is also seen in their gene expression profiles, where they express definitive mesoderm markers such as *Brachyury/T* and *Mixl1* and definitive endoderm markers such as *Gata6* and *Sox17* (Brons *et al.*, 2007; Tesar *et al.*, 2007; Bernemann *et al.*, 2011; Kojima *et al.*, 2014).

1.2.4. L-Proline-mediated activation of mammalian target of rapamycin is required but not sufficient for differentiation of mouse ESCs

An intermediate/formative stage of pluripotency between naïve and primed can be produced from naïve mouse ESCs with MEDII (conditioned medium derived from the human hepatocarcinoma cell line, HepG2) to homogeneously differentiate them to early primitive ectoderm-like cells (EPLCs) (Figure 1.1) (Rathjen *et al.*, 1999; Rodda *et al.*, 2002). EPLCs are analogous to primitive ectoderm, in that they have upregulated expression of primitive ectoderm cell markers *Fgf5* and *Dnmt3b* and downregulate ICM cell markers *Rex1*, *Gbx2* and *Nanog* while maintaining pluripotency markers such as *Oct4* and alkaline phosphatase activity (Figure 1.1) (Rathjen *et al.*, 1999; Pelton *et al.*, 2002). Therefore, culturing mouse ESCs with MEDII to produce EPLCs recapitulates *in vivo* primitive ectoderm formation from ICM.

L-Proline is one of the main active components in MEDII driving mouse ESC-to-EPLC transition (Washington *et al.*, 2010). Culturing mouse ESCs with >0.1 mM L-Proline produces EPLCs in the same manner as MEDII, as indicated by faster proliferation rates and gene expression changes associated with the transition (Rathjen *et al.*, 1999; Washington *et al.*, 2010; Casalino *et al.*, 2011). Upon addition of MEDII or L-Proline, signalling activity occurs that promotes differentiation over self-renewal. In part, this is due to a combination of L-Proline-induced activation of MAPK/ERK and mammalian target of rapamycin (mTOR) pathways (Figure 1.2). Activation of mTOR is necessary for L-Proline-mediated mouse ESC differentiation. However, mTOR signalling is not sufficient as other amino acids that activate mTOR, such as Glycine and L-Leucine, fail to induce gene expression changes associated with EPLC formation (Washington *et al.*, 2010; Tan *et al.*, 2011). Inhibiting MAPK/ERK using U0126 prevents gene expression changes associated with L-Proline-mediated EPLC formation (Lonic, 2007; Washington *et al.*, 2010; Tan *et al.*, 2011). Therefore, mTOR must work together with MAPK/ERK signalling to produce EPLCs from mouse ESCs (Burdon *et al.*, 2002; Gangloff *et al.*, 2004; Lonic, 2007; Niwa *et al.*, 2009; Washington *et al.*, 2010).

1.2.5. L-Proline-mediated activation of mTOR and MAPK/ERK pathway activities are required for neural differentiation

Culturing mouse ESCs with MEDII or L-Proline also promotes neural differentiation from mouse ESCs via embryologically relevant intermediate stages including *Fgf5*⁺ primitive ectoderm, *Penk1*⁺/*Pard6*⁺ definitive ectoderm and *Sox1*⁺ neuroectoderm (Rathjen *et al.*, 2002; Rodda *et al.*, 2002; Harvey *et al.*, 2010; Shparberg *et al.*, 2019a). In fact, EPLCs appear to be primed for neural differentiation, as indicated by upregulation of genes associated with neurogenesis (Glover, 2018), as opposed to EpiSCs which appear to be more primed for mesendoderm lineage commitment (Brons *et al.*, 2007; Tesar *et al.*, 2007; Bernemann *et al.*, 2011; Kojima *et al.*, 2014; Glover, 2018). Furthermore, L-Proline-mediated mTOR and

MAPK/ERK pathway activities not only are required for mouse ESC differentiation to EPLCs, but also contribute to neural differentiation (Shparberg *et al.*, 2019b; Glover *et al.*, 2022).

1.3. Metabolism in stem-cell fate

1.3.1. *Metabolism maintaining pluripotency and loss of pluripotency*

ICM cells in the blastocyst proliferate and differentiate under hypoxic conditions of the fallopian tubes and uterus (~5% O₂) and embryos rely more heavily on anaerobic metabolism (the conversion of glucose to lactate), to produce ATP than on oxidative phosphorylation (OxPhos) of glucose via the electron transport system (Figure 1.4) (Oh *et al.*, 2005; Cho *et al.*, 2006; Rehman, 2010; Ochocki and Simon, 2013). This metabolic preference is also reflected *in vitro* in highly proliferative ESCs, which rely more on aerobic glycolysis (i.e., the conversion of glucose to lactate in the presence of 20% O₂) to produce ATP – a phenomenon known as the Warburg effect (Figure 1.4) (Liu *et al.*, 2015; Sperber *et al.*, 2015; Harvey *et al.*, 2019). Consistent with this, ESCs have immature mitochondria, as assessed by their morphology (Oh *et al.*, 2005; Cho *et al.*, 2006; Mandal *et al.*, 2010). The transition of ESCs-to-EpiSCs is marked by changes in metabolic profile to a more glycolytic one (Zhou *et al.*, 2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022).

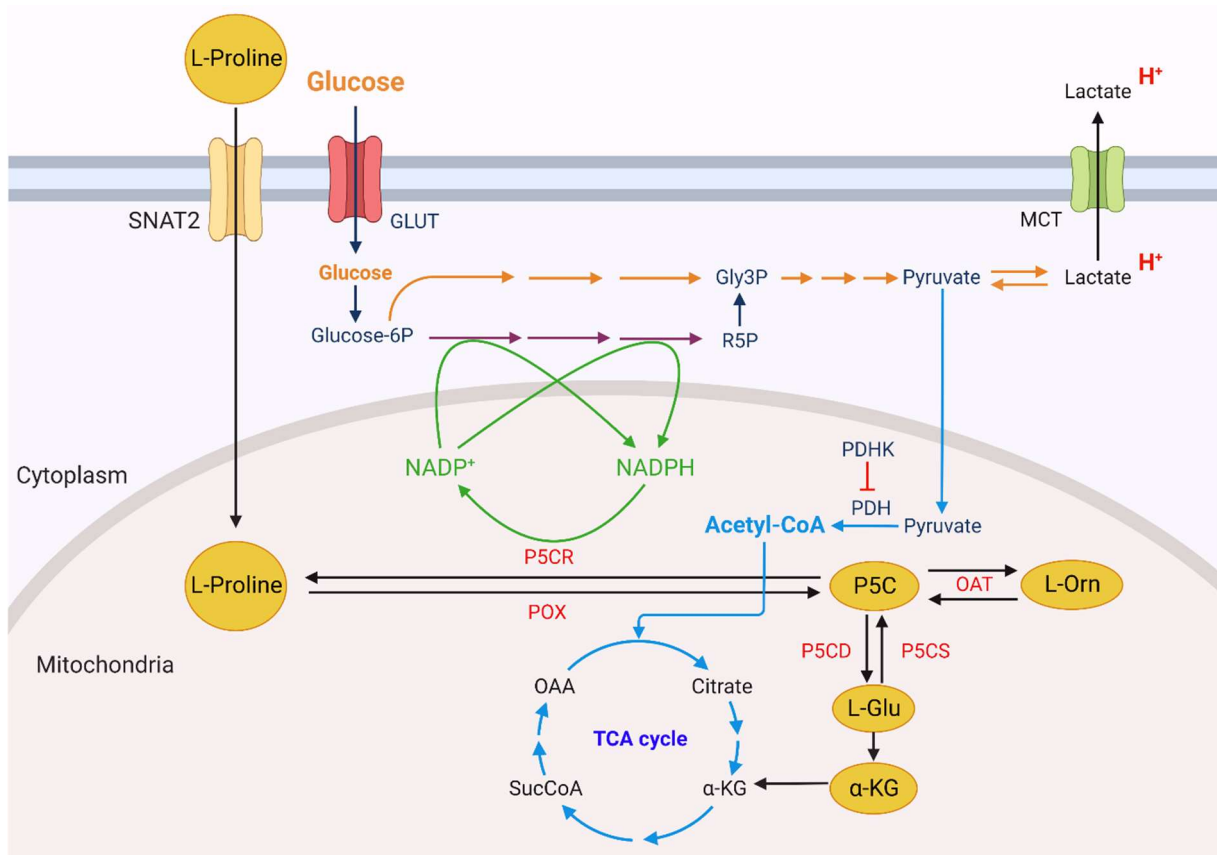


Figure 1.4. L-Proline mediates anaerobic and aerobic metabolism in mouse ESCs. L-Proline enters mESCs via the SNAT2 transporter and moves into the mitochondria. L-proline gets metabolised through the proline cycle: L-Proline produces pyrroline-5-carboxylate (P5C) via proline oxidase (POX) and converts back to L-Proline by via P5C reductase (P5CR). P5CR converts NADPH into $NADP^+$ which feeds into various reactions of the pentose phosphate pathway (purple arrows) and is linked to anaerobic metabolism (orange arrows). P5C can also be converted to L-Glutamate (L-Glu) through P5C dehydrogenase (P5CD), which can be converted back to P5C via P5C synthase (P5CS). L-Ornithine (L-Orn) also produces P5C via ornithine aminotransferase (OAT). L-Glu can be converted into alpha-ketoglutarate (α -KG) contributing to the TCA cycle and, consequently, oxidative phosphorylation. Pyruvate produced from glycolysis can contribute to the TCA cycle by production of acetyl-CoA, catalysed by pyruvate dehydrogenase (PDH) (blue arrows). PDH kinase (PDHK) inactivates PDH. GLUT: glucose transporter; MCT: monocarboxylate transporter; Glucose-6P: glucose 6-phosphate; R5P: ribose 5-phosphate; Gly3P: glyceraldehyde 3-phosphate; SucCoA: succinyl CoA; OAA: oxaloacetate. Figure produced on Biorender and adapted from (Kim *et al.*, 2006; Tanner, 2008; Phang *et al.*, 2010; Casalino *et al.*, 2011; Rodrigues *et al.*, 2015).

Low oxygen environments result in the stabilisation of hypoxia inducible factor-1 α (HIF1 α), which translocates into the nucleus to induce transcription of pluripotency genes *Oct4*, *Sox2* and *Nanog* (Forristal *et al.*, 2010) and key glycolytic genes (Ke and Costa, 2006). Glycolytic genes include glucose transporters *Glut1* and *Glut2*, and pyruvate dehydrogenase kinases (PDHKs) which limit pyruvate dehydrogenases catalysing acetyl-CoA production from pyruvate for the tricarboxylic acid (TCA) cycle (Figure 1.4) (Kim *et al.*, 2006; Prigione *et al.*, 2014). In ESCs, induction of HIF1 α and its downstream targets activates the glycolytic program and promotes pluripotency (Forristal *et al.*, 2010; Fiorenza and Rava, 2019). As ESCs differentiate to somatic cells (i.e., lose their pluripotency), their metabolic preference changes from anaerobic metabolism to OxPhos, in part, due to maturation of mitochondria and the increasing energy demand of differentiating cells (Rehman, 2010; Varum *et al.*, 2011; Agathocleous *et al.*, 2012; Panopoulos *et al.*, 2012; Knobloch and Jessberger, 2017; Arrázola *et al.*, 2019; Tsogtbaatar *et al.*, 2020).

1.3.2. Amino acids and lineage commitment

1.3.2.1. Amino acids strongly influence ESC function and differentiation

Amino acids play critical roles as signalling molecules that regulate transcriptional networks and epigenetics to influence ESC function and differentiation. For instance, L-Glutamine and L-Threonine contribute to maintaining self-renewal and proliferation of mouse ESCs (Ryu and Han, 2011; Chen and Wang, 2013; Ryu *et al.*, 2015; Kilberg *et al.*, 2016). In contrast, L-Proline drives mouse ESC differentiation to EPLCs (Washington *et al.*, 2010; Casalino *et al.*, 2011; Comes *et al.*, 2013; D’Aniello *et al.*, 2015).

Removal of L-Glutamine from mouse ESC culture causes downregulation of expression of pluripotency factors such as *Oct4*; upregulation of the expression of the mesoderm marker gene *Brachyury/T* and trophectoderm marker genes *Cdx2*, *Fgf4* and *Gata3*, and decreases expression of cell-cycle regulatory proteins, cyclin D1/CDK4 and cyclin E/CDK2 (Ryu *et al.*,

2015). This response to L-Glutamine depends on PI3K/AKT pathway activity, which subsequently stimulates the mTOR pathway via protein kinase C activation (Ryu *et al.*, 2015). The effects of L-Glutamine rely on its conversion to L-Glutamate. Inhibition of this conversion by blocking glutaminase activity using compound 968 results a reduction of *Oct4* and cell-cycle regulatory protein expression in mouse ESCs, which in turn reduces the activity of the pluripotency transcriptional network and proliferation, respectively (Ryu *et al.*, 2015). L-Glutamate can be converted to alpha-ketoglutarate (α -KG) which contributes to the TCA cycle (Figure 1.4). A high α -KG-to-succinate ratio is thus maintained, causing an overall increase in histone acetylation and DNA hypomethylation through inhibition of DNA methyltransferases, Dnmt3a and Dnmt3b. Some of these histone and DNA modifications occur in the *Oct4* gene, resulting in increased Oct4 expression and thus pluripotency maintenance. L-Glutamine-mediated signalling and epigenetic effects together contribute to mouse ESC self-renewal while preparing the cells for differentiation when the right signals are available (Ryu *et al.*, 2015; Shparberg *et al.*, 2019b).

Restricting L-Threonine in culture causes mouse ESCs to reduce proliferation, downregulate expression of self-renewal markers and increase expression of trophectoderm and mesoderm markers (Ryu and Han, 2011). Mouse ESC survival relies on threonine dehydrogenase activity to convert L-Threonine to Glycine for one-carbon metabolism, and to acetyl-CoA for energy production (Wang *et al.*, 2009; Kilberg *et al.*, 2016). Threonine dehydrogenase activity is rapidly downregulated upon mouse ESC differentiation (Wang *et al.*, 2009; Shyh-Chang, Ng *et al.*, 2013). L-Threonine also activates signalling pathways including PI3K/AKT, MAPK/ERK and mTOR, contributing to proliferation and self-renewal of mouse ESCs (Ryu and Han, 2011). L-Threonine metabolism is also coupled to S-adenosylmethionine (SAM) metabolism and is required to maintain a high SAM: S-adenosylhomocysteine (SAH) ratio (Shyh-Chang, Ng *et al.*, 2013). SAM is a substrate used for protein methylation, including histone methyltransferases used for epigenetic modifications, whereas SAH inhibits this action.

Thus, a high SAM:SAH ratio in mouse ESCs results in increasing histone H3 lysine 4 trimethylation (H3K4me3), leading to cell-cycle progression and self-renewal, and is essential for maintaining pluripotency (Bernstein *et al.*, 2006; Wang *et al.*, 2009, 2011; Shyh-Chang, N *et al.*, 2013).

In contrast to L-Glutamine and L-Threonine, L-Proline initiates a different outcome in mouse ESCs. L-Proline is a conditionally non-essential amino acid. This means L-Proline is produced in mouse ESCs but not to a sufficient concentration to drive all differentiation events (Wu, 2013; Wu *et al.*, 2013). Endogenous L-Proline synthesis is modulated and maintained via the amino-acid starvation response (AAR) pathway involving the Gcn2-Eif2 α -Atf4 circuit (i.e., general control non-repressed 2 - Eukaryotic Initiation Factor 2 α - Activating Transcription Factor 4) (D'Aniello *et al.*, 2015). Mouse ESCs starved of L-Proline results in Proline-free prolyl-tRNA binding to and activating the Gcn2 protein kinase, phosphorylating Eif2 α which leads to increased translation of Atf4. Atf4 upregulates gene expression of enzymes for endogenous L-Proline synthesis from L-Glutamate (D'Aniello *et al.*, 2015). A negative feedback loop then occurs where endogenous L-Proline that is produced reduces the amount of Proline-free prolyl-tRNA. As a result, mouse ESC identity is maintained and cell proliferation is limited (D'Aniello *et al.*, 2015; Tsogtbaatar *et al.*, 2020). Adding L-Proline exogenously overwhelms and shuts down the AAR pathway, shifting the balance of pathway activities to favour mouse ESC differentiation (D'Aniello *et al.*, 2015) (Figure 1.2). Adding L-Proline to mouse ESCs also increases trimethylation of histone H3 lysine 9 (H3K9me3) and H3 lysine 36 (H3K36me3), inducing changes in the epigenome and remodels the mouse ESC transcriptome to induce the EPLC transition (Comes *et al.*, 2013; Harvey *et al.*, 2016).

1.3.2.2. *L-Proline metabolism mediates anaerobic and aerobic metabolism in mESCs*

L-Proline enters mESCs via the SNAT2 transporter (Tan *et al.*, 2011). It is metabolised into pyrroline-5-carboxylate (P5C) by proline oxidase in the mitochondria. P5C can be converted back to L-Proline via P5C reductase. This conversion is redox coupled to the pentose phosphate

pathway (PPP) (Figure 1.4) (Tanner, 2008; Pandhare *et al.*, 2009; Phang *et al.*, 2010; Casalino *et al.*, 2011). The PPP is required for protein and nucleotide production and is coupled with anaerobic metabolism (i.e., the conversion of glucose to lactate) (Figure 1.4) to generate energy in the form of ATP in highly proliferative cells such as cancer cells and mESCs (Liu *et al.*, 2015).

Like L-Proline, adding L-Ornithine onto mouse ESCs drives differentiation of mESCs to EPLCs, as shown by gene expression changes associated with the transition (Casalino *et al.*, 2011). This probably occurs via the conversion of L-Ornithine to P5C by ornithine amino transferase which is then converted to L-Proline by P5C reductase (Figure 1.4). P5C can also be converted into L-Glutamate via P5C dehydrogenase, which can convert into α -KG to contribute to the TCA cycle, and eventually to OxPhos to produce ATP (Figure 1.4) (Phang *et al.*, 2010). Overall, this demonstrates L-Proline metabolism is capable of stimulating either the anaerobic or aerobic metabolic pathways (Figure 1.4).

1.3.2.3. Changes in metabolic preference across mouse ESC-to-neural lineage commitment

The overarching metabolic profile changes from anaerobic metabolism to OxPhos when ESCs lose their pluripotency (Rehman, 2010; Varum *et al.*, 2011; Agathocleous *et al.*, 2012; Panopoulos *et al.*, 2012; Knobloch and Jessberger, 2017; Arrázola *et al.*, 2019; Tsogtbaatar *et al.*, 2020). Continued culture of L-Proline on mouse ESCs not only drives differentiation to EPLCs, but also induces neural differentiation (Shparberg *et al.*, 2019a; Glover *et al.*, 2022). With neural differentiation, the metabolic switch correlates with decreased expression of glycolysis enzymes lactate dehydrogenase (LDH) and hexokinase 2 (HK2) and increased expression of master transcriptional regulators of mitochondrial biogenesis peroxisome proliferator-activated receptor α (PPAR α), PPAR α coactivator-1 α and mitochondrial transcription factor A (Chen, C.-T. *et al.*, 2010; Cheng *et al.*, 2012; O'Brien *et al.*, 2015; Agostini *et al.*, 2016; Arrázola *et al.*, 2019). This results in an upregulation of mitochondrial function (Chen, C.-T. *et al.*, 2010; Cheng *et al.*, 2012; O'Brien *et al.*, 2015; Agostini *et al.*,

2016; Arrázola *et al.*, 2019). Reduction in aerobic glycolysis is shown to be essential for neural progenitor-to-neural cell differentiation, indicating a metabolic preference of OxPhos over anaerobic metabolism in neurons (Zheng *et al.*, 2016). PI3K/AKT, mTOR and MAPK/ERK pathway activities are involved in mitochondrial biogenesis and function during neural differentiation and are involved in mouse ESC differentiation in response to L-Proline (Figure 1.2) (Lonic, 2007; Washington *et al.*, 2010; Tan *et al.*, 2011; Agostini *et al.*, 2016). Thus, metabolic mechanisms are worth investigating within the intermediate cell types (primitive ectoderm, definitive ectoderm and neurectoderm) that lie between mouse ESCs and mature neural cells in response to L-Proline. Changes in the metabolic profile can be verified using the Seahorse Extracellular Flux assays (Martano *et al.*, 2019). Metabolic pathways relevant in definitive ectoderm and neurectoderm can be revealed by detecting metabolites using liquid chromatography/mass spectrometry (Martano *et al.*, 2019).

1.4. Transcriptional mechanisms in stem-cell fate

1.4.1. *Core and extended pluripotency networks maintain ESC self-renewal and pluripotency*

The molecular circuitry regulating pluripotency and self-renewal in ESCs involves a core network of transcription factors comprised of Oct4, Nanog and Sox2 (Boyer *et al.*, 2005; Chambers *et al.*, 2007; Xue *et al.*, 2011; Glover, 2018; Shparberg *et al.*, 2019b). The concentrations, interactions, and functions of these proteins all determine the activity of this core network (Mitsui *et al.*, 2003; Niwa *et al.*, 2009; Thomson *et al.*, 2011; Xue *et al.*, 2011). These transcription factors regulate at least 353 genes, including their own as well as "extended pluripotency network" genes like *Fgf4*, *Rex1* and *Klf4* (Boyer *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005; Xue *et al.*, 2011). This extended network helps sustain the core network activity and suppresses lineage commitment pathways and the expression of differentiation-associated genes (Boyer *et al.*, 2005; Xue *et al.*, 2011; Shparberg *et al.*, 2019b). For example, Oct4 inhibits

NeuroG1, which is associated with neurogenesis (Boyer *et al.*, 2005; Jaenisch and Young, 2008). Downregulation of Nanog protein causes downregulation of proteins associated with the pluripotency network, including itself, *Sall4*, *Oct4* and *Esrrb* (Lu *et al.*, 2009). The downregulation of Nanog in ESCs also provides a temporary opportunity to prime for differentiation (Chambers *et al.*, 2007; MacArthur *et al.*, 2009). Differentiation-associated genes have bivalent histone modifications, meaning they have both active (H3K4me3) and repressive (H3K27me3) epigenetic marks (Figure 1.5). As a result, these genes are silent in ESCs whilst at the same time being primed for transcription under the right conditions (Bernstein *et al.*, 2006).

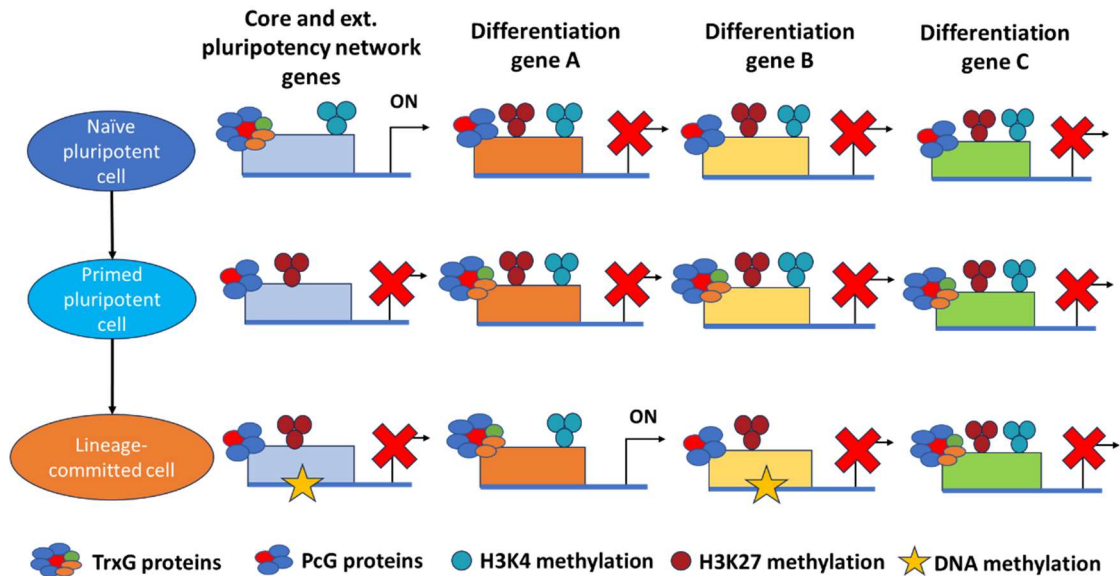


Figure 1.5 Recruitment of polycomb and trithorax group proteins to repress and express genes during cell-fate commitment. In the cell types denoted on the left-hand side, pluripotent cells transition across a spectrum and become primed before lineage commitment to a particular cell fate. **Core and extended (ext.) pluripotency network genes:** In naïve pluripotent cells, core pluripotency genes, such as *Oct4*, and ext. pluripotency network genes, such as *Rex1*, carry the active H3K4 trimethylation mark mediated by trithorax group (TrxG) proteins. These marks maintain open chromatin and the genes are expressed. In primed pluripotent cells, the genes are methylated at H3K27 mediated by polycomb group (PcG) proteins. This repressive mark and the bound PcG contributes to chromatin compaction and reduced expression, but the genes are poised to be responsive to differentiation stimuli. In lineage-committed cells, PcG proteins Ezh1/2 catalyse DNA methylation of core and ext. pluripotency genes on top of chromatin compaction to ensure full silencing. For differentiation-associated genes, there are three ways that PcG and TrxG proteins coordinate to mediate repression and expression of genes: **Differentiation gene A:** In naïve pluripotent cells, these genes (such as neuroectoderm marker *Sox1*) are bivalent; i.e., they have the H3K27 and H3K4 trimethylation marks. Recruitment of PcG proteins mediates gene repression in numerous ways, including chromatin compaction and transcription interference preventing RNA polymerase II activity or SWI/SNF accessibility to promoter regions. In primed pluripotent cells, these genes have acquired TrxG proteins to open chromatin. However, the bivalent chromatin structure keeps these genes in a repressed state but are prepared for lineage-specific transcription. In lineage-committed cells, most genes with these bivalent domains are resolved to contain only active H3K4 trimethylation marks mediated by TrxG proteins. Chromatin remain open and RNA polymerase II activity and/or SWI/SNF have access to promoter regions to contribute to gene expression. **Differentiation gene B:** In naïve pluripotent cells, these genes (such as mesendoderm markers *Brachyury* and *Mixl1*) are bivalent with PcG-mediated chromatin compaction and transcription interference, and so remain repressed. In primed pluripotent cells, these genes have open chromatin mediated by TrxG proteins to prepare to respond to differentiation stimuli. In lineage-committed cells, DNA methylation, combined with trimethylation of H3K27, chromatin compaction and promoter regions marked by PcG proteins ensures these genes are completely silenced. **Differentiation gene C:** In naïve pluripotent cells, some genes (such as transcription factors regulating hematopoietic development *Gata1* and *Hes1*), are bivalent and remain repressed by PcG protein-mediated chromatin compaction and transcription interference. In primed pluripotent cells, PcG proteins are displaced in response to certain differentiation stimuli and these genes acquire TrxG proteins to mediate and maintain chromatin opening. In lineage-committed cells, these genes have chromatin that remains open to prepare to respond to other differentiation stimuli at the right time and location, allowing for multilineage priming; i.e., multipotency. By default, without differentiation stimuli, differentiation-associated genes not bound by TrxG proteins are subjected to PcG-mediated repression to maintain long-term memory of expression profiles upon replication. Figure derived from: (Santos *et al.*, 2002; Ringrose and Paro, 2004; Bernstein *et al.*, 2006; Boyer *et al.*, 2006; Sampath *et al.*, 2008; Wu *et al.*, 2008; Lu *et al.*, 2009; MacArthur *et al.*, 2009; Christophersen and Helin, 2010; De Gobbi *et al.*, 2011; Herz *et al.*, 2013; Kingston and Tamkun, 2014; Lau and So, 2015; Sze *et al.*, 2017; Li and Belmonte, 2017; Takahashi *et al.*, 2018; Argelaguet *et al.*, 2019; Brand *et al.*, 2019; Luo *et al.*, 2020; Tam and Ho, 2020).

When differentiation-associated genes are silenced, they are occupied by polycomb group (PcG) proteins, which act as transcription repressors, helping to maintain ESC pluripotency (Figure 1.5 and 1.6) (Ringrose and Paro, 2004; Boyer *et al.*, 2006; MacArthur *et al.*, 2009). PcG proteins form polycomb-repressive complexes 1 (PRC1) and 2 (PRC2). PRC1 contains 4 essential subunits: The E3 ligase ring finger protein 1A or 1B (RING1A/B), a posterior sex comb ortholog from the polycomb group RING finger protein family (PCGF1-6), a polyhomeotic homolog (PHC1-3) and a chromobox protein (Cbx2/4/6-8) (Figure 1.6A). PRC2 consists of 3 essential subunits: Enhancer of zeste 1 (EZH1) or its homolog EZH2, embryonic ectoderm development (EED) protein and suppressor of zeste 12 (SUZ12) (Morey and Helin, 2010; Morey *et al.*, 2012; Di Croce and Helin, 2013; Aranda *et al.*, 2015; Santanach *et al.*, 2017) (Figure 1.6B). Both complexes are associated with the repressive epigenetic mark H3K27me3 since PRC2 represses transcription of genes by catalysing trimethylation of H3K27. H3K27me3 acts as a docking site for Cbx protein core units in PRC1. PRC1 then catalyses mono-ubiquitination of histone H2A at lysine 119 by RING1A/B activity. Other PRC1 components BMI1, MEL-18, and Cbx family proteins have an affinity for H3K27me3 (Figure 1.6A) (Ku *et al.*, 2008; Morey and Helin, 2010; Morey *et al.*, 2012; Di Croce and Helin, 2013). This provides a mechanism in ESCs for these complexes to co-repress 512 genes, including genes from the Fox, Sox, Gata and Tbx transcription factor families that are associated with development (Boyer *et al.*, 2006).

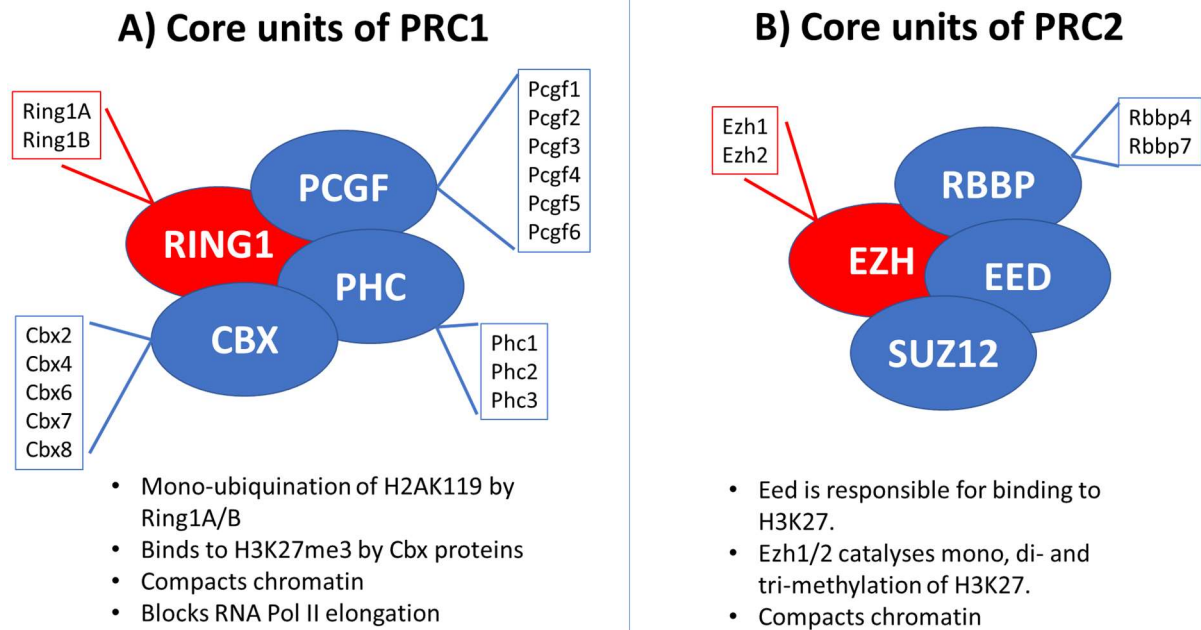


Figure 1.6 Core units of A) polycomb-repressive complexes 1 (PRC1) and B) PRC2 and their mammalian homologs. Enzymatic components are depicted in red. Core components are depicted in blue. The mammalian homologs are depicted in boxes. Functions of each PRC are listed under each complex. Different homologs in mammals allow multiple forms of PRC1 and PRC2 complexes. RING1: Real interesting new gene 1; Pcgf: Posterior sex comb ortholog from the Polycomb group RING finger protein family; Phc: Polyhomeotic homolog; Cbx: chromobox protein; Ezh: Enhancer of zeste; RBBP: Retinoblastoma binding protein; EED: Embryonic ectoderm development protein; SUZ12: Suppressor of zeste 12. Figure adapted from: (Ku *et al.*, 2008; Morey and Helin, 2010; Morey *et al.*, 2012; Di Croce and Helin, 2013; Lau and So, 2015).

Core pluripotency network transcription factors Oct4, Nanog and Sox2 and their genes remain expressed to maintain pluripotency. Trithorax group (TrxG) proteins occupy these genes, maintaining active gene expression (Figure 1.5 and 1.7). This is achieved via TrxG-mediated trimethylation of H3K4 (H3K4me3) mark, which is associated with expression, and other posttranslational modifications associated with TrxG complexes (Kingston and Tamkun, 2014; Lau and So, 2015). TrxG proteins form a more heterogeneous group of complexes than PcG proteins (Figure 1.7) (Schuettengruber *et al.*, 2007, 2011; Lau and So, 2015). They are categorized into 2 groups, based on their targeted modifications: Histone modifiers and chromatin remodelling.

Histone modifiers contain a SET protein. In mammals, SET has 6 homologs: SET1A/B, MLL1/2 and MLL3/4 (Miller *et al.*, 2001) (Figure 1.7). Each of these proteins form a complex associated with trithorax-related SET protein (COMPASS) and catalyses histone methyltransferase activity to methylate H3K4 (Lau and So, 2015). All 6 COMPASS's share common subunits that include WD repeat-containing protein 5 (WDR5), retinoblastoma binding protein 5 (RBBP5), absent, small, or homeotic discs 2-like protein (ASH2L) and Dumpy-30 (Dpy-30) (Figure 1.7) (Dou *et al.*, 2006; Southall *et al.*, 2009; Ang *et al.*, 2011; Jiang *et al.*, 2011; Piunti and Shilatifard, 2016). RBBP5 and ASH2L stimulates histone methyltransferase activity, WDR5 is required to maintain complex integrity and substrate binding, while Dpy-30 is responsible for regulating methylation of H3K4 (Dou *et al.*, 2006; Southall *et al.*, 2009; Ernst and Vakoc, 2012). COMPASS complexes also contain specific subunits that may confer different functions due to their roles in regulating recruitment to different target genes and interactions with other proteins (Figure 1.7) (Wu *et al.*, 2008; Herz *et al.*, 2013; Lau and So, 2015).

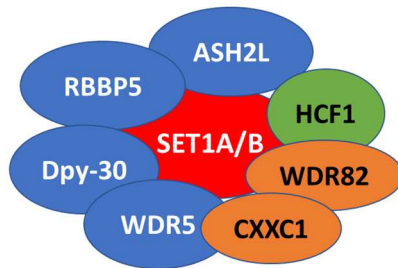
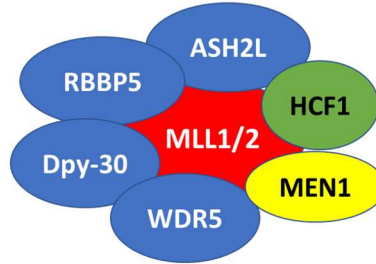
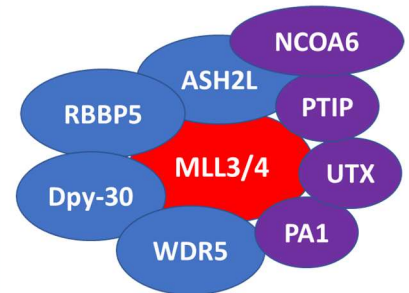
A) SET1A/B complex**B) MLL1/2 complex****C) MLL3/4 complex**

Figure 1.7 COMPASS complexes are composed of trithorax proteins. Mammals have 6 COMPASS complexes, which can be categorized into **A) SET1A/B**, **B) MLL1/2** and **C) MLL3/4**. The SET domain enzymatic component that catalyses methylation of H3K4 is depicted in red. Core components are depicted in blue. Complex-specific subunits are indicated by the unique colours. SET1A/B: Su[*var*]3-9, Enhancer of Zeste, Trithorax 1A/B; MLL1/2: Mixed-lineage leukemia 1/2; MLL3/4: Mixed-lineage leukemia 3/4; WDR5: WD repeat-containing protein 5; RBBP5: Retinoblastoma binding protein 5; ASH2L: Absent, small, or homeotic discs 2-like protein; Dpy-30: Dumpy-30; HCF1: Host cell factor 1; WDR82: WD repeat-containing protein 82; CXXC1: CXXC finger 1; MEN1: multiple endocrine neoplasia 1; NCOA6: nuclear receptor coactivator 6; PTIP: PAX transcription activation domain– interacting protein; UTX: Ubiquitously transcribed X chromosome tetratricopeptide repeat protein; PA1: PTIP-associated 1. Figure derived from: (Dou *et al.*, 2006; Wu *et al.*, 2008; Herz *et al.*, 2013; Lau and So, 2015; Piunti and Shilatifard, 2016).

Chromatin remodelling TrxG complexes are ATP-dependent and contain a catalytic ATPase subunit (Schuettengruber *et al.*, 2011; Lau and So, 2015). They use hydrolysis of ATP to disrupt DNA-histone contacts, resulting in changes in position and composition of nucleosomes on chromatin (Hargreaves and Crabtree, 2011; Lau and So, 2015). This is mediated by either exchanging histone variants, complete eviction of nucleosomes from chromatin, or sliding nucleosomes along DNA (Hargreaves and Crabtree, 2011; Lau and So, 2015). These various mechanisms allow varying accessibility of DNA to cellular proteins and nucleosome positioning, which can induce gene repression (Hargreaves and Crabtree, 2011; Lau and So, 2015). In ESCs, most COMPASS subunits are not required for self-renewal and maintaining pluripotency except ASH2L and WDR5. ASH2L is essential to maintain open chromatin and pluripotency in mouse ESCs while WDR5 directly interacts with Oct4 (Ang *et al.*, 2011; Wan *et al.*, 2013; Tsai *et al.*, 2019). Subunits from the SWI/SNF chromatin remodelling TrxG complex occupies promoters of core pluripotency genes *Oct4*, *Nanog* and *Sox2*; and *Stat3* and *Smad1*, reflecting their role in LIF and BMP signalling for self-renewal and pluripotency (Figure 1.2) (Ying *et al.*, 2003; Ho *et al.*, 2009; Hirai *et al.*, 2011; Lau and So, 2015).

Overall, PcG and TrxG proteins coordinate to maintain gene expression patterns during development: PcG proteins repress gene expression via chromatin compaction while TrxG proteins antagonise this by mediating chromatin opening at specific loci determined by lineage-specific transcription factors (Figure 1.5) (Brand *et al.*, 2019). Pluripotency is maintained by stabilising the expression of core pluripotency genes *Oct4*, *Nanog* and *Sox2* and other target genes such as extended pluripotency network genes *Rex1* and *Klf4* that mediate self-renewal, whereas differentiation-associated transcription factors have their targeted loci repressed via PcG proteins. TrxG proteins are recruited to specific loci to establish cellular memory that allows transcription of the same set of gene expression patterns in daughter cells upon replication. This results in ESCs maintaining pluripotency and self-renewal upon replication.

1.4.2. Multiple molecular networks work interdependently to drive ESC differentiation and lineage commitment

1.4.2.1. Transcriptional control works interdependently with epigenetics, translational control, and signalling to drive cellular activity during embryo development

The transition to different cell fates is driven by transcription factors, which works with other regulatory mechanisms including epigenetics, translational control and signalling (Sampath *et al.*, 2008; Lu *et al.*, 2009; MacArthur *et al.*, 2009; Martynoga *et al.*, 2012; Tam and Ho, 2020). Transcription factors have access to enhancer sites to drive cell fate in a context-specific manner, partly due to the dynamic nature of DNA methylation mediated by epigenetic mechanisms (Figure 1.5) (Santos *et al.*, 2002; Luo *et al.*, 2020). Loss of pluripotency is associated with changes in epigenetic-transcriptional molecular activity and increasing protein translation efficiency (Sampath *et al.*, 2008; Lu *et al.*, 2009; Argelaguet *et al.*, 2019; Tam and Ho, 2020). Pluripotent cells have a generalised global repressive landscape, which is followed by lineage-specific epigenetic changes upon gastrulation (Argelaguet *et al.*, 2019; Luo *et al.*, 2020). Loss of methylation and increased chromatin activity in regulatory elements are associated with expression of transcriptional factors that determine lineage commitment and specificity. This priming of transcription factors via epigenetic modifications is what drives the emergence of the three germ layers during gastrulation (Mohammed *et al.*, 2017; Argelaguet *et al.*, 2019; Tam and Ho, 2020). For instance, DNA methylation enzyme Ten–Eleven–Translocation DNA dioxygenase 1 (Tet1) coordinates with pluripotency transcription factors such as Oct4, and differentiation-inducing signalling pathways such as Nodal-Smad2/3 and Wnt/ β -catenin signalling to drive mesendoderm and neurectoderm lineage commitment choice (Funa *et al.*, 2015; Wang *et al.*, 2017; Luo *et al.*, 2020). Tet1 acts as an upstream negative regulator of Wnt/ β -catenin signalling which then favours neural lineage commitment (Luo *et al.*, 2020). Tet1 also indirectly regulates chromatin activity of Oct4, allowing proper lineage commitment to the neurepithelium by preventing Nodal and Wnt signalling (Luo *et al.*, 2020).

In addition, Tet1 coordinates with zinc finger protein of the cerebellum 2 and basic helix–loop–helix (bHLH) transcription factors Tcf12 and NeuroD1 to occupy loci in the epiblast and regulate epiblast cell fates (Luo *et al.*, 2020).

PcG and TrxG proteins establish an antagonistic relationship to maintain pluripotency and self-renewal in ESCs (Section 1.4.1). For differentiation to occur, the PcG-TrxG antagonistic relationship must be disrupted in order for ESCs to temporarily “forget” their current cell fate and establish a new one. This can happen by providing the right internal and/or external signals to shift this relationship (Brand *et al.*, 2019).

PcG proteins compact chromatin at specific gene loci, allowing these genes to be "avoided" by developmental programs and repress certain alternate cell fates (Figure 1.5). For instance, PRC2 complexes ensure efficient specification towards the definitive ectoderm lineage, which acts via PRC2 core subunit Ezh2 to repress mesendoderm-specific genes during human ESC-derived embryoid body formation (Collinson *et al.*, 2016; Shan *et al.*, 2017). PRC1 complexes drive ESC specification to the mesodermal lineage via its core subunit PCGF2 by positively regulating mesoderm-specific transcription factors and repressing pluripotency and ectodermal genes (Morey *et al.*, 2015). Cbx2-containing PRC1 complexes repress trophoblast-specific genes but Cbx4-containing PRC1 complexes do not do this. However, Cbx2- and Cbx4-containing PRC1 complexes commonly repress mesoderm and endoderm lineage genes (Morey *et al.*, 2012). PRC1 also can act to suppress expression of signalling pathways (Brand *et al.*, 2019). For example, the PCGF6-containing PRC1 complex represses the SMAD2-TGF β signalling pathway that would favour mesendoderm lineage commitment and so serve to drive neural differentiation from ESCs (Yao *et al.*, 2018).

TrxG proteins help establish and maintain cell-specific gene expression patterns during differentiation (Figure 1.5). COMPASS complexes have evolved to create active gene expression memory at loci with differentiation-associated genes (Wang *et al.*, 2012, 2016; Sze

et al., 2017). For instance, proliferating ESCs become primed for differentiation by Set1A-COMPASS complexes marking differentiation-associated genes with H3K4me3 to create bivalently marked promoters (Figure 1.5) (Sze *et al.*, 2017). Upon differentiation, MLL3/4-COMPASS complexes target transcription factor-dependent enhancers of lineage-specific genes to allow activation of gene expression programs such as *Brachyury* and *Wnt3* to contribute to mesodermal-lineage commitment (Figure 1.5) (Wang *et al.*, 2012, 2016). SWI/SNF complexes are also essential to mediate ESC differentiation to different lineages, depending on their subunits. For example, BAF250A-containing SWI/SNF complexes are essential for mesodermal and endodermal lineage commitment (Gao *et al.*, 2008). Overexpression of SWI/SNF complex subunit BAF170 favours ectodermal lineage commitment over endoderm formation in human ESCs (Wade *et al.*, 2015). Overall, various PcG and TrxG mechanisms work together with signalling and transcription to allow ESCs to partially erase their cell-fate memory in order to differentiate (Figure 1.5) (Brand *et al.*, 2019).

Transcriptional activities also directly influence signalling to drive cell fate (Martynoga *et al.*, 2012; Tam and Ho, 2020). For instance, a major binding partner of Id proteins in mouse pluripotent cells is the bHLH transcription factor E2A (Rao *et al.*, 2020). E2A gene products regulate transcription of target genes via homodimerisation or heterodimerisation with class II bHLH transcription factors, like proneural factors *Ascl1* and *NeuroG1/2* (Murre *et al.*, 1989; Imayoshi and Kageyama, 2014; Le Dréau *et al.*, 2018; Rao *et al.*, 2020). E2A is located upstream from neural lineage-associated genes including *Sox1* and *Foxd4*. It instructs the promotion of neural lineage commitment in pluripotent cells by promoting transcription of neural lineage-associated genes while suppressing Nodal signalling activity, which directs differentiation to mesendoderm (Rao *et al.*, 2020). Transcriptional activity changes in sequential waves across developmental time, not just simply a “one after the other” scenario (Telley *et al.*, 2016; Marcy and Raineteau, 2019). Overall, transcriptional mechanisms play an important part with other regulatory networks to drive differentiation and lineage commitment.

1.4.2.2. *A better understanding of the transcriptome is required to understand cell-fate specification.*

Regulatory control of transcriptional activity during embryo development is required to drive pluripotent cells through multiple metastable states, lose pluripotency, and undergo appropriately timed and located lineage commitment toward somatic cell types in the face of the many other ‘noisy’ molecular processes taking place (Sampath *et al.*, 2008; Lu *et al.*, 2009; MacArthur *et al.*, 2009). For example, early transcriptional waves in newly produced mouse neurons *in vivo* instruct sequencing and pacing of neuronal differentiation events (Telley *et al.*, 2016). These waves help guide newborn neurons to their final fate. These transcriptional waves can be used as a means to identify neural cell-fate specification (Kuegler *et al.*, 2010; Spangler *et al.*, 2018; Marcy and Raineteau, 2019). Analysing diverging gene expression profiles can uncover potential mechanisms that lead to different cell fates (Spangler *et al.*, 2018; Tam and Ho, 2020). For example, in the process of neurogenesis, all proliferative transcripts are repressed first, then neurogenic transcripts repressed and finally neuronal transcripts induced after cell division (Telley *et al.*, 2016). Rn7SK, a highly conserved small nuclear RNA, which is transcribed by RNA Polymerase III, mediates a global transcriptional regulation programme that has an important role in driving ESC-to-neural cell transition (Bazi *et al.*, 2018), whereas *Brachyury* and *Wnt3* expression patterns indicate a mesodermal-lineage commitment (Wang *et al.*, 2012, 2016).

Transcription factors of the core pluripotency network Oct4, Sox2 and Nanog regulate the transcriptional network necessary to maintain a gene expression profile to sustain self-renewal and pluripotency (Boyer *et al.*, 2005; Xue *et al.*, 2011; Shparberg *et al.*, 2019b). These transcription factors also occupy gene promoters for glycolytic enzymes such as hexokinase 2 (*Hk2*) and pyruvate kinase M2 (*Pkm2*) in ESCs (Kim *et al.*, 2015). Maintenance of glycolysis by overexpression of *Hk2* and *Pkm2* causes mESCs to maintain pluripotency and self-renew, even in the absence of LIF (Kim *et al.*, 2015). Addition of 3-bromopyruvate, an analogue of

pyruvate that is used to inhibit HK2 activity to inhibit glycolysis, causes mESCs to switch from glycolysis to OxPhos and lose pluripotency even when cultured in the presence of LIF (Rodrigues *et al.*, 2015). The non-coding RNA Lncenc1 also regulates and enhances transcription of glycolysis-associated genes by interacting with RNA-binding proteins polypyrimidine tract-binding protein 1 (PTBP1) and heterogeneous nuclear ribonucleoprotein K (HNRNPK) (Sun *et al.*, 2018; Nishimura *et al.*, 2019). Ablation of Lncenc1 in mESCs results in poor colony formation, significantly reduces glycolysis-associated gene expression, >50% decrease in glucose consumption and lactate production, and decreases expression of core pluripotency genes. This indicates Lncenc1, PTBP1 and HNRNPK are all necessary to maintain self-renewal in ESCs (Sun *et al.*, 2018; Nishimura *et al.*, 2019) and shows that transcriptional mechanisms can directly influence metabolism and vice versa to influence cell fate (Son and Han, 2018; Nishimura *et al.*, 2019). Therefore, it is worth investigating if metabolic processes are driving differentiation by mediating changes in gene expression profiles. An understanding of the transcriptome will reveal potential mechanisms that drive cell-fate specification from pluripotent cells.

1.4.3. Profiling the transcriptome can be used to predict mechanisms, including metabolic mechanisms, driving differentiation of mESCs to neural-cell fate

1.4.3.1. Single-cell RNAseq may reveal novel molecular mechanisms on differentiation while accounting for heterogeneity and revealing new cell types and states

Analysis of the transcriptome across ESC differentiation will provide an overview of potential mechanisms driving cell-fate specification. DNA microarray analysis using EBs as an *in vitro* model to reconstruct molecular mechanisms driving embryo development (Bratt-Leal *et al.*, 2009; Lu *et al.*, 2009; MacArthur *et al.*, 2009) show that EBs often express genes transiently in a manner that recapitulates normal development from primitive ectoderm to gastrulation to early cell-lineage commitment (Dvash *et al.*, 2004). These transcriptional changes also indicate a complementary relationship with the proteome (Fathi *et al.*, 2009). Traditional microarray

analysis reveals the appearance and disappearance of mRNAs. However, a combination of microarray and RNA sequencing (RNAseq) reveals non-coding RNAs also play a role in driving development through mediating gene expression programming (Dinger *et al.*, 2008; Pauli *et al.*, 2011; Hu *et al.*, 2012; Bazi *et al.*, 2018; Fico *et al.*, 2019). For example, microRNAs in the Let-7 family and pluripotency factor LIN28 establish the antagonistic switch between pluripotency and differentiation, where in mESCs LIN28 is on and Let-7 is off, while in differentiating cells the opposite occurs (reviewed by Pauli *et al.*, 2011). LIN28 is an RNA-binding protein that binds to Let-7 at the terminal loop region to inhibit Let-7 activity to maintain pluripotency (Newman *et al.*, 2008; Piskounova *et al.*, 2008; Melton *et al.*, 2010). With insufficient LIN28, Let-7 represses multiple pluripotency and self-renewal genes to initiate somatic cell differentiation (Melton *et al.*, 2010).

Bulk RNAseq is useful in looking at global transcriptome changes and profiles. For instance, bulk RNAseq revealed decreased expression of pluripotency marker genes, including *Nanog*, *Pou5f1* and *Esrrb*, in mouse EBs over 6 days of differentiation (Spangler *et al.*, 2018). This mimics the transcriptional changes in *in vivo* embryogenesis from E3.5–6.5 and priming for the production of the post-implantation epiblast state (Spangler *et al.*, 2018). Bulk RNAseq of mouse EPLCs shows a transcriptional profile that diverges from EpiSCs derived from the same ESC line: EPLCs are already primed for neuroectoderm commitment while EpiSCs are primed for mesendoderm commitment (Glover, 2018).

Bulk RNAseq cannot easily account for heterogeneity within the tested sample. Different cell types may escape detection or be detected as noise, and heterogeneity of expression within cells of the same type cannot be determined. RNAseq at the single-cell level largely overcomes these problems and can help identify previously uncharacterised cell types or states (Kumar *et al.*, 2017; Mohammed *et al.*, 2017; Spangler *et al.*, 2018; Zeng *et al.*, 2018; Argelaguet *et al.*, 2019; Nishimura *et al.*, 2019; Tam and Ho, 2020). For instance, does a portion of the cells in an EB remain pluripotent even under conditions of spontaneous differentiation? This question

cannot be addressed at the bulk RNAseq level. Day 6 EBs do contain a subset of cells having a pluripotent stem cell signature, including *Zfp42* expression (Spangler *et al.*, 2018; Tam and Ho, 2020). Single-cell RNAseq of EB stages reflecting embryos at E3.5–6.5 also reveals clusters corresponding to various stages of embryo development (Spangler *et al.*, 2018). Day 0 EB population is significantly enriched in E3.5 ICM and E4.5 epiblast clusters. Day 4 EB population, has the highest enrichment in E6.5 primitive streak clusters (Spangler *et al.*, 2018). Single-cell RNAseq can determine the relationship between cell states and their developmental progress within a narrow timeframe (reviewed by Kumar *et al.*, 2017; Tam and Ho, 2020). For example, single-cell RNAseq shows changes in transcriptional profiles during human inner cell mass to an early-to-late epiblast stage (Warrier *et al.*, 2018). The study reflects a crucial developmental timepoint where transcriptional rewiring is required for human ICM cells to leave the naïve pluripotency state and begin the primed pluripotency state (Warrier *et al.*, 2018). Single-cell RNAseq has also shown decreasing signalling activity, increasing translational machinery and regulatory switches in transcription factors, metabolism and energy sources across adult hippocampal quiescent neural stem cells to neurogenesis initiation (Shin *et al.*, 2015). All proliferative transcripts were repressed first, then neurogenic transcripts were repressed, and neuronal transcripts were induced after cell division. These transcripts occur in waves, not simply a “one after the other” scenario, to instruct sequencing and pacing of neurogenesis events. These waves help guide newly developed neurons to their final fate (Telley *et al.*, 2016). As there are a lack of markers for definitive ectoderm, single-cell RNAseq may help identify these cell types across neural lineage commitment (Loebel *et al.*, 2003; Tam *et al.*, 2006; Arnold and Robertson, 2009; Harvey *et al.*, 2010; Li *et al.*, 2013). However, care must be taken in regards to identification of these new cell subpopulations, as the number of parameters must be defined per step of the pipeline (Kiselev *et al.*, 2017). This itself can be time-consuming and difficult to optimise, as it risks producing false positives and negatives with respect to cell subpopulations (Kiselev *et al.*, 2017).

1.4.3.2. *Bioinformatics pipelining is required to process raw RNAseq data to generate hypotheses for further studies*

Raw bulk or single-cell RNAseq data must go through various pipelining processes before anything meaningful can be interpreted (Koch *et al.*, 2018). The RNAseq process involves RNA, either as total RNA, enriched for mRNA, or with ribosomal RNA depleted, converted into cDNA (Koch *et al.*, 2018). The cDNA samples are then fragmented and have adapters and indexes ligated onto the cDNA fragments for sequencing (Koch *et al.*, 2018). These fragments are to be sequenced or “read” using a high-throughput platform to generate raw RNAseq read datasets (Koch *et al.*, 2018).

These raw sequences then undergo various pipelining processes. First, raw sequences must be assessed for their read quality using FastQC (Andrews, 2010). Each base in the sequence is analysed to see if it has been read correctly; i.e., the base-calling error is low. With good quality reads, the raw sequences are then aligned and mapped to a reference genome and sequencing reads are quantified. For instance, alignment and quantification of reads can be done at once using Spliced Transcripts Alignment to a Reference software (Dobin *et al.*, 2013). Raw reads are aligned to a reference genome and mapped to genes using an assigned gene annotation of choice; e.g., from Ensembl. Reads are then quantified to produce a raw counts table, which is subjected to bioinformatic tools and programs to do the following: produce principle component analysis or Uniform Manifold Approximation and Projection plots from bulk and single-cell RNAseq raw counts, respectively, to reduce the number of dimensions for easier visualisation and identification of variability between and within groups; display heat maps to visualise patterns of gene expression (i.e., genes that have been up- and/or downregulated); run differential expression analysis to count the number of reads within each gene (Koch *et al.*, 2018; McInnes *et al.*, 2018). The differences in dimensionality reduction between bulk and single-cell RNAseq is due to higher variability between samples in single-cell compared to bulk RNAseq. The use of Uniform Manifold Approximation and Projection can account for this

higher variability than principle component analysis (Koch *et al.*, 2018; McInnes *et al.*, 2018). Differentially expressed genes (DEGs) are clustered by grouping them based on their distance relative to each other in an unbiased manner (Kumar *et al.*, 2017; Koch *et al.*, 2018). These clusters are useful to help identify different cell types or states across a developmental timeframe based on a group of genes and to select genes for further study (Kumar *et al.*, 2017; Argelaguet *et al.*, 2019; Nowotschin *et al.*, 2019; Tam and Ho, 2020). ToppGene can be used to annotate these clusters and assign biological relevance (Chen *et al.*, 2009). With ToppGene, DEG clusters can be analysed for Kyoto Encyclopedia of Genes and Genomes pathways and Gene Ontology (GO) classifications, which has been used to identify markers as reference points for development (Chen *et al.*, 2009; Kuegler *et al.*, 2010). GO classifications are also further categorized into biological processes and molecular functions (Chen *et al.*, 2009). The significance of the assigned genes for a GO classification of interest is assessed using a hypergeometric test, where the probability that an overlap between the number of genes associated with the GO classification and the number of genes in the targeted set could occur by chance, taking into account the number of genes in the background; i.e., the other genes not associated with the GO classification (Falcon and Gentleman, 2008; Koch *et al.*, 2018). It should be noted that the tools used to produce comparisons between DEG clusters do not account for multiple hypothesis testing, and thus do not actually reflect the true relative expression changes (MacArthur *et al.*, 2009; Koch *et al.*, 2018). Identification of these processes, however, are simply hypotheses of the possible drivers of changes between these clusters (MacArthur *et al.*, 2009; Koch *et al.*, 2018). Other methods must be used to validate these processes. For instance, western blots or qPCR should be used to validate the expression of key genes (Lu *et al.*, 2009; MacArthur *et al.*, 2009; Glover, 2018). Ultimately, hypothesised pathway(s) causing certain biological function(s) should be validated by functional studies such as using pathway inhibitors or gene knockout studies (Lu *et al.*, 2009; MacArthur *et al.*, 2009; Glover, 2018; Koch *et al.*, 2018).

1.5. Mechanisms driving mESCs to a neural fate during development may highlight mechanisms responsible for the onset of neurodegenerative diseases later in life

Neurodegenerative disease is the umbrella term for diseases caused by the degeneration of neurons. For example, motor neuron disease (MND) results from degeneration of motor neurons. In 2019, there was around 270 000 reported cases (including familial and sporadic forms) of MND worldwide and over 39 000 deaths (Park *et al.*, 2022), with case numbers and number of deaths due to MND predicted to increase (Xu *et al.*, 2020; McCluskey *et al.*, 2022; Park *et al.*, 2022). Although familial cases of MND, where there is an inherited MND-associated gene that passes on from one generation to the next, have been studied extensively, much more limited knowledge is available about mechanisms causing sporadic MND, where a combination of both spontaneous genetic mutations and environmental factors trigger onset.

MND is a genetically complex disease because multiple associated genes contribute (Sterneckert *et al.*, 2014; Teoh *et al.*, 2017; Karagiannis and Inoue, 2020). While late onset is common, early onset triggered anywhere from birth to adulthood also occurs (Teoh *et al.*, 2017). This also indicates that the mechanisms that trigger degeneration of neurons in MND is complex and multifactorial.

1.5.1. MND-associated genes and proteins and their contribution to the mechanism of neural degeneration

The prevalence of sporadic MND cases is increasing as the population ages (Xu *et al.*, 2020; McCluskey *et al.*, 2022). This begs the question: what triggers the genetic and protein mutation(s) profile in the first place? Mutations of certain genes and proteins are associated with the onset of MND (Sterneckert *et al.*, 2014; Karagiannis and Inoue, 2020). These mutations have been used as diagnostic tools, and studies using next generation sequencing in part aim to

define MND subtypes and associate them with various phenotypes (Teoh *et al.*, 2017; Karagiannis and Inoue, 2020). The following subsections briefly describe the effects of some MND-associated genes and proteins within the physiological context, and the effect of these mutations in the context of neural maintenance and MND. Chapter 5 (Section 5.1) goes into further detail of these effects.

1.5.1.1. Fused in sarcoma (Fus) and Transactivating response DNA-binding protein of 43 kDa (Tdp-43) are MND-associated proteins involved in RNA metabolism

Fus and Tdp-43 are RNA binding proteins that are associated with MND (Ling *et al.*, 2013; Renton *et al.*, 2014; Biscarini *et al.*, 2018). These proteins also function as regulators of RNA metabolism (Mirra *et al.*, 2017; Biscarini *et al.*, 2018; Harley *et al.*, 2021). RNA metabolism involves processes that contribute to transcription and translation, including post-transcriptional processes, RNA processing, pre-mRNA splicing, mRNA export, localization, turnover, and translational control. Common pathophysiologies of neurodegenerative diseases include abnormalities in RNA processing and protein homeostasis and quality control (Teoh *et al.*, 2017; Biscarini *et al.*, 2018). Mutations of *Fus* and *Tardbp*, which encode Fus and Tdp-43, respectively, are markers of MND (Ricketts *et al.*, 2014; Black *et al.*, 2017; Toth and Atkin, 2018; Harley *et al.*, 2021). Mutations of these genes cause mislocalisation of these RNA binding proteins, which have major impacts downstream on protein homeostasis and production that disrupt nervous system development and neural-cell maintenance, contributing to the onset of MND (Biscarini *et al.*, 2018; Harley *et al.*, 2021). Dysfunctional epigenetic enzymes driving histone and chromatin modifications disrupt proper FUS and TDP-43 protein expression and activity, contributing to neurodegeneration (Bennett *et al.*, 2019; Gimenez *et al.*, 2023).

1.5.1.2. Optineurin (Optn) and its association with trafficking and cell-cycle regulation

Optn is a protein ubiquitously expressed in mammalian organs (Li *et al.*, 1998; Rezaie and Sarfarazi, 2005; Rezaie *et al.*, 2005; Ying and Yue, 2012). It undergoes post-translational modifications that impact on its function; e.g., TANK-binding kinase 1 (Tbk-1) is one kinase

capable of phosphorylating Optn (Toth and Atkin, 2018). Optn-mediated onset of MND is potentially due to mutated Optn protein-protein interactions or protein binding to non-physiological partners, triggering a toxic gain-of-function mechanism (Sirohi *et al.*, 2013; Toth and Atkin, 2018). Neurons are susceptible to dysfunctional cellular trafficking due to their long and complex cell shape and so suboptimal vesicular trafficking due to mutated Optn will have detrimental effects on the functions of endoplasmic reticulum and the Golgi apparatus, compromising protein homeostasis and cell-cycle maintenance (Akizuki *et al.*, 2013; Toth and Atkin, 2018). However, links between epigenetic phenomena, *Optn* mutations and neurodegeneration are not, at present, well described, potentially due to the ubiquity of Optn expression (Li *et al.*, 1998; Rezaie and Sarfarazi, 2005; Rezaie *et al.*, 2005; Ying and Yue, 2012).

1.5.1.3. Survival motor neuron 1 (Smn1) and its coordinated role in RNA metabolism

Smn1 is associated with spinal muscular dystrophy, the most commonly inherited form of MND (Teoh *et al.*, 2017). The *Smn1* protein regulates programmed cell death (Schränk *et al.*, 1997; Parker *et al.*, 2008) and *Smn1*-deficient mouse embryos develop normally to the morula stage but then degenerate and die prior to implantation (Schränk *et al.*, 1997; Sephton *et al.*, 2010). *Smn1* is required for small nuclear ribonucleoprotein assembly for RNA splicing (Fischer *et al.*, 2011; Huo *et al.*, 2014). *Smn1* also coordinates with the *Fus* protein in pre-mRNA splicing (Mirra *et al.*, 2017; Biscarini *et al.*, 2018). Mutations of the *Smn1* gene leads to a loss-of-function phenotype (Schränk *et al.*, 1997; Sephton *et al.*, 2010; Sternecker *et al.*, 2014; Teoh *et al.*, 2017). Thus, mutated *Smn1* leads to depleted levels of functional *Smn1* protein, disrupting *Smn1*-regulated processes including assembly of small nuclear ribonucleoproteins and coordinated pre-mRNA splicing with the *Fus* protein. Mutated *Smn1* and *Fus* proteins have been reported in spinal muscular atrophy and other cases of MND (Mirra *et al.*, 2017). Dysfunctional chromatin remodelling in *Smn* genes leads to premature transcription termination

and genomic instability, contributing to neurodegeneration (Yanling Zhao *et al.*, 2016; Bennett *et al.*, 2019).

1.5.1.4. Superoxide dismutase (Sod1) and its role as an antioxidant

Oxidative stress is a major contributor to neurodegeneration (Halliwell, 2006; Spencer *et al.*, 2011; Ferreira *et al.*, 2016). It impairs neurotransmitter release from the pre-synaptic motor nerve terminal of the neuromuscular junction, which is essential for communication between the brain and muscles for movement (Spencer *et al.*, 2011; Ferreira *et al.*, 2016). In the context of MND, some mutations of *Sod1* contribute to Sod1 protein aggregation, leading to increased apoptosis and compromised production of neurofilaments resulting in altered neuronal morphology (Ferreira *et al.*, 2016; Black *et al.*, 2017; Karagiannis and Inoue, 2020). Defective Sod1 protein also contributes to an increase in reactive oxygen species production in mitochondria and abnormal mitochondrial morphology (Black *et al.*, 2017). Patients suffering from amyotrophic lateral sclerosis, a subset of MND, display decreased mitochondrial function and impairment in Ca^{2+} homeostasis, contributing to oxidative damage in the spinal cord (Sterneckert *et al.*, 2014). *SOD1* has a non-methylated promoter region in amyotrophic lateral sclerosis cases, indicating epigenetic-mediated contributions to disease onset (Oates and Pamphlett, 2007; Figueroa-Romero *et al.*, 2012; Jimenez-Pacheco *et al.*, 2017). Thus, selected *Sod1* mutations probably contribute to the onset of MNDs through imbalances in antioxidant activity and increased reactive oxygen species levels.

1.5.1.5. Vesicle-associated membrane protein-associated protein B and C (Vapb) and its coordinated role in Sod1-linked MND

Vapb is a protein responsible for regulating endoplasmic reticulum stress, the unfolded protein response, and Ca^{2+} homeostasis (Qiu *et al.*, 2013). The protein is abundant in motor neurons and is involved in the pathogenesis of Sod1-linked MND in humans (Teuling *et al.*, 2007). Reduced Vapb levels lead to the reduction of endoplasmic reticulum anchoring of lipid-binding proteins, resulting in motor neuron degeneration (Teuling *et al.*, 2007). Mutations of *Vapb* lead

to increase in Vapb aggregates in the spinal cord (Teuling *et al.*, 2007). However, this is not sufficient for the onset of motor neuron degeneration, as overexpression of mutant *Vapb* alone does not cause Sod1-linked MND (Qiu *et al.*, 2013). Abnormal Vapb clusters are present in peripheral blood mononuclear cells in sporadic ALS relative to healthy controls, which was hypothesised to be a consequence of epigenetic modifications resulting in incorrect protein localisation (Masala *et al.*, 2018; Cadoni *et al.*, 2020).

1.5.2. Developmental Origin of Health and Disease (DOHaD): A potential trigger of sporadic MND later in life

The DOHaD Hypothesis states that adverse impacts during *in-utero* development can result in changes in physiology and metabolism, resulting in increased risk of certain diseases later in life (De Boo and Harding, 2006; Heindel and Vandenberg, 2015; Ferreira *et al.*, 2016; Goyal *et al.*, 2019). This link is well established, for example, between reduced birthweight and increased risk of heart disease, stroke, diabetes and hypertension in adulthood, and included contributions from adverse environmental factors during postnatal growth (De Boo and Harding, 2006; Goyal *et al.*, 2019). Prenatal development is sensitive to environmental stressors, in part due to the sensitivity of epigenetic marks during differentiation (Heindel and Vandenberg, 2015). The resulting alterations to the epigenome can predispose newly formed cells and tissues to disease and dysfunction (Heindel and Vandenberg, 2015; Singh *et al.*, 2018; Almeida *et al.*, 2019; Goyal *et al.*, 2019). During development, once pluripotent cells commit towards multipotency, the tissue epigenome is relatively stable and less sensitive to environmental influences (Heindel and Vandenberg, 2015) so it's possible that epigenomic triggers for the onset of MND later in life occur very early in development – possibly as early as the preimplantation stages or post-implantation prior to and during embryonic development of the nervous system.

1.6. Concluding remarks

There is increasing appreciation that metabolic processes influence other molecular mechanisms, including signalling and transcriptional changes, to drive development, consistent with an emerging paradigm that changes in metabolism are a driver of development and differentiation and not just a passenger to it. L-Proline-mediated mechanisms driving early embryonic development, including neural cell fate, is now an archetypal example (Rathjen *et al.*, 2002; Washington *et al.*, 2010; D’Aniello *et al.*, 2015; Glover, 2018; Shparberg, 2018) though the mechanisms remain poorly understood. Understanding these mechanisms should result in improved protocols that increase efficiency of neural-cell production for stem-cell therapies, and neurodevelopmental disease models. These mechanisms can also be studied to determine which, if any, trigger onset of diseases later in life, such as MND.

1.7. Overall aims and hypotheses

The overall aim of this thesis is to determine the L-Proline-mediated mechanisms that contribute to driving pluripotent mouse ESCs to neural fate, and determine if these contributions may play a factor in late onset MND.

L-Proline-mediated metabolic processes will be investigated to determine drivers of the ES-to-EPLC transition. Inhibitors of various L-Proline-mediated metabolic pathways will be used to assess the contribution of these metabolic processes towards the ES-to-EPLC transition by analysing the emergent properties (i.e. colony morphology, proliferation rates, gene expression profiles and overall metabolic statuses). These mechanisms will further our understanding of the metabolic processes that produce pluripotent primitive ectoderm, the precursor to gastrulation.

Mechanisms driving fate across the pluripotency spectrum will be investigated by interrogating the transcriptomic profile of various stem-cell states: ground, naïve, EPLCs and primed EpiSCs. Focus will include candidate genes and protein-protein interactions predicted to directly drive cell fate and identity. This information will provide hypothesized drivers of pluripotency maintenance.

Mechanisms driving neural fate beyond the EPLC stage will also be investigated through interrogating the transcriptomic profile across cell populations representing definitive ectoderm and neurectoderm. Again, focus will include candidate genes and protein-protein interactions predicted to directly drive neural fate and identity. The possible influence of metabolic processes and PcG and TrxG proteins on the pluripotency spectrum and neural fate will also be interrogated. This information will provide insight into some of the metabolic- and epigenetic-mediated mechanisms in maintenance and loss of pluripotency and maintaining neural fate.

The L-Proline-mediated ESC-to-neural fate transition will be studied as a potential predictor of MND onset, which utilises the concept of the DOHaD Hypothesis. Expression

profiles of MND-associated genes and proteins will be analysed across various stages of the ESC-to-neural fate transition. This will help us understand the importance of cell-fate decisions as early as the setup of the nervous system in the context of disease onset.

Thus, the general hypotheses for this thesis are:

- Inhibiting L-Proline-mediated metabolic processes will prevent the ES-to-EPLC transition, as shown via changes in colony morphology, proliferation rates and gene expression profiles.
- As ESCs differentiate to EPLCs, the metabolic status will shift in response to L-Proline.
- Candidate genes that maintain pluripotency and drive neural fate will include genes encoding for transcriptional activity, metabolic enzymes, and PcG and TrxG proteins.
- MND-associated genes and proteins contribute to cell-fate decisions across the pluripotency spectrum and drive neural fate.

1.8. References

Agathocleous, M., Love, N. K., Randlett, O., Harris, J. J., Liu, J., Murray, A. J., and Harris, W. A. (2012) 'Metabolic differentiation in the embryonic retina', *Nature Cell Biology*, 14(8), pp. 859–864. doi: 10.1038/ncb2531.

Agostini, M., Romeo, F., Inoue, S., Niklison-Chirou, M. V, Elia, A. J., Dinsdale, D., Morone, N., Knight, R. A., Mak, T. W., and Melino, G. (2016) 'Metabolic reprogramming during neuronal differentiation', *Cell Death & Differentiation*, 23(9), pp. 1502–1514. doi: 10.1038/cdd.2016.36.

Akizuki, M., Yamashita, H., Uemura, K., Maruyama, H., Kawakami, H., Ito, H., and Takahashi, R. (2013) 'Optineurin suppression causes neuronal cell death via NF- κ B pathway', *Journal of Neurochemistry*. John Wiley & Sons, Ltd, 126(6), pp. 699–704. doi: <https://doi.org/10.1111/jnc.12326>.

Almeida, D. L., Pavanello, A., Saavedra, L. P., Pereira, T. S., de Castro-Prado, M. A. A., and de Freitas Mathias, P. C. (2019) 'Environmental monitoring and the developmental origins of health and disease', *Journal of Developmental Origins of Health and Disease*. 2019/05/27. Cambridge University Press, 10(6), pp. 608–615. doi: DOI: 10.1017/S2040174419000151.

Anderson, C. and Stern, C. D. (2016) 'Organizers in development', in *Current Topics in Developmental Biology*. Academic Press, pp. 435–454. doi: 10.1016/bs.ctdb.2015.11.023.

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data [Online]'. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Ang, S.-L. and Rossant, J. (1994) 'HNF-3 β is essential for node and notochord formation in mouse development', *Cell*. Cell Press, 78(4), pp. 561–574. doi: 10.1016/0092-8674(94)90522-3.

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y.,

Darr, H., Chang, B., Wang, J., Rendl, M., Bernstein, E., Schaniel, C., and Lemischka, I. R. (2011) ‘Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network’, *Cell*. Elsevier, 145(2), pp. 183–197. doi: 10.1016/j.cell.2011.03.003.

Aranda, S., Mas, G., and Di Croce, L. (2015) ‘Regulation of gene transcription by polycomb proteins’, *Science Advances*, 1(11), p. e1500737. doi: 10.1126/sciadv.1500737.

Argelaguet, R., Clark, S. J., Mohammed, H., Stapel, L. C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C. W., Smallwood, S., Ibarra-Soria, X., Buettner, F., Sanguinetti, G., Xie, W., Krueger, F., Göttgens, B., Rugg-Gunn, P. J., Kelsey, G., Dean, W., Nichols, J., Stegle, O., Marioni, J. C., and Reik, W. (2019) ‘Multi-omics profiling of mouse gastrulation at single-cell resolution’, *Nature*, 576(7787), pp. 487–491. doi: 10.1038/s41586-019-1825-8.

Arnold, S. J. and Robertson, E. J. (2009) ‘Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(2), pp. 91–103. doi: 10.1038/nrm2618.

Arrázola, M. S., Andraini, T., Szelechowski, M., Mouledous, L., Arnauné-Pelloquin, L., Davezac, N., Belenguer, P., Rampon, C., and Miquel, M.-C. (2019) ‘Mitochondria in developmental and adult neurogenesis’, *Neurotoxicity Research*, 36(2), pp. 257–267. doi: 10.1007/s12640-018-9942-y.

Artus, J., Piliszek, A., and Hadjantonakis, A.-K. (2011) ‘The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17’, *Developmental Biology*, 350(2), pp. 393–404. doi: 10.1016/j.ydbio.2010.12.007.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMahon, A. P., Harland, R. M., Rossant, J., and De Robertis, E. M. (2000)

‘The organizer factors Chordin and Noggin are required for mouse forebrain development’, *Nature*, 403(6770), pp. 658–661. doi: 10.1038/35001072.

Bazi, Z., Bertacchi, M., Abasi, M., Mohammadi-Yeganeh, S., Soleimani, M., Wagner, N., and Ghanbarian, H. (2018) ‘Rn7SK small nuclear RNA is involved in neuronal differentiation’, *Journal of Cellular Biochemistry*, 119(4), pp. 3174–3182. doi: 10.1002/jcb.26472.

Bennett, S. A., Tanaz, R., Cobos, S. N., and Torrente, M. P. (2019) ‘Epigenetics in amyotrophic lateral sclerosis: A role for histone post-translational modifications in neurodegenerative disease’, *Translational Research*, 204, pp. 19–30. doi: 10.1016/j.trsl.2018.10.002.

Bernemann, C., Greber, B., Ko, K., Sternecker, J., Han, D. W., Araúzo-Bravo, M. J., and Schöler, H. R. (2011) ‘Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features’, *Stem Cells*. England, 29(10), pp. 1496–1503. doi: 10.1002/stem.709.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) ‘A bivalent chromatin structure marks key developmental genes in embryonic stem cells’, *Cell*, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Biscarini, S., Caputo, D., Peruzzi, G., Lu, L., Colantoni, A., Santini, T., Shneider, N. A., Caffarelli, E., Laneve, P., and Bozzoni, I. (2018) ‘Characterization of the lncRNA transcriptome in mESC-derived motor neurons: Implications for FUS-ALS’, *Stem Cell Research*, 27, pp. 172–179. doi: <https://doi.org/10.1016/j.scr.2018.01.037>.

Black, H. A., Leighton, D. J., Cleary, E. M., Rose, E., Stephenson, L., Colville, S., Ross, D., Warner, J., Porteous, M., Gorrie, G. H., Swingler, R., Goldstein, D., Harms, M. B., Connick,

P., Pal, S., Aitman, T. J., and Chandran, S. (2017) 'Genetic epidemiology of motor neuron disease-associated variants in the Scottish population', *Neurobiology of Aging*. 2016/12/21. Elsevier, 51, pp. 178.e11-178.e20. doi: 10.1016/j.neurobiolaging.2016.12.013.

De Boo, H. A. and Harding, J. E. (2006) 'The developmental origins of adult disease (Barker) hypothesis', *Australian and New Zealand Journal of Obstetrics and Gynaecology*. John Wiley & Sons, Ltd, 46(1), pp. 4–14. doi: <https://doi.org/10.1111/j.1479-828X.2006.00506.x>.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 122(6), pp. 947–956. doi: 10.1016/j.cell.2005.08.020.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*, 441(7091), pp. 349–353. doi: 10.1038/nature04733.

Brand, M., Nakka, K., Zhu, J., and Dilworth, F. J. (2019) 'Polycomb/Trithorax antagonism: Cellular memory in stem cell fate and function', *Cell Stem Cell*, 24(4), pp. 518–533. doi: <https://doi.org/10.1016/j.stem.2019.03.005>.

Bratt-Leal, A. M., Carpenedo, R. L., and McDevitt, T. C. (2009) 'Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation', *Biotechnology Progress*, 25(1), pp. 43–51. doi: 10.1021/bp.139.

Brazil, D. P., Church, R. H., Surae, S., Godson, C., and Martin, F. (2015) 'BMP signalling: Agony and antagonism in the family', *Trends in Cell Biology*, 25(5), pp. 249–264. doi: 10.1016/J.TCB.2014.12.004.

Briscoe, J. and Ericson, J. (1999) 'The specification of neuronal identity by graded sonic hedgehog signalling', *Seminars in Cell & Developmental Biology*, 10(3), pp. 353–362. doi: 10.1006/scdb.1999.0295.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., and Vallier, L. (2007) 'Derivation of pluripotent epiblast stem cells from mammalian embryos', *Nature*. England, 448(7150), pp. 191–195. doi: 10.1038/nature05950.

Bryja, V., Bonilla, S., and Arenas, E. (2006) 'Derivation of mouse embryonic stem cells', *Nature Protocols*, 1(4), pp. 2082–2087. doi: 10.1038/nprot.2006.355.

Burdon, T., Smith, A., and Savatier, P. (2002) 'Signalling, cell cycle and pluripotency in embryonic stem cells', *Trends in Cell Biology*, 12(9), pp. 432–438. doi: 10.1016/S0962-8924(02)02352-8.

Cadoni, M. P. L., Biggio, M. L., Arru, G., Secchi, G., Orrù, N., Clemente, M. G., Sechi, G., Yamoah, A., Tripathi, P., Orrù, S., Manetti, R., and Galleri, G. (2020) 'VAPB ER-aggregates, a possible new biomarker in ALS pathology', *Cells*, 9(1), p. 164. doi: 10.3390/cells9010164.

Camus, A., Perea-Gomez, A., Moreau, A., and Collignon, J. (2006) 'Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo', *Developmental Biology*, 295(2), pp. 743–755. doi: <https://doi.org/10.1016/j.ydbio.2006.03.047>.

Cartwright, P., McLean, C., Sheppard, A., Rivett, D., Jones, K., and Dalton, S. (2005) 'LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism', *Development*, 132(5), pp. 885–896. doi: 10.1242/dev.01670.

Casalino, L., Comes, S., Lambazzi, G., De Stefano, B., Filosa, S., De Falco, S., De Cesare, D., Minchiotti, G., and Patriarca, E. J. (2011) 'Control of embryonic stem cell metastability by

L-Proline catabolism', *Journal of Molecular Cell Biology*, 3(2), pp. 108–122. doi: 10.1093/jmcb/mjr001.

Casanova, E. A., Bürki, K., and Cinelli, P. (2011) 'Molecular mechanisms of pluripotency in murine embryonic stem cells', in Atwood, C. (ed.) *Embryonic Stem Cells: The Hormonal Regulation of Pluripotency and Embryogenesis*. InTech, pp. 27–60. doi: 10.5772/14219.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007) 'Nanog safeguards pluripotency and mediates germline development', *Nature*, 450(7173), pp. 1230–1234. doi: 10.1038/nature06403.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006) 'Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway', *Developmental Cell*. Cell Press, 10(5), pp. 615–624. doi: 10.1016/j.devcel.2006.02.020.

Chen, C.-T., Hsu, S.-H., and Wei, Y.-H. (2010) 'Upregulation of mitochondrial function and antioxidant defense in the differentiation of stem cells', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1800(3), pp. 257–263. doi: 10.1016/j.bbagen.2009.09.001.

Chen, C. W., Liu, C. S., Chiu, I. M., Shen, S. C., Pan, H. C., Lee, K. H., Lin, S. Z., and Su, H. L. (2010) 'The signals of FGFs on the neurogenesis of embryonic stem cells', *Journal of Biomedical Science*, 17(1), p. 33. doi: 10.1186/1423-0127-17-33.

Chen, G. and Wang, J. (2013) 'Threonine metabolism and embryonic stem cell self-renewal', *Current Opinion in Clinical Nutrition and Metabolic Care*, 17(1), pp. 80–85. doi: 10.1097/MCO.0000000000000007.

Chen, J., Bardes, E. E., Aronow, B. J., and Jegga, A. G. (2009) 'ToppGene Suite for gene list enrichment analysis and candidate gene prioritization', *Nucleic Acids Research*, 37(suppl_2), pp. W305–W311. doi: 10.1093/nar/gkp427.

Cheng, A., Wan, R., Yang, J.-L., Kamimura, N., Son, T. G., Ouyang, X., Luo, Y., Okun, E., and Mattson, M. P. (2012) 'Involvement of PGC-1 α in the formation and maintenance of neuronal dendritic spines', *Nature Communications*, 3(1), p. 1250. doi: 10.1038/ncomms2238.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H., and Beachy, P. A. (1996) 'Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function', *Nature*, 383(6599), pp. 407–413. doi: 10.1038/383407a0.

Cho, Y. M., Kwon, S., Pak, Y. K., Seol, H. W., Choi, Y. M., Park do, J., Park, K. S., and Lee, H. K. (2006) 'Dynamic changes in mitochondrial biogenesis and antioxidant enzymes during the spontaneous differentiation of human embryonic stem cells', *Biochemical and Biophysical Research Communications*, 348(4), pp. 1472–1478. doi: 10.1016/j.bbrc.2006.08.020.

Christophersen, N. S. and Helin, K. (2010) 'Epigenetic control of embryonic stem cell fate', *Journal of Experimental Medicine*, 207(11), pp. 2287–2295. doi: 10.1084/jem.20101438.

Collinson, A., Collier, A. J., Morgan, N. P., Sienerth, A. R., Chandra, T., Andrews, S., and Rugg-Gunn, P. J. (2016) 'Deletion of the Polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells', *Cell Reports*, 17(10), pp. 2700–2714. doi: 10.1016/j.celrep.2016.11.032.

Comes, S., Gagliardi, M., Laprano, N., Fico, A., Cimmino, A., Palamidessi, A., De Cesare, D., De Falco, S., Angelini, C., Scita, G., Patriarca, E. J., Matarazzo, M. R., and Minchiotti, G. (2013) 'L-Proline induces a mesenchymal-like invasive program in embryonic stem cells by remodeling H3K9 and H3K36 methylation', *Stem Cell Reports*, 1(4), pp. 307–321. doi: 10.1016/j.stemcr.2013.09.001.

Di Croce, L. and Helin, K. (2013) 'Transcriptional regulation by polycomb group proteins', *Nature Structural & Molecular Biology*, 20(10), pp. 1147–1155. doi:

10.1038/nsmb.2669.

D’Aniello, C., Fico, A., Casalino, L., Guardiola, O., Di Napoli, G., Cermola, F., De Cesare, D., Tatè, R., Cobellis, G., Patriarca, E. J., and Minchiotti, G. (2015) ‘A novel autoregulatory loop between the Gcn2-Atf4 pathway and L-Proline metabolism controls stem cell identity’, *Cell Death And Differentiation*, 22, pp. 1094–1105. doi: 10.1038/cdd.2015.24.

Dalton, S. (2015) ‘Linking the cell cycle to cell fate decisions’, *Trends in Cell Biology*, 25(10), pp. 592–600. doi: 10.1016/j.tcb.2015.07.007.

Deepak, B. and Sudipto, R. (2013) ‘Left–right asymmetry: cilia stir up new surprises in the node’, *Open Biology*. Royal Society, 3(5), p. 130052. doi: 10.1098/rsob.130052.

Di-Gregorio, A., Sancho, M., Stuckey, D. W., Crompton, L. A., Godwin, J., Mishina, Y., and Rodriguez, T. A. (2007) ‘BMP signalling inhibits premature neural differentiation in the mouse embryo’, *Development*, 134(18), pp. 3359–3369. doi: 10.1242/dev.005967.

Ding, Q., Motoyama, J., Gasca, S., Mo, R., Sasaki, H., Rossant, J., and Hui, C. C. (1998) ‘Diminished Sonic hedgehog signaling and lack of floor plate differentiation in Gli2 mutant mice’, *Development*, 125(14), pp. 2533–2543. Available at: <http://dev.biologists.org/content/125/14/2533>.

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., Askarian-Amiri, M. E., Ru, K., Solda, G., Simons, C., Sunkin, S. M., Crowe, M. L., Grimmond, S. M., Perkins, A. C., and Mattick, J. S. (2008) ‘Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation’, *Genome Research*, 18(9), pp. 1433–1445. doi: 10.1101/gr.078378.108.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) ‘STAR: Ultrafast universal RNA-seq aligner’, *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985) 'The in vitro development of blastocyst derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium', *Journal of Embryology and Experimental Morphology*, 87, pp. 27–45.

Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) 'Regulation of MLL1 H3K4 methyltransferase activity by its core components', *Nature Structural & Molecular Biology*, 13(8), pp. 713–719. doi: 10.1038/nsmb1128.

Le Dréau, G., Escalona, R., Fueyo, R., Herrera, A., Martínez, J. D., Usieto, S., Menendez, A., Pons, S., Martinez-Balbas, M. A., and Marti, E. (2018) 'E proteins sharpen neurogenesis by modulating proneural bHLH transcription factors' activity in an E-box-dependent manner', *eLife*. England, 7. doi: 10.7554/eLife.37267.

Dvash, T., Mayshar, Y., Darr, H., McElhaney, M., Barker, D., Yanuka, O., Kotkow, K. J., Rubin, L. L., Benvenisty, N., and Eiges, R. (2004) 'Temporal gene expression during differentiation of human embryonic stem cells and embryoid bodies', *Human Reproduction*, 19(12), pp. 2875–2883. doi: 10.1093/humrep/deh529.

Ernst, P. and Vakoc, C. R. (2012) 'WRAD: Enabler of the SET1-family of H3K4 methyltransferases', *Briefings in Functional Genomics*, 11(3), pp. 217–226. doi: 10.1093/bfpg/els017.

Evans, M. J. and Kaufman, M. H. (1981) 'Establishment in culture of pluripotential cells from mouse embryos', *Nature*, 292(5819), pp. 154–156. doi: 10.1038/292154a0.

Falcon, S. and Gentleman, R. (2008) 'Hypergeometric testing used for gene set enrichment analysis', in Gentleman, Robert, Hornik, K., and Parmigiani, G. (eds) *Bioconductor Case Studies. Use R!* New York, NY: Springer, New York, NY, pp. 207–220. doi: 10.1007/978-

Fathi, A., Pakzad, M., Taei, A., Brink, T. C., Pirhaji, L., Ruiz, G., Sharif Tabe Bordbar, M., Gourabi, H., Adjaye, J., Baharvand, H., and Hosseini Salekdeh, G. (2009) 'Comparative proteome and transcriptome analyses of embryonic stem cells during embryoid body-based differentiation', *Proteomics*, 9(21), pp. 4859–4870. doi: 10.1002/pmic.200900003.

Ferreira, D. J. S., Sellitti, D. F., and Lagranha, C. J. (2016) 'Protein undernutrition during development and oxidative impairment in the central nervous system (CNS): Potential factors in the occurrence of metabolic syndrome and CNS disease', *Journal of Developmental Origins of Health and Disease*. Cambridge University Press, 7(5), pp. 513–524. doi: 10.1017/S2040174416000246.

Fico, A., Fiorenzano, A., Pascale, E., Patriarca, E. J., and Minchiotti, G. (2019) 'Long non-coding RNA in stem cell pluripotency and lineage commitment: functions and evolutionary conservation', *Cellular and Molecular Life Sciences*, 76(8), pp. 1459–1471. doi: 10.1007/s00018-018-3000-z.

Figuerola-Romero, C., Hur, J., Bender, D. E., Delaney, C. E., Cataldo, M. D., Smith, A. L., Yung, R., Ruden, D. M., Callaghan, B. C., and Feldman, E. L. (2012) 'Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis', *PLoS ONE*, 7(12), p. e52672. doi: 10.1371/journal.pone.0052672.

Fiorenza, M. T. and Rava, A. (2019) 'The TCL1 function revisited focusing on metabolic requirements of stemness', *Cell Cycle*, 18(22), pp. 3055–3063. doi: 10.1080/15384101.2019.1672465.

Fiorenzano, A., Pascale, E., D'Aniello, C., Acampora, D., Bassalart, C., Russo, F., Andolfi, G., Biffoni, M., Francescangeli, F., Zeuner, A., Angelini, C., Chazaud, C., Patriarca, E. J., Fico, A., and Minchiotti, G. (2016) 'Cripto is essential to capture mouse epiblast stem cell

and human embryonic stem cell pluripotency', *Nature Communications*. England, 7(1), p. 12589. doi: 10.1038/ncomms12589.

Fischer, U., Englbrecht, C., and Chari, A. (2011) 'Biogenesis of spliceosomal small nuclear ribonucleoproteins', *WIREs RNA*. John Wiley & Sons, Ltd, 2(5), pp. 718–731. doi: <https://doi.org/10.1002/wrna.87>.

Forristal, C. E., Wright, K. L., Hanley, N. A., Oreffo, R. O. C., and Houghton, F. D. (2010) 'Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions', *Reproduction*, 139(1), pp. 85–97. doi: 10.1530/REP-09-0300.

Funa, N. S., Schachter, K. A., Lerdrup, M., Ekberg, J., Hess, K., Dietrich, N., Honoré, C., Hansen, K., and Semb, H. (2015) 'β-Catenin regulates primitive streak induction through collaborative interactions with SMAD2/SMAD3 and OCT4', *Cell Stem Cell*, 16(6), pp. 639–652. doi: 10.1016/j.stem.2015.03.008.

Galonska, C., Ziller, M. J., Karnik, R., and Meissner, A. (2015) 'Ground state conditions induce rapid reorganization of core pluripotency factor binding before global epigenetic reprogramming', *Cell Stem Cell*. United States, 17(4), pp. 462–470. doi: 10.1016/j.stem.2015.07.005.

Gangloff, Y., Mueller, M., Dann, S. G., Svoboda, P., Sticker, M., Spetz, J., Um, H., Brown, E. J., Cereghini, S., Thomas, G., Kozma, S. C., and Um, S. H. (2004) 'Disruption of the mouse mTOR gene leads to early postimplantation lethality and prohibits embryonic stem cell development', *Molecular and Cellular Biology*, 24(21), pp. 9508–9516. doi: 10.1128/MCB.24.21.9508.

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., and Wang, Z. (2008) 'ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component

BAF250a', *Proceedings of the National Academy of Sciences*, 105(18), pp. 6656–6661. doi: 10.1073/pnas.0801802105.

Gimenez, J., Spalloni, A., Cappelli, S., Ciaiola, F., Orlando, V., Buratti, E., and Longone, P. (2023) 'TDP-43 epigenetic facets and their neurodegenerative implications', *International Journal of Molecular Sciences*, 24(18), p. 13807. doi: 10.3390/ijms241813807.

Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.

Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) 'L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell', in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.

De Gobbi, M., Garrick, D., Lynch, M., Vernimmen, D., Hughes, J. R., Goardon, N., Luc, S., Lower, K. M., Sloane-Stanley, J. A., Pina, C., Soneji, S., Renella, R., Enver, T., Taylor, S., Jacobsen, S. E. W., Vyas, P., Gibbons, R. J., and Higgs, D. R. (2011) 'Generation of bivalent chromatin domains during cell fate decisions', *Epigenetics & Chromatin*, 4(1), p. 9. doi: 10.1186/1756-8935-4-9.

Goyal, D., Limesand, S. W., and Goyal, R. (2019) 'Epigenetic responses and the developmental origins of health and disease', *Journal of Endocrinology*. Bristol, UK: Bioscientifica Ltd, 242(1), pp. T105–T119. doi: 10.1530/JOE-19-0009.

Graf, U., Casanova, E. A., and Cinelli, P. (2011) 'The role of the leukemia inhibitory factor (LIF) - pathway in derivation and maintenance of murine pluripotent stem cells', *Genes*, 2(1), pp. 280–297. doi: 10.3390/genes2010280.

Greber, B., Wu, G., Bernemann, C., Joo, J. Y., Han, D. W., Ko, K., Tapia, N., Sabour, D., Sternecker, J., Tesar, P., and Schöler, H. R. (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. doi: <https://doi.org/10.1016/j.stem.2010.01.003>.

Greber, B., Lehrach, H., and Adjaye, J. (2007) 'Fibroblast growth factor 2 modulates transforming growth factor β signaling in mouse embryonic fibroblasts and human ESCs (hESCs) to support hESC self-renewal', *Stem Cells*. England, 25(2), pp. 455–464. doi: 10.1634/stemcells.2006-0476.

Greber, B., Lehrach, H., and Adjaye, J. (2008) 'Control of early fate decisions in human ES cells by distinct states of TGF β pathway activity', *Stem Cells and Development*. Mary Ann Liebert, Inc., publishers, 17(6), pp. 1065–1078. doi: 10.1089/scd.2008.0035.

Hackett, J. A., Dietmann, S., Murakami, K., Down, T. A., Leitch, H. G., and Surani, M. A. (2013) 'Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency', *Stem Cell Reports*, 1(6), pp. 518–531. doi: <https://doi.org/10.1016/j.stemcr.2013.11.010>.

Hackett, J. A. and Surani, M. A. (2014) 'Regulatory principles of pluripotency: From the ground state up', *Cell Stem Cell*, 15(4), pp. 416–430. doi: <https://doi.org/10.1016/j.stem.2014.09.015>.

Halliwell, B. (2006) 'Oxidative stress and neurodegeneration: where are we now?', *Journal of Neurochemistry*. John Wiley & Sons, Ltd, 97(6), pp. 1634–1658. doi: <https://doi.org/10.1111/j.1471-4159.2006.03907.x>.

Hargreaves, D. C. and Crabtree, G. R. (2011) 'ATP-dependent chromatin remodeling: genetics, genomics and mechanisms', *Cell Research*, 21(3), pp. 396–420. doi: 10.1038/cr.2011.32.

Harley, J., Hagemann, C., Serio, A., and Patani, R. (2021) 'TDP-43 and FUS mislocalization in VCP mutant motor neurons is reversed by pharmacological inhibition of the

VCP D2 ATPase domain', *Brain Communications*. Oxford University Press, 3(3), p. fcab166. doi: 10.1093/braincomms/fcab166.

Harvey, A., Caretti, G., Moresi, V., Renzini, A., and Adamo, S. (2019) 'Interplay between metabolites and the epigenome in regulating embryonic and adult stem cell potency and maintenance', *Stem Cell Reports*, 13(4), pp. 573–589. doi: 10.1016/j.stemcr.2019.09.003.

Harvey, A. J., Rathjen, J., and Gardner, D. K. (2016) 'Metaboloepigenetic regulation of pluripotent stem cells', *Stem Cells International*, 2016, pp. 1–15. doi: 10.1155/2016/1816525.

Harvey, N. T., Hughes, J. N., Lonic, A., Yap, C., Long, C., Rathjen, P. D., and Rathjen, J. (2010) 'Response to BMP4 signalling during ES cell differentiation defines intermediates of the ectoderm lineage', *Journal of Cell Science*, 123(10), pp. 1796–1804. doi: 10.1242/jcs.047530.

Hayashi, K., Lopes, S. M. C. de S., Tang, F., and Surani, M. A. (2008) 'Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states', *Cell Stem Cell*, 3(4), pp. 391–401. doi: <https://doi.org/10.1016/j.stem.2008.07.027>.

Heindel, J. J. and Vandenberg, L. N. (2015) 'Developmental origins of health and disease: A paradigm for understanding disease cause and prevention', *Current Opinion in Pediatrics*, 27(2), pp. 248–253. doi: 10.1097/MOP.0000000000000191.

Herz, H. M., Garruss, A., and Shilatfard, A. (2013) 'SET for life: Biochemical activities and biological functions of SET domain-containing proteins', *Trends in Biochemical Sciences*, 38(12), pp. 621–639. doi: 10.1016/j.tibs.2013.09.004.

Hirai, H., Karian, P., and Kikyo, N. (2011) 'Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor', *Biochemical Journal*, 438(1), pp. 11–23. doi: 10.1042/BJ20102152.

Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., and Crabtree, G. R. (2009) ‘An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network’, *Proceedings of the National Academy of Sciences*, 106(13), pp. 5187–5191. doi: 10.1073/pnas.0812888106.

Hu, W., Alvarez-Dominguez, J. R., and Lodish, H. F. (2012) ‘Regulation of mammalian cell differentiation by long non-coding RNAs’, *EMBO Reports*, 13(11), pp. 971–983. doi: 10.1038/embor.2012.145.

Huang, Y., Osorno, R., Tsakiridis, A., and Wilson, V. (2012) ‘In vivo differentiation potential of epiblast stem cells revealed by chimeric embryo formation’, *Cell Reports*, 2(6), pp. 1571–1578. doi: <https://doi.org/10.1016/j.celrep.2012.10.022>.

Huo, Q., Kayikci, M., Odermatt, P., Meyer, K., Michels, O., Saxena, S., Ule, J., and Schümperli, D. (2014) ‘Splicing changes in SMA mouse motoneurons and SMN-depleted neuroblastoma cells: Evidence for involvement of splicing regulatory proteins’, *RNA Biology*. Taylor & Francis, 11(11), pp. 1430–1446. doi: 10.1080/15476286.2014.996494.

Imayoshi, I. and Kageyama, R. (2014) ‘bHLH factors in self-renewal, multipotency, and fate choice of neural progenitor cells’, *Neuron*. Elsevier, 82(1), pp. 9–23. doi: 10.1016/j.neuron.2014.03.018.

Inman, G. J., Nicolás, F. J., Callahan, J. F., Harling, J. D., Gaster, L. M., Reith, A. D., Laping, N. J., and Hill, C. S. (2002) ‘SB-431542 is a potent and specific inhibitor of Transforming Growth Factor- β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7’, *Molecular Pharmacology*, 62(1), pp. 65–74. doi: 10.1124/mol.62.1.65.

Ishitobi, H., Wakamatsu, A., Liu, F., Azami, T., Hamada, M., Matsumoto, K., Kataoka, H., Kobayashi, M., Choi, K., Nishikawa, S., Takahashi, S., and Ema, M. (2011) ‘Molecular

basis for Flk1 expression in hemato-cardiovascular progenitors in the mouse’, *Development*, 138(24), pp. 5357 LP – 5368. doi: 10.1242/dev.065565.

Jaenisch, R. and Young, R. (2008) ‘Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming’, *Cell*, 132(4), pp. 567–582. doi: 10.1016/j.cell.2008.01.015.

Jeong, Y. and Epstein, D. J. (2003) ‘Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node’, *Development*, 130, pp. 3891–3902. doi: 10.1242/dev.00590.

Jiang, H., Shukla, A., Wang, X., Chen, W., Bernstein, B. E., and Roeder, R. G. (2011) ‘Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains’, *Cell*, 144(4), pp. 513–525. doi: 10.1016/j.cell.2011.01.020.

Jimenez-Pacheco, A., Franco, J. M., Lopez, S., Gomez-Zumaquero, J. M., Magdalena Leal-Lasarte, M., Caballero-Hernandez, D. E., Cejudo-Guillén, M., and Pozo, D. (2017) ‘Epigenetic mechanisms of gene regulation in amyotrophic lateral sclerosis’, in Delgado-Morales, R. (ed.) *Neuroepigenomics in Aging and Disease*. Springer International Publishing, pp. 255–275. doi: 10.1007/978-3-319-53889-1_14.

Jones, E. A. V., Crotty, D., Kulesa, P. M., Waters, C. W., Baron, M. H., Fraser, S. E., and Dickinson, M. E. (2002) ‘Dynamic in vivo imaging of postimplantation mammalian embryos using whole embryo culture’, *Genesis*, 34(4), pp. 228–235. doi: 10.1002/gene.10162.

Karagiannis, P. and Inoue, H. (2020) ‘ALS, a cellular whodunit on motor neuron degeneration’, *Molecular and Cellular Neuroscience*, 107, p. 103524. doi: <https://doi.org/10.1016/j.mcn.2020.103524>.

Ke, Q. and Costa, M. (2006) ‘Hypoxia-Inducible Factor-1 (HIF-1)’, *Molecular Pharmacology*, 70(5), pp. 1469–1480. doi: 10.1124/mol.106.027029.

Kilberg, M. S., Terada, N., and Shan, J. (2016) ‘Influence of amino acid metabolism on

embryonic stem cell function and differentiation’, *Advances in Nutrition: An International Review Journal*, 7(4), pp. 780S-789S. doi: 10.3945/an.115.011031.

Kim, H., Wu, J., Ye, S., Tai, C.-I., Zhou, X., Yan, H., Li, P., Pera, M., and Ying, Q.-L. (2013) ‘Modulation of β -catenin function maintains mouse epiblast stem cell and human embryonic stem cell self-renewal’, *Nature Communications*, 4(1), p. 2403. doi: 10.1038/ncomms3403.

Kim, H., Jang, H., Kim, T. W., Kang, B. H., Lee, S. E., Jeon, Y. K., Chung, D. H., Choi, J., Shin, J., Cho, E. J., and Youn, H. D. (2015) ‘Core pluripotency factors directly regulate metabolism in embryonic stem cell to maintain pluripotency’, *Stem Cells*, 33(9), pp. 2699–2711. doi: 10.1002/stem.2073.

Kim, J. W., Tchernyshyov, I., Semenza, G. L., and Dang, C. V. (2006) ‘HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia’, *Cell Metabolism*, 3(3), pp. 177–185. doi: 10.1016/j.cmet.2006.02.002.

Kingston, R. E. and Tamkun, J. W. (2014) ‘Transcriptional regulation by Trithorax-group proteins’, *Cold Spring Harbor Perspectives in Biology*, 6(10), p. a019349. doi: 10.1101/cshperspect.a019349.

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R., and Hemberg, M. (2017) ‘SC3: consensus clustering of single-cell RNA-seq data’, *Nature Methods*, 14(5), pp. 483–486. doi: 10.1038/nmeth.4236.

Knobloch, M. and Jessberger, S. (2017) ‘Metabolism and neurogenesis’, *Current Opinion in Neurobiology*, 42, pp. 45–52. doi: 10.1016/j.conb.2016.11.006.

Koch, C. M., Chiu, S. F., Akbargpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., and Winter, D. R. (2018) ‘A beginner’s guide to analysis of RNA sequencing data’, *American*

Journal of Respiratory Cell and Molecular Biology, 59(2), pp. 145–157. doi: 10.1165/rcmb.2017-0430TR.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., de Alencastro, G., Logan, G. J., Teber, E. T., Tam, O. H., Stutz, M. D., Alexander, I. E., Pickett, H. A., and Tam, P. P. L. (2014) ‘The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak’, *Cell Stem Cell*, 14(1), pp. 107–120. doi: <https://doi.org/10.1016/j.stem.2013.09.014>.

Krupinski, P., Chickarmane, V., and Peterson, C. (2011) ‘Simulating the mammalian blastocyst - molecular and mechanical interactions pattern the embryo’, *PLoS Computational Biology*, 7(5), pp. 1–11. doi: 10.1371/journal.pcbi.1001128.

Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S., Adli, M., Kasif, S., Ptaszek, L. M., Cowan, C. A., Lander, E. S., Koseki, H., and Bernstein, B. E. (2008) ‘Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains’, *PLoS Genetics*. 2008/10/31, 4(10), p. e1000242. doi: 10.1371/journal.pgen.1000242.

Kuegler, P. B., Zimmer, B., Waldmann, T., Baudis, B., Ilmjarv, S., Hescheler, J., Gaughwin, P., Brundin, P., Mundy, W., Bal-Price, A. K., Schrattenholz, A., Krause, K.-H., van Thriell, C., Rao, M. S., Kadereit, S., and Leist, M. (2010) ‘Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing’, *ALTEX: Alternatives to Animal Experimentation*, 27(1), pp. 16–42.

Kumar, P., Tan, Y., and Cahan, P. (2017) ‘Understanding development and stem cells using single cell-based analyses of gene expression’, *Development*, 144, pp. 17–32. doi: 10.1242/dev.133058.

Kunath, T., Saba-El-Leil, M. K., Almousailleakh, M., Wray, J., Meloche, S., and Smith,

A. (2007) 'FGF stimulation of the Erk1/2 signalling cascade triggers transition of pluripotent embryonic stem cells from self-renewal to lineage commitment.', *Development*. England, 134(16), pp. 2895–2902. doi: 10.1242/dev.02880.

Lau, P. N. I. and So, C. W. E. (2015) 'Polycomb and Trithorax factors in transcriptional and epigenetic regulation', in Huang, S., Litt, M. D., and Blakey, C. A. (eds) *Epigenetic Gene Expression and Regulation*. Elsevier Science & Technology, pp. 63–94. doi: 10.1016/B978-0-12-799958-6.00004-4.

Lee, K. J., Dietrich, P., and Jessell, T. M. (2000) 'Genetic ablation reveals that the roof plate is essential for dorsal interneuron specification', *Nature*, 403(6771), pp. 734–740. doi: 10.1038/35001507.

Li, L., Liu, C., Biechele, S., Zhu, Q., Song, L., Lanner, F., Jing, N., and Rossant, J. (2013) 'Location of transient ectodermal progenitor potential in mouse development', *Development*, 140(22), pp. 4533–4543. doi: 10.1242/dev.092866.

Li, M. and Belmonte, J. C. I. (2017) 'Ground rules of the pluripotency gene regulatory network', *Nature Reviews Genetics*, 18(3), pp. 180–191. doi: 10.1038/nrg.2016.156.

Li, Y., Kang, J., and Horwitz, M. S. (1998) 'Interaction of an adenovirus E3 14.7-kilodalton protein with a novel tumor necrosis factor alpha-inducible cellular protein containing leucine zipper domains', *Molecular and cellular biology*. United States, 18(3), pp. 1601–1610. doi: 10.1128/MCB.18.3.1601.

Li, Z., Fei, T., Zhang, J., Zhu, G., Wang, L., Lu, D., Chi, X., Teng, Y., Hou, N., Yang, X., Zhang, H., Han, J.-D. J., and Chen, Y.-G. (2012) 'BMP4 signaling acts via Dual-Specificity Phosphatase 9 to control ERK activity in mouse embryonic stem cells', *Cell Stem Cell*, 10(2), pp. 171–182. doi: 10.1016/j.stem.2011.12.016.

Liem, K. F., Tremml, G., and Jessell, T. M. (1997) 'A role for the roof plate and its

resident TGF β -related proteins in neuronal patterning in the dorsal spinal cord', *Cell*, 91(1), pp. 127–138. doi: 10.1016/S0092-8674(01)80015-5.

Ling, S.-C., Polymenidou, M., and Cleveland, D. W. (2013) 'Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis', *Neuron*, 79(3), pp. 416–438. doi: <https://doi.org/10.1016/j.neuron.2013.07.033>.

Liu, W., Hancock, C. N., Fischer, J. W., Harman, M., and Phang, J. M. (2015) 'Proline biosynthesis augments tumor cell growth and aerobic glycolysis: involvement of pyridine nucleotides', *Scientific Reports*, 5(1), p. 17206. doi: 10.1038/srep17206.

Liu, W., Deng, C., Godoy-Parejo, C., Zhang, Y., and Chen, G. (2019) 'Developments in cell culture systems for human pluripotent stem cells', *World Journal of Stem Cells*. United States, 11(11), pp. 968–981. doi: 10.4252/wjsc.v11.i11.968.

Loebel, D. A. ., Watson, C. M., De Young, R. A., and Tam, P. P. L. (2003) 'Lineage choice and differentiation in mouse embryos and embryonic stem cells', *Developmental Biology*, 264(1), pp. 1–14. doi: 10.1016/S0012-1606(03)00390-7.

Lonic, A. (2007) *Molecular mechanism of L-proline induced EPL-cell formation*. University of Adelaide.

Lu, C. C., Brennan, J., and Robertson, E. J. (2001) 'From fertilization to gastrulation: axis formation in the mouse embryo', *Current Opinion in Genetics & Development*, 11(4), pp. 384–392. doi: [https://doi.org/10.1016/S0959-437X\(00\)00208-2](https://doi.org/10.1016/S0959-437X(00)00208-2).

Lu, R., Markowitz, F., Unwin, R. D., Leek, J. T., Airoidi, E. M., MacArthur, B. D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L. A., Troyanskaya, O. G., Whetton, A. D., and Lemischka, I. R. (2009) 'Systems-level dynamic analyses of fate change in murine embryonic stem cells', *Nature*, 462(7271), pp. 358–362. doi: 10.1038/nature08575.

Luo, X., van der Veer, B. K., Sun, L., Bartocetti, M., Boretto, M., Vankelecom, H.,

Khoueiry, R., and Koh, K. P. (2020) ‘Coordination of germ layer lineage choice by TET1 during primed pluripotency’, *Genes and Development*, 34(7–8), pp. 598–618. doi: 10.1101/gad.329474.119.

MacArthur, B. D., Ma’ayan, A., and Lemischka, I. R. (2009) ‘Systems biology of stem cell fate and cellular reprogramming’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(10), pp. 672–681. doi: 10.1038/nrm2766.

Mandal, S., Lindgren, A., and Srivastava, A. (2010) ‘Mitochondrial function controls proliferation and early differentiation potential of embryonic stem cells’, *Stem Cells*, 29(310), pp. 486–495. doi: 10.1002/stem.590.

Marcy, G. and Raineteau, O. (2019) ‘Contributions of single-cell approaches for probing heterogeneity and dynamics of neural progenitors throughout life: Concise review’, *Stem Cells*, 37(11), pp. 1381–1388. doi: 10.1002/stem.3071.

Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., and Stunnenberg, H. G. (2012) ‘The transcriptional and epigenomic foundations of ground state pluripotency’, *Cell*, 149(3), pp. 590–604. doi: <https://doi.org/10.1016/j.cell.2012.03.026>.

Martano, G., Borroni, E. M., Lopci, E., Cattaneo, M. G., Mattioli, M., Bachi, A., Decimo, I., and Bifari, F. (2019) ‘Metabolism of stem and progenitor cells: Proper methods to answer specific questions’, *Frontiers in Molecular Neuroscience*, 12, p. 151. doi: 10.3389/fnmol.2019.00151.

Martin, G. R. (1981) ‘Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.’, *Proceedings of the National Academy of Sciences of the United States of America*. United States, 78(12), pp. 7634–7638. doi: 10.1073/pnas.78.12.7634.

Martynoga, B., Drechsel, D., and Guillemot, F. (2012) ‘Molecular control of neurogenesis: A view from the mammalian cerebral cortex’, *Cold Spring Harbor Perspectives in Biology*, 4(10), p. a008359. doi: 10.1101/cshperspect.a008359.

Masala, A., Sanna, S., Esposito, S., Rassu, M., Galioto, M., Zinellu, A., Carru, C., Carri, M. T., Iaccarino, C., and Crosio, C. (2018) ‘Epigenetic changes associated with the expression of amyotrophic lateral sclerosis (ALS) causing genes’, *Neuroscience*, 390, pp. 1–11. doi: 10.1016/j.neuroscience.2018.08.009.

McCluskey, G., Duddy, W., Haffey, S., Morrison, K., Donaghy, C., and Duguez, S. (2022) ‘Epidemiology and survival trends of motor neurone disease in Northern Ireland from 2015 to 2019’, *European Journal of Neurology*. John Wiley & Sons, Ltd, 29(3), pp. 707–714. doi: 10.1111/ene.15172.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018) ‘UMAP: Uniform Manifold Approximation and Projection’, *Journal of Open Source Software*, 3(29), p. 861. doi: 10.21105/joss.00861.

McMahon, J. A., Takada, S., Zimmerman, L. B., Fan, C.-M., Harland, R. M., and McMahon, A. P. (1998) ‘Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite’, *Genes & Development*, 12(10), pp. 1438–1452. doi: 10.1101/gad.12.10.1438.

Melton, C., Judson, R. L., and Blueloch, R. (2010) ‘Opposing microRNA families regulate self-renewal in mouse embryonic stem cells’, *Nature*, 463(7281), pp. 621–626. doi: 10.1038/nature08725.

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., and Shilatifard, A. (2001) ‘COMPASS: A complex of proteins associated with a trithorax-related SET domain protein’, *Proceedings of the National Academy of Sciences*,

98(23), pp. 12902–12907. doi: 10.1073/pnas.231473398.

Minchiotti, G., D’Aniello, C., Fico, A., De Cesare, D., and Patriarca, E. J. (2022) ‘Capturing transitional pluripotency through proline metabolism’, *Cells*, 11(14), p. 2125. doi: 10.3390/cells11142125.

Mirra, A., Rossi, S., Scaricamazza, S., Di Salvio, M., Salvatori, I., Valle, C., Rusmini, P., Poletti, A., Cestra, G., Carri, M. T., and Cozzolino, M. (2017) ‘Functional interaction between FUS and SMN underlies SMA-like splicing changes in wild-type hFUS mice.’, *Scientific reports*. England, 7(1), p. 2033. doi: 10.1038/s41598-017-02195-0.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) ‘The homeoprotein Nanog is required for maintenance of pluripotency in mouse ppiblast and ES cells’, *Cell*, 113(5), pp. 631–642. doi: 10.1016/S0092-8674(03)00393-3.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., and Reik, W. (2017) ‘Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation’, *Cell Reports*, 20(5), pp. 1215–1228. doi: 10.1016/j.celrep.2017.07.009.

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., and Di Croce, L. (2012) ‘Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells’, *Cell Stem Cell*, 10(1), pp. 47–62. doi: 10.1016/j.stem.2011.12.006.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E. P., Bruneau, B. G., and Di Croce, L. (2015) ‘Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms’, *Cell Stem Cell*. Elsevier, 17(3), pp. 300–315. doi: 10.1016/j.stem.2015.08.009.

Morey, L. and Helin, K. (2010) ‘Polycomb group protein-mediated repression of

transcription', *Trends in Biochemical Sciences*, 35(6), pp. 323–332. doi: 10.1016/j.tibs.2010.02.009.

Murre, C., McCaw, P. S., and Baltimore, D. (1989) 'A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins', *Cell*. MIT,DEPT BIOL,CAMBRIDGE,MA 02139, 56(5), pp. 777–783. doi: 10.1016/0092-8674(89)90682-X.

Newman, M. A., Thomson, J. M., and Hammond, S. M. (2008) 'Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing', *RNA*, 14(8), pp. 1539–1549. doi: 10.1261/rna.1155108.

Nichols, J. and Smith, A. (2009) 'Naive and primed pluripotent states', *Cell Stem Cell*, 4(6), pp. 487–492. doi: 10.1016/j.stem.2009.05.015.

Nichols, J. and Smith, A. (2012) 'Pluripotency in the embryo and in culture', *Cold Spring Harbor Perspectives in Biology*, 4(8), pp. a008128–a008128. doi: 10.1101/cshperspect.a008128.

Nishimura, K., Fukuda, A., and Hisatake, K. (2019) 'Mechanisms of the metabolic shift during somatic cell reprogramming', *International Journal of Molecular Sciences*, 20(9), p. 2254. doi: 10.3390/ijms20092254.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009) 'A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells', *Nature*, 460(7251), pp. 118–122. doi: 10.1038/nature08113.

Nowotschin, S., Setty, M., Kuo, Y.-Y., Liu, V., Garg, V., Sharma, R., Simon, C. S., Saiz, N., Gardner, R., Boutet, S. C., Church, D. M., Hoodless, P. A., Hadjantonakis, A.-K., and Pe'er, D. (2019) 'The emergent landscape of the mouse gut endoderm at single-cell resolution', *Nature*, 569(7756), pp. 361–367. doi: 10.1038/s41586-019-1127-1.

O'Brien, L. C., Keeney, P. M., and Bennett, J. P. (2015) 'Differentiation of human neural stem cells into motor neurons stimulates mitochondrial biogenesis and decreases glycolytic flux', *Stem Cells and Development*, 24(17), pp. 1984–1994. doi: 10.1089/scd.2015.0076.

Oates, N. and Pamphlett, R. (2007) 'An epigenetic analysis of SOD1 and VEGF in ALS', *Amyotrophic Lateral Sclerosis*, 8(2), pp. 83–86. doi: 10.1080/17482960601149160.

Ochocki, J. D. and Simon, M. C. (2013) 'Nutrient-sensing pathways and metabolic regulation in stem cells', *Journal of Cell Biology*, 203(1), pp. 23–33. doi: 10.1083/jcb.201303110.

Oh, S. K., Kim, H. S., Ahn, H. J., Seol, H. W., Kim, Y. Y., Park, Y. Bin, Yoon, C. J., Kim, D.-W., Kim, S. H., and Moon, S. Y. (2005) 'Derivation and characterization of new human embryonic stem cell lines: SNUhES1, SNUhES2, and SNUhES3', *Stem Cells*, 23(2), pp. 211–219. doi: 10.1634/stemcells.2004-0122.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005) 'Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells', *Journal of Biological Chemistry*, 280(7), pp. 5307–5317. doi: 10.1074/jbc.M410015200.

Osteil, P., Studdert, J., Wilkie, E., Fossat, N., and Tam, P. P. L. (2016) 'Generation of genome-edited mouse epiblast stem cells via a detour through ES cell-chimeras', *Differentiation*. Elsevier, 91(4–5), pp. 119–125. doi: 10.1016/j.diff.2015.10.004.

Pandhare, J., Cooper, S. K., Donald, S. P., and Phang, J. M. (2009) 'Regulation and function of proline oxidase under nutrient stress', *Journal of Cellular biochemistry*, 107(4), pp. 759–768. doi: 10.1002/jcb.22174.

Panopoulos, A. D., Yanes, O., Ruiz, S., Kida, Y. S., Diep, D., Tautenhahn, R., Herrerías, A., Batchelder, E. M., Plongthongkum, N., Lutz, M., Berggren, W. T., Zhang, K., Evans, R. M., Siuzdak, G., and Belmonte, J. C. I. (2012) 'The metabolome of induced pluripotent stem

cells reveals metabolic changes occurring in somatic cell reprogramming’, *Cell Research*, 22(1), pp. 168–177. doi: 10.1038/cr.2011.177.

Park, J., Kim, J.-E., and Song, T.-J. (2022) ‘The global burden of motor neuron disease: An analysis of the 2019 Global Burden of Disease Study’, *Frontiers in Neurology*. Switzerland, 13, p. 864339. doi: 10.3389/fneur.2022.864339.

Parker, G. C., Li, X., Anguelov, R. A., Toth, G., Cristescu, A., and Acsadi, G. (2008) ‘Survival motor Neuron protein regulates apoptosis in an in vitro model of spinal muscular atrophy’, *Neurotoxicity Research*, 13(1), pp. 39–48. doi: 10.1007/BF03033366.

Pauklin, S. and Vallier, L. (2013) ‘The cell-cycle state of stem cells determines cell fate propensity’, *Cell*, 155(1), pp. 135–147. doi: <https://doi.org/10.1016/j.cell.2013.08.031>.

Pauklin, S. and Vallier, L. (2015) ‘Activin/Nodal signalling in stem cells’, *Development*, 142(4), pp. 607–619. doi: 10.1242/dev.091769.

Pauli, A., Rinn, J. L., and Schier, A. F. (2011) ‘Non-coding RNAs as regulators of embryogenesis’, *Nature Reviews Genetics*, 12(2), pp. 136–149. doi: 10.1038/nrg2904.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., and Rathjen, P. D. (2002) ‘Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development’, *Journal of Cell Science*, 115(2), pp. 329–339. doi: 10.1073/pnas.95.9.5082.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., Song, L., Tam, P. P. L., Han, J.-D. J., and Jing, N. (2016) ‘Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo’, *Developmental Cell*, 36(6), pp. 681–697. doi: <https://doi.org/10.1016/j.devcel.2016.02.020>.

Phang, J. M., Liu, W., and Zabirnyk, O. (2010) ‘Proline metabolism and microenvironmental stress’, *Annual Review of Nutrition*, 30(1), pp. 441–463. doi:

Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P., and Gregory, R. I. (2008) 'Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28', *Journal of Biological Chemistry*, 283(31), pp. 21310–21314. doi: 10.1074/jbc.C800108200.

Piunti, A. and Shilatifard, A. (2016) 'Epigenetic balance of gene expression by polycomb and COMPASS families', *Science*, 352(6290), p. aad9780. doi: 10.1126/science.aad9780.

Placzek, M. and Briscoe, J. (2005) 'The floor plate: multiple cells, multiple signals', *Nature Reviews Neuroscience*, 6(3), pp. 230–240. doi: 10.1038/nrn1628.

Prigione, A., Rohwer, N., Hoffmann, S., Mlody, B., Drews, K., Bukowiecki, R., Blümlein, K., Wanker, E. E., Ralser, M., Cramer, T., and Adjaye, J. (2014) 'HIF1 α modulates cell fate reprogramming through early glycolytic shift and upregulation of PDK1–3 and PKM2', *Stem Cells*. John Wiley & Sons, Ltd, 32(2), pp. 364–376. doi: 10.1002/stem.1552.

Qi, X., Li, T.-G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G.-Q. (2004) 'BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(16), pp. 6027–6032. doi: 10.1073/pnas.0401367101.

Qiu, L., Qiao, T., Beers, M., Tan, W., Wang, H., Yang, B., and Xu, Z. (2013) 'Widespread aggregation of mutant VAPB associated with ALS does not cause motor neuron degeneration or modulate mutant SOD1 aggregation and toxicity in mice', *Molecular Neurodegeneration*, 8(1), p. 1. doi: 10.1186/1750-1326-8-1.

Rao, C., Malaguti, M., Mason, J. O., and Lowell, S. (2020) 'The transcription factor E2A drives neural differentiation in pluripotent cells', *Development*, 147(12), p. dev184093. doi:

10.1242/dev.184093.

Rathjen, J., Lake, J.-A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999) 'Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors', *Journal of Cell Science*, 112(5), pp. 601–612. doi: 10.1242/jcs.004613.

Rathjen, J., Haines, B. P., Hudson, K. M., Nesci, A., Dunn, S., and Rathjen, P. D. (2002) 'Directed differentiation of pluripotent cells to neural lineages: homogeneous formation and differentiation of a neurectoderm population', *Development*, 129(11), pp. 2649–2661. Available at: <http://dev.biologists.org/content/129/11/2649>.

Rehman, J. (2010) 'Empowering self-renewal and differentiation: the role of mitochondria in stem cells', *Journal of Molecular Medicine*, 88(10), pp. 981–986. doi: 10.1007/s00109-010-0678-2.

Renton, A. E., Chiò, A., and Traynor, B. J. (2014) 'State of play in amyotrophic lateral sclerosis genetics', *Nature Neuroscience*, 17(1), pp. 17–23. doi: 10.1038/nn.3584.

Rezaie, T., Waitzman, D. M., Seeman, J. L., Kaufman, P. L., and Sarfarazi, M. (2005) 'Molecular cloning and expression profiling of optineurin in the rhesus monkey', *Investigative Ophthalmology & Visual Science*, 46(7), pp. 2404–2410. doi: 10.1167/iovs.04-1243.

Rezaie, T. and Sarfarazi, M. (2005) 'Molecular cloning, genomic structure, and protein characterization of mouse optineurin', *Genomics*, 85(1), pp. 131–138. doi: <https://doi.org/10.1016/j.ygeno.2004.10.011>.

Ricketts, T., McGoldrick, P., Fratta, P., de Oliveira, H. M., Kent, R., Phatak, V., Brandner, S., Blanco, G., Greensmith, L., Acevedo-Arozena, A., and Fisher, E. M. C. (2014) 'A nonsense mutation in mouse Tardbp Affects TDP43 alternative splicing activity and causes limb-clasping and body tone defects', *PLOS ONE*. Public Library of Science, 9(1), p. e85962.

Ringrose, L. and Paro, R. (2004) 'Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins', *Annual Review of Genetics*. Annual Reviews, 38(1), pp. 413–443. doi: 10.1146/annurev.genet.38.072902.091907.

Rodda, S. J., Kavanagh, S. J., Rathjen, J., and Rathjen, P. D. (2002) 'Embryonic stem cell differentiation and the analysis of mammalian development', *The International Journal of Developmental Biology*, 46(4), pp. 449–458. doi: 10.1387/IJDB.12141431.

Rodrigues, A. S., Pereira, S. L., Correia, M., Gomes, A., Perestrelo, T., and Ramalho-Santos, J. (2015) 'Differentiate or die: 3-bromopyruvate and pluripotency in mouse embryonic stem cells', *PLoS ONE*, 10(8), p. e0135617. doi: 10.1371/journal.pone.0135617.

Ryu, J. M., Lee, S. H., Seong, J. K., and Han, H. J. (2015) 'Glutamine contributes to maintenance of mouse embryonic stem cell self-renewal through PKC-dependent downregulation of HDAC1 and DNMT1/3a', *Cell Cycle*. Taylor & Francis, 14(20), pp. 3292–3305. doi: 10.1080/15384101.2015.1087620.

Ryu, J. M. and Han, H. J. (2011) 'L-Threonine regulates G1/S phase transition of mouse embryonic stem cells via PI3K/Akt, MAPKs, and mTORC pathways', *Journal of Biological Chemistry*, 286(27), pp. 23667–23678. doi: 10.1074/jbc.M110.216283.

Sampath, P., Pritchard, D. K., Pabon, L., Reinecke, H., Schwartz, S. M., Morris, D. R., and Murry, C. E. (2008) 'A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation', *Cell Stem Cell*, 2(5), pp. 448–460. doi: 10.1016/j.stem.2008.03.013.

Santanach, A., Blanco, E., Jiang, H., Molloy, K. R., Sansó, M., LaCava, J., Morey, L., and Di Croce, L. (2017) 'The polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity', *Nature Communications*, 8(1), p. 1235. doi: 10.1038/s41467-

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002) 'Dynamic reprogramming of DNA methylation in the early mouse embryo', *Developmental Biology*, 241(1), pp. 172–182. doi: 10.1006/dbio.2001.0501.

Schrank, B., Götz, R., Gunnensen, J. M., Ure, J. M., Toyka, K. V, Smith, A. G., and Sendtner, M. (1997) 'Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences of the USA, 94(18), pp. 9920–9925. doi: 10.1073/pnas.94.18.9920.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007) 'Genome regulation by Polycomb and Trithorax proteins', *Cell*, 128(4), pp. 735–745. doi: 10.1016/j.cell.2007.02.009.

Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. (2011) 'Trithorax group proteins: switching genes on and keeping them active', *Nature Reviews Molecular Cell Biology*, 12(12), pp. 799–814. doi: 10.1038/nrm3230.

Sephton, C. F., Good, S. K., Atkin, S., Dewey, C. M., Mayer III, P., Herz, J., and Yu, G. (2010) 'TDP-43 is a developmentally regulated protein essential for early embryonic development', *Journal of Biological Chemistry*. Elsevier, 285(9), pp. 6826–6834. doi: 10.1074/jbc.M109.061846.

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., Huang, K., Liu, Y., Wang, X., Chen, Q., Zhang, J., Shang, B., Li, S., Shi, X., Liao, B., Zhang, C., Lai, K., Zhong, X., Shu, X., Wang, J., Yao, H., Chen, J., Pei, D., and Pan, G. (2017) 'PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs', *Nature Communications*, 8, p. 672. doi: 10.1038/s41467-017-00668-4.

Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., Enikolopov, G., Nauen, D. W., Christian, K. M., Ming, G., and Song, H. (2015) ‘Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis’, *Cell Stem Cell*, 17(3), pp. 360–372. doi: 10.1016/j.stem.2015.07.013.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019a) ‘Embryoid body differentiation of mouse embryonic stem cells into neurectoderm and neural progenitors’, in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019b) ‘Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells’, *Frontiers in Physiology*, 10, p. 705. doi: 10.3389/fphys.2019.00705.

Shyh-Chang, Ng, Locasale, J. W., Lyssiotis, C. A., Zheng, Y., Teo, R. Y., Ratanasirintrao, S., Zhang, J., Onder, T., Unternaehrer, J. J., Zhu, H., Asara, J. M., Daley, G. Q., and Cantley, L. C. (2013) ‘Influence of threonine metabolism on S-adenosylmethionine and histone methylation’, *Science*, 339(6116), p. 222. doi: 10.1126/science.1226603.

Shyh-Chang, N, Daley, G. Q., and Cantley, L. C. (2013) ‘Stem cell metabolism in tissue development and aging’, *Development*, 140(12), pp. 2535–2547. doi: 10.1242/dev.091777.

Singh, A. K., Halder-Sinha, S., Clement, J. P., and Kundu, T. K. (2018) ‘Epigenetic modulation by small molecule compounds for neurodegenerative disorders’, *Pharmacological Research*, 132, pp. 135–148. doi: <https://doi.org/10.1016/j.phrs.2018.04.014>.

Sirohi, K., Chalasani, M. L. S., Sudhakar, C., Kumari, A., Radha, V., and Swarup, G. (2013) ‘M98K-OPTN induces transferrin receptor degradation and RAB12-mediated

autophagic death in retinal ganglion cells', *Autophagy*, 9(4), pp. 510–527. doi: 10.4161/auto.23458.

Smith, J. R., Vallier, L., Lupo, G., Alexander, M., Harris, W. A., and Pedersen, R. A. (2008) 'Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm', *Developmental Biology*. Academic Press, 313(1), pp. 107–117. doi: 10.1016/j.ydbio.2007.10.003.

Snow, M. H. L. (1977) 'Gastrulation in the mouse: Growth and regionalization of the epiblast', *Development*, 42(1), pp. 293–303. Available at: <http://dev.biologists.org/content/42/1/293>.

Son, G. and Han, J. (2018) 'Roles of mitochondria in neuronal development', *BMB Reports*. 2018/11/30, 51(11), pp. 549–556. doi: 10.5483/BMBRep.2018.51.11.226.

Southall, S. M., Wong, P.-S., Odho, Z., Roe, S. M., and Wilson, J. R. (2009) 'Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks', *Molecular Cell*, 33(2), pp. 181–191. doi: <https://doi.org/10.1016/j.molcel.2008.12.029>.

Spangler, A., Su, E. Y., Craft, A. M., and Cahan, P. (2018) 'A single cell transcriptional portrait of embryoid body differentiation and comparison to progenitors of the developing embryo', *Stem Cell Research*, 31, pp. 201–215. doi: 10.1016/j.scr.2018.07.022.

Spencer, W. A., Jeyabalan, J., Kichambre, S., and Gupta, R. C. (2011) 'Oxidatively generated DNA damage after Cu(II) catalysis of dopamine and related catecholamine neurotransmitters and neurotoxins: Role of reactive oxygen species.', *Free radical biology & medicine*. United States, 50(1), pp. 139–147. doi: 10.1016/j.freeradbiomed.2010.10.693.

Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., Fischer, K. A., Devi, A., Detraux, D., Gu, H., Battle, S. L., Showalter, M., Valensisi, C., Bielas, J. H., Ericson, N.

G., Margaretha, L., Robitaille, A. M., Margineantu, D., Fiehn, O., Hockenbery, D., Blau, C. A., Raftery, D., Margolin, A. A., Hawkins, R. D., Moon, R. T., Ware, C. B., and Ruohola-Baker, H. (2015) 'The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition', *Nature Cell Biology*, 17(12), pp. 1523–1535. doi: 10.1038/ncb3264.

Sternecker, J. L., Reinhardt, P., and Schöler, H. R. (2014) 'Investigating human disease using stem cell models', *Nature Reviews Genetics*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15(9), pp. 625–639. doi: 10.1038/nrg3764.

Sun, Z., Zhu, M., Lv, P., Cheng, L., Wang, Q., Tian, P., Yan, Z., and Wen, B. (2018) 'The Long noncoding RNA Lncenc1 maintains naive states of mouse ESCs by promoting the glycolysis pathway', *Stem Cell Reports*, 11(3), pp. 741–755. doi: 10.1016/j.stemcr.2018.08.001.

Suter, D. M., Tirefort, D., Julien, S., and Krause, K.-H. (2009) 'A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells', *Stem Cells*, 27(1), pp. 49–58. doi: 10.1634/stemcells.2008-0319.

Sze, C. C., Cao, K., Collings, C. K., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Chen, F. X., Morgan, M. A., Wang, L., and Shilatifard, A. (2017) 'Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation', *Genes & Development*, 31(17), pp. 1732–1737. doi: 10.1101/gad.303768.117.

Takahashi, S., Kobayashi, S., and Hiratani, I. (2018) 'Epigenetic differences between naïve and primed pluripotent stem cells', *Cellular and Molecular Life Sciences*, 75(7), pp. 1191–1203. doi: 10.1007/s00018-017-2703-x.

Tam, P. P. L. and Behringer, R. R. (1997) ‘Mouse gastrulation: The formation of a mammalian body plan’, *Mechanisms of Development*, 68(1–2), pp. 3–25. doi: 10.1016/S0925-4773(97)00123-8.

Tam, P. P. L. and Ho, J. W. K. (2020) ‘Cellular diversity and lineage trajectory: Insights from mouse single cell transcriptomes’, *Development*, 147(2), pp. 1–7. doi: 10.1242/dev.179788.

Tam, P. P. L. and Loebel, D. A. F. (2007) ‘Gene function in mouse embryogenesis: get set for gastrulation’, *Nature Reviews Genetics*, 8(5), pp. 368–381. doi: 10.1038/nrg2084.

Tam, P. P. L., Loebel, D. A. F., and Tanaka, S. S. (2006) ‘Building the mouse gastrula: signals, asymmetry and lineages’, *Current Opinion in Genetics & Development*, 16(4), pp. 419–425. doi: 10.1016/J.GDE.2006.06.008.

Tan, B. S. N., Lonic, A., Morris, M. B., Rathjen, P. D., and Rathjen, J. (2011) ‘The amino acid transporter SNAT2 mediates L-proline-induced differentiation of ES cells’, *American Journal of Physiology-Cell Physiology*, 300(6), pp. 1270–1279. doi: 10.1152/ajpcell.00235.2010.

Tanner, J. J. (2008) ‘Structural biology of proline catabolism’, *Amino Acids*, 35(4), pp. 719–730. doi: 10.1007/s00726-008-0062-5.

Telley, L., Govindan, S., Prados, J., Stevant, I., Nef, S., Dermitzakis, E., Dayer, A., and Jabaudon, D. (2016) ‘Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex’, *Science*, 351(6280), pp. 1443–1446. doi: 10.1126/science.aad8361.

Teoh, H. L., Carey, K., Sampaio, H., Mowat, D., Roscioli, T., and Farrar, M. (2017) ‘Inherited paediatric motor neuron disorders: Beyond spinal muscular atrophy’, *Neural Plasticity*. Hindawi, 2017, pp. 1–22. doi: 10.1155/2017/6509493.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., and McKay, R. D. G. (2007) 'New cell lines from mouse epiblast share defining features with human embryonic stem cells', *Nature*, 448(7150), pp. 196–199. doi: 10.1038/nature05972.

Teuling, E., Ahmed, S., Haasdijk, E., Demmers, J., Steinmetz, M. O., Akhmanova, A., Jaarsma, D., and Hoogenraad, C. C. (2007) 'Motor neuron disease-associated mutant vesicle-associated membrane protein-associated protein (VAP) B recruits wild-type VAPs into endoplasmic reticulum-derived tubular aggregates', *Journal of Neuroscience*, 27(36), pp. 9801–9815. doi: 10.1523/JNEUROSCI.2661-07.2007.

Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A., and Ramanathan, S. (2011) 'Pluripotency factors in embryonic stem cells regulate differentiation into germ layers.', *Cell*, 145(6), pp. 875–889. doi: 10.1016/j.cell.2011.05.017.

Tosolini, M. and Jouneau, A. (2015) 'Acquiring ground state pluripotency: Switching mouse embryonic stem cells from serum/LIF medium to 2i/LIF medium', in Turksen, K. (ed.) *Methods in Molecular Biology*. New York, NY: Springer New York, pp. 41–48. doi: 10.1007/7651_2015_207.

Toth, R. P. and Atkin, J. D. (2018) 'Dysfunction of optineurin in amyotrophic lateral sclerosis and glaucoma', *Frontiers in Immunology*, 9, p. 1017. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2018.01017>.

Tsai, P.-H., Chien, Y., Wang, M.-L., Hsu, C.-H., Laurent, B., Chou, S.-J., Chang, W.-C., Chien, C.-S., Li, H.-Y., Lee, H.-C., Huo, T.-I., Hung, J.-H., Chen, C.-H., and Chiou, S.-H. (2019) 'Ash2l interacts with Oct4-stemness circuitry to promote super-enhancer-driven pluripotency network', *Nucleic Acids Research*, 47(19), pp. 10115–10133. doi: 10.1093/nar/gkz801.

Tsogtbaatar, E., Landin, C., Minter-Dykhouse, K., and Folmes, C. D. L. (2020) 'Energy metabolism regulates stem cell pluripotency', *Frontiers in Cell and Developmental Biology*, 8, p. 87. doi: 10.3389/fcell.2020.00087.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L. E., Trotter, M. W. B., Cho, C. H.-H., Martinez, A., Rugg-Gunn, P., Brons, G., and Pedersen, R. A. (2009) 'Activin/Nodal signalling maintains pluripotency by controlling Nanog expression', *Development*, 136(8), pp. 1339–1349. doi: 10.1242/dev.033951.

Vallier, L., Touboul, T., Chng, Z., Brimpari, M., Hannan, N., Millan, E., Smithers, L. E., Trotter, M., Rugg-Gunn, P., Weber, A., and Pedersen, R. A. (2009) 'Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways', *PLoS ONE*. Edited by P. Callaerts. United States, 4(6), p. e6082. doi: 10.1371/journal.pone.0006082.

Varum, S., Rodrigues, A. S., Moura, M. B., Momcilovic, O., Easley IV, C. A., Ramalho-Santos, J., van Houten, B., and Schatten, G. (2011) 'Energy metabolism in human pluripotent stem cells and their differentiated counterparts', *PLoS ONE*, 6(6), pp. 1–15. doi: 10.1371/journal.pone.0020914.

Wade, S. L., Langer, L. F., Ward, J. M., and Archer, T. K. (2015) 'MiRNA-mediated regulation of the SWI/SNF chromatin remodeling complex controls pluripotency and endodermal differentiation in human eSCs', *Stem Cells*, 33(10), pp. 2925–2935. doi: 10.1002/stem.2084.

Wan, M., Liang, J., Xiong, Y., Shi, F., Zhang, Y., Lu, W., He, Q., Yang, D., Chen, R., Liu, D., Barton, M., and Songyang, Z. (2013) 'The Trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells', *Journal of Biological Chemistry*, 288(7), pp. 5039–5048. doi: 10.1074/jbc.M112.424515.

Wang, C., Lee, J.-E., Cho, Y.-W., Xiao, Y., Jin, Q., Liu, C., and Ge, K. (2012) 'UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity', *Proceedings of the National Academy of Sciences*, 109(38), pp. 15324–15329. doi: 10.1073/pnas.1204166109.

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T. S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016) 'Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition', *Proceedings of the National Academy of Sciences*, 113(42), pp. 11871–11876. doi: 10.1073/pnas.1606857113.

Wang, J., Alexander, P., Wu, L., Hammer, R., Cleaver, O., and McKnight, S. L. (2009) 'Dependence of mouse embryonic stem cells on Threonine catabolism', *Science*, 325(5939), pp. 435–439. doi: 10.1126/science.1173288.

Wang, J., Alexander, P., and McKnight, S. L. (2011) 'Metabolic specialization of mouse embryonic stem cells', *Cold Spring Harbor Symposia on Quantitative Biology*, 76, pp. 183–193. doi: 10.1101/sqb.2011.76.010835.

Wang, Q., Zou, Y., Nowotschin, S., Kim, S. Y., Li, Q. V, Soh, C.-L., Su, J., Zhang, C., Shu, W., Xi, Q., Huangfu, D., Hadjantonakis, A.-K., and Massagué, J. (2017) 'The p53 family coordinates Wnt and Nodal inputs in mesendodermal differentiation of embryonic stem cells', *Cell Stem Cell*, 20(1), pp. 70–86. doi: <https://doi.org/10.1016/j.stem.2016.10.002>.

Warrier, S., Taelman, J., Tilleman, L., Van der Jeught, M., Duggal, G., Lierman, S., Popovic, M., Van Soom, A., Peelman, L., Van Nieuwerburgh, F., Deforce, D., Chuva de Sousa Lopes, S. M., De Sutter, P., and Heindryckx, B. (2018) 'Transcriptional landscape changes during human embryonic stem cell derivation', *Molecular Human Reproduction*, 24(11), pp. 543–555. doi: 10.1093/molehr/gay039.

Washington, J. M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M. D., Hamra, N.,

Semendric, L., Tan, B. S. N., Lake, J. A., Keough, R. A., Morris, M. B., and Rathjen, P. D. (2010) 'L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture', *American Journal of Physiology-Cell Physiology*, 298(5), pp. C982–C992. doi: 10.1152/ajpcell.00498.2009.

Wine-Lee, L., Ahn, K. J., Richardson, R. D., Mishina, Y., Lyons, K. M., and Crenshaw, E. B. (2004) 'Signaling through BMP type 1 receptors is required for development of interneuron cell types in the dorsal spinal cord', *Development*, 131(21), pp. 5393–5403. doi: 10.1242/dev.01379.

Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. (2011) 'Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation', *Nature Cell Biology*. England, 13(7), pp. 838–845. doi: 10.1038/ncb2267.

Wray, J., Kalkan, T., and Smith, A. G. (2010) 'The ground state of pluripotency', *Biochemical Society Transactions*, 38(4), pp. 1027–1032. doi: 10.1042/BST0381027.

Wu, G., Wu, Z., Dai, Z., Yang, Y., Wang, W., Liu, C., Wang, B., Wang, J., and Yin, Y. (2013) 'Dietary requirements of “nutritionally non-essential amino acids” by animals and humans', *Amino Acids*, 44(4), pp. 1107–1113. doi: 10.1007/s00726-012-1444-2.

Wu, G. (2013) 'Functional amino acids in nutrition and health', *Amino Acids*, 45(3), pp. 407–411. doi: 10.1007/s00726-013-1500-6.

Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008) 'Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS', *Molecular and Cellular Biology*, 28(24), pp. 7337–7344. doi: 10.1128/MCB.00976-08.

Xu, L., Liu, T., Liu, L., Yao, X., Chen, L., Fan, D., Zhan, S., and Wang, S. (2020) 'Global

variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis', *Journal of Neurology*, 267(4), pp. 944–953. doi: 10.1007/s00415-019-09652-y.

Xu, R.-H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., and Thomson, J. A. (2005) 'Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells', *Nature Methods*, 2(3), pp. 185–190. doi: 10.1038/nmeth744.

Xue, K., Ng, J.-H., and Ng, H.-H. (2011) 'Mapping the networks for pluripotency', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1575), pp. 2238–2246. doi: 10.1098/rstb.2011.0005.

Yanling Zhao, D., Gish, G., Braunschweig, U., Li, Y., Ni, Z., Schmitges, F. W., Zhong, G., Liu, K., Li, W., Moffat, J., Vedadi, M., Min, J., Pawson, T. J., Blencowe, B. J., and Greenblatt, J. F. (2016) 'SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination', *Nature*, 529(7584), pp. 48–53. doi: 10.1038/nature16469.

Yao, M., Zhou, X., Zhou, J., Gong, S., Hu, G., Li, J., Huang, K., Lai, P., Shi, G., Hutchins, A. P., Sun, H., Wang, H., and Yao, H. (2018) 'PCGF5 is required for neural differentiation of embryonic stem cells', *Nature Communications*, 9(1), p. 1463. doi: 10.1038/s41467-018-03781-0.

Ybot-Gonzalez, P., Cogram, P., Gerrelli, D., and Copp, A. J. (2002) 'Sonic hedgehog and the molecular regulation of mouse neural tube closure', *Development*, 129(10), pp. 2507–2517. Available at: <http://dev.biologists.org/content/129/10/2507>.

Ying, H. and Yue, B. Y. J. T. (2012) 'Cellular and molecular biology of optineurin', in Jeon, K. W. B. T.-I. R. of C. and M. B. (ed.) *International Review of Cell and Molecular Biology*. Academic Press, pp. 223–258. doi: 10.1016/B978-0-12-394305-7.00005-7.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003) 'BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3', *Cell*, 115, pp. 281–292. doi: 10.1016/S0092-8674(03)00847-X.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008) 'The ground state of embryonic stem cell self-renewal', *Nature*, 453(7194), pp. 519–523. doi: 10.1038/nature06968.

Zeng, Z., Miao, N., and Sun, T. (2018) 'Revealing cellular and molecular complexity of the central nervous system using single cell sequencing', *Stem Cell Research & Therapy*, 9(1), p. 234. doi: 10.1186/s13287-018-0985-z.

Zheng, X., Boyer, L., Jin, M., Mertens, J., Kim, Y., Ma, L., Ma, L., Hamm, M., Gage, F. H., and Hunter, T. (2016) 'Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation', *eLife*, 5, p. e13374. doi: 10.7554/eLife.13374.

Zhou, H., Li, W., Zhu, S., Joo, J. Y., Do, J. T., Xiong, W., Kim, J. B., Zhang, K., Schöler, H. R., and Ding, S. (2010) 'Conversion of mouse epiblast stem Cells to an earlier pluripotency state by small molecules', *Journal of Biological Chemistry*. Elsevier, 285(39), pp. 29676–29680. doi: 10.1074/jbc.C110.150599.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C. A., Horwitz, M. S., Hockenberg, D., Ware, C., and Ruohola-Baker, H. (2012) 'HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition', *The EMBO Journal*. Nature Publishing Group, 31(9), pp. 2103–2116. doi: 10.1038/emboj.2012.71.

Chapter 2: Materials and Methods

2.1. Cell Culture

2.1.1. Materials

Reagent stocks and cell culture components

All water-based powders were prepared in MilliQ water (MQ) unless specified.

- 0.1% porcine gelatin [Sigma: #G1890-100G] in phosphate buffered saline (PBS)
- 0.25% Trypsin/ ethylenediaminetetraacetic acid (EDTA) (1x) [Gibco: #25200-072]
- 0.05% Trypsin/EDTA (1x) [Gibco: #25300-054]
- β -mercaptoethanol (β -mec) [Sigma] 0.1 M stock in PBS
- CHIR99021 [Sigma #SML1046] 3 mM stock in dimethylsulfoxide (DMSO)
- Collagenase P [Roche #11213865001] 1mg/mL stock in PBS
- F12 nutrient mixture (1x) [Sigma #N6760]
- F12 nutrient mixture (HAM) (1x) [Gibco Ref #ME16225L1; Lot #UNS11622901, no L-Proline, no L-Glutamine]
- Fetal bovine serum (FBS) [AusGenex #FBS500-S or Gibco #16000-044]
- Dulbecco's Modified Eagle Medium (DMEM) [Gibco: #12800-017 or Sigma: #D7777]
- Insulin-transferrin-sodium selenite (ITSS) (100x) [Sigma: #I1884]
- Knockout serum replacement (KOSR) [Gibco: #10828:028]
- Leukaemia inhibitory factor (LIF) [Millipore: #ESG1107 or Neuromics: #PR80000 or StemCell Technologies #78056] 10^7 original stock aliquoted into 10^6 U/mL working stocks in PBS
- L-Proline: [Sigma: #P5607] 0.1 M working stock
- PD0325901 [Sigma #PZ0162] 1 mM stock in DMSO
- Penicillin/streptomycin (pen-strep) (100x) containing 5000 U/ml Penicillin G Sodium Salt [Sigma: #P3032-10MU], 5 mg/ml streptomycin sulfate [Sigma: #S-6501] in PBS
- PBS [Gibco: #21600-010 or Amresco: #E404]
- SB431542 [MedChemExpress: #HY-10431] 2.6 mM stock in DMSO

- Trypan blue: 0.4 g Trypan blue, 0.81 g NaCl, 0.06 g KH₂PO₄ in 95 ml MQ, pH: 7.2-7.3
- TrypLE (1x) [Gibco #12605010]

Chemicals

- 6-aminonicotinamide (6AN) [Sigma #A68203] 9 mM and 30 mM working stocks in DMSO
- Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone (FCCP) [ABCAM #ab120081] (20 mM) working stock in DMSO
- 3,4-Dehydro-L-Proline (DHP) [Chem-Supply #D3825] 120 mM working stock
- Dichloroacetic acid (DCA) [Sigma #347795]
- Tetrahydro-2-furoic acid (THFA) [Sigma #341517] 250 mM working stock
- Mitomycin C [Sigma #M0503]
- Y27632 [MedChem Express #HY-10583] in DMSO

Plasticware

- 0.22 µm 150 ml bottle top filters [Corning: #431161]
- 0.22 µm filter unit [Millipore: #SLGP33RS]
- 10 cm culture plates [Corning: #430591 or BD Falcon: #351029]
- 10 cm suspension culture petri dishes [Technoplas: #S9014UV20]
- 12-well culture plates [Corning #5313]
- 15 cm culture plates [Corning #430599]
- 24-well culture plates [Corning #3527]
- 48-well culture plates [Corning #3548 or BD Falcon #353078]
- 6 cm culture plates [Falcon: #353002]
- 6-well culture plates [Corning #3516]
- Agilent Seahorse unbuffered XF assay medium [In Vitro Technologies #SEA103335-100]
- Agilent Seahorse XF calibrant [In vitro Technologies #103059-000 or #100840-000]

- Agilent Seahorse XFe96 FluxPaks [In Vitro Technologies #SEA102416100]
- Seahorse XF96 cell culture microplates [In Vitro Technologies #101085-004]
- T-150 culture flask [Corning #431465]
- T-175 culture flask [Corning #CLS431080]
- T-25 culture flask [Sarstedt: Lot #83.1810.002]

2.1.2. Cell culture media

Cell culture media with final concentrations of components are listed in Table 2.1. Volumes used for cell culture were dependent on plate/flask surface area (Table 2.2).

Table 2.1. Cell culture media for mESC, HepG2, MEF and EpiSC maintenance, as well as adherent culture and embryoid body culture experiments.

Adherent culture medium	
mESC self-renewal medium	90% (v/v) DMEM, 10% (v/v) FBS, 1× pen-strep, 0.1 mM β -mec, 1000 U/mL LIF
mESC freezing medium	40% (v/v) mESC self-renewal medium, 40% (v/v) FBS, 20% (v/v) DMSO
EPLC medium	90% (v/v) DMEM, 10% (v/v) FBS, 1× pen-strep, 0.1 mM β -mec, 330 U/mL LIF, 0.4 mM L-Pro
2i medium	90% (v/v) DMEM, 10% (v/v) FBS, 1× pen-strep, 0.1 mM β -mec, 1000 U/mL LIF, 1 μ M PD0325901, 3 μ M CHIR99021
MEF and EpiSC culture medium	
MEF culture medium	90% (v/v) DMEM, 10% (v/v) FBS, 1x pen-strep, 1× glutamax, 1× NEAA
MEF freezing medium	60% (v/v) MEF culture medium, 20% (v/v) FBS, 20% (v/v) DMSO
MEF conditioning medium	MEFs were grow in 80% (v/v) KO-DMEM, 20% (v/v) KOSR, 1x pen-strep, 0.1 mM β -mec, 1× glutamax, 1× NEAA. Media was collected after 24 h
Complete EpiSC medium	MEF conditioning medium supplemented with 20 ng/mL Activin A and 10 ng/mL Fgf2
Embryoid body culture medium	
Basal medium	90% (v/v) DMEM, 10% (v/v) FBS, 1× pen-strep, 0.1 mM β -mec
HepG2 culture medium	90% (v/v) DMEM, 1× pen-strep, 10% (v/v) FBS

Complete MEDII (MEDII medium)	HepG2 cells were grown for 4-5 days in HepG2 culture medium for conditioning. This medium was then collected and supplemented with 50% Basal medium, 0.1 mM β -mec and 1 mM L-Gln
L-Proline-mediated differentiation medium (Day 0-5 embryoid body suspension culture)	80% (v/v) DMEM, 10% (v/v) FBS, 1x pen-strep, 0.1 mM β -mec, 10% (v/v) KOSR, 0.4 mM L-Pro
L-Proline-mediated differentiation medium (Day 6-9 embryoid body suspension culture)	80% (v/v) DMEM, 10% (v/v) FBS, 1x pen-strep, 0.1 mM β -mec, 10% (v/v) KOSR, 0.4 mM L-Pro, 10 μ M SB431542
Serum-free medium (Day 10-15 embryoid body adherent culture)	50% (v/v) DMEM, 50% (v/v) F12 nutrient mixture (HAM) supplemented with 1x pen-strep, 1x ITSS
Serum-free + L-Proline medium (Day 10-15 embryoid body adherent culture)	50% (v/v) DMEM, 50% (v/v) F12 nutrient mixture supplemented with 1x pen-strep, 1x ITSS, $\pm$ 0.25 mM L-Pro

Table 2.2 Plate sizes, seeding densities, and media volumes for maintenance of adherent cells.

mESC maintenance			mESC passaging		
Plate type	Seeding density (cells)	Medium volume (mL)	Trypsin volume (μ L)	Volume to deactivate trypsin (μ L)	Resuspended in for cell counting (μ L)
6 cm	$1-5 \times 10^5$	5	350	650	1000
10 cm	$5-10 \times 10^5$	10	1000	1000	1000
6-well	$1.5-2.5 \times 10^5$	2	250	250	500
12-well	$2-5 \times 10^4$	1	100	200	300
24-well	$2-5 \times 10^4$	0.5	100	200	300
HepG2 cell maintenance			HepG2 cell passaging		
Plate type	Seeding density (cells)	Medium volume (mL)	Trypsin volume (mL)	Volume to deactivate trypsin (mL)	Resuspended in for cell counting (mL)
T-25	1×10^6	5	0.35	0.65	1
T-175	4×10^6	50	5	20	5

2.1.3. Cell lines

Wild type D3 feeder-independent mESCs were derived from the ICM of pre-implanted blastocysts of the 129/Sv mouse strain (Doetschman *et al.*, 1985). HepG2 cells were derived as well-differentiated human hepatocarcinoma cells from the liver tissue of a 15-year-old Caucasian male (Knowles *et al.*, 1980).

A2 feeder-independent mESCs were obtained from Prof. Patrick Tam and Dr. Pierre Osteil (Children's Medical Research Institute, Westmead, NSW, Australia) and were used by Dr. Hannah Glover for RNA sequencing (RNAseq) (Section 2.4). A2 mESCs were generated from the E14Tg2a mouse ESCs (originally derived from 129/Ola mice) and have the XY karyotype (Wutz *et al.*, 2002; Iacovino *et al.*, 2011). This cell line has a doxycycline-inducible Cre-LoxP modification in the constitutively expressed *HPRT* locus. This allows for recombination after plasmid incorporation (Wutz *et al.*, 2002; Iacovino *et al.*, 2011). This cell line also has a tetracycline-inducible promoter in the *Rosa26* locus. This allows for inducible gene expression (Chu *et al.*, 2016).

A2-derived EpiSCs were also obtained from Prof. Patrick Tam and Dr Pierre Osteil and used by Dr. Hannah Glover for RNAseq (Section 2.4). These cells were derived by incorporating the A2 mESC line into the blastocyst, then implanted into a pseudopregnant mouse. After blastocysts were implanted, they were allowed to develop into embryos, then isolated and maintained as an EpiSC line. These were maintained on a feeder layer of mouse embryonic fibroblasts (MEFs) (Section 2.1.8). A2 EpiSCs have the same genetic background as A2 mESCs (Osteil *et al.*, 2016).

2.1.4. Thawing mESCs and HepG2 cells

mESCs and HepG2 cells were removed from liquid nitrogen storage and thawed gradually in a 37°C water bath. Cells were gently transferred and resuspended into mESC medium and HepG2 medium, respectively (Table 2.1 and 2.2). Cells were centrifuged for 2 min at 1200 rpm, the

supernatant removed, and cells gently resuspended in 5 mL mESC medium or HepG2, respectively (Table 2.1). mESCs were seeded directly onto a pre-gelatinised (0.1% porcine gelatin) 6 cm cell culture plate containing mESC medium (Table 2.2). HepG2 cells were seeded onto a T-25 or T-75 flask containing HepG2 medium (Table 2.2). Cells were incubated in a 5% CO₂, 21% O₂, 37°C humidified incubator. Cells were passaged for 2-3 consecutive days before experimental use or for maintenance as per Section 2.1.5.

2.1.5. Mouse ESCs and HepG2 cell maintenance

Cells were maintained by passaging when confluency reached 50-70% (Glover *et al.*, 2022). Briefly, medium was removed and cells were washed twice with PBS. Cells were trypsinised with addition of the appropriate volume of 0.25% trypsin/EDTA (Table 2.2), swirled for 20-30 s for mESCs or incubated for 5 min for HepG2 cells. The trypsin was deactivated by addition of mESC medium for mESCs or HepG2 culture medium for HepG2 cells (Table 2.2). Cells were triturated and single-cell suspended. The cell suspension was centrifuged at 14,000 rpm for 10 s for mESCs or 1200 rpm for 2 min for HepG2 cells. Cells were gently resuspended in sterile PBS. A cell count was performed using a haemocytometer; 10 µL of cell suspension was mixed with 40 or 90 µL 0.4% Trypan Blue. Dead cells, identified by uptake of Trypan Blue, were not counted as part of the live-cell count. The mESC suspension was replated at a cell density of $1-5 \times 10^5$ live cells in fresh mESC medium onto pre-gelatinised cell culture plates (Table 2.1 and 2.2). The HepG2 cell suspension was replated at a cell density of 10^6 live cells in HepG2 culture medium into T-25 flasks for maintenance and/or for the production of conditioned medium (Table 2.1 and 2.2 and Section 2.1.6). Cells were incubated in a 5% CO₂, 21% O₂, 37°C humidified incubator.

2.1.6. Conditioned medium from HepG2 cells (MEDII)

HepG2 cells were passaged as per section 2.1.5 and cells replated at 4×10^6 live cells into T-175 flasks. Cells were cultured for 4-5 days (Table 2.1) and the medium collected and sterile

filtered through a 0.22 µm filter, after which it was supplemented with 50% Basal medium, 0.1 mM β-mec and 1 mM L-Gln and stored at -20°C for up to 6 months, or 4°C for a month (Table 2.1). Cells were passaged and either maintained as per Section 2.1.5 or conditioned again.

2.1.7. EPLC formation and ground-state mouse ESC maintenance from mouse ESCs

EPLCs and ground-state mESCs were derived from both D3 and A2 mESC lines cultured as per Section 2.1.5. After live-cell counts, mESCs were replated into EPLC medium and 2i medium, respectively, in 6 cm plates (Table 2.1) (Ying *et al.*, 2008; Glover *et al.*, 2022). Cells were passaged every 2-3 days when confluency reached 50-70% in the same way as mESCs as per section 2.1.5.

2.1.8. Mouse embryonic fibroblast (MEF) derivation and culture and epiblast stem cell (EpiSC) culture

MEFs and their culture, freezing, and conditioning media (Table 2.1) were prepared by Dr. Hannah Glover in collaboration with A/Prof. Stuart Fraser (Bosch Institute, University of Sydney) (Glover, 2018). To derive MEFs, mice were sacrificed and the 12.5 dpc embryos removed from the uterine horns. The fetal liver and heads were removed, then embryos were washed with PBS. Embryos were manually dissected in 0.25% trypsin until they formed a suspension. Cells in this suspension were transferred into tissue culture flasks containing ~1 embryo per 10-15 cm² and cultured in a humidified incubator at 5% CO₂, 21% O₂ at 37°C. Once confluency reached 50-70%, medium was removed and washed twice with cold sterile PBS. 12.5 µL 0.5% trypsin/cm² was added to culture plates for 1 min before deactivating with an equal volume of MEF culture medium (Table 2.1). Cells were removed via trituration and resuspended into a sterile microfuge tube for cell counting with a haemocytometer as per Section 2.1.4. Cells were centrifuged for 15 s and were either frozen down in MEF freezing

medium (Table 2.1) at 3×10^6 cells/vial or replated at 10^4 cells/cm² in MEF culture medium in a humidified incubator at 5% CO₂, 21% O₂ at 37°C.

For use as a feeder layer, MEFs were thawed into T-150 flasks and cultured until 50-70% confluent. Cells were passaged the same way as deriving MEFs. $>125 \mu\text{L}/\text{cm}^2$ fresh MEF culture medium was added to pre-gelatinized (0.1% porcine gelatin) cell-culture plates. MEFs were replated at 10^4 cells/cm² and cultured in a humidified incubator at 5% CO₂, 21% O₂ at 37°C. Upon reaching 100% confluency, 10 $\mu\text{g}/\text{mL}$ mitomycin C was added to culture medium for 2.5 h to mitotically arrest MEFs. Medium was aspirated and cells were washed twice in MEF culture medium. MEFs were then available for use as a feeder layer for routine EpiSC culture and experiments.

For EpiSC culture, complete EpiSC medium was derived by growing inactivated MEFs in MEF culture medium for 24 h, then supplementing with 20 ng/mL Activin A and 10 ng/mL Fgf2 (Table 2.1). EpiSCs were seeded onto MEF feeder layer and cultured in a humidified incubator at 5% CO₂, 21% O₂ at 37 °C. EpiSC colonies were removed from the MEF feeder layer using $12.5 \mu\text{L}/\text{cm}^2$ of 1 mg/mL collagenase for 1 min and collected into a sterile 15 mL tube before adding an equal volume of TrypLE to dissociate the colonies into single cells. Cells were centrifuged for 20 s at 12,000 rpm, resuspended into 1 mL complete EpiSC medium, counted on the haemocytometer and seeded at 5×10^3 cells/cm² on a fresh MEF feeder layer in complete EpiSC medium. This medium was supplemented with 50 μM Y27632 in the first 24 h after passaging to promote cell survival.

2.1.9. Neural cell formation from mESCs

Embryoid bodies (EBs) were derived from mESCs in suspension either in hanging drops (Wang and Yang, 2008; Ohnuki and Kurosawa, 2013) or petri dishes (Shparberg *et al.*, 2019) for the first 3 days of culture. As hanging drops, mESCs were passaged and diluted to 5×10^4 cells/mL in Complete MEDII, Basal medium, or L-Proline-mediated differentiation medium (Table 2.1)

in parallel (Figure 2.1). Hanging drops of mESCs in either Basal medium, complete MEDII or L-Proline-mediated differentiation medium (Table 2.1) were produced by aliquoting 20 μ L (containing 1000 cells) onto the underside of 15 cm tissue culture plate lids. The lids were carefully and quickly inverted and placed onto 15 cm tissue culture plates (section 2.1.1) containing 10 mL PBS. For suspension in 10 cm petri dishes, mESCs were passaged and counted (Section 2.1.5); 5×10^5 cells per petri dish in either Basal medium, Complete MEDII or L-Proline-mediated differentiation medium (Table 2.1). Hanging drops and EBs in suspension culture were incubated at 5% CO₂, 21% O₂, 37°C for 3 days before continuing suspension culture in either Complete MEDII, Basal medium or L-Proline-mediated differentiation medium in 10 cm petri dishes for up to an additional 6 days in a 5% CO₂, 21% O₂, 37°C humidified incubator (Table 2.1) (Shparberg *et al.*, 2019). For EBs cultured in L-Proline-mediated protocol, medium was replaced with the appropriate L-Proline-mediated differentiation medium (Table 2.1) on Days 5 and 7 (Figure 2.1). Medium was replaced by swirling the EBs into the middle of the plate so old medium could be aspirated from the sides. At the end of Day 9, individual EBs were transferred to gelatinized (0.1% porcine gelatin) 48-well plates (2-3 EBs per well) for adherent culture in serum-free or serum-free + L-Proline medium (Table 2.1). EBs were incubated as above for a further 6 days. Medium was replaced on Day 12 and the experiment was terminated on Day 15 (Figure 2.1) (Shparberg *et al.*, 2019).

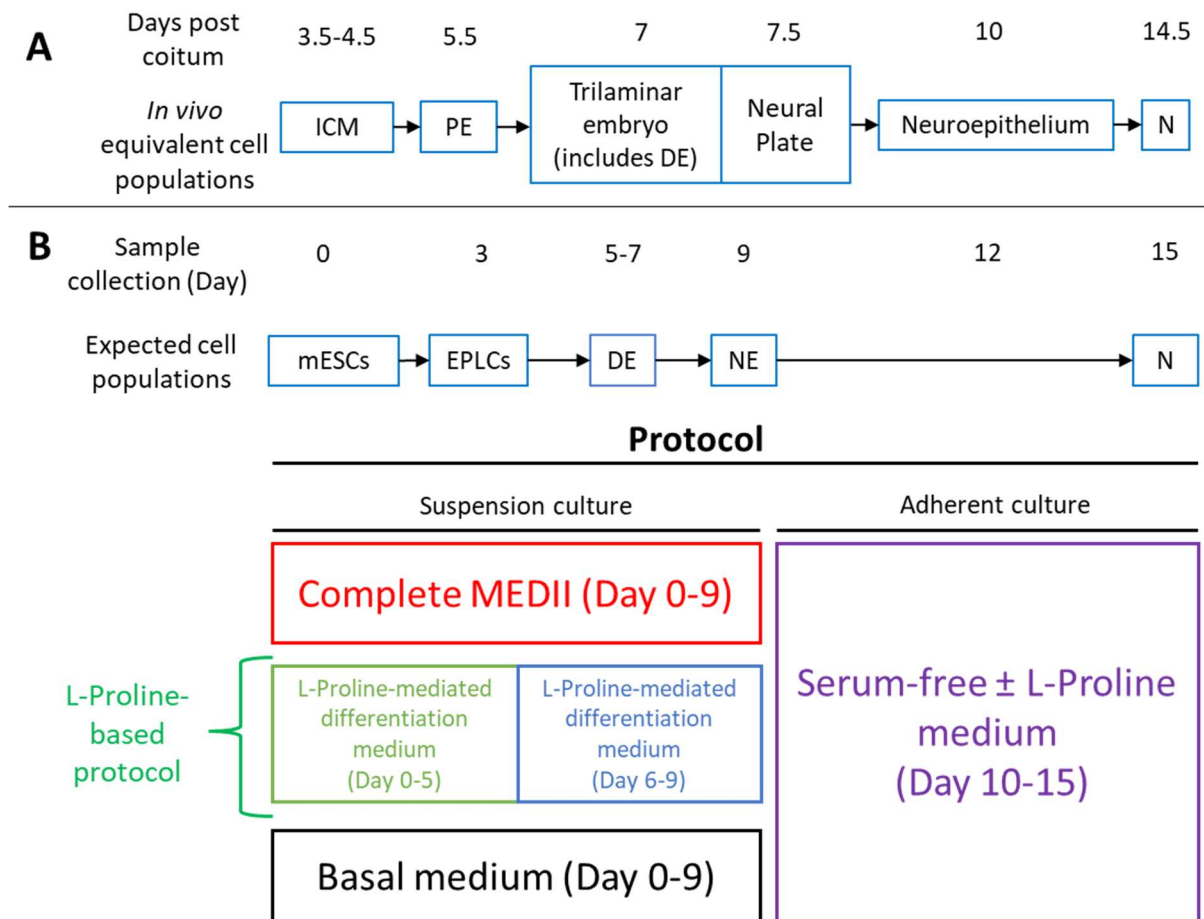


Figure 2.1 L-Proline-based protocol to model neural differentiation from mouse ESCs.

A. *In vivo*, the pluripotent inner cell mass cell population (ICM) is formed 3.5–4.5 days post coitum (dpc), which develops further into another pluripotent cell primitive ectoderm (PE) by 5.5 dpc. The PE is primed for gastrulation to form the trilaminar embryo which contains the 3 germ layers by 7 dpc. The DE develops further in response to form the neur ectoderm (NE), which acts as a precursor to form the neural plate by 7.5 dpc. The neural plate undergoes formation of the neural tube. The neuroepithelium in the neural tube at 10 dpc develops into different neural cell types (N) by 14.5 dpc. **B.** Development of neural cells from mouse ESCs occurs stepwise via multiple cell populations *in vitro*. Mouse ESCs differentiate to a second pluripotent cell population called early primitive ectoderm-like cells (EPLCs). Between Days 5–7, the DE-like population is formed transiently, before differentiating into NE by Day 9. To recapitulate these cell populations across Days 0–9, mESCs are cultured in suspension as embryoid bodies (EBs) for 9 days in one of (i) MEDII (conditioned medium from the hepatocarcinoma cell line HepG2), (ii) L-Proline-mediated differentiation medium for Days 0–5 before switching to L-Proline-mediated differentiation medium plus 10 μ M Nodal inhibitor SB431542 for Days 6–9, or (ii) basal medium (which allows for spontaneous differentiation) (see Table 2.1 for media formulations). After this, EBs are transferred to adherent culture to further differentiate to form neural cells (N) from Days 10–15 using serum-free medium (see Table 2.1 for media formulations). Figure adapted from published protocols (Shparberg *et al.*, 2019).

2.2. Seahorse extracellular flux (XF) assay

The Seahorse XF assay was used to measure the oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) simultaneously in real time as indicators of aerobic and anaerobic metabolism, respectively. Oxygen consumption results primarily from mitochondria utilising it in the final reaction of the electron transport chain to produce adenosine triphosphate (ATP). Alternatively, the production of lactic acid from glucose via glycolysis causes cells to release protons, which leads to a change in pH, resulting in an ECAR.

The XF assay sensor cartridge contains 4 injection ports. Each of these injection ports was used to inject specific compounds directly onto live cells. Raw measurements (oxygen concentration in pmol and changes in pH) were converted into OCR and ECAR, respectively, using the Wave Software (v 2.6.1).

The Seahorse XF assay was used to determine immediate changes in OCR and ECAR in ground-state mESCs, naïve-state mESCs, and EPLCs in response to injections of XF medium (control), 0.4 mM, 1 mM and 3 mM L-Proline and 0.37 μ M FCCP. FCCP uncouples oxidative phosphorylation and is used to determine the maximum respiratory capacity of cells (Nicholls *et al.*, 2010; Silverthorn *et al.*, 2010; Chaban *et al.*, 2014).

2.2.1. Hydrating the sensor cartridge

The XF analyser, controller and workstation was turned on at least 2 days prior to running the protocol to allow all items to warm up to 37°C. The sensor cartridge was hydrated as per manufacturer's instructions. Briefly, 24 h prior to running the assay, each well of the 96-well utility plate was filled with 200 μ L sterile MQ. The sensor cartridge was lowered to allow the sensors to be completely submerged before incubating overnight for at least 12 h at 37°C in a non-CO₂ humidified incubator. On the day of the assay, MQ was replaced with 200 μ L XF calibrant pre-warmed to 37°C in a non-CO₂ environment. The sensor cartridge was incubated

further for another 45-60 min at 37°C in a non-CO₂ environment before injection ports were loaded (Section 2.2.3).

2.2.2. Culturing and seeding cells

mESCs (ground state and naïve state) and EPLCs were cultured in pre-gelatinized (0.1% porcine gelatin) 6-well plates for 4 days before the assay was performed. 48 h prior to the assay, mESCs and EPLCs were trypsinized and seeded into pre-gelatinized (0.1% porcine gelatin) XF96 well microplates. 2×10^4 mESCs (ground state and naïve state) and 1.5×10^4 EPLCs were suspended in 80 µL medium into each well, with the lower seeding density for EPLCs than mESCs to account for faster doubling time. The 4 corner wells of the XF96 well microplates were filled with XF assay medium only as per manufacturer's instructions. The plate was left at room temperature for 1 h to allow cells to seed down evenly across the surface of the wells before incubating in a 37°C, 21% O₂, 5% CO₂ incubator for at least 48 h.

On the day of the assay, cells underwent a confluency check using the IncuCyte ZOOM before cell culture medium was removed, and the wells washed 2× with PBS pre-warmed to 37°C in a non-CO₂ environment before being replaced with unbuffered XF assay medium (Section 2.2.3) pre-warmed to 37°C in the non-CO₂ environment. Cells were incubated at 37°C in a non-CO₂ environment for 1-2 h before running the assay (Section 2.2.4).

2.2.3. Preparation of XF assay medium and injections

XF assay medium is unbuffered so as not to interfere with ECAR measurements. 24 h prior running the assay, the XF assay medium was supplemented with 1 mM sodium pyruvate, 2 mM L-Gln and 1 mM glucose. Injectable compounds containing L-Proline or FCCP were prepared with XF medium as the diluent before storing overnight at 4°C. On the day of the assay, supplemented XF assay medium and injectable compounds had their pH adjusted to 7.4 at 37°C before loading them into the appropriate ports, making sure all ports were filled for all wells.

2.2.4. Running the protocol and analysis

The sensor cartridge and utility plate containing XF calibrant (Section 2.7.1) was placed in the XF analyser and was left to calibrate for 1 h as per manufacturer's instructions. Cells in the XF96 well microplate (section 2.7.2) were rechecked for confluency using the IncuCyte ZOOM. Post-calibration, the utility plate was replaced with the XF96 well microplate (section 2.7.2) to run the assay.

Oxygen concentration and pH in each well was measured over 2 min and the values used to compute OCR and ECAR, respectively. These measurements were repeated at 7-minute intervals. 5 OCR and ECAR measurements were taken before the first port injection to establish baseline values. 5 more measurements were taken after each of the injections from the ports.

The details of the 7-minute assay cycle were as follows:

- i) Mixing (3 min) – the sensor cartridge was repeatedly raised and lowered in each well to mix. This evens out treatment, oxygen and proton concentrations in the medium.
- ii) Measuring (2 min) – the sensor cartridge is lowered 200 μm above the layer of cells, creating a microchamber of trapped medium. 12 measurements of oxygen concentration and pH changes are taken at this stage.
- iii) Resting (2 min) – the sensor cartridge is raised after 12 measurements, which allows the trapped media to mix with the rest of the medium and to re-equilibrate oxygen concentration and pH with the rest of the medium in the well.

Seahorse data was analysed on the Wave Software (v 2.6.1), which converts oxygen concentrations to OCR values and raw pH to ECAR values. Linear regression was used to convert the slope produced by the 12 points of measurement to the corresponding metabolic rate. Confluence (%) measured by the IncuCyte ZOOM was converted to decimal format and used to normalise OCR and ECAR values to a scale factor of 0.5. OCR and ECAR values were

excluded if the raw measurements were non-linear as were data from wells with <20% confluence.

2.3. Gene Expression

2.3.1. Materials

Molecular biology reagents and plasticware

- Nuclease-free H₂O [Baxter #AHF7114]
- TRIzol/Tri Reagent [Sigma: #T9424 or #93289]
- Glycogen [Sigma: #G1767]
- Chloroform [Sigma: #C2432]
- Isopropanol [Sigma: #I9516]
- Absolute Ethanol [Sigma #E7023]
- RQ1 RNase-Free DNase kit [Promega: #M6101], contains the following:
 - RNase-free DNase I [Promega: #M610A]
 - 10x Reaction Buffer [Promega: #M198A]
 - Stop solution [Promega: #M199A]
- High-Capacity cDNA Reverse Transcription Kit [Applied Biosystems: #4368813]
- 96-well MicroAmp Fast Optical PCR plate [Applied Biosystems: #4346906]
- 96-well MicroAmp Fast Optical PCR reaction plate seal [Applied Biosystems: #4311971]
- TaqMan™ Fast Advanced Master Mix [Applied Biosystems: #4444556]
- PCR tubes/Eppendorf tubes

Taqman™ gene expression assays of genes of interest are described in Table 2.3. The Taqman™ assay identification number used to amplify the splice variant was based on previously determined primers (Glover, 2018; Shparberg, 2018; Glover *et al.*, 2022).

Table 2.3. Taqman™ gene expression assays for qPCR of selected *Mus musculus* genes. All primer sets were pre-made and selected from the ThermoFisher database. 18s ribosomal RNA was used as the housekeeping gene. VIC: 2'-chloro-7'-phenyl-1,4-dichloro-6-carboxyfluorescein; FAM: Fluorescein amidite.

Gene	Assay identification	Assay Design	Amplicon length (bp)	Reporter	Catalogue #
<i>18s</i>	Hs03003631_g1	Non-Standard Annotation	69	VIC	4448489
<i>Brachyury/T</i>	Mm01318252_m1	Probe spans exons	57	FAM	4331182
<i>Dnmt3b</i>	Mm01246232_g1	Probe spans exons	108	FAM	4351372
<i>Fgf5</i>	Mm00438918_m1	Probe spans exons	67	FAM	4331182
<i>Fus</i>	Mm01271304_m1	Probe spans exons	95	FAM	4448892
<i>Mixl1</i>	Mm00489085_m1	Probe spans exons	74	FAM	4331182
<i>Nanog</i>	Mm02384862_g1	Probe spans exons	93	FAM	4331182
<i>Nestin</i>	Mm00450205_m1	Probe spans exons	72	FAM	4331182
<i>Oct4</i>	Mm00658129_gH	Probe spans exons	148	FAM	4331182
<i>Optn</i>	Mm01333241_m1	Probe spans exons	88	FAM	4448892
<i>Pard6b</i>	Mm00480520_m1	Probe spans exons	76	FAM	4331182
<i>Penk1</i>	Mm01212875_m1	Probe spans exons	57	FAM	4331182
<i>Rex1</i>	Mm03053975_g1	Both primers and probe map within a single exon	111	FAM	4331182
<i>Smn1</i>	Mm01282975_m1	Probe spans exons	58	FAM	4448892
<i>Sod1</i>	Mm01344233_g1	Probe spans exons	71	FAM	4453320
<i>Sox1</i>	Mm00486299_s1	Both primers and probe map within a single exon	71	FAM	4331182
<i>Stella</i>	Mm01184198_g1	Probe spans exons	89	FAM	4331182
<i>Tardbp</i>	Mm01257503_g1	Probe spans exons	108	FAM	4448892

2.3.2. RNA isolation and purification

RNA was isolated and purified using TRIzol-chloroform protocol with samples kept on ice (Joglekar *et al.*, 2010). Briefly, samples were collected in TRIzol (50 μ L/single EB or 150 μ L/10 EBs or 500 μ L/mESC cell suspension), flash frozen in liquid nitrogen and stored on ice (if used immediately) or at -80°C until required. Upon thawing from -80°C , samples were allowed to be left at room temperature to promote nucleoprotein complex dissociation. Glycogen was added to 1 μ g/ μ L per sample. Chloroform was then added to TRIzol (1:5), and the sample was shaken vigorously by hand for 1 min before being centrifuged at 4°C for 15 min at 12,000 g. The aqueous phase (containing RNA) of each sample was transferred to another tube and isopropanol was added and mixed by inverting 10 times. The samples were incubated at room temperature for 15 min before centrifuging at 4°C for 15 min at 12,000 g. Samples were then washed with freshly made 75% ethanol, vortexed briefly to mix before centrifuging as above. Supernatant was removed before air-drying at room temperature for <-5 min. Nuclease-free water (50 μ L/sample) was added, the RNA resuspended thoroughly, and incubated at $55-60^{\circ}\text{C}$ for 10 min to promote precipitation of nucleic acids before either being quantified for RNA (Section 2.3.3) or stored at -80°C until required.

2.3.3. RNA quantification and assessment

The concentration of purified RNA samples were measured using the NanoDrop spectrophotometer (ThermoFisher Scientific) using an RNA extinction coefficient of 40 μ g/mL.

The machine was blanked using nuclease-free water (Section 2.3.1). Absorbance was measured at 260 nm (A_{260}) and 280 nm (A_{280}) to detect nucleic acid and proteins, respectively. Organic compounds like ethanol and phenol have strong absorbance at 230 nm and 270 nm, respectively. The $A_{260}:A_{280}$ ratio was used to assess purity, with a ratio of 1.7-2.0 indicating little protein contamination.

2.3.4. DNase treatment

Up to 3 µg purified RNA (determined in Section 2.3.3) was taken into a new tube for DNA decontamination. Each sample was diluted with nuclease-free water (Section 2.3.1) up to 24 µL total volume. DNA decontamination was done using the RQ1 RNase-Free DNase kit (Promega) as per manufacturer's instructions. In addition to producing samples treated with DNase, samples were made without DNase treatment by adding an equivalent amount of RNase-free water (Section 2.3.1). Sample purity was compared to its replicate sample with no DNase treatment using the NanoDrop spectrophotometer (ThermoFisher Scientific). Samples with an $A_{260}:A_{280}$ ratio of 1.7-2.0 were used to produce cDNA (Section 2.3.5).

2.3.5. cDNA production

Up to 1 µg purified, DNase-treated RNA was reverse transcribed to produce cDNA for qPCR. Production of cDNA was done using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) following the manufacturer's instructions. In addition to producing samples containing reverse transcriptase (+RT), samples were produced whereby RT was replaced with an equivalent amount of nuclease-free water (−RT) (Section 2.3.1) to account for residual genomic DNA. All samples were placed in the Eppendorf Master Cycler and thermal cycling conditions were as follows:

- 25°C for 10 min (annealing)
- 37°C for 120 min (elongation)
- 85°C for 5 min (inactivation)
- 4°C hold (or store on ice)

cDNA samples were then diluted to 33.3 ng/µL based on the original RNA concentration obtained from NanoDrop then stored at −20°C for long term storage or 4°C for immediate qPCR (Section 2.3.6).

2.3.6. Quantitative real-time polymerase chain reaction (qPCR) using Taqman™ gene expression assay

cDNA samples (both +RT and –RT samples) were prepared for quality assessment and qPCR in a 5 µL reaction volume containing:

- 2.5 µL TaqMan Fast Advanced Master Mix
- 0.25 µL 20x gene expression assay primer of interest/probe mix (Table 2.3)
- 0.25 µL 20x gene expression assay *18s* primer/probe mix (Table 2.3)
- 1 µL nuclease-free water (Section 2.3.1)
- 1 µL cDNA template (33.3 ng total)

A –RT mix containing all –RT samples (final concentration of 33.3 ng/µL) was produced. A no-template control was also produced whereby an equivalent volume of nuclease-free water was used to replace cDNA.

The quality of each cDNA sample was assessed by the cycle threshold (Ct) for the reference gene *18s* ribosomal RNA. If the Ct value was above the expected range (9-10 cycles), new cDNA was produced and retested. qPCR for the different genes (Table 2.3) was performed in duplicate. Samples were loaded into 96-well FAST PCR reaction plate and sealed (Applied Biosystems).

The Applied Biosystems QuantStudio FAST ViiA7 Real-Time PCR system was set up to run the following program:

- 95°C for 20 s (Holding)

Then 40 cycles of:

- 95°C for 3 s (Denaturing)
- 60°C for 30 s (Annealing)

Raw sample Ct values for each gene of interest were normalised to raw *18s* Ct values by the following:

$$\Delta Ct_{(normalised\ gene\ of\ interest)} = Average\ Ct\ value_{(gene\ of\ interest)} - Raw\ Ct\ value_{(18s)}$$

A fold change of the condition of interest relative to the assigned control was calculated using the following equation:

$$\Delta\Delta Ct = \Delta Ct_{(normalised\ gene\ of\ interest\ condition)} - \Delta Ct_{(normalised\ gene\ of\ interest\ control)}$$

$$Fold\ change = 2^{-\Delta\Delta Ct}$$

2.4. RNA sequencing (RNAseq) analysis

2.4.1. Sample preparation

Samples were prepared, processed, quality checked and underwent cDNA production by Dr. Hannah Glover (Glover, 2018). Quality of total RNA was checked and quantified using the Agilent 2100 Bioanalyser by Dr Hannah Glover (Glover, 2018). An Agilent RNA 6000 Nanochip was prepared as per manufacturer's instructions. RNA samples and RNA ladder were loaded into the Nanochip and quantified using the Agilent 2100 Bioanalyser. Samples used for RNAseq had a mean ($\pm$ SEM) RNA Integrity Number (RIN) of 9.8 ± 0.04 (mean $\pm$ SEM, n=21).

2.4.2. Sequencing

Library construction and sequencing was sent to Macrogen (South Korea). Library construction was done using TruSeq RNA Sample Prep Kit v2 (Illumina #RS-122-2001). Samples were sequenced using paired-end 100 bp fragments using HiSeq 4000 (Illumina). Read depth ranged from ~37 to ~57 million reads per sample.

2.4.3. RNAseq bioinformatics pipeline

Pipelining was performed using Artemis' High Performance Computing (HPC) and Galaxy interface. Raw RNAseq files were checked for quality using FastQC (v 0.11.8) (Figure 2.2) and

visualised using MultiQC. Paired-end sequences were aligned to reference genome *Mus musculus* mm10 using HISAT2 (v 2.1.0) with default settings (Figure 2.2) (Kim *et al.*, 2019). Alignments were visualised using the Integrated Genomics Viewer (IGV) (Figure 2.2). HISAT2 outputs were counted for the number of reads using FeatureCounts from the Subread package (v 1.5.1) and the *Mus musculus* GRCm39 annotation GTF file from Gencode (Figure 2.2). Differential expression and principle component analysis (PCA) plots were done using the DESeq2 (v 1.33.4) pipeline with default settings on Rstudio (v 4.1.1) (Figure 2.2) (Anders and Huber, 2010; Love *et al.*, 2014). Raw counts were normalised using DESeq2's median of ratios to account for sequencing depth and RNA composition (Anders and Huber, 2010; Love *et al.*, 2014).

Published single-cell RNAseq data (accession number GEO: GSE100597) (Mohammed *et al.*, 2017) underwent PCA using the Scater library and single-cell consensus clustering (SC3) previously to identify different cell clusters (Kiselev *et al.*, 2017; Mohammed *et al.*, 2017). Differential expression analysis was done using the DESeq2 (v 1.33.4) pipeline, with the data fitted to a Gamma-Poisson distribution as the model accounts for sampling variability of single-cell RNAseq counts (Ahlmann-Eltze and Huber, 2020).

Raw counts were also normalised using fragments per kilobase of exon per million fragments mapped (FPKM) to account for sequencing depth and gene length using a published R package (Figure 2.2) (Trapnell *et al.*, 2010; Alhendi, 2019). These normalised counts were then used on Python (v 3.8.2) to run the Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE) pipeline to prioritize genes with cell-type specific regulatory functions (Figure 2.2) (Shim *et al.*, 2020). Repressing expression of genes that guide cell-fate decisions, and pluripotency and lineage commitment are marked by H3K27me3 (see Sections 1.4.1 and 1.4.2.1 and Figure 1.5). The TRIAGE pipeline utilises this epigenomic mark to transform normalised RNAseq counts to discordance scores (DS) (Figure 2.2). This score depends on 2 things: i) the extent of gene expression; ii) the degree of association of H3K27me3

marks (or the tendency of gene repression) observed across the genome (Shim *et al.*, 2020).

This is calculated as the following:

$$\begin{aligned} &Discordance\ score_{(gene\ of\ interest)} \\ &= Original\ expression_{(gene\ of\ interest)} \\ &\times Repressive\ Tendency\ Score_{(gene\ of\ interest)} \end{aligned}$$

The Repressive Tendency Scores derived from the *Mus musculus* genome were built into the Python scripts that were initially analysed from EpiMap data (Shim *et al.*, 2020). Genes were considered differentially expressed with a Benjamini-Hochberg adjusted $P < 0.05$, with different $\log_2$ fold change thresholds as described in the figure legends.

2.4.4. Biological interpretation of TRIAGE-transformed genes

The top 100 genes based on their ranked DS relative to the mouse genome were categorized using the gene clustering method TRIAGE-Parser on Python (v 3.8.2) (Figure 2.2) (Sun *et al.*, 2023). This method allows grouping or clustering of genes based on their “biological processes” gene ontology term. Identification of these clusters were used to predict protein-protein interactions (PPIs) (Figure 2.2) and mechanisms that determine cell fate and cell identity.

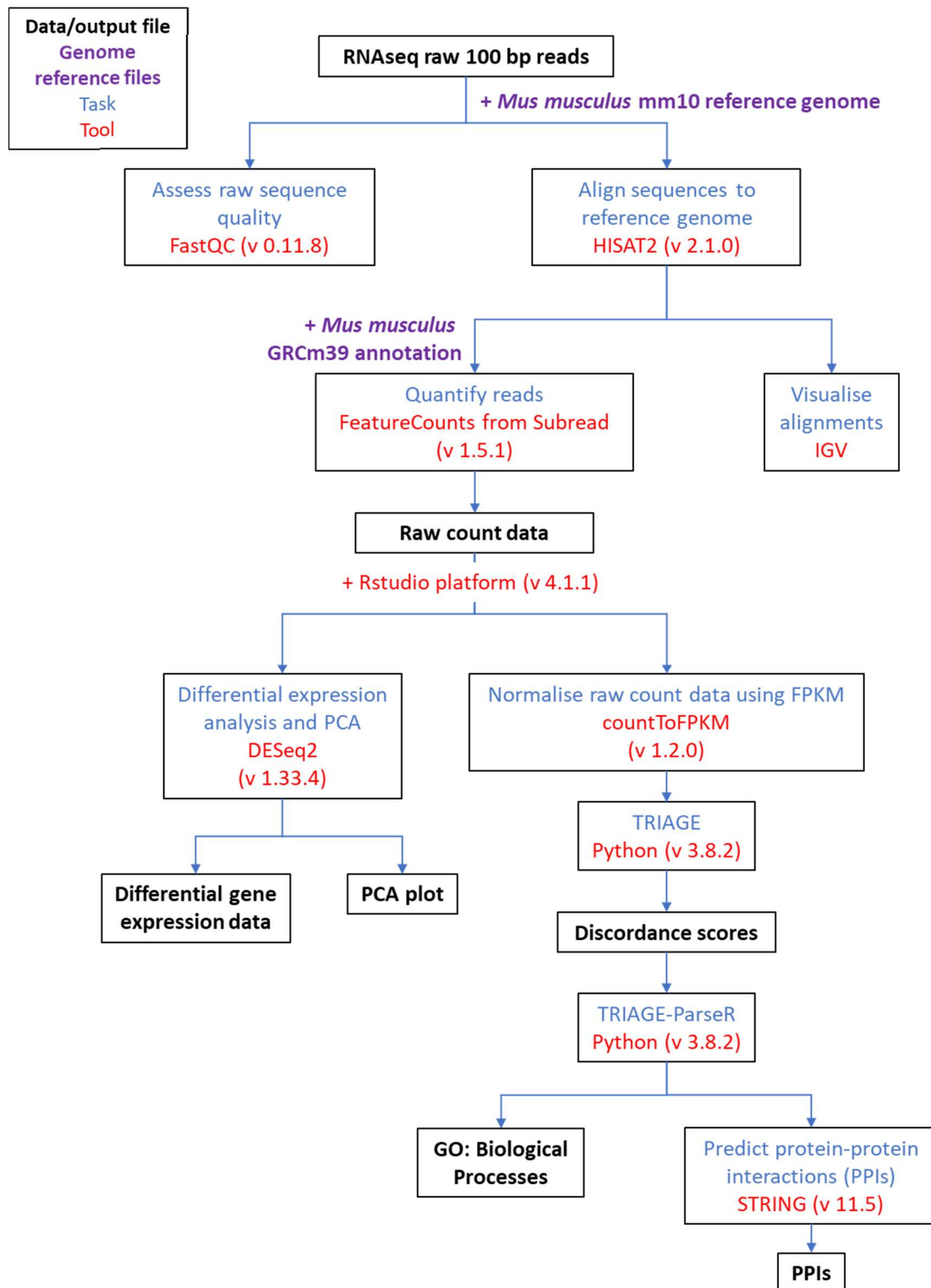


Figure 2.2 RNA sequencing (RNAseq) bioinformatics pipeline. Data files (both raw and output) are in black. RNAseq 100 bp reads were subjected to pipelining on Artemis' HPC and Galaxy interface. RNAseq 100 bp reads were aligned using the *Mus musculus* mm10 reference genome (purple). Each step of the pipeline is outlined (blue), with the annotation file acquired from Gencode (purple). Data processing tools used in each pipelining step is shown with their corresponding versions (red). Raw count data was subjected for further pipelining on Rstudio platform. Figure adapted from Glover, 2019.

2.5. Microarray analysis

cDNA samples were produced from EBs by Dr J. Rathjen (University of Adelaide) then hybridised onto the *Mus musculus* 22k Affymetrix platform. The samples were as follows (also see Figure 2.3):

D3 naïve-state mESCs were cultured in MEDII (Section 2.1.6 and Table 2.1) in suspension as EBs up to Day 3, 6 and 9 to produce EPLCs, definitive ectoderm and neurectoderm, respectively. EB samples were collected on those respective days. In a different set of experiments, D3 naïve-state mESCs were cultured in MEDII (Section 2.1.6 and Table 2.1) in suspension as EBs up to Day 3 to produce EPLCs, then left to spontaneously differentiate in basal medium (Table 2.1) up to Day 5 and 7, with EB samples collected on those respective days. Total RNA was harvested from these EB samples with Trizol (Section 2.3.1) and subjected to RNA extraction, purification and quantification procedures before producing cDNA for microarray generation.

Microarray data (Affymetrix 22k *Mus musculus* microarray) was generated from these samples across the 4 independent experiments and days, as per Affymetrix instructions (see also Figure 2.3). The data was normalised to take into account differences in cDNA concentration. Code was produced by Dr. J. Rathjen and Dr. C. Wilkinson (University of Adelaide) to work up the raw data into graphical form. Code was run on the R interface (v 2.2.1) and raw data was extracted using Rstudio (v 4.1.1). From these datasets, differential expression of genes was calculated and also subjected to TRIAGE transformation and analysis (Section 2.4.3 and 2.4.4).

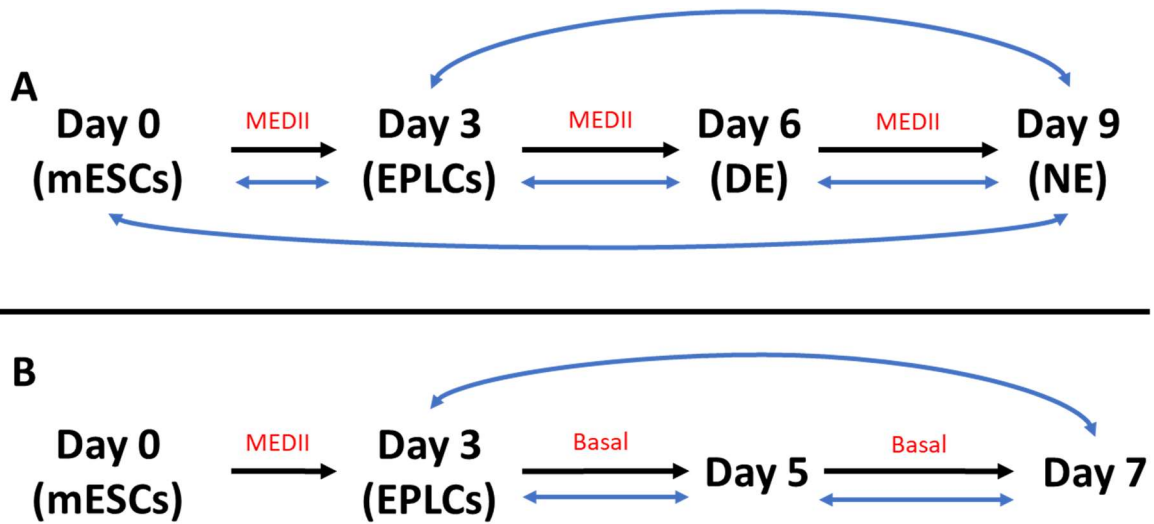


Figure 2.3 Hybridisation strategy used for microarray datasets. Microarray was performed on samples from 2 different but related experiments. For both **A** and **B**, D3 mouse embryonic stem cells (mESCs) at the naïve state (Day 0) represents the starting population maintained in mESC medium (see Table 2.1 for medium formulation). D3 naïve mESCs were cultured for 3 days in suspension using MEDII to produce early primitive ectoderm-like cells (EPLCs) (black arrows with condition in red). **A** Day 3 (EPLCs) were continued to be cultured in suspension to produce Day 6 (DE: definitive ectoderm) and Day 9 (NE: neur ectoderm) EBs in MEDII (black arrows with condition in red). All pairwise comparisons, except Day 6 EBs relative to Day 0, were performed (blue double arrows). **B** Day 3 (EPLCs) derived from Day 0 samples from a separate set of experiments were allowed to spontaneously differentiate to Day 5 and Day 7 in basal medium (see Table 2.1 for medium formulation) (black arrows with conditions in red). All pairwise comparisons, except Day 0 samples relative to Day 3, 5 and 7, were performed (blue double arrows). Both sets of hybridisations (blue double arrows from both **A** and **B**) were repeated 4 times each from different passage numbers (n=4 independent experiments).

2.6. Western blot analysis

2.6.1. Materials

- Acetic acid [Sigma #A6283]
- Acrylamide: bis-acrylamide 40% (w/v) (37.5:1) [Astral Scientific #BIOA0008]
- Ammonium persulphate (APS) 10% (w/v) [Chem-Supply #A2098]
- Bovine serum albumin (BSA) [Sigma #A9418]
- Bromophenol Blue [Sigma #B8026]
- Cell Lysis Buffer (10X) [Cell Signaling Technology #9803]
- Dithiothreitol (DTT) (1 mM working stock) [Life Technologies #P2325]
- Glycerol [Sigma #G5516]
- Glycine [Sigma #G7126]
- Micro BCA™ Protein Assay Kit [Thermofisher Scientific #23235]
- Mini-PROTEAN® TGX™ Precast Gels (4-20%) [BioRad #456-1093]
- Molecular weight markers [Bio-Rad #161-0374]
- Nitrocellulose membrane, 0.45 µm [GE Healthcare #10600002]
- Odyssey Blocking Buffer [LI-COR #927-50000]
- PBS [Gibco: #21600-010 or Amresco: #E404]
- Phenylmethanesulfonyl fluoride (PMSF) (100 mM in ethanol/isopropanol) [Sigma #P7626]
- Ponceau S (0.1% (w/v) in 5% (v/v) acetic acid) [Sigma #3504]
- Precision Plus Protein Dual Colour Standards [Bio-Rad #161-0374]
- Sodium dodecyl sulfate (SDS) (10% (w/v)) [Amresco #0227]
- Sodium chloride [Sigma #S5886]
- Tetramethylethylenediamine (TEMED) [Bio-Rad #1610801]
- Tris Base [Amresco #0497]
- Tween-20 [Astral Scientific #BIO0777]
- Whatman cellulose chromatography paper (3MM) [GE Healthcare #3030917]

Primary and secondary antibodies are listed in Tables 2.4 and 2.5, respectively. The percentage of polyacrylamide use in gels depended on the range of sizes of the proteins of interest. Gel formulations are shown in Table 2.6 and those for Western blot buffers and solutions in Table 2.7.

Table 2.4 Primary antibodies used for Western blot analysis.

Primary Antibody	Species and specificity	Dilution	Size (kDa)	Company	Catalogue #
Fus	C-term Rabbit polyclonal	1:10,000	73	ThermoFisher (Bethyl)	A300-294A
Optn	C-term Rabbit polyclonal	1:1000	70-74	Cayman Chemical	100000
Smn1	Mouse monoclonal	1:5000	35	BD Biosciences	610647
Sod1	Goat polyclonal	1:1000	19	R&D systems	AF3787
Tardbp	Rabbit polyclonal	1:1000	43	ProteinTech	12892-1-AP
Vapb	Rabbit polyclonal	1:1000	27-30	ProteinTech	14477-1-AP
α -tubulin	Mouse monoclonal	1:5000	50-52	Developmental Studies Hybridoma Bank	E7
β -actin	Mouse monoclonal	1:5,000-1:10,000	42	Sigma	a5441

Table 2.5 Secondary antibodies used for Western blot analysis.

Fluorophore	Species	Dilution	Company	Catalogue #
IRDye 680	Donkey anti-mouse IgG	1:5000	LI-COR®	926-32213
IRDye 800	Donkey anti-rabbit IgG	1:5000	LI-COR®	926-68022
IRDye 680	Donkey anti-goat IgG	1:20,000	LI-COR®	925-68024

2.6.2. Gel preparation

Table 2.6 Separating gel components for different acrylamide/bis-acrylamide % (v/v).

Reagent	8% (v/v) acrylamide/bis-acrylamide	10% (v/v) acrylamide/bis-acrylamide	12% (v/v) acrylamide/bis-acrylamide
40% (w/v) acrylamide/bis-acrylamide	2.0 mL	2.5 mL	3.0 mL
1.5 M TrisCl, pH 8.8	2.5 mL	2.5 mL	2.5 mL
MQ	5.29 mL	4.79 mL	4.29 mL
10% (w/v) SDS	100 μ L	100 μ L	100 μ L
10% (w/v) APS	100 μ L	100 μ L	100 μ L
TEMED	10 μ L	10 μ L	10 μ L
Protein size detection	50-100 kDa	20-45 kDa	< 20 kDa

Table 2.7 Western blot buffers and solutions and their components. DDT, dithiothreitol; SDS, sodium dodecyl sulfate; TEMED, tetramethylethylenediamine; TBS, Tris-buffered saline; TBST, 0.1% (v/v) Tween-20 in Tris-buffered saline.

Stacking Gel	MQ, 4% (v/v) acrylamide/bis-acrylamide, 125 mM TrisCl (pH 6.8), 0.1% (w/v) SDS, 0.05% (v/v) APS, 0.1% (v/v) TEMED
Running Buffer (5x)	MQ, 125 mM Tris Base (pH 8.3), 960 mM Glycine, 0.5% (w/v) SDS
Transfer Buffer (1x)	MQ, 20% (v/v) Ethanol, 25 mM Tris Base, 192 mM Glycine
TBS	10 mM TrisCl pH 7.4, 150 mM NaCl
TBST	10 mM TrisCl pH 7.4, 150 mM NaCl, 0.1% of 10% (v/v) Tween 20
Loading Buffer (6x)	65.8 mM TrisCl (pH 6.8), 2.1% SDS, 26.3% (w/v) glycerol, 0.01% Bromophenol Blue, 1 mM DTT (added as fresh liquid)

2.6.3. Cell lysis for protein quantification and Western blot

Medium was aspirated from cultured cells followed by 2x cold PBS washes. Cell Lysis Buffer (Section 2.6.1) was added to a concentration of 1 μ L/40 000 cells, immediately followed by addition of 1 μ L PMSF (100 mM)/100 μ L Cell Lysis Buffer. Cells were scraped from the plate before transferring them to pre-chilled microfuge tubes. Cells were spun at 12 000 rpm for 10 min at 4°C. Supernatant was transferred to new pre-chilled microfuge tubes and stored at -80°C until required.

For suspension culture, 50-100 EBs were collected into pre-chilled microfuge tubes, washed 2x with cold PBS before adding 100 μ L Cell Lysis Buffer (Section 2.6.1) per 100 EBs, immediately followed by 1 μ L PMSF (100 mM)/100 μ L Cell Lysis Buffer. Cells were spun down at 12 000 rpm for 10 min at 4°C. Supernatant was transferred to new, pre-chilled microfuge tubes and stored on ice used to produce dilutions in MQ (1 in 100, 1 in 200 and 1 in 500). The original concentration of supernatant was stored at -80°C until required. The diluted samples were stored on ice to be quantified immediately using the Micro BCA™ Protein Assay kit, as per manufacturer's instructions, to determine the volume of the original concentration required for 30 μ g protein.

2.6.4. Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis was conducted to separate proteins based on their molecular weight. Separating gel (Table 2.6) was poured between 2 glass plates spaced 1.5 mm apart and sealed with Parafilm at the base. MQ was gently overlayed using a P1000 pipette and the polyacrylamide allowed to polymerise for 15 min. Excess MQ was poured off and stacking gel (Table 2.7) was poured onto the separating gel. A 10-well comb was inserted between the plates before allowing the stacking gel to polymerise for 15 min. Once polymerised, the comb was removed, and wells were washed out with 1x running buffer in MQ. 5 µL Bio-Rad molecular weight markers was loaded into the first well and last well. Protein samples (5 µL for adherent culture samples) or 30 µg protein from suspension culture (as determined by Micro BCA™ Protein Assay) were mixed into 6x loading buffer containing DDT (Table 2.7) before being loaded into the wells. Electrophoresis was carried out using the BioRad Mini Power Pack at 100 V, 0.15 Amps, followed by 120 V after proteins had passed halfway through the gel for a total of 120 min.

2.6.5. Protein transfer and quality check

After SDS-PAGE, the proteins were transferred onto a nitrocellulose membrane (pore size 0.45 µm) to be stained. The transfer sandwich consisted of the following layers: 1 fibre pad, 2 sheets Whatman chromatography paper, the SDS-PAGE gel, nitrocellulose membrane, 2 sheets oWhatman chromatography paper and 1 fibre pad. Bubbles were rolled out of the sandwich 1 layer at a time using a plastic roller. The transfer sandwich was put into a chamber filled with 1x transfer buffer (Section 2.6.2) and cooled with ice. Transfer was conducted at 100 V using a Biorad Mini Power Pack for 1.5-3 h depending on size of the protein(s) of interest.

After protein transfer, the membrane was removed from the apparatus and stained with 0.1% (w/v) Ponceau S with 5% (v/v) acetic acid for 1-5 min. Protein bands were revealed by rinsing with TBS and to check transfer quality.

2.6.6. Membrane blocking and antibody staining

Membrane was blocked using TBS + 3% (w/v) skim milk powder and incubated for 2 h at room temperature or overnight on a rocker at 4°C. Blocking solution was then poured off and the membrane was washed 3X with TBS while leaving the last wash on the membrane.

For primary antibody staining, TBS was replaced with a fresh aliquot containing primary antibody at the required dilution in Odyssey blocking buffer + TBST (1:1), covered in foil and incubated for 2 h or overnight on a rocker at 4°C. Primary antibody was removed and the membrane washed 3X with TBST while keeping the last wash on the membrane. The membrane was covered with foil and placed on a rocker at room temperature for 30 min. TBST was removed and the appropriate fluorophore-conjugated secondary antibody was then added at the required dilution in fresh Odyssey blocking buffer + TBST (1:1) and incubated for 2 h at room temperature on a rocker. Secondary antibody was removed, the membrane washed 3X with TBST while leaving the last wash on the membrane, covered with foil and placed on a rocker for 30 min at room temperature. TBST was removed then the membrane was washed 2X with MQ to decrease background for imaging on the LI-COR® Odyssey infrared scanner and quantification of band intensities using the Odyssey Software (v3).

2.7. Microscopy

2.7.1. Phase contrast microscopy

Images of EBs in hanging drops were taken on Days 0 and 3 using the Axiocam 105 camera on a Zeiss Stemi 305 stereomicroscope, under 10x, 20x or 30x magnification. For Days 5, 7, 9, 10, 12 and 15, EBs were imaged with phase-contrast microscopy using 10x, 20x or 40x objectives on an Olympus IX-81 Inverted Fluorescence Microscope or the EVOS FL Auto Cell Imaging System (ThermoFisher Scientific). Adherent cultures were imaged on the culture plate using 10x, 20x or 40x objectives on the same microscopes.

2.7.2. Time-lapse microscopy with IncuCyte ZOOM

D3 mESCs were seeded at 18 000 cells/well onto pre-gelatinized (0.1% porcine gelatin) 12-well plates. Cells were incubated at 37°C, 5% CO₂, 21% O₂ for 96 h and imaged every 2 h with the IncuCyte ZOOM (Essen BioScience). The application was programmed to take 9 images per well and quantified confluency per image. Each image was assigned a random file name for unbiased scoring (Section 2.7.4).

2.7.3. Colony morphology scoring

Colonies were scored based on a pre-determined scale of differentiation (Figure 2.4): Colonies were scored as mESCs (1), morphology intermediate between ESCs and EPLCs (2), and morphology consistent with being EPLCs (3).

For scoring neural features on seeded down EBs at day 15, the presence of neurites, axonal bundles, neural progenitors and neural rosettes was assessed (Figure 2.5) (Shparberg *et al.*, 2019). Other features associated with mesendoderm lineage commitment, including beating cardiomyocytes, were also identified.

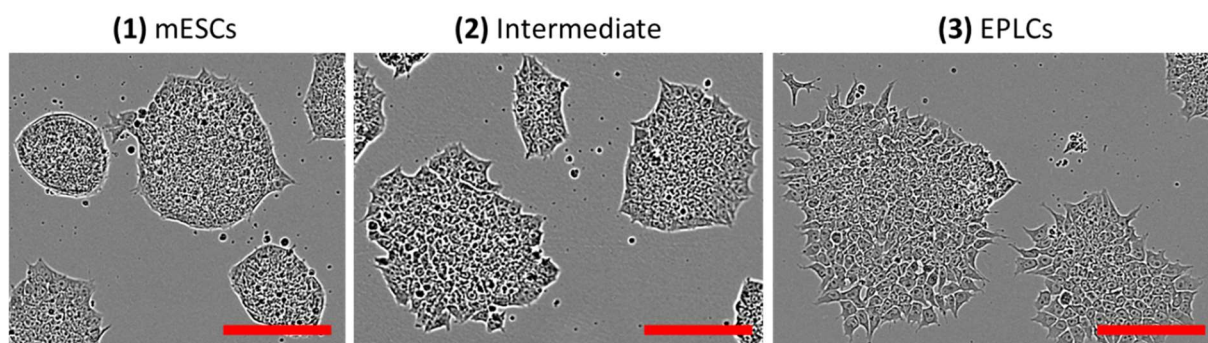


Figure 2.4 Representative reference images for scoring colony morphology. Colonies were scored as **(1)** mESCs, **(2)** intermediate between mESCs and EPLCs, or **(3)** EPLCs. Undifferentiated colonies were scored as having round, domed colonies and where individual cells could not be distinguished, consistent with them being mouse ESCs (mESCs). Colony morphologies intermediate between mESCs and EPLCs were scored based on having irregular borders and/or mostly visible cells within the colony. Colonies were also scored displaying irregular/jagged borders with extensions, visible cells breaking out of the colony and displaying a monolayer-like structure, consistent with them being EPLCs. Images were taken on IncuCyte ZOOM at 10x magnification. Scale bar = 150 μm .

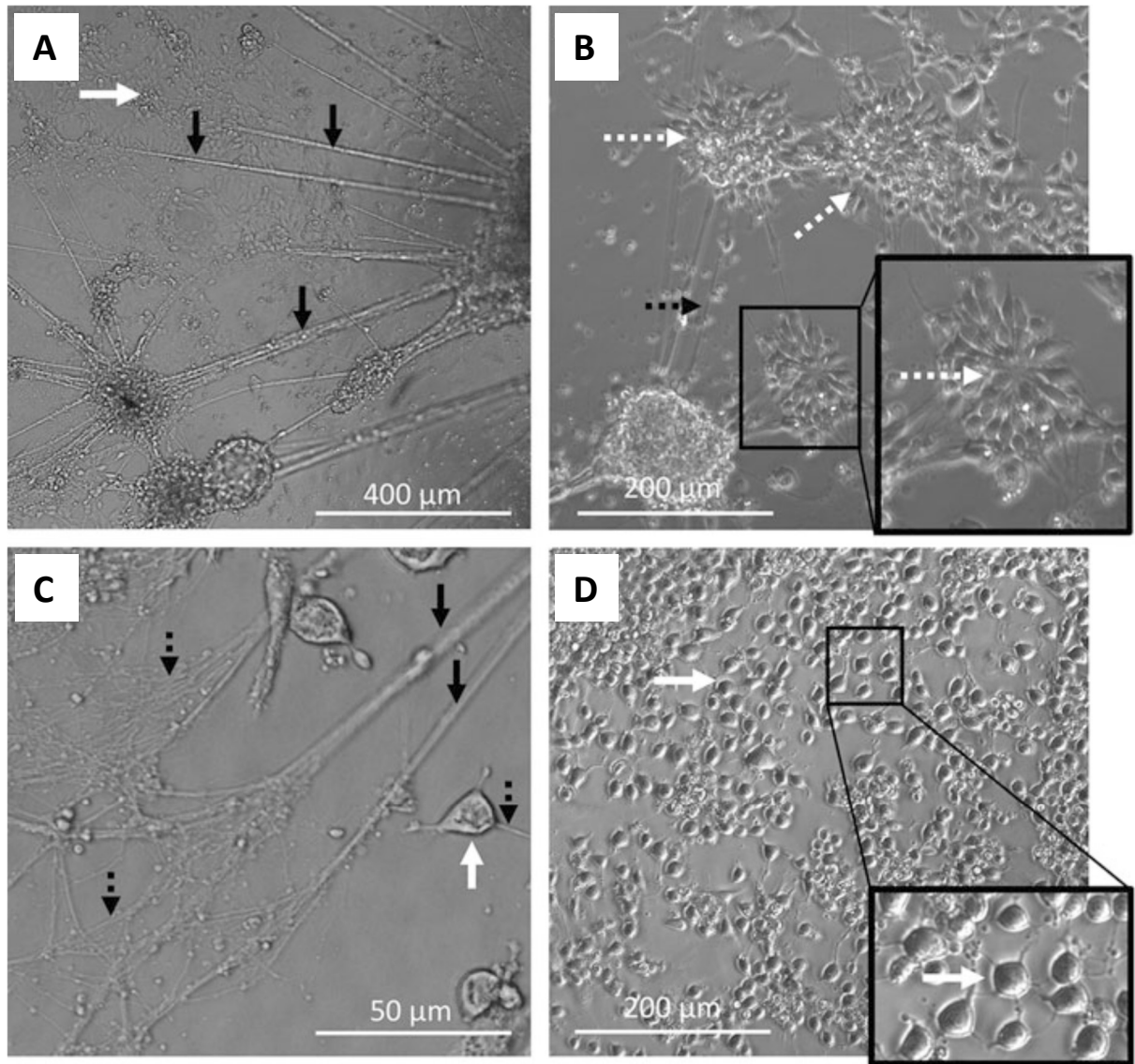


Figure 2.5 Representative images of neural features used for phenotypic scoring of EBs in adherent culture at Day 15. Neural features included the following: branched/unbranched neurites (broken black arrows); axonal projections/bundles (solid black arrows); neural rosettes (broken white arrows); and neural progenitors (solid white arrows). Original images were taken by Dr. R. Shparberg using phase contrast with a 10X objective. Scale bars in **A-D** are as shown. Figure adapted from Shparberg *et. al.* 2019a.

2.7.4. Doubling time calculations

Percentage confluence over time was calculated using the IncuCyte ZOOM software and converted into doubling time by fitting the data with the exponential growth function in GraphPad Prism v9:

$$Y = Y_0 * \exp (k * X)$$

where:

- Y_0 is the Y intercept, expressed in the same units as Y (% confluence)
- k is the rate constant, expressed as the reciprocal of the X axis unit (time in h)

From the rate constant, doubling time (DT) was determined by the following formula:

$$DT = \frac{\ln (2)}{k}$$

2.8. Statistical analyses

For routine statistical analysis and *post hoc* tests, GraphPad Prism v9 was used. For RNAseq and microarray analyses, DESeq2's Wald test was used for hypothesis testing when comparing between 2 groups, with the null hypothesis being no difference in log₂ fold change. *P* values obtained from DESeq2 and microarray analyses were then adjusted for multiple testing using the Benjamini-Hochberg method. These adjusted *P* values were used to determine significant differences in relative gene expression while accounting for false discovery rates (Anders and Huber, 2010; Koch *et al.*, 2018). *P* values acquired from the TRIAGE pipeline is a combined *P* value, which was calculated by using Fisher's method (based on Chi-square distribution) from 2 separate statistical tests: 1) the probability of observing a repressive tendency score of a given gene (calculated from discordance scores) by chance; 2) the probability of observing an expression value of the same gene by chance (Shim *et al.*, 2020). Dunnett's, Tukey's, and Sidak's *post hoc* tests were used as described in the figure legends. Heatmaps were produced on GraphPad Prism v9 and edited on Adobe Illustrator.

2.9. References

- Ahlmann-Eltze, C. and Huber, W. (2020) ‘glmGamPoi: Fitting Gamma-Poisson generalized linear models on single cell count data’, *Bioinformatics*, 36(24), pp. 5701–5702. doi: 10.1093/bioinformatics/btaa1009.
- Alhendi, A. S. N. (2019) ‘countToFPKM: Convert Counts to Fragments per Kilobase of Transcript per Million (FPKM)’. Available at: <https://cran.r-project.org/package=countToFPKM>.
- Anders, S. and Huber, W. (2010) ‘Differential expression analysis for sequence count data’, *Genome Biology*, 11(10), p. R106. doi: 10.1186/gb-2010-11-10-r106.
- Chaban, Y., Boekema, E. J., and Dudkina, N. V (2014) ‘Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation’, *Biochimica et Biophysica Acta - Bioenergetics*, 1837(4), pp. 418–426.
- Chu, V. T., Weber, T., Graf, R., Sommermann, T., Petsch, K., Sack, U., Volchkov, P., Rajewsky, K., and Kühn, R. (2016) ‘Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes’, *BMC Biotechnology*, 16(1), p. 4. doi: 10.1186/s12896-016-0234-4.
- Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, R. (1985) ‘The in vitro development of blastocyst derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium’, *Journal of Embryology and Experimental Morphology*, 87, pp. 27–45.
- Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.
- Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) ‘L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell’, in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.
- Iacovino, M., Bosnakovski, D., Fey, H., Rux, D., Bajwa, G., Mahen, E., Mitanoska, A., Xu, Z., and Kyba, M. (2011) ‘Inducible cassette exchange: A rapid and efficient system enabling conditional gene expression in embryonic stem and primary cells’, *Stem Cells*, 29(10), pp. 1580–1588. doi: 10.1002/stem.715.
- Joglekar, M. V., Wei, C., and Hardikar, A. A. (2010) ‘Quantitative Estimation of Multiple miRNAs and mRNAs from a Single Cell’, *Cold Spring Harbor Protocols*, 2010(8), p. pdb.prot5478. doi: 10.1101/pdb.prot5478.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019) ‘Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype’, *Nature Biotechnology*, 37(8), pp. 907–915. doi: 10.1038/s41587-019-0201-4.

Kiselev, V. Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., Natarajan, K. N., Reik, W., Barahona, M., Green, A. R., and Hemberg, M. (2017) 'SC3: consensus clustering of single-cell RNA-seq data', *Nature Methods*, 14(5), pp. 483–486. doi: 10.1038/nmeth.4236.

Knowles, B. B., Howe, C. C., and Aden, D. P. (1980) 'Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen', *Science*, 209(4455), pp. 497 LP – 499. doi: 10.1126/science.6248960.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., and Winter, D. R. (2018) 'A beginner's guide to analysis of RNA sequencing data', *American Journal of Respiratory Cell and Molecular Biology*, 59(2), pp. 145–157. doi: 10.1165/rcmb.2017-0430TR.

Love, M. I., Huber, W., and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), p. 550. doi: 10.1186/s13059-014-0550-8.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., and Reik, W. (2017) 'Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation', *Cell Reports*, 20(5), pp. 1215–1228. doi: 10.1016/j.celrep.2017.07.009.

Nicholls, D. G., Darley-USmar, V. M., Wu, M., Jensen, P. B., Rogers, G. W., and Ferrick, D. A. (2010) 'Bioenergetic profile experiment using C2C12 myoblast cells', *Journal of Visualized Experiments: JoVE*, (46), pp. 3–7. doi: 10.3791/2511.

Ohnuki, Y. and Kurosawa, H. (2013) 'Effects of hanging drop culture conditions on embryoid body formation and neuronal cell differentiation using mouse embryonic stem cells: Optimization of culture conditions for the formation of well-controlled embryoid bodies', *Journal of Bioscience and Bioengineering*, 115(5), pp. 571–574. doi: 10.1016/j.jbiosc.2012.11.016.

Osteil, P., Studdert, J., Wilkie, E., Fossat, N., and Tam, P. P. L. (2016) 'Generation of genome-edited mouse epiblast stem cells via a detour through ES cell-chimeras', *Differentiation*. Elsevier, 91(4–5), pp. 119–125. doi: 10.1016/j.diff.2015.10.004.

Shim, W. J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., Peng, G., Jing, N., Wang, Y., Kellis, M., Tam, P. P. L., Smith, A., Piper, M., Christiaen, L., Nguyen, Q., Bodén, M., and Palpant, N. J. (2020) 'Conserved epigenetic regulatory logic infers genes governing cell identity', *Cell Systems*, 11(6), pp. 625–639. doi: <https://doi.org/10.1016/j.cels.2020.11.001>.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019) 'Embryoid body differentiation of

mouse embryonic stem cells into neurectoderm and neural progenitors’, in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Silverthorn, D. U., Johnson, B. R., Ober, W. C., Garrison, C. W., and Silverthorn, A. C. (2010) *Human physiology: an integrated approach*. 5th edn. Edited by D. Espinoza et al. San Francisco, USA: Pearson Education Inc.

Sun, Y., Shim, W. J., Shen, S., Sinniah, E., Pham, D., Su, Z., Mizikovsky, D., White, M. D., Ho, J. W. K., Nguyen, Q., Bodén, M., and Palpant, N. J. (2023) ‘Inferring cell diversity in single cell data using consortium-scale epigenetic data as a biological anchor for cell identity’, *Nucleic Acids Research*, p. gkad307. doi: 10.1093/nar/gkad307.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) ‘Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation’, *Nature Biotechnology*, 28(5), pp. 511–515. doi: 10.1038/nbt.1621.

Wang, X. and Yang, P. (2008) ‘In vitro Differentiation of Mouse Embryonic Stem (mES) Cells Using the Hanging Drop Method’, *JoVE*. MyJoVE Corp, (17), p. e825. doi: doi:10.3791/825.

Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002) ‘Chromosomal silencing and localization are mediated by different domains of Xist RNA’, *Nature Genetics*, 30(2), pp. 167–174. doi: 10.1038/ng820.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008) ‘The ground state of embryonic stem cell self-renewal’, *Nature*, 453(7194), pp. 519–523. doi: 10.1038/nature06968.

Chapter 3: Investigating cell-fate drivers and L-Proline-mediated metabolic processes across the pluripotency spectrum

3.1. Introduction

Amino acids stimulate cell signalling and regulate gene expression to facilitate and drive embryogenesis and differentiation (Wu, 2009; Washington *et al.*, 2010; Shparberg *et al.*, 2019b). Embryogenesis requires pluripotent ICM cells to transition into another pluripotent cell population called primitive ectoderm, the cellular substrate for gastrulation and the generation of the 3 multipotent germ layers which go on to form all cell types of the embryo proper (Rathjen *et al.*, 1999; Pelton *et al.*, 2002; Nichols and Smith, 2009; Krupinski *et al.*, 2011). Culturing mESCs, the *in vitro* equivalent of the ICM, with L-Proline causes differentiation to EPLCs, the *in vitro* equivalent of primitive ectoderm (Rathjen *et al.*, 2003; Washington *et al.*, 2010).

L-Proline acts as a growth factor to activate cell signalling pathways such as MAPK/ERK, P38 and amino acid-sensing mammalian target of rapamycin (mTOR) to induce the mESC-to-EPLC transition (Lonic, 2007; Washington *et al.*, 2010; Glover, 2018). In addition, metabolism of L-Proline contributes to stimulating the transition (Casalino *et al.*, 2011; Comes *et al.*, 2013; D’Aniello *et al.*, 2015, 2017). In particular, metabolism of L-Proline stimulates the phenotypic transition of mESCs towards a mesenchymal-like state, which is reversible upon removal of L-Proline or addition of antioxidants (Casalino *et al.*, 2011; Comes *et al.*, 2013; D’Aniello *et al.*, 2017).

In mESCs, L-Proline biosynthesis is regulated by the AAR pathway, which suppresses production of L-Proline such that mESC identity and self-renewal is maintained (D’Aniello *et al.*, 2015). Upon exogenous addition of L-Proline, the AAR pathway is overwhelmed, which leads to the loss of controlled L-Proline production (D’Aniello *et al.*, 2015) and promotes mESCs to transition to EPLCs (Comes *et al.*, 2013; D’Aniello *et al.*, 2015, 2017) and their further directed differentiation to neural cell fate (Rathjen *et al.*, 2002, 2003; Maye *et al.*, 2004; Shparberg *et al.*, 2019a).

L-Proline enters cells via the sodium-coupled neutral amino acid transporter 2 (SNAT2) (Figure 3.1) (Tan *et al.*, 2011). Once in the cytoplasm, it is metabolised in the proline cycle to pyrroline-5-carboxylate (P5C) via proline oxidase (POX) (Figure 3.1) (Phang *et al.*, 2010). The proline cycle is redox coupled to the pentose phosphate pathway (PPP) (Figure 3.1), which stimulates glycolysis in highly proliferative cells like cancer cells and ESCs (Phang *et al.*, 2010; Hancock *et al.*, 2016). ESCs prefer anaerobic metabolism to support their growth and survival in comparison to somatic mammalian cells which generally prefer to generate ATP by aerobic metabolism (Rehman, 2010; Varum *et al.*, 2011; Panopoulos *et al.*, 2012). Similar to ESCs, when the blastocyst, with its ICM, implants into the uterine epithelium, glucose becomes the major energy source for the embryo and is preferentially metabolised anaerobically to generate ATP (Folmes *et al.*, 2012; Lees *et al.*, 2017).

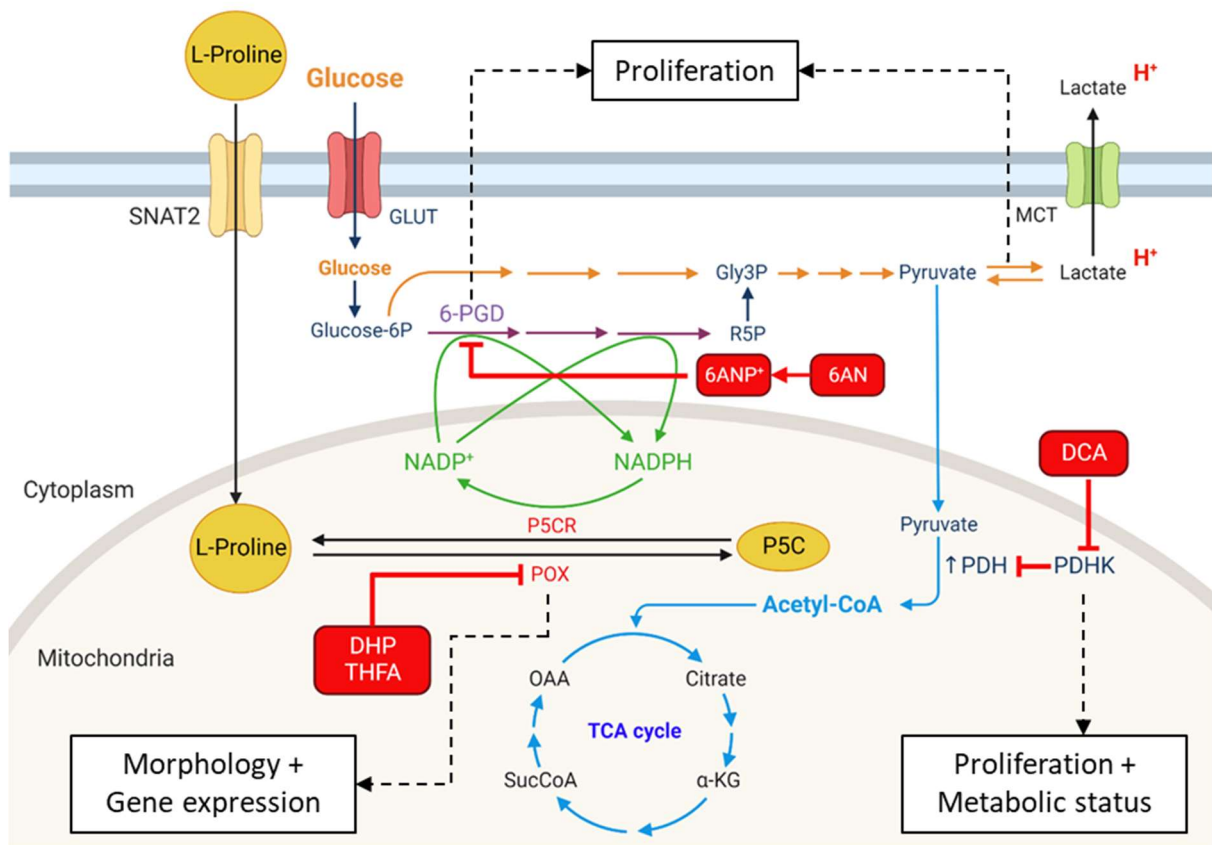


Figure 3.1 Metabolism in ESCs, including proposed emergent properties from pathways active during ES-to-EPLC transition, and the role of metabolic inhibitors dichloroacetate (DCA), L-tetrahydro-2-furoic acid (THFA), 3,4 dehydro-L-Proline (DHP) and 6-aminonicotinamide (6AN). Glucose is transported into ESCs via GLUT transporters and converted into glucose-6-phosphate (Glucose-6P), then to pyruvate through intermediate steps (orange arrows). Pyruvate is used to generate energy in the form of ATP by being converted to lactate for anaerobic metabolism (orange arrows) or acetyl-CoA for aerobic metabolism via mitochondrial respiration (blue arrows). DCA inhibits pyruvate dehydrogenase kinase (PDHK), alleviating phosphorylation of PDH that would otherwise cause its inhibition. This increases the rate of pyruvate entry into mitochondrial respiration and effectively reduces anaerobic metabolism of glucose to lactate. L-Proline enters mammalian cells via SNAT2 and is metabolized via the proline cycle in which L-Proline is converted to pyrroline-5-carboxylic acid (P5C) by proline oxidase (POX). P5C can be converted back to L-Proline by P5C-reductase (P5CR). L-Proline metabolism can be reduced by DHP or THFA inhibition of POX. The proline cycle is redox coupled to the pentose phosphate pathway (PPP, purple arrows) for protein and nucleotide synthesis, where NADP^+ can be used as an oxidizing agent at various stages. PPP activity is limited by the availability of NADP^+ to act on rate-limiting enzyme, 6-phosphogluconate dehydrogenase (6-PGD). 6AN inhibits the PPP by first being converted to 6-amino- NADP^+ (6ANP^+) by NAD^+ glycohydrolase, which then inhibits 6-PGD. The proposed emergent properties during the L-Proline-mediated ES-to-EPLC transition that each of the metabolic pathways affect and studied in this Chapter are shown in white boxes. MCT: monocarboxylate transporter, Gly3P: glyceraldehyde-3-phosphate, R5P: ribose-5-phosphate, OAA: oxaloacetate, α -KG: alpha-ketoglutarate, SucCoA: succinyl-coA. Adapted from: (Köhler *et al.*, 1970; Michelakis *et al.*, 2008; Casalino *et al.*, 2011; Rawat *et al.*, 2012; Tsouko *et al.*, 2014; De Preter *et al.*, 2016; Hancock *et al.*, 2016).

As the embryo continues to develop the ICM transitions to the primed primitive ectoderm. EpiSC cell lines are equivalent to this primed state. Therefore, they lie further along the pluripotency spectrum than ground- and naïve-state mESCs (Hackett *et al.*, 2013; Kojima *et al.*, 2014; Peng *et al.*, 2016). EpiSCs rely even more on anaerobic metabolism to generate ATP than ESCs (Zhou *et al.*, 2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022).

Glycolysis also produces large amounts of NADPH which help maintain nucleic acid and protein synthesis necessary for the high proliferation rates of the early, rapidly growing embryo (Figure 3.1) (Folmes *et al.*, 2012; Tsouko *et al.*, 2014; Lees *et al.*, 2017). Rapidly proliferating ESCs display similar requirements (Rodrigues *et al.*, 2015) and although it has not been specifically tested, addition of L-Proline may stimulate glycolysis along with the PPP to meet the even higher proliferation rates that accompany the transition of ESCs along the pluripotency spectrum to EPLCs (Figure 3.1) (Aral and Kamoun, 1997; Phang *et al.*, 2008; Casalino *et al.*, 2011; Liu *et al.*, 2015; De Preter *et al.*, 2016).

A comparison of the metabolic profiles of ground- and naïve-state mESCs, along with EPLCs and primed EpiSCs, will help show how cells across the pluripotency spectrum respond to exogenously added L-Proline (Figure 3.2). To facilitate this, the proline cycle can be disrupted using the POX inhibitors 3,4-dehydro-L-proline (DHP) or L-tetrahydro-2-furoic acid (THFA) (Figure 3.1 and 3.2). DHP is a substrate analogue of L-Proline and competitively inhibits POX. DHP prevents proline-induced morphological changes in mESCs at a three-fold molar excess over L-Proline (Figure 3.1) (Casalino *et al.*, 2011). THFA is also a competitive inhibitor of POX (Figure 3.1) (Zhu *et al.*, 2002; Krishnan *et al.*, 2008; Elia *et al.*, 2017).

Dichloroacetate (DCA) can be added to cultured mESCs in the presence of L-Proline to assess the involvement of anaerobic metabolism on the L-Proline-mediated transition of mESCs to EPLCs. DCA inhibits pyruvate dehydrogenase kinase (PDHK) and therefore alleviates inhibition of pyruvate dehydrogenase (PDH) to allow pyruvate to enter mitochondrial

respiration (Figure 3.1) (Michelakis *et al.*, 2008; Sanchez *et al.*, 2013; Rodrigues *et al.*, 2015; De Preter *et al.*, 2016). In effect, this leads to a reduction in glycolysis to lactate and, therefore, anaerobic metabolism.

6-aminonicotinamide (6AN) inhibits the PPP by acting on 6-phosphogluconate dehydrogenase through an intermediate (Figure 3.1) (Köhler *et al.*, 1970; Rawat *et al.*, 2012; De Preter *et al.*, 2016). Therefore, it can be used to determine the role of the PPP on changes in colony morphology, proliferation rates and gene expression profile associated with the mESC-to-EPLC transition.

In this Chapter, we use experimental and bioinformatics approaches to examine L-Proline-mediated metabolic changes across the pluripotency spectrum from ground-state mESCs to primed-state EpiSCs, which includes the metastable intermediate populations of naïve-state mESCs and EPLCs (Figure 3.2).

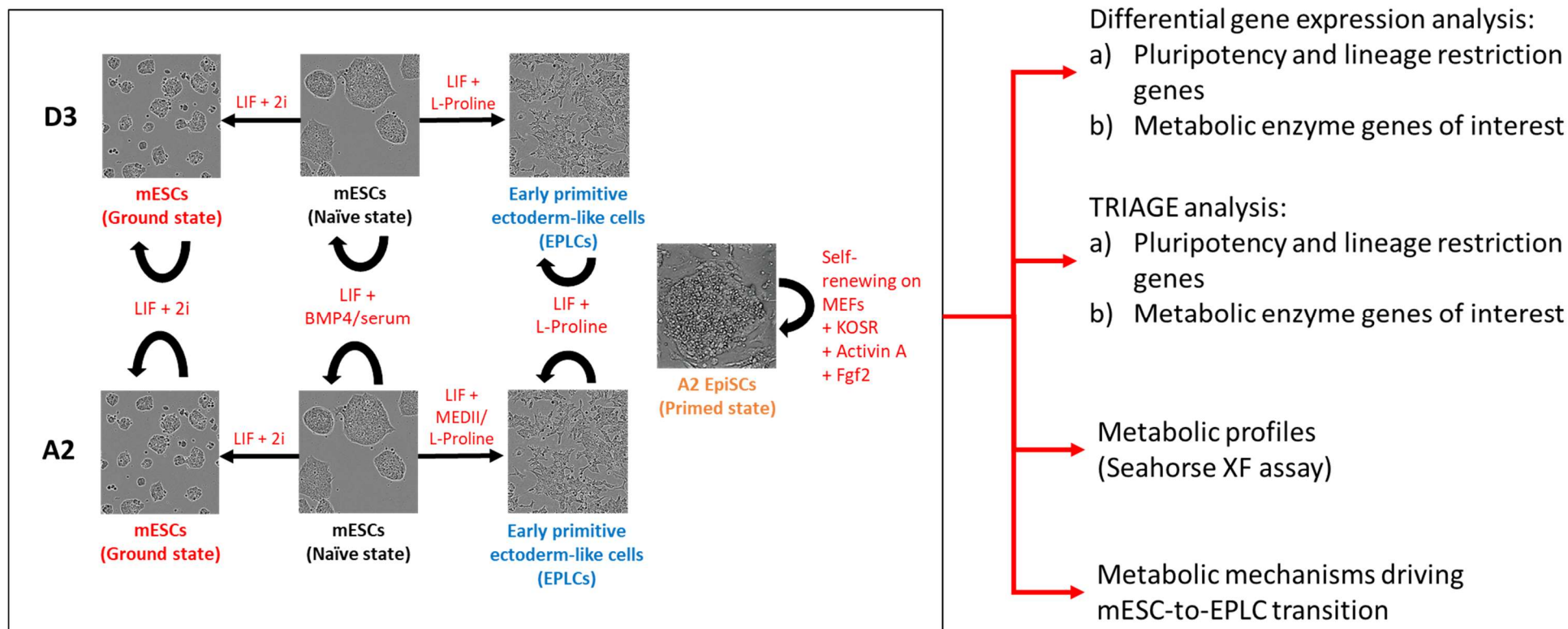


Figure 3.2 Bioinformatics and experimental approaches used to determine L-Proline-mediated metabolic changes across the pluripotency spectrum. D3 and A2 naïve-state mESCs were either maintained in *mESC medium* (LIF + BMP4/serum), cultured in *2i medium* (LIF + 2i) to produce and maintain ground state or cultured in *EPLC medium* (LIF + L-Proline) to produce and maintain EPLCs. A2 EpiSCs were derived from A2 mESCs (Osteil *et al.* 2016) and maintained on MEFs in *Complete EpiSC medium* supplemented with KOSR (+ KOSR + Activin A + Fgf2). For bioinformatics approaches, RNA samples (n=3) were subjected to RNAseq (by Dr H. Glover. Also see Methods Section 2.4). Raw counts were acquired and underwent differential gene expression analysis (DESeq2) and TRIAGE analysis (Shim *et al.* 2020) to interrogate pluripotency and lineage restriction genes. Metabolic enzyme genes of interest were also interrogated via the same analyses. For experimental approaches, D3 ground- and naïve-state mESCs and D3 EPLCs were subjected to metabolic profiling via the Seahorse Extracellular Flux (XF) assay protocol. The mESC-to-EPLC transition was also assessed using metabolic inhibitors (see Figure 3.1) in the presence of L-Proline. 2i = MAPK inhibitor PD0325091 and the GSK-3 β inhibitor CHIR99021.

3.2. Hypotheses and aims

In this Chapter, the following hypotheses will be tested:

1. L-Proline added to pluripotent cells will result in changes in the expression of pluripotency-associated genes and metabolism genes, as shown by bioinformatics analysis of RNAseq data.
2. Interrogating genes encoding enzymes associated with metabolism will provide candidates for those genes which play a role in cell identity and cell-fate decisions, as shown by Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE) analysis.
3. The addition of L-Proline to pluripotent cells will increase anaerobic metabolism across the pluripotency spectrum, as measured by the Seahorse Extracellular Flux (XF) assay.
4. The mESC-to-EPLC transition requires the breakdown of L-Proline to P5C by POX, as determined by the use of the POX inhibitors DHP and THFA in the presence of L-Proline.
5. An L-Proline-mediated increase in the activity of the PPP and anaerobic glycolysis is required for the mESC-to-EPLC transition, as determined by the 6-PGD inhibitor, 6AN, and the PDHK inhibitor, DCA, in the presence of L-Proline.

3.3. Results

3.3.1. *RNAseq analysis of cells at various stages along the pluripotency spectrum from ground state to primed state*

The pluripotency spectrum is a dynamic, reversible set of states, which ranges from ground state to naïve state to a primed state of pluripotency, the last of which is represented by primitive ectoderm in the embryo and cell lines derived from primitive ectoderm known as EpiSCs (Hackett *et al.*, 2013). An intermediate/formative stage of pluripotency exists between naïve and primed states in the form of EPLCs (Figure 3.2) (Rathjen *et al.*, 1999; Rodda *et al.*, 2002).

Briefly, ground-state mESCs are derived from naïve-state mESCs by culturing in LIF + 2i medium, where 2i represents the MAPK inhibitor PD0325091, and the GSK-3 β inhibitor CHIR99021 (see Section 2.1). Naïve-state mESCs are maintained in LIF + BMP4/serum medium (see Section 2.1). EPLCs are derived from naïve-state mESCs by culturing in LIF + L-Proline medium (see Section 2.1). EpiSCs are derived from naïve-state mESCs from the A2 cell line by chimera production (Osteil *et al.*, 2016). A2 naïve mESCs were incorporated into the blastocyst, then implanted into pseudopregnant mice. After blastocysts were implanted, they developed into embryos, which at 6.5 dpc, were isolated and maintained as the self-renewing A2 EpiSC line on a MEF feeder layer in a culture medium containing Fgf2, activin A and KOSR (see Section 2.1) (Osteil *et al.*, 2016).

Samples from these stages of pluripotency were collected and RNAseq was performed by Dr. Hannah Glover, as described in Figure 3.2 and 3.3, over 3 independent experiments (n=3) for D3 and A2 cell lines to obtain transcriptome profiles. We continued bioinformatics interrogation of this RNAseq dataset, starting with principle component analysis (Section 3.3.1.1).

3.3.1.1. Principle component analysis of RNAseq data

Raw RNAseq data produced by Dr Hannah Glover was processed by me, starting with a principle component analysis (PCA). Samples from the same state of pluripotency should cluster together, while samples at different stages across the pluripotency spectrum should cluster separately. Figure 3.3 shows the PCA plot of samples from the D3 and A2 cell lines, which cluster close together at the ground, naïve and EPLC stages. Each of the clusters at these stages are separate from one another and also separate from primed EpiSC state (Figure 3.3).

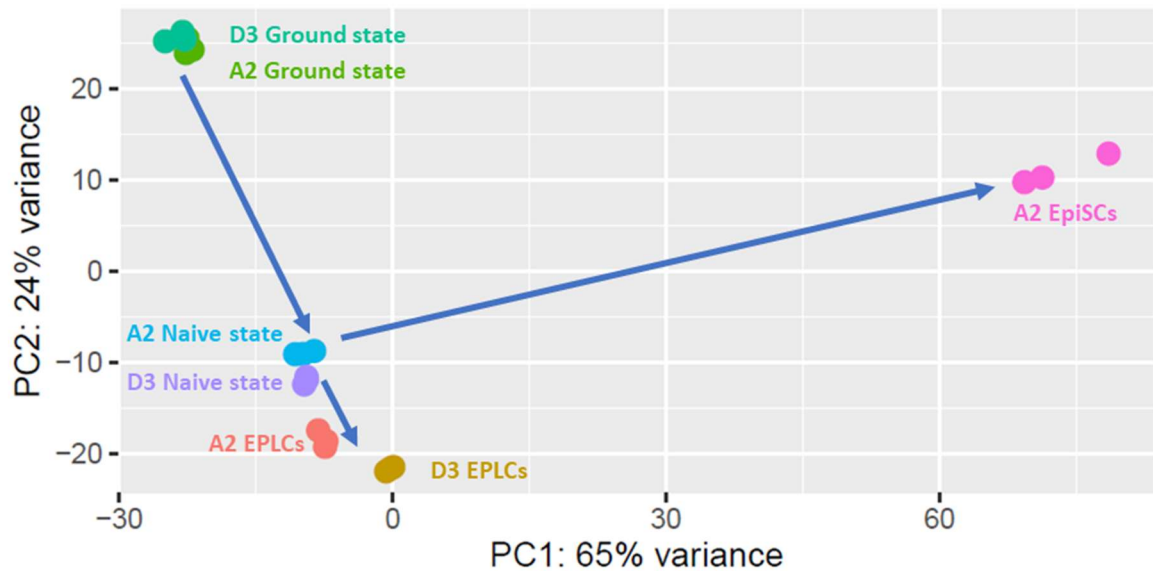


Figure 3.3 Principle components analysis (PCA) of cells from the D3 and A2 cell lines at various states along the pluripotency spectrum. Naïve-state mESCs were maintained as per Figure 3.2 and Methods (see Section 2.1). Ground-state mESCs were generated from naïve-state mESCs as shown in Figure 3.2 and Methods (see Section 2.1). EPLCs were generated from naïve-state mESCs by the addition of L-Proline (see Section 2.1). The A2 EpiSCs represent the primed pluripotency state. 3 independently derived mRNA samples ($n = 3$) from each cell line for each stage along the pluripotency continuum were subjected to RNAseq (see Section 2.4). RNAseq output was pipelined (see Section 2.4.3). Briefly, RNAseq output was quality tested using FastQC, then aligned using HISAT2 to the *Mus musculus* mm10 genome. The number of reads was quantified using FeatureCounts and the *Mus musculus* GRCm39 annotation file from Gencode. The DESeq2 pipeline was used for differential gene expression analysis. PCA for each sample was plotted using this DESeq2 pipeline on Rstudio. The arrows indicate the proposed relationship between cell types along the pluripotency spectrum.

3.3.1.2. Differential gene expression analysis across the pluripotency spectrum

Differential gene expression analysis was also performed on each of the states along the pluripotency spectrum. Significant changes in expression were defined as having a Benjamini-Hochberg P value or false discovery rate (FDR) <0.05 , a value which helps reduce false positives, regardless of the value of $\log_2$ fold changes in expression (Koch *et al.*, 2018).

Figure 3.4 shows changes in expression of various groups of genes associated with pluripotency and early stages of development. The results are shown as heat maps, where expression of ground-state mESCs, EPLCs and EpiSCs are compared to naïve mESCs. Asterisks on the heat maps indicate that gene expression is significantly different from the expression in the naïve state.

Figure 3.4 shows that genes of the core pluripotency network, *Oct4*, *Nanog* and *Sox2*, are expressed across the pluripotency spectrum but expression tends to decrease towards the primed end of the spectrum.

Similar results are seen for the genes of the extended pluripotency network with the exception of *Id1* and *Id2*, whose highest expression is at the naïve state.

The first lineage commitment in the developing embryo is the formation of ICM cells and the outer layer of trophoctoderm, which goes on to form the placenta. Expression of genes associated with the trophoctoderm lineage (*Cdx2*, *Gata3* and *Tead4*) were upregulated in ground state relative to naïve state and EPLCs, consistent with an early decision point in the pluripotency spectrum for commitment to this lineage.

On the other hand, the expression of markers for primitive ectoderm (*Dnmt3b*, *Fgf5*, *Otx2* and *Lefty2*) increase across the pluripotency spectrum, as expected (Rathjen *et al.*, 1999; Pelton *et al.*, 2002; Washington *et al.*, 2010; Glover, 2018; Glover *et al.*, 2022).

Similarly, expression of markers for neurectoderm (*Pax6* and *Sox1*), surface ectoderm (*Krt18*) and definitive endoderm (*Gata6* and *Sox17*) tend to increase across the pluripotency spectrum, perhaps indicating preparation for lineage commitment to these cell types following eventual loss of pluripotency.

Of note is that EPLCs show upregulated expression of the definitive ectoderm germ layer markers *Pard6b* and *Penk1* (Harvey *et al.*, 2010), at least in the D3 line, and downregulated expression of the mesendoderm marker, *T*, in both lines. On the other hand, EpiSCs show no change in the expression of these definitive ectoderm markers and have upregulated expression of both *T* and *Eomes* mesendoderm markers. This is an initial indication that EPLCs represent a branchpoint in the pluripotency spectrum (Figure 3.3), which is becoming committed to the formation of the ectoderm germ layer, while EpiSCs are at a point in development which is preparing to form the other two germ layers.

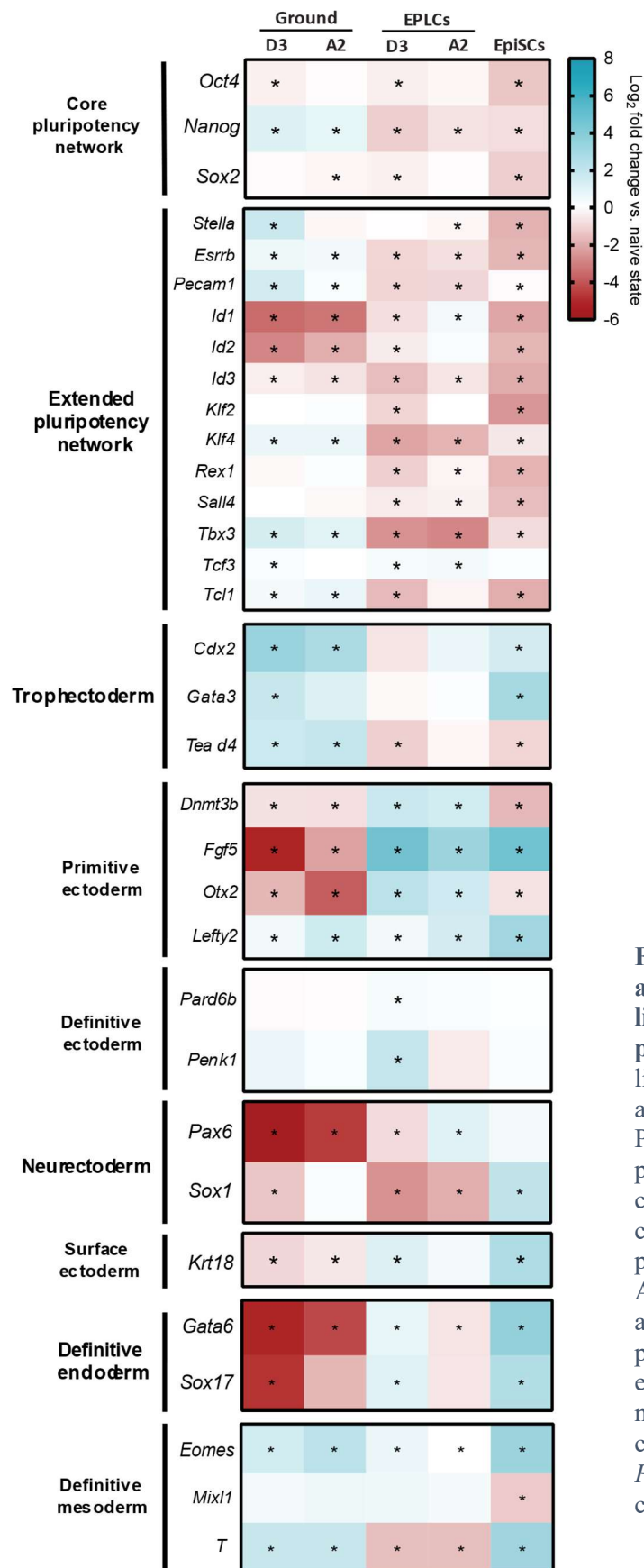


Figure 3.4 Expression of genes associated with pluripotency and lineage restriction across the ESC pluripotency spectrum. D3 and A2 ESC lines were analysed for gene expression at the ground state, naïve state and the L-Proline-induced EPLC stages of the pluripotency spectrum following the collection of RNASeq data (n = 3 for each cell line at each stage of pluripotency, as per Figure 3.2). In addition, data for the A2 EpiSCs (n = 3) were collected and analysed (as per Figure 3.2). Results are plotted as heat maps and changes in gene expression (mean log₂ fold change) normalised to the naïve state for the same cell line. * Benjamini-Hochberg adjusted $P < 0.05$, and indicates a significant change in expression.

3.3.1.3. TRIAGE transformation of RNAseq data

The same RNAseq raw data then underwent TRIAGE transformation to discover candidate genes that contribute to cell fate and identity across the pluripotency spectrum (Figure 3.5) (Shim *et al.*, 2020). The TRIAGE method utilises the repressive H3K27me3 mark to quantify discordance between gene-repressive tendency and expression, also known as the discordance score (DS). A high DS rank (rank = 1-1359) indicates transient expression of genes, inferring a role in regulating cell fate and determining cell identity (Figure 3.5) (Shim *et al.*, 2020). Medium-low (rank = 1360-2718) to low DS (rank >2718) indicates ubiquitous and consistent expression of genes and roles associated with structural integrity and housekeeping, respectively, and so probably not driving cell fate (Figure 3.5) (Shim *et al.*, 2020).

Figure 3.5 shows an example of implementing the TRIAGE method using D3 ground-state mESCs. Panel A shows DS plotted against genes ranked by DS. The 10 genes with the highest DS are highlighted in red and appear at the left side of the plot.

Panel B shows DS plotted against gene expression levels in these ground-state cells. Note that the 10 highlighted genes have very high DS compared to the vast majority of genes whose DS lies very close to zero. The range of expression of the 10 highlighted genes is very wide. Some, such as *Hoxc12*, appear to be highly expressed while others, such as *Cnpy1*, have much lower expression.

Ground-state D3 cells express 8866 genes. However, Panel C shows that of the 10 genes with the highest DS even the highest expresser, *Esrrb*, is expressed at only 2% of the level of the most highly expressed gene in ground-state D3 cells, *Rplp1* (encoding for a ribosomal protein). *Igf2bp1* is expressed at only 0.8% of the level of *Rplp1*. The other 8 genes with the highest DS are expressed at levels between 0.05 and 0.5% of *Rplp1* expression. These low levels of expression are consistent with the function of the products of these 10 genes: 7 are genes for transcription factors, 2 are involved in the cell signalling, and 1 is involved in cell

cycling. Thus, the functions of this group of 10 genes with the highest DS are consistent with their involvement in determining cell fate.

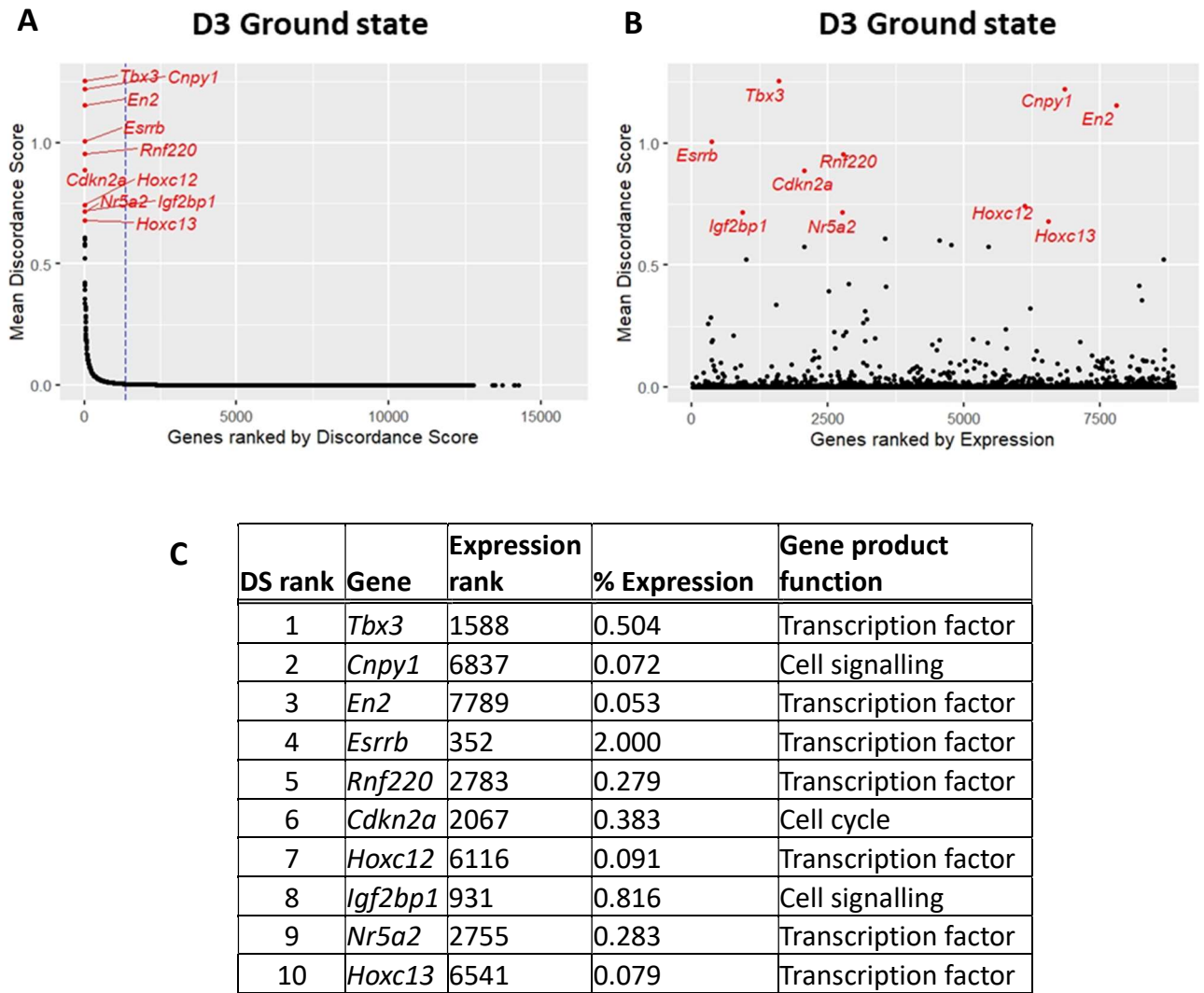
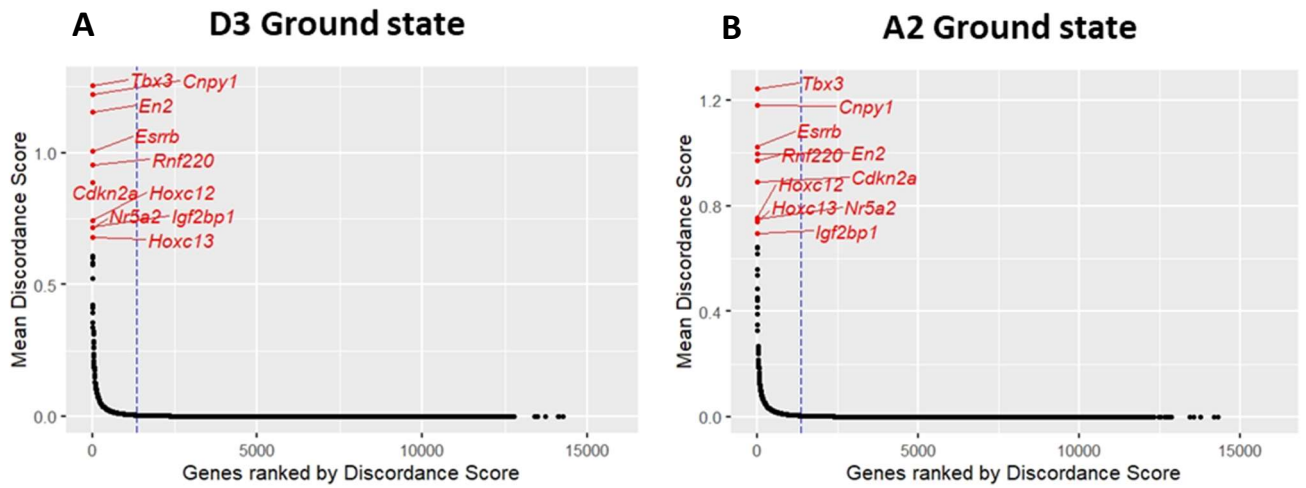


Figure 3.5 D3 ground-state mESC RNAseq data transformed using Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE). Mean Discordance Scores (DS) from the TRIAGE transformation were calculated from the results obtained for the three D3 ground-state ESC samples subjected to RNASeq analysis. **A.** Mean DS plotted against the ranked DS values for all genes expressed in the D3 ground-state cells, with DS rank threshold (1359) shown by the dotted blue line. The 10 genes with the highest DS are highlighted in red and named. Raw FeatureCount reads from the RNAseq analysis were normalised to fragments per kilobase of transcript per million mapped reads (FPKM) to account for sequencing depth and gene length, and used as a measure of relative gene expression. An arbitrary cut-off of ≥ 3 FPKM was used to define genes that were expressed; a total of 8866 genes in this case. **B.** Mean DS plotted against the ranked gene expression (highest to lowest) for all genes expressed in D3 ground-state cells. **C.** Table of the 10 genes with the highest DS and their percentage expression compared to most highly expressed gene in D3 ground-state cells, *Rplp1*. The function of each gene product is also shown.

Figure 3.6 shows a comparison of TRIAGE results for D3 ground-state cells with A2 ground-state cells. Panels C and D show that the 10 genes with the highest DS are identical in both cell lines (with small differences in order), and their percentage of expression compared to the most highly expressed gene are very similar, indicating there is very close similarity in the properties of these cell lines in the ground state, as measured by these parameters.



C

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Tbx3</i>	1588	0.504	Transcription factor
2	<i>Cnpy1</i>	6837	0.072	Cell signalling
3	<i>En2</i>	7789	0.053	Transcription factor
4	<i>Esrrb</i>	352	2.000	Transcription factor
5	<i>Rnf220</i>	2783	0.279	Transcription factor
6	<i>Cdkn2a</i>	2067	0.383	Cell cycle
7	<i>Hoxc12</i>	6116	0.091	Transcription factor
8	<i>Igf2bp1</i>	931	0.816	Cell signalling
9	<i>Nr5a2</i>	2755	0.283	Transcription factor
10	<i>Hoxc13</i>	6541	0.079	Transcription factor

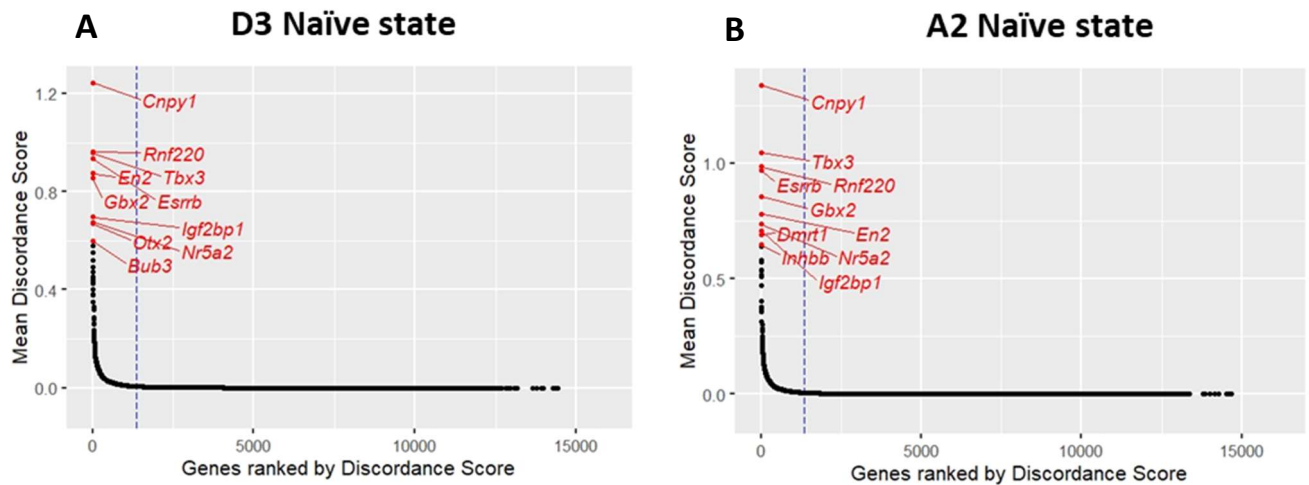
D

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Tbx3</i>	1560	0.479	Transcription factor
2	<i>Cnpy1</i>	6464	0.077	Cell signalling
3	<i>Esrrb</i>	324	1.964	Transcription factor
4	<i>En2</i>	8714	0.038	Transcription factor
5	<i>Rnf220</i>	2667	0.271	Transcription factor
6	<i>Cdkn2a</i>	1825	0.405	Cell cycle
7	<i>Hoxc12</i>	5918	0.091	Transcription factor
8	<i>Nr5a2</i>	2356	0.310	Transcription factor
9	<i>Hoxc13</i>	6212	0.083	Transcription factor
10	<i>Igf2bp1</i>	1106	0.639	Cell signalling

Figure 3.6 D3 and A2 ground-state mESC RNAseq data transformed using TRIAGE. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three ground-state mESC samples from both D3 and A2 cell lines subjected to RNASeq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in the **A.** D3 ground-state cells and **B.** A2 ground-state cells, with DS rank threshold (1359) shown on both graphs by the dotted blue line. The 10 genes with the highest DS are highlighted in red and named. Raw FeatureCount reads from the RNAseq analysis were normalised to FPKM to account for sequencing depth and gene length, and used as a measure of relative gene expression. An arbitrary cut-off of ≥ 3 FPKM was used to define genes that were expressed; a total of 8866 genes for D3 cell line, and 8936 for the A2 cell line. The 10 genes with the highest DS and their percentage expression compared to most highly expressed gene in **C.** D3 ground-state cells, *Rplp1* or **D.** A2 ground-state cells, *mt-Col*. The function of each gene product is also shown.

Figure 3.7 shows the same comparison of TRIAGE results for D3 naïve-state cells and A2 naïve-state cells. 8 out of 10 genes with the highest DS are identical in both cell lines. *Otx2* and *Bub3* appear in the top 10 for D3 naïve-state cells but do not appear in the A2 list. However, they are still ranked highly in terms of DS coming in at positions 11 and 12, respectively. Similarly, *Dmrt1* and *Inhbb* appear in the top 10 for A2 naïve-state cells but do not appear in the D3 list. However, they are still ranked highly in terms of DS coming in at positions 12 and 14, respectively. Again, this indicates that the two cell lines are behaving in a very similar fashion at this stage of the pluripotency spectrum.

7 genes (*Tbx3*, *Cpny1*, *En2*, *Esrrb*, *Rnf220*, *Igf2bp1* and *Nr5a2*) that appear in the top 10 DS list for ground-state cells also appear in the top 10 list for naïve-state cells. 3 genes (*Cdkn2a*, *Hoxc12*, *Hoxc13*) which appear in the list for ground-state cells don't appear in the list for naïve-state cells. These 3 genes come in at DS positions 11, 143, and 56 in D3 naïve cells, respectively, and positions 14, 126 and 33, in A2 naïve cells, respectively. Similarly, 5 genes (*Gbx2*, *Otx2*, *Bub3*, *Dmrt1*, *Inhbb*) which appear in the top 10 DS list for naïve cells come in at DS positions 15, 25, 14, 19 and 63 in D3 ground-state cells, respectively, and positions 11, 113, 14, 20 and 22 in A2 ground-state cells, respectively.



C

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Cnpy1</i>	6050	0.105	Cell signalling
2	<i>Rnf220</i>	2763	0.327	Transcription factor
3	<i>Tbx3</i>	3515	0.243	Transcription factor
4	<i>Esrrb</i>	594	1.567	Transcription factor
5	<i>En2</i>	9355	0.037	Transcription factor
6	<i>Gbx2</i>	2066	0.453	Transcription factor
7	<i>Igf2bp1</i>	1161	0.808	Cell signalling
8	<i>Nr5a2</i>	3198	0.273	Transcription factor
9	<i>Otx2</i>	6983	0.078	Transcription factor
10	<i>Bub3</i>	1879	0.503	Cell cycle

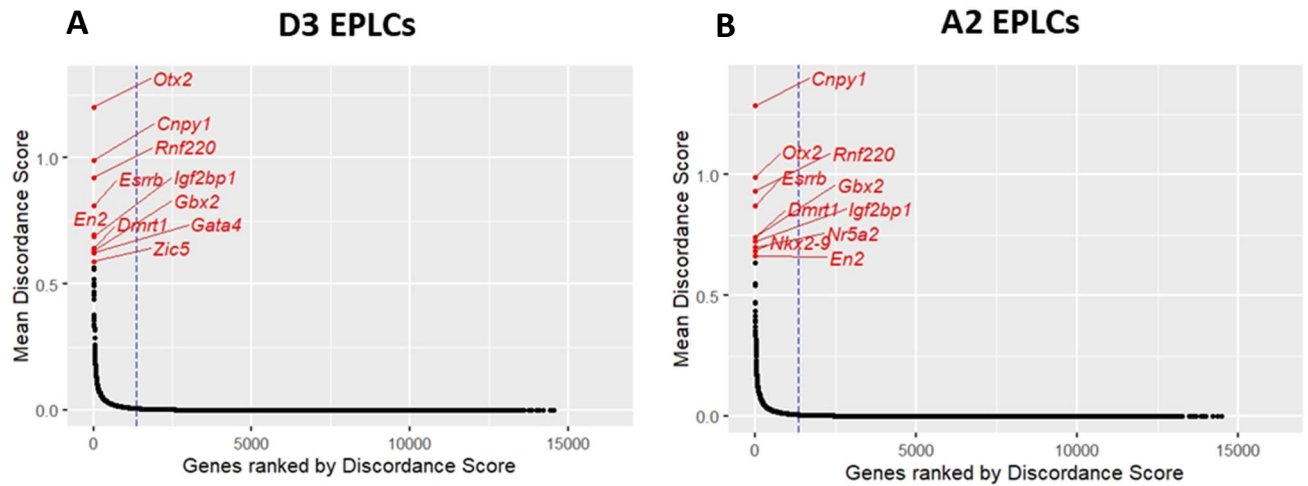
D

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Cnpy1</i>	5552	0.132	Cell signalling
2	<i>Tbx3</i>	2765	0.340	Transcription factor
3	<i>Rnf220</i>	2516	0.375	Transcription factor
4	<i>Esrrb</i>	440	1.981	Transcription factor
5	<i>Gbx2</i>	1972	0.483	Transcription factor
6	<i>En2</i>	10159	0.032	Transcription factor
7	<i>Nr5a2</i>	2478	0.381	Transcription factor
8	<i>Igf2bp1</i>	1043	0.913	Cell signalling
9	<i>Dmrt1</i>	5060	0.154	Transcription factor
10	<i>Inhbb</i>	1127	0.853	Transcription factor

Figure 3.7 D3 and A2 naïve-state mESC RNAseq data transformed using TRIAGE. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three naïve-state mESC samples from both D3 and A2 cell lines subjected to RNASeq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in the **A.** D3 naïve-state cells and **B.** A2 naïve-state cells, with DS rank threshold (1359) shown on both graphs by the dotted blue line. The 10 genes with the highest DS are highlighted in red and named. Raw FeatureCount reads from the RNAseq analysis were normalised to FPKM to account for sequencing depth and gene length, and used as a measure of relative gene expression. An arbitrary cut-off of ≥ 2 FPKM was used to define genes that were expressed; a total of 10,049 genes for D3 cell line, and 10,258 for the A2 cell line. The 10 genes with the highest DS and their percentage expression compared to most highly expressed gene in **C.** D3 naïve-state cells, *Rplp1* or **D.** A2 naïve-state cells, *mt-Col*. The function of each gene product is also shown.

Figure 3.8 shows the same comparison of TRIAGE results for D3 EPLCs with A2 EPLCs. 8 out of 10 genes with the highest DS are identical in both cell lines. *Gata4* and *Zic5* appear in the top 10 for D3 EPLCs but do not appear in the A2 list. However, they are still ranked highly in terms of DS coming in at positions 21 and 15, respectively. Similarly, *Nr5a2* and *Nkx2-9* appear in the top 10 for A2 EPLCs but do not appear in the D3 list. However, they are still ranked highly in terms of DS coming in at positions 13 and 16, respectively. Again, this indicates that the two cell lines are behaving in a very similar fashion at this stage of the pluripotency spectrum.

8 genes (*Otx2*, *Cnpy1*, *Rnf220*, *Esrrb*, *En2*, *Igf2bp1*, *Dmrt1* and *Gbx2*) that appear in the top 10 DS list for EPLCs also appear in the top 10 list for ground- and naïve-state cells. 2 genes (*Gata4* and *Zic5*) which appear in the top 10 for D3 EPLCs appear in D3 ground- and naïve-state cells with DS ranks within the top 50. 2 genes (*Nkx2-9* and *Nr5a2*) which appear in the top 10 DS list for A2 EPLCs also appear in A2 ground- and naïve-state cells with DS ranks within the top 30.



C

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Otx2</i>	2331	0.342	Transcription factor
2	<i>Cnpy1</i>	7830	0.056	Cell signalling
3	<i>Rnf220</i>	3061	0.251	Transcription factor
4	<i>Esrrb</i>	1061	0.739	Transcription factor
5	<i>En2</i>	10584	0.022	Transcription factor
6	<i>Igf2bp1</i>	1178	0.677	Cell signalling
7	<i>Dmrt1</i>	5733	0.107	Transcription factor
8	<i>Gbx2</i>	4555	0.153	Transcription factor
9	<i>Gata4</i>	11115	0.018	Transcription factor
10	<i>Zic5</i>	4536	0.155	Transcription factor

D

DS rank	Gene	Expression rank	% Expression	Gene product function
1	<i>Cnpy1</i>	5775	0.105	Cell signalling
2	<i>Otx2</i>	3948	0.187	Transcription factor
3	<i>Rnf220</i>	2937	0.265	Transcription factor
4	<i>Esrrb</i>	755	1.010	Transcription factor
5	<i>Dmrt1</i>	4384	0.161	Transcription factor
6	<i>Gbx2</i>	3020	0.256	Transcription factor
7	<i>Igf2bp1</i>	916	0.847	Cell signalling
8	<i>Nr5a2</i>	2843	0.275	Transcription factor
9	<i>Nkx2-9</i>	4363	0.162	Transcription factor
10	<i>En2</i>	10691	0.021	Transcription factor

Figure 3.8 D3 and A2 EPLC RNAseq data transformed using TRIAGE. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three EPLC samples from both D3 and A2 cell lines subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in the **A.** D3 EPLCs and **B.** A2 EPLCs, with DS rank threshold (1359) shown on both graphs by the dotted blue line. The 10 genes with the highest DS are highlighted in red and named. Raw FeatureCount reads from the RNAseq analysis were normalised to FPKM to account for sequencing depth and gene length, and used as a measure of relative gene expression. An arbitrary cut-off of ≥ 1 FPKM was used to define genes that were expressed; a total of 12,025 genes for D3 cell line, and 12,030 for the A2 cell line. The 10 genes with the highest DS and their percentage expression compared to most highly expressed gene, *Rplp1*, in **C.** D3 EPLCs or **D.** A2 EPLCs. The function of each gene product is also shown.

Figure 3.9 shows TRIAGE results for EpiSCs. 6 genes (*Cnpy1*, *Tbx3*, *Cdkn2a*, *Rnf220*, *En2* and *Esrrb*) that appear in the top 10 DS list for EpiSCs also appear in the top 10 list for A2 ground- and naïve-state cells and EPLCs. The 4 genes (*Gata4*, *Pitx2*, *Hoxb13* and *En1*) that uniquely appear in the top 10 for EpiSCs appear in A2 ground- and naïve-state cells and EPLCs with DS ranks in the top 600.

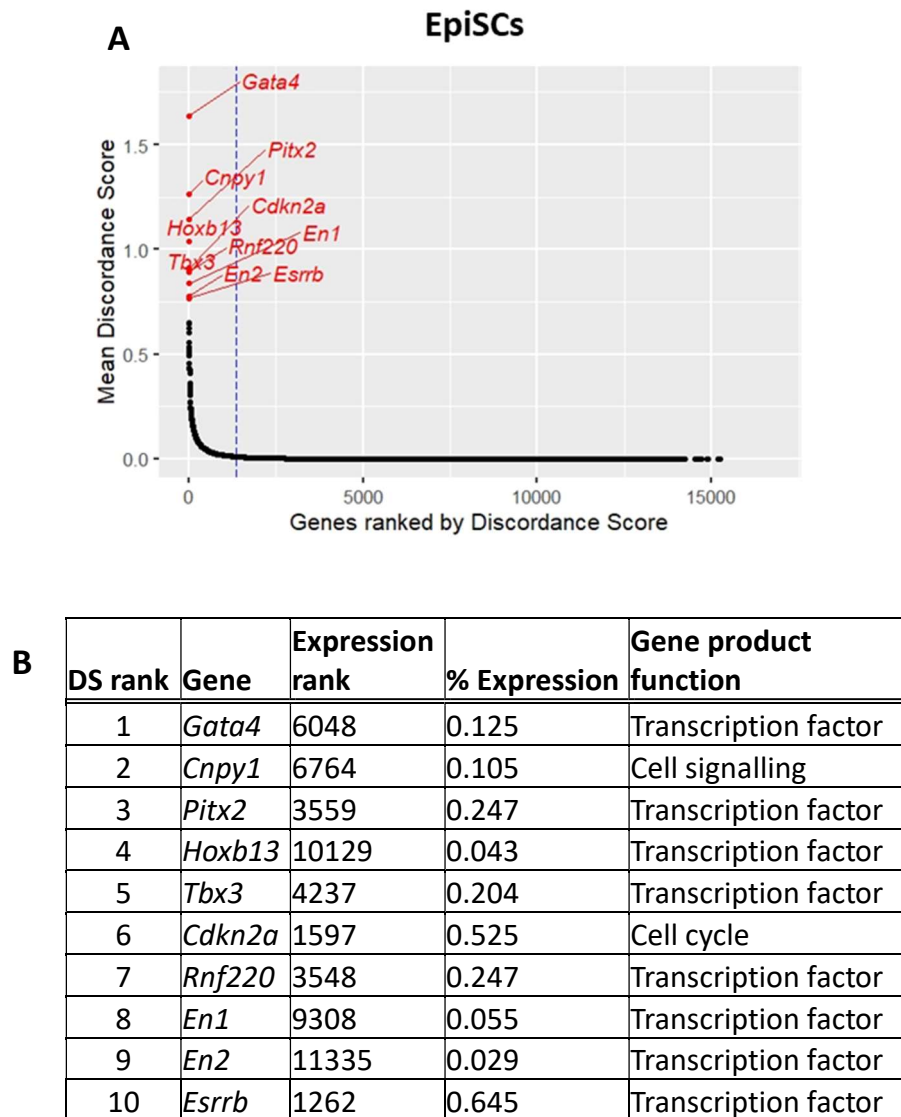


Figure 3.9 EpiSC RNAseq data transformed using TRIAGE. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three EpiSC samples from the A2 cell line subjected to RNAseq analysis. **A.** Mean DS were plotted against the ranked DS values for all genes expressed, with DS rank threshold (1359) shown by the dotted blue line. The 10 genes with the highest DS are highlighted in red and named. Raw FeatureCount reads from the RNAseq analysis were normalised to FPKM to account for sequencing depth and gene length, and used as a measure of relative gene expression. An arbitrary cut-off of ≥ 2 FPKM was used to define genes that were expressed; a total of 11,481 genes. **B.** The 10 genes with the highest DS and their percentage expression compared to most highly expressed gene, *mt-Col*. The function of each gene product is also shown.

3.3.1.4. *Gene ontology and pathway analysis of top 100 DS genes*

The mouse genes with the top 100 DS were converted to their respective human gene orthologs and subjected to TRIAGE-Parser for gene clustering analysis (Sun *et al.*, 2023). The method allows unbiased clustering of these genes and were assigned a random number to distinguish between them. Clusters are then defined by using gene ontology (GO) enrichment analysis for the “biological process” GO term, with significance denoted as Benjamini-Hochberg adjusted P or false-discovery rate (FDR) set at <0.05 to account for false positives. Protein-protein interactions (PPIs) were also predicted for each cluster using the STRING database with default settings (Szklarczyk *et al.*, 2019). PPIs predicted include functional and physical protein associations. A PPI enrichment P value <0.001 indicates these associations, and therefore the gene clusters, are biologically meaningful (Sun *et al.*, 2023).

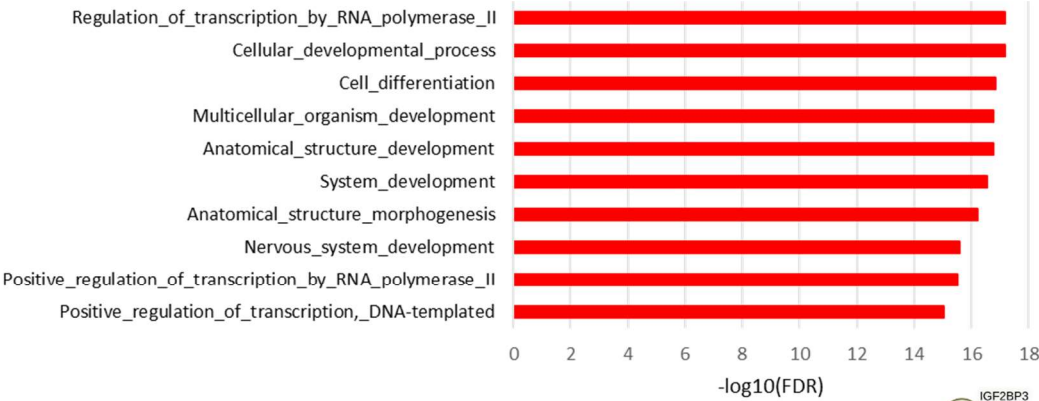
Figure 3.10 shows results from TRIAGE-Parser analysis for samples from the D3 cell line. 2 clusters were identified in the ground state (Cluster 1 and 2), which is reduced to 1 cluster at the naïve mESC and EPLC stages each (Cluster 1).

For ground-state cells, Cluster 2 consists of *HOX* genes, and contains GO terms associated with patterning and skeletal system morphogenesis. Most of the top 100 genes are in Cluster 1, which contains an extensive list of GO terms associated with transcription, metabolism, development and lineage commitment, including nervous system development as shown by their Benjamini-Hochberg adjusted P values.

For naïve-state cells and EPLCs, each Cluster contains extensive lists of GO terms associated with transcription, metabolism, development and lineage commitment. All clusters across each pluripotent cell state are biologically meaningful with PPI enrichment P values <0.001 .

A. D3 Ground state

Gene Cluster 1 GO: Biological Processes



Gene Cluster 1

Gene	Rank by mean DS
<i>TBX3</i>	1
<i>CNPY1</i>	2
<i>EN2</i>	3
<i>ESRRB</i>	4
<i>RNF220</i>	5
<i>IGF2BP1</i>	8
<i>NR5A2</i>	9
<i>NGFR</i>	11
<i>GRIK3</i>	12
<i>PPP2R2C</i>	13

Gene Cluster 1
PPI enrichment $P < 1.0^{-16}$

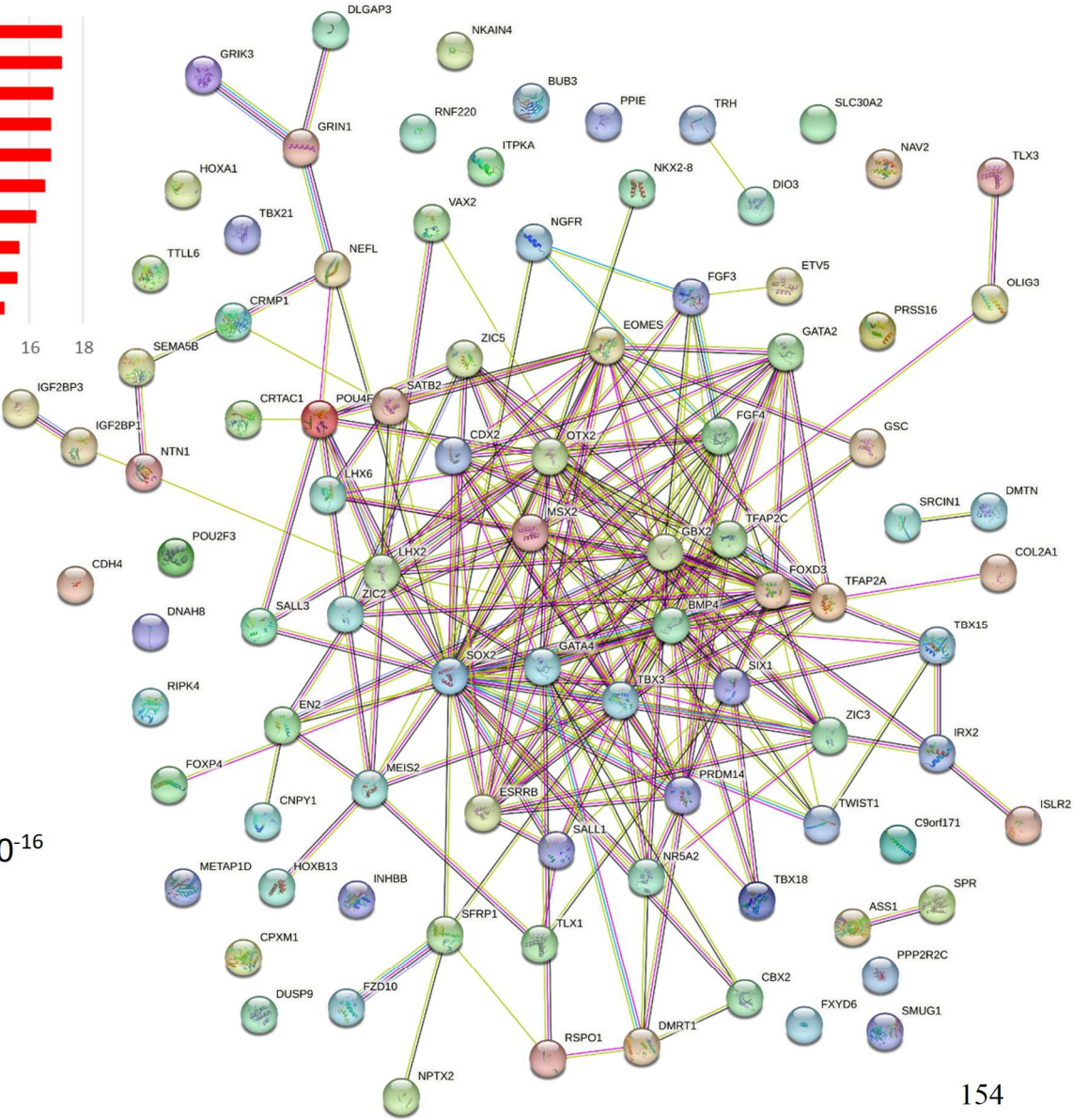
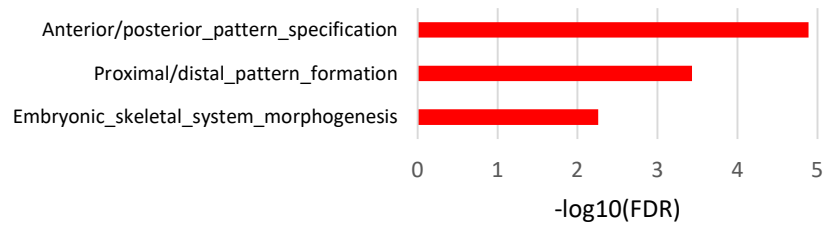


Figure 3.10A continued...

A. D3 Ground state

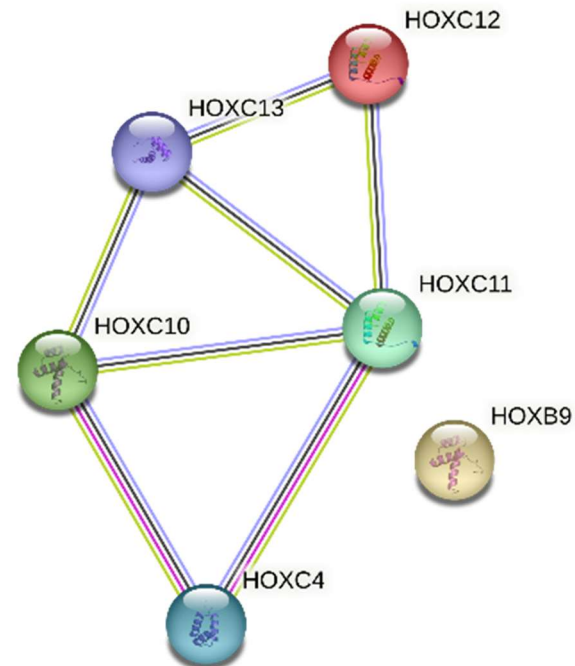
Gene Cluster 2 GO: Biological Processes



Gene Cluster 2

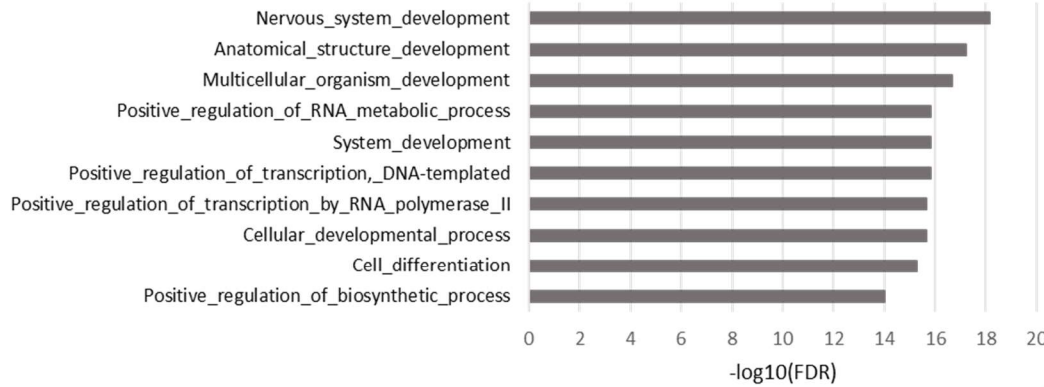
Gene	Rank by mean DS
<i>HOXC12</i>	7
<i>HOXC13</i>	10
<i>HOXC4</i>	21
<i>HOXC10</i>	27
<i>HOXC11</i>	62
<i>HOXB9</i>	96

Gene Cluster 2
PPI enrichment $P < 1.0^{-16}$



B. D3 Naïve state

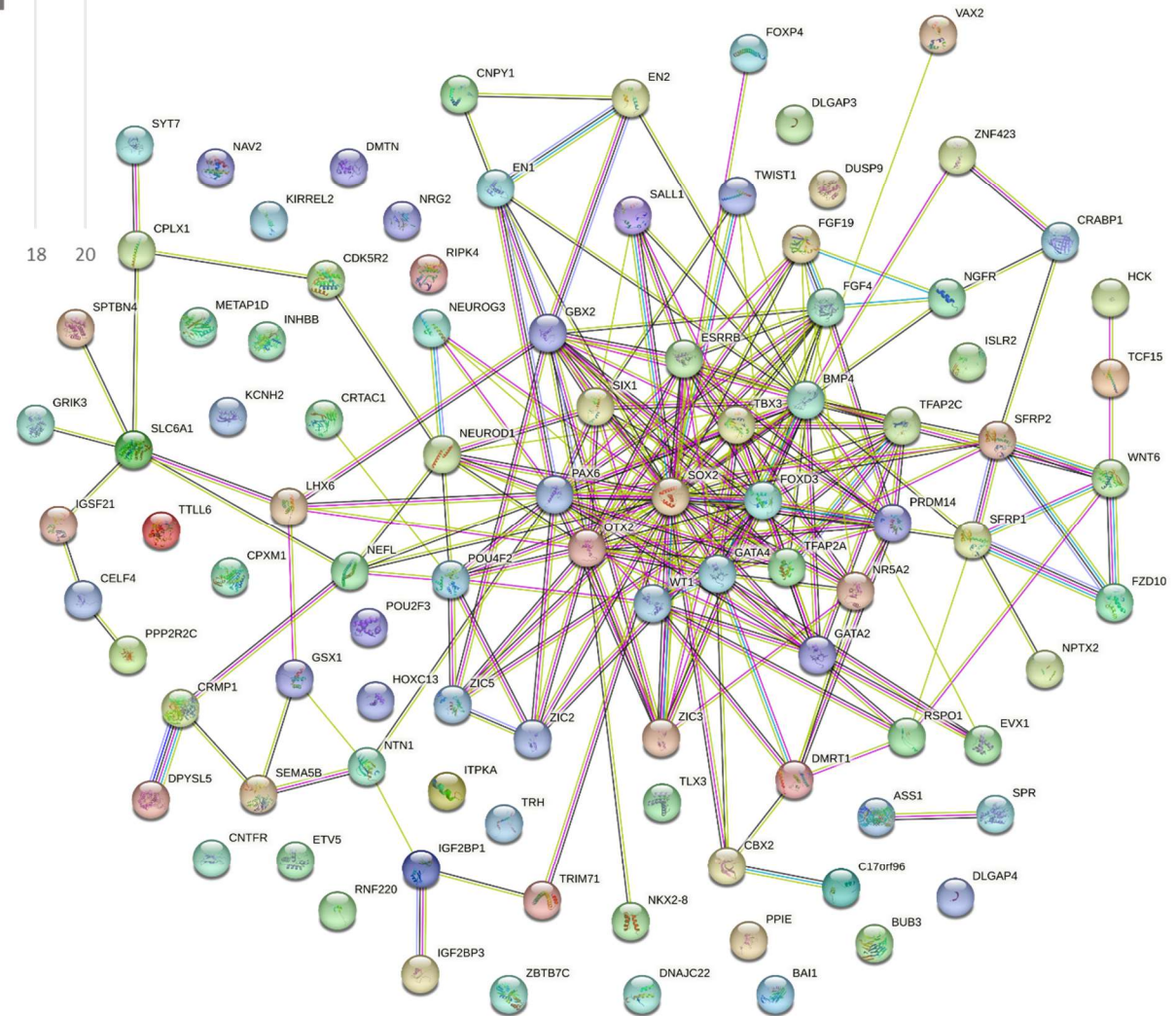
Gene Cluster 1 GO: Biological Processes



Gene Cluster 1

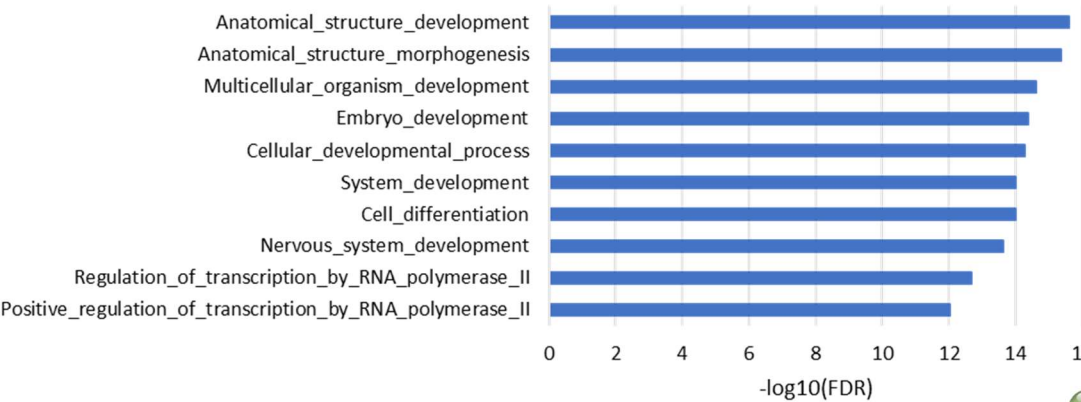
Gene	Rank by mean DS
<i>CNPY1</i>	1
<i>RNF220</i>	2
<i>TBX3</i>	3
<i>ESRRB</i>	4
<i>EN2</i>	5
<i>GBX2</i>	6
<i>IGF2BP1</i>	7
<i>NR5A2</i>	8
<i>OTX2</i>	9
<i>BUB3</i>	10

PPI enrichment $P < 1.0^{-16}$



C. D3 EPLCs

Gene Cluster 1 GO: Biological Processes



Gene Cluster 1

Gene	Rank by mean DS
OTX2	1
CNPY1	2
RNF220	3
ESRRB	4
EN2	5
IGF2BP1	6
DMRT1	7
GBX2	8
GATA4	9
ZIC5	10

PPI enrichment $P < 1.0^{-16}$

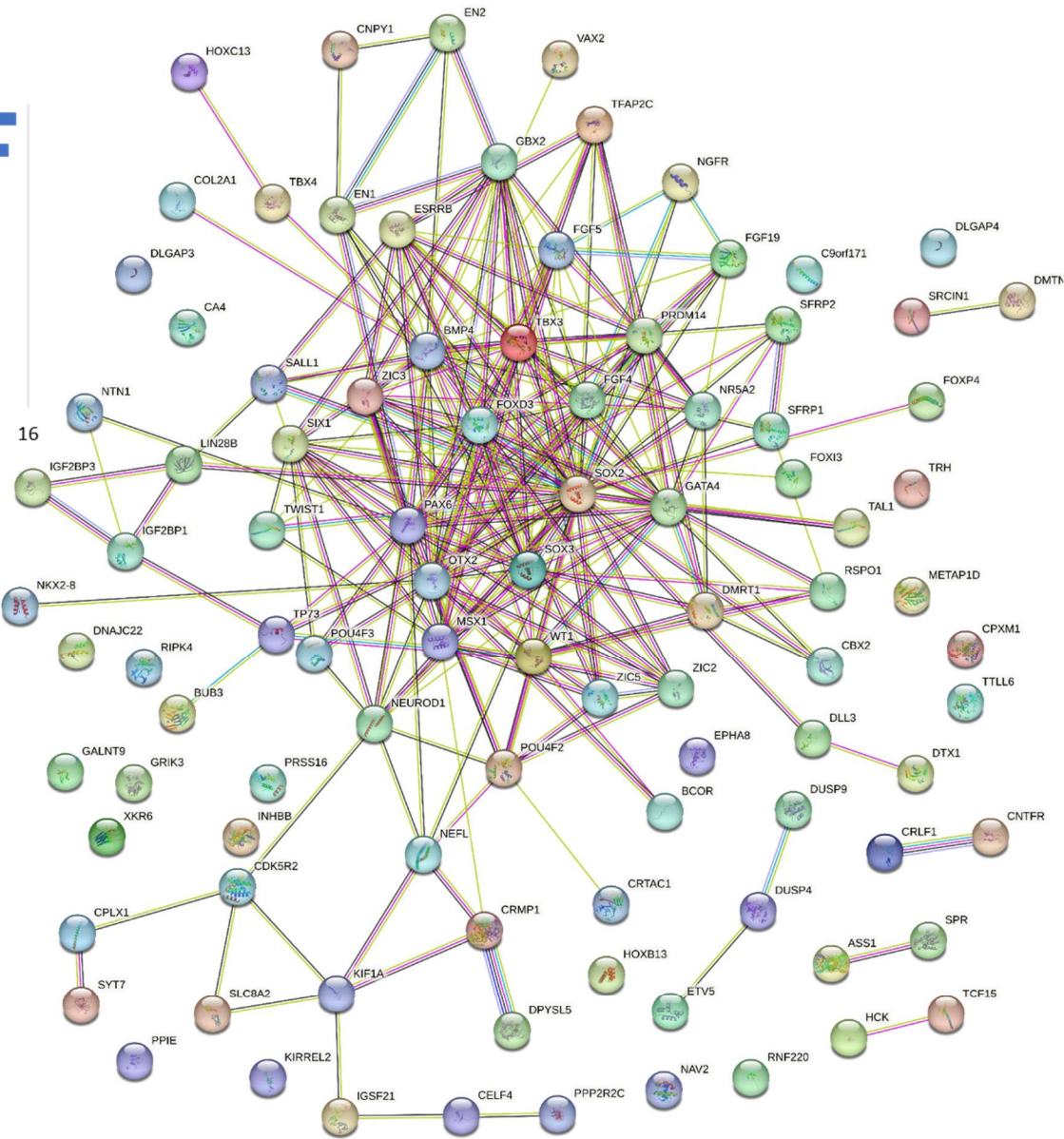


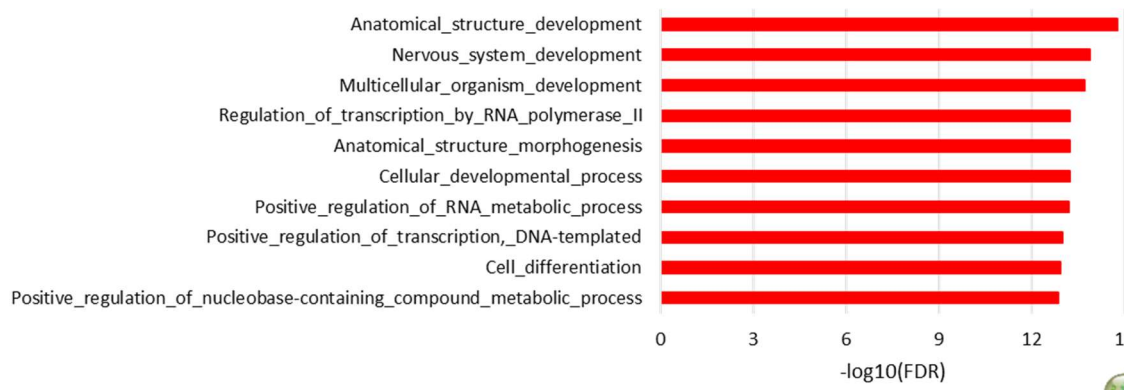
Figure 3.10 Gene ontology (GO) analysis of the top 100 ranked DS genes using TRIAGE-Parser for the D3 cell line to identify biological processes associated with the A. ground state, B. naïve state and C. EPLCs. RNAseq data was subjected to TRIAGE transformation to produce DS ranked across the entire *Mus musculus* genome. The top 100 ranked DS genes underwent TRIAGE-Parser to both categorize these genes into clusters and undergo GO enrichment analysis on the same clusters to identify their GO: biological process terms. Clusters were identified by the percentage chance the gene is in that cluster using an arbitrary cut-off of $\geq 95\%$. The top 10 terms or less, based on their Benjamini-Hochberg adjusted P value < 0.05 , or their false discovery rate (FDR), were graphed for each cluster. Protein-protein interactions (PPIs) for each cluster were predicted using the STRING database (v 11.5) with default settings. PPI enrichment FDR P value < 0.001 indicates significance and, therefore, biologically meaningful PPIs.

Figure 3.11 shows results from TRIAGE-Parser analysis for samples from the A2 cell line, including EpiSCs. The number of clusters identified across the ground, naïve and EPLC stages were the same as the D3 cell line: 2 clusters were identified in the ground state (Cluster 1 and 2), followed by 1 cluster each for naïve and EPLC stages (Cluster 1). A2 ground state also has the same type of terms under the same respective cluster, with Cluster 2 consisting of *HOX* genes and the GO term “anterior/posterior pattern specification”, and Cluster 1 with most terms associated with transcription, metabolism and development, including nervous system development. A2 naïve and EPLC states also have the same type of terms under the same respective cluster as D3 naïve and EPLC states, including terms associated with nervous system. All clusters are predicted to be biologically meaningful with PPI enrichment P values <0.001 .

Figure 3.11D shows TRIAGE-Parser results for the primed EpiSC state. 2 clusters were identified (Cluster 1 and 2). However, only Cluster 1 is functionally meaningful, as no GO terms are significant for Cluster 2 which consists of mostly *HOX* genes. This is also reflected in the PPI enrichment P values, where Cluster 1 has a PPI enrichment P value <0.001 , while Cluster 2 has a PPI enrichment P value >0.001 .

A. A2 Ground state

Gene Cluster 1 GO: Biological Processes



Gene Cluster 1

Gene	Rank by mean DS
<i>TBX3</i>	1
<i>CNPY1</i>	2
<i>ESRRB</i>	3
<i>EN2</i>	4
<i>RNF220</i>	5
<i>NR5A2</i>	8
<i>IGF2BP1</i>	10
<i>GBX2</i>	11
<i>GRIK3</i>	12
<i>NGFR</i>	13

Gene Cluster 1
PPI enrichment $P < 1.0^{-16}$

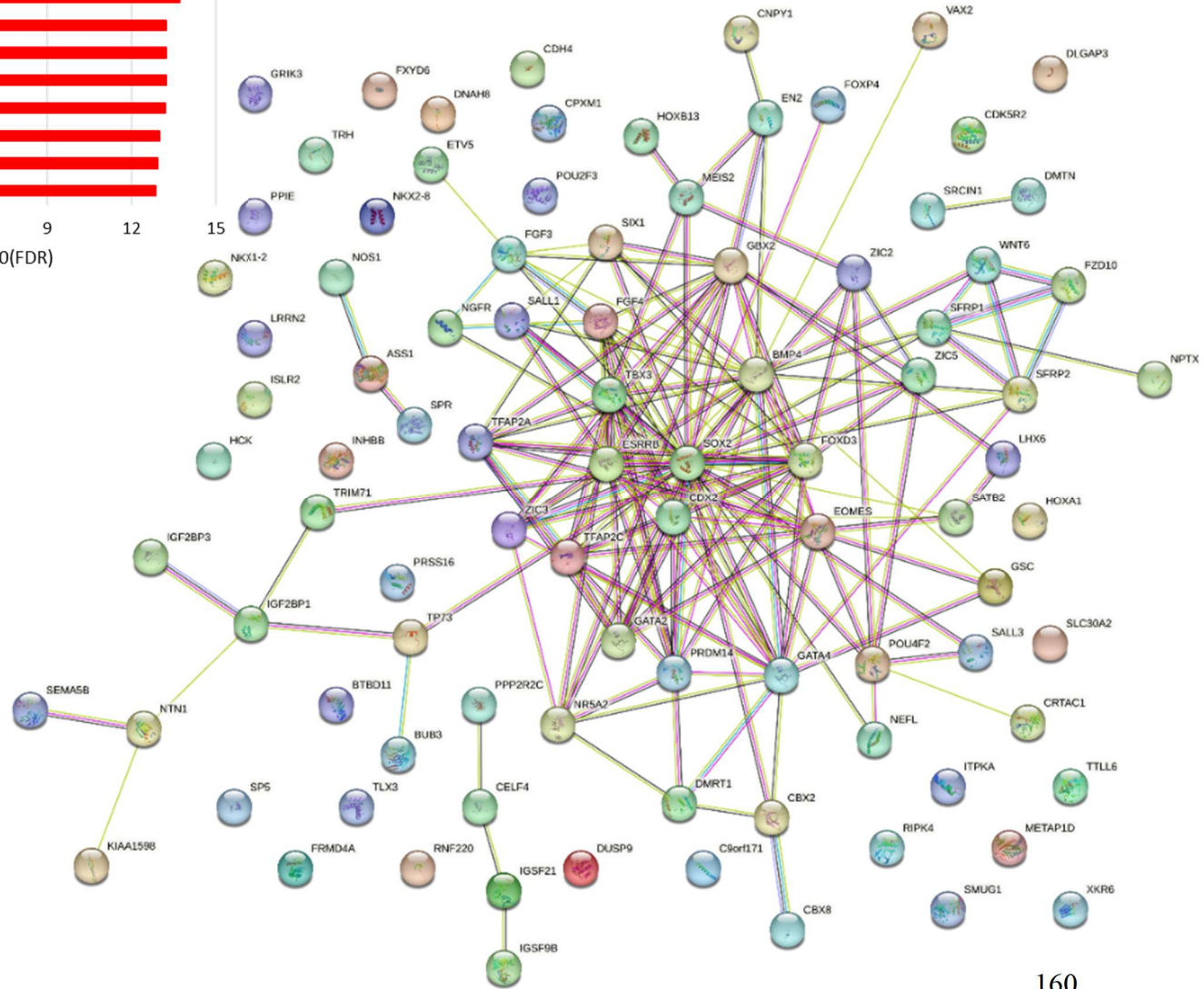


Figure 3.11A continued...

A. A2 Ground state

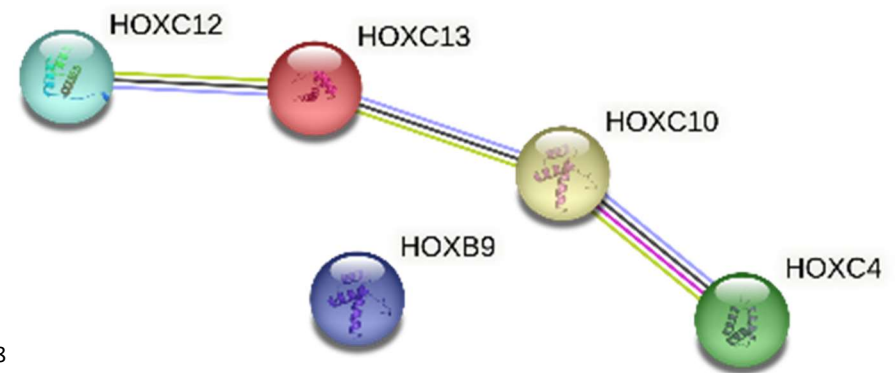
Gene Cluster 2 GO: Biological Processes



Gene Cluster 2

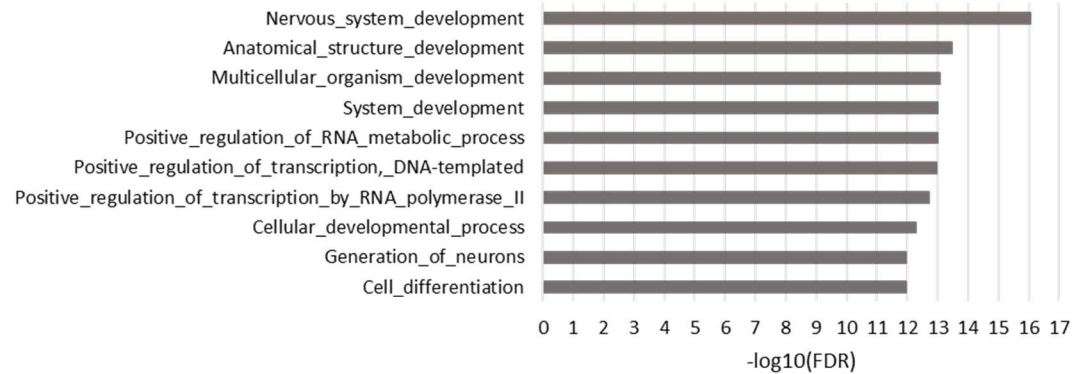
Gene	Rank by mean DS
<i>HOXC12</i>	7
<i>HOXC13</i>	9
<i>HOXC4</i>	35
<i>HOXC10</i>	46
<i>HOXB9</i>	85

Gene Cluster 2
PPI enrichment $P = 3.41^{-08}$



B. A2 Naïve state

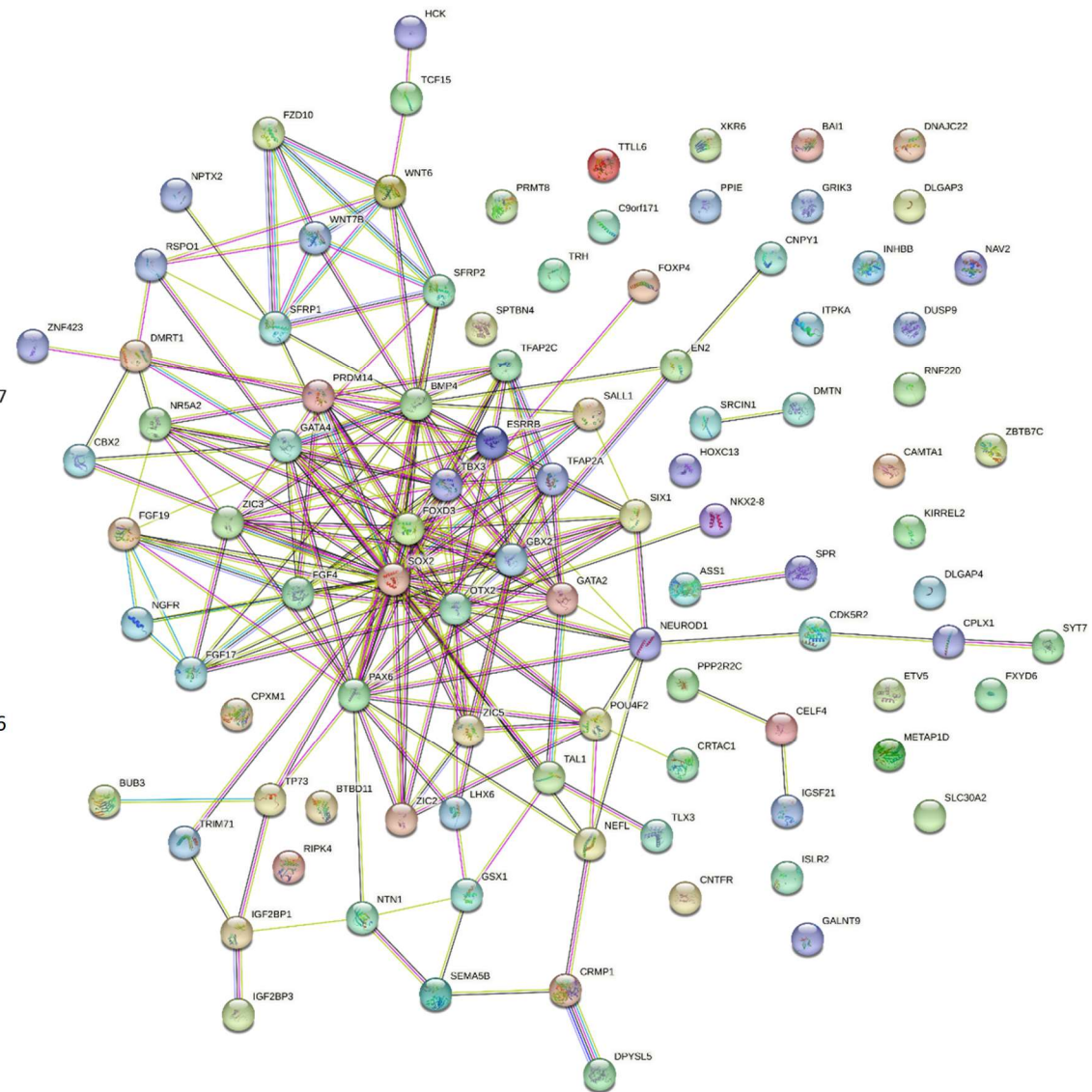
Gene Cluster 1 GO: Biological Processes



Gene Cluster 1

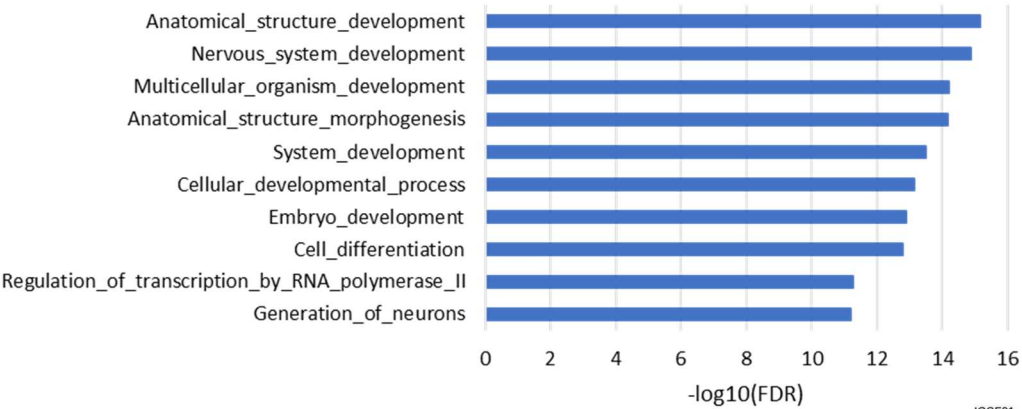
Gene	Rank by mean DS
<i>CNPY1</i>	1
<i>TBX3</i>	2
<i>RNF220</i>	3
<i>ESRRB</i>	4
<i>GBX2</i>	5
<i>EN2</i>	6
<i>NR5A2</i>	7
<i>IGF2BP1</i>	8
<i>DMRT1</i>	9
<i>INHBB</i>	10

PPI enrichment $P < 1.0^{-16}$



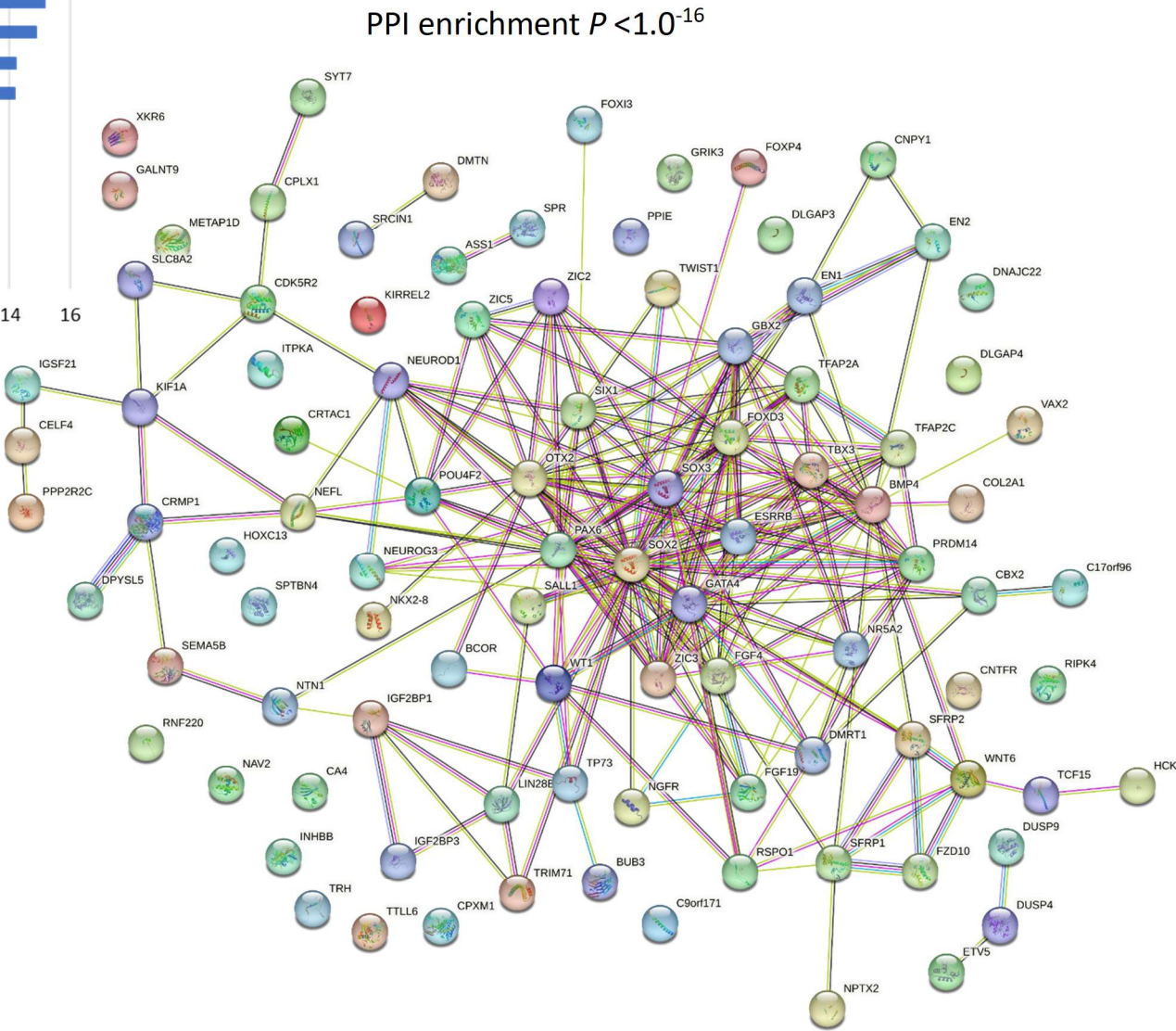
C. A2 EPLCs

Gene Cluster 1 GO: Biological Processes



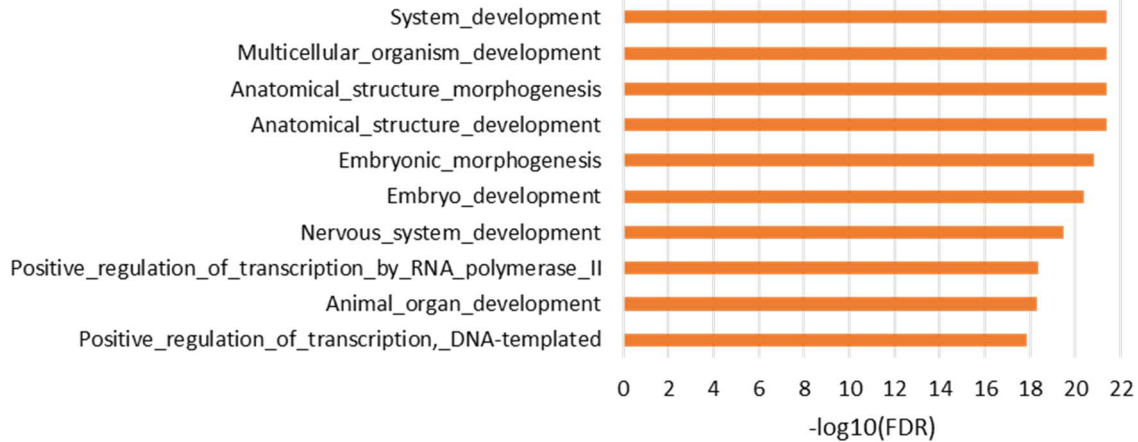
Gene Cluster 1

Gene	Rank by mean DS
CNPY1	1
OTX2	2
RNF220	3
ESRRB	4
DMRT1	5
GBX2	6
IGF2BP1	7
NR5A2	8
NKX2-8	9
EN2	10



D. EpiSCs

Gene Cluster 1 GO: Biological Processes

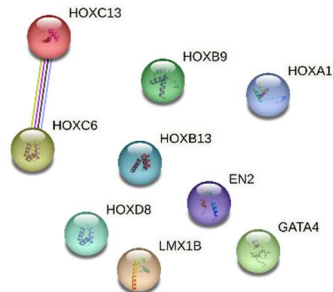


Gene Cluster 2

Gene	Rank by mean DS
GATA4	1
HOXB13	4
EN2	9
HOXA1	23
HOXC13	63
HOXB9	80
HOXC6	85
HOXD8	88
LMX1B	100

Gene Cluster 2

PPI enrichment $P = 0.0971$



Gene Cluster 1

Gene	Rank by mean DS
CNPY1	2
TBX3	5
RNF220	7
EN1	8
ESRRB	10
IGF2BP1	11
DMRT1	12
OLIG3	13
GBX2	14
CNTFR	15

Gene Cluster 1

PPI enrichment $P < 1.0^{-16}$

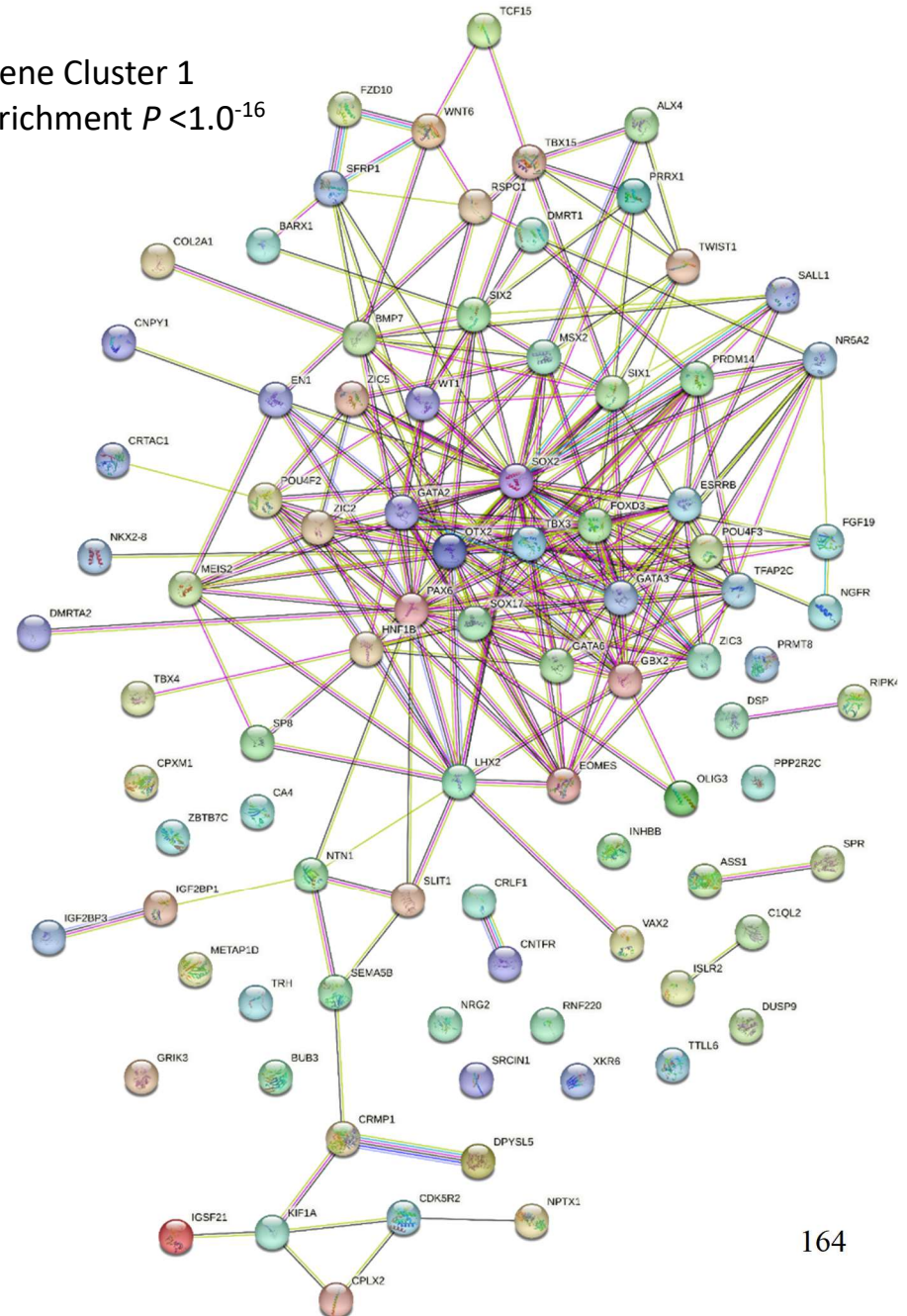


Figure 3.11 Gene ontology (GO) analysis of the top 100 ranked DS genes using TRIAGE-Parser for the A2 cell line to identify biological processes associated with the A. ground state, B. naïve state, C. EPLCs and D. EpiSCs. RNAseq data was subjected to TRIAGE transformation using Python (v 3.8.2) to produce discordance scores (DS) ranked across the entire *Mus musculus* genome. The top 100 genes based on their DS then underwent TRIAGE-Parser to both categorize these genes into clusters and undergo GO enrichment analysis on the same clusters to identify their GO: biological process terms. Clusters were identified by the percentage chance the gene is in that cluster using an arbitrary cut-off of $\geq 95\%$. The top 10 terms or less, based on their Benjamini-Hochberg adjusted P value <0.05 , or their false discovery rate (FDR), were graphed for each cluster. Protein-protein interactions (PPIs) for each cluster were predicted using the STRING database (v 11.5) with default settings. PPI enrichment FDR P value <0.001 indicates significance and, therefore, biologically meaningful PPIs.

3.3.2. *Relative expression and TRIAGE analysis of L-Proline metabolism-associated genes in cells across the pluripotency spectrum*

L-Proline metabolites can feed into both anaerobic and aerobic metabolism (Figure 3.1) (Phang *et al.*, 2008; Pandhare *et al.*, 2009; Casalino *et al.*, 2011; Phang, 2012; Hancock *et al.*, 2016). L-Proline is also known to drive differentiation of pluripotent mESCs to a second pluripotent cell population called EPLCs. This *in vitro* transition recapitulates the ICM-to-primitive ectoderm transition *in vivo* (Rathjen *et al.*, 2003; Washington *et al.*, 2010; Glover *et al.*, 2022). The *in vivo* transition is important, as primitive ectoderm is ‘primed’ for gastrulation, where pluripotency is lost and the three multipotent germ layers are formed, which go on to generate all the cells of the developing embryo and adult (Snow, 1977; Tam and Behringer, 1997).

To explore this further, Figure 3.12 shows differential gene expression analysis of RNAseq data generated for mESCs and EpiSCs, focused on the expression of genes encoding for enzymes associated with L-Proline metabolism. Significant differences in expression are denoted by an asterisk, which represents a Benjamini-Hochberg adjusted *P* value <0.05, regardless of log₂ fold changes in expression, as any relative changes may be biologically meaningful (Koch *et al.*, 2018).

Expression of these L-Proline catabolism genes is most similar between naïve cells and EPLCs, with few significant differences in expression. EPLCs, however, do have significantly higher expression of *Slc38a2*, which encodes for the SNAT2 transporter that allows L-Proline to enter cells, and possibly *Prodh*, which encodes for the rate limiting enzyme, POX.

On the other hand, the ground-state profile is very different to the naïve-state profile. Expression of almost all these catabolism genes in the ground state is either upregulated or downregulated compared to the naïve state, with upregulation of the expression of genes responsible for glutathione and α -KG production, and downregulation of *Glul*, responsible for L-Glutamine production.

The expression profile of EpiSCs vs naïve state is also distinct, with almost all EpiSCs genes expressed significantly less than in the naïve state.

Figure 3.13 shows where these L-Proline metabolism genes lie in the DS ranking following TRIAGE transformation. All these genes across all stages of pluripotency and cell lines have a DS rank >1359, which suggests L-Proline catabolism is not directly responsible for cell fate (Shim *et al.*, 2020). L-Proline catabolism may, however, affect the expression of key genes which are responsible for cell fate.

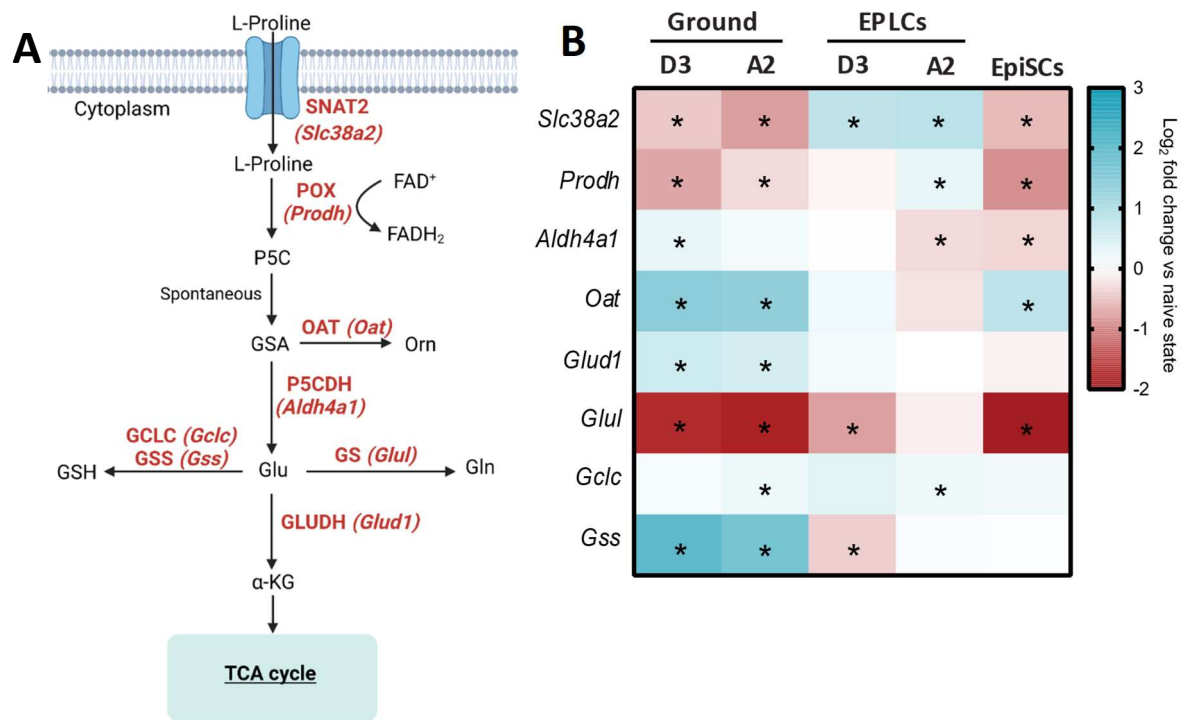


Figure 3.12 Relative expression of genes encoding enzymes associated with proline catabolism across the pluripotency spectrum. A. L-Proline enters the cell via the SNAT2 transporter and is metabolised into pyrroline-5-carboxylate (P5C) by proline oxidase (POX). P5C spontaneously converts to glutamate-5-semialdehyde (GSA), and from there to glutathione (GSH) or L-Glutamine (Gln) or α -ketoglutarate (α -KG). Enzymes are shown in red with corresponding genes in parentheses. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and for A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: SNAT2: Sodium-coupled neutral amino acid transporter 2; *Slc38a2*: Solute Carrier Family 38 Member 2; POX: proline oxidase; *Prodh*: proline dehydrogenase; OAT: Ornithine aminotransferase; P5CDH: pyrroline-5-carboxylate dehydrogenase; *Aldh4a1*: Aldehyde Dehydrogenase 4 Family, Member A1; *Gclc*: Glutamate-cysteine ligase catalytic subunit; *Gss*: Glutathione synthetase; GS: Glutamine synthase; *Glul*: Glutamate-ammonia ligase; *Glud1* (GLUDH): Glutamate dehydrogenase 1; Glu: L-Glutamate.

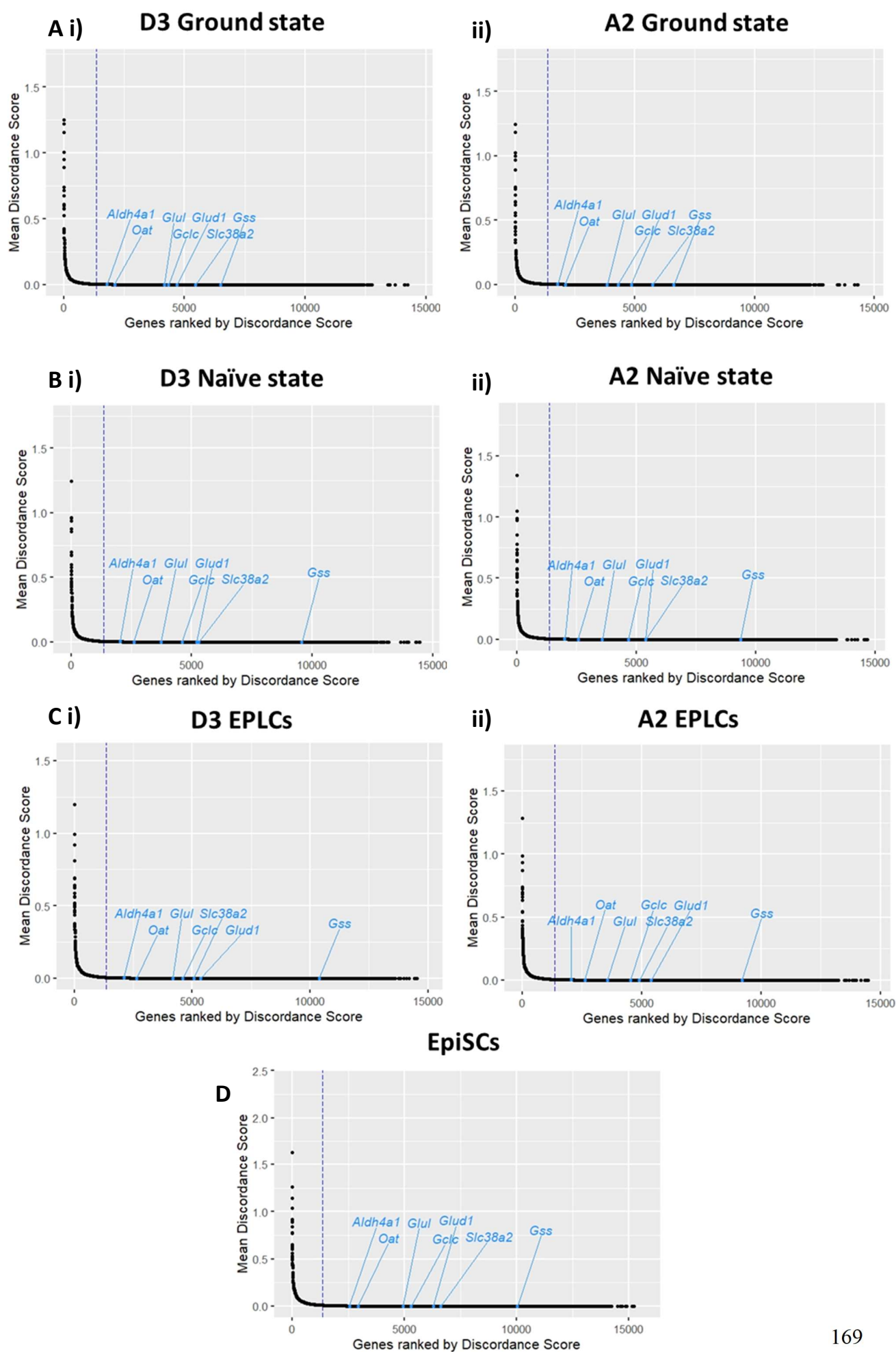


Figure 3.13 TRIAGE transformation of genes encoding enzymes associated with proline catabolism across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. Proline catabolism genes are labelled. Note no DS could be calculated for the *Prodh* gene, and thus was not ranked.

Figure 3.14 shows differential changes in expression for genes encoding enzymes associated with L-Proline biosynthesis across the pluripotency spectrum.

EPLCs have downregulated expression of both *Aldh18a1* and *Pycr1* compared to naïve cells, suggesting they undertake reduced L-Proline synthesis in the face of exogenously added L-Proline.

On the other hand, ground state cells have stable expression of these genes compared to naïve cells. The expression of *Oat* is upregulated suggesting these ground state cells may preferentially generate L-Proline synthesis from ornithine.

The expression profile of EpiSCs vs naïve state is much more distinct. Expression of almost all these genes is either upregulated or downregulated compared to the naïve state, with upregulation of the expression of genes responsible for glutamic semialdehyde and glutamate production, and downregulation of *Pycr1*, responsible for L-Proline synthesis.

TRIAGE transformation of genes encoding enzymes associated with L-Proline biosynthesis shows all genes having DS rank >1359 across all stages of pluripotency and cell lines (Figure 3.15).

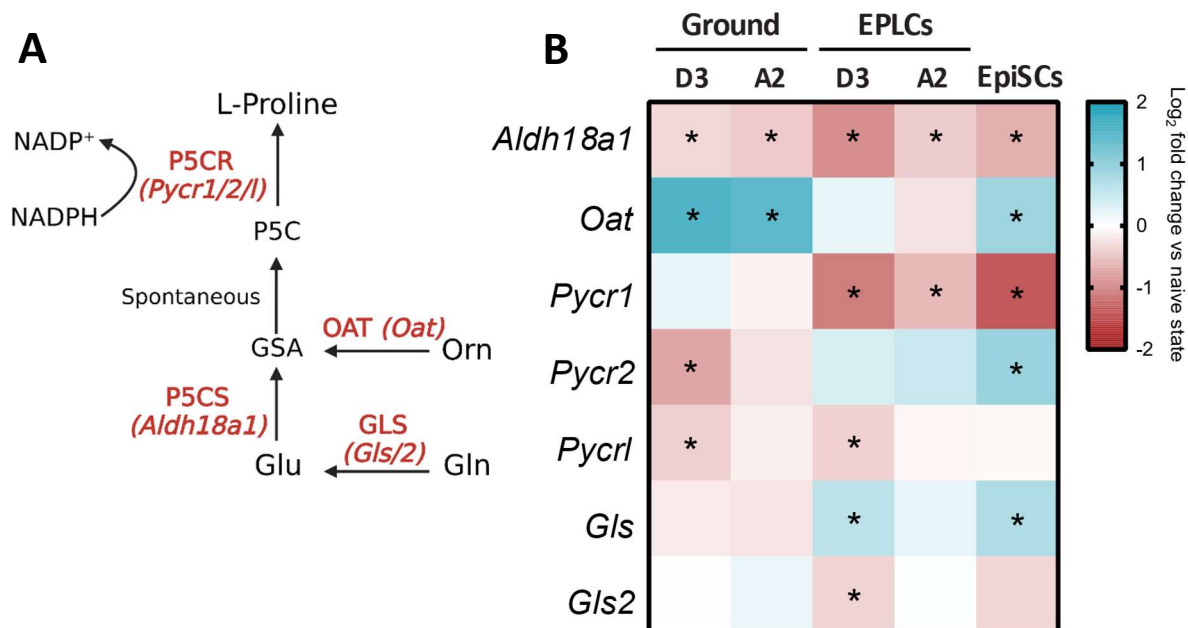


Figure 3.14 Relative expression of genes encoding enzymes associated with proline biosynthesis across the pluripotency spectrum. **A.** L-Proline can be produced from either L-Glutamine (Gln) or L-Ornithine (Orn). Enzymes are shown in red with corresponding genes in parentheses. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: *Pycr1/2/l*: Pyrroline-5-carboxylate reductase 1/2/-like; *Oat*: ornithine aminotransferase, mitochondrial; GSA: glutamic semialdehyde; *Aldh18a1*: Aldehyde dehydrogenase 18 family, member A1; *Gls/2*: Glutaminase/glutaminase 2.

Figure 3.15 TRIAGE transformation of genes encoding enzymes associated with proline biosynthesis across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. Proline biosynthesis genes are labelled.

3.3.3. *Relative expression and TRIAGE analysis of genes associated with glycolysis in cells across the pluripotency spectrum*

The metabolism of L-Proline also mediates other metabolic processes via the proline cycle, which include some processes involved in glycolysis (Figure 3.1) (Phang et al., 2008; Pandhare et al., 2009; Casalino et al., 2011; Phang, 2012; Hancock et al., 2016). Therefore, we focused on interrogating genes encoding for enzymes associated with glycolysis across the pluripotency spectrum and in response to exogenously added L-Proline.

Figure 3.16 shows differential expression of glycolysis genes in ground-state mESCs to EpiSCs. EPLCs have a distinct profile to naïve cells, generally having upregulated expression of glycolysis enzyme genes.

Ground state also has a distinct profile to naïve cells: There is upregulated expression of genes near the start of the glycolysis pathway compared to naïve, and downregulated expression further downstream.

EpiSCs also have a distinct profile to naïve cells, with most glycolysis genes being downregulated.

TRIAGE transformation of genes encoding enzymes associated with glycolysis show all genes having DS rank >1359 across all stages of pluripotency and cell lines, except for *Slc16a3* (the monocarboxylate transporter of lactate) in EPLCs (Figure 3.17).

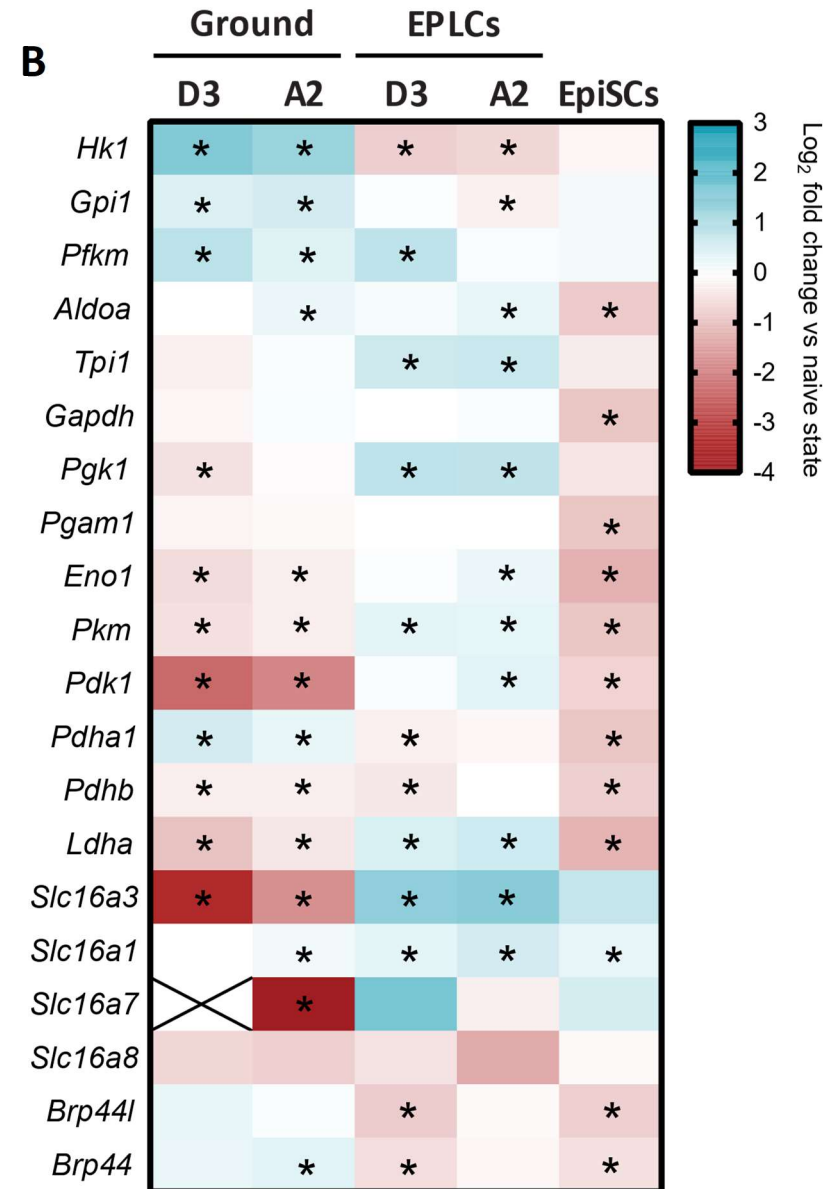
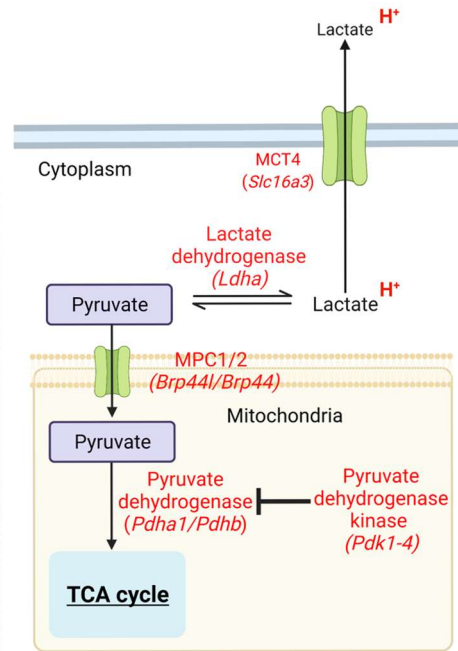
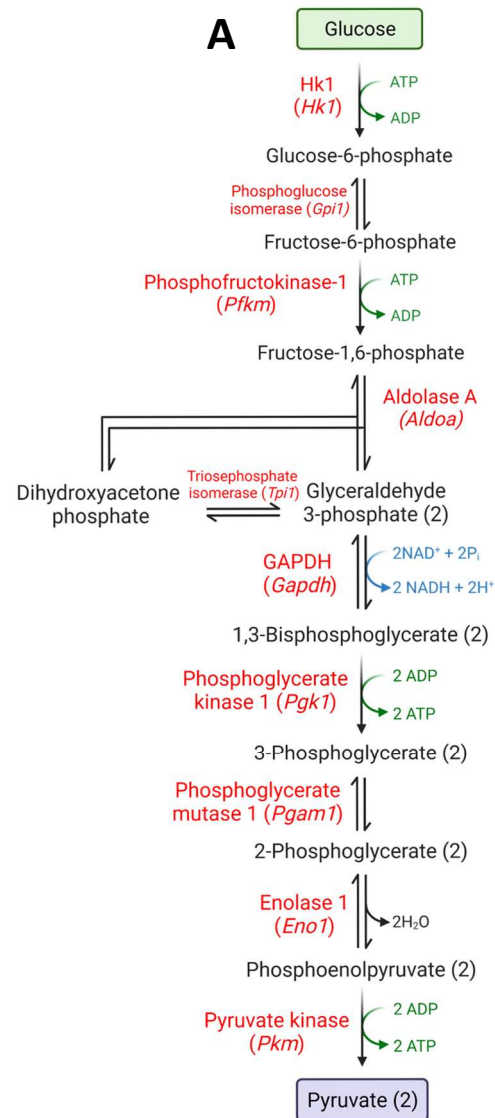


Figure 3.16 Expression of genes encoding enzymes associated with glycolysis across the pluripotency spectrum. **A.** Glucose gets converted into two pyruvate molecules via intermediates. Pyruvate either converts to lactate, which gets exported out of the cell via monocarboxylate transporter 4 (MCT4) or transported into the mitochondria via mitochondrial pyruvate carrier 1/2 (MPC1/2) to contribute to the TCA cycle. Enzymes are shown in red with corresponding genes in parentheses. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and for A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$. *Hk1*: Hexokinase 1.

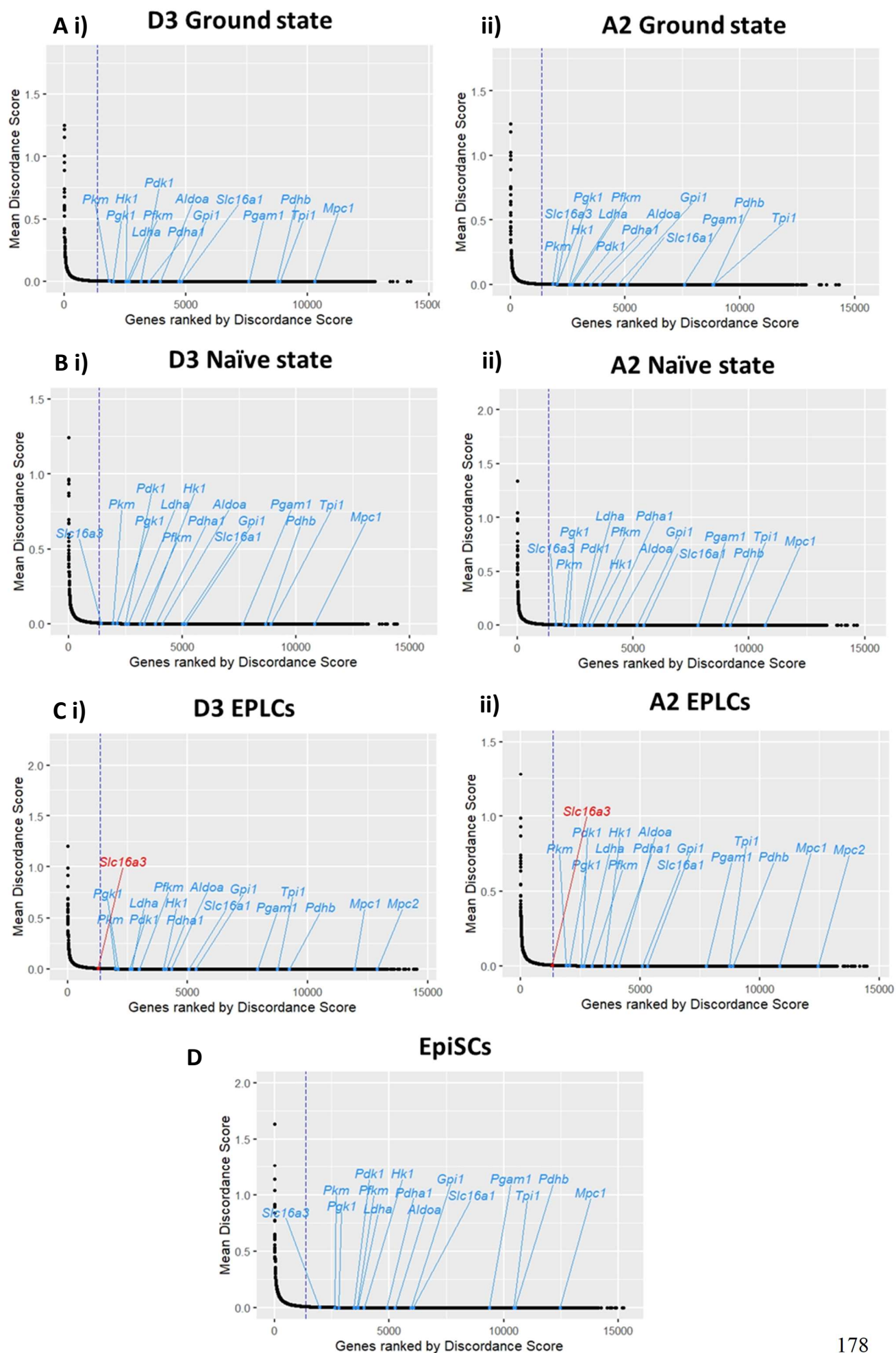


Figure 3.17 TRIAGE transformation of genes encoding enzymes associated with glycolysis across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. Glycolysis genes are labelled, with blue genes indicating a DS rank > 1359 , and red genes indicating a DS rank ≤ 1359 .

3.3.4. *Relative expression and TRIAGE analysis of genes associated with the pentose phosphate pathway in cells across the pluripotency spectrum*

The proline cycle is also redox coupled to the oxidative phase of the pentose phosphate pathway (PPP) (Figure 3.1) (Phang *et al.*, 2010; Hancock *et al.*, 2016). In addition, the PPP contributes to protein and nucleotide synthesis in highly proliferative cells, including cancer cells and ESCs (DeBerardinis *et al.*, 2008; Ito and Suda, 2014; Phang *et al.*, 2015; Harvey *et al.*, 2016; Browne, 2017; Tohyama *et al.*, 2017; Shparberg *et al.*, 2019b; Tsogtbaatar *et al.*, 2020). Thus, we interrogated genes encoding for enzymes associated with the PPP across the pluripotency spectrum and in response to exogenously added L-Proline.

Figure 3.18 shows the expression of these PPP genes is most similar between naïve cells and EPLCs, with few significant differences in expression. However, EPLCs show upregulation of rate-limiting enzyme gene *G6pdx*, indicating increasing PPP activity in response to exogenously added L-Proline.

Expression of these PPP genes is also similar in ground state relative to naïve state but there is some changes in expression of genes in the top half of the pathway.

In contrast, the expression profile of EpiSCs vs naïve state is more distinct. Most genes are either significantly upregulated or downregulated in EpiSCs, including now in the non-oxidative phase.

TRIAGE transformation of genes encoding enzymes associated with the PPP show all genes having DS rank >1359 across all stages of pluripotency and cell lines, except for *Tktl1* in EpiSCs (Figure 3.19).

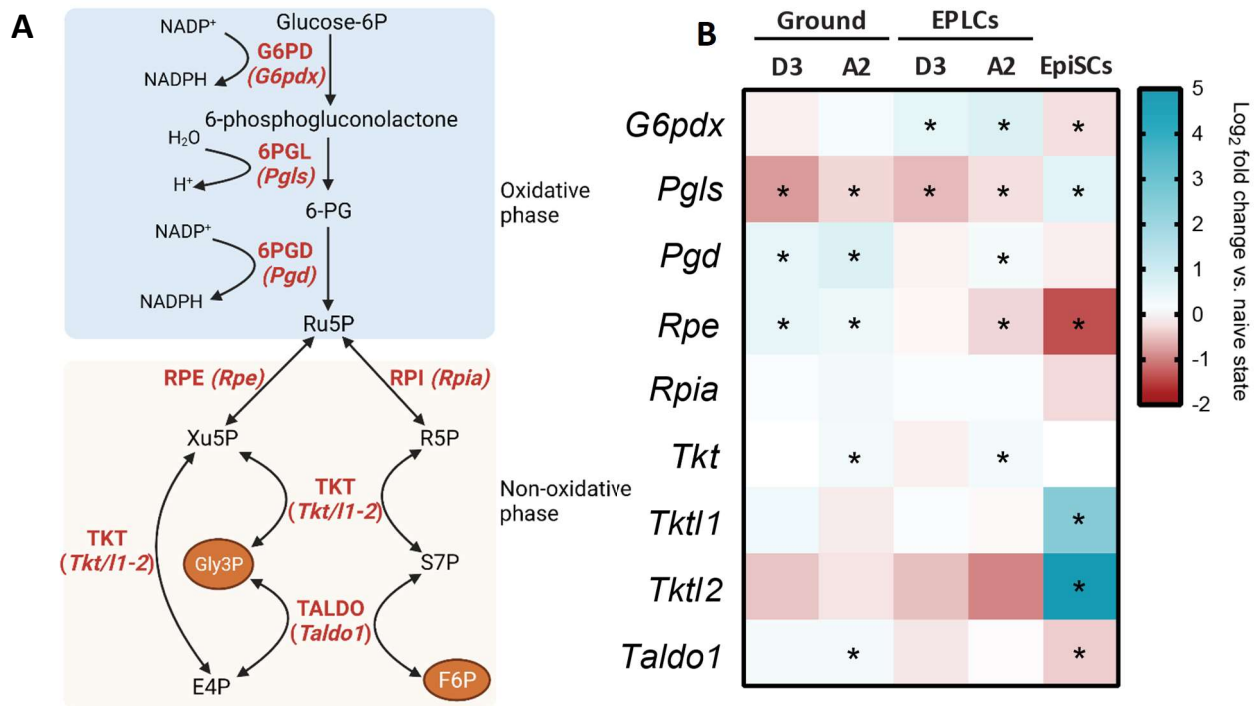


Figure 3.18 Expression of genes encoding enzymes associated with oxidative and non-oxidative phases of the PPP across the pluripotency spectrum. **A.** The PPP and all the associated enzymes, which are shown in red with corresponding genes in parentheses. Metabolite intermediates in orange circles are involved in glycolysis. NADP^+ and H_2O are used as oxidizing agents in the oxidative phase. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and for A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: Glucose-6P: glucose-6-phosphate; 6-PG: 6-phosphogluconate; Ru5P: ribulose-5-phosphate; Xu5P: xylulose-5-phosphate; Gly3P: glyceraldehyde 3-phosphate; E4P: erythrose-4-phosphate; R5P: ribose-5-phosphate; S7P: sedoheptulose-7-phosphate; F6P: fructose-6-phosphate; G6PD (*G6pdx*): glucose-6-phosphate dehydrogenase; 6PGL (*Pgls*): 6-phosphogluconolactonase; 6PGD (*Pgd*): 6-phosphogluconate dehydrogenase; RPE: ribulose phosphate 3-epimerase/ phosphopentose epimerase; RPI (*Rpia*): ribulose phosphate 3-epimerase/ phosphopentose epimerase; TKT: transketolase; *Tktl1-2*: transketolase-like 1/2; TALDO (*Taldo1*): transaldolase.

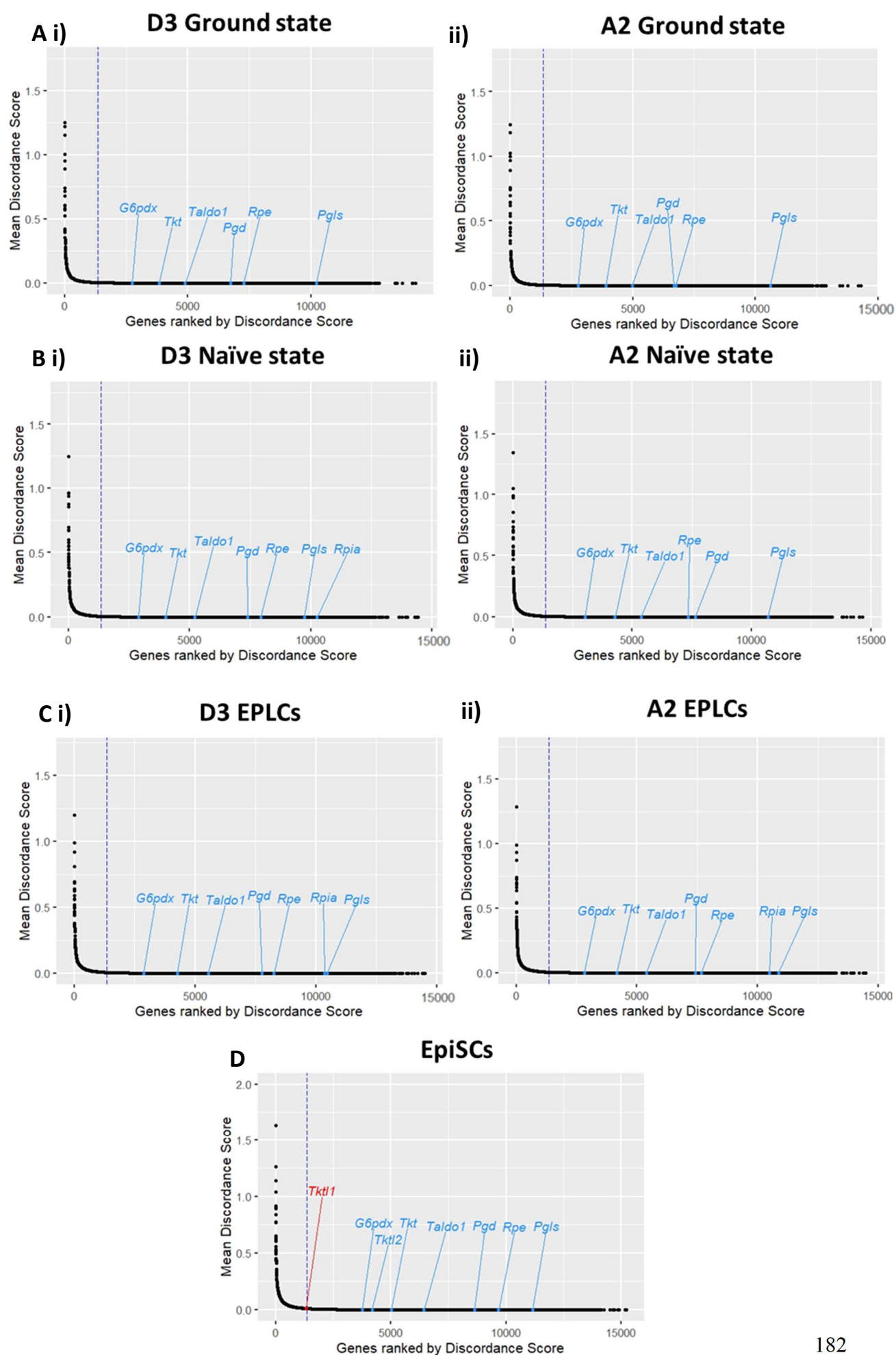


Figure 3.19 TRIAGE transformation of genes encoding enzymes associated with oxidative and non-oxidative phases of the PPP. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. PPP genes are labelled, with blue genes indicating a DS rank > 1359 , and red genes indicating a DS rank ≤ 1359 .

3.3.5. *Relative expression and TRIAGE analysis of genes associated with the TCA cycle and the electron transport chain in cells across the pluripotency spectrum*

The TCA cycle and electron transport chain (ETC) form the basis of aerobic metabolism in mitochondria. As pluripotent cells transition from ground state to primed, there is an increased reliance on anaerobic metabolism (Zhou *et al.*, 2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022). The proline cycle also can mediate TCA cycle activity by enhancing production of pyruvate from glucose (Figure 3.1). Therefore, we interrogated genes encoding for enzymes involved in the TCA cycle and complexes involved in the ETC across the pluripotency spectrum.

Figure 3.20 shows differential gene expression analysis of RNAseq data generated for mESCs and EpiSCs, focused on the expression of genes encoding for enzymes associated with the TCA cycle.

Ground-state mESCs and EPLCs show a distinct profile to naïve state, with some genes having upregulated expression and some downregulated.

EpiSCs also display a distinct profile. The expression for most genes is downregulated. This includes the expression of *Pcx*, the enzyme for which, pyruvate carboxylase, allows entry of pyruvate into the TCA cycle through formation of oxaloacetate and *Pdha1*, the enzyme of which, pyruvate dehydrogenase, allows entry of pyruvate into the cycle through formation of acetyl-CoA (Figure 3.1 and 3.20A).

TRIAGE transformation of genes encoding enzymes and complexes associated with the TCA cycle show all genes, except for *Idh3g*, having DS ranks >1359 across all stages of pluripotency and cell lines (Figure 3.21).

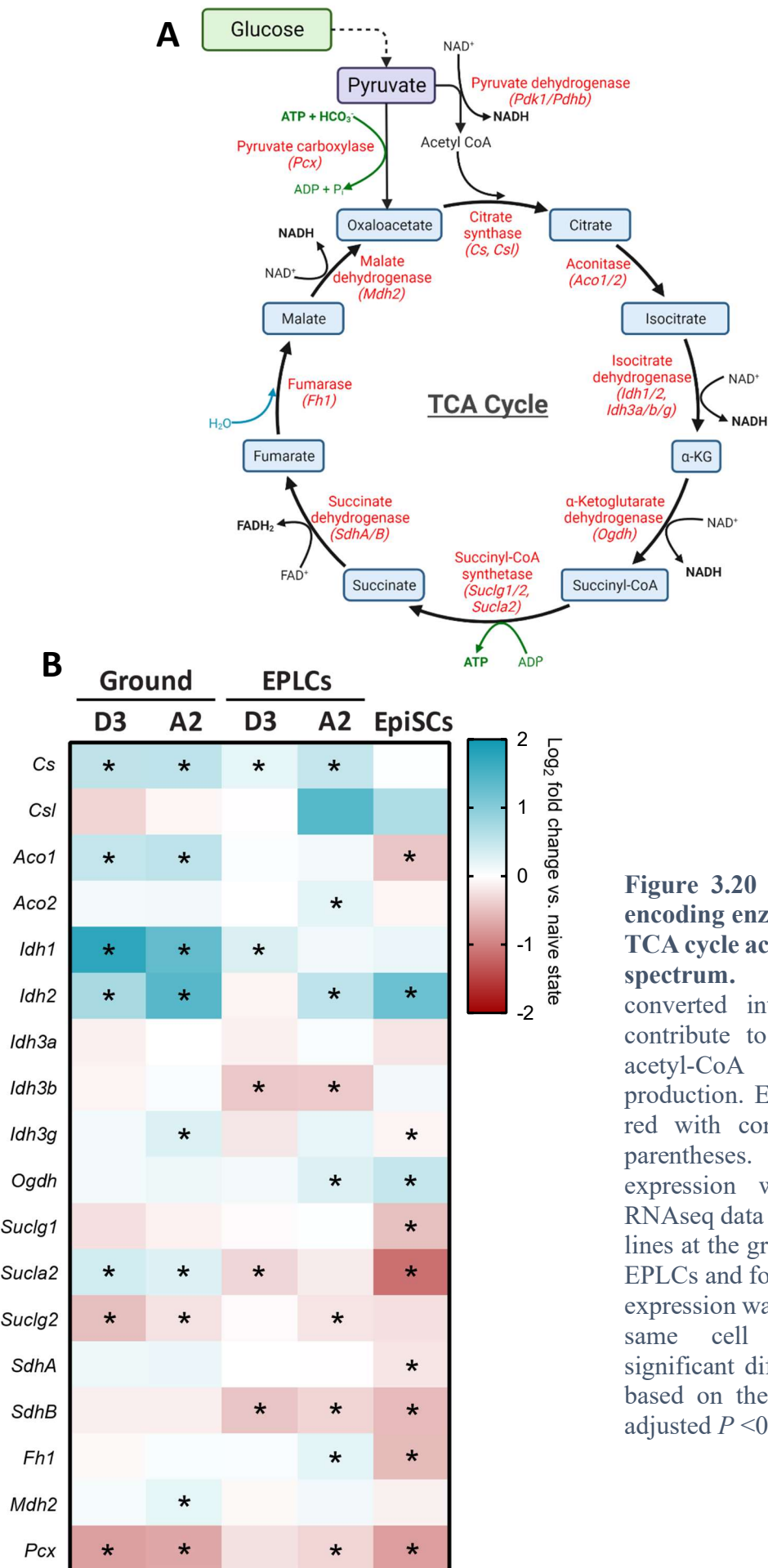


Figure 3.20 Expression of genes encoding enzymes associated with TCA cycle across the pluripotency spectrum. **A.** Glucose gets converted into pyruvate to then contribute to the TCA cycle via acetyl-CoA or oxaloacetate production. Enzymes are shown in red with corresponding genes in parentheses. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and for A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$.

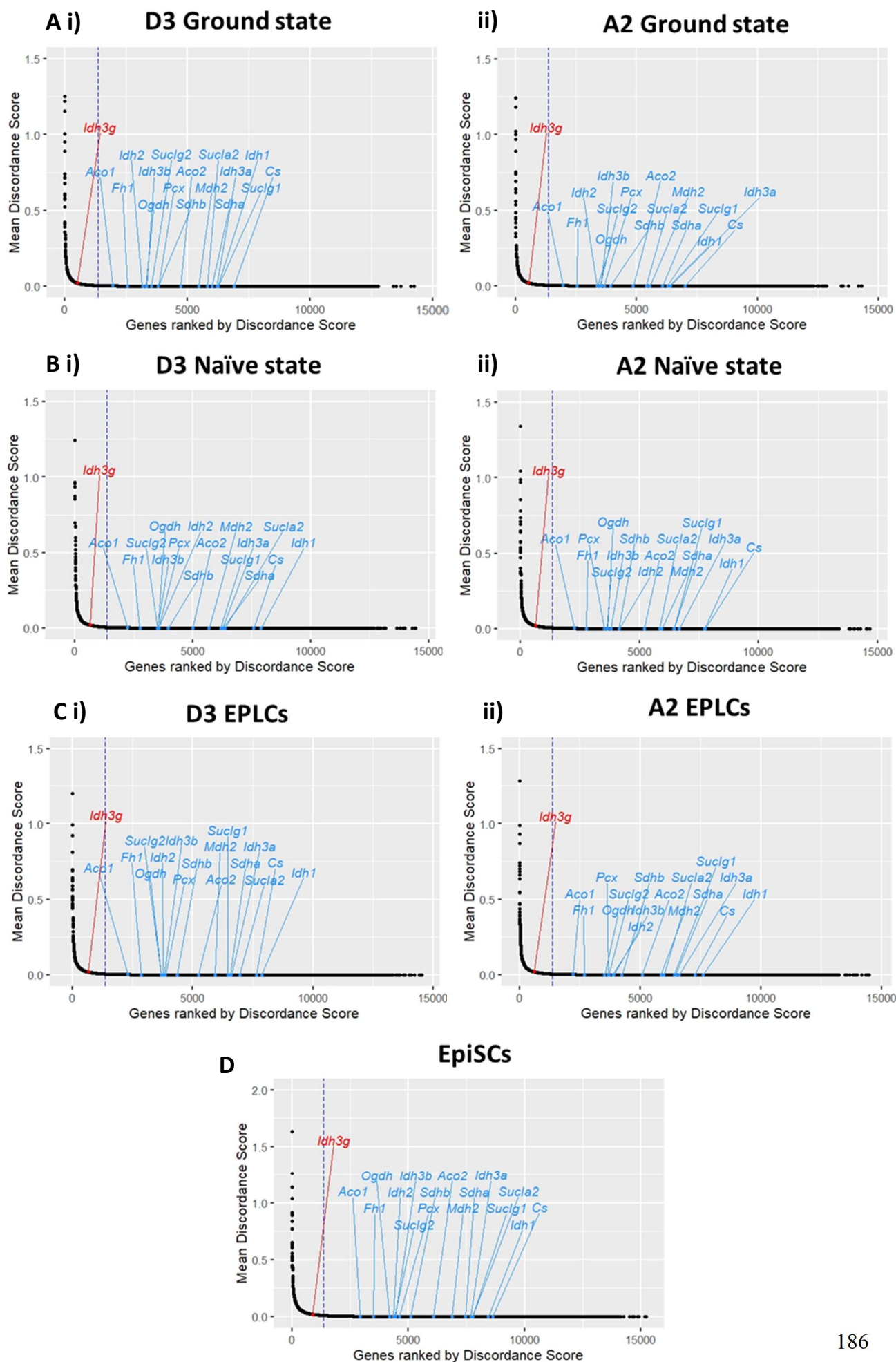


Figure 3.21 TRIAGE transformation of genes encoding enzymes associated with the TCA cycle across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. TCA cycle genes are labelled, with blue genes indicating a DS rank > 1359 , and red genes indicating a DS rank ≤ 1359 .

Figure 3.22 shows differential changes in expression for genes encoding enzymes and complexes associated with the ETC across the pluripotency spectrum.

Here it's apparent that expression of these genes is, in general, highest in the naïve state and 'falls away' towards the ground-state and EPLC/EpiSC ends of the pluripotency spectrum. The exception is *Ucp2* but in fact the uncoupling protein, which is the product of this gene, is responsible for lowering mitochondrial membrane potential and activity. This suggests that ground state and EpiSCs are utilising increased expression of *Ucp2* to limit ETC-mediated production of ATP, and this has the additional benefit of reducing oxidative stress (Rousset *et al.*, 2004; Jastroch *et al.*, 2010; Zhang *et al.*, 2011).

In addition, the expression of *Prodh*, whose product is POX, is also downregulated in ground-state and EpiSCs. POX converts L-Proline to P5C and in addition produces high-energy electrons for the ETC in the form of FADH₂. This, in these same two states of the pluripotency spectrum, ETC activity, and oxidative stress, may be lowered through reduced expression of the POX gene.

TRIAGE transformation of genes encoding enzymes associated with the ETC shows all genes, except for *Ndufa10*, having DS ranks >1359 across all stages of pluripotency and cell lines (Figure 3.23).

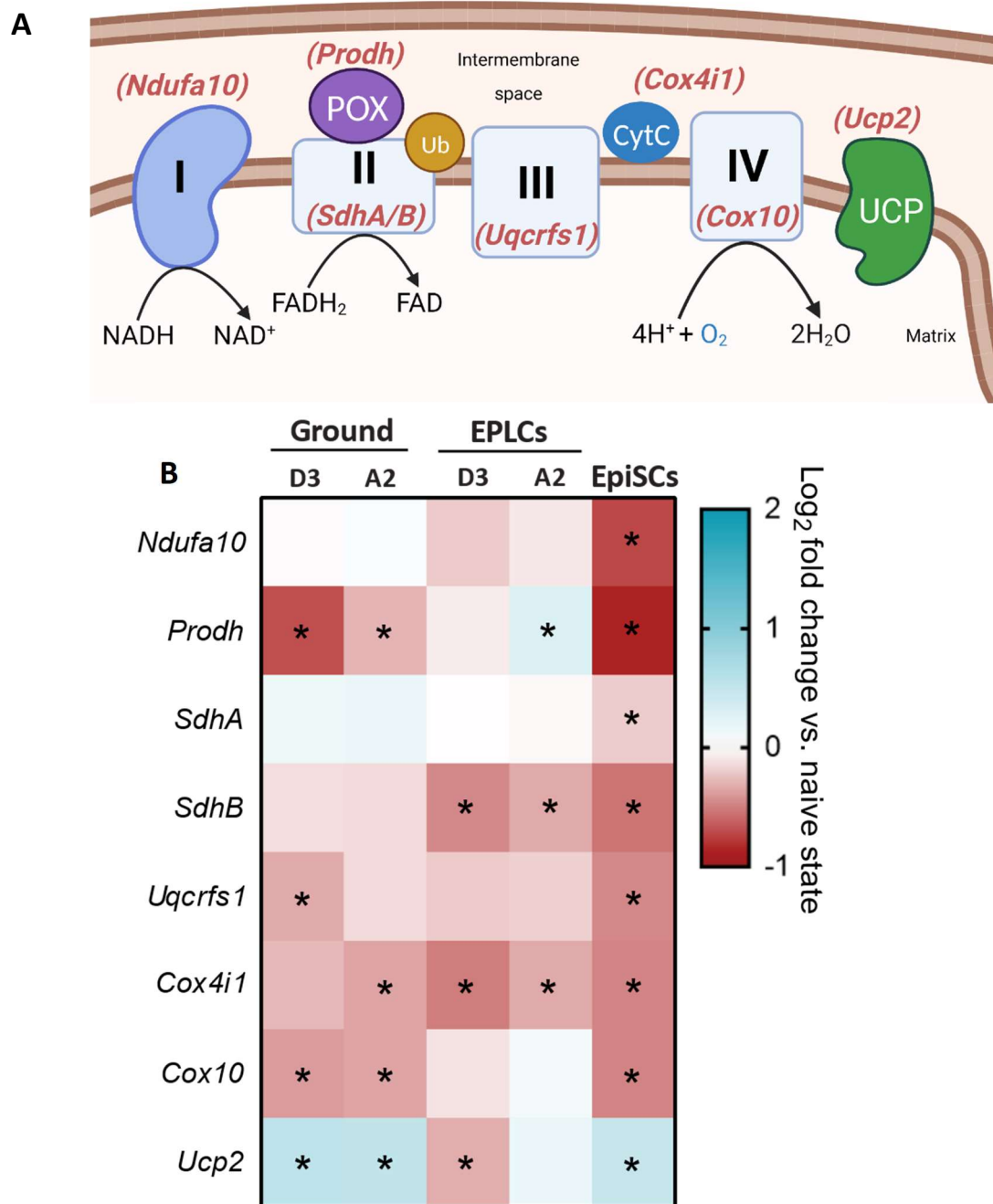


Figure 3.22 Expression of genes encoding complexes and enzymes involved in the ETC across the pluripotency spectrum. **A.** NADH produced from the TCA cycle is used in Complex I to initiate the proton gradient. Components of the ETC include other complexes (II-IV), as well as proline oxidase (POX) and cytochrome-c oxidase (CytC) activity, and uncoupling protein (UCP). Enzymes are shown in red with corresponding genes in parentheses. **B.** Relative gene expression was calculated from RNAseq data for the D3 and A2 cell lines at the ground and naïve states, EPLCs and for A2 EpiSCs. Relative expression was compared within the same cell line. * Indicates significant difference in expression based on the Benjamini-Hochberg adjusted $P < 0.05$.

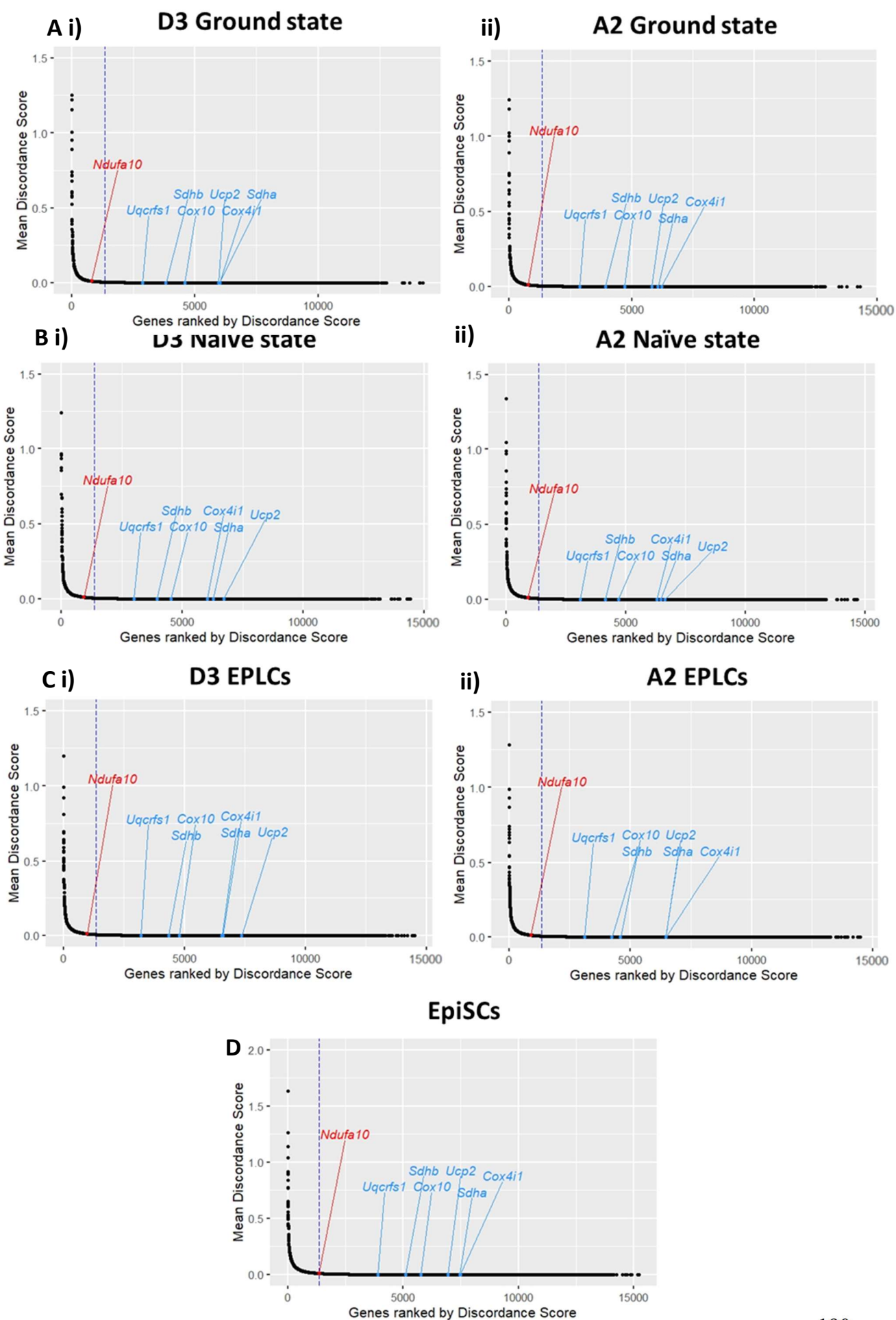


Figure 3.23 TRIAGE transformation of genes encoding complexes and enzymes associated with the ETC across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. ETC genes are labelled, with blue genes indicating a DS rank > 1359 , and red genes indicating a DS rank ≤ 1359 .

3.3.6. *Metabolic profiles of ground- and naïve-state mESCs and EPLCs in response to L-Proline*

As we previously mentioned, L-Proline metabolites can feed into both anaerobic and aerobic processes (Figure 3.1) (Phang *et al.*, 2008; Pandhare *et al.*, 2009; Casalino *et al.*, 2011; Phang, 2012; Hancock *et al.*, 2016). ESCs have a preference for anaerobic metabolism to support their growth and survival in comparison to somatic mammalian cells which generally prefer to generate ATP by aerobic metabolism (Rehman, 2010; Varum *et al.*, 2011; Panopoulos *et al.*, 2012). However, the transition of ESCs to primed EpiSCs is marked by changes in metabolic profile to a more glycolytic one (Zhou *et al.*, 2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022). This aligns with the *in vivo* process: When the blastocyst, with its ICM, implants into the uterine epithelium, glucose becomes the major energy source for the embryo and is preferentially metabolised anaerobically to generate ATP (Folmes *et al.*, 2012; Lees *et al.*, 2017).

We then questioned the following: How does metabolism change from the ground to the formative/intermediate (EPLC) stages of the pluripotency spectrum? How do these stages of pluripotency respond to exogenously added L-Proline? Does the L-Proline-mediated mESC-to-EPLC transition stimulate anaerobic or aerobic processes? To answer these questions, we used the Seahorse extracellular flux assay to compare metabolic profiles of ground- and naïve-state mESCs and EPLCs (Figure 3.2). The assay will help show how cells across the pluripotency spectrum respond to exogenously added L-Proline (Figure 3.2). We also profile their basal metabolic rates and maximum respiratory capacities to capture how these cells metabolise at the “baseline level” and their maximum possible response following uncoupling of the ETC, respectively (Divakaruni *et al.*, 2014; Martano *et al.*, 2019).

Changes in oxygen consumption rates (OCR) and extracellular acidification rates (ECAR) were calculated and are indicative of aerobic and anaerobic metabolism, respectively (Divakaruni *et al.*, 2014).

Figure 3.24A shows typical raw values of oxygen concentration and extracellular pH. 12 measurements of each are taken over 2 min at 7 min intervals, giving rise to a typical saw-tooth pattern.

Panel B shows OCR and ECAR plots calculated from the raw values. 5 values of OCR and ECAR are taken before any injections into the wells so that baseline values for each stage of pluripotency can be measured. The last 3 measurements before the first injection are averaged to give these baseline values. Values are corrected for differences in confluence as calculated from IncuCyte images. Note that there are large differences in baseline OCR (but not ECAR) values for the different pluripotency states with the ground state value being ~300 pmol/min, naïve state ~200 pmol/min and EPLCs ~100 pmol/min (also see Figure 3.25A).

Panel B also shows 3 consecutive injections of L-Proline to give final concentration of 0.4, 1 and 3 mM, followed by a final injection of FCCP to uncouple the ETC so that spare mitochondrial capacity can be measured.

Panels C and D, show two different types of corrected OCR and ECAR values. In C, ΔOCR (%) and ΔECAR (%) show the percentage change in value compared to averaged baseline value (i.e., the average of the three values before the first injection). In D, $\Delta\Delta\text{OCR}$ (%) and $\Delta\Delta\text{ECAR}$ (%) values make a further correction by subtracting the ΔOCR (%) and ΔECAR (%) values obtained with injection of XF medium alone from the ΔOCR (%) and ΔECAR (%) values with injection of L-Proline (or FCCP).

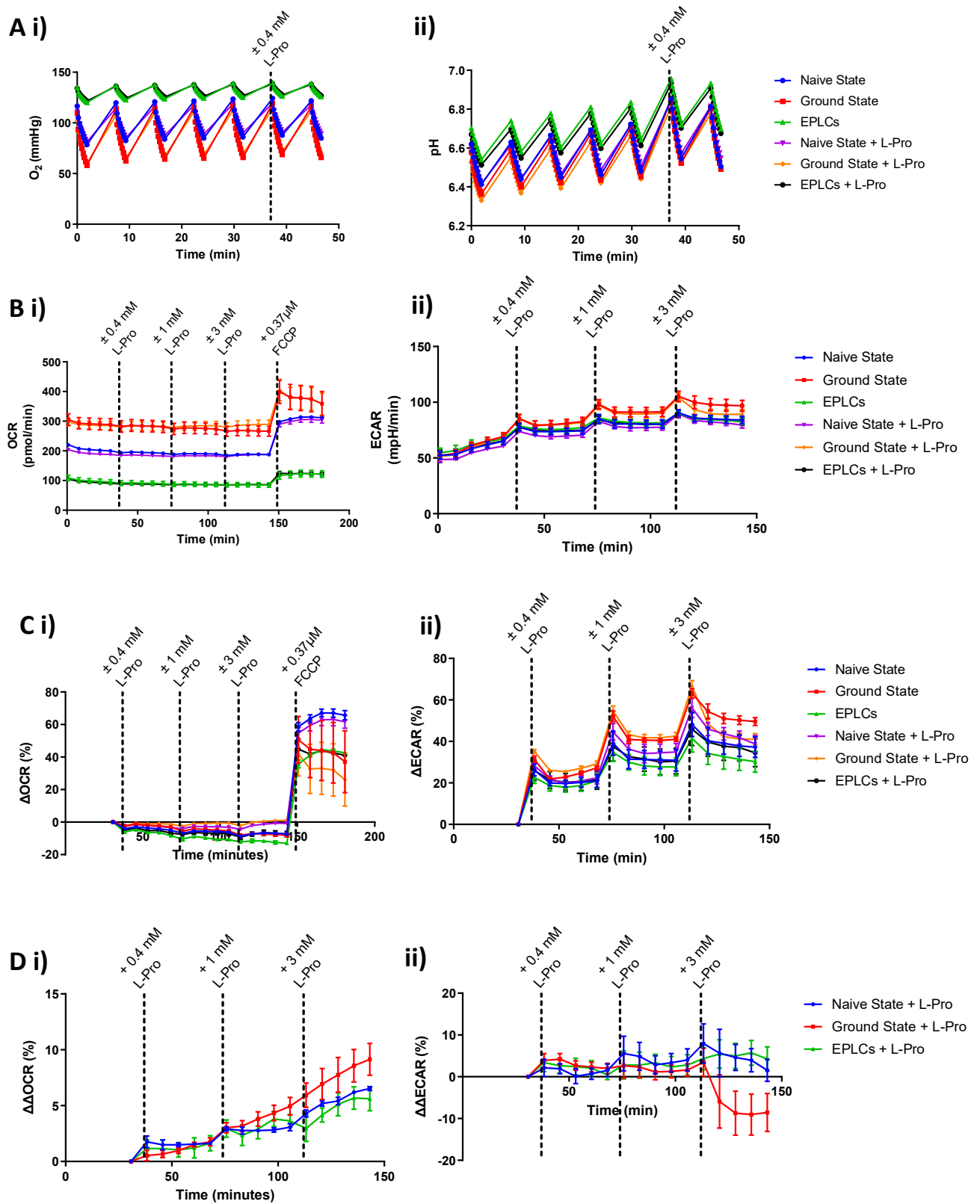


Figure 3.24 Measurement of oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) in ground- and naïve-state mESCs and EPLCs using the Seahorse extracellular flux assay. **A.** Raw Seahorse data showing changes in **i)** oxygen concentration and **ii)** pH with time. 12 measurements are taken over 2 min at 7 min intervals. The plots show typical results for ground- and naïve-state mESCs and EPLCs. At 37 min, XF medium alone or XF medium + L-Pro (to give a final concentration of 0.4 mM) was injected into the wells, as shown in the key to the right. **B.** Typical plots of OCR and ECAR against time. To calculate OCR and ECR, the first and last point in each set of 12 points in **A** is removed and then linear regressed. After 5 OCR/ECAR measurements, wells were injected at the times shown either with XF medium alone or XF medium + L-Pro (to give final concentrations of L-Pro of 0.4, 1 and 3 mM). OCR was also measured after a final injection of the ETC uncoupler, FCCP (0.37 μ M final). Values are corrected for differences in confluence as measured from IncuCyte images. **C.** Plots of OCR and ECAR converted into **i)** Δ OCR (%) and **ii)** Δ ECAR (%), respectively. These values represent the percentage change from the mean of the 3 measurements before the first injection. For the FCCP injection, it represents the percentage change from the mean of the last 3 measurements after the 3rd injection. **D.** Plots of Δ OCR (%) and Δ ECAR (%) converted to produce **i)** $\Delta\Delta$ OCR (%) and **ii)** $\Delta\Delta$ ECAR (%). In the conversion, the Δ OCR (%) and Δ ECAR (%) values with XF medium injection are subtracted from those values for L-Proline (or FCCP) injection. This corrects for the fact that injection of medium alone can have a small effect, particularly for ECAR. Error bars in **B-D** represent SEM across $n \geq 4$ independent experiments.

Figure 3.25 shows baseline OCR and ECAR values for ground- and naïve-state mESCs and EPLCs. ECAR values are not significantly different between the states but OCR values drop by a third as ground state transitions to naïve state and another third for EPLCs.

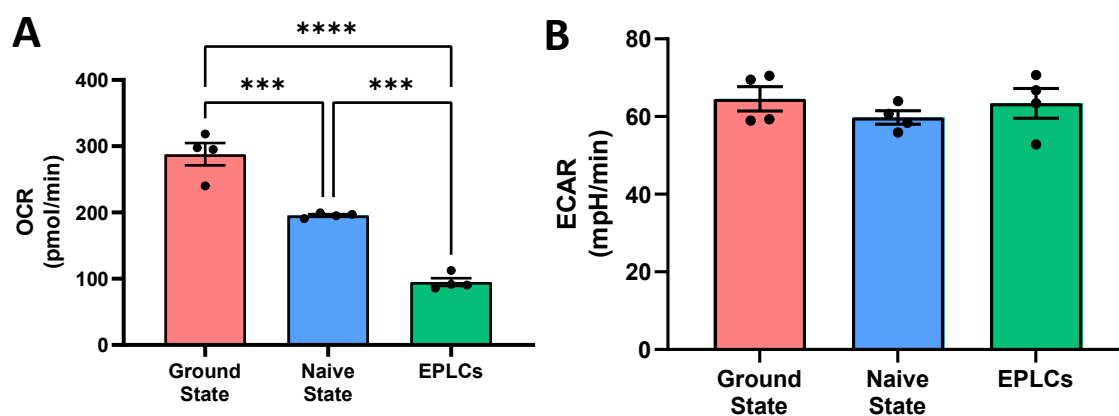


Figure 3.25 Basal metabolic rates of ground and naïve state ESCs and EPLCs. D3 mESCs were used to produce ground- and naïve-state mESCs and EPLCs (see Figure 3.2). 5 baseline metabolic rate measurements were made, and an average of the last 3 metabolic rates were used to produce **A.** average OCR and **B.** average ECAR. Error bars represent mean $\pm$ SEM from $n = 4$ independent experiments. Statistical analysis was conducted using a one-way ANOVA with Tukey's *post hoc* test. *** $P < 0.001$, **** $P < 0.0001$.

Figure 3.26 shows the corrected OCR and ECAR values for injections of L-Proline and the uncoupler, FCCP. None of the cells across the pluripotency spectrum responded to injection of L-Proline at any of the three concentrations (Panels A–C). Panel D shows there are, however, significant differences in mitochondrial spare capacity. Ground and naïve mESCs have ~60% spare capacity, whereas EPLCs have only ~40%.

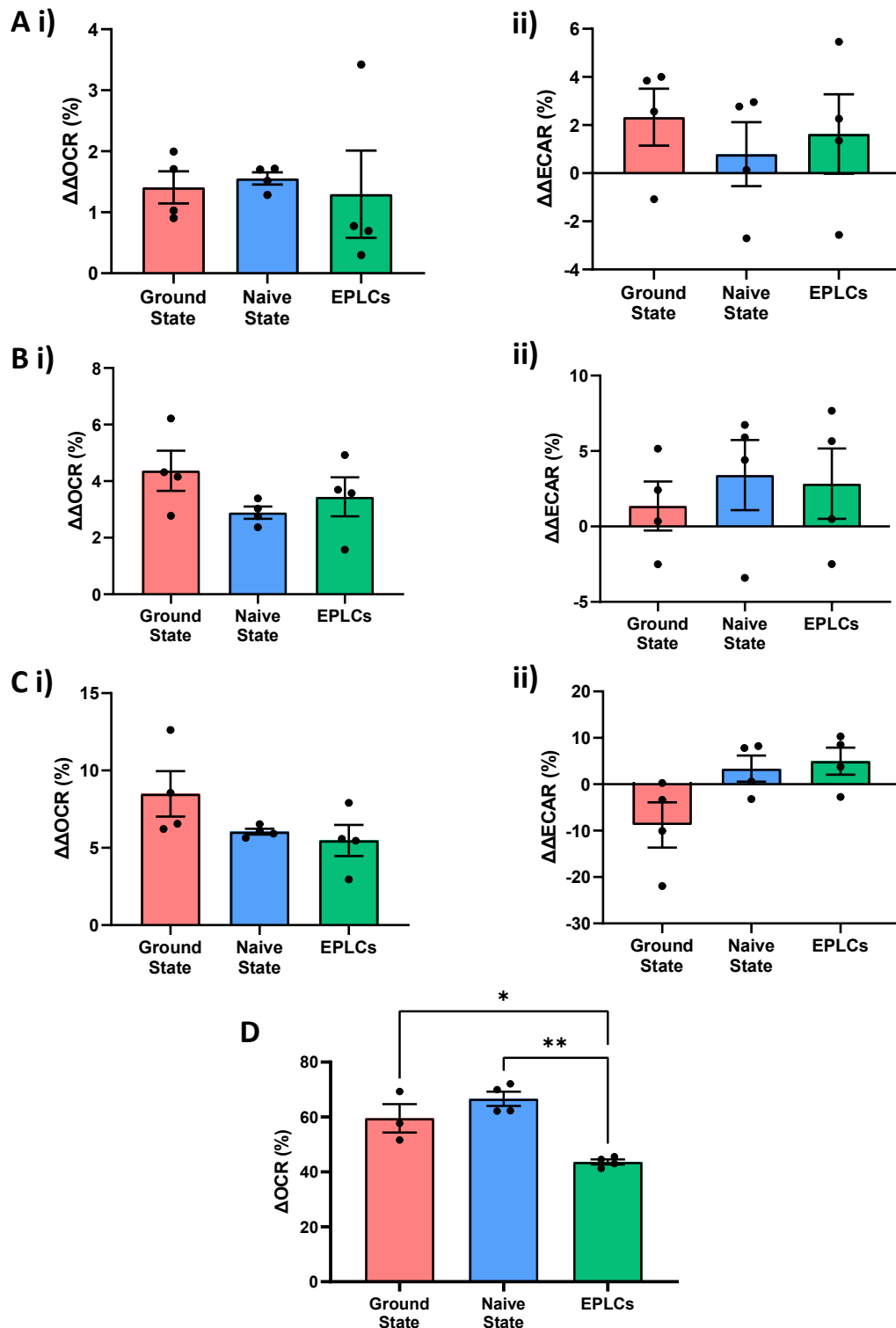


Figure 3.26 Corrected OCR and ECAR values of cells across the pluripotency spectrum in response to acute injections of L-Proline and mitochondrial uncoupler FCCP. D3 mESCs were used to produce ground- and naïve-state mESCs and EPLCs (see Figure 3.2). The last 3 values before the first injection were averaged to produce the baseline values, which were then used to calculate **i)** $\Delta\Delta OCR$ (%) and **ii)** $\Delta\Delta ECAR$ (%), which represent the percentage change compared to the baseline value corrected for any changes that occur when XF medium alone is injected. **A.** 1st injection (0.4 mM L-Proline). **B.** 2nd injection (1 mM L-Proline). **C.** 3rd injection (3 mM L-Proline). **D.** Spare respiratory capacity following injection of FCCP. Bars represent mean $\pm$ SEM, $n \geq 3$ independent experiments. Statistical analysis was conducted using a one-way ANOVA with a Tukey's *post hoc* test. * $P < 0.05$.

3.3.7. *The role of L-Proline metabolism in the mESC-to-EPLC transition*

We used 3 parameters – changes in colony morphology, doubling time and gene expression – to assess the role of L-Proline metabolism in mESC-to-EPLC transition (Rathjen *et al.*, 1999, 2002; Washington *et al.*, 2010; Casalino *et al.*, 2011). Given that we wanted to inhibit the key enzyme POX, which catalyses L-Proline to P5C, with inhibitors that must be dissolved in DMSO, we first tested if the presence of DMSO affected the L-Proline-mediated transition.

Figure 3.27 shows that when mESCs self-renew in the naïve state, neither a reduction in the concentration of LIF from 1000 U/mL to 330 U/mL nor the addition of 0.1% DMSO to culture in 330 U/mL LIF had any significant effect on colony morphology or cell doubling time. Most colonies remain round and dome-shaped, typical of self-renewing naïve mESCs (Casalino *et al.*, 2011; Glover, 2018). Some were scored as “Intermediate” because they have small protrusions from the edge, while only a very small percentage were scored as flattened monolayers with irregular edges, which are typical of EPLC colonies (Casalino *et al.*, 2011; Glover, 2018).

In terms of doubling time, note that Panel Ci shows there are two growth phases: The first in which the cells are seeding onto the tissue culture plate, and the second, exponential phase, from ~32 h onwards, which was used for calculating doubling time.

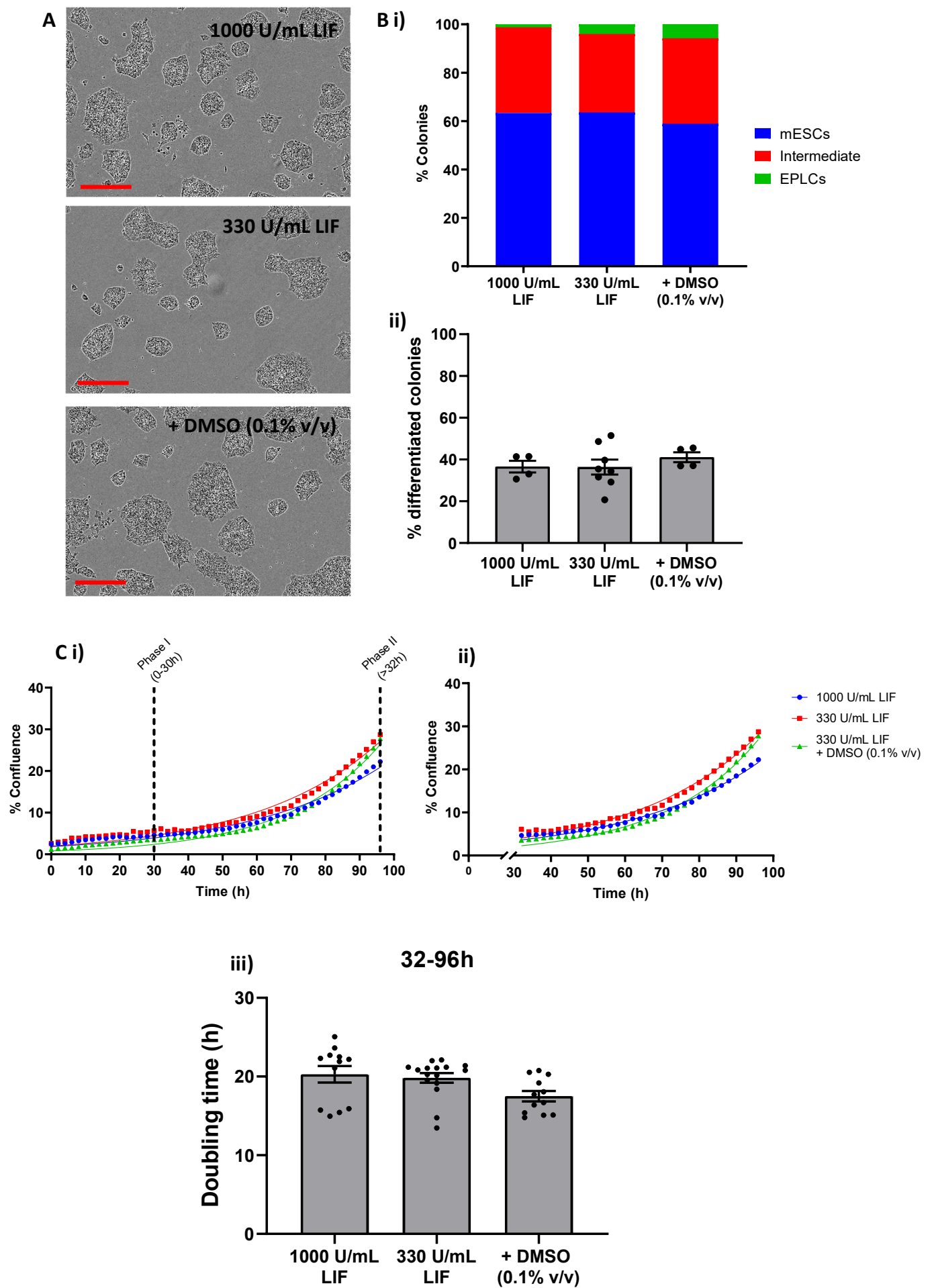


Figure 3.27 Colony morphology and doubling time profiles of naïve mESCs cultured in LIF and 0.1% DMSO vehicle control. D3 mESCs were cultured in 1000 U/mL LIF or 330 U/mL LIF $\pm$ 0.1% DMSO for 4 days. **A.** Representative images of each condition using IncuCyte ZOOM at 10x magnification at 4 days. Scale bars = 300 μ m. **B. i)** Plots of colony morphology scores. In **ii)**, % differentiated colonies refers to colonies with have an EPLC morphology plus those whose morphology lies between mESCs and EPLCs. Data analysed using one-way ANOVA with Dunnett's *post hoc* test relative to 330 U/mL LIF only condition. Bars represent mean $\pm$ SEM, $n \geq 4$. **C.** Growth curves plotted as % confluence against time. The solid lines in **i)** and **ii)** show the regressed fit of the doubling time equation (see Section 2.7.4). **iii)** Bar graph of the doubling time from 32-96 h for each condition. Data are analysed using a one-way ANOVA with Dunnett's *post hoc* test compared to 330 U/mL LIF only condition. Bars represent mean $\pm$ SEM, $n \geq 12$.

Figure 3.28 shows no significant changes in expression of pluripotency, ground-state or EPLC marker genes in the presence of DMSO, except for *Fgf5*. However, the difference is small (~1.4 fold), confined to this one gene, and not likely to be biologically meaningful given that none of the other genes are affected and also given that change in expression of *Fgf5* in moving from naïve ESCs to EPLCs is of the order of 50-fold (Glover *et al.*, 2022). This indicates we can proceed to the use of POX inhibitors without being concerned with effects of the vehicle control.

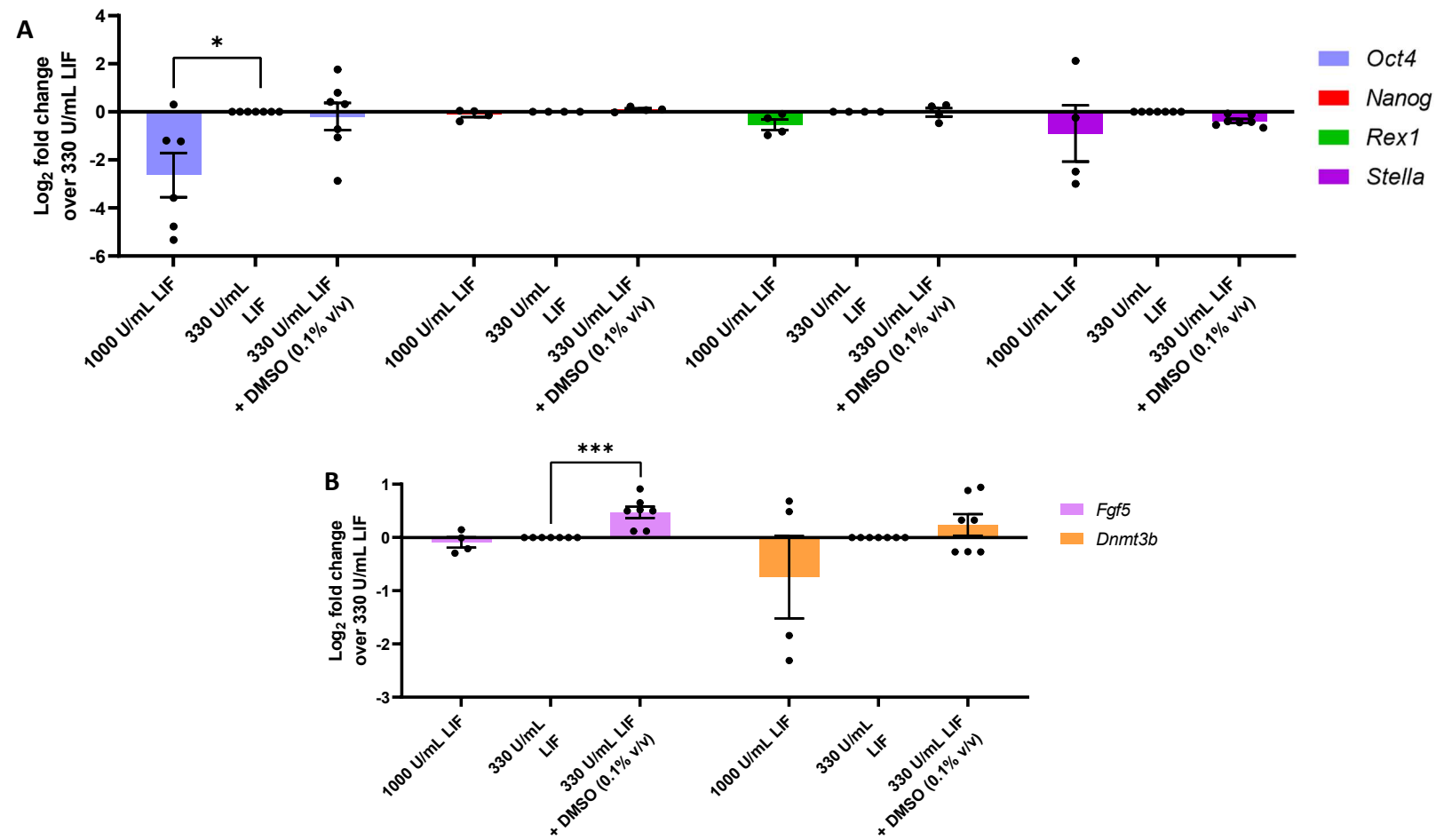


Figure 3.28 Gene expression profiles of naïve mESCs cultured in LIF and 0.1% DMSO vehicle control. D3 mESCs were cultured in 1000 U/mL LIF or 330 U/mL LIF $\pm$ 0.1% DMSO for 4 days. Samples were subjected to qPCR, with gene expression normalised to *18s* as the reference gene. Relative expression compared to 330 U/mL LIF only of: **A.** Pluripotency-associated genes *Oct4*, *Nanog* and *Rex1* and ground-state marker *Stella*. **B.** EPLC markers *Fgf5* and *Dnmt3b*. One-way ANOVA was used with Dunnett's *post hoc* test. Bars represent mean $\pm$ SEM, $n \geq 4$, * $P < 0.05$, *** $P < 0.001$.

3.3.8. *Effect of POX inhibitors on the L-Proline-mediated mESC-to-EPLC transition*

Both DHP and THFA are competitive inhibitors of POX and therefore should prevent POX-mediated conversion of L-Proline to P5C (Tanner, 2008; Phang *et al.*, 2010; Casalino *et al.*, 2011).

Figure 3.29 shows that the addition of L-Proline promotes differentiation to EPLC-like colony morphology and reduces doubling time, as found in other studies (Washington *et al.*, 2010; Casalino *et al.*, 2011; Tan *et al.*, 2011).

The addition of DHP or THFA further promoted the formation of L-Proline-induced colony morphology (Panel B) but did not affect L-Proline-mediated doubling time (Panel C).

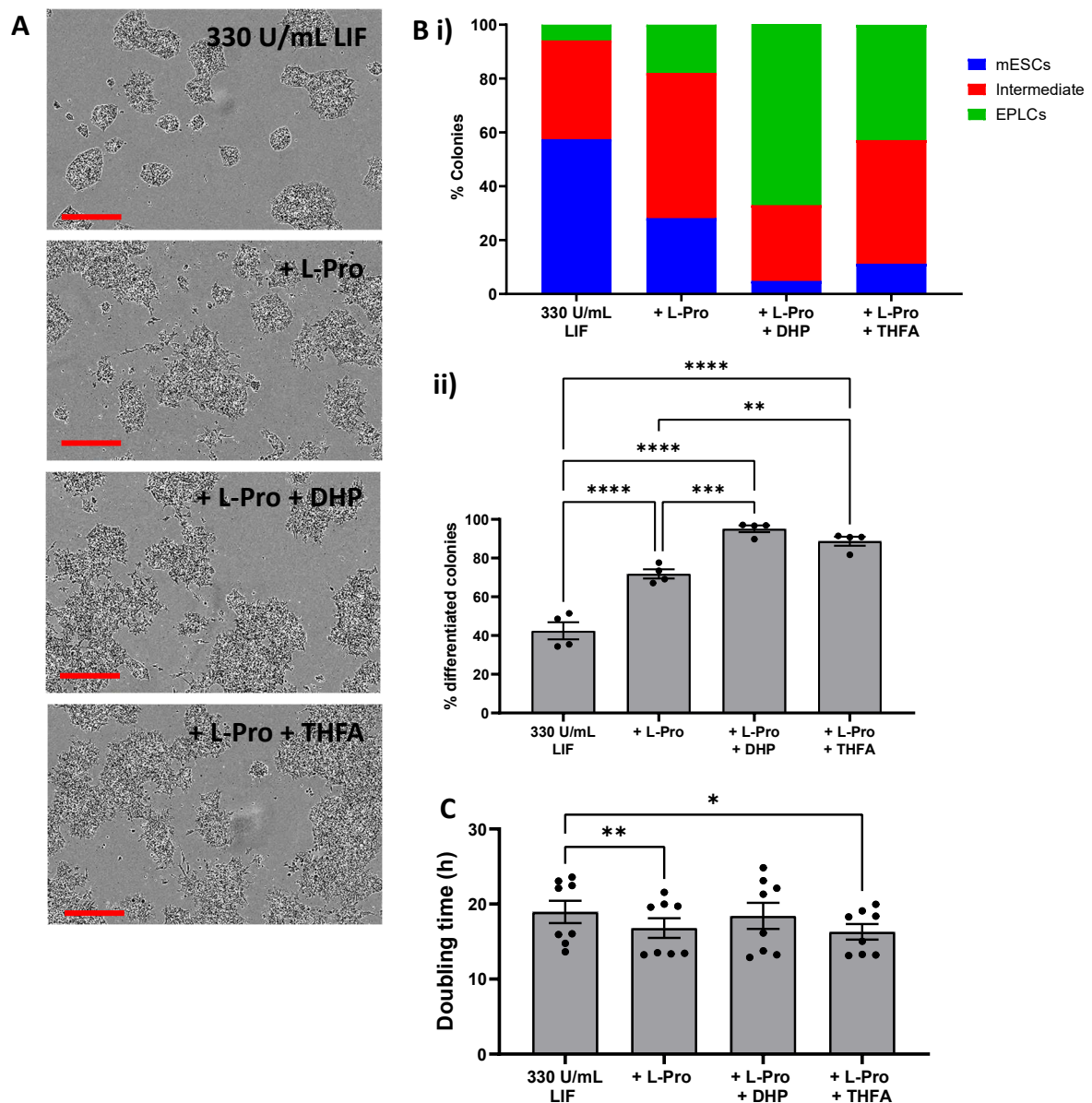


Figure 3.29 Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline with or without POX inhibitors. D3 mESCs were cultured in 330 U/mL LIF ± 0.4 mM L-Proline (L-Pro) or LIF + L-Pro + either 1.2 mM DHP or 0.4 mM THFA for 4 days. **A.** Representative images of each condition using IncuCyte ZOOM at 10x magnification. Scale bars = 300 μ m. **B. i)** Plots of colony morphology scores. In **ii)**, % differentiated colonies refers to colonies with have an EPLC morphology plus those whose morphology lies between mESCs and EPLCs. Data analysed using a one-way ANOVA with Tukey's *post hoc* test. Bars represent mean $\pm$ SEM, $n = 4$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ **C.** Bar graph of doubling times between 32 and 96 h based on % cell confluence measured from IncuCyte images. The growth curves were fitted by regression using an exponential growth curve equation (see Section 2.7.4). Data were analysed using a repeated measures one-way ANOVA with Tukey's *post hoc* test relative to 330 U/mL LIF. Bars represent mean $\pm$ SEM, $n = 8$, * $P < 0.05$, ** $P < 0.01$.

Figure 3.30 shows the effect of DHP and THFA on gene expression. The addition of DHP or THFA to L-Proline-treated cells had no effect on pluripotency marker expression (Figure 3.30A) compared to L-Proline alone, but downregulates *Fgf5* expression (Figure 3.30B).

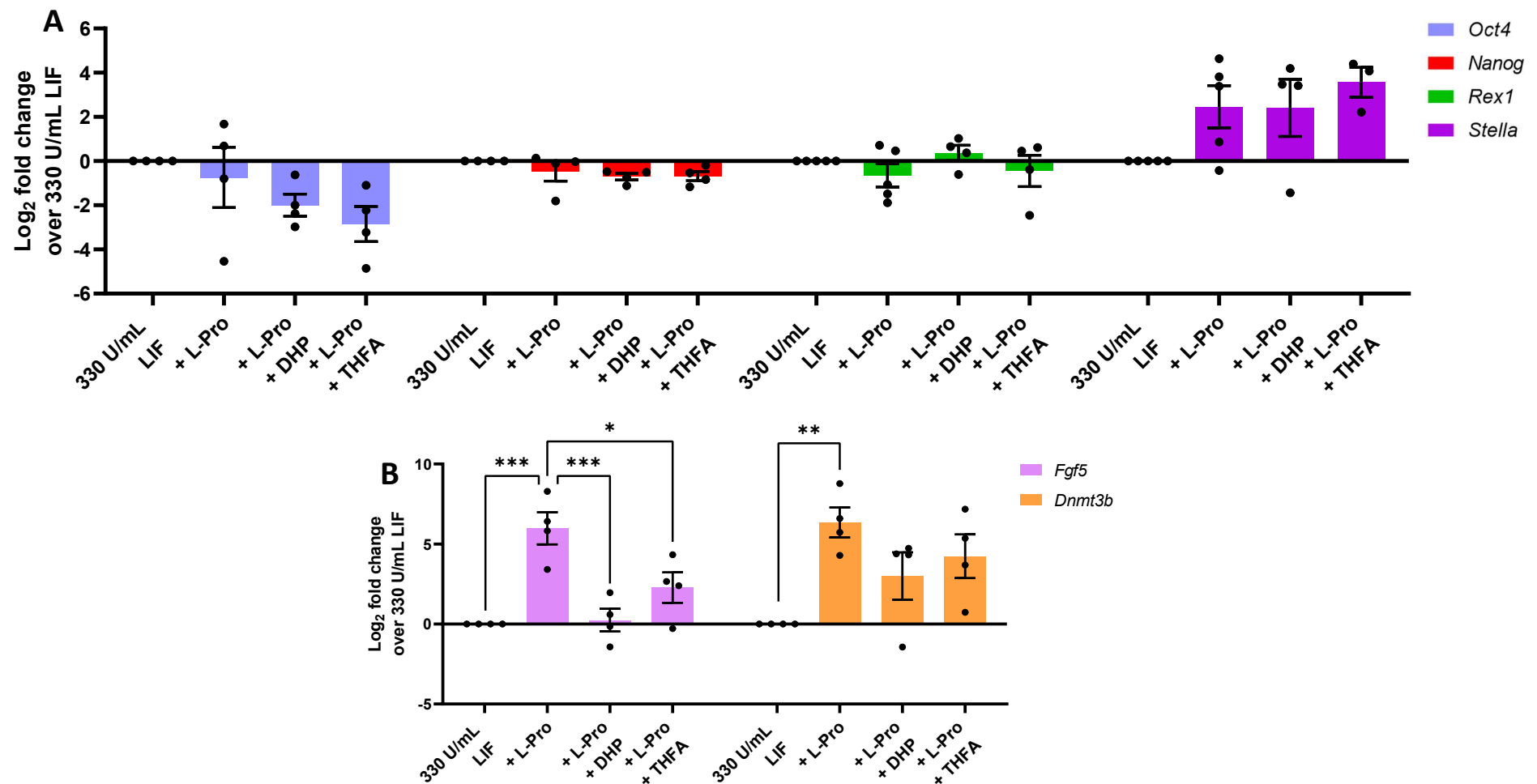


Figure 3.30 Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± DHP ± THFA. D3 mESCs were cultured in 330 U/mL LIF ± 0.4 mM L-Proline (L-Pro) or LIF + L-Pro + either 1.2 mM DHP or 0.4 mM THFA for 4 days. Samples were subjected to qPCR, with gene expression normalised to *18s* as the reference gene. Relative expression compared to 330 U/mL LIF only of: **A.** Pluripotency-associated genes *Oct4*, *Nanog* and *Rex1* and ground-state marker *Stella*. **B.** EPLC markers *Fgf5* and *Dnmt3b*. One-way ANOVA was used with Tukey's *post hoc* test. Bars represent mean ± SEM, $n \geq 4$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

3.3.9. *L-Proline effects on colony morphology, doubling time and gene expression profiles in the presence of DCA*

Pyruvate produced from glycolysis can be converted to lactate (resulting in anaerobic metabolism) or converted into acetyl-CoA via PDH, which then enters the TCA cycle and contributes to aerobic metabolism (Figure 3.1) (Michelakis *et al.*, 2008; Sanchez *et al.*, 2013; Rodrigues *et al.*, 2015).

DCA can be used to suppress anaerobic metabolism by inhibiting PDHK. Inhibiting PDHK in turn relieves its inhibition of PDH and thus promotes conversion of pyruvate to acetyl-CoA and entry into the TCA cycle (Figure 3.1) (Michelakis *et al.*, 2008; Sanchez *et al.*, 2013; Rodrigues *et al.*, 2015).

Figure 3.31 shows that DCA promotes EPLC colony morphology when added along with L-Proline compared to L-Proline alone but increases doubling time.

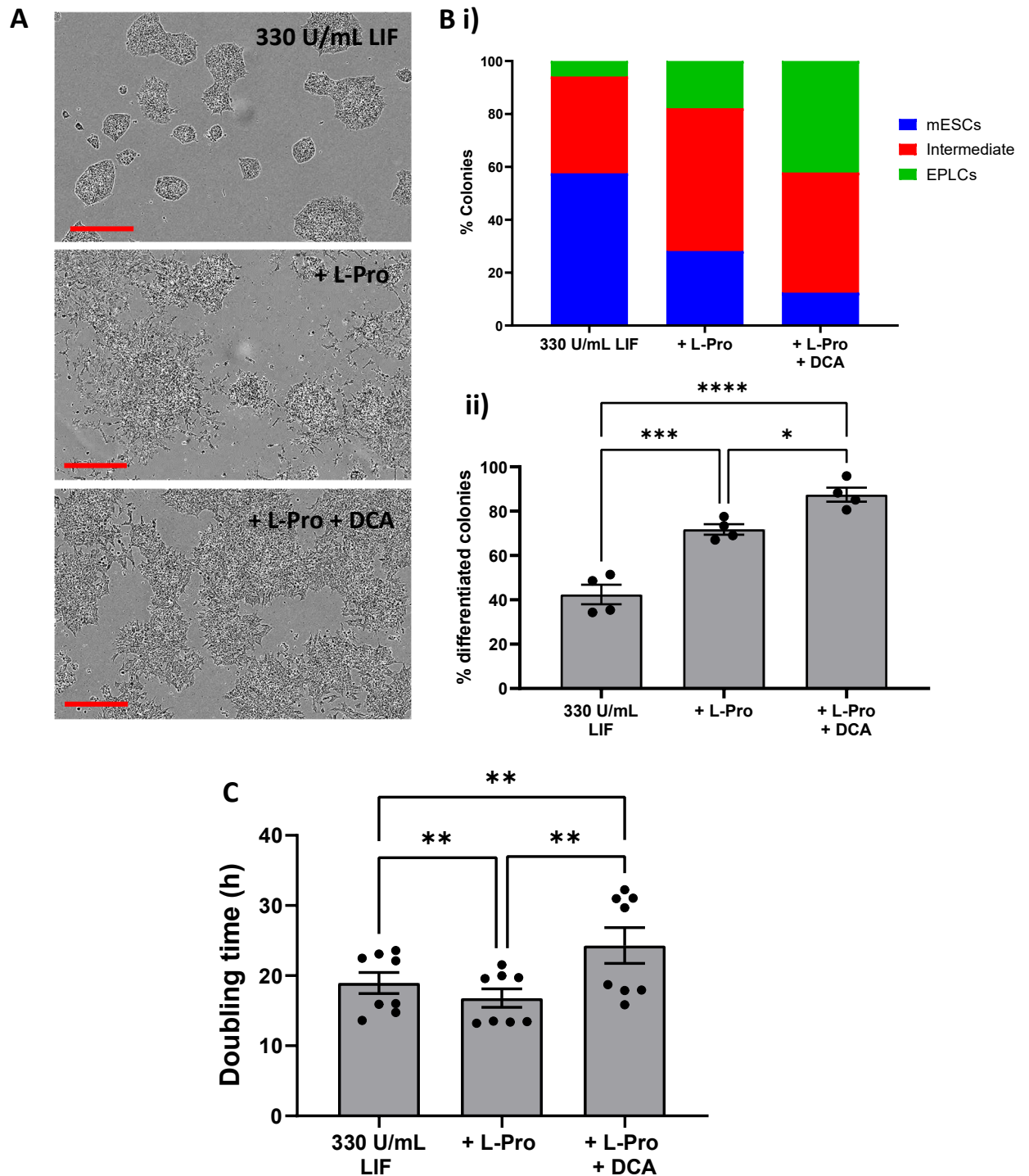


Figure 3.31 Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline with or without suppression of anaerobic metabolism by DCA. D3 mESCs were cultured in 330 U/mL LIF ± 0.4 mM L-Proline (L-Pro) ± 30 mM DCA for 4 days. **A.** Representative images of each condition using IncuCyte ZOOM at 10x magnification. Scale bars = 300 μ m. **B. i)** Plots of colony morphology scores. In **ii)**, % differentiated colonies refers to colonies with have an EPLC morphology plus those whose morphology lies between mESCs and EPLCs. Data analysed using a one-way ANOVA with Tukey's *post hoc* test. Bars represent mean $\pm$ SEM, $n = 4$, $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$ **C.** Bar graph of doubling time between 32 and 96 h based on % cell confluence measured from IncuCyte images. The growth curves were fitted by regression using an exponential growth curve equation (see Section 2.7.4). Data were analysed using a repeated measures one-way ANOVA with Tukey's *post hoc* test relative to 330 U/mL LIF. Bars represent mean $\pm$ SEM, $n = 8$, $**P < 0.01$

Figure 3.32 shows that DCA had no effect on L-Proline-mediated gene expression changes. In particular, the expression of the EPLC markers, *Fgf5* and *Dnmt3b*, which are upregulated in the presence of L-Proline are not affected by the addition of DCA.

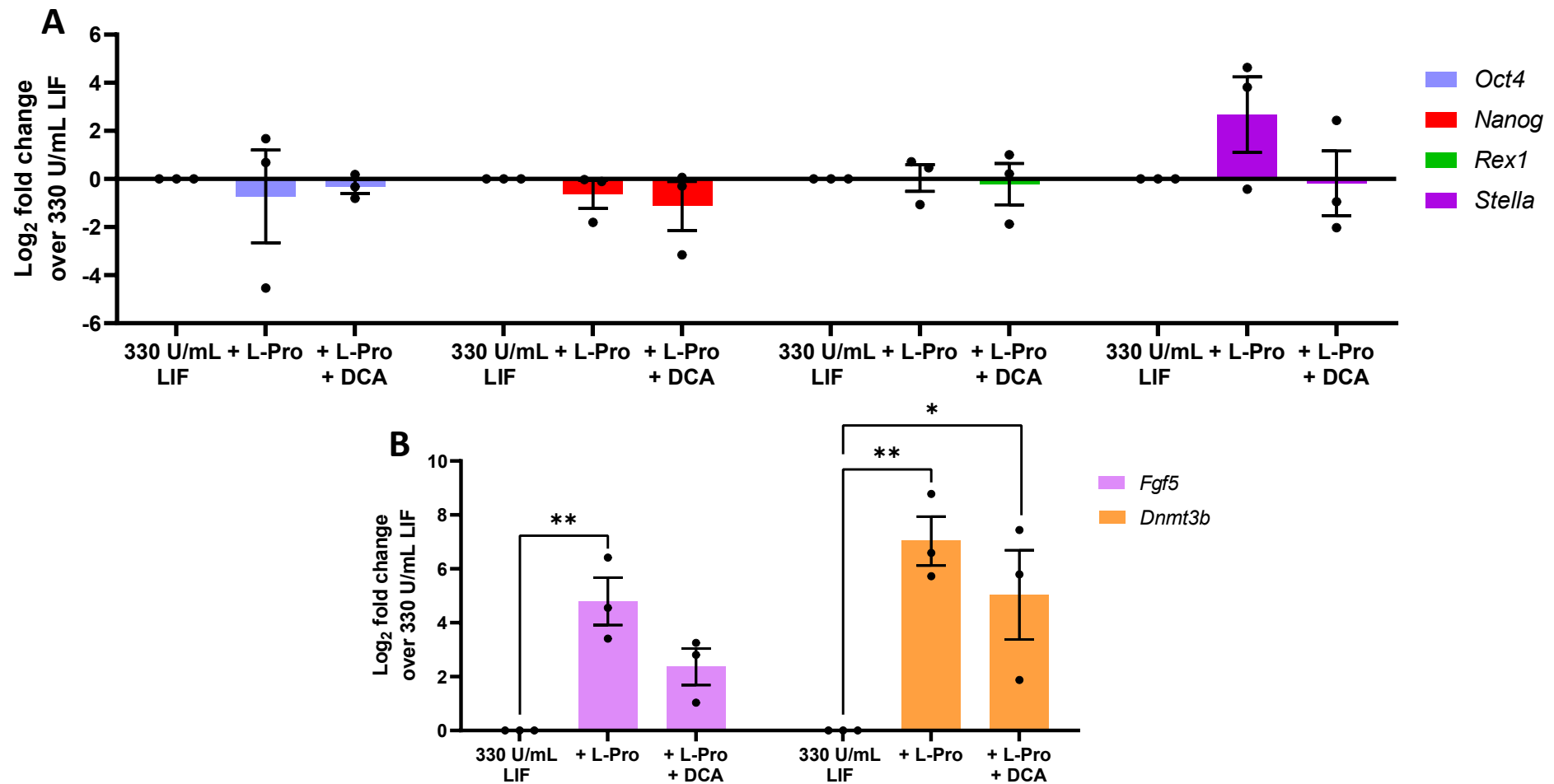


Figure 3.32 Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± DCA. D3 mESCs were cultured in 330 U/mL LIF ± 0.4 mM L-Proline (L-Pro) ± 30 mM DCA for 4 days. Samples were subjected to qPCR, with gene expression normalised to *18s* as the reference gene. Relative expression compared to 330 U/mL LIF only of: **A.** Pluripotency-associated genes *Oct4*, *Nanog* and *Rex1* and ground-state marker *Stella*. **B.** EPLC markers *Fgf5* and *Dnmt3b*. One-way ANOVA was used with Tukey's *post hoc* test. Bars represent mean ± SEM, $n \geq 4$, * $P < 0.05$, ** $P < 0.01$.

3.3.10. *The L-Proline effects on colony morphology, doubling time and gene expression profiles in the presence of 6AN*

6AN is an inhibitor of the first, and rate-limiting, enzyme of the PPP, 6-PGD (Figure 3.1) (Köhler *et al.*, 1970; Peters *et al.*, 2008).

Figure 3.33 shows dose culture results for 6AN and its dose-dependent response on doubling time on self-renewing naïve mESCs in 0.1% DMSO (vehicle control). Doubling time remained unaffected up to 3 μ M, and this concentration was then used for subsequent experiments.

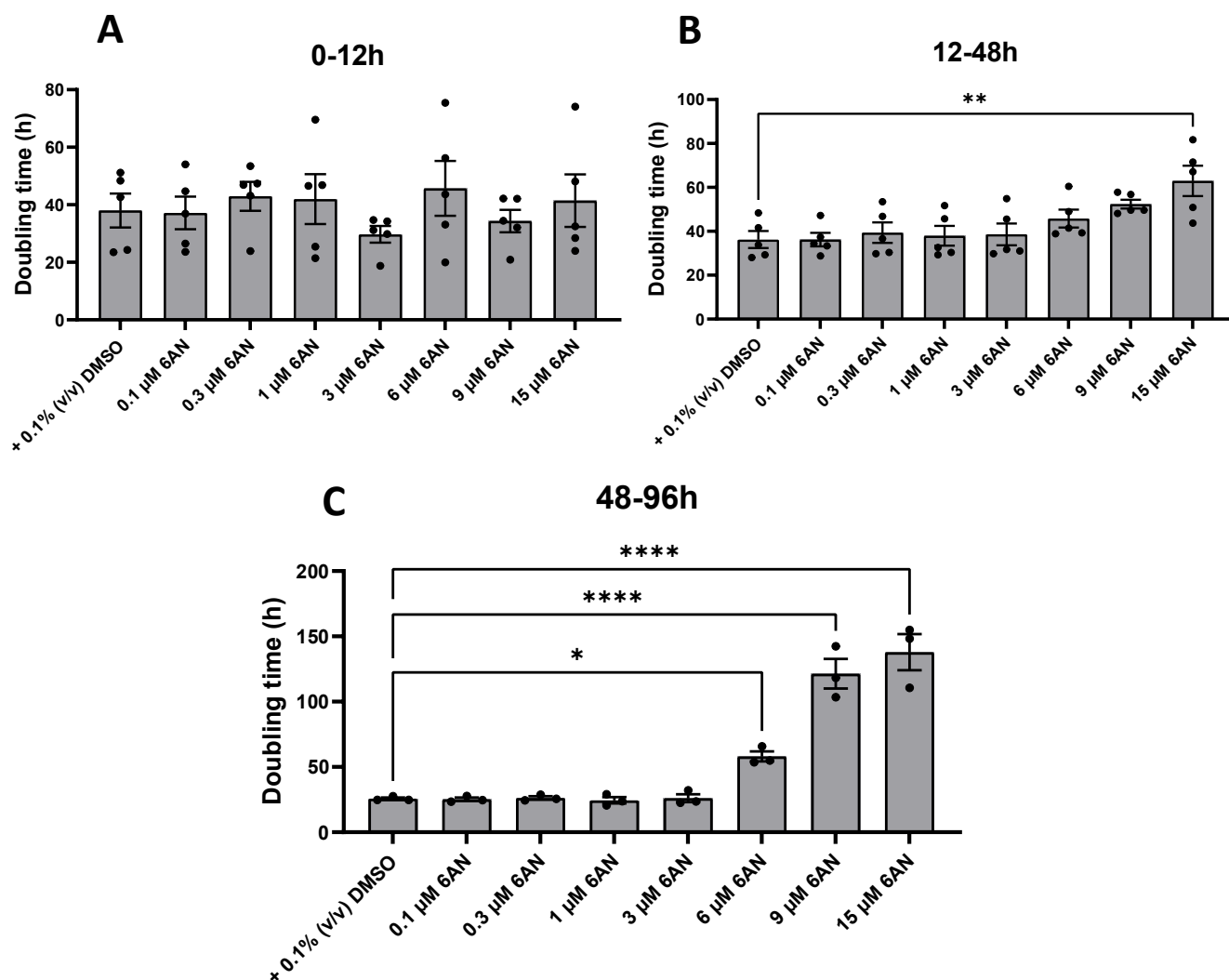


Figure 3.33 Dose culture of 6AN using doubling time to determine the optimal concentration. D3 mESCs were cultured in 330 U/mL LIF + 0.1% DMSO or 6AN (0-15 μM) for 4 days. Bar graph of doubling times between **A.** 0-12h, **B.** 12-48h and **C.** 48-96h based on % cell confluence measured from IncuCyte images. The growth curves were fitted by regression using an exponential growth curve equation (see Section 2.7.4). Data were analysed using a one-way ANOVA with Dunnett's *post hoc* test relative to 330 U/mL LIF + 0.1% DMSO. Bars represent mean ± SEM, $n \geq 3$. * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$.

Figure 3.34 shows that the addition of 3 μ M 6AN doesn't affect colony morphology when added alone to self-renewing naïve mESCs but it does significantly inhibit the colony morphology change induced by L-Proline, although the reduction is quite small (~85% to ~72%). 6AN alone doesn't affect doubling time but it does increase doubling time when added to L-Proline-treated cells.

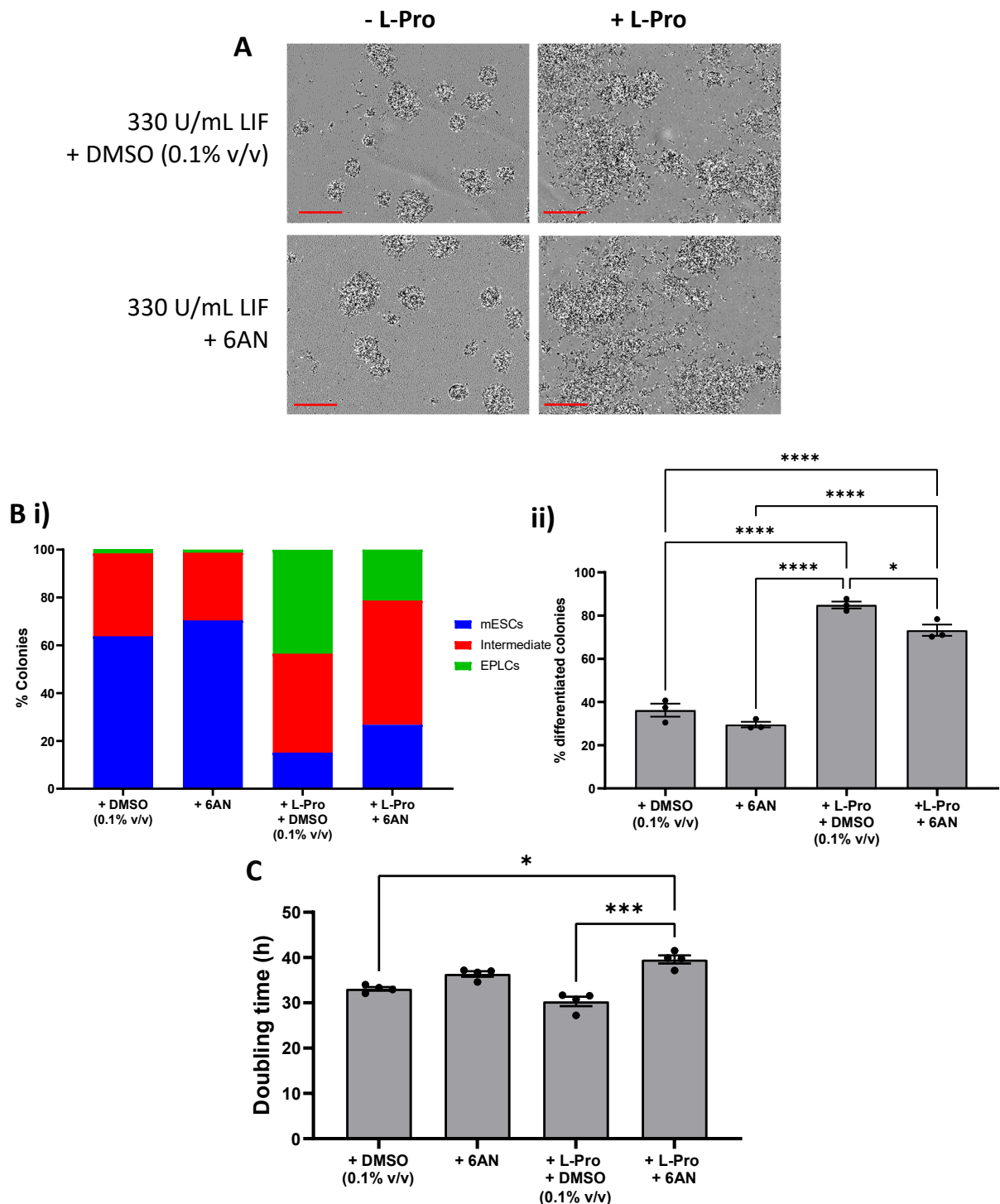


Figure 3.34 Colony morphology and doubling time profiles of naïve mESCs cultured in LIF ± L-Proline ± 6AN. D3 mESCs were cultured in 330 U/mL LIF + 0.1% (v/v) DMSO vehicle control or 3 μ M 6AN ± 0.4 mM L-Proline (L-Pro) for 4 days. **A.** Representative images of each condition using IncuCyte ZOOM at 10x magnification. Scale bars = 300 μ m. **B. i)** Plots of colony morphology scores. In **ii)**, % differentiated colonies refers to colonies with have an EPLC morphology plus those whose morphology lies between mESCs and EPLCs. Data analysed using a one-way ANOVA with Tukey's *post hoc* test. Bars represent mean $\pm$ SEM, $n = 3$, * $P < 0.05$, **** $P < 0.0001$ **C.** Bar graph of doubling times between 32 and 96 h based on % cell confluence measured from IncuCyte images. The growth curves were fitted by regression using an exponential growth curve equation (see Section 2.7.4). Data were analysed using a repeated measures one-way ANOVA with Tukey's *post hoc* test. Bars represent mean $\pm$ SEM, $n = 4$, * $P < 0.05$, *** $P < 0.001$.

Figure 3.35 shows the effect of 6AN on gene expression. Again, the addition of L-Proline increases the expression of definitive ectoderm markers *Fgf5* and *Dnmt3b*, as expected. However, the addition of 6AN along with L-Proline doesn't alter the gene expression profile.

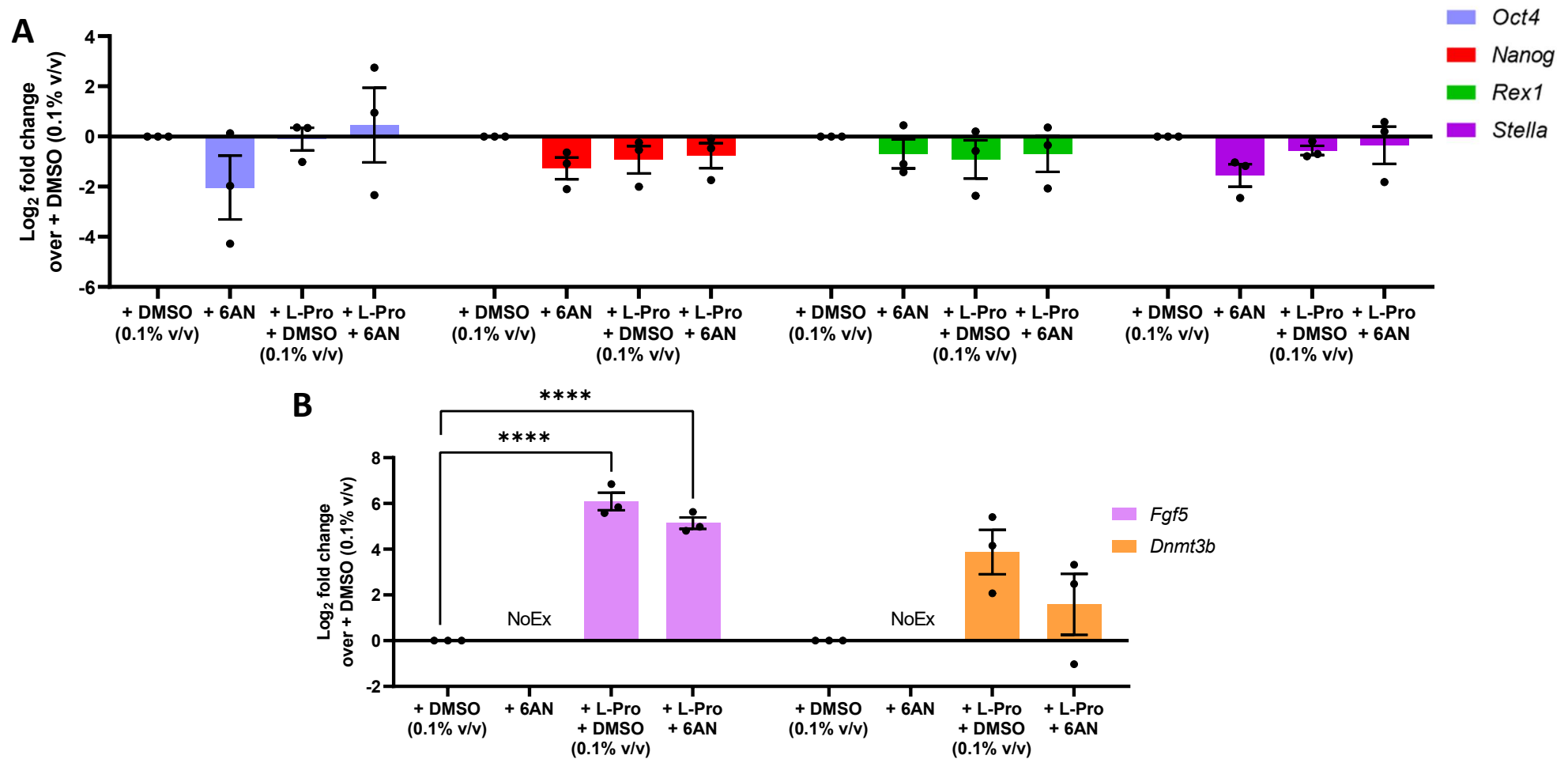


Figure 3.35 Gene expression profiles of naïve mESCs cultured in LIF ± L-Proline ± 6AN. D3 mESCs were cultured in 330 U/mL LIF + 0.1% (v/v) DMSO vehicle control or 3 μ M 6AN $\pm$ 0.4 mM L-Proline (L-Pro) for 4 days. Samples were subjected to qPCR, with gene expression normalised to *18s* as the reference gene. Relative expression compared to 330 U/mL LIF + 0.1% (v/v) DMSO of: **A.** Pluripotency-associated genes *Oct4*, *Nanog* and *Rex1* and ground-state marker *Stella*. **B.** EPLC markers *Fgf5* and *Dnmt3b*. One-way ANOVA was used with Tukey's *post hoc* test. Bars represent mean $\pm$ SEM, n = 3, **** P < 0.0001. NoEx = no expression.

3.4. Discussion

In this Chapter, we sought to elucidate L-Proline-mediated metabolic processes that drive cell fate across the pluripotency spectrum. L-Proline added to cancer cells increases anaerobic metabolism (a process known as aerobic glycolysis or the Warburg effect) and it does so in cooperation with the PPP (Phang *et al.*, 2008, 2010; Pandhare *et al.*, 2009; Patra and Hay, 2014). ESCs also use these mechanisms to maintain their pluripotency (Rodrigues *et al.*, 2015; Hu *et al.*, 2016; Intlekofer and Finley, 2019; Tsogtbaatar *et al.*, 2020).

3.4.1. *Differential gene expression analysis and TRIAGE transformation provides candidate genes and predict protein-protein interactions driving cell fate across pluripotency*

From the RNAseq results, PCA showed EPLCs cluster independently from ground- and naïve-state mESCs and from the primed (EpiSC) state (Figure 3.2). This indicates EPLCs are a unique cell type within the pluripotency spectrum, and is consistent with their altered gene expression profile compared to naïve mESCs, which includes downregulation of the expression of core and extended pluripotency gene networks and upregulation of the expression of primitive ectoderm markers (Figure 3.4) both of which indicate EPLCs are moving towards a more primed state of pluripotency similar to that of primitive ectoderm of the implanted E5.5 embryo (Rathjen *et al.*, 1999, 2003; Pelton *et al.*, 2002; Washington *et al.*, 2010).

On the other hand, ground-state mESCs upregulate the expression of many (though not all) pluripotency genes compared to naïve mESCs (Figure 3.4), consistent with them being less responsive to pro-differentiation signals and with the core and extended pluripotency circuits being interconnected and self-reinforcing (Boyer *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005; Casanova *et al.*, 2011; Xue *et al.*, 2011; Shparberg *et al.*, 2019b). Further reinforcement of ground-state status is seen with the reduced expression of markers further along the pluripotency spectrum (in particular, EPLC markers) as well as many markers of lineage

commitment (in particular, those for neurectoderm, surface ectoderm, and the extraembryonic primitive endoderm) (Figure 3.4). Thus, based on their gene expression profiles, ground-state cells are ‘less fated’ to lineage commitment.

Consistent with this, in the embryo, at the early blastocyst stage (E3.5), gene/protein expression in the ICM shows a ‘salt-and-pepper’ pattern with some cells expressing *Nanog* and the others expressing *Gata6* in stochastic fashion (Pelton *et al.*, 2002; Chazaud *et al.*, 2006; Artus *et al.*, 2011). Primitive endoderm is then formed by E4-4.5 (i.e., just prior to blastocyst hatching and implantation) from the *Gata6*-expressing cells, which move to form the primitive endoderm lining of the blastocoel (Pelton *et al.*, 2002; Chazaud *et al.*, 2006; Artus *et al.*, 2011). Our results (Figure 3.4) show that ground-state cells express ~32-fold less *Gata6* compared to naïve cells, suggesting they represent a population of pluripotent cells less mature than the ICM of the E3.5 embryo.

TRIAGE transformation of the RNAseq data (Figures 3.5–3.9) shows that the top 10 ranked genes by DS consist largely of transcription factors but also includes cell cycle and signalling pathway genes. 5 of the top 10 are the same across all 4 cell types, including *Tbx3* and *Esrrb*, which are part of the extended pluripotency network (Ivanova *et al.*, 2006; Xue *et al.*, 2011; Shparberg *et al.*, 2019b). This strongly suggests these genes are required to maintain pluripotency, although it doesn’t exclude the possibility that each is also required for the transition of one pluripotent cell type to another. If the latter is true, this is similar to, for example, the mechanisms of action of *Sox2*: There is an absolute requirement for the expression of *Sox2* to maintain pluripotency (Okumura-Nakanishi *et al.*, 2005; Nakatake *et al.*, 2006; Niwa *et al.*, 2009; Casanova *et al.*, 2011) but its expression is also required to drive development to definitive ectoderm, neurectoderm and neural stem-cell populations (Kuegler *et al.*, 2010; Ng *et al.*, 2013; Schaefer and Lengerke, 2020).

There is ‘drift’ in the composition of the top 10 DS with movement along the pluripotency spectrum. For example, *Pitx2*, *Hoxb13*, and *En1* are only found in the top 10 for EpiSCs and *Gata4* is found only in EpiSCs and EPLCs. To what extent these genes are required to drive the transition to the more primed end of the pluripotency spectrum remains to be determined.

Cluster and GO analysis of the top 100 DS-ranked genes was undertaken using TRIAGE-Parser (Figures 3.10–3.11). The top 10 significant GO terms included terms associated with transcription across all pluripotency states and cell lines (Figures 3.10–3.11). This is not surprising given the nature of TRIAGE analysis (Shim *et al.*, 2020; Sun *et al.*, 2023). Significant GO terms also included some metabolic processes beyond the top 10 (data not shown, further discussed in Section 3.4.2). PPIs predicted include physical and functional interactions. This means these interactions represent a collection of evidence based on *in vitro* and *in vivo* methods (i.e. protein fusion, co-expression, co-occurrence and others) and computational predictions (i.e. text mining and database searches) (Szklarczyk *et al.*, 2019; Sun *et al.*, 2023). Although rather coarse-grained, the results suggest that EpiSCs are more geared towards commitment to mesendoderm compared to EPLCs. To test this, the differentiation potential of these two populations can be examined further by quantifying their capacity to commit to each of the 3 pluripotent germ layers. Gene expression analysis of L-Proline-derived EPLCs do indeed indicate they represent a population of cells equivalent to the anterior (nervous system forming) end of the embryo (Glover, 2018).

Where two GO/PPI clusters exist – as is the case with ground-state cells – the circuits may act in parallel, and independently, to help drive development. As an analogous example, failure of the rostral end of the neural tube to close during early development results in anencephaly but many other parallel programs, such as limb development, proceed largely, if not completely, normally.

Though each gene/PPI cluster is, generally speaking, highly complex and interconnected, those at the centre of the clusters are more likely to represent master regulators of development and therefore among the first choices for further experiments; e.g., by knockdown and knockout.

For instance, TFAP2A is shown to be a master regulator at ground to EPLC stages, but not the EpiSC stage. TFAP2A is a transcription factor, which is active in the neural crest in the embryo for craniofacial development (Schorle *et al.*, 1996; Van Otterloo *et al.*, 2018, 2022). TFAP2A interacts with core pluripotency transcription factors OCT4 and SOX2 to reprogram human ESCs to a neural crest stem cell fate (Hovland *et al.*, 2022). Knockout of this transcription factor would potentially result in a preference to EpiSC formation over EPLCs.

Another example is OTX2, which is at the centre of D3 ground, naïve and EPLC states (Figure 3.10). OTX2 a transcription factor essential for brain, cerebellum and retinal development (Beby and Lamonerie, 2013). Otx2 also interacts with Oct4 to establish epigenetic changes during the exit from ground state pluripotency (Yang *et al.*, 2014). We hypothesise Otx2 knockdown will prevent these epigenetic changes and maintain the ground state.

3.4.2. Selected genes encoding metabolic enzymes may drive cell fate and identity across the pluripotency spectrum

We used differential gene expression analysis and TRIAGE transformation to gain insight into what role various metabolic pathways may play across the pluripotency spectrum (Figure 3.12-3.23).

Generally speaking, changes in expression of many genes in these pathways do change between the ground- and naïve states and EPLCs but very few have a DS rank which puts them into the top 1359 genes, and therefore likely to be key drivers promoting lineage progression.

For example, in terms of L-Proline catabolism, ground-state cells compared to naïve cells have reduced expression of the *Slc38a2*, coding for the plasma-membrane L-Proline transporter

SNAT2 (Figure 3.12). If this change in expression is mapped to cellular uptake, this indicates ground-state ESCs have a reduced requirement for L-Proline uptake. In fact, L-Proline is a non-essential amino acid and so can be synthesised endogenously but its concentration is kept low in these cells (~25 μ M) in the L-Proline-free medium used here through a feedback loop involving the amino-acid response pathway, and amino-acid starvation reduces expression of amino-acid transporters such as SNAT2 (Tan *et al.*, 2011, 2016; D’Aniello *et al.*, 2015; Glover, 2018). Thus, this profile would help promote self-renewal of these ground-state cells, as reduced uptake of L-Proline and suppressed endogenous production keeps intracellular concentrations of L-Proline at a level below that needed to induce differentiation (D’Aniello *et al.*, 2015; Glover, 2018).

Downstream of SNAT2, ground-state cells have decreased expression of *Glul*, and increased expression of both *Glud1* and *Gss* compared to naïve cells. This would map to the following with respect to L-Proline catabolism, respectively: (i) Reduced production of glutamine, (ii) increased production of acetyl-CoA and flux into the TCA cycle (promoting aerobic metabolism), and (iii) increased GSH production. This is consistent with ground-state ESCs being more reliant on aerobic metabolism than cells further along the pluripotency spectrum (Khoa *et al.*, 2020; Tsogtbaatar *et al.*, 2020; Diamante and Martello, 2022) but this also generates more reactive oxygen species, which can be kept under homeostatic control by increased production of GSH.

Several other aspects of the expression profiles are in keeping with this scenario: (i) Ground-state mESCs also have increased expression of *Oat*, which would lead to more L-Ornithine being converted to GSA (Figure 3.14). Since the expression of other L-Proline biosynthesis genes are generally suppressed in ground-state cells compared to naïve cells (Figure 3.14), the increased GSA might also be expected to flux through the TCA cycle and produce more GSH, consistent with a greater reliance on aerobic metabolism and the need to control levels of reactive oxygen species. (ii) Expression of *Pdk1*, the gene for the kinase which

inhibits PDH, is decreased, while that for the PDH gene itself, *Pdha1*, which converts pyruvate to acetyl-CoA, is increased (Figure 3.16). (iii) The expression of *Slc16a3*, the gene for the monocarboxylate transport of lactate out of the cell is also decreased, in keeping with reduced anaerobic metabolism via glycolysis. (iv) The expression of genes in the first ‘quarter’ of the TCA cycle is also increased (Figure 3.20). What is not consistent with this scenario is the generally reduced expression of ETC genes in ground-state cells compared to naïve cells (Figure 3.22).

In any case, each of these possibilities requires confirmation through pathway flux and enzyme assays. However, our results do highlight obvious candidates for further testing.

Naïve mESCs are further along the pluripotency spectrum and therefore more ‘prepared’ to receive exogenous signal leading to frank priming and loss of pluripotency. Consistent with this, they have upregulated expression of the L-Proline transporter gene, *Slc38a2*, compared to ground-state cells (Figure 3.12). When transition-inducing L-Proline is added to culture medium, the amino-acid response pathway is overwhelmed in naïve cells and, as a consequence, intracellular L-Proline concentrations increase (Tan *et al.*, 2011), expression of *Slc38a2* will increase (Figure 3.12), and the naïve state will progress to EPLCs (Washington *et al.*, 2010; Casalino *et al.*, 2011; Glover, 2018).

Once at the EPLC stage, not only is the expression of the SNAT2 gene increased even further (Figure 3.12), presumably to take advantage of the added extracellular L-Proline, but synthesis of L-Proline from the key precursor, L-Glutamate, appears to be reduced as evidenced by reduced expression of the genes for the two required enzymes (Figure 3.14): (i) *Aldh18a1*, the enzyme for which converts L-Glutamate to GSA, and (ii) *Pycr1*, the enzyme for which converts P5C to L-Proline as part of the proline cycle.

In addition, it also appears that the balance shifts to anaerobic metabolism over aerobic since in EPLCs compared to naïve mESCs: (i) The expression for *Pdk1* increases and that for

Pdhal decreases, consistent with reduced production of acetyl-CoA and flux through the TCA cycle. (ii) The expression of *Slc16a3*, the gene for transport of lactate out of the cell increases (Figure 3.16). These results are consistent with increased reliance on anaerobic metabolism as priming across the pluripotency spectrum continues.

These trends more or less continue for primed EpiSCs, again consistent with primed cells having even more reliance on anaerobic metabolism than those earlier in the pluripotency spectrum (Zhou *et al.*, 2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022). In addition, expression of the SNAT2 gene is again reduced in EpiSCs compared to naïve mESCs (Figure 3.12). We speculate this is because these primed cells are geared to form mesendoderm (Bernemann *et al.*, 2011), which doesn't require L-Proline, as opposed to EPLCs, which are geared towards neurectoderm, which does require L-Proline (Harvey *et al.*, 2010; Washington *et al.*, 2010; Li *et al.*, 2013; Glover, 2018; Shparberg, 2018). In addition, the expression of *Ndufa10*, the gene for Complex I in the ETC, is strongly downregulated (Figure 3.22) and as this is one of the main entry points for high-energy electrons into the ETC, reduced expression culminating in reduced Complex I production is likely to lead to reduced aerobic metabolism, consistent with the energy requirements of this cell type.

Of all the genes in all the pathways examined here, only 4 had a TRIAGE-transformed DS ranking suggesting they play a direct role in cell fate. These results suggest each of these genes is worthy of further work to establish their importance in driving cells across the pluripotency spectrum and perhaps allowing them to exit pluripotency. These 4 genes are:

(i) The lactate transporter gene, *Slc16a3*. It has a high DS only in EPLCs (Figure 3.17), and this is where its expression is also strongly increased (Figure 3.16). It's possible that increased expression of this gene in this part of the pluripotency spectrum is a key driver of the shift to more reliance on anaerobic metabolism.

(ii) *Tkt11*, the product for which is the enzyme in the PPP that catalyses the bi-directional conversion of xylulose-5-phosphate to glyceraldehyde-3-phosphate and erythrose-4-phosphate, and ribose-5-phosphate to sedoheptulose-7-phosphate (Figure 3.18). It has a high DS only in EpiSCs (Figure 3.19), and this is where its expression is increased (Figure 3.18). Increased expression of this gene may be required for primed cells to exit pluripotency. The enzyme regulates the carbon flux between the PPP and glycolysis in both directions (Ge *et al.*, 2020). Therefore, the enzyme allows the PPP to directly influence the flux of metabolites between itself and glycolysis.

(iii) *Ndufa10*, the gene for Complex I in the ETC. It has a high DS in all 4 cell types of the pluripotency spectrum (Figure 3.23) but its expression is decreased in EpiSCs (Figure 3.22), as noted above. Again, it's possible this gene is required for primed cells to exit pluripotency.

(iv) *Idh3g*, the product for which is the enzyme in the TCA cycle that catalyses the conversion of isocitrate to α -KG (Figure 3.20). It also has a high DS in all 4 cell types of the pluripotency spectrum, and its expression doesn't change much. It's possible this gene is required to maintain pluripotency. The enzyme catalyses one of the main entry points for high-energy electrons into the ETC, as well as producing α -KG, which contribute to epigenetic modifications (Xiao *et al.*, 2012; Ryu *et al.*, 2015).

3.4.3. Ground and naïve states and EPLCs have different basal aerobic profiles as measured by the Seahorse extracellular flux assay but do not respond to acute exposure to L-Proline

Basal rates of anaerobic metabolism (ECAR) don't change in cells across the pluripotency spectrum (Figure 3.25B) but that for aerobic metabolism drops by two-thirds (Figure 3.25A). This indicates that reliance shifts to anaerobic metabolism, and that the cells become metabolically quieter, as priming across the pluripotency spectrum proceeds. This aligns with shifts in anaerobic/aerobic metabolism observed when ESCs transition to EpiSC (Zhou *et al.*,

2012; Fiorenzano *et al.*, 2016; Minchiotti *et al.*, 2022). However, the idea of ‘quiet metabolism’ is new. This idea is analogous to the Quiet Embryo Hypothesis, which proposes that embryos with lower overall levels of metabolism have improved development and viability (Leese *et al.*, 2008).

None of the cell types responded to acute injections of L-Proline in terms of anaerobic or aerobic metabolism (Figure 3.26A-C). This indicates, for example, that any L-Proline-mediated effects on metabolism are not at this gross level of anaerobic/aerobic metabolism or that the time-frame required is over several hours as opposed to the 1-2 h over which a Seahorse experiment can be run.

Spare respiratory capacity indicates the reserve a cell type has to produce ATP via aerobic metabolism (Divakaruni *et al.*, 2014). A high spare capacity indicates that the cell type can better meet a high acute demand for the production of ATP via aerobic metabolism (Divakaruni *et al.*, 2014). The spare capacity of these pluripotent cells does change, with that for EPLCs dropping by around a third compared to ground and naïve states (Figure 3.26D). This is consistent with EPLCs having increased reliance of anaerobic metabolism over aerobic metabolism to meet any acute demand in ATP production. Ground and naïve states’ higher spare capacities indicate the greater reliance of these pluripotency states on aerobic metabolism. These results, together with the higher basal levels of OCR in ground-state ESCs, again indicate reliance on aerobic metabolism drops as cells progress along the pluripotency spectrum (Yanes *et al.*, 2010; Sperber *et al.*, 2015; Tohyama *et al.*, 2017; Vlaski-Lafarge *et al.*, 2019; Khoa *et al.*, 2020; Tsogtbaatar *et al.*, 2020; Diamante and Martello, 2022). EpiSCs may demonstrate a similar metabolic profile to EPLCs as EpiSCs are at the primed end of the pluripotency spectrum. This can be tested but it would need to be done in the absence of a layer of MEF feeder layer cells normally used to support EpiSC self-renewal and growth.

3.4.4. *Inhibition of POX-mediated conversion of L-Proline to P5C has different effects on parameters associated with the mESC-to-EPLC transition*

Inhibition of POX in the presence of L-Proline on mESCs inhibited morphological change associated with transition to EPLCs (Casalino *et al.*, 2011) but we see the reverse effect: Inhibition of POX by DHP or THFA in ESCs exposed to L-Proline *promotes* changes in colony morphology associated with the transition (Figure 3.29-3.30). Culture conditions in the two studies are very similar. However, we use a lower concentration of LIF (330 U/mL instead of 1000 U/mL) and this can affect colony morphology change: The more LIF, the less colony morphology change (Morris, unpublished results). This highlights that cooperation between growth factors (in this case, LIF and L-Proline) to elicit biological change is dependent on their relative concentrations.

POX inhibition failed to have any effect on the L-Proline-mediated reduction in doubling time but it did inhibit L-Proline mediated increase in expression of the key EPLC marker, *Fgf5*. Collectively, these results show that metabolism of L-Proline by POX have different effects on different parameters of the transition: Colony morphology change is accentuated, doubling time is unaffected, and changes in gene expression profile were observed. This is consistent with the idea that specific mechanisms of action (in this case, L-Proline metabolism) contribute to some but not all aspects of the complex processes involved in changes in cell identity. Similarly, the various signalling pathways involved in the ES-to-EPLC transition (MEK/ERK, PI3K/Akt, mTOR) contribute to different aspects of the transition, as shown by the use of selective pathway inhibitors (Lonic, 2007; Glover *et al.*, 2023).

3.4.5. *Anaerobic metabolism and the PPP drive some aspects of the mESC-to-EPLC transition*

Similar to the POX inhibitor results above, the use of DCA (which effectively suppresses anaerobic metabolism by accelerating aerobic metabolism) and 6AN (which inhibits the PPP),

have different effects on the parameters used to measure the ES-to-EPLC transition. DCA, like the POX inhibitors, accentuates the L-Proline-mediated colony morphology change, but it reverses the reduced doubling time that L-Proline induces, and has no effect on gene expression change (Figures 3.31–3.32). On the other hand, 6AN (partially) reverses L-Proline-mediated colony morphology change, reverses the reduced doubling time that L-Proline induces, and has no effect on gene expression change (Figures 3.34–3.35).

These results show that flux via anaerobic/aerobic metabolism pathways and the PPP contribute to the ES-to-EPLC transition, and it's possible that this is because L-Proline catabolism is linked to these (Figure 3.1).

Measurement of actual pathway flux would be valuable to examine the biological complexity of the roles of these pathways. This would include taking timed samples following the addition of L-Proline and measuring the activity of rate-limiting enzymes, and quantifying the movement of metabolites of L-Proline into various metabolic pathways using radiolabelled forms of the molecule in combination with LC-MS.

3.5. References

Aral, B. and Kamoun, P. (1997) 'The proline biosynthesis in living organisms', *Amino Acids*, 13(3–4), pp. 189–217. doi: 10.1007/BF01372588.

Artus, J., Piliszek, A., and Hadjantonakis, A.-K. (2011) 'The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17', *Developmental Biology*, 350(2), pp. 393–404. doi: 10.1016/j.ydbio.2010.12.007.

Beby, F. and Lamonerie, T. (2013) 'The homeobox gene Otx2 in development and disease', *Experimental Eye Research*. England, 111, pp. 9–16. doi: 10.1016/j.exer.2013.03.007.

Bernemann, C., Greber, B., Ko, K., Sterneckert, J., Han, D. W., Araúzo-Bravo, M. J., and Schöler, H. R. (2011) 'Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features', *Stem Cells*. England, 29(10), pp. 1496–1503. doi: 10.1002/stem.709.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 122(6), pp. 947–956. doi: 10.1016/j.cell.2005.08.020.

Browne, K. (2017) *Anaerobic metabolism stimulates differentiation in mouse embryonic stem cells*. University of Sydney.

Casalino, L., Comes, S., Lambazzi, G., De Stefano, B., Filosa, S., De Falco, S., De Cesare, D., Minchiotti, G., and Patriarca, E. J. (2011) 'Control of embryonic stem cell metastability by L-Proline catabolism', *Journal of Molecular Cell Biology*, 3(2), pp. 108–122. doi: 10.1093/jmcb/mjr001.

Casanova, E. A., Bürki, K., and Cinelli, P. (2011) 'Molecular mechanisms of pluripotency

in murine embryonic stem cells’, in Atwood, C. (ed.) *Embryonic Stem Cells: The Hormonal Regulation of Pluripotency and Embryogenesis*. InTech, pp. 27–60. doi: 10.5772/14219.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006) ‘Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway’, *Developmental Cell*. Cell Press, 10(5), pp. 615–624. doi: 10.1016/j.devcel.2006.02.020.

Comes, S., Gagliardi, M., Laprano, N., Fico, A., Cimmino, A., Palamidessi, A., De Cesare, D., De Falco, S., Angelini, C., Scita, G., Patriarca, E. J., Matarazzo, M. R., and Minchiotti, G. (2013) ‘L-Proline induces a mesenchymal-like invasive program in embryonic stem cells by remodeling H3K9 and H3K36 methylation’, *Stem Cell Reports*, 1(4), pp. 307–321. doi: 10.1016/j.stemcr.2013.09.001.

D’Aniello, C., Fico, A., Casalino, L., Guardiola, O., Di Napoli, G., Cermola, F., De Cesare, D., Tatè, R., Cobellis, G., Patriarca, E. J., and Minchiotti, G. (2015) ‘A novel autoregulatory loop between the Gcn2-Atf4 pathway and L-Proline metabolism controls stem cell identity’, *Cell Death And Differentiation*, 22, pp. 1094–1105. doi: 10.1038/cdd.2015.24.

D’Aniello, C., Habibi, E., Cermola, F., Paris, D., Russo, F., Fiorenzano, A., Di Napoli, G., Melck, D. J., Cobellis, G., Angelini, C., Fico, A., Bluelloch, R., Motta, A., Stunnenberg, H. G., De Cesare, D., Patriarca, E. J., and Minchiotti, G. (2017) ‘Vitamin C and L-Proline antagonistic effects capture alternative states in the pluripotency continuum’, *Stem Cell Reports*. Elsevier, 8(1), pp. 1–10. doi: 10.1016/j.stemcr.2016.11.011.

DeBerardinis, R. J., Lum, J. J., Hatzivassiliou, G., and Thompson, C. B. (2008) ‘The biology of cancer: metabolic reprogramming fuels cell growth and proliferation’, *Cell Metabolism*. Cell Press, 7(1), pp. 11–20. doi: 10.1016/j.cmet.2007.10.002.

Diamante, L. and Martello, G. (2022) ‘Metabolic regulation in pluripotent stem cells’,

Current Opinion in Genetics & Development, 75, p. 101923. doi: <https://doi.org/10.1016/j.gde.2022.101923>.

Divakaruni, A. S., Paradyse, A., Ferrick, D. A., Murphy, A. N., and Jastroch, M. (2014) ‘Analysis and interpretation of microplate-based oxygen consumption and pH data’, in *Methods in Enzymology*. United States, pp. 309–354. doi: 10.1016/B978-0-12-801415-8.00016-3.

Elia, I., Broekaert, D., Christen, S., Boon, R., Radaelli, E., Orth, M. F., Verfaillie, C., Grünewald, T. G. P., and Fendt, S.-M. (2017) ‘Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells’, *Nature Communications*, 8(1), p. 15267. doi: 10.1038/ncomms15267.

Fiorenzano, A., Pascale, E., D’Aniello, C., Acampora, D., Bassalart, C., Russo, F., Andolfi, G., Biffoni, M., Francescangeli, F., Zeuner, A., Angelini, C., Chazaud, C., Patriarca, E. J., Fico, A., and Minchiotti, G. (2016) ‘Cripto is essential to capture mouse epiblast stem cell and human embryonic stem cell pluripotency’, *Nature Communications*. England, 7(1), p. 12589. doi: 10.1038/ncomms12589.

Folmes, C. D. L., Dzeja, P. P., Nelson, T. J., and Terzic, A. (2012) ‘Metabolic Plasticity in Stem Cell Homeostasis and Differentiation’, *Cell Stem Cell*, 11(5), pp. 596–606. doi: 10.1016/j.stem.2012.10.002.

Ge, T., Yang, J., Zhou, S., Wang, Y., Li, Y., and Tong, X. (2020) ‘The role of the pentose phosphate pathway in diabetes and cancer’, *Frontiers in Endocrinology*, 11. doi: 10.3389/fendo.2020.00365.

Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.

Glover, H. J., Holliday, H., Shparberg, R. A., Winkler, D., Day, M., and Morris, M. B. (2023) ‘Signalling pathway crosstalk stimulated by L-Proline drives mouse embryonic stem

cells to primitive-ectoderm-like cells’, *Development*, 150(20), p. dev201704. doi: 10.1242/dev.201704.

Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) ‘L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell’, in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.

Hackett, J. A., Dietmann, S., Murakami, K., Down, T. A., Leitch, H. G., and Surani, M. A. (2013) ‘Synergistic mechanisms of DNA demethylation during transition to ground-state pluripotency’, *Stem Cell Reports*, 1(6), pp. 518–531. doi: <https://doi.org/10.1016/j.stemcr.2013.11.010>.

Hancock, C. N., Liu, W., Alvord, W. G., and Phang, J. M. (2016) ‘Co-regulation of mitochondrial respiration by proline dehydrogenase/oxidase and succinate’, *Amino Acids*, 48(3), pp. 859–872. doi: 10.1007/s00726-015-2134-7.

Harvey, A. J., Rathjen, J., and Gardner, D. K. (2016) ‘Metaboloepigenetic regulation of pluripotent stem cells’, *Stem Cells International*, 2016, pp. 1–15. doi: 10.1155/2016/1816525.

Harvey, N. T., Hughes, J. N., Lonic, A., Yap, C., Long, C., Rathjen, P. D., and Rathjen, J. (2010) ‘Response to BMP4 signalling during ES cell differentiation defines intermediates of the ectoderm lineage’, *Journal of Cell Science*, 123(10), pp. 1796–1804. doi: 10.1242/jcs.047530.

Hovland, A. S., Bhattacharya, D., Azambuja, A. P., Pramio, D., Copeland, J., Rothstein, M., and Simoes-Costa, M. (2022) ‘Pluripotency factors are repurposed to shape the epigenomic landscape of neural crest cells’, *Developmental Cell*, 57(19), pp. 2257–2272.e5. doi: <https://doi.org/10.1016/j.devcel.2022.09.006>.

Hu, C., Fan, L., Cen, P., Chen, E., Jiang, Z., and Li, L. (2016) ‘Energy metabolism plays

a critical role in stem cell maintenance and differentiation’, *International Journal of Molecular Sciences*. Edited by W. Bloch. MDPI, 17(2), p. 253. doi: 10.3390/ijms17020253.

Intlekofer, A. M. and Finley, L. W. S. (2019) ‘Metabolic signatures of cancer cells and stem cells’, *Nature Metabolism*, 1(2), pp. 177–188. doi: 10.1038/s42255-019-0032-0.

Ito, K. and Suda, T. (2014) ‘Metabolic requirements for the maintenance of self-renewing stem cells’, *Nature Reviews Molecular Cell Biology*, 15(4), pp. 243–256. doi: 10.1038/nrm3772.

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I. R. (2006) ‘Dissecting self-renewal in stem cells with RNA interference’, *Nature*, 442(7102), pp. 533–538. doi: 10.1038/nature04915.

Jastroch, M., Divakaruni, A. S., Mookerjee, S., Treberg, J. R., and Brand, M. D. (2010) ‘Mitochondrial proton and electron leaks’, *Essays in Biochemistry*, 47, pp. 53–67. doi: 10.1042/bse0470053.

Khoa, L. T. P., Tsan, Y.-C., Mao, F., Kremer, D. M., Sajjakulnukit, P., Zhang, L., Zhou, B., Tong, X., Bhanu, N. V., Choudhary, C., Garcia, B. A., Yin, L., Smith, G. D., Saunders, T. L., Bielas, S. L., Lyssiotis, C. A., and Dou, Y. (2020) ‘Histone acetyltransferase MOF blocks acquisition of quiescence in ground-state ESCs through activating fatty acid oxidation’, *Cell Stem Cell*, 27(3), pp. 441–458.e10. doi: <https://doi.org/10.1016/j.stem.2020.06.005>.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., and Winter, D. R. (2018) ‘A beginner’s guide to analysis of RNA sequencing data’, *American Journal of Respiratory Cell and Molecular Biology*, 59(2), pp. 145–157. doi: 10.1165/rcmb.2017-0430TR.

Köhler, E., Barrach, H. J., and Neubert, D. (1970) ‘Inhibition of NADP dependent oxidoreductases by the 6-aminonicotinamide analogue of NADP’, *FEBS Letters*, 6(3), pp. 225–

228. doi: 10.1016/0014-5793(70)80063-1.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., de Alencastro, G., Logan, G. J., Teber, E. T., Tam, O. H., Stutz, M. D., Alexander, I. E., Pickett, H. A., and Tam, P. P. L. (2014) 'The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak', *Cell Stem Cell*, 14(1), pp. 107–120. doi: <https://doi.org/10.1016/j.stem.2013.09.014>.

Krishnan, N., Dickman, M. B., and Becker, D. F. (2008) 'Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress', *Free Radical Biology and Medicine*. 2007/11/12, 44(4), pp. 671–681. doi: 10.1016/j.freeradbiomed.2007.10.054.

Krupinski, P., Chickarmane, V., and Peterson, C. (2011) 'Simulating the mammalian blastocyst - molecular and mechanical interactions pattern the embryo', *PLoS Computational Biology*, 7(5), pp. 1–11. doi: 10.1371/journal.pcbi.1001128.

Kuegler, P. B., Zimmer, B., Waldrmann, T., Baudis, B., Ilmjarv, S., Hescheler, J., Gaughwin, P., Brundin, P., Mundy, W., Bal-Price, A. K., Schrattenholz, A., Krause, K.-H., van Thriell, C., Rao, M. S., Kadereit, S., and Leist, M. (2010) 'Markers of murine embryonic and neural stem cells, neurons and astrocytes: reference points for developmental neurotoxicity testing', *ALTEX: Alternatives to Animal Experimentation*, 27(1), pp. 16–42.

Lees, J. G., Gardner, D. K., and Harvey, A. J. (2017) 'Pluripotent stem cell metabolism and mitochondria: beyond ATP', *Stem Cells International*, 2017, pp. 1–17. doi: 10.1155/2017/2874283.

Leese, H. J., Baumann, C. G., Brison, D. R., McEvoy, T. G., and Sturmey, R. G. (2008) 'Metabolism of the viable mammalian embryo: Quietness revisited', *Molecular Human Reproduction*, 14(12), pp. 667–672. doi: 10.1093/molehr/gan065.

Li, L., Liu, C., Biechele, S., Zhu, Q., Song, L., Lanner, F., Jing, N., and Rossant, J. (2013) 'Location of transient ectodermal progenitor potential in mouse development', *Development*, 140(22), pp. 4533–4543. doi: 10.1242/dev.092866.

Liu, W., Hancock, C. N., Fischer, J. W., Harman, M., and Phang, J. M. (2015) 'Proline biosynthesis augments tumor cell growth and aerobic glycolysis: involvement of pyridine nucleotides', *Scientific Reports*, 5(1), p. 17206. doi: 10.1038/srep17206.

Lonic, A. (2007) *Molecular mechanism of L-proline induced EPL-cell formation*. University of Adelaide.

Martano, G., Borroni, E. M., Lopci, E., Cattaneo, M. G., Mattioli, M., Bachi, A., Decimo, I., and Bifari, F. (2019) 'Metabolism of stem and progenitor cells: Proper methods to answer specific questions', *Frontiers in Molecular Neuroscience*, 12, p. 151. doi: 10.3389/fnmol.2019.00151.

Maye, P., Becker, S., Siemen, H., Thorne, J., Byrd, N., Carpentino, J., and Gabel, L. (2004) 'Hedgehog signaling is required for the differentiation of ES cells into neurectoderm', *Developmental Biology*, 265(1), pp. 276–290. doi: <https://doi.org/10.1016/j.ydbio.2003.09.027>.

Michelakis, E. D., Webster, L., and Mackey, J. R. (2008) 'Dichloroacetate (DCA) as a potential metabolic-targeting therapy for cancer', *British Journal of Cancer*, 99(7), pp. 989–994. doi: 10.1038/sj.bjc.6604554.

Minchiotti, G., D'Aniello, C., Fico, A., De Cesare, D., and Patriarca, E. J. (2022) 'Capturing transitional pluripotency through proline metabolism', *Cells*, 11(14), p. 2125. doi: 10.3390/cells11142125.

Nakatake, Y., Fukui, N., Iwamatsu, Y., Masui, S., Takahashi, K., Yagi, R., Yagi, K., Miyazaki, J., Matoba, R., Ko, M. S. H., and Niwa, H. (2006) 'Klf4 cooperates with Oct3/4 and Sox2 to activate the Lefty1 core promoter in embryonic stem cells', *Molecular and Cellular*

Biology, 26(20), pp. 7772 LP – 7782. doi: 10.1128/MCB.00468-06.

Ng, S.-Y., Bogu, G. K., Soh, B. S., and Stanton, L. W. (2013) ‘The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis’, *Molecular Cell*, 51(3), pp. 349–359. doi: <https://doi.org/10.1016/j.molcel.2013.07.017>.

Nichols, J. and Smith, A. (2009) ‘Naive and primed pluripotent states’, *Cell Stem Cell*, 4(6), pp. 487–492. doi: 10.1016/j.stem.2009.05.015.

Niwa, H., Ogawa, K., Shimosato, D., and Adachi, K. (2009) ‘A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells’, *Nature*, 460(7251), pp. 118–122. doi: 10.1038/nature08113.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005) ‘Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells’, *Journal of Biological Chemistry*, 280(7), pp. 5307–5317. doi: 10.1074/jbc.M410015200.

Osteil, P., Studdert, J., Wilkie, E., Fossat, N., and Tam, P. P. L. (2016) ‘Generation of genome-edited mouse epiblast stem cells via a detour through ES cell-chimeras’, *Differentiation*. Elsevier, 91(4–5), pp. 119–125. doi: 10.1016/j.diff.2015.10.004.

Van Otterloo, E., Li, H., Jones, K. L., and Williams, T. (2018) ‘AP-2 α and AP-2 β cooperatively orchestrate homeobox gene expression during branchial arch patterning’, *Development*, 145(2), p. dev157438. doi: 10.1242/dev.157438.

Van Otterloo, E., Milanda, I., Pike, H., Thompson, J. A., Li, H., Jones, K. L., and Williams, T. (2022) ‘AP-2 α and AP-2 β cooperatively function in the craniofacial surface ectoderm to regulate chromatin and gene expression dynamics during facial development’, *eLife*. Edited by P. Trainor and K. S. E. Cheah. eLife Sciences Publications, Ltd, 11, p. e70511. doi: 10.7554/eLife.70511.

Pandhare, J., Cooper, S. K., Donald, S. P., and Phang, J. M. (2009) ‘Regulation and

function of proline oxidase under nutrient stress', *Journal of Cellular biochemistry*, 107(4), pp. 759–768. doi: 10.1002/jcb.22174.

Panopoulos, A. D., Yanes, O., Ruiz, S., Kida, Y. S., Diep, D., Tautenhahn, R., Herrerías, A., Batchelder, E. M., Plongthongkum, N., Lutz, M., Berggren, W. T., Zhang, K., Evans, R. M., Siuzdak, G., and Belmonte, J. C. I. (2012) 'The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming', *Cell Research*, 22(1), pp. 168–177. doi: 10.1038/cr.2011.177.

Patra, K. C. and Hay, N. (2014) 'The pentose phosphate pathway and cancer', *Trends in Biochemical Sciences*. Elsevier Current Trends, 39(8), pp. 347–354. doi: 10.1016/j.tibs.2014.06.005.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., and Rathjen, P. D. (2002) 'Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development', *Journal of Cell Science*, 115(2), pp. 329–339. doi: 10.1073/pnas.95.9.5082.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., Song, L., Tam, P. P. L., Han, J.-D. J., and Jing, N. (2016) 'Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo', *Developmental Cell*, 36(6), pp. 681–697. doi: <https://doi.org/10.1016/j.devcel.2016.02.020>.

Peters, A. K., Steemans, M., Hansen, E., Mesens, N., Verheyen, G. R., and Vanparys, P. (2008) 'Evaluation of the embryotoxic potency of compounds in a newly revised high throughput embryonic stem cell test', *Toxicological Sciences*, 105(2), pp. 342–350. doi: 10.1093/toxsci/kfn126.

Phang, J. M., Donald, S. P., Pandhare, J., and Liu, Y. (2008) 'The metabolism of proline, a stress substrate, modulates carcinogenic pathways', *Amino Acids*, 35(4), pp. 681–690. doi:

10.1007/s00726-008-0063-4.

Phang, J. M. (2012) 'Proline metabolism and cancer', *Frontiers in Bioscience*, 17(1), p. 1835. doi: 10.2741/4022.

Phang, J. M., Liu, W., Hancock, C. N., and Fischer, J. W. (2015) 'Proline metabolism and cancer: emerging links to glutamine and collagen', *Current Opinion in Clinical Nutrition and Metabolic Care*, 18(1), pp. 71–77. doi: 10.1097/MCO.0000000000000121.

Phang, J. M., Liu, W., and Zabirnyk, O. (2010) 'Proline metabolism and microenvironmental stress', *Annual Review of Nutrition*, 30(1), pp. 441–463. doi: 10.1146/annurev.nutr.012809.104638.

De Preter, G., Neveu, M.-A., Danhier, P., Brisson, L., Payen, V. L., Porporato, P. E., Jordan, B. F., Sonveaux, P., and Gallez, B. (2016) 'Inhibition of the pentose phosphate pathway by dichloroacetate unravels a missing link between aerobic glycolysis and cancer cell proliferation', *Oncotarget*. Impact Journals LLC, 7(3), pp. 2910–2920. doi: 10.18632/oncotarget.6272.

Rathjen, J., Lake, J.-A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999) 'Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors', *Journal of Cell Science*, 112(5), pp. 601–612. doi: 10.1242/jcs.004613.

Rathjen, J., Haines, B. P., Hudson, K. M., Nesci, A., Dunn, S., and Rathjen, P. D. (2002) 'Directed differentiation of pluripotent cells to neural lineages: homogeneous formation and differentiation of a neurectoderm population', *Development*, 129(11), pp. 2649–2661. Available at: <http://dev.biologists.org/content/129/11/2649>.

Rathjen, J., Washington, J. M., Bettess, M. D., and Rathjen, P. D. (2003) 'Identification of a biological activity that supports maintenance and proliferation of pluripotent cells from the

primitive ectoderm of the mouse’, *Biology of Reproduction*, 69(6), pp. 1863–1871. doi: 10.1095/biolreprod.103.017384.

Rawat, D. K., Hecker, P., Watanabe, M., Chettimada, S., Levy, R. J., Okada, T., Edwards, J. G., and Gupte, S. A. (2012) ‘Glucose-6-phosphate dehydrogenase and NADPH redox regulates cardiac myocyte L-type calcium channel activity and myocardial contractile function’, *PLoS ONE*. Edited by R. Moreno-Sanchez, 7(10), p. e45365. doi: 10.1371/journal.pone.0045365.

Rehman, J. (2010) ‘Empowering self-renewal and differentiation: the role of mitochondria in stem cells’, *Journal of Molecular Medicine*, 88(10), pp. 981–986. doi: 10.1007/s00109-010-0678-2.

Rodda, S. J., Kavanagh, S. J., Rathjen, J., and Rathjen, P. D. (2002) ‘Embryonic stem cell differentiation and the analysis of mammalian development’, *The International Journal of Developmental Biology*, 46(4), pp. 449–458. doi: 10.1387/IJDB.12141431.

Rodrigues, A. S., Correia, M., Gomes, A., Pereira, S. L., Perestrelo, T., Sousa, M. I., and Ramalho-Santos, J. (2015) ‘Dichloroacetate, the pyruvate dehydrogenase complex and the modulation of mESC pluripotency’, *PLoS ONE*. Edited by P. K. Singh. San Francisco, CA USA: Public Library of Science, 10(7), p. e0131663. doi: 10.1371/journal.pone.0131663.

Rousset, S., Alves-Guerra, M.-C., Mozo, J., Miroux, B., Cassard-Doulcier, A.-M., Bouillaud, F., and Ricquier, D. (2004) ‘The biology of mitochondrial uncoupling proteins’, *Diabetes*, 53, pp. S130–S135. Available at: http://diabetes.diabetesjournals.org/content/53/suppl_1/S130.abstract.

Ryu, J. M., Lee, S. H., Seong, J. K., and Han, H. J. (2015) ‘Glutamine contributes to maintenance of mouse embryonic stem cell self-renewal through PKC-dependent downregulation of HDAC1 and DNMT1/3a’, *Cell Cycle*. Taylor & Francis, 14(20), pp. 3292–

3305. doi: 10.1080/15384101.2015.1087620.

Sanchez, W. Y., McGee, S. L., Connor, T., Mottram, B., Wilkinson, A., Whitehead, J. P., Vuckovic, S., and Catley, L. (2013) 'Dichloroacetate inhibits aerobic glycolysis in multiple myeloma cells and increases sensitivity to bortezomib', *British Journal of Cancer*, 108(8), pp. 1624–1633. doi: 10.1038/bjc.2013.120.

Schaefer, T. and Lengerke, C. (2020) 'SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond', *Oncogene*, 39(2), pp. 278–292. doi: 10.1038/s41388-019-0997-x.

Schorle, H., Meier, P., Buchert, M., Jaenisch, R., and Mitchell, P. J. (1996) 'Transcription factor AP-2 essential for cranial closure and craniofacial development', *Nature*. England, 381(6579), pp. 235–238. doi: 10.1038/381235a0.

Shim, W. J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., Peng, G., Jing, N., Wang, Y., Kellis, M., Tam, P. P. L., Smith, A., Piper, M., Christiaen, L., Nguyen, Q., Bodén, M., and Palpant, N. J. (2020) 'Conserved epigenetic regulatory logic infers genes governing cell identity', *Cell Systems*, 11(6), pp. 625–639. doi: <https://doi.org/10.1016/j.cels.2020.11.001>.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019a) 'Embryoid body differentiation of mouse embryonic stem cells into neurectoderm and neural progenitors', in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019b) 'Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells', *Frontiers in*

Physiology, 10, p. 705. doi: 10.3389/fphys.2019.00705.

Snow, M. H. L. (1977) ‘Gastrulation in the mouse: Growth and regionalization of the epiblast’, *Development*, 42(1), pp. 293–303. Available at: <http://dev.biologists.org/content/42/1/293>.

Sperber, H., Mathieu, J., Wang, Y., Ferreccio, A., Hesson, J., Xu, Z., Fischer, K. A., Devi, A., Detraux, D., Gu, H., Battle, S. L., Showalter, M., Valensisi, C., Bielas, J. H., Ericson, N. G., Margaretha, L., Robitaille, A. M., Margineantu, D., Fiehn, O., Hockenbery, D., Blau, C. A., Raftery, D., Margolin, A. A., Hawkins, R. D., Moon, R. T., Ware, C. B., and Ruohola-Baker, H. (2015) ‘The metabolome regulates the epigenetic landscape during naive-to-primed human embryonic stem cell transition’, *Nature Cell Biology*, 17(12), pp. 1523–1535. doi: 10.1038/ncb3264.

Sun, Y., Shim, W. J., Shen, S., Sinniah, E., Pham, D., Su, Z., Mizikovsky, D., White, M. D., Ho, J. W. K., Nguyen, Q., Bodén, M., and Palpant, N. J. (2023) ‘Inferring cell diversity in single cell data using consortium-scale epigenetic data as a biological anchor for cell identity’, *Nucleic Acids Research*, p. gkad307. doi: 10.1093/nar/gkad307.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. von (2019) ‘STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets’, *Nucleic Acids Research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Tam, P. P. L. and Behringer, R. R. (1997) ‘Mouse gastrulation: The formation of a mammalian body plan’, *Mechanisms of Development*, 68(1–2), pp. 3–25. doi: 10.1016/S0925-4773(97)00123-8.

Tan, B. S. N., Lonic, A., Morris, M. B., Rathjen, P. D., and Rathjen, J. (2011) ‘The amino

acid transporter SNAT2 mediates L-proline-induced differentiation of ES cells', *American Journal of Physiology-Cell Physiology*, 300(6), pp. 1270–1279. doi: 10.1152/ajpcell.00235.2010.

Tan, B. S. N., Rathjen, P. D., Harvey, A. J., Gardner, D. K., and Rathjen, J. (2016) 'Regulation of amino acid transporters in pluripotent cell populations in the embryo and in culture; novel roles for sodium-coupled neutral amino acid transporters', *Mechanisms of Development*. Elsevier B.V., 141, pp. 32–39. doi: 10.1016/j.mod.2016.06.003.

Tanner, J. J. (2008) 'Structural biology of proline catabolism', *Amino Acids*, 35(4), pp. 719–730. doi: 10.1007/s00726-008-0062-5.

Tohyama, S., Tanosaki, S., Someya, S., Fujita, J., and Fukuda, K. (2017) 'Manipulation of pluripotent stem cell metabolism for clinical application', *Current Stem Cell Reports*. Cham: Springer International Publishing, 3(1), pp. 28–34. doi: 10.1007/s40778-017-0073-9.

Tsogtbaatar, E., Landin, C., Minter-Dykhouse, K., and Folmes, C. D. L. (2020) 'Energy metabolism regulates stem cell pluripotency', *Frontiers in Cell and Developmental Biology*, 8, p. 87. doi: 10.3389/fcell.2020.00087.

Tsouko, E., Khan, A. S., White, M. A., Han, J. J., Shi, Y., Merchant, F. A., Sharpe, M. A., Xin, L., and Frigo, D. E. (2014) 'Regulation of the pentose phosphate pathway by an androgen receptor–mTOR-mediated mechanism and its role in prostate cancer cell growth', *Oncogenesis*. Nature Publishing Group, 3(5), pp. e103–e103. doi: 10.1038/oncsis.2014.18.

Varum, S., Rodrigues, A. S., Moura, M. B., Momcilovic, O., Easley IV, C. A., Ramalho-Santos, J., van Houten, B., and Schatten, G. (2011) 'Energy metabolism in human pluripotent stem cells and their differentiated counterparts', *PLoS ONE*, 6(6), pp. 1–15. doi: 10.1371/journal.pone.0020914.

Vlaski-Lafarge, M., Loncaric, D., Perez, L., Labat, V., Debeissat, C., Brunet de la Grange,

P., Rossignol, R., Ivanovic, Z., and Bœuf, H. (2019) 'Bioenergetic changes underline plasticity of murine embryonic stem cells', *Stem Cells*. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY, 37(4), pp. 463–475. doi: 10.1002/stem.2965.

Washington, J. M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M. D., Hamra, N., Semendric, L., Tan, B. S. N., Lake, J. A., Keough, R. A., Morris, M. B., and Rathjen, P. D. (2010) 'L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture', *American Journal of Physiology-Cell Physiology*, 298(5), pp. C982–C992. doi: 10.1152/ajpcell.00498.2009.

Wu, G. (2009) 'Amino acids: metabolism, functions and nutrition', *Amino Acids*, 37(1), pp. 1–17. doi: 10.1007/s00726-009-0269-0.

Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., Liu, L., Liu, Y., Yang, C., Xu, Y., Zhao, S., Ye, D., Xiong, Y., and Guan, K.-L. (2012) 'Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors', *Genes & Development*. 2012/06/07. Cold Spring Harbor Laboratory Press, 26(12), pp. 1326–1338. doi: 10.1101/gad.191056.112.

Xue, K., Ng, J.-H., and Ng, H.-H. (2011) 'Mapping the networks for pluripotency', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1575), pp. 2238–2246. doi: 10.1098/rstb.2011.0005.

Yanes, O., Clark, J., Wong, D. M., Patti, G. J., Sánchez-Ruiz, A., Benton, H. P., Trauger, S. A., Despons, C., Ding, S., and Siuzdak, G. (2010) 'Metabolic oxidation regulates embryonic stem cell differentiation', *Nature Chemical Biology*. Nature Publishing Group, 6(6), pp. 411–417. doi: 10.1038/nchembio.364.

Yang, S.-H., Kalkan, T., Morissroe, C., Marks, H., Stunnenberg, H., Smith, A., and Sharrocks, A. D. (2014) 'Otx2 and Oct4 drive early enhancer activation during embryonic stem

cell transition from naive pluripotency', *Cell Reports*, 7(6), pp. 1968–1981. doi: <https://doi.org/10.1016/j.celrep.2014.05.037>.

Zhang, J., Khvorostov, I., Hong, J. S., Oktay, Y., Vergnes, L., Nuebel, E., Wahjudi, P. N., Setoguchi, K., Wang, G., Do, A., Jung, H.-J., McCaffery, J. M., Kurland, I. J., Reue, K., Lee, W.-N. P., Koehler, C. M., and Teitell, M. A. (2011) 'UCP2 regulates energy metabolism and differentiation potential of human pluripotent stem cells', *The EMBO journal*. England, 30(24), pp. 4860–4873. doi: 10.1038/emboj.2011.401.

Zhou, W., Choi, M., Margineantu, D., Margaretha, L., Hesson, J., Cavanaugh, C., Blau, C. A., Horwitz, M. S., Hockenbery, D., Ware, C., and Ruohola-Baker, H. (2012) 'HIF1 α induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition', *The EMBO Journal*. Nature Publishing Group, 31(9), pp. 2103–2116. doi: 10.1038/emboj.2012.71.

Zhu, W., Gincherman, Y., Docherty, P., Spilling, C. D., and Becker, D. F. (2002) 'Effects of proline analog binding on the spectroscopic and redox properties of PutA', *Archives of Biochemistry and Biophysics*, 408(1), pp. 131–136. doi: 10.1016/S0003-9861(02)00535-0.

Chapter 4: Transcription, epigenetics and L-Proline-mediated metabolism involved in neural lineage commitment from pluripotent mouse ESCs

4.1. Introduction

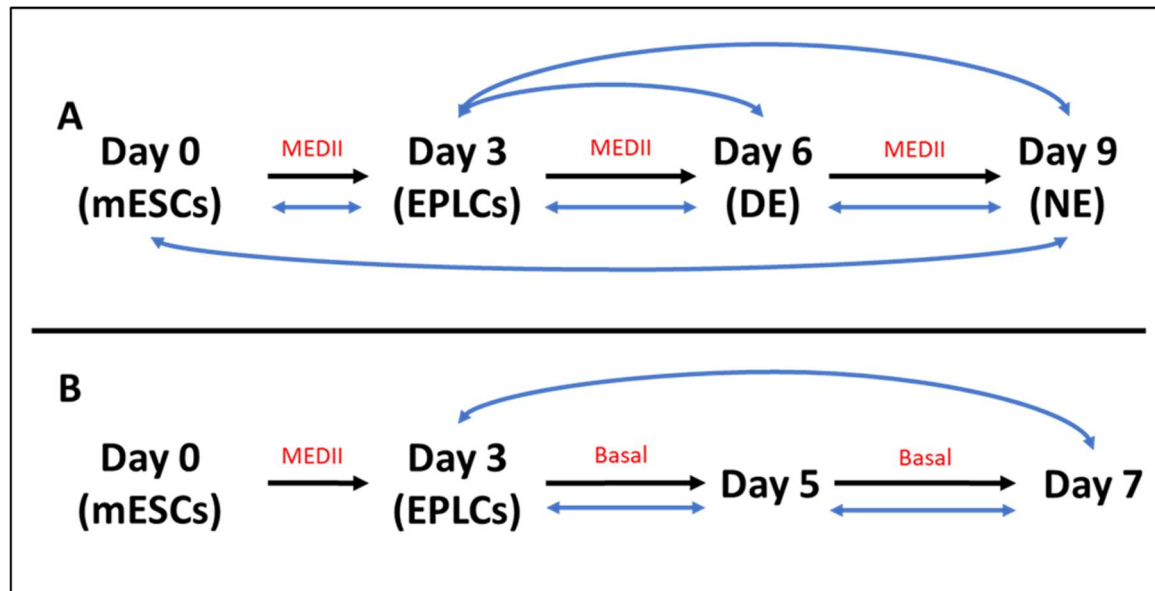
Development of the central nervous system *in vivo* requires transition from the pluripotent inner cell mass to pluripotent primitive ectoderm, which is the precursor to gastrulation and the formation of the three multipotent germ layers: definitive endoderm, definitive mesoderm and definitive ectoderm (Rathjen *et al.*, 1999; Pelton *et al.*, 2002; Nichols and Smith, 2009; Krupinski *et al.*, 2011). The definitive ectoderm further differentiates to become the multipotent cell populations of surface ectoderm and neurectoderm (Inman *et al.*, 2002; Li *et al.*, 2013). It is the multipotent neurectoderm which will set up the central and peripheral nervous systems (Rathjen *et al.*, 2002; Maye *et al.*, 2004; Suter *et al.*, 2009; Dang and Tropepe, 2010; Li *et al.*, 2013). Thus, this multi-step process is essential for the setup of the nervous system and mESCs, which are an *in vitro* model of mammalian development, can be used to study mechanisms at play (Rodda *et al.*, 2002; Shparberg *et al.*, 2019a; Glover *et al.*, 2022).

The conditioned medium MEDII has been used to drive differentiation of mESCs to neural cells (Rathjen *et al.*, 2002; Rodda *et al.*, 2002; Harvey *et al.*, 2010) and later the exogenous addition of L-Proline to culture medium was shown to achieve the same result (Shparberg, 2018). Understanding the mechanisms by which L-Proline effects this differentiation *in vitro* should help shed light on how its presence in uterine and oviductal fluid promotes neural development *in vivo* (Aguilar and Reyley, 2005; Phang *et al.*, 2010, 2015; Krupinski *et al.*, 2011; Patriarca *et al.*, 2021). So far it is known that L-Proline can act via acute activation of signalling pathways (Gangloff *et al.*, 2004; Lonic, 2007; Washington *et al.*, 2010; D’Aniello *et al.*, 2015), through mechanisms involving changes to the epigenetic landscape (Comes *et al.*, 2013; D’Aniello *et al.*, 2017; Harvey *et al.*, 2019; Patriarca *et al.*, 2021), and via some of its metabolites (Phang, Pandhare, *et al.*, 2008; Phang *et al.*, 2010; Casalino *et al.*, 2011; Kilberg *et al.*, 2016; Phang, 2019; Patriarca *et al.*, 2021).

In terms of signalling pathways, changes in pathway activities drive cell fate. In mESCs, leukemia inhibitor factor (LIF) activity and bone morphogenic protein 4 (BMP4) signalling

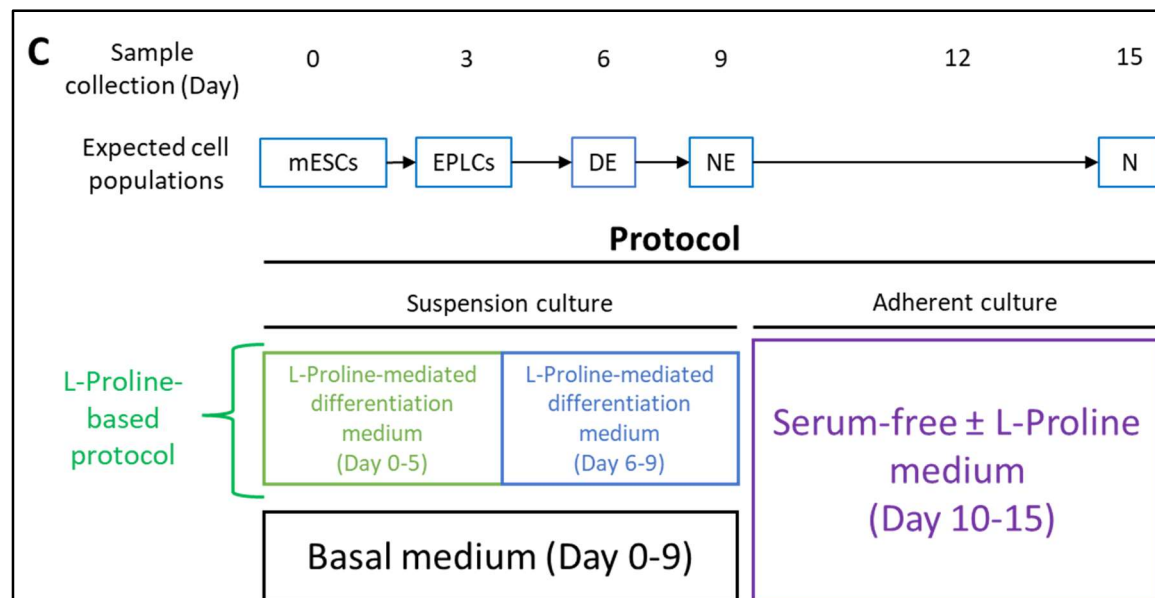
pathway favour self-renewal and maintenance of pluripotency over differentiation (Section 1.2.1). Addition of L-Proline in the presence of LIF and BMP4 tips this balance to promote differentiation via activation of MAPK/ERK and mTOR pathways (Section 1.2.1) (Lonic, 2007; Washington *et al.*, 2010; Tan *et al.*, 2011). As a result, mESCs transition to EPLCs and thereafter to neural differentiation (Shparberg *et al.*, 2019a; Glover *et al.*, 2022).

Although the transcriptional profiles of mESCs and EPLCs have been analysed (Glover, 2018), other intermediate states of neural fate commitment (definitive ectoderm and neurectoderm) are much more poorly understood. This chapter will use: (i) Microarray analysis of mESCs that had undergone differentiation to neurectoderm in MEDII to identify gene expression changes associated with neural fate commitment. (ii) An L-Proline-based protocol using embryoid bodies and hanging drop culture conditions to promote neural differentiation (Shparberg *et al.*, 2019a). (iii) Transcriptional Regulatory Inference Analysis from Gene Expression (TRIAGE) analysis to discover candidate genes that drive neural cell fate (Figure 4.1) (Shim *et al.*, 2020). These approaches can also be used to identify candidate markers of transient cell types, such as definitive ectoderm, for which only two candidate markers, *Penk1* and *Pard6b*, are known (Harvey *et al.*, 2010). This contrasts with the number of well-established markers of mesoderm and endoderm germ layers e.g., *Mixl1*, *T*, *Flk1* and *Sox17* (Arnold and Robertson, 2009; Ishitobi *et al.*, 2011).



Differential gene expression analysis

TRIAGE analysis



Gene expression analysis

Phenotype scoring

Figure 4.1 MEDII-based (A, B) and L-Proline-based (C) protocols used to generate cell types as mESCs differentiate towards neural fate via embryologically relevant cell types. **A.** D3 mESCs (Day 0) were cultured in suspension as embryoid bodies (EBs) in MEDII to produce samples EPLCs (Day 3), the germ layer definitive ectoderm (DE; Day 6) and neurectoderm (NE; Day 9). **B.** As a control to the directed neural differentiation in **A**, different starting populations of D3 mESCs were cultured in MEDII out to Day 3 then switched to Basal medium to allow spontaneous differentiation to all cell types out to Day 5 and Day 7. For **A** and **B**, cDNA samples from individual experiments (n=4) from the cells at the various days were produced, normalised to take into account differences in cDNA concentration, and used to generate microarray data using the *Mus musculus* 22k Affymetrix platform (by Dr J. Rathjen). Hybridisation strategies are shown as double-headed blue arrows. Code in R was produced to work the raw microarray data to graphical form (by Dr C. Wilkinson). Raw data was used for differential gene expression analysis and TRIAGE analysis (by myself) throughout this chapter. **C.** The L-Proline-based protocol (Shparberg *et al.* 2019a) used to generate the same cell types as for the MEDII-based protocol in **A**. D3 mESCs (Day 0) were cultured as hanging drops (1000 cells/20 μ L) in either *L-Proline-mediated differentiation medium* (Day 0-5) or Basal medium up to Day 3, before removal from the hanging drops and continued suspension culture in the respective media. At Day 6 of the L-Proline-based protocol, the medium was changed to *L-Proline-mediated differentiation medium* (Day 6-9), which still contains L-Proline and also SB431542, an inhibitor of mesendoderm production. At Day 9, EBs were transferred to adherent culture in *Serum-free $\pm$ L-Proline medium* (Day 10-15). Samples were collected on Days 0, 3, 6 and 9 for gene expression analysis using qPCR, while EB samples up to Day 15 were scored for neural cell types.

In terms of epigenetic changes, histone modifications drive cell fate (Kingston and Tamkun, 2014; Brand *et al.*, 2019; Harvey *et al.*, 2019; Shim *et al.*, 2020) including via the polycomb group (PcG) and trithorax group (TrxG) protein complexes (Ying, Nichols, *et al.*, 2003; Ringrose and Paro, 2004; Boyer *et al.*, 2006; Ho *et al.*, 2009; MacArthur *et al.*, 2009; Hirai *et al.*, 2011; Lau and So, 2015; Brand *et al.*, 2019). For example, pluripotency is maintained by silencing of differentiation-associated genes and lineage commitment pathways via PcG-mediated transcriptional repression (Ringrose and Paro, 2004; Boyer *et al.*, 2006; MacArthur *et al.*, 2009), while TrxG proteins occupy promoter regions of core pluripotency genes *Oct4*, *Nanog* and *Sox2* to maintain their expression (Ang *et al.*, 2011; Wan *et al.*, 2013; Kingston and Tamkun, 2014; Lau and So, 2015; Tsai *et al.*, 2019). TrxG proteins also occupy promoter regions of genes encoding proteins associated with LIF and BMP signalling to help maintain self-renewal and pluripotency (Ying, Nichols, *et al.*, 2003; Ho *et al.*, 2009; Hirai *et al.*, 2011; Lau and So, 2015). For differentiation to occur, disruption of PcG and TrxG activity is required upon exposure to the right external and/or internal signals (Gao *et al.*, 2008; Morey *et al.*, 2015; Wade *et al.*, 2015; Collinson *et al.*, 2016; Shan *et al.*, 2017; Yao *et al.*, 2018; Brand *et al.*, 2019).

In the case of neural lineage commitment of mESCs, exposure to exogenously added L-Proline disrupts the AAR pathway and activates mTOR and MAPK/ERK signalling pathways, which trigger mESCs to prime themselves for differentiation to EPLCs and neural cell fate (Lonic, 2007; D’Aniello *et al.*, 2015; Shparberg *et al.*, 2019b, 2019a; Glover *et al.*, 2022). How does PcG and TrxG expression respond to exogenously added L-Proline? Does expression of PcG and TrxG complex proteins change such that there is shift from maintaining pluripotency to committing to a neural fate? This Chapter will investigate changes in expression of genes encoding for PcG and TrxG components across the pluripotency spectrum and on towards the production of neurectoderm. Figure 4.1 shows the experimental protocol for generating the various cell types.

RNAseq data (from Chapter 3) and microarray data (Figure 4.1A and 4.1B) will be used to quantify expression of genes which contribute components of PcG and TrxG protein complexes. TRIAGE analysis (Shim *et al.*, 2020) will be used determine if components of these complexes are candidates for controlling cell fate and identity.

A new paradigm in development is that metabolism also contributes to cell fate. For example, mESCs rely on L-Threonine conversion to Glycine, catalysed by threonine dehydrogenase, for survival and self-renewal (Wang *et al.*, 2009; Kilberg *et al.*, 2016). Downregulation of threonine dehydrogenase activity causes mESCs to differentiate (Wang *et al.*, 2009; Shyh-Chang, Ng *et al.*, 2013). L-Threonine restriction in mESC culture causes reduction in proliferation rates, represses self-renewal markers and upregulates expression of trophoctoderm and mesoderm markers (Ryu and Han, 2011). L-Threonine metabolism is also coupled to epigenetic modifications, including maintenance of the S-adenosylmethionine (SAM):S-adenosylhomocysteine (SAH) ratio (Shyh-Chang, Ng *et al.*, 2013). A high SAM:SAH ratio in mESCs results in increasing trimethylation of histone H3 lysine 4 (H3K4me3), which is essential for maintenance of proliferation, self-renewal and pluripotency (Bernstein *et al.*, 2006; Wang *et al.*, 2009, 2011; Shyh-Chang, N *et al.*, 2013). mESCs also rely on L-Glutamine for pluripotency maintenance and cell-cycle regulation (Ryu *et al.*, 2015). These effects depend on L-Glutamine's conversion to L-Glutamate, catalysed by glutaminase, as inhibition of this enzyme represses expression of core pluripotency gene *Oct4* and cell-cycle regulatory proteins in mESCs (Ryu *et al.*, 2015). In addition, L-Glutamate's conversion to alpha-ketoglutarate (α -KG) contributes to the α -KG-to-succinate ratio (Ryu *et al.*, 2015). A high α -KG-to-succinate ratio results to inhibition of DNA methyltransferases, Dnmt3a and Dnmt3b (Ryu *et al.*, 2015; Shparberg *et al.*, 2019b). This leads to an increase in histone acetylation and DNA hypomethylation, which occurs on the *Oct4* gene, resulting to increasing Oct4 production and, therefore, core pluripotency activity (Ryu *et al.*, 2015; Shparberg *et al.*, 2019b).

On the other hand, L-Proline stimulates a different outcome in mESCs. Endogenous synthesis of L-Proline is maintained via the amino-acid starvation response (AAR) pathway, which keeps endogenous production of L-Proline low (~25 μ M). This helps maintain mESC identity and limits cell proliferation (D’Aniello *et al.*, 2015; Tsogtbaatar *et al.*, 2020). Exogenous addition of L-Proline overwhelms the AAR pathway, via uptake of L-Proline into the mESCs by SNAT2, shifting the balance of signalling pathway activities for mESCs to now promote differentiation and increased proliferation (Tan *et al.*, 2011; D’Aniello *et al.*, 2015). The excess L-Proline undergoes catabolism, catalysed by proline oxidase (POX), resulting in a change in mESC colony morphology to that of EPLCs (Casalino *et al.*, 2011; Glover, 2018). POX inhibition results in inhibition of these morphological changes (Casalino *et al.*, 2011), although this depends on LIF concentrations (Chapter 3). Exogenous L-Proline added to mESCs also increases trimethylation of histone H3 lysine 9 (H3K9me3) and H3 lysine 36 (H3K36me3), inducing changes in the epigenome and remodels the mESC transcriptome to permit induction of the formation of EPLC (Comes *et al.*, 2013; Harvey *et al.*, 2016).

It’s not known, however, if L-Proline-mediated changes in metabolism continue to play a role in driving EPLCs to a neural cell fate. This Chapter will use microarray analysis and TRIAGE to investigate the expression of genes encoding for metabolic enzymes as mESCs commit to a neural fate (Figure 4.1).

4.2. Aims and hypotheses

This chapter will focus on:

- (i) Using an L-Proline-based differentiation protocol on mESCs to drive neural differentiation, which incorporates a hanging drop methodology.
- (ii) Analysing a microarray data set of mESCs differentiated to neurectoderm using the conditioned medium MEDII. The analysis will include quantifying relative changes in gene expression and TRIAGE transformation.

We hypothesise that these approaches will help identify candidate genes that control differentiation to neural fate via embryologically relevant intermediate cell types, including genes encoding for PcG and TrxG proteins and metabolic enzymes.

4.3. Results

4.3.1. Differential expression analysis of genes associated with pluripotency and lineage restriction in MEDII-treated embryoid bodies

Changes in the transcriptome are indicative of changing cell identity, and can be used to identify mechanisms driving cell fate (Glover, 2018). For this Chapter, differential expression analysis of genes associated with pluripotency and lineage restriction were performed on ESCs driven to a neural fate in the conditioned medium MEDII, and compared to MEDII-treated ESCs left to spontaneously differentiate (Figure 4.1A, 4.1B and 4.2).

D3 naïve-state mESCs (Day 0) were cultured in suspension as EBs in MEDII up to Day 3, 6 and 9, representative of EPLCs, DE and NE, respectively (Figure 4.1A) (Dr J. Rathjen and Dr C. Wilkinson, University of Adelaide). A different set of D3 naïve-state mESCs were cultured in suspension as EBs in MEDII up to Day 3, then switched to Basal medium to allow spontaneous differentiation to all cell types up to Day 5 and 7 (Figure 4.1B) (Dr J. Rathjen and Dr C. Wilkinson, University of Adelaide). These samples were used to produce separate but related microarray datasets (Figure 4.1A-B) (Dr. J. Rathjen and Dr. C. Wilkinson, University of Adelaide).

We interrogated these datasets to look at gene expression profiles of core and extended pluripotency network genes, and lineage restriction genes (Figure 4.2). To limit false positives (i.e. reduce the false discovery rate (FDR)) and restrict analysis of genes to those whose expression changed significantly over the course of differentiation, we used a Benjamini-Hochberg adjusted $P < 0.05$, regardless of $\log_2$ fold change (Koch *et al.*, 2018).

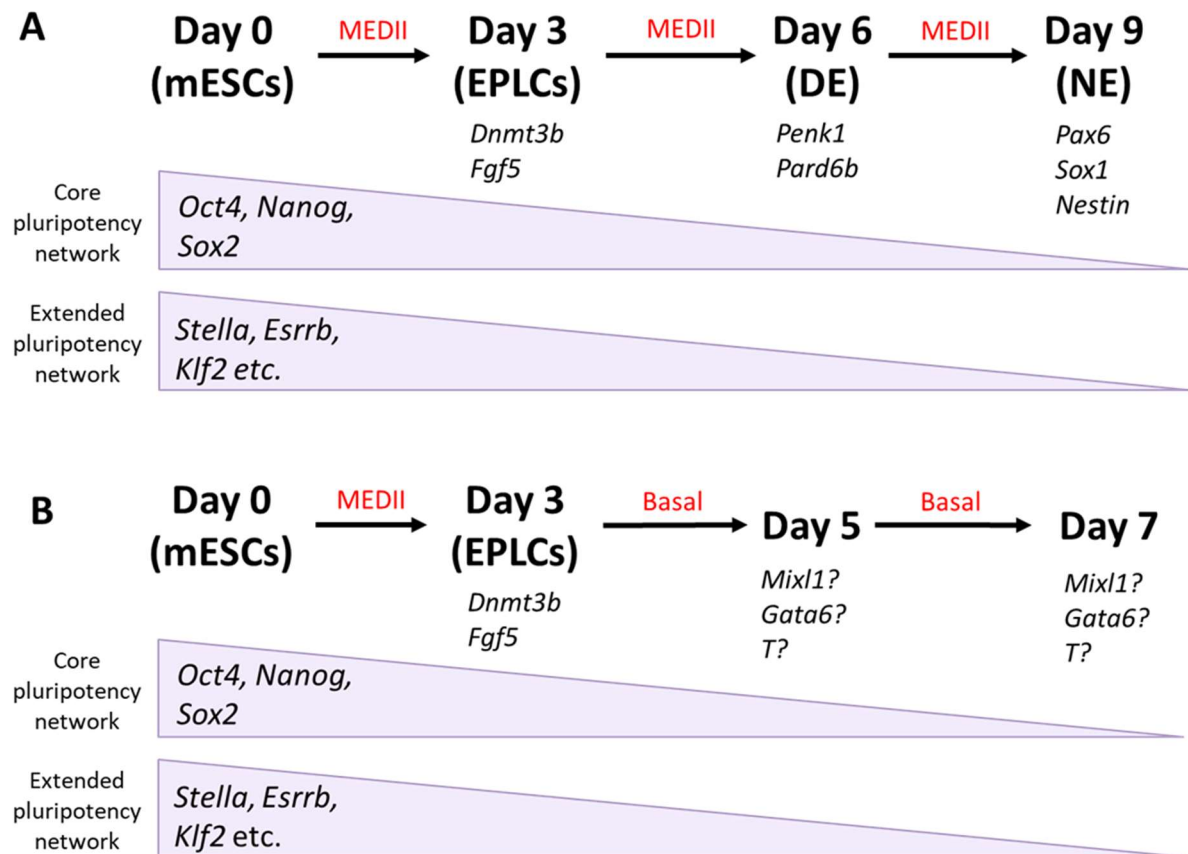


Figure 4.2 Pluripotency network and lineage restriction genes in a MEDII-based protocol driving mESCs to neural fate. **A.** D3 mESCs (Day 0) were cultured in suspension as EBs in MEDII to Day 3, 6 and 9, representative of EPLCs, the germ layer definitive ectoderm (DE) and neurectoderm (NE), respectively, as per Figure 4.1A. **B.** Different sets of D3 mESCs were cultured in MEDII out to Day 3 then switched to Basal medium to allow spontaneous differentiation to all cell types out to Day 5 and 7, as per Figure 4.1B. Microarray data was generated (by Dr J. Rathjen and Dr C. Wilkinson). The raw data was used by us for differential gene expression analysis. Some gene markers of interest are listed under each intermediate stage of neural fate commitment. The expected changes in expression of core and some extended pluripotency network genes are shown as mESCs differentiate. It is unknown if mesendoderm markers, such as *Mixl1*, *Gata6* and *T*, will be expressed in Day 5 and 7 samples.

Figure 4.3A shows that for MEDII-treated naïve mESCs, the expression of core and most extended pluripotency network genes was downregulated as they differentiated over the course of 9 days, consistent with the loss of pluripotency.

There is also significant upregulation of expression of EPLC markers at Day 3 and 6, definitive ectoderm markers (*Pard6b* and *Penk1*) at Day 6, then neurectoderm markers (*Pax6* and *Sox1*) by Day 9. These results are consistent with a ‘temporal wave’ of differentiation as the naïve mESCs differentiate via embryologically relevant intermediates.

There is no change in expression of the surface ectoderm marker, which is a branch point of differentiation at the definite ectoderm stage, suggesting that those cells that do become definitive ectoderm proceed to neurectoderm instead of surface ectoderm (Figure 4.3A). On the other hand, there is upregulation of the expression of germ layer markers for mesoderm and endoderm, suggesting that while MEDII promotes neural commitment, there is some off-pathway differentiation.

Figure 4.3B helps confirm these findings. Here, expression is measured relative to Day 3 (EPLCs) for both the MEDII-treated and spontaneously differentiated cells in Basal medium. The MEDII-treated cells show strong downregulation of the primitive ectoderm markers by Day 9 compared to spontaneously differentiated cells, and also show upregulation of the expression of definitive ectoderm and neurectoderm markers, which the spontaneously differentiated cells do not.

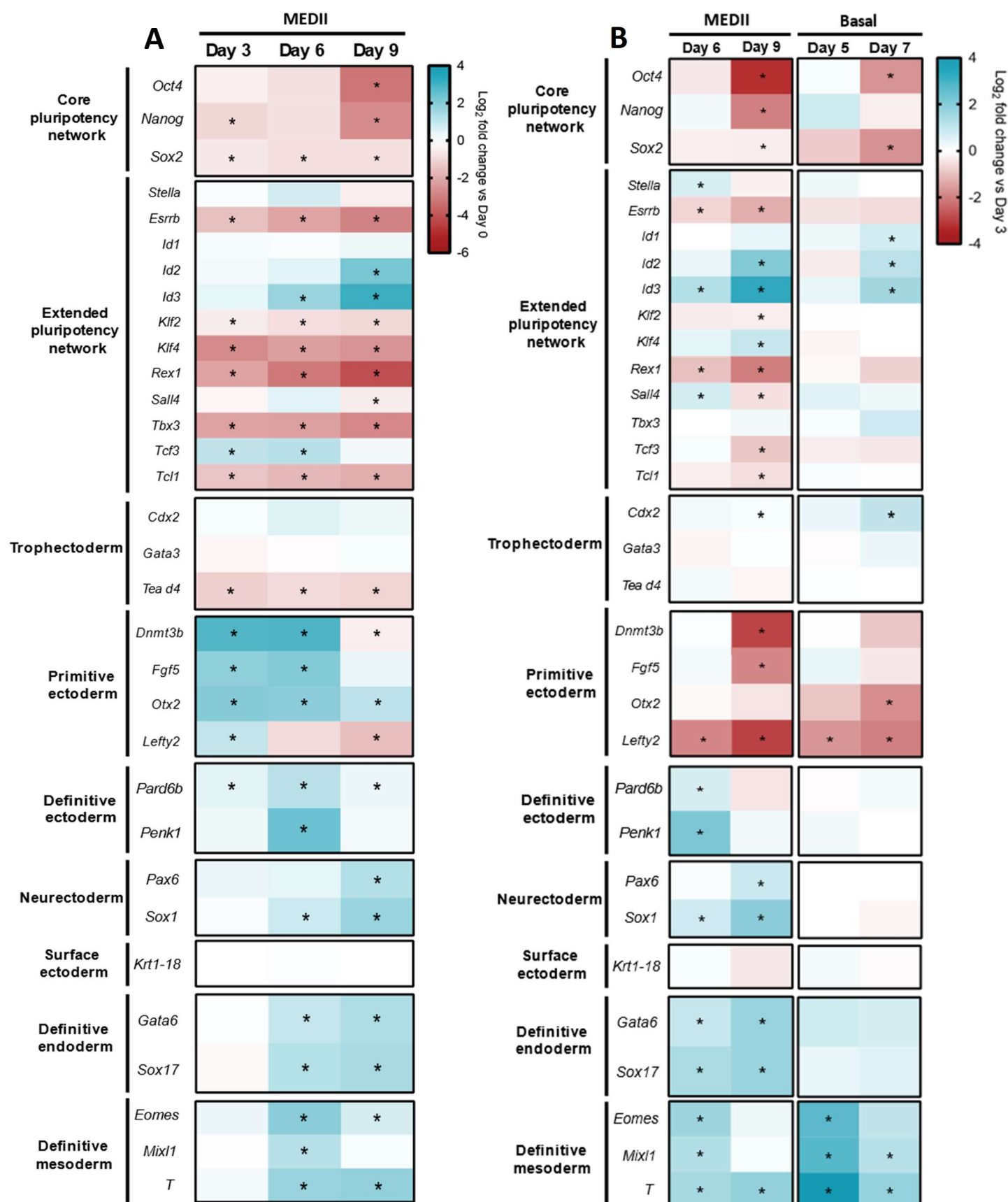


Figure 4.3 Differential expression analysis of genes associated with pluripotency and lineage restriction in MEDII-treated and spontaneously differentiated naïve ESCs. The differentiation protocol is shown in Figure 4.1A and 4.1B. cDNA samples (n=4) (produced by Dr J. Rathjen) were hybridized onto the *Mus musculus* 22k Affymetrix platform. Microarray data was generated and normalized to take into account differences in cDNA concentration across the experiments and days. Raw data was extracted on Rstudio by Dr C. Wilkinson. We selected the marker genes for this analysis and generated the heatmaps using Graphpad Prism. **A.** Relative expression for MEDII-treated cells compared to naïve mESCs (Day 0). **B.** Relative expression of MEDII-treated cells and spontaneously differentiated cells compared to Day 3 MEDII-treated cells (EPLCs). * Indicates significant change in expression as measured by a Benjamini-Hochberg adjusted $P < 0.05$.

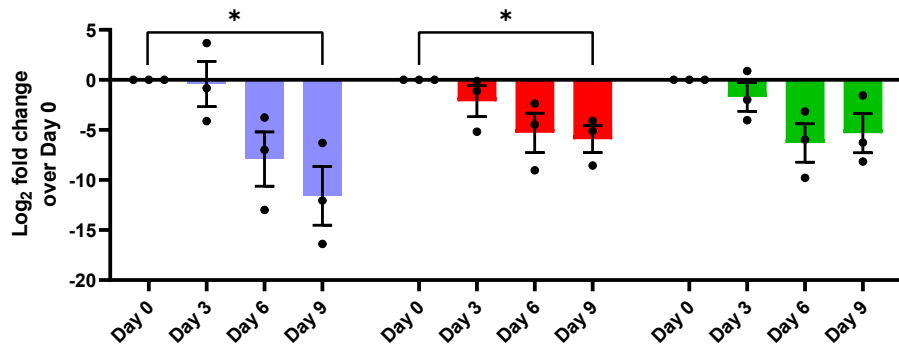
4.3.2. Gene expression profile and phenotype changes across neural lineage commitment

L-Proline is one of the bioactive components in MEDII and promotes the transition of naïve mESCs to EPLCs in adherent culture (Washington *et al.*, 2010). More recently, L-Proline has been shown to promote the differentiation of naïve mESCs all the way out to mature neural cells when cells are grown in suspension as embryoid bodies (EBs) (Shparberg, 2018; Shparberg *et al.*, 2019a).

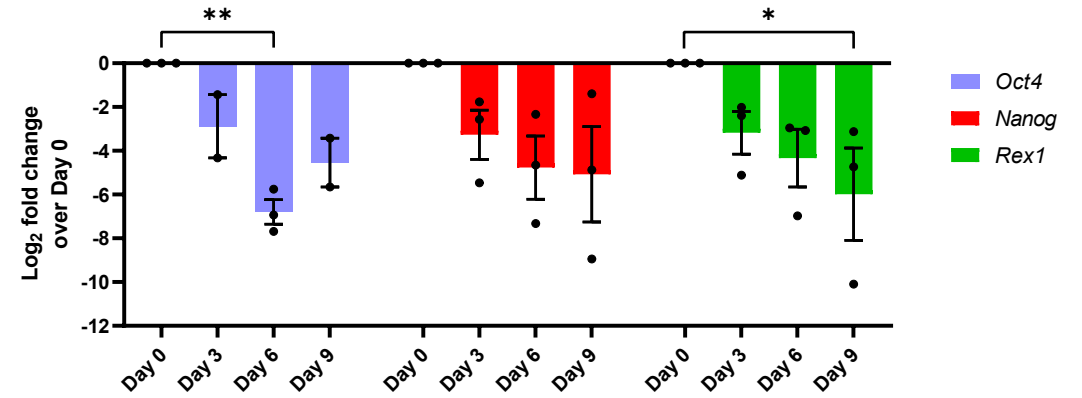
Here, we modified this L-Proline-based suspension protocol by incorporating a hanging drop step (Ohnuki and Kurosawa, 2013) in the first 3 days of suspension culture to control cell number and proliferation. Figure 4.1C shows the protocol. The rationale to adjust the first 3 days of suspension culture to include hanging drop was that cell number and size of EBs are factors affecting efficiency of directed neural differentiation using other (non-proline-based) protocols (Ohnuki and Kurosawa, 2013).

Figure 4.4 shows qPCR results for this modified protocol as the EBs differentiated over 9 days in the presence of L-Proline or allowed to spontaneously differentiate by the removal of L-Proline after Day 3. Most of the expected temporal changes in gene expression for pluripotency factors, EPLC, definitive ectoderm and neurectoderm markers as the cell differentiated were not observed with the hanging drop modification as they were with the L-Proline protocol without hanging drop (Shparberg, 2018; Shparberg *et al.*, 2019a) or with MEDII-treated cells (Figure 4.3). This may be because the expression of many of these markers is very restricted in time and the hanging drop samples may not have been collected at the best days to observe this. However, the spontaneously differentiated cells did show reduced expression of the neurectoderm marker *Nestin* at Day 9, whereas the L-Proline-treated cells at least maintained the expression of this marker (Figure 4.4C). Spontaneously differentiated cells also failed to express *Dnmt3b* or *Sox1* (data not shown).

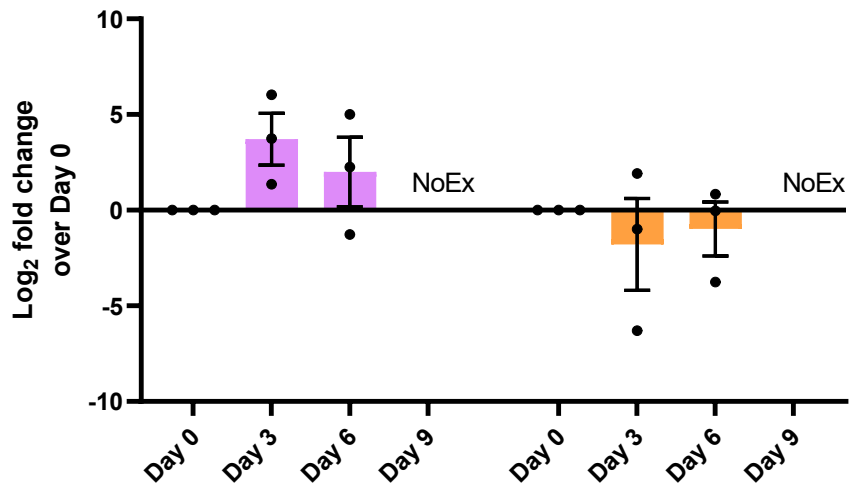
A i) + L-Proline



ii) Basal



B i) + L-Proline



ii) Basal

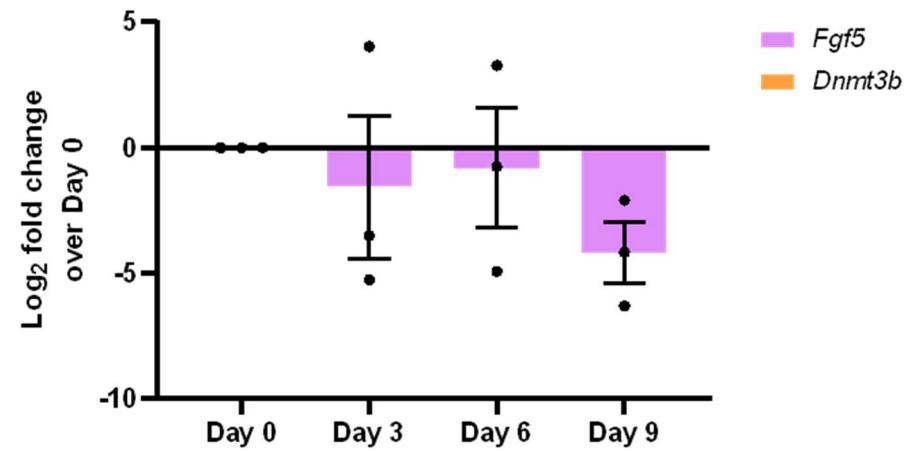
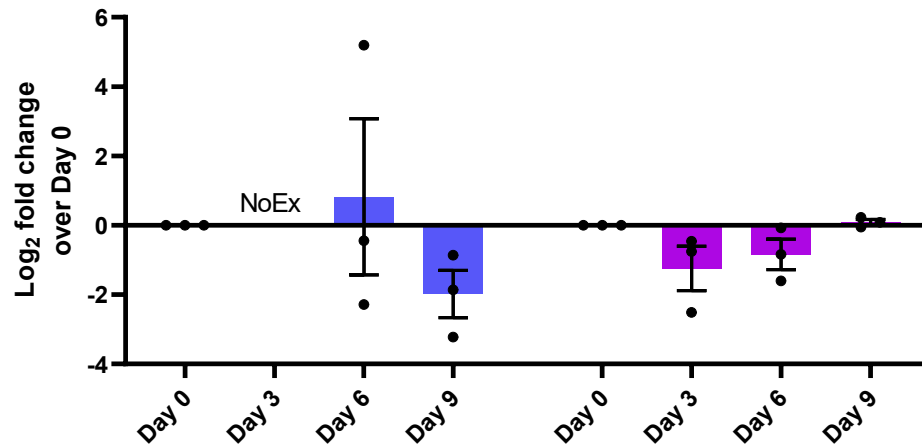
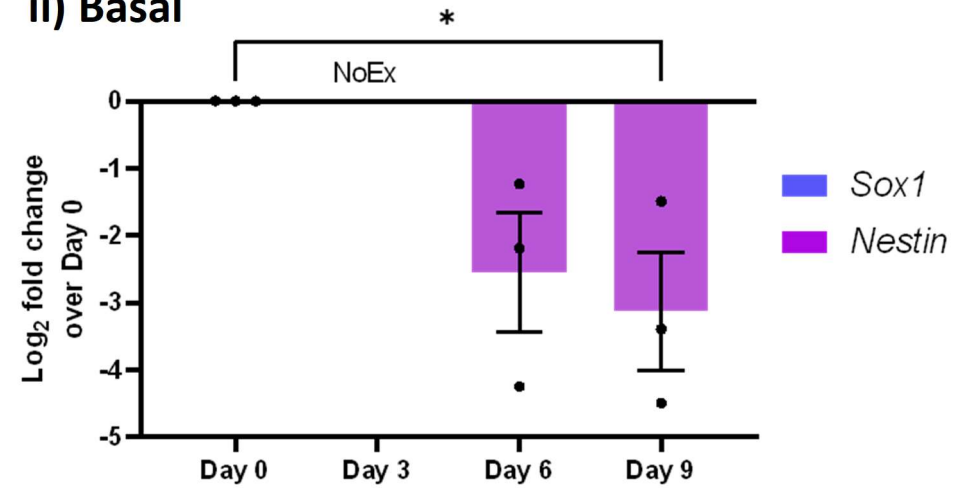


Figure 4.4 continued...

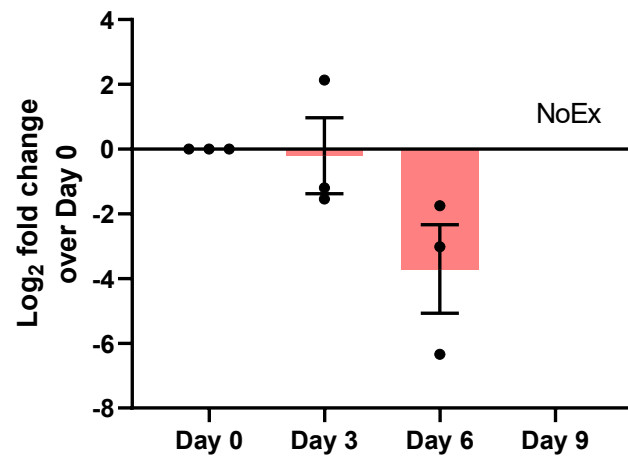
C i) + L-Proline



ii) Basal



D i) + L-Proline



ii) Basal

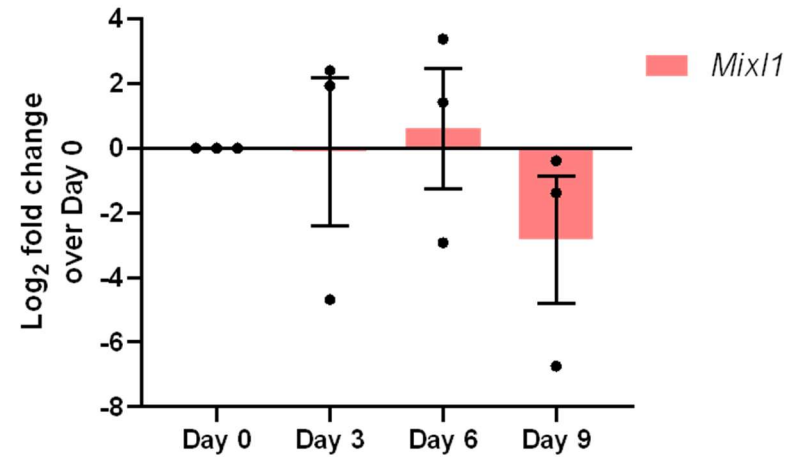


Figure 4.4 Gene expression profiles of EBs cultured by the hanging drop method in L-Proline for neural lineage commitment or in Basal medium to allow spontaneous differentiation. EBs derived from naïve-state mESCs (Day 0) were produced using hanging drop method in either **i)** L-Proline-based or **ii)** Basal medium for 9 days, as per Figure 4.1C. On Days 0, 3, 6 and 9, samples were collected, processed and subjected to qPCR. Results for each sample were normalised to the expression of *18s* as the housekeeping gene, and then shown relative to expression of the naïve state (Day 0) as a log₂ fold change. **A.** Pluripotency-associated genes, *Oct4*, *Nanog* and *Rex1*. **B.** EPLC markers, *Fgf5* and *Dnmt3b*. **C.** Genes associated with neural lineage commitment, *Sox1* and *Nestin*. **D.** Mesendoderm lineage commitment marker, *Mixl1*. Statistical analysis is by one-way ANOVA with Dunnett's *post hoc* test relative to Day 0. Bars represent mean $\pm$ SEM, $n \geq 3$, * $P < 0.05$, ** $P < 0.01$. NoEx: not expressed.

Despite not showing the expected marker-gene expression profile, when these L-Proline-treated hanging drop EBs were seeded down at Day 9 and allowed to grow out to Day 15, 50-70% of the EBs showed neural features (Figure 4.5), significantly higher than for spontaneously differentiated EBs (~10%) (Figure 4.5).

In addition, L-Proline was added to the serum-free medium at this late stage (Day 9) to see if it improved the percentage of EBs which showed neural features, but it did not do so (Figure 4.5).

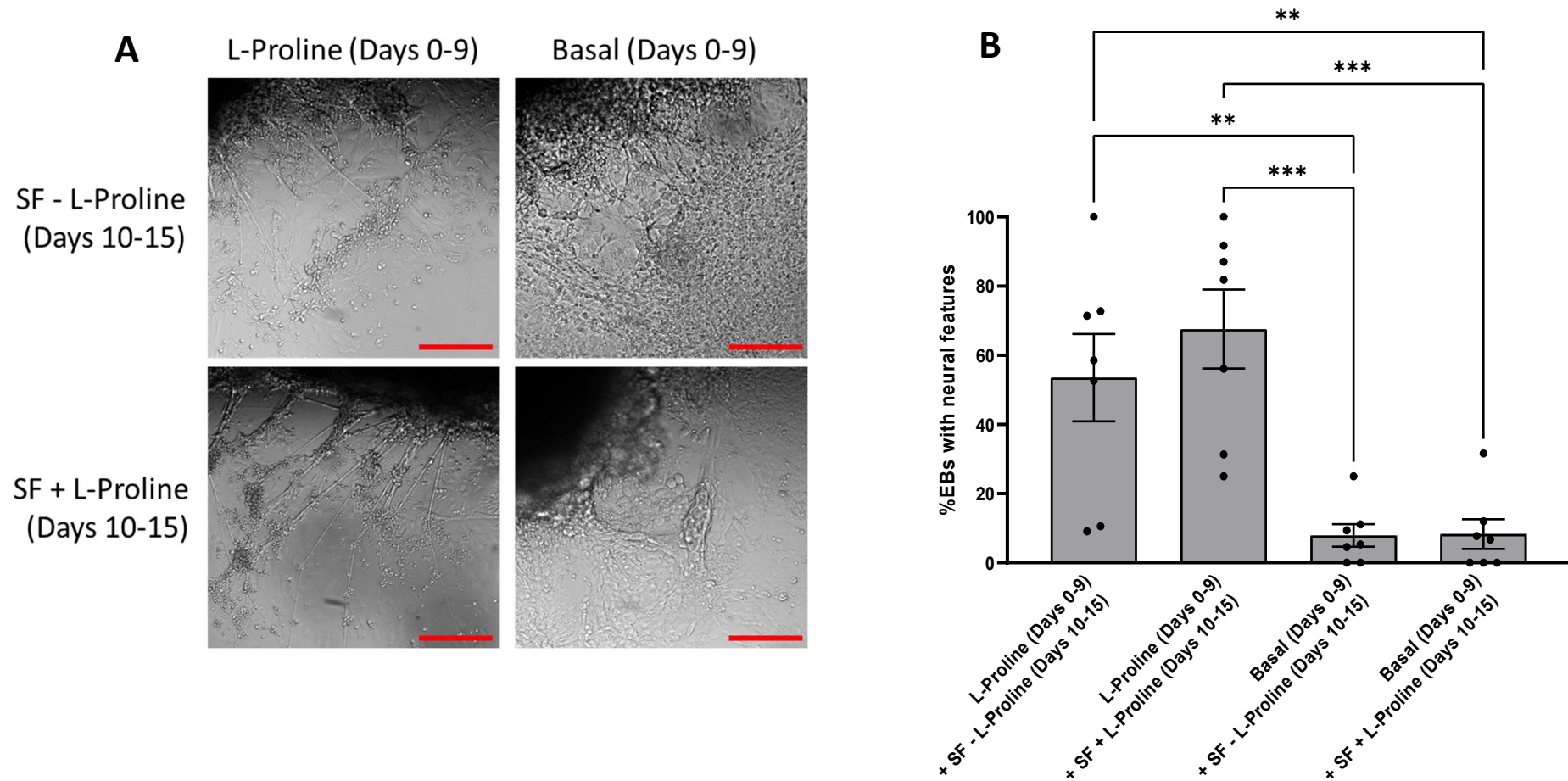


Figure 4.5 EBs scored for neural features following seeding down at Day 9 after suspension culture. EBs derived from D3 mESCs produced in suspension culture (Days 0-9) were transferred to serum-free medium (SF) $\pm$ L-Proline in adherent culture for another 6 days up to Day 15, as per Figure 4.1C. **A.** Representative images of EB morphology at Day 15, imaged at 10x magnification, showing in this case numerous neurites and axonal bundles in the L-Proline protocol but not in the spontaneous differentiation protocol. Scale bar = 200 μ m. **B.** EBs were scored based on a set of morphological images predetermined previously in a published protocol (Shparberg *et al.* 2019a). Statistical analysis is by one-way ANOVA with Tukey's *post-hoc*. Bars represent mean $\pm$ SEM, $n = 7$, ** $P < 0.01$ *** $P < 0.001$.

4.3.3. *TRIAGE transformation provides candidate genes controlling differentiation of mESC to neural fate*

Part of initiating changes in cell fate requires epigenetic changes through histone modifications (Kingston and Tamkun, 2014; Brand *et al.*, 2019; Harvey *et al.*, 2019; Shim *et al.*, 2020). Histone modifications activate or repress genes to guide cell fate decisions. H3K27me3 is a histone mark that indicates gene repression. Selective absence of H3K27me3 domains can be used as predictors of cell-type-specific and transient gene expression driving cell-fate decisions (Shim *et al.*, 2020). TRIAGE utilises this repressive mark to prioritise genes associated with cell fate and identity over genes associated with housekeeping and cellular structure (Shim *et al.*, 2020).

Normally, TRIAGE is advantageous in that analysis is usually isolated to just the sample of interest and does not require a comparison with the gene expression profile of a different cell type along, for example, a differentiation pathway (Shim *et al.*, 2020). However, in this case, the microarray data available to us only shows relative changes in gene expression compared to a selected day along the differentiation process. Here, we use TRIAGE to transform absolute fold changes relative to Day 3 (EPLCs) (Figure 4.1A and 4.1B) to calculate discordance scores (DS). This score depends on 2 things: (i) the extent of gene expression and (ii) the degree of association of H3K27me3 marks (or the tendency of gene repression) observed across the genome (Shim *et al.*, 2020).

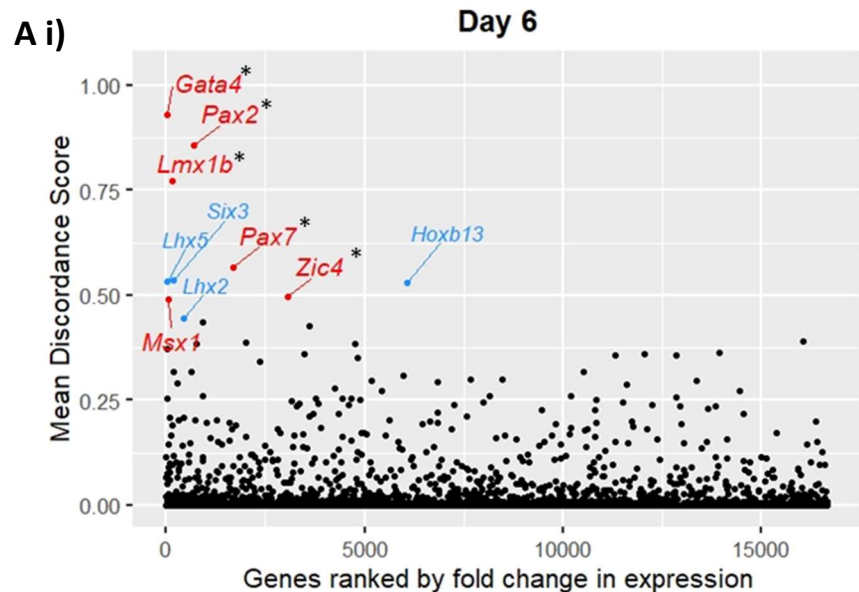
DS were ranked in descending order across the entire mouse microarray dataset; i.e., highest DS is ranked first. A DS rank ≤ 1359 is considered to be a gene likely associated with cell fate and cell identity (Shim *et al.*, 2020). Beyond this rank a gene is likely associated with housekeeping and structural maintenance (Shim *et al.*, 2020).

Figure 4.6A shows DS for Day 6 (definitive ectoderm) and Day 9 (neurectoderm) cells following treatment with MEDII plotted against the absolute fold change in expression relative

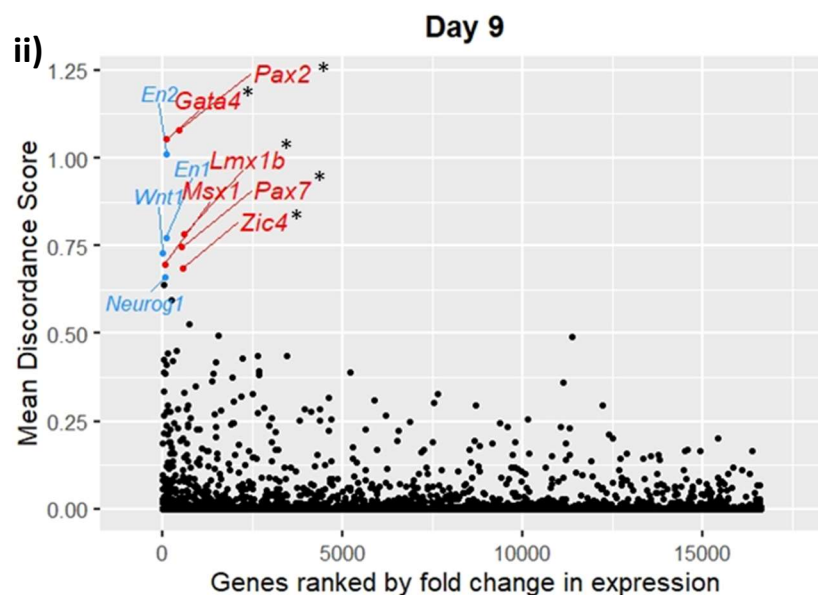
to that at Day 3 (EPLCs). 6 genes (*Gata4*, *Pax2*, *Lmx1b*, *Pax7*, *Zic4* and *Msx1*) that appear in the top 10 DS ranks are common between Day 6 and 9, while 4 are unique to each day. The 4 genes that appear on Day 6 (*Hoxb13*, *Lhx2*, *Lhx5* and *Six3*) also appear on Day 9 at DS ranks 15, 12, 16 and 13, respectively. The 4 genes that appear on Day 9 (*En1*, *En2*, *Wnt1* and *Neurog1*) also appear on Day 6 at DS ranks 11, 13, 81 and 28, respectively.

Figure 4.6B shows DS for Day 5 and Day 7 for spontaneously differentiated cells plotted against the absolute fold change in expression relative to that at Day 3 (EPLCs). 8 genes (*Gata4*, *Pax2*, *Lmx1b*, *Hoxb13*, *Pax7*, *Alx4*, *En2* and *Zic4*) that appear in the top 10 DS ranks are common between Day 5 and 7, while 2 are unique to each day. The 2 genes that appear on Day 5 (*En1* and *Pax5*) also appear on Day 7 at DS ranks 18 and 15, respectively. The 2 genes that appear on Day 7 (*Hand1* and *Cdx2*) also appear on Day 5 at DS ranks 36 and 19, respectively.

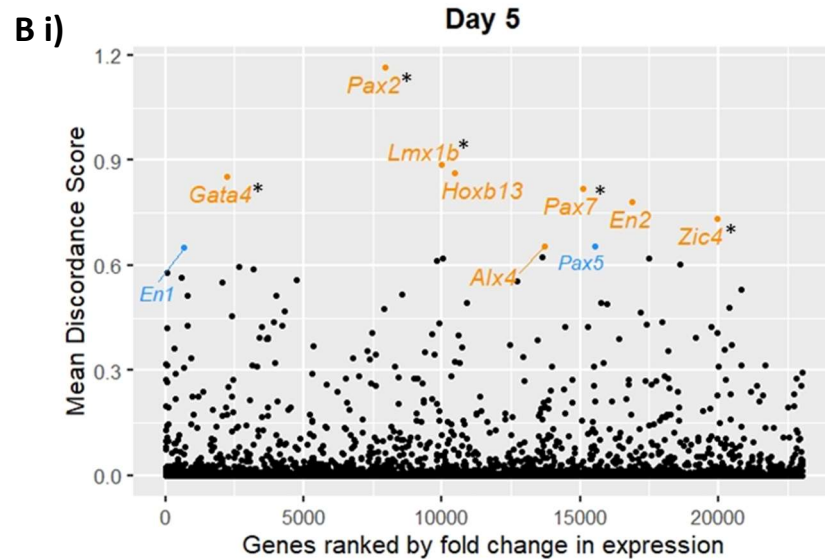
Across all these days and treatments, 4 genes in the top 10 DS-ranked are common, which all encode for transcription factors (*Gata4*, *Pax2*, *Pax7* and *Zic4*).



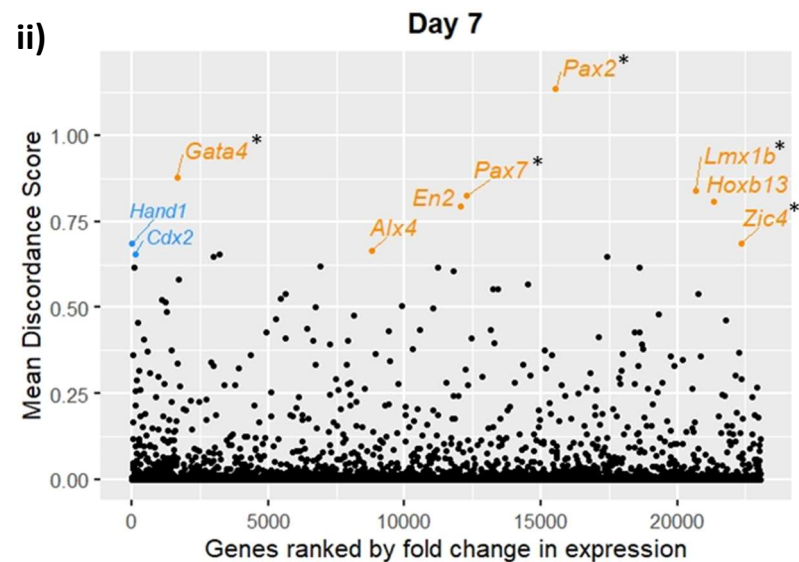
DS rank	Gene	Rank by fold change in expression	Gene product function
1	<i>Gata4</i>	26	Transcription factor
2	<i>Pax2</i>	694	Transcription factor
3	<i>Lmx1b</i>	169	Transcription factor
4	<i>Pax7</i>	1690	Transcription factor
5	<i>Six3</i>	189	Transcription factor
6	<i>Lhx5</i>	27	Transcription factor
7	<i>Hoxb13</i>	6074	Transcription factor
8	<i>Zic4</i>	3071	Transcription factor
9	<i>Msx1</i>	53	Transcription factor
10	<i>Lhx2</i>	443	Transcription factor



DS rank	Gene	Rank by fold change in expression	Gene product function
1	<i>Pax2</i>	448	Transcription factor
2	<i>Gata4</i>	105	Transcription factor
3	<i>En2</i>	110	Transcription factor
4	<i>Lmx1b</i>	601	Transcription factor
5	<i>En1</i>	109	Transcription factor
6	<i>Pax7</i>	547	Transcription factor
7	<i>Wnt1</i>	1	Cell signalling
8	<i>Msx1</i>	58	Transcription factor
9	<i>Zic4</i>	560	Transcription factor
10	<i>Neurog1</i>	91	Transcription factor



DS rank	Gene	Rank by fold change in expression	Gene product function
1	<i>Pax2</i>	7949	Transcription factor
2	<i>Lmx1b</i>	9999	Transcription factor
3	<i>Hoxb13</i>	10459	Transcription factor
4	<i>Gata4</i>	2216	Transcription factor
5	<i>Pax7</i>	15080	Transcription factor
6	<i>En2</i>	16886	Transcription factor
7	<i>Zic4</i>	19948	Transcription factor
8	<i>Pax5</i>	15524	Transcription factor
9	<i>Alx4</i>	13731	Transcription factor
10	<i>En1</i>	687	Transcription factor



DS rank	Gene	Rank by fold change in expression	Gene product function
1	<i>Pax2</i>	15540	Transcription factor
2	<i>Gata4</i>	1644	Transcription factor
3	<i>Lmx1b</i>	20656	Transcription factor
4	<i>Pax7</i>	12290	Transcription factor
5	<i>Hoxb13</i>	21315	Transcription factor
6	<i>En2</i>	12030	Transcription factor
7	<i>Hand1</i>	4	Transcription factor
8	<i>Zic4</i>	22363	Transcription factor
9	<i>Alx4</i>	8775	Transcription factor
10	<i>Cdx2</i>	114	Transcription factor

Figure 4.6 TRIAGE transformation of microarray data for mESCs differentiated in MEDII to neural-cell fate (A) or allowed to spontaneously differentiate (B). D3 mESCs were cultured in suspension as embryoid bodies (EBs) in MEDII out to Day 3. These EBs continued to be cultured in **A.** MEDII out to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm) or switched (with a different set of Day 3 EBs) to **B.** Basal medium to allow spontaneous differentiation out to **i)** Day 5 and **ii)** Day 7 (as per Figure 4.1A and 4.1B). cDNA samples (n=4) were used to produce the microarray data. Raw data was extracted on Rstudio, then fold changes relative to Day 3 (EPLCs) were calculated. These results were TRIAGE transformed to produce DS plotted against ranked absolute fold change relative to Day 3 (EPLCs). For clarity, only the top 10 ranked DS genes are shown. Genes in common in for MEDII-treated cells at Day 6 and 9 are shown in red. Genes in common in spontaneously differentiated cells at Day 5 and 7 are shown in orange. Genes in blue indicate genes unique to that day and treatment. * indicates genes in the top 10 DS common across all 4 samples. The function of each gene product is also shown.

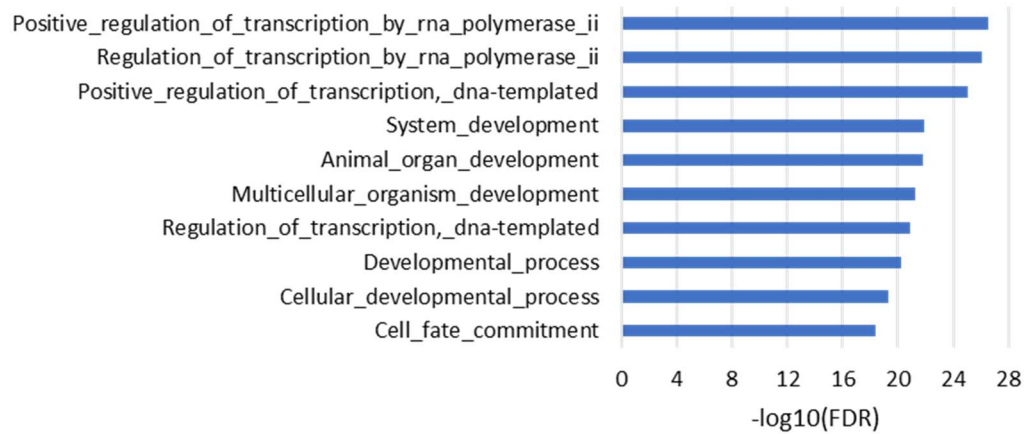
For further interpretation of the TRIAGE-transformed data, the top 100 genes based on DS rank were converted to their respective human gene orthologs and underwent gene clustering analysis using TRIAGE-Parser (Sun *et al.*, 2023). The method allows unbiased grouping of these genes into clusters, which are assigned a random number to distinguish between them. These numbered clusters are then defined using gene ontology (GO) enrichment analysis (Sun *et al.*, 2023). Protein-protein interactions (PPIs) were also predicted for each cluster using the STRING database (Szklarczyk *et al.*, 2019). A PPI enrichment P value <0.001 indicates these interactions are biologically meaningful (Sun *et al.*, 2023).

Figure 4.7A shows that for MEDII-treated cells, Day 6 samples grouped into 5 gene clusters (assigned as Cluster 1-5). Clusters 1-3 and Cluster 5 contained significant GO terms (FDR < 0.05). The Clusters have extensive lists of GO terms associated with transcription, metabolism, development, and lineage commitment, including development of the nervous system by their Benjamini-Hochberg adjusted P values beyond the top 10. All 4 gene clusters are biologically significant with Clusters 1, 2 and 5 having PPI enrichment P values $<1.0^{-16}$, and Cluster 3 with PPI enrichment P value <0.001 .

Figure 4.7B shows that for MEDII-treated cells, Day 9 samples grouped into 3 gene clusters (assigned as Cluster 1-3). Cluster 1 and 3 contained significant GO terms (FDR < 0.05). Again, both clusters contain GO terms associated with transcription, metabolism, development, and lineage commitment (Benjamini-Hochberg adjusted P values beyond the top 10), including nervous system development. Both clusters are biologically significant having PPI enrichment P values $<1.0^{-16}$.

A. Day 6

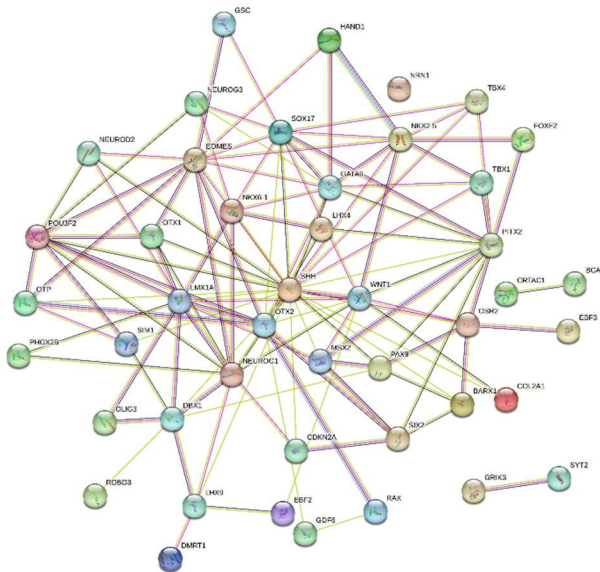
Gene Cluster 1 GO: Biological Processes



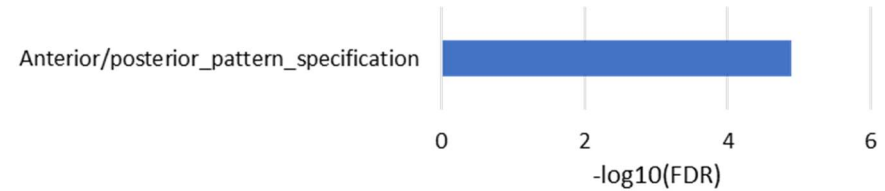
Gene Cluster 1

Gene	Rank by mean DS
<i>DMRT1</i>	97
<i>EOMES</i>	45
<i>PHOX2B</i>	67
<i>GATA6</i>	70
<i>SOX17</i>	17
<i>WNT1</i>	81
<i>SYT2</i>	73
<i>GSC</i>	86
<i>CDKN2A</i>	89
<i>CRTAC1</i>	54

PPI enrichment $P < 1.0^{-16}$



Gene Cluster 2 GO: Biological Process



Gene Cluster 2

Gene	Rank by mean DS
<i>HOXC13</i>	50
<i>HOXC10</i>	19
<i>HOXC4</i>	24
<i>HOXC12</i>	66
<i>HOXC5</i>	77
<i>HOXC9</i>	84

PPI enrichment $P < 1.0^{-16}$

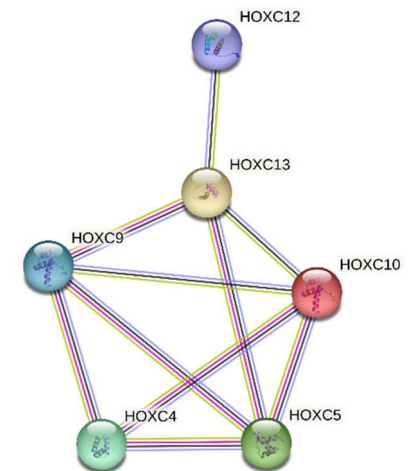
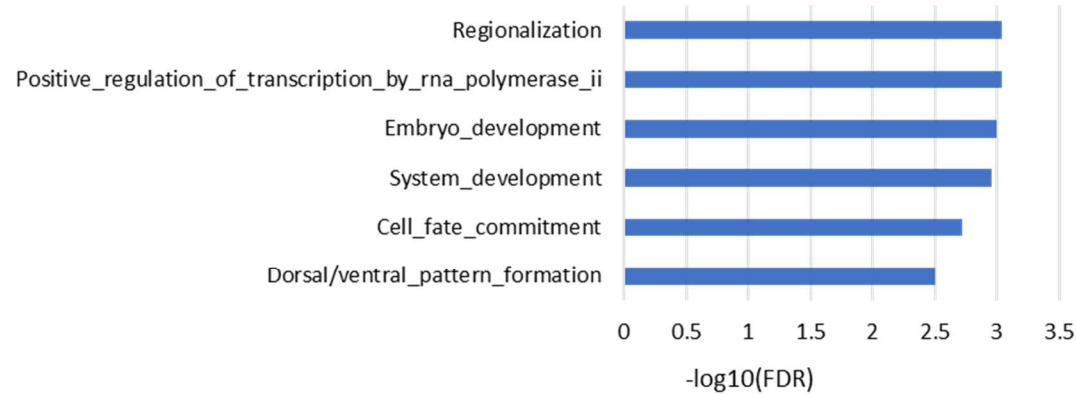


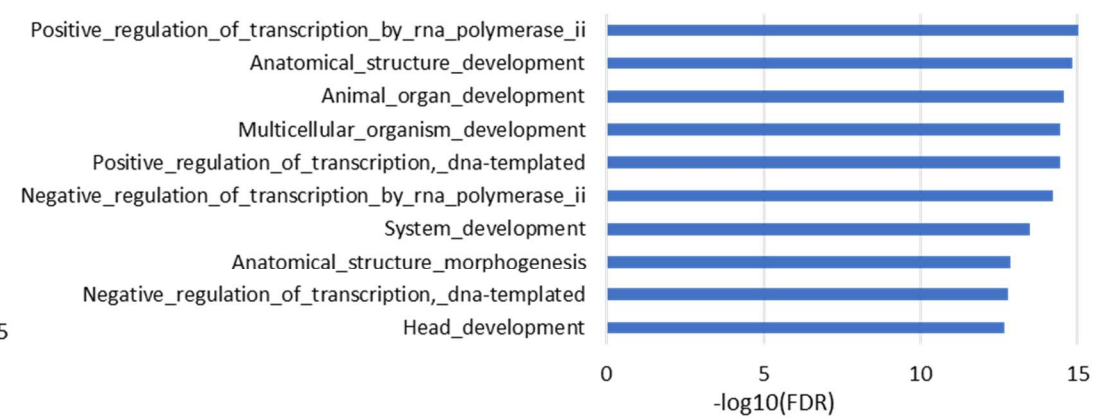
Figure 4.7A continued...

A. Day 6

Gene Cluster 3 GO: Biological Processes



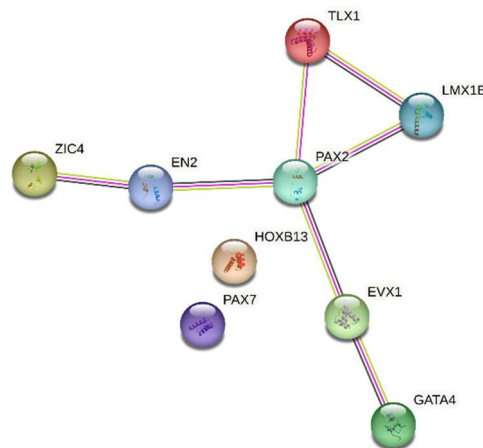
Gene Cluster 5 GO: Biological Processes



Gene Cluster 3

Gene	Rank by mean DS
ZIC4	8
PAX7	4
LMX1B	3
GATA4	1
TLX1	99
PAX2	2
EN2	13
HOXB13	7
EVX1	69

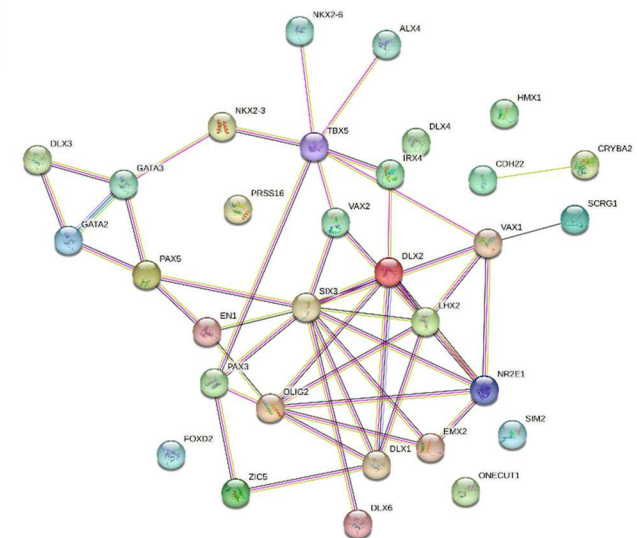
PPI enrichment $P = 1.05^{-08}$



Gene Cluster 5

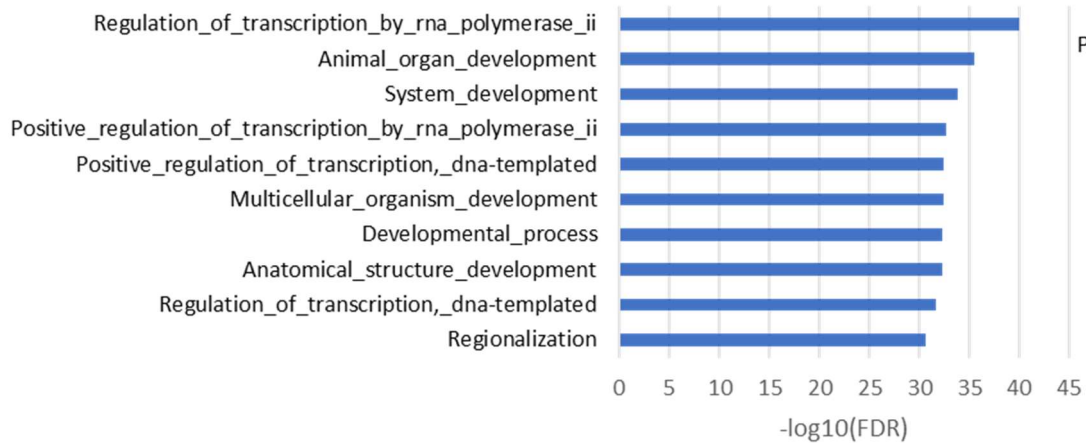
Gene	Rank by mean DS
HMX1	20
PRSS16	80
CDH22	15
TBX5	32
PAX5	12
DLX4	29
SIM2	21
DLX3	62
FOXD2	64
ZIC5	98

PPI enrichment $P < 1.0^{-16}$

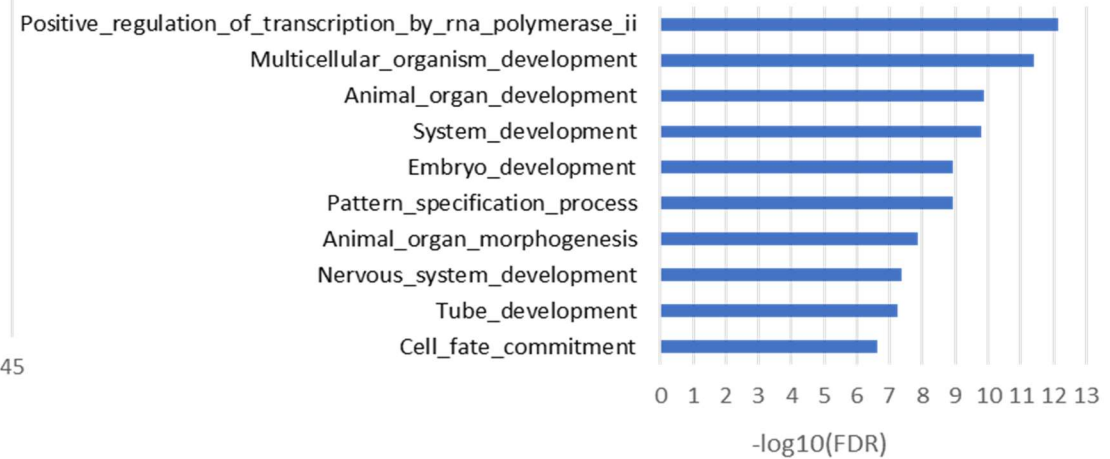


B. Day 9

Gene Cluster 1 GO: Biological Processes



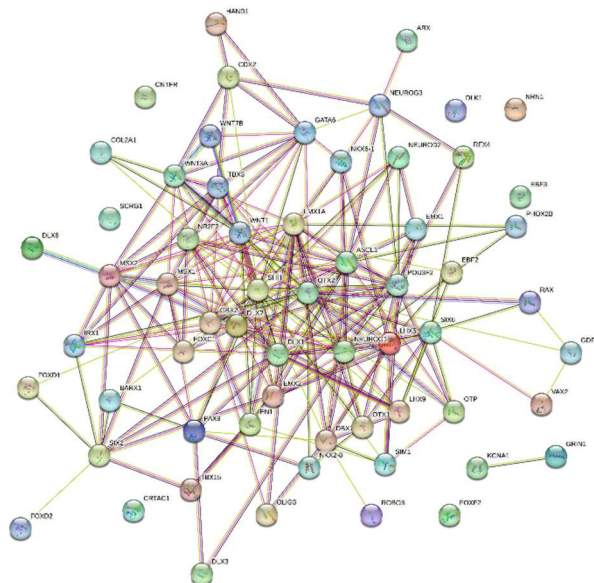
Gene Cluster 3 GO: Biological Processes



Gene Cluster 1

Gene	Rank by mean DS
NR2F2	68
SCRG1	67
NKX2-8	52
TBX3	62
PHOX2B	87
LMX1A	61
NEUROG1	10
PAX9	55
COL2A1	21
SIX6	100

PPI enrichment $P < 1.0^{-16}$



Gene Cluster 3

Gene	Rank by mean DS
GATA4	2
EN2	3
PAX2	1
HOXB13	15
LMX1B	4
PAX7	6
ZIC4	9
PAX5	14
CDH22	20
NKX2-3	69

PPI enrichment $P < 1.0^{-16}$

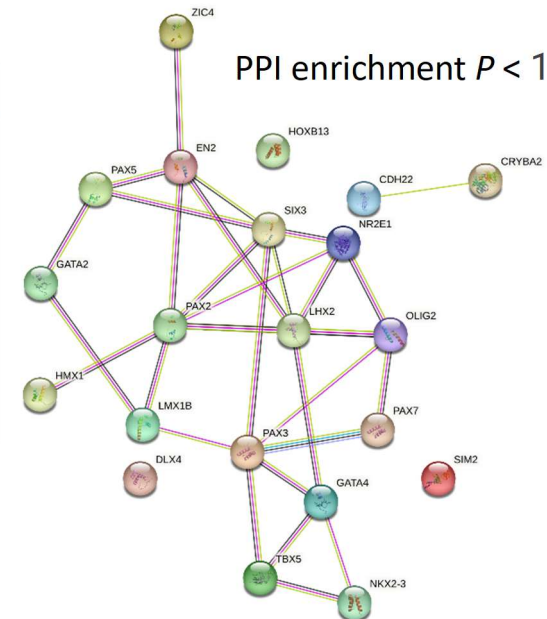
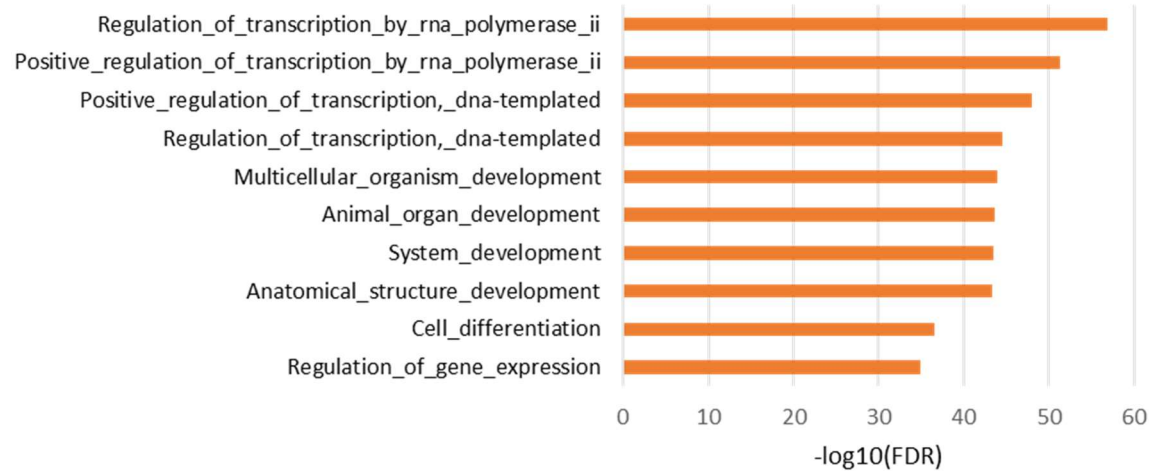


Figure 4.7 TRIAGE-Parser analysis of genes with the top 100 DS rank for cells at Day 6 and 9 differentiated in MEDII to neural fate. D3 mESCs were cultured in suspension as EBs in MEDII out to Day 3 (EPLCs), with culture continued in MEDII out to **A.** Day 6 (definitive ectoderm) and **B.** Day 9 (neurectoderm), as per Figure 4.1A. cDNA samples (n=4) were used to produce microarray data from which fold changes in gene expression were calculated relative to Day 3. The fold changes were TRIAGE transformed to produce discordance scores (DS) and ranked by mean DS across n=4. The top 100 ranked DS underwent TRIAGE-Parser to identify GO terms (FDR < 0.05), genes in the clusters (confidence threshold $\geq 95\%$) and predicted protein-protein interactions (PPIs) (STRING v 11.5) with default settings. PPI enrichment FDR *P* value < 0.001 indicates significance and, therefore, biologically meaningful PPIs. Only the top 10 significant GO terms are shown for simplicity.

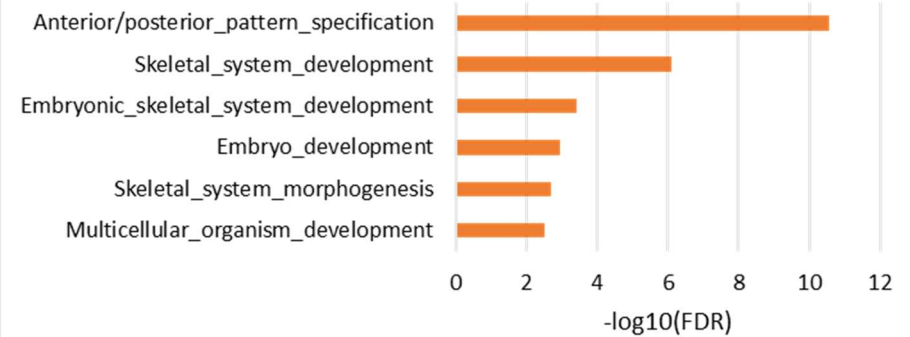
Figure 4.8 shows the results for Day 3 (EPLCs) allowed to spontaneously differentiate in Basal medium out to Day 5 and 7. While some of the GO terms are shared with the MEDII-treated cells, by and large the genes in the clusters and the PPIs they are associated with are very different.

A. Day 5

Gene Cluster 1 GO: Biological Processes

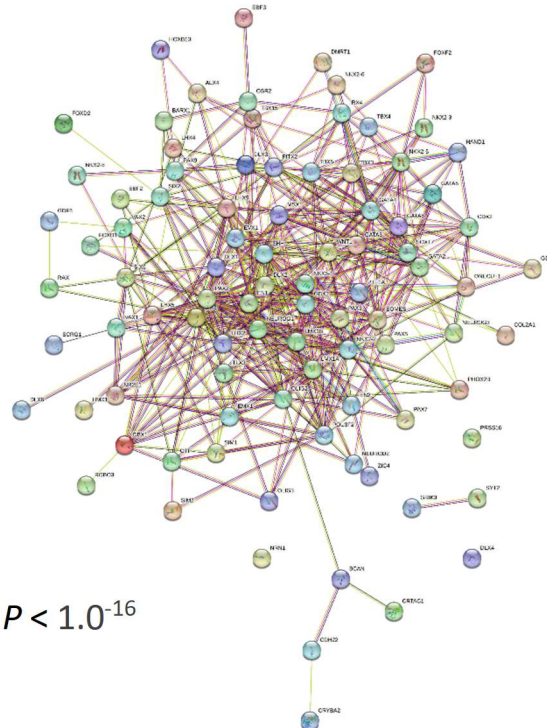


Gene Cluster 2 GO: Biological Processes



Gene Cluster 1

Gene	Rank by mean DS
PAX2	1
LMX1B	2
GATA4	3
HOXB13	4
PAX7	5
EN2	6
MSX1	7
ZIC4	8
EN1	9
ALX4	10

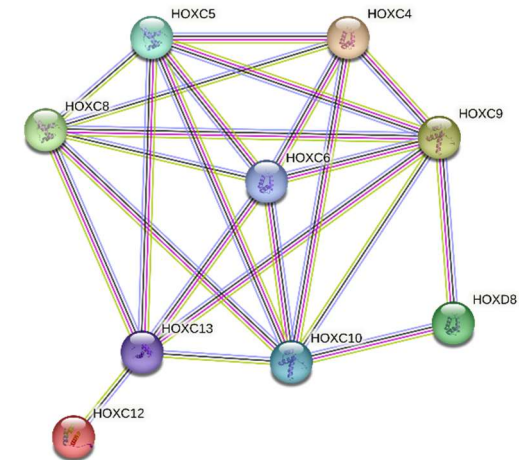


PPI enrichment $P < 1.0^{-16}$

Gene Cluster 2

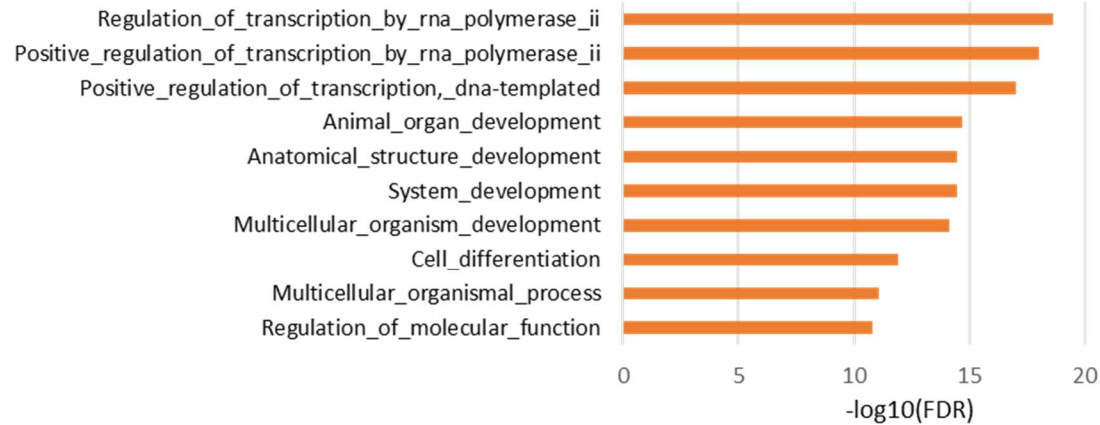
Gene	Rank by mean DS
HOXC4	26
HOXC5	87
HOXC6	96
HOXC8	94
HOXC9	85
HOXC10	13
HOXC12	49
HOXC13	57
HOXD8	22

PPI enrichment $P < 1.0^{-16}$

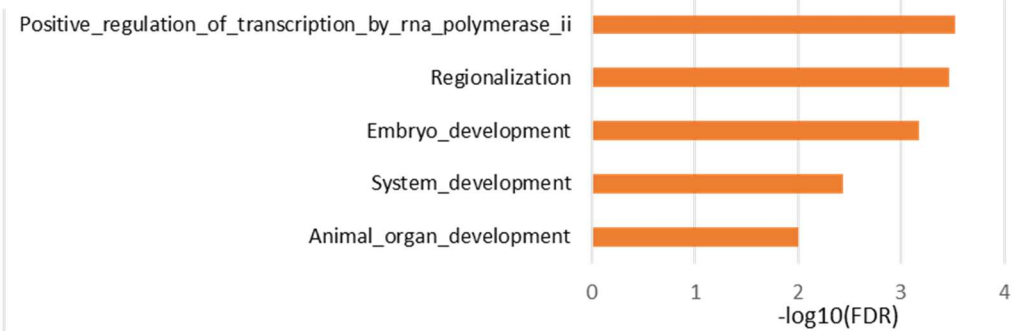


B. Day 7

Gene Cluster 2 GO: Biological Processes



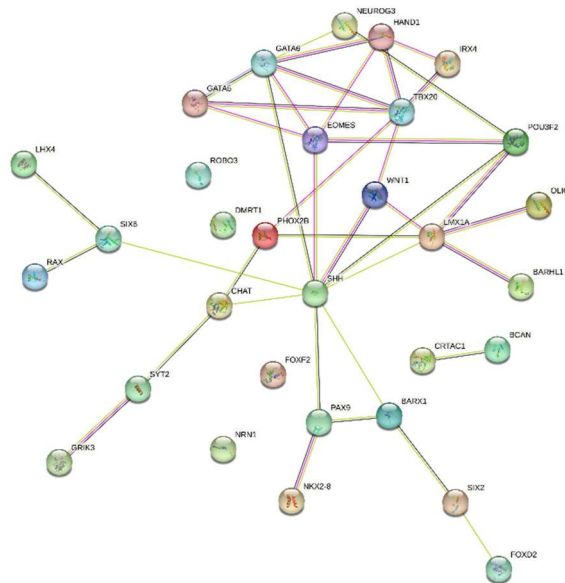
Gene Cluster 4 GO: Biological Processes



Gene Cluster 2

Gene	Rank by mean DS
<i>GATA5</i>	97
<i>PHOX2B</i>	59
<i>NKX2-8</i>	91
<i>NEUROG3</i>	81
<i>SYT2</i>	83
<i>PAX9</i>	42
<i>POU3F2</i>	90
<i>OLIG3</i>	49
<i>WNT1</i>	98
<i>FOXD2</i>	80

PPI enrichment $P < 1.0 \cdot 10^{-16}$



Gene Cluster 4

Gene	Rank by mean DS
<i>PAX2</i>	2
<i>GATA4</i>	3
<i>LMX1B</i>	7
<i>EN2</i>	8
<i>HOXB13</i>	9
<i>HOXD8</i>	16
<i>EVX1</i>	57
<i>TLX1</i>	79

PPI enrichment $P = 2.19 \cdot 10^{-8}$

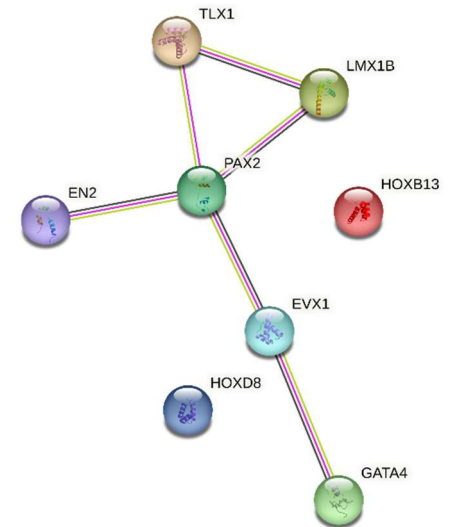
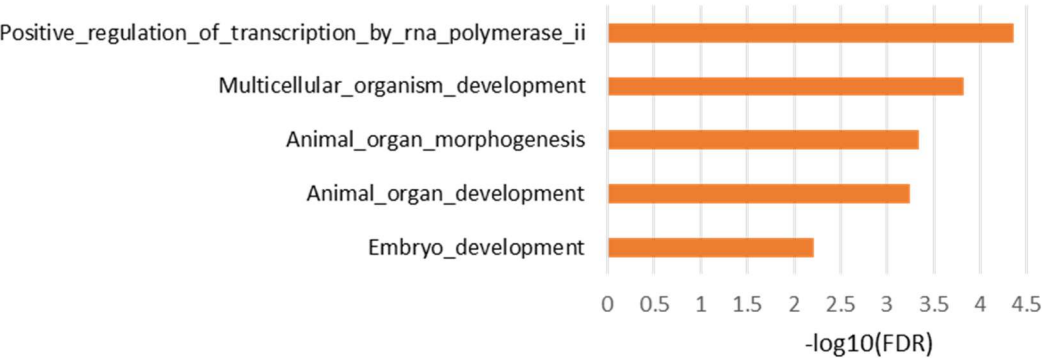


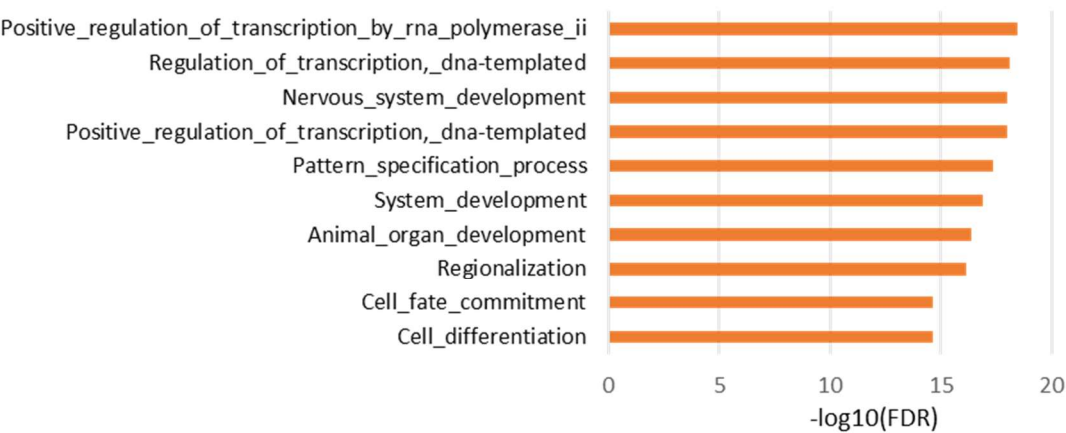
Figure 4.8B continued...

B. Day 7

Gene Cluster 5 GO: Biological Processes



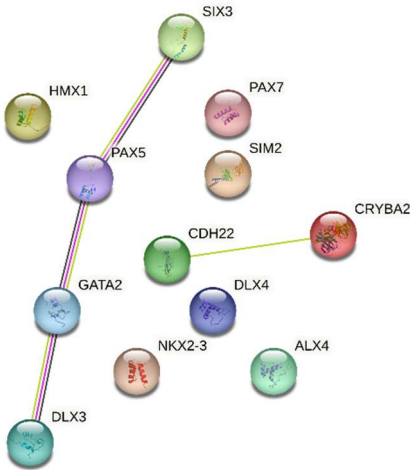
Gene Cluster 6 GO: Biological Processes



Gene Cluster 5

Gene	Rank by mean DS
PAX7	6
DLX4	36
HMX1	26
CDH22	15
DLX3	53
PAX5	14
CRYBA2	86
NKX2-3	68
GATA2	48
SIX3	11

PPI enrichment $P = 0.000272$



Gene Cluster 6

Gene	Rank by mean DS
TBX5	25
ZIC4	13
PRSS16	82
DLX6	21
MSX2	32
ONECUT1	24
NR2E1	33
PAX3	63
FOXD1	96
SCR1	55

PPI enrichment $P < 1.0^{-16}$

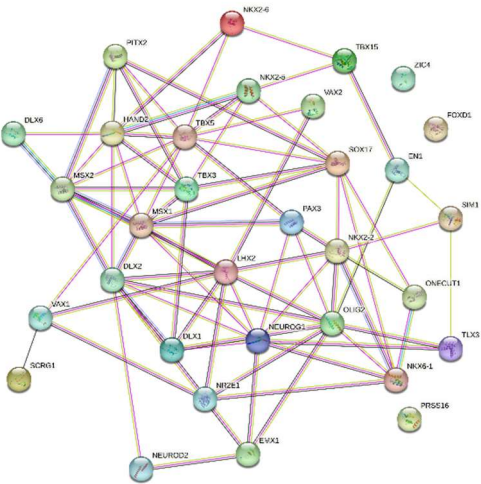


Figure 4.8 TRIAGE-Parser analysis of genes with the top 100 DS rank for EPLCs left to spontaneously differentiate to Day 5 and Day 7. D3 mESCs were cultured in suspension as EBs in MEDII out to Day 3 (EPLCs), with culture switched to Basal medium out to A. Day 5 and B. Day 7, as per Figure 4.1B. cDNA samples (n=4) were used to produce microarray data from which fold changes in gene expression were calculated relative to Day 3. The fold changes were TRIAGE transformed to produce discordance scores (DS) and ranked by mean DS across n=4. The top 100 ranked DS underwent TRIAGE-Parser to identify GO terms (FDR < 0.05), genes in the clusters (confidence threshold $\geq 95\%$) and predicted protein-protein interactions (PPIs) (STRING v 11.5) with default settings. PPI enrichment FDR *P* value < 0.001 indicates significance and, therefore, biologically meaningful PPIs. Only the top 10 significant GO terms are shown for simplicity.

4.3.4. *Relative expression and TRIAGE analysis of genes associated with PcG and TrxG proteins across pluripotency and neural lineage commitment*

L-Proline, one of the bioactive components in MEDII (Washington *et al.*, 2010), can act via mechanisms involving changes to the epigenetic landscape (Comes *et al.*, 2013; D’Aniello *et al.*, 2017; Harvey *et al.*, 2019; Patriarca *et al.*, 2021). Changes in the epigenetic landscape are achieved via mechanisms which include PcG and TrxG protein complex activity (Ying, Nichols, *et al.*, 2003; Ringrose and Paro, 2004; Boyer *et al.*, 2006; Ho *et al.*, 2009; MacArthur *et al.*, 2009; Hirai *et al.*, 2011; Lau and So, 2015; Brand *et al.*, 2019). As cell fate is, in part, determined by these activities, we used differential gene expression analysis to interrogate genes encoding for PcG and TrxG proteins across (i) the pluripotency spectrum, which includes ground and naïve states, EPLCs and primed EpiSCs, based on the RNAseq data sets (see Figure 3.2) and (ii) the EPLC, definitive ectoderm (Day 6) and neurectoderm (Day 9) stages of MEDII-mediated neural differentiation, based on the microarray data sets (see Figure 4.1A and 4.1B). Significance is denoted with a Benjamini-Hochberg adjusted $P < 0.05$ and a $\log_2$ fold change ≥ 1 or ≤ -1 to restrict analysis to genes with large changes in expression.

PcG proteins form polycomb-repressive complexes 1 (PRC1) and 2 (PRC2). PRC1 contains 4 essential subunits: RING1A/B, a posterior sex comb ortholog from the polycomb group RING finger protein family (PCGF1-6), a polyhomeotic homolog (PHC1-3) and a chromobox protein (Cbx2/4/6-8). PRC2 consists of 3 essential subunits: EZH1 or its homolog EZH2, an embryonic ectoderm development (EED) protein and a SUZ12 protein (Morey and Helin, 2010; Morey *et al.*, 2012; Di Croce and Helin, 2013; Aranda *et al.*, 2015; Santanach *et al.*, 2017).

Figure 4.9A shows changes in gene expression for PcG proteins across the pluripotency spectrum from ground-state mESCs to primed EpiSCs in the D3 and A2 cell lines based on the RNAseq data sets. Changes in gene expression are compared to naïve mESCs.

EPLCs demonstrate a very similar expression profile for these genes compared to naïve cells. In contrast, ground state cells demonstrate a distinct expression profile compared to naïve state. Upregulated expression of 4 PcG genes is observed; *Cbx4*, *Cbx8*, *Pcgfl* and *Bmi1*, which represent genes from two different core components of the PcG complex. EpiSCs also demonstrate a distinct expression profile compared to naïve state. 4 PcG genes are upregulated in EpiSCs (*Cbx4*, *Cbx8*, *Phc2* and *Bmi1*), and 2 (*Phc1* and *Eed*) are downregulated.

Figure 4.9B shows changes in gene expression for PcG proteins differentiated from EPLCs (Day 3) out to definitive ectoderm and neurectoderm with MEDII (Day 6 and 9) versus spontaneous differentiation with Basal medium (out to Day 5 and 7).

At Day 9 in MEDII, the expression of *Cbx7* and *Phc1* is downregulated, while that for *Phc2* and *Bim1* is upregulated compared to Day 3 (EPLCs). No change is seen for the spontaneously differentiated cells, except that expression of *Phc1* is downregulated.

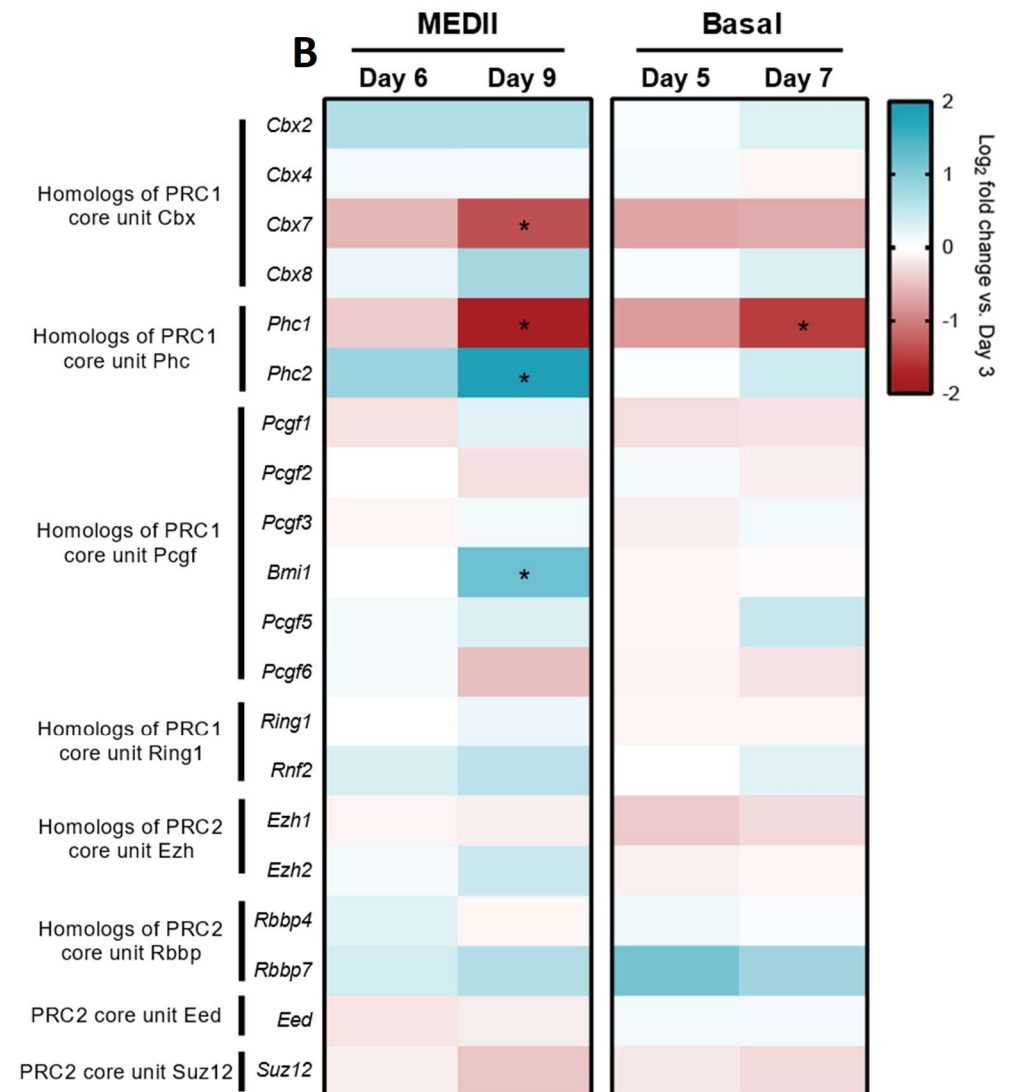
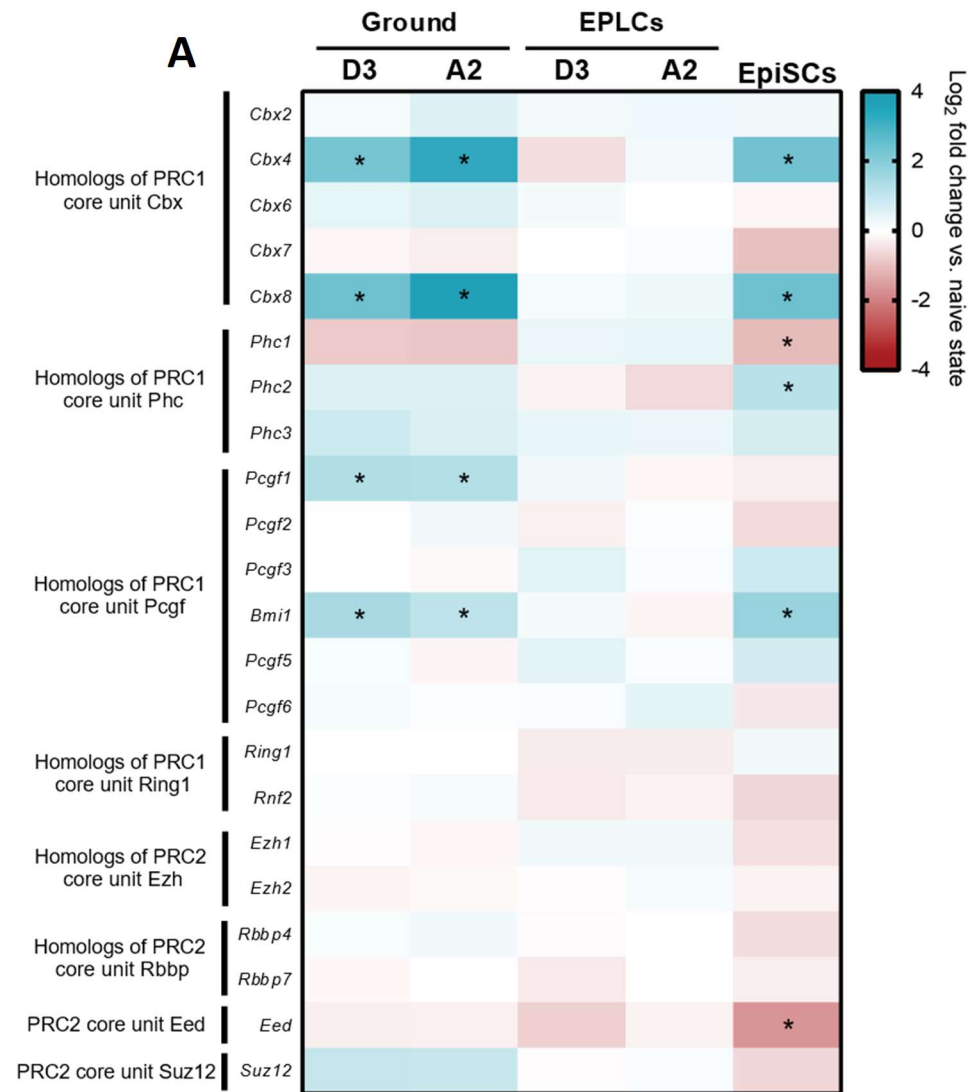


Figure 4.9 Changes in PcG core component gene expression (A) across the pluripotency spectrum and (B) in MEDII-directed neural differentiation and spontaneously differentiated cells. Changes in expression of genes of the homologs that encode units of polycomb repressive complexes 1 (PRC1) and 2 (PRC2) are shown relative to the expression of **A.** naïve cells and **B.** Day 3 (EPLCs). **A.** D3 and A2 mESCs were used to produce naïve-state mESCs, ground-state mESCs and EPLCs, as per Figure 3.2. A2 primed EpiSCs were maintained, as per Figure 3.2. Samples (n=3) were subjected to RNAseq from which relative expressions were calculated. **B.** D3 mESCs were subjected to MEDII-mediated differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to Day 6 and Day 9, or allowed to spontaneously differentiate in Basal medium to Day 5 and Day 7. cDNA samples (n=4) were used to produce microarray data, as per Figure 4.1. * indicates a Benjamini-Hochberg adjusted $P < 0.05$ and $\log_2$ fold change ≥ 1 or ≤ -1 .

TrxG proteins include a SET protein (SET1A/B, MLL1/2 or MLL3/4), which form a complex associated with trithorax-related SET protein (COMPASS) each (Lau and So, 2015). All COMPASS's share common core subunits that include WD repeat-containing protein 5 (WDR5), retinoblastoma binding protein 5 (RBBP5), absent, small, or homeotic discs 2-like protein (ASH2L) and Dumpy-30 (Dpy-30) (Dou *et al.*, 2006; Southall *et al.*, 2009; Ang *et al.*, 2011; Jiang *et al.*, 2011; Piunti and Shilatifard, 2016). RBBP5 and ASH2L stimulates histone methyltransferase activity, WDR5 is required to maintain complex integrity and substrate binding, while Dpy-30 is responsible for regulating methylation of H3K4 (Dou *et al.*, 2006; Southall *et al.*, 2009; Ernst and Vakoc, 2012). COMPASS complexes also contain specific subunits, such as Host cell factor 1 (HCF1) and nuclear receptor coactivator 6 (NCOA6) that confer different functions due to their roles in regulating recruitment to different target genes and interactions with other proteins (Wu *et al.*, 2008; Herz *et al.*, 2013; Lau and So, 2015).

Figure 4.10A shows changes in gene expression for TrxG proteins across the pluripotency spectrum from ground-state mESCs to primed EpiSCs in the D3 and A2 cell lines based on the RNAseq data sets. Changes in gene expression are compared to naïve mESCs.

EPLCs demonstrate a similar expression profile for these genes compared to naïve cells. Ground state also demonstrate a similar expression profile compared to naïve state, except for downregulated *Setd1b* (A2 only) and upregulated *Ncoa6*. EpiSCs demonstrate downregulation of some COMPASS subunits.

Figure 4.10B shows changes in gene expression for TrxG proteins differentiated from EPLCs (Day 3) out to definitive ectoderm and neurectoderm with MEDII (Day 6 and 9) versus spontaneous differentiation with Basal medium (out to Day 5 and 7). No change is seen for both conditions.

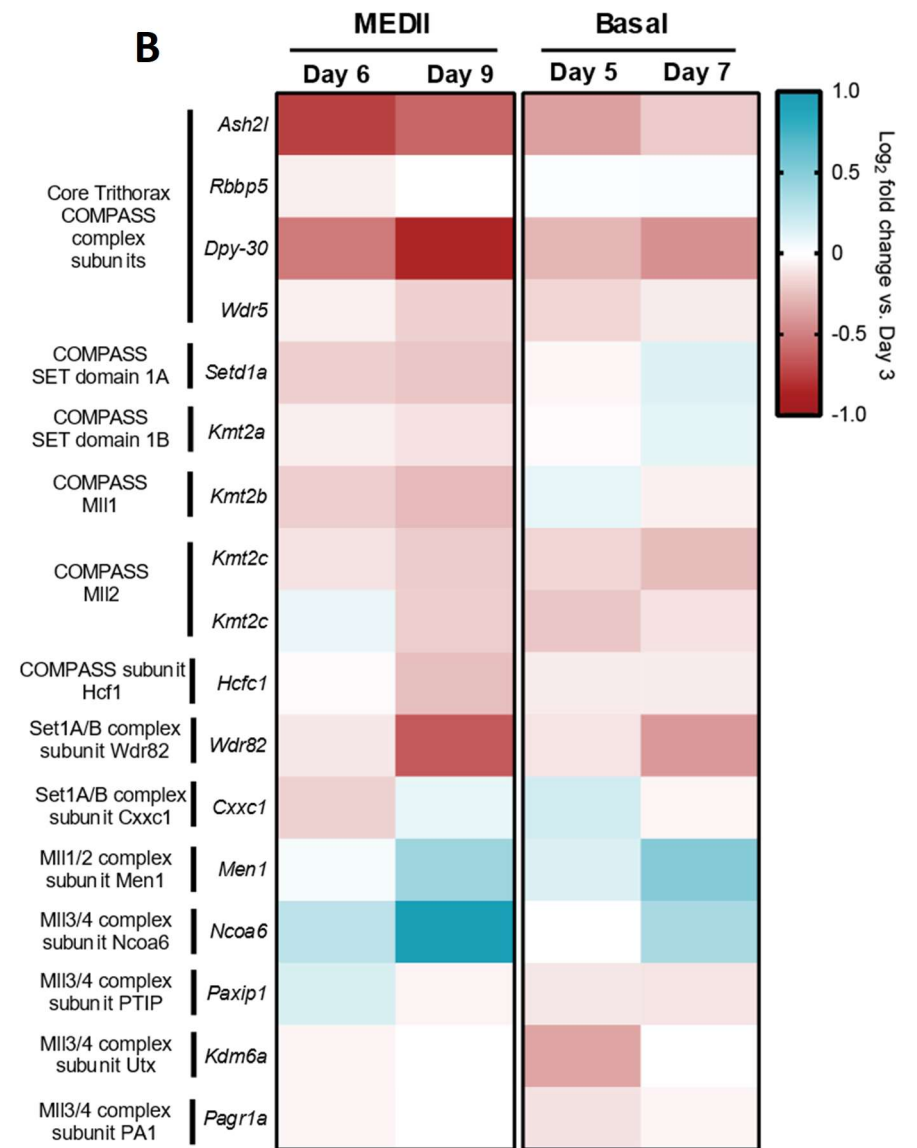
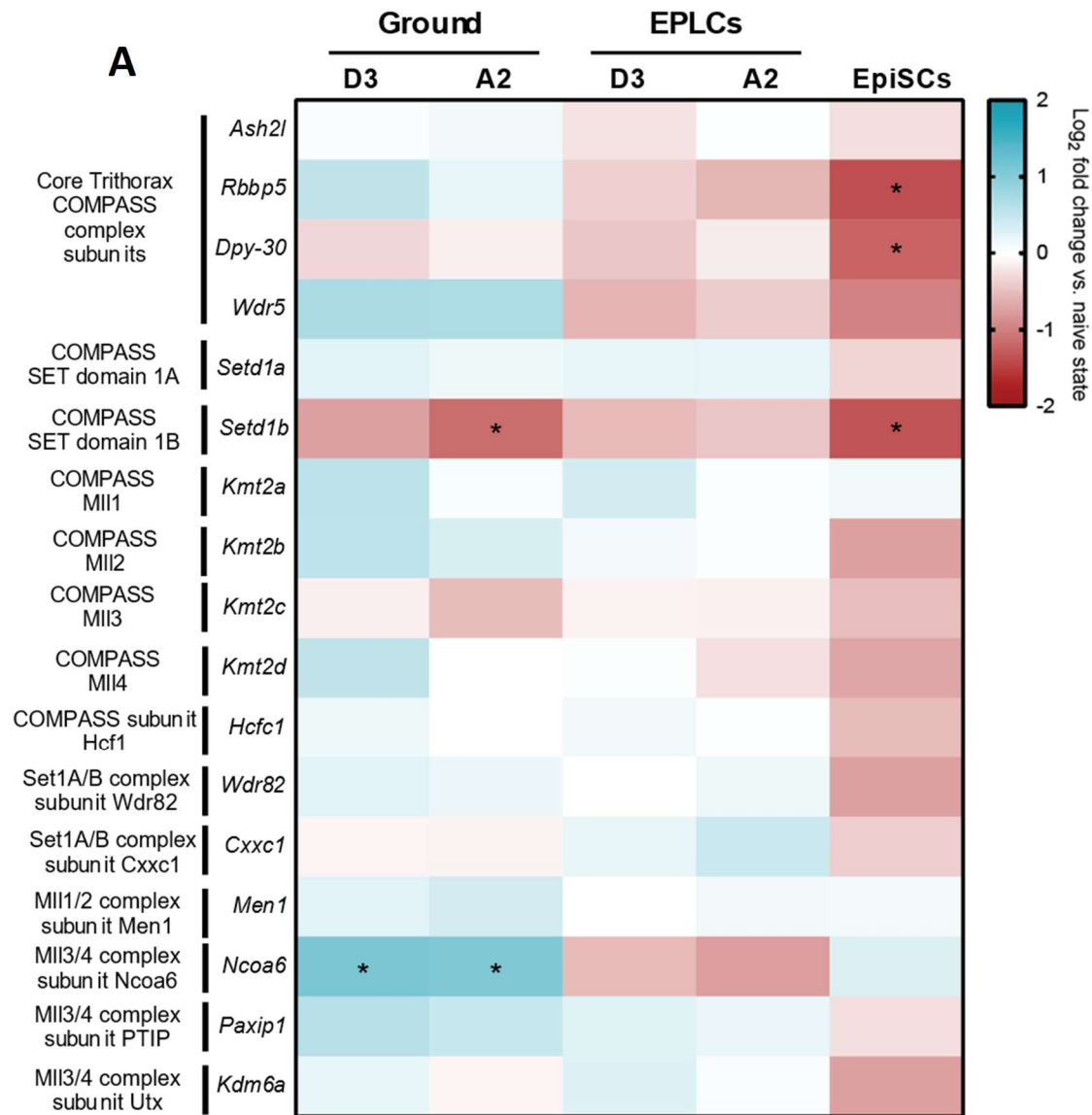


Figure 4.10 Changes in TrxG gene expression (A) across the pluripotency spectrum and (B) in MEDII-directed neural differentiation and spontaneously differentiated cells. Changes in expression of genes of the homologs that encode units of TrxG proteins that make COMPASS complexes are shown relative to the expression of **A.** naïve cells and **B.** Day 3 (EPLCs). **A.** D3 and A2 mESCs were used to produce naïve-state mESCs, ground-state mESCs and EPLCs, as per Figure 3.2. A2 primed EpiSCs were maintained, as per Figure 3.2. Samples (n=3) were subjected to RNAseq from which relative expressions were calculated. **B.** D3 mESCs were subjected to MEDII-mediated differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to Day 6 and Day 9, or allowed to spontaneously differentiate in Basal medium to Day 5 and Day 7. cDNA samples (n=4) were used to produce microarray data, as per Figure 4.1. * indicates a Benjamini-Hochberg adjusted $P < 0.05$ and log2 fold change ≥ 1 or ≤ -1 .

4.3.5 TRIAGE analysis of genes associated with PcG and TrxG proteins across the pluripotency continuum and neural lineage commitment

We then used TRIAGE to transform raw counts from the RNAseq dataset to DS. DS were ranked in descending order across the entire mouse genome; i.e., highest DS is ranked first. A gene with a DS rank ≤ 1359 is likely associated with cell fate and cell identity (Shim *et al.*, 2020). Beyond this rank a gene is likely associated with housekeeping and structural maintenance (Shim *et al.*, 2020).

Figure 4.11 shows DS highlighting PcG genes graphed in this manner for each of the 4 stages of pluripotency in order; i.e., ground state, naïve state, EPLCs, and primed EpiSCs. Figure 4.12 shows the same graphs as 4.11 except highlighting DS of TrxG genes. PcG and TrxG genes in red have a DS rank ≤ 1359 .

Several PcG genes (*Cbx2*, *Rbbp7*, *Ring1*, *Phc2* and *Bmi1*) have a high DS rank and appear at all stages of pluripotency but for TrxG genes only one, *Hcfc1*, representing a COMPASS subunit, has a high DS rank, and it also appears at all stages of the pluripotency continuum, except for EpiSCs.

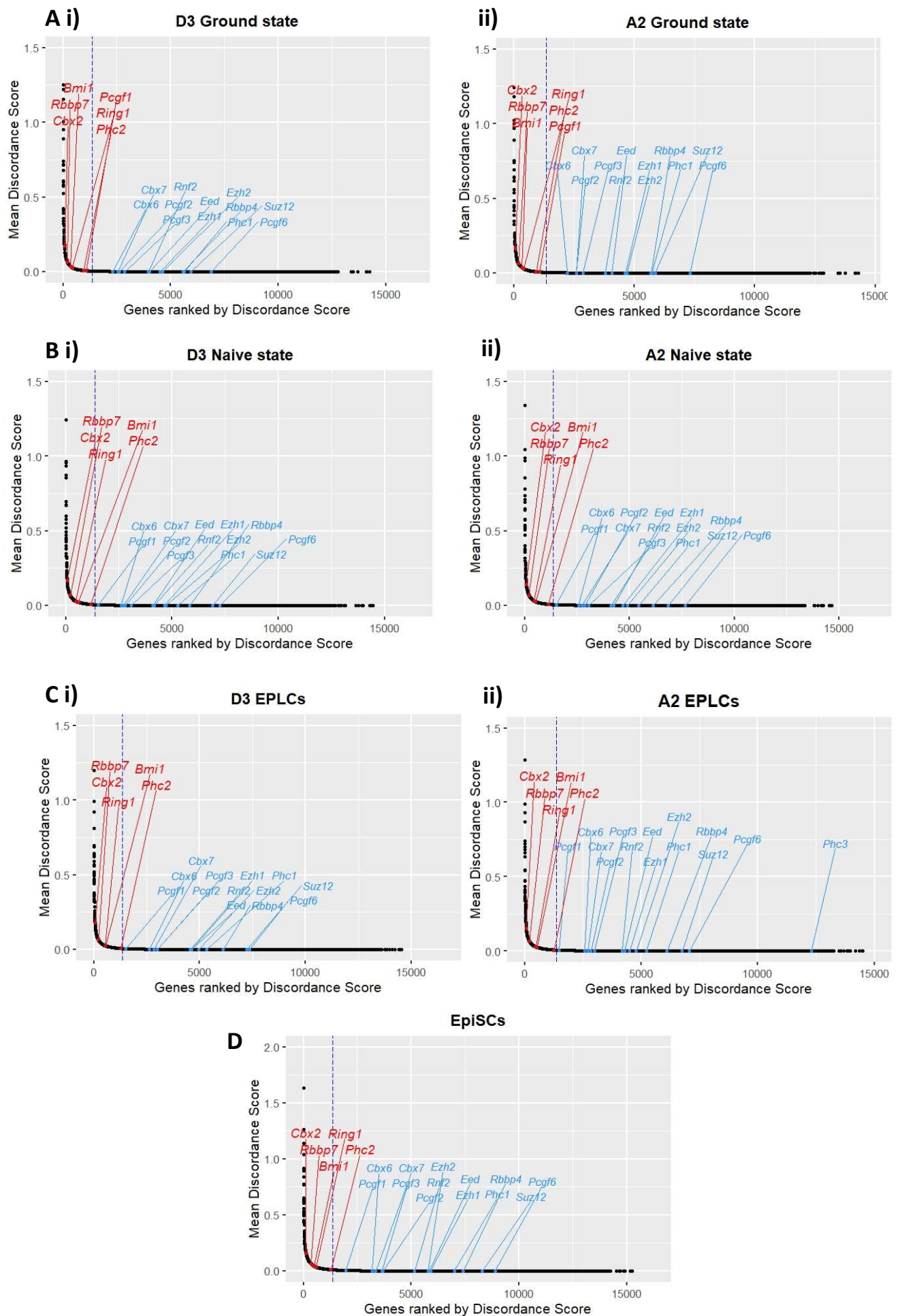


Figure 4.11 TRIAGE-transformed PcG genes across the pluripotency spectrum. **i)** D3 and **ii)** A2 mESCs were used to produce **A.** ground-state mESCs, **B.** naïve-state mESCs, and **C.** EPLCs. **D.** shows primed A2 EpiSCs. Samples (n=3) were subjected to RNAseq and raw counts were acquired. Raw counts were TRIAGE transformed to produce discordance scores (DS) and these DS are plotted against DS rank. The blue dashed line indicates the ranked DS threshold, where DS rank ≤ 1359 . PcG genes in red have a DS rank ≤ 1359 . PcG genes in blue indicate DS rank > 1359 .

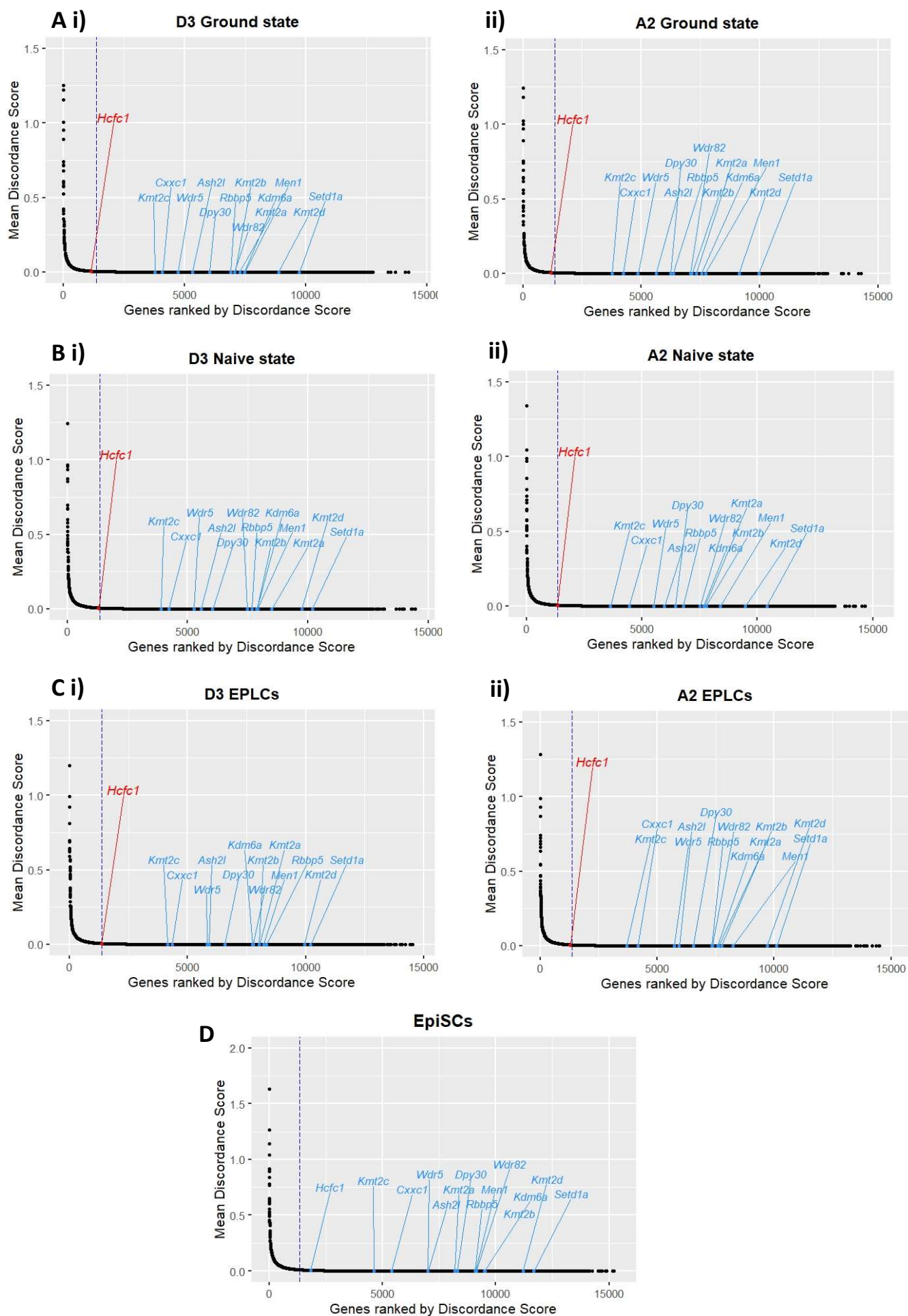


Figure 4.12 TRIAGE-transformed TrxG genes across the pluripotency spectrum. **i)** D3 and **ii)** A2 mESCs were used to produce **A.** ground-state mESCs, **B.** naïve-state mESCs, and **C.** EPLCs. **D.** shows primed A2 EpiSCs. Samples (n=3) were subjected to RNAseq and raw counts were acquired. Raw counts were TRIAGE transformed to produce discordance scores (DS) and these DS are plotted against DS rank. The blue dashed line indicates the ranked DS threshold, where DS rank ≤ 1359 . TrxG genes in red have a DS rank ≤ 1359 . TrxG genes in blue indicate DS rank > 1359 .

In terms of neural lineage commitment, TRIAGE was applied to transform fold changes relative to Day 3 (EPLCs). Figure 4.13 and 4.14 show the DS plotted against DS rank highlighting PcG and TrxG genes, respectively.

For MEDII-treated cells, several PcG genes have DS ranks ≤ 1359 but only one, *Phc2*, does not appear in the highly ranked PcG genes in spontaneously differentiating cells (Figure 4.13). No TrxG genes have a high DS (Figure 4.14).

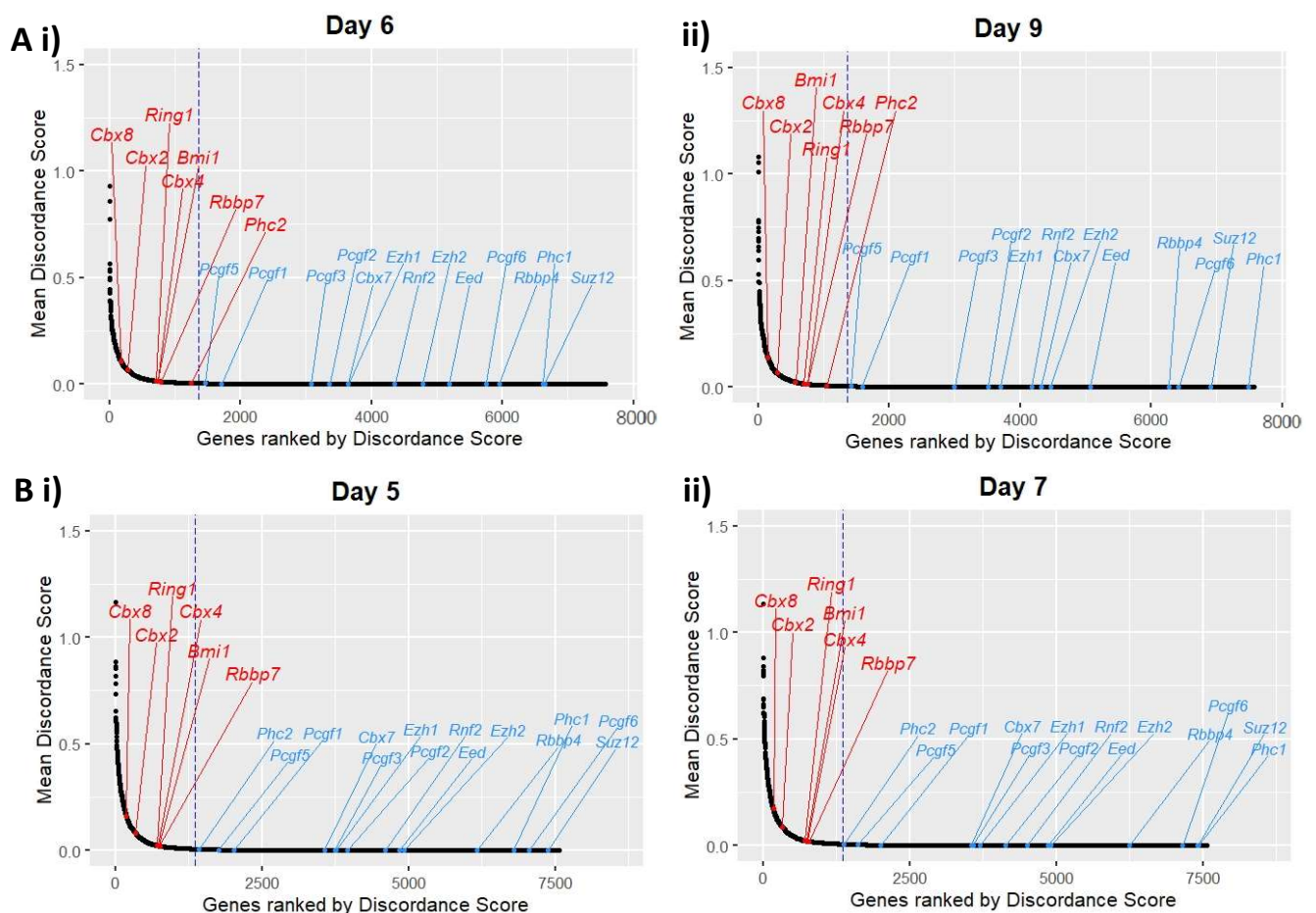


Figure 4.13 TRIAGE-transformed PcG genes across intermediate stages of neural fate commitment. D3 mESCs were cultured in suspension as EBs in MEDII medium (MEDII) out to Day 3. These EBs were continued to be cultured in **A.** MEDII out to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm) or switched (with a different set of Day 3 EBs) to **B.** Basal medium to spontaneously differentiate out to **i)** Day 5 and **ii)** Day 7 (as per Figure 4.1A and 4.1B). cDNA samples ($n=4$) were used to produce microarray data. Raw data was extracted on Rstudio, then fold changes relative to Day 3 were TRIAGE transformed to produce discordance scores (DS). DS of PcG genes are plotted against DS rank ($n=4$). Red genes indicate a DS rank ≤ 1359 , while blue genes indicate a DS rank > 1359 .

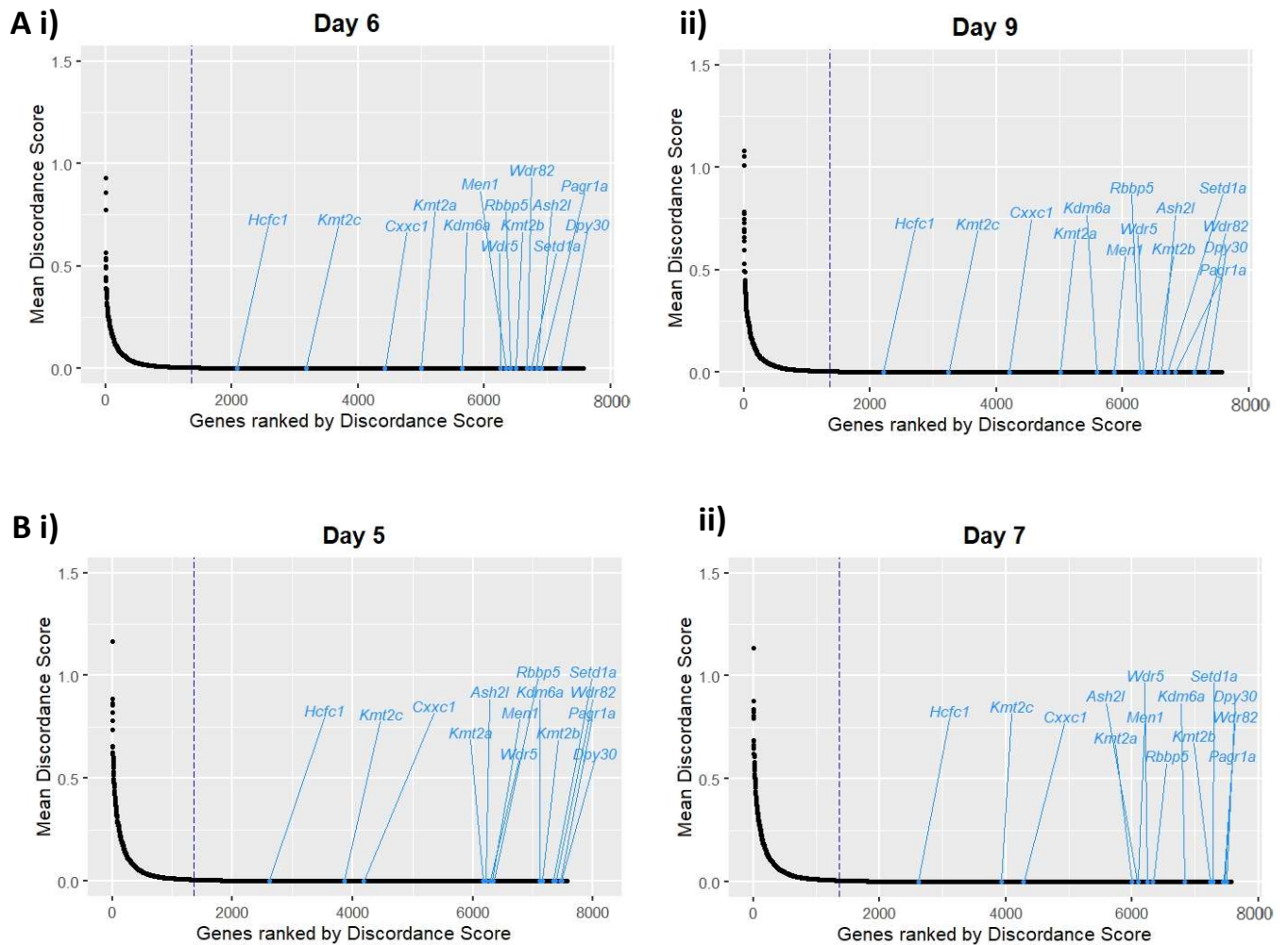


Figure 4.14 TRIAGE-transformed TrxG genes across intermediate stages of neural fate commitment. D3 mESCs were cultured in suspension as EBs in MEDII medium (MEDII) out to Day 3. These EBs were continued to be cultured in **A**. MEDII out to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm) or switched (with a different set of Day 3 EBs) to **B**. Basal medium to spontaneously differentiate out to **i)** Day 5 and **ii)** Day 7 (as per Figure 4.1A and 4.1B). cDNA samples (n=4) were used to produce microarray data. Raw data was extracted on Rstudio, then fold changes relative to Day 3 were TRIAGE transformed to produce discordance scores (DS). DS of TrxG genes are plotted against DS rank (n=4). Blue genes indicate a DS rank >1359.

4.3.5. Relative expression and TRIAGE analysis of genes associated with L-Proline metabolism across neural lineage commitment

Exogenously added L-Proline can replace MEDII to drive differentiation of mESCs to neural cell types (Shparberg, 2018; Shparberg *et al.*, 2019a).

A new paradigm in development is that metabolism also contributes to cell fate and L-Proline metabolised through the proline cycle can drive changes in cell identity (Phang, Pandhare, *et al.*, 2008; Phang *et al.*, 2010; Casalino *et al.*, 2011; Kilberg *et al.*, 2016; Phang, 2019; Patriarca *et al.*, 2021).

With this rationale, we used differential gene expression analysis and TRIAGE transformation of the MEDII microarray data to examine if various metabolic pathways, including those associated with the metabolism of L-Proline metabolism, may contribute to neural cell fate. To account for FDRs, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of $\log_2$ fold change (Koch *et al.*, 2018).

Figure 4.15A shows the pathway of uptake of L-Proline into mESCs and subsequent catabolism to various products. Figure 4.15B shows the relative changes in expression compared to Day 3 (EPLCs) of the genes in this pathway for both MEDII-directed and spontaneous differentiation. MEDII-treated cells at Day 9 (neurectoderm) show downregulation of the expression of the L-Proline transporter, perhaps indicating a reduced need to take up extracellular L-Proline at this stage of development.

Figure 4.15C shows that none of the genes in this pathway have a high DS rank, although due to limitations of the TRIAGE pipeline, *Prodh*, the gene encoding for rate-limiting enzyme POX, could not be TRIAGE transformed, and thus was not ranked (data not shown).

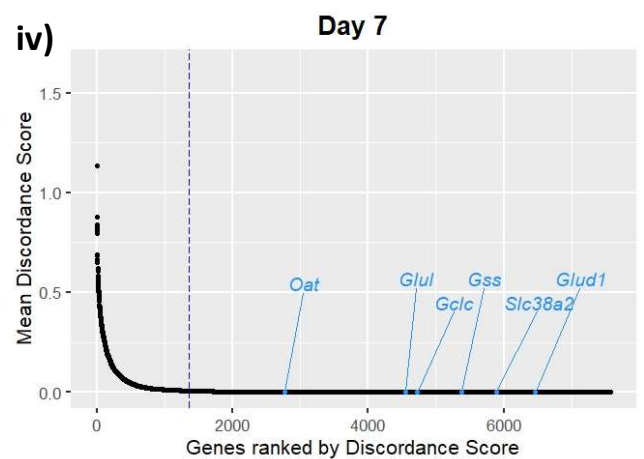
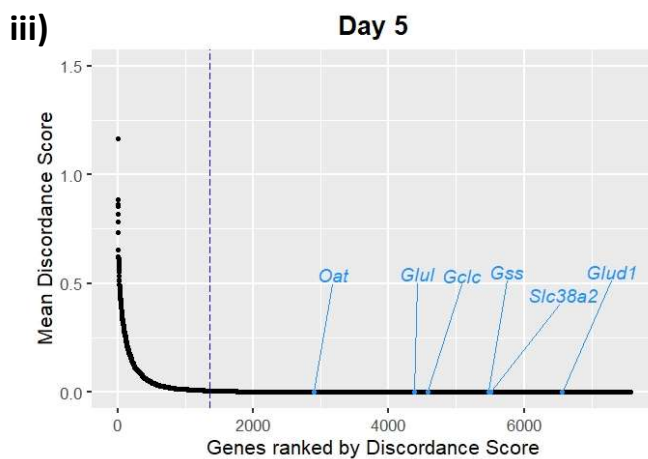
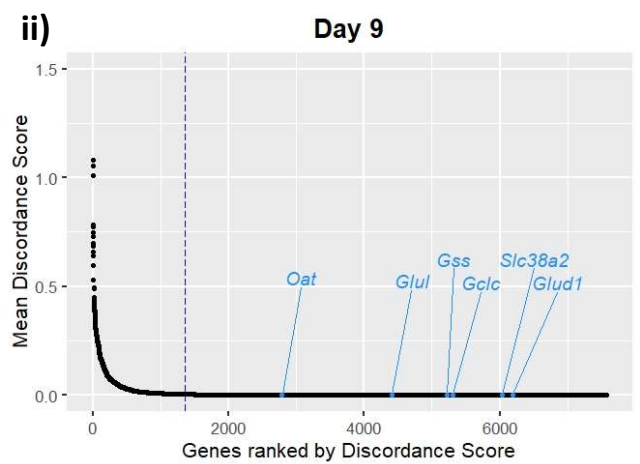
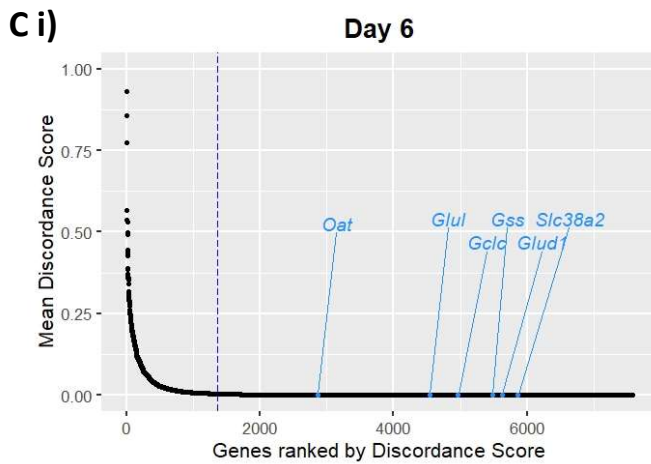
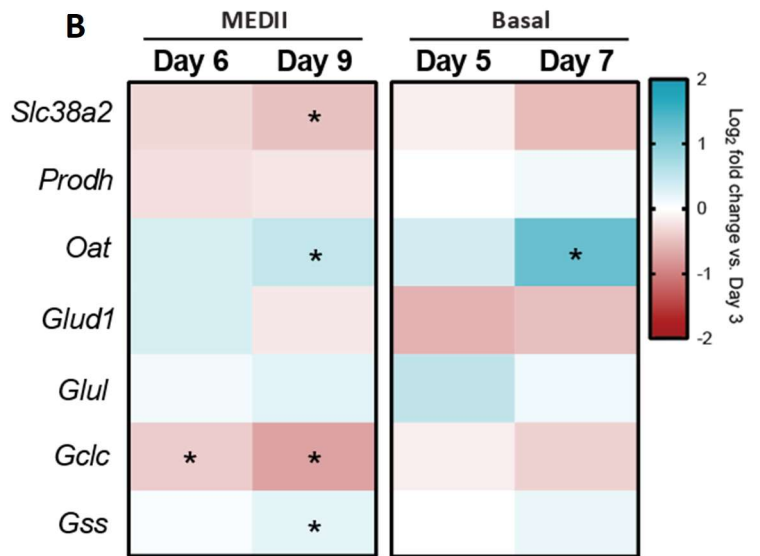
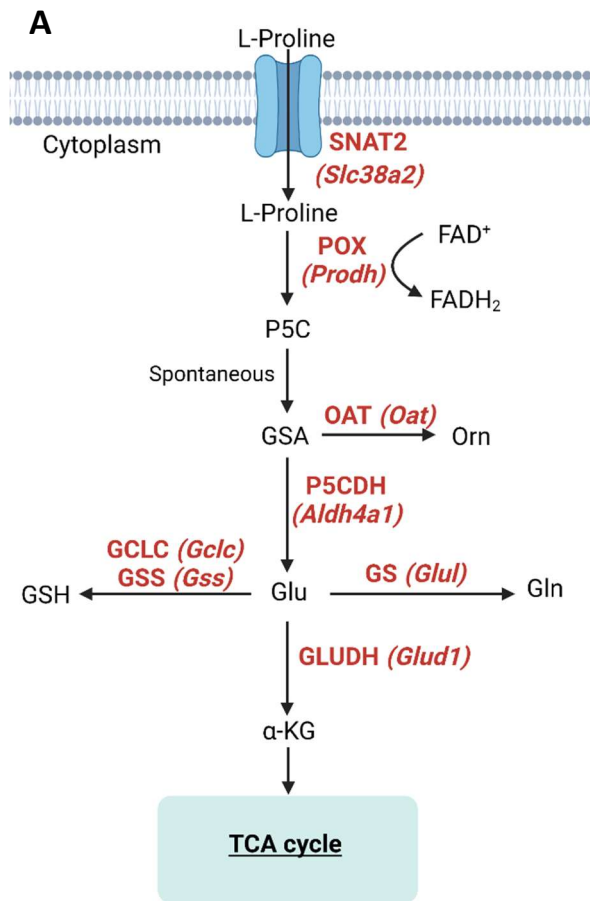


Figure 4.15 Relative expression and TRIAGE transformation of proline catabolism genes across neural lineage commitment. **A.** L-Proline enters the cell via the SNAT2 transporter and is metabolised to various products. Enzymes are shown in red and the corresponding genes in italics in parentheses. **B-C.** D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm), or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to **iii)** Day 5 and **iv)** Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1. Microarray data underwent **B.** differential gene analysis with changes in gene expression made relative to Day 3 (EPLCs) or **C.** TRIAGE analysis, with the discordance scores (DS) plotted against DS rank. Blue genes indicate a DS rank >1359. * Indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: SNAT2: Sodium-coupled neutral amino acid transporter 2; *Slc38a2*: Solute Carrier Family 38 Member 2; POX: proline oxidase; *Prodh*: proline dehydrogenase; OAT: Ornithine aminotransferase; P5CDH: pyrroline-5-carboxylate dehydrogenase; *Aldh4a1*: Aldehyde Dehydrogenase 4 Family, Member A1; *Gclc*: Glutamate-cysteine ligase catalytic subunit; *Gss*: Glutathione synthetase; GS: Glutamine synthase; *Glul*: Glutamate-ammonia ligase; *Glud1* (GLUDH): Glutamate dehydrogenase 1; Glu: L-Glutamate; Gln: L-Glutamine; GSH: Glutathione.

Figure 4.16A shows the process of L-Proline biosynthesis from other amino acid sources. Figure 4.16B shows the relative changes in expression compared to Day 3 (EPLCs) of the genes in this pathway for both MEDII-directed and spontaneous differentiation. MEDII-treated cells at Day 9 (neurectoderm) show upregulation of L-Proline biosynthesis genes, perhaps indicating an increased production of endogenous L-Proline at this stage of development.

Figure 4.16C shows that none of the genes in this pathway have a high DS rank.

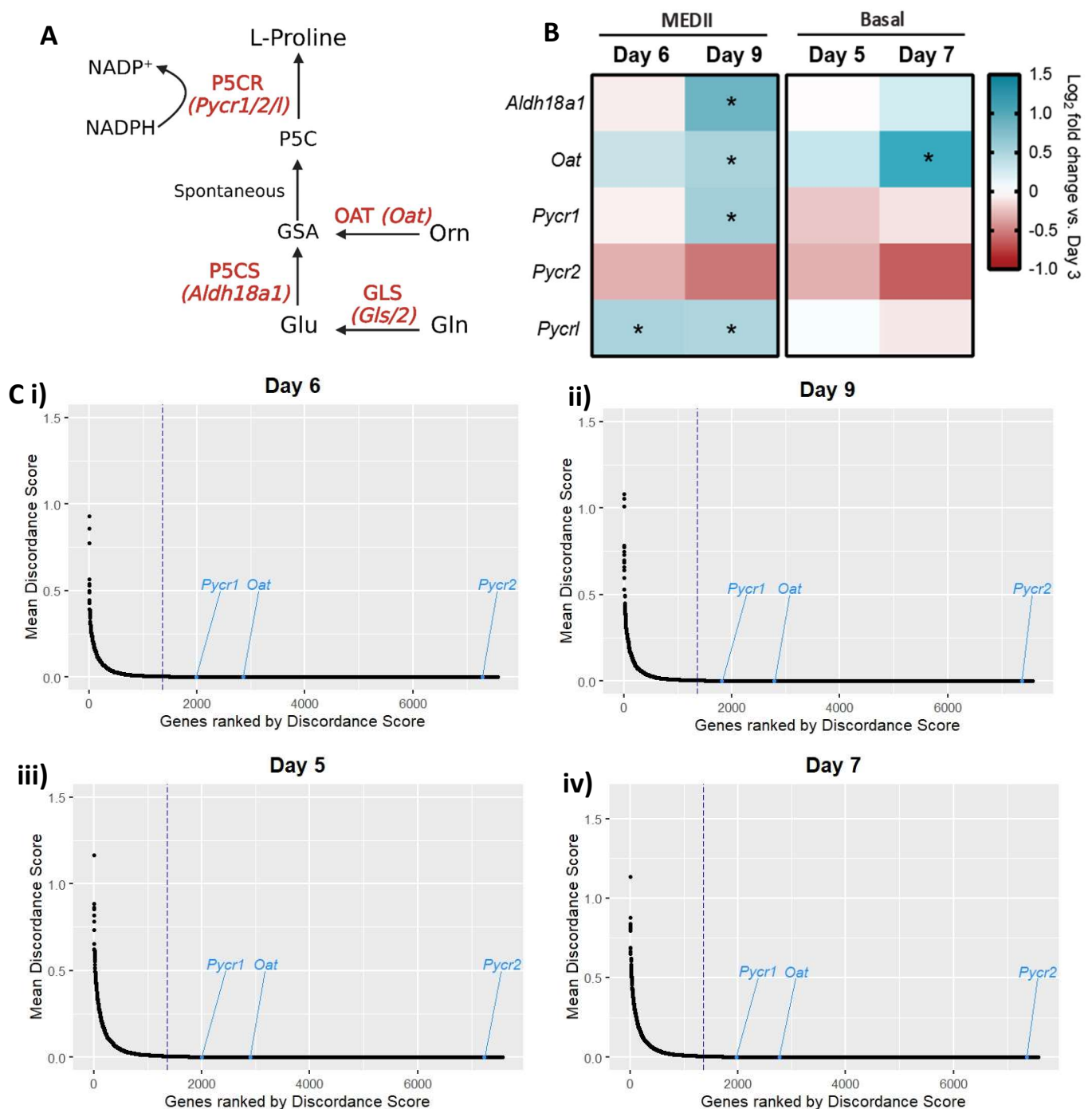


Figure 4.16 Relative expression and TRIAGE transformation of proline biosynthesis genes across neural lineage commitment. **A.** Some amino acid sources of L-Proline can be produced from either L-Glutamine (Gln) or L-Ornithine (Orn). Enzymes are shown in red and the corresponding genes in italics in parentheses. **B-C.** D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm), or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to **iii)** Day 5 and **iv)** Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1A and 4.1B. Microarray data underwent **B.** differential gene analysis with changes in gene expression made relative to Day 3 (EPLCs) or **C.** TRIAGE analysis, with the discordance scores (DS) plotted against DS rank. Blue genes indicate a DS rank >1359. * Indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: *Aldh18a1*: Aldehyde dehydrogenase 18 family, member A1; *Oat*: ornithine aminotransferase, mitochondrial; *Pycr1/2/l*: Pyrroline-5-carboxylate reductase 1/2/-like.

4.3.6. Relative expression and TRIAGE analysis of genes associated with glycolysis across neural lineage commitment

The metabolism of L-Proline also mediates other metabolic processes, which include some processes involved in glycolysis (also see Figure 3.1). With this rationale, we used differential gene expression analysis and TRIAGE transformation of the MEDII microarray data to examine if glycolysis may contribute to neural cell fate (Figure 4.18). We included monocarboxylate transporter genes *Slc16a1/7/8* in our analysis to determine the extent of expression of other monocarboxylate transporters beyond the lactate transporter gene *Slc16a3* (Halestrap, 2013). We are also interested in the *Hif1 α* gene because the transcription factor in ESCs promotes transcription of glycolytic genes such as glucose transporter genes *Glut1* and *Glut2*, and pyruvate dehydrogenase kinases *Pdk1-4* (Ke and Costa, 2006; Kim *et al.*, 2006; Forristal *et al.*, 2010; Prigione *et al.*, 2014; Fiorenza and Rava, 2019). To account for FDRs, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of log₂ fold change (Koch *et al.*, 2018).

Figure 4.17A shows processes involved in glycolysis. Figure 4.17B shows the relative changes in expression compared to Day 3 (EPLCs) of the genes in this pathway for both MEDII-directed and spontaneous differentiation. Expression of glycolysis genes are upregulated in the differentiation to neurectoderm.

Figure 4.17C shows none of the genes in this pathway have a high DS rank.

Figure 4.17 continued...

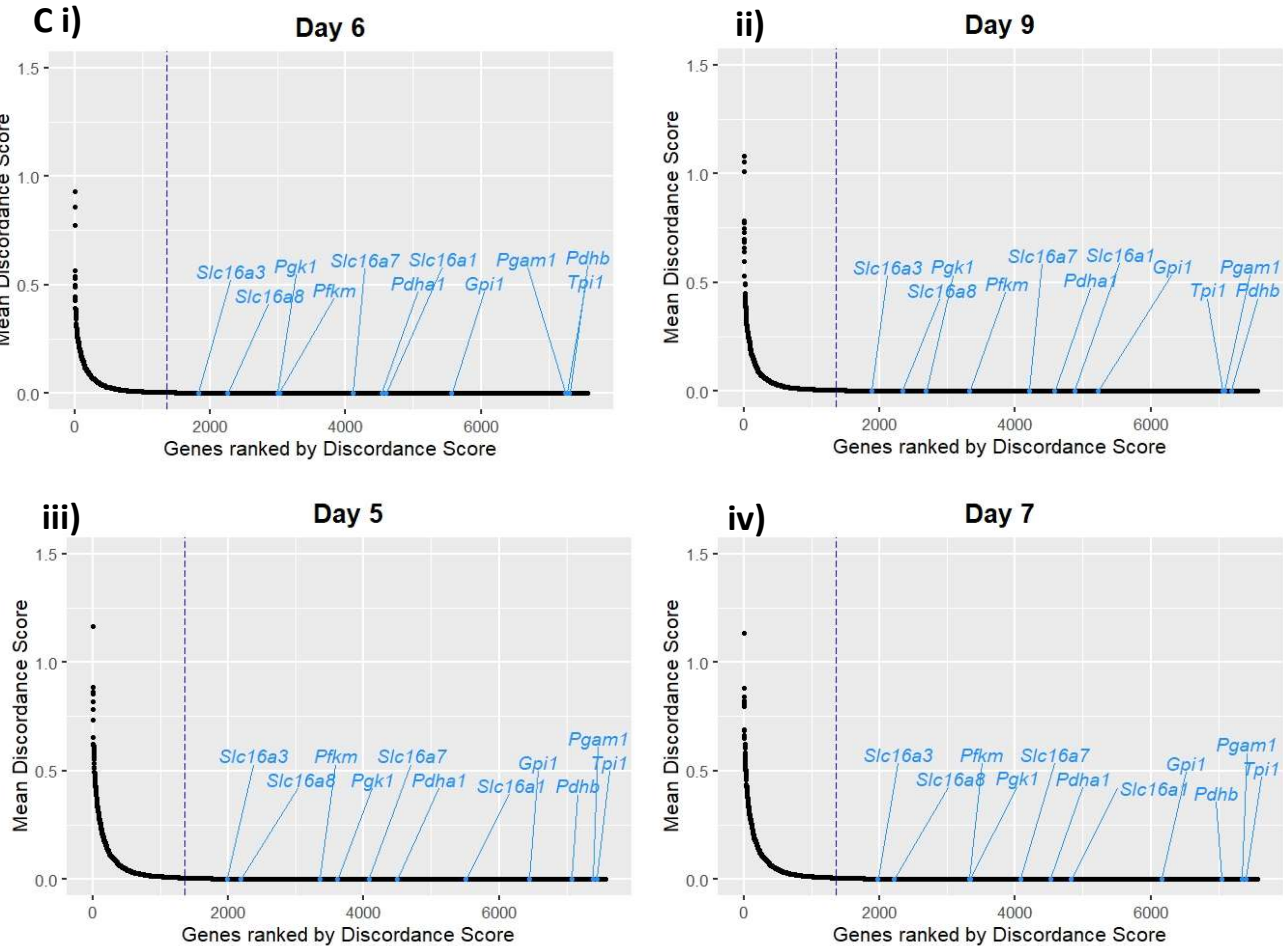


Figure 4.17 Relative expression and TRIAGE transformation of glycolysis genes across neural lineage commitment. A. The glycolysis pathway and metabolite intermediates that converts glucose into two pyruvate molecules. Enzymes are shown in red and the corresponding genes in italics in parentheses **B-C.** D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm), or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to **iii)** Day 5 and **iv)** Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1A and 4.1B. Microarray data underwent **B.** differential gene analysis with changes in gene expression made relative to Day 3 (EPLCs) or **C.** TRIAGE analysis, with the discordance scores (DS) plotted against DS rank. Blue genes indicate a DS rank >1359. * Indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$. *Hk1*: hexokinase 1.

4.3.7. Relative expression and TRIAGE analysis of genes associated with pentose phosphate pathway across neural lineage commitment

Metabolism of L-Proline is also redox coupled to the oxidative phase of the pentose phosphate pathway (PPP) (see also Figure 3.1) (Phang *et al.*, 2010; Hancock *et al.*, 2016). Glyceraldehyde-3-phosphate and fructose-6-phosphate are the metabolites that are produced from the subsequent non-oxidative phase of the PPP which contribute towards glycolysis (Figure 3.1 and 4.18A) (Ge *et al.*, 2020).

With this rationale, we used differential gene expression analysis and TRIAGE transformation of the MEDII microarray data to examine if the oxidative and/or non-oxidative phases of the PPP contribute to neural cell fate. To account for FDRs, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of $\log_2$ fold change (Koch *et al.*, 2018).

Figure 4.18A shows processes involved in the PPP (oxidative and non-oxidative phases). Figure 4.18B shows the relative changes in expression compared to Day 3 (EPLCs) of these genes for both MEDII-directed and spontaneous differentiation. Consistent downregulation of rate-limiting enzyme *G6pdx* during differentiation to neurectoderm is shown, perhaps indicating reduced PPP activity during this process.

Figure 4.18C shows none of the genes in this pathway have a high DS rank, except for *Tktl1*.

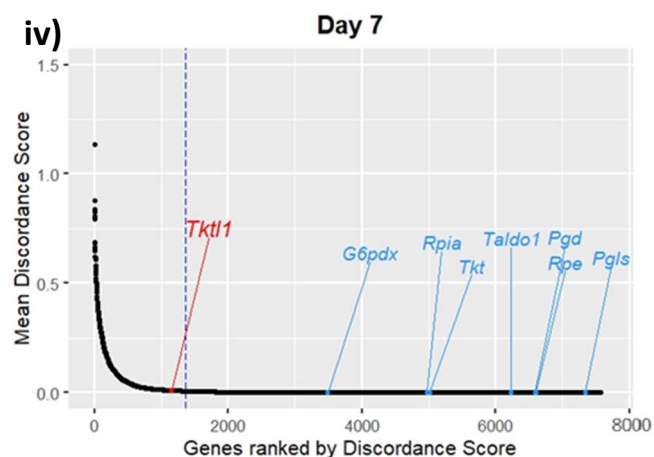
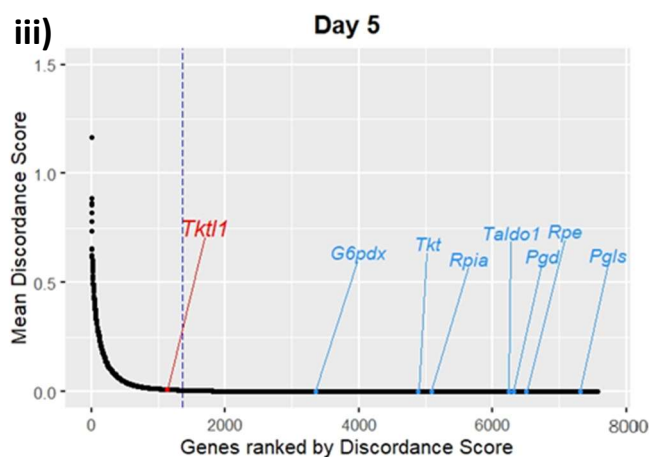
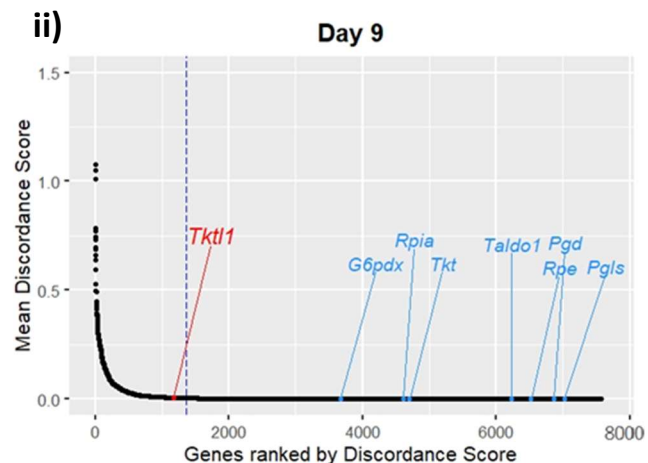
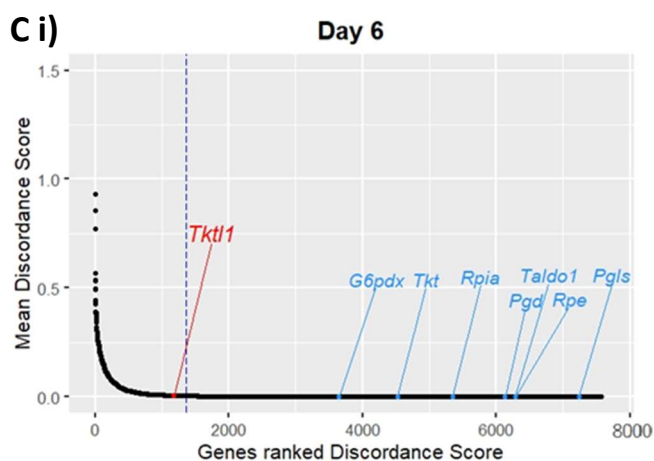
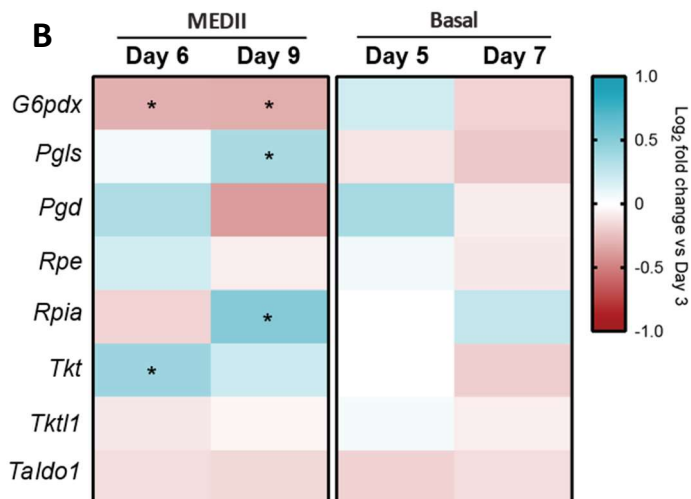
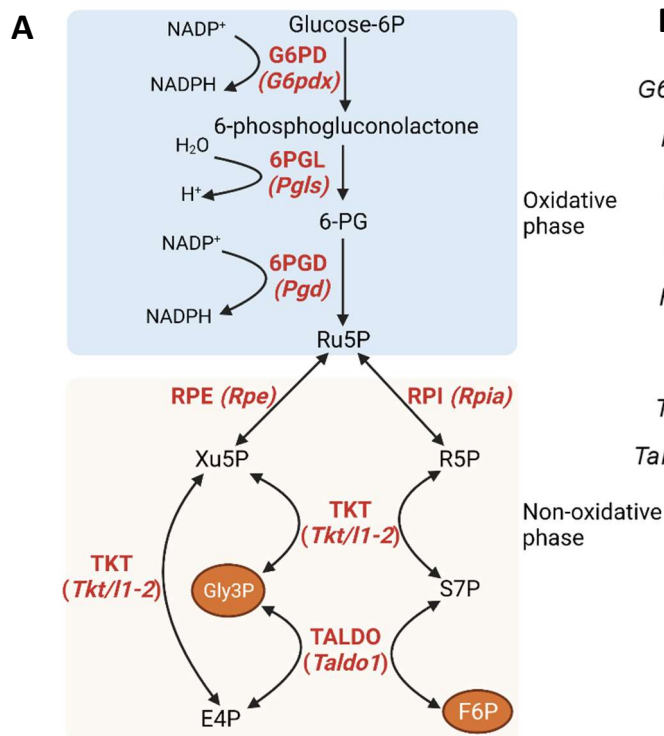


Figure 4.18 Relative expression and TRIAGE transformation of PPP genes across neural lineage commitment. **A.** The PPP (oxidative and non-oxidative phases) and its metabolite intermediates. Enzymes are shown in red and the corresponding genes in italics in parentheses. **B-C.** D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to **i)** Day 6 and **ii)** Day 9, or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to **iii)** Day 5 and **iv)** Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1A and 4.1B. Microarray data underwent **B.** differential gene analysis or **C.** TRIAGE analysis. PPP genes were ranked by fold changes relative to Day 3 (Expression) and mean DS across n=4 were plotted. The *Tktl1* gene in red indicates a DS rank ≤ 1359 . Blue genes indicate a DS rank > 1359 . * Indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: Glucose-6P: glucose-6-phosphate; 6-PG: 6-phosphogluconate; Ru5P: ribulose-5-phosphate; Xu5P: xylulose-5-phosphate; Gly3P: glyceraldehyde 3-phosphate; E4P: erythrose-4-phosphate; R5P: ribose-5-phosphate; S7P: sedoheptulose-7-phosphate; F6P: fructose-6-phosphate; G6PD (*G6pdx*): glucose-6-phosphate dehydrogenase; 6PGL (*Pgls*): 6-phosphogluconolactonase; 6PGD (*Pgd*): 6-phosphogluconate dehydrogenase; RPE: ribulose phosphate 3-epimerase/ phosphopentose epimerase; RPI (*Rpia*): ribulose phosphate 3-epimerase/ phosphopentose epimerase; TKT: transketolase; *Tktl1-2*: transketolase-like 1/2; TALDO (*Taldo1*): transaldolase.

4.3.8. *Relative expression and TRIAGE analysis of genes associated with the TCA cycle across neural lineage commitment*

As pluripotent cells commit to a particular lineage, preference of metabolic processes changes from anaerobic metabolism to aerobic metabolism to keep up with higher energy demands for specialised functions (Rehman, 2010; Varum *et al.*, 2011; Agathocleous *et al.*, 2012; Panopoulos *et al.*, 2012; Knobloch and Jessberger, 2017; Arrázola *et al.*, 2019; Tsogtbaatar *et al.*, 2020). The proline cycle also has the capacity to mediate the TCA cycle through production of pyruvate (see also Figure 1.4 and 3.1).

With this rationale, we used differential gene expression analysis and TRIAGE transformation of the MEDII microarray data to examine if the TCA cycle contributes to neural cell fate. To account for FDRs, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of $\log_2$ fold change (Koch *et al.*, 2018).

Figure 4.19A shows the processes involved in the TCA cycle. Figure 4.19B shows the relative changes in expression compared to Day 3 (EPLCs) of the genes in the TCA cycle for both MEDII-directed and spontaneous differentiation. A mixed expression profile is shown upon MEDII-directed differentiation.

Figure 4.19C shows none of the genes, except for *Idh3g*, have high DS ranks.

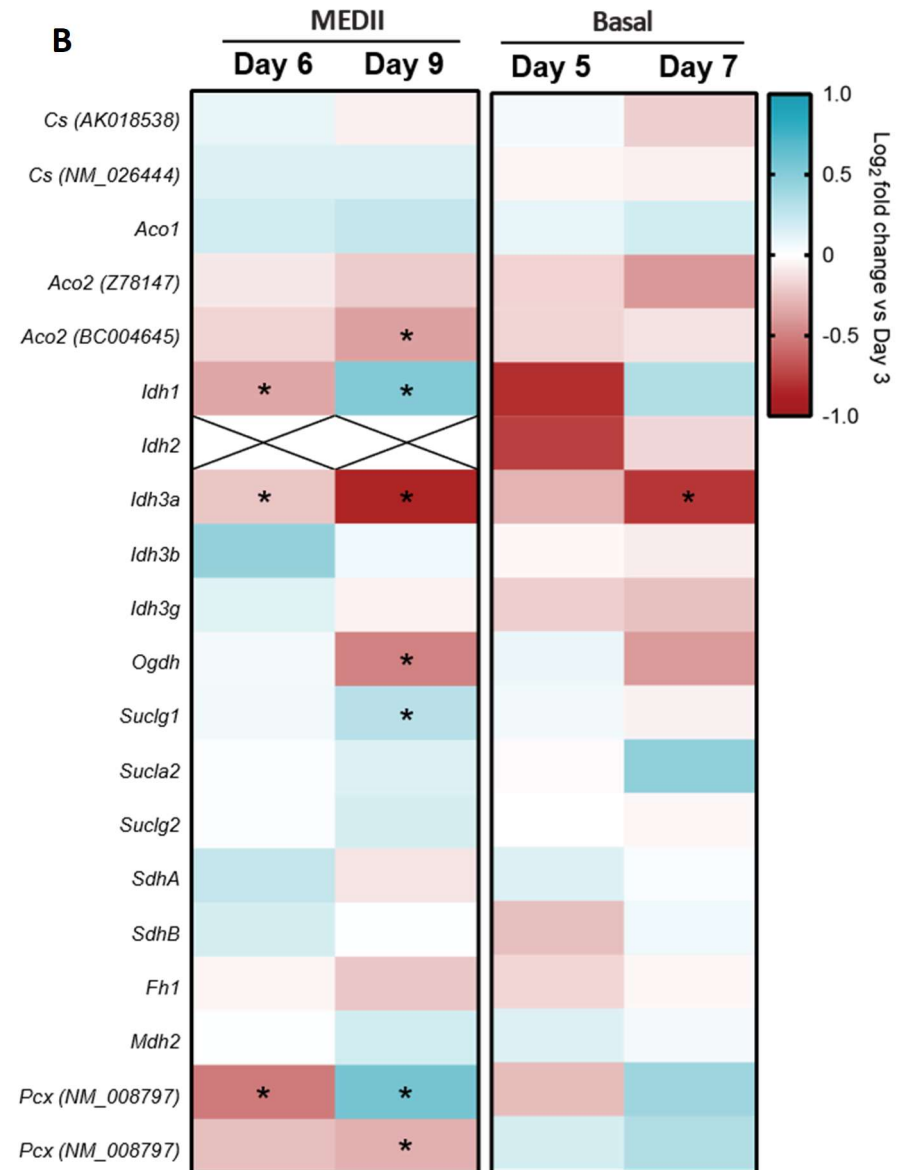
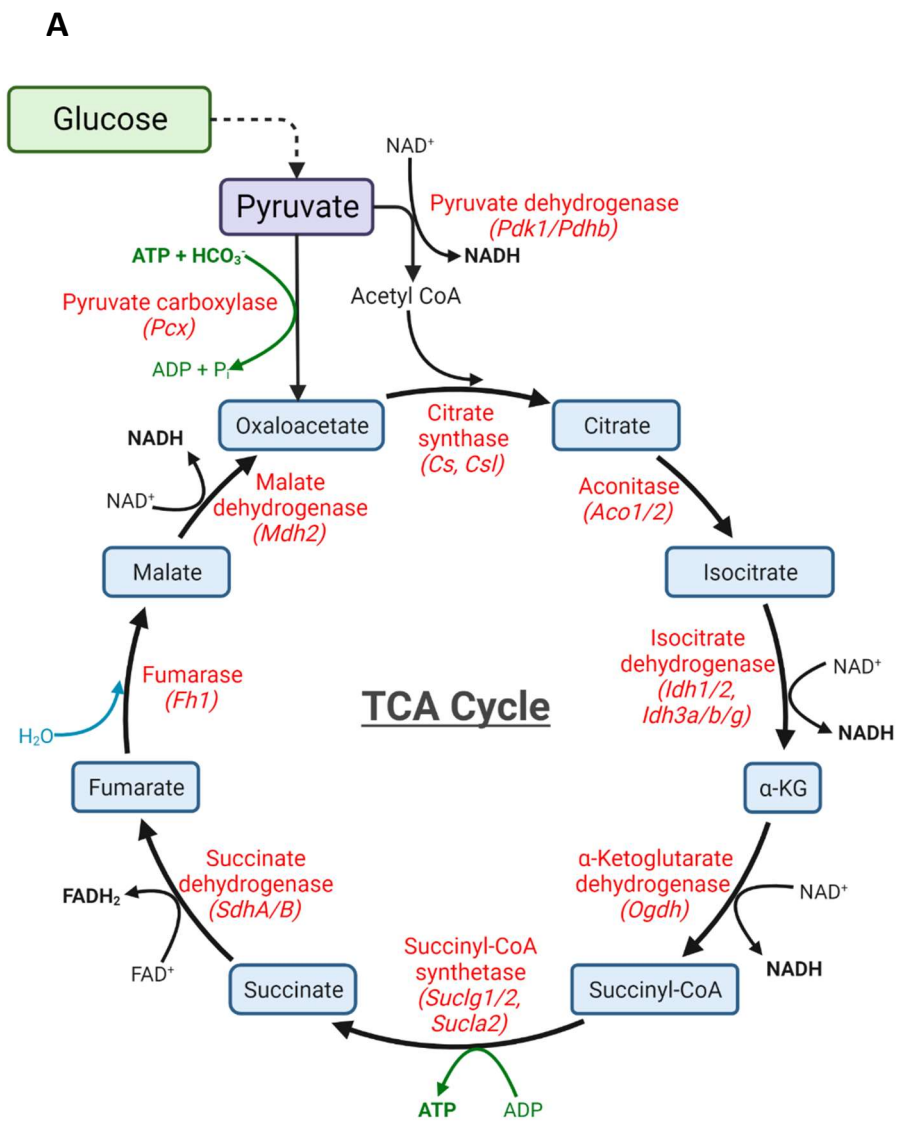


Figure 4.19 continued...

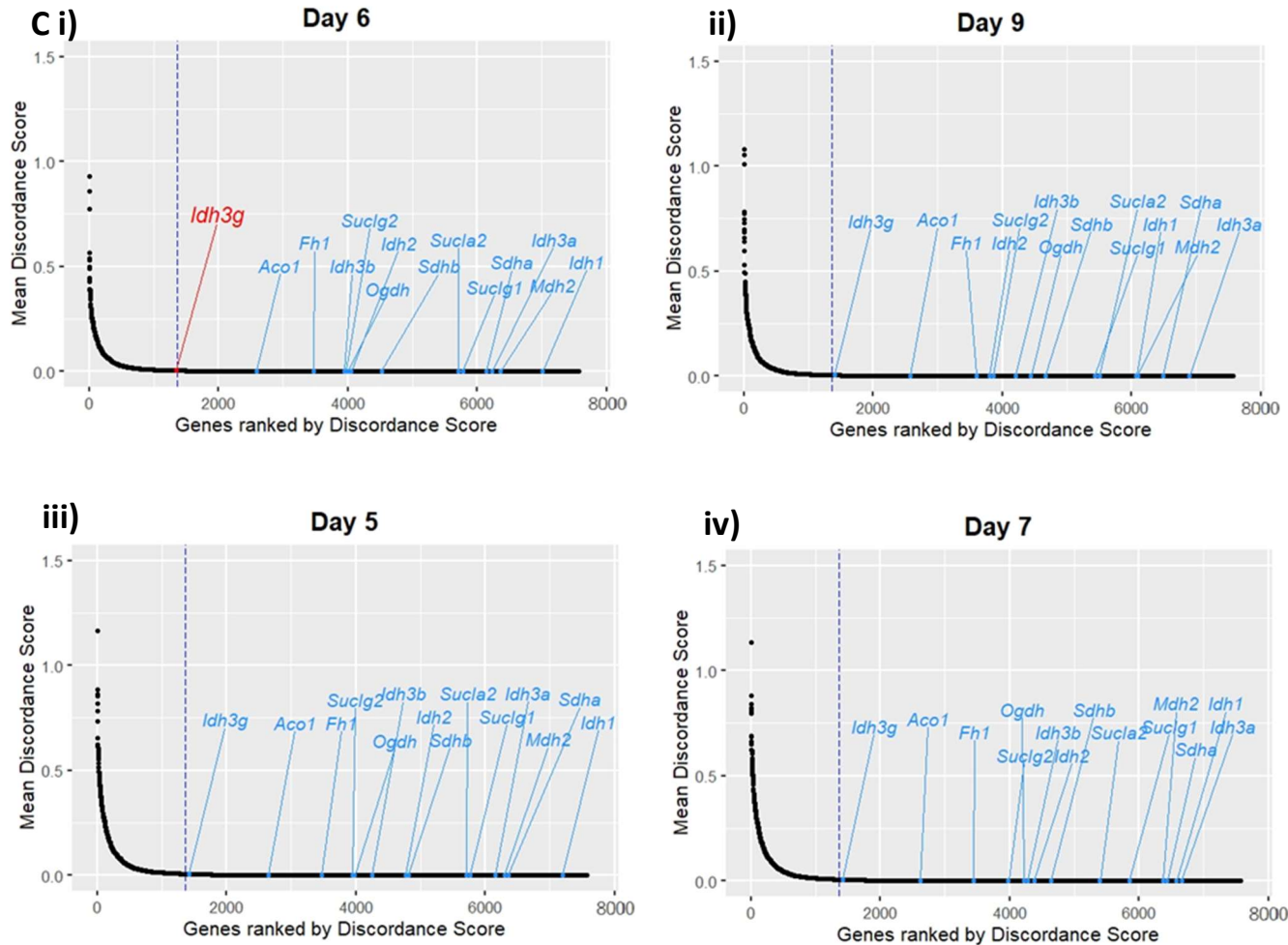


Figure 4.19 Relative expression and TRIAGE transformation of TCA cycle genes across neural lineage commitment. A. The TCA cycle and all the metabolite intermediates. Enzymes are shown in red and the corresponding genes in italics in parentheses. B-C. D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to i) Day 6 and ii) Day 9, or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to iii) Day 5 and iv) Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1A and 4.1B. Microarray data underwent B. differential gene analysis or C. TRIAGE analysis. PPP genes were ranked by fold changes relative to Day 3 (Expression) and mean DS across n=4 were plotted. The *Idh3g* gene in red indicates a DS rank ≤ 1359 . Blue genes indicate a DS rank > 1359 . * indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$.

4.3.9. Relative expression and TRIAGE analysis of genes associated with the electron transport chain across neural lineage commitment

NADH, produced from the TCA cycle, is utilised in the electron transport chain (ETC) to produce ATP (Figure 4.20A). Proline oxidase, the rate limiting step of L-Proline metabolism that catalyses L-Proline conversion to P5C, also mediates the mitochondrial ETC by donating electrons through an intervening flavoprotein into complex II (Figure 4.20A) (Aral and Kamoun, 1997; Phang, Donald, *et al.*, 2008; Pallag *et al.*, 2022). Cytochrome C accepts these electrons to contribute to generation of ATP (Figure 4.20A) (Phang, Donald, *et al.*, 2008; Hancock *et al.*, 2016). Proline oxidase also shares its localisation with the ETC in the inner membrane of the mitochondria (Figure 4.20A) (Hancock *et al.*, 2016; Pallag *et al.*, 2022). In addition, upon differentiation, there is an increase in energy demand for specialised functions upon lineage commitment (Chen *et al.*, 2010; Rehman, 2010; Varum *et al.*, 2011; Agathocleous *et al.*, 2012; Panopoulos *et al.*, 2012; Knobloch and Jessberger, 2017; Arrázola *et al.*, 2019; Tsogtbaatar *et al.*, 2020). In keeping up with this higher energy demand, mitochondria mature to accommodate for more efficient ETC activity to produce ATP.

With this rationale, we used differential gene expression analysis and TRIAGE transformation of the MEDII microarray data to examine if the ETC contributes to neural cell fate. To account for FDRs, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of $\log_2$ fold change (Koch *et al.*, 2018).

Figure 4.20A shows complexes and enzymes that contribute to the ETC. Figure 4.20B shows the relative changes in expression compared to Day 3 (EPLCs) of the genes in the ETC for both MEDII-directed and spontaneous differentiation. Only the uncoupling protein gene *Ucp2* is downregulated in differentiation to neurectoderm.

Figure 4.20C shows none of the genes in the ETC has a high DS rank.

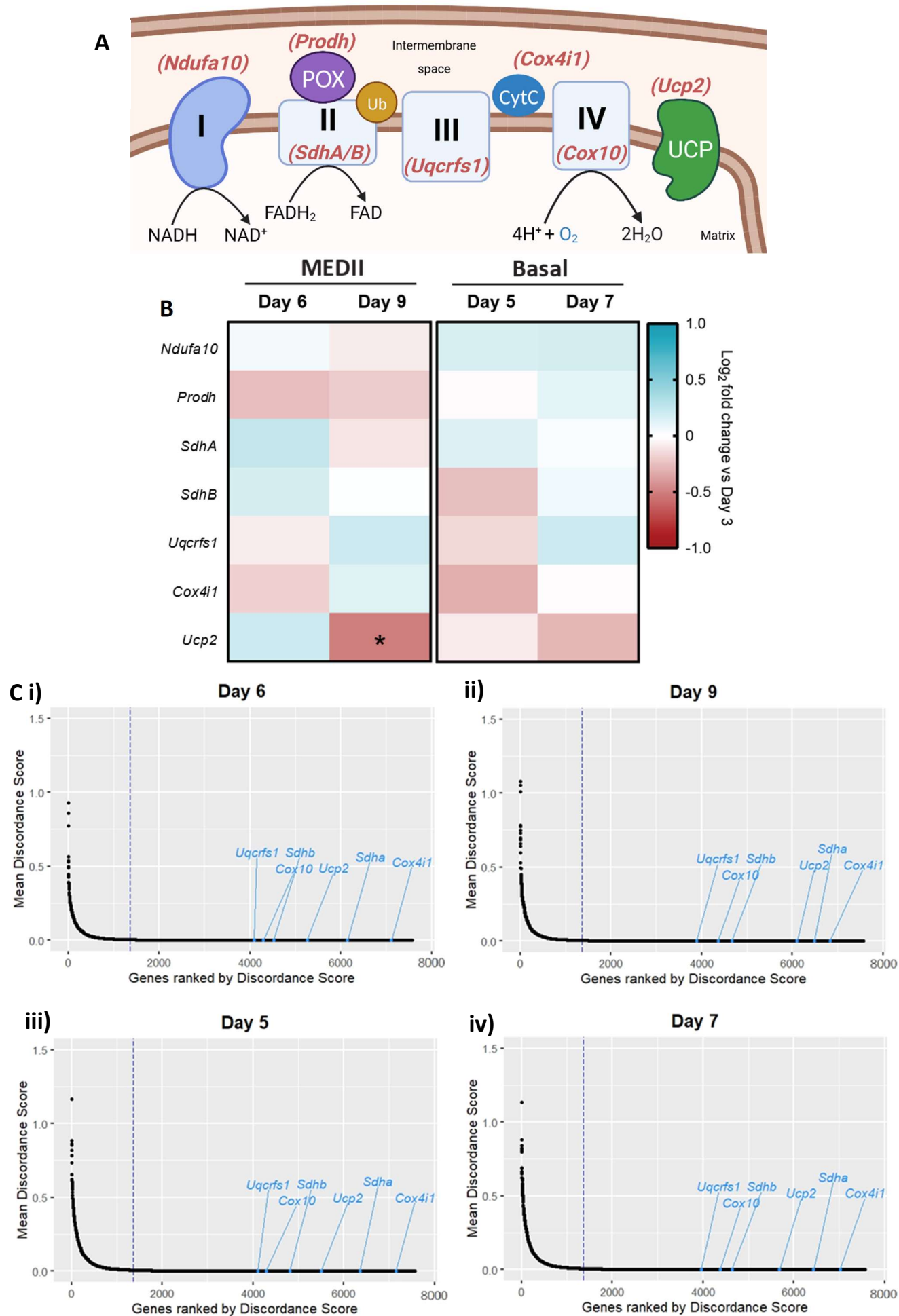


Figure 4.20 Relative expression and TRIAGE transformation of ETC genes across neural lineage commitment. The ETC and some of the associated complexes (I-IV), as well as proline oxidase (POX) and cytochrome-c oxidase (CytC) activity. Uncoupling protein (UCP), is known to lower mitochondrial membrane potential as a negative feedback mechanism in preventing oxidative stress accumulation. The genes encoding for each complex are shown in italics in parentheses. **B-C.** D3 mESCs were subjected to MEDII differentiation in suspension as EBs to Day 3 (EPLCs) and either continued in MEDII to **i)** Day 6 and **ii)** Day 9, or spontaneous differentiation in Basal medium (with a different set of Day 3 EBs) to **iii)** Day 5 and **iv)** Day 7. Samples (n=4) were subjected to produce microarray data, as per Figure 4.1A and 4.1B. Microarray data underwent **B.** differential gene analysis or **C.** TRIAGE analysis. PPP genes were ranked by fold changes relative to Day 3 (Expression) and mean DS across n=4 were plotted. Blue genes indicate a DS rank > 1359. * indicates significance as measured by a Benjamini-Hochberg adjusted $P < 0.05$. Abbreviations: *Ndufa10* (NADH): NADH dehydrogenase [ubiquinone] 1; *Prodh*: Proline (oxidase) dehydrogenase; *SdhA*: succinate dehydrogenase subunit A, flavoprotein (Fp); *SdhB*: succinate dehydrogenase subunit B, iron sulfur (Ip); *Uqcrrf1*: Complex III Rieske subunit; *Cox4i1*: cytochrome C oxidase subunit IV; *Cox10*: heme A: farnesyltransferase cytochrome c oxidase assembly factor 10.

4.4. Discussion

For successful organism development, determining cell fate and establishing cell identity are important processes for pluripotent cells to commit ultimately to a range of lineages. Mechanisms that drive cell fate and identity include several classes of genes including, for example, transcription factors whose expression can be repressed and activated, respectively, by PcG and TrxG protein complexes (Ringrose and Paro, 2004; Boyer *et al.*, 2006; MacArthur *et al.*, 2009; Kingston and Tamkun, 2014; Lau and So, 2015). These complexes include PRC1 and PRC2 that are composed of PcG proteins (Figure 1.5), and COMPASS complexes that are composed of TrxG proteins (Figure 1.6) (Ringrose and Paro, 2004, 2007; Boyer *et al.*, 2006; MacArthur *et al.*, 2009; Kingston and Tamkun, 2014; Lau and So, 2015; Luo *et al.*, 2020). More recently, metabolic processes are more frequently being recognised as drivers of cell fate (Shyh-Chang, N *et al.*, 2013; O'Brien *et al.*, 2015; Agostini *et al.*, 2016; Zheng *et al.*, 2016; Arrázola *et al.*, 2019) and presumably coordination of the activity of all these mechanisms (and others) contribute to directing cell fate of pluripotent stem cells (Okazaki and Maltepe, 2006).

In the previous Chapter, we focused on investigating cell-fate drivers and L-Proline-mediated metabolic processes that maintain cells in a pluripotent state. In this Chapter, we focused on how cells exit the pluripotent state to start committing to neural fate. Results from this Chapter have helped to identify candidate genes (such as those which encode for transcription factors, cell cycle control and cell signalling) and metabolic mechanisms which drive mESCs to neural cell fate when treated with L-Proline or MEDII.

4.4.1. MEDII and L-Proline primes mESCs to commit to neural fate as soon as mESCs leaves the naïve state

We took microarray datasets and looked at changes in gene expression across the whole mouse genome for MEDII-treated cells, which are lineage-directed to neural fate, and spontaneously differentiated cells, which can differentiate to all cell types. Figure 4.3 shows that MEDII-

treated cells lose pluripotency over time, as measured by reduced expression of core pluripotency and extended pluripotency genes, as they differentiate from naïve mESCs to pluripotent EPLCs (Day 3) to germ layer definitive ectoderm (Day 6) and neurectoderm (Day 9). This is consistent with the fact that robust expression of extended pluripotency genes supports expression of core pluripotency genes (Boyer *et al.*, 2005; Okumura-Nakanishi *et al.*, 2005; Casanova *et al.*, 2011; Xue *et al.*, 2011; Shparberg *et al.*, 2019b), and so reduced expression of extended pluripotency genes will result in eventual loss of pluripotency. Although some extended pluripotency genes have significantly upregulated, other downregulated extended pluripotency genes contribute to a cancelling out effect which allows priming of EBs to differentiate.

There is also temporally regulated increase in the expression of EPLC markers (at Day 3 and 6), definitive ectoderm markers (at Day 6) and neurectoderm markers (at Day 9). These results indicate a ‘temporal wave’ of directed differentiation to neural cells via embryologically relevant intermediate cell types. The results are consistent with the finding that when MEDII-treated Day 9 EBs are seeded down and allowed to grow for a further 6 days, most of the EBs show neural-cell features (Rathjen *et al.*, 1999, 2002). Similar morphological results for EBs are achieved when L-Proline replaces MEDII (Figure 4.5) (Shparberg *et al.*, 2019a) indicating that 400 μ M L-Proline, which is present at similar concentration in MEDII (Washington *et al.*, 2010), is capable of promoting neural lineage commitment.

Once formed, definitive ectoderm represents a bifurcation point in which cells commit either to neurectoderm or surface ectoderm (Harvey *et al.*, 2010; Li *et al.*, 2013). However, there is no change in the expression of surface ectoderm markers in MEDII-treated cells (Figure 4.3A). This indicates MEDII directs definitive ectoderm preferentially to neurectoderm rather than surface ectoderm, and in fact directing definitive ectoderm to surface ectoderm requires exogenous addition of BMP4 (Rathjen *et al.*, 2002; Harvey *et al.*, 2010).

Similarly, there are no changes in the expression of trophoctoderm markers with MEDII (Figure 4.3A), indicating that it does not direct mESCs to this extraembryonic cell type, which ultimately goes on to form the placenta.

Collectively these results are consistent with the idea that the amino acid L-Proline acts in a growth factor-like manner to promote lineage commitment to neural cells. However, Figure 4.5 shows that L-Proline is not required for directing neurectoderm to mature neural cell types, since its addition to serum-free medium at the neurectoderm stage made no significant difference to the percentage of EBs showing neural features at Day 15 compared to its absence. Rather, L-Proline promotes neural commitment at earlier stages; specifically, in the mESC-to-EPLC transition (Rathjen *et al.*, 1999; Washington *et al.*, 2010; Casalino *et al.*, 2011) and the differentiation of EPLCs to definitive ectoderm and neurectoderm (Figure 4.3) (Shparberg, 2018; Shparberg *et al.*, 2019a).

The results are also consistent with L-Proline (and L-Gln) acting in a growth factor-like manner to stimulate development in pre-implantation mouse embryos out to the hatching blastocyst stage (Chatot *et al.*, 1989; Erbach *et al.*, 1994; Lane and Gardner, 1997; Morris *et al.*, 2020; Hardy, 2023). The number of growth factors is limited in number and so their re-use at different stages of development is to be expected.

MEDII-treated mESCs also upregulated expression of definitive mesoderm and definitive endoderm markers at Day 6 and 9 (Figure 4.3A). This likely reflects off-pathway differentiation and is consistent with the fact that not all EBs scored at Day 15 after seeding down, and not all cells in each seeded-down EB, show neural features (Figure 4.5) (Rathjen *et al.*, 2002). There is some evidence that with L-Proline-treated mESCs, the addition of the mesendoderm inhibitor, SB431542, can prevent some of this off-pathway differentiation to these germ layers (Shparberg, 2018).

Generally speaking, the temporal changes in gene expression for pluripotency and specific-cell markers seen for MEDII-directed differentiation weren't observed, or were inconsistently observed, for spontaneously differentiating cells, indicating the asynchronous, heterogeneous nature of differentiation in Basal medium, where lineage commitment is not directed (Figure 4.3B). In particular, there is no consistent change in the expression of primitive ectoderm markers and no change in the expression of definitive ectoderm or neurectoderm markers that are seen with the MEDII-treated cells (Figure 4.3A). There is, however, upregulation of the expression of gene markers for definitive mesoderm suggesting that spontaneously differentiating cells have some preference (likely not exclusive) for lineage commitment to this germ layer.

We modified the L-Proline protocol for directed differentiation of mESCs to neural cells by adding a hanging drop step when the EBs were first formed (Figure 4.1C). The purpose of this was to generate single EBs, which will be of very uniform size compared to the very wide range of EB sizes that result when large numbers of EBs form in bulk culture (Ohnuki and Kurosawa, 2013; Shparberg, 2018; Shparberg *et al.*, 2019a). In turn, this might be expected to reduce off-pathway differentiation to cell types other than those on differentiating towards neural cell types.

Although incorporation of the hanging drop still led to a large percentage of EBs having neural features at Day 15 (Figure 4.5), it did not improve results compared to the previous protocol without hanging drop (Shparberg, 2018; Shparberg *et al.*, 2019a). Nor were the time-dependent changes in marker expression over the course of the 9 days of differentiation observed with hanging drop (Figure 4.4) as they were without it (Shparberg, 2018; Shparberg *et al.*, 2019a).

One possibility is that with the more uniformly sized EBs from hanging drop, differentiation to the various intermediate cell populations (primitive ectoderm, definitive

ectoderm, neurectoderm) *en route* to mature neural cell types is more synchronous. In culture, and in the embryo, these cell types are transient and individual cells may spend only a few hours as each type before differentiating further (Stavridis and Smith, 2003; Ying, Stavridis, *et al.*, 2003; Rao *et al.*, 2020). If this is the case, then measuring marker gene expression on samples collected every day (instead of only Day 0, 3, 6 and 9) may shed light on this issue. This would also allow some transient markers such as *Sox1*, *Penk1* and *Pard6b*, to be captured during differentiation (Suter *et al.*, 2009; Harvey *et al.*, 2010). Composition of these hanging drop EBs can be assessed by using single-cell RNAseq. Bulk RNAseq and pipelining of L-Proline-mediated differentiation of EBs can be used to validate relative gene expression changes that were derived from the MEDII microarray dataset.

It's also clear that while the addition of L-Proline alone leads to very efficient lineage commitment to neural cell types, the process is not homogeneous. Two obvious strategies present for trying to improve efficiency. One is to combine L-Proline with other factors known to induce neural differentiation in mESCs, such as retinoic acid (Bibel *et al.*, 2004, 2007; Lang *et al.*, 2004; Engberg *et al.*, 2010; Stavridis *et al.*, 2010; Tonge and Andrews, 2010). Another is to use a chemically defined culture medium, thus excluding FBS. In particular, FBS contains BMP4, which helps maintain self-renewal of mESCs (Ying, Nichols, *et al.*, 2003; Qi *et al.*, 2004; Hirai *et al.*, 2011) and therefore may prevent (or at least slow down) the progression of mESCs to neural fate. BMP4 also promotes differentiation to mesendoderm (Inman *et al.*, 2002; Li *et al.*, 2013), and definitive ectoderm differentiation to surface ectoderm (Harvey *et al.*, 2010), so its presence may work against the actions of L-Proline. The L-Proline-based protocol used in this Chapter contains KOSR to reduce BMP4-mediated effects even in the presence of FBS (Shparberg, 2018; Shparberg *et al.*, 2019a). Combining L-Proline with BMP4 antagonists, such as chordin, noggin and/or follistatin, with Nodal inhibitor SB431542 to limit mesendoderm differentiation would be helpful (Bachiller *et al.*, 2000; Inman *et al.*, 2002; Li *et al.*, 2013; Brazil *et al.*, 2015; Shparberg *et al.*, 2019a). Single-cell RNAseq of EBs undergoing

this modified protocol may prove useful to validate the homogeneity of the process, and discover markers of transient cell populations (Kumar *et al.*, 2017; Mohammed *et al.*, 2017; Spangler *et al.*, 2018; Zeng *et al.*, 2018; Argelaguet *et al.*, 2019; Nishimura *et al.*, 2019; Tam and Ho, 2020) such as definitive ectoderm.

4.4.2. *TRIAGE transformation of genes shows candidate transcription factors directly driving neural fate commitment*

The microarray data was also transformed using TRIAGE to predict specific genes that drive cell fate and identity (Shim *et al.*, 2020). Candidate genes are selected based on their DS. Figure 4.6 identifies the top 10 ranked genes based on DS for both the MEDII-treated EPLCs and the spontaneously differentiated EPLCs. All of the top ten are genes for transcription factors (except for *Wnt1*) and are known to be involved in embryo development (Figure 4.6).

Across both days of both conditions, the set of top 10 candidate genes is very similar (Figure 4.6). On this basis, it's difficult to say if these genes help drive generalised loss of pluripotency and differentiation rather than directed differentiation to neural fate. The situation is complicated by the fact that many of the spontaneously differentiating cells will adopt a neural fate as well as a range of other fates.

However, note that the expression of the top 10 candidates in cells directed to a neural fate with MEDII is high compared to other genes in the genome; i.e., they appear on the plots in Figure 4.6A close to, or quite close to, the y-axis. On the other hand, the expression of the top 10 candidates for the spontaneously differentiated cells (Figure 4.6B) are scattered across the whole range of expression of genes in the genome. This indicates that the top candidates for the MEDII-treated cells are indeed key genes involved in neural cell fate, and is consistent with other research on directed differentiation of pluripotent cells showing that this requires upregulation of the expression of specific transcription factors (Artus *et al.*, 2011; Telley *et al.*, 2016; Spangler *et al.*, 2018; Oh and Jang, 2019; Rao *et al.*, 2020). In particular, *Neurog1* is in

the list for MEDII-treated cells at Day 9 (neurectoderm) and the gene's product, neurogenin1, is a transcription factor and key regulator of neuronal fate (Ma *et al.*, 1998; Guillemot, 2005; Kim *et al.*, 2011; Onoguchi *et al.*, 2012).

Each of these candidates can be tested further for their involvement, if any, in lineage directed differentiation to neural fate, for example by gene expression knockdown using siRNA or gene knockout. Ultimately, a gene knockout in an animal model may prove useful.

The top 100 DS-ranked genes were then subjected to cluster analysis using TRIAGE-Parser followed by STRING analysis (Figures 4.7 and 4.8).

In MEDII-directed differentiation of EBs, 4 clusters were identified by Day 6 (definitive ectoderm), followed by a decrease in the number of clusters to only 2 by Day 9 (neurectoderm) (Figure 4.7A and 4.7B). 3 out of 4 clusters at Day 6 (Cluster 1, 3 and 5) focus towards transcriptional activity in combination with cell-fate commitment (Figure 4.7A). 2 of the clusters at Day 6 (Cluster 1 and 5) focus towards lineage progression, which includes neural lineage commitment (Figure 4.7A). Cluster 2 only contains *HOX* genes with only the GO term “anterior/posterior pattern specification”, indicating a focus on body segmentation and embryonic patterning (Figure 4.7A) (Hunt, 1992; Tam and Behringer, 1997; Rathjen *et al.*, 2002). This cluster may work cooperatively with another gene cluster (Cluster 3), which contains GO terms also associated with body segmentation and embryonic patterning (“regionalization” and “dorsal/ventral pattern formation”) (Figure 4.7A).

Clusters 1 and 3 at Day 9 focus more on lineage commitment, with nervous system development being in the top 10 significant GO terms (Figure 4.7B). Both clusters contain GO terms related to transcription (Figure 4.7B). This means as MEDII differentiation progresses, pluripotent cells start to commit to multipotent cell lineages and the number of gene clusters decreases as they control more neural lineage-specific processes. Transcription of related genes contribute to neural lineage commitment. These results are consistent with work done

previously using MEDII on mESCs to model neural lineage commitment (Rathjen *et al.*, 2002, 2003; Harvey *et al.*, 2010).

In comparison to spontaneously differentiated EBs, 2 clusters were identified by Day 5, which increased to 4 by Day 7 (Figure 4.8A and 4.8B). GO enrichment analysis of Day 5 spontaneously differentiated EBs identified significant GO: biological processes contain terms associated with nervous system development beyond the top 10 (Cluster 1), followed by mesoderm commitment (i.e., skeletal development) and regionalisation as emphasised with *HOX* genes (Cluster 2) (Figure 4.8A). Most genes in the top 100 DS list are contained in Cluster 1, which focus on germ layer lineage progression, while Cluster 2 contained only *HOX* genes which focus on body segmentation and embryonic patterning, as indicated by the GO term “anterior/posterior pattern specification” (Figure 4.8A). The *HOX* gene cluster aligns with work by others, as Hox transcription factors are drivers of positional information to determine the anterior-posterior body plan (Hunt, 1992; Tam and Behringer, 1997; Rathjen *et al.*, 2002).

By Day 7, cluster and GO enrichment analysis showed 2 clusters associated with transcription, organism development, and germ layer lineage progression (Cluster 2 and 6), and another 2 clusters associated with regionalisation and organism development that is not necessarily lineage commitment-associated (Cluster 4 and 5) (Figure 4.8B). Overall, this means as spontaneous differentiation progresses, or in the embryo itself, complexity increases as the foundation of more body systems are established, and so more gene clusters are established that are responsible for driving development.

These results also highlight differences between neural lineage-directed commitment vs undirected lineage commitment from pluripotent stem cells: A decrease in cluster number with MEDII indicates directed lineage progression, while an increase in the number of clusters in spontaneously differentiated cells indicates a complex progression of development across multiple lineages.

The clustering, GO enrichment terms and PPI analyses only provide clues in the potential developmental processes these gene clusters control. In addition, the TRIAGE-Parser method requires the conversion of mouse genes to human orthologs for clustering analysis. Thus, PPIs are human-based and may not translate to the mouse model. Therefore, future work should utilise the identified gene clusters to siRNA knockout or knockdown target genes to test their relevance to development. A criterion to keep in mind when choosing a gene to knockout or knockdown is one with many PPIs with other genes. For example, PAX2 is a known transcription factor associated with hindbrain and kidney development (Lancaster *et al.*, 2013; Xia *et al.*, 2013, 2014). However, PPI analysis has shown it is associated with EN2, TLX1, LMX1B and EVX1 in Cluster 4 in Day 7 spontaneously differentiated EBs (Figure 4.8B). These transcription factors are associated with various lineages, particularly neural (EN2, LMX1B and EVX1), spleen (TLX1) and eye development (LMX1B) and anterior-posterior patterning during gastrulation (EVX1) (Rathjen *et al.*, 2002; Dang and Tropepe, 2010; Doucet-Beaupré *et al.*, 2015; Bell *et al.*, 2016; Juárez-Morales *et al.*, 2016; Poudel *et al.*, 2022). siRNA knockdown of the *Pax2* gene would be a potential target to test the relevance of this master regulator of this gene cluster in development. Knockdowns or knockouts of genes encoding proteins on the edges of clusters, or no PPIs, like *HOXB13*, would potentially have no effect on the developmental progress, at least at this particular time of cell-fate determination under these particular conditions (Figure 4.8B). Using this interrogative approach improves current reductive methods to test relevance of genes associated with cell fate and identity during embryonic development.

4.4.3. *PcG genes drive cell fate and identity across neural fate commitment, but TrxG genes do not*

Drivers of cell fate and identity during embryo development and neural lineage commitment in part requires gene repression and activation via PcG and TrxG protein complexes (Boyer *et al.*, 2006; Ringrose and Paro, 2007; MacArthur *et al.*, 2009; Aranda *et al.*, 2015; Brand *et al.*, 2019).

The antagonistic mechanisms of PcG and TrxG complexes allow cells to ‘remember’ what gene expression patterns and molecular mechanisms are required to establish a particular cell type (Ringrose and Paro, 2004, 2007; Brand *et al.*, 2019). PcG proteins repress gene expression via chromatin compaction while TrxG proteins antagonise this by opening chromatin at specific loci determined by lineage-specific transcription factors (Ying, Nichols, *et al.*, 2003; Ringrose and Paro, 2004; Boyer *et al.*, 2006; Ho *et al.*, 2009; MacArthur *et al.*, 2009; Hirai *et al.*, 2011; Lau and So, 2015; Brand *et al.*, 2019). Thus, we used differential expression analysis and TRIAGE transformation of genes encoding PcG and TrxG proteins to examine if their expression plays a role in driving cells across the pluripotency spectrum and into lineage commitment.

In terms of PcG complex genes, EPLCs demonstrated stable expression relative to naïve state (Figure 4.9A). However, ground state and primed EpiSCs demonstrated upregulation of core PRC1 subunits relative to naïve state (*Cbx4/8*, *Pcgf1* and *Bmi1*) (Figure 4.9A). Primed EpiSCs also demonstrate downregulation of PRC1 core subunit *Phc1* and PRC2 core subunit gene *Eed* (Figure 4.9A). As some genes for PRC1/2 core subunits (*Cbx2/4/8*, *Bmi1*, *Rbbp7*, *Pcgf1*, *Ring1* and *Phc2*) were also identified in the top-scoring DS-ranked genes across the pluripotency spectrum and neural lineage commitment, these subunits may play an important role in controlling the transition of mESCs from the ‘immature’ ground state to the primed EpiSC state (Figure 4.9A and 4.11), and exit out of pluripotency to neural fate, at least on a transcriptional level (Figure 4.9B and 4.13).

Other research shows that maintenance of pluripotency and self-renewal depends on the Cbx7-containing PRC1 complex which suppresses pro-differentiation Cbx protein activities, including Cbx2 and Cbx4 (Morey *et al.*, 2012; Santanach *et al.*, 2017). However, our results show no significant change in expression level *Cbx2/6/7*, and only *Cbx2* is found in the top-ranked DS genes (Figure 4.9A and 4.11). Thus, maintaining *Cbx2* and not *Cbx4* is required for transition along the pluripotency spectrum (Morey *et al.*, 2012).

In addition, the protein products of *Phc1* and *Eed* silence expression of differentiation-associated genes in self-renewing ESCs (Pasini *et al.*, 2007; Casanova *et al.*, 2011), so downregulation of their expression in primed EpiSCs (Figure 4.9A) is consistent with preparation for the loss of pluripotency.

TrxG proteins are not generally required in ESCs to maintain pluripotency, except for Ash2l and Wdr5 (Ang *et al.*, 2011; Wan *et al.*, 2013; Lau and So, 2015; Tsai *et al.*, 2019). Ash2l and Wdr5 are 2 of the 4 core components of all COMPASS complexes, the others being Rbbp5 and Dpy-30 (Dou *et al.*, 2006; Wu *et al.*, 2008; Herz *et al.*, 2013; Lau and So, 2015; Piunti and Shilatifard, 2016). Ash2l and Wdr5 contribute to the core pluripotency circuitry by recruitment of Oct4, Nanog and Sox2 to drive upregulation of expression of stemness genes such as *Klf4* and *c-Myc* (Ang *et al.*, 2011; Wan *et al.*, 2013; Lau and So, 2015; Tsai *et al.*, 2019).

Our differential gene expression analysis shows significant downregulation of *Rbbp5*, *Dpy-30* and *Setd1b* as cells progress from the ground to the primed state (Figure 4.10A), which is consistent with pluripotent cells preparing to their pluripotency.

MLL3/4 is one of the methyltransferases of the SET1-family that are involved in gene regulatory networks (Ernst and Vakoc, 2012; Wang *et al.*, 2016). The expression of the gene encoding for its subunit, *Ncoa6*, is significantly upregulated in the ground state relative to naïve state (Figure 4.10A). *Hcfc1*, which encodes for the subunit Hcfc1 as part of SET1A/B and MLL1/2 COMPASS complexes, is the only gene to appear in the top-ranked DS across the pluripotency spectrum (Figure 4.12) (Herz *et al.*, 2013; Lau and So, 2015). It's possible that one or both of these genes plays a role in the transition of cells across the pluripotency spectrum and regulate the expression should prove useful in this regard.

This leads to some of the limitations of this interrogation process: (i) Relative changes in gene expression do not necessarily correlate with actual changes in histone modifications or correlate to function or mechanisms directly, and (ii) the activity of PcG and TrxG complexes

are not captured at the transcriptome level but rather the proteome/protein activity and epigenome (histone modification and DNA methylation) levels. Coordination between various regulatory mechanisms, including epigenetics, transcription, signalling and protein-protein interactions are required to drive cell fate choices (Lu *et al.*, 2009; MacArthur *et al.*, 2009; Martynoga *et al.*, 2012; Luo *et al.*, 2020). Investigating the epigenetic status of cells across the pluripotency spectrum as well as EBs directed to the neural lineage using ChIPSeq pipelining should prove valuable to understanding this complex interplay between various levels of regulatory control.

4.4.4. Metabolism mediates other mechanisms driving neural fate commitment

When pluripotent cells commit to a particular lineage, energy production shifts from anaerobic to aerobic metabolism due to increasing complexity and energy demands (Rehman, 2010; Varum *et al.*, 2011; Agathocleous *et al.*, 2012; Panopoulos *et al.*, 2012; O'Brien *et al.*, 2015; Hu *et al.*, 2016; Knobloch and Jessberger, 2017; Arrázola *et al.*, 2019; Tsogtbaatar *et al.*, 2020). In terms of differentiation to neural cells, mitochondrial biogenesis increases to accommodate this increasing demand (O'Brien *et al.*, 2015; Agostini *et al.*, 2016; Zheng *et al.*, 2016). In keeping with this, neural stem cells have a preference for oxidative phosphorylation and fatty acid oxidation (Hu *et al.*, 2016; Zheng *et al.*, 2016; Arrázola *et al.*, 2019). L-Proline, the active component in MEDII, drives neural lineage commitment of mESCs and is metabolised via the proline cycle (Rathjen *et al.*, 2002, 2003; Phang, Donald, *et al.*, 2008; Pandhare *et al.*, 2009; Washington *et al.*, 2010; Casalino *et al.*, 2011; Shparberg, 2018; Glover *et al.*, 2022; Pallag *et al.*, 2022).

Thus, we looked at the changes in gene expression of a range of metabolic processes, which directly and indirectly included the metabolism of L-Proline as EBs were directed to neural cells with MEDII to see if some of these play a direct role in this lineage commitment (Figures 4.15-4.20). The data were also transformed using TRIAGE to assist in this process (Figures 4.15-4.20).

Overall, there are few changes in gene expression across the metabolic pathways examined. L-Proline uptake and catabolism may decrease by Day 9 (neurectoderm), with concomitant increase in L-Proline production (Figures 4.15 And 4.16) but none of the genes in these pathways has a high DS rank. This doesn't mean the balance of activity between these two metabolic pathways isn't important for lineage progression since, for example, increased uptake and metabolism of L-Proline following its exogenous addition might be handled by spare capacity in pathway activity and/or control of activity at the post-translational level. Thus, it would appear that while metabolism of L-Proline is required for at least some aspects of lineage progression of ESCs (Casalino *et al.*, 2011; Comes *et al.*, 2013; D'Aniello *et al.*, 2015), the pathways themselves are not central to this process at the gene expression level but instead that control lies downstream.

In terms of the other pathways examined, gene expression for many glycolysis genes is upregulated at Day 9 (neurectoderm) (Figure 4.17) perhaps to increase flux through to the TCA cycle and electron transport chain to allow increased ATP production required for differentiation. However, changes in expression of genes for the TCA cycle (figure 4.19) and ETC (Figure 4.20) are modest.

For the PPP, expression of the *G6pdx*, the gene for G6PD, which is the rate-limiting enzyme of the pathway, is downregulated at Day 6 (definitive ectoderm) and Day 9 (neurectoderm) (Figure 4.18). This may indicate a reduced dependence on this pathway as differentiation to neural cells progresses. Measurement of actual pathway activity using cytosolic extracts together with enzyme activity assays and/or metabolic flux assays with LC-MS would help shed light on this issue.

Limitations of the current approach to interrogate metabolic processes include the following: (i) TRIAGE relies on the prior knowledge of absence and presence of broad H3K27me3 marks on genes to derive repressive tendency scores; (ii) the 22k mouse Affymetrix

platform not detecting all genes, or extra fold changes of the same genes indicative of splice variants; (iii) TRIAGE is limited to only protein-coding genes; and (iv) results are speculative. Like the limitations discussed previously (Section 4.4.3), future work must be done to verify expression of genes of interest, and capture differential gene expression profiles of L-Proline-mediated EB production using bulk RNAseq (Glover, 2018; Shparberg *et al.*, 2019a; Glover *et al.*, 2022). A snapshot of a cell population's metabolic state can be captured at the metabolomic level using LC-MS (Martano *et al.*, 2015, 2019). Metabolomics analysis can also help discover what metabolites drive cell fate across neural vs spontaneous differentiation of EBs. A snapshot of the metabolic preference in live EBs can be captured using the Seahorse XF assay. Metabolic flux analysis can be used to compliment interrogation of metabolic states across neural lineage commitment (Martano *et al.*, 2019). This will determine what specific metabolic pathways metabolites, as identified in metabolomics analysis, are active between different stages of neural lineage commitment (Martano *et al.*, 2019). This can be achieved using radioactive tracers on substrates of interest (De Bock *et al.*, 2013; Purnell, 2018; Martano *et al.*, 2019). For example, [³H]-L-Proline can be used to determine the fate in which L-Proline is metabolised during mESC-to-neural transition vs. spontaneous differentiation (Purnell, 2018). [¹³C]-glucose can be used to trace how glucose is metabolised, whether through glycolysis, PPP, and/or the TCA cycle, across both conditions and days (De Bock *et al.*, 2013; Martano *et al.*, 2019).

So far, interrogation of genes associated with metabolism suggests the following: at the first stage of neural lineage commitment, pluripotent cells transition from a bivalent metabolic profile to a glycolytic one upon differentiation to cells equivalent to neurectoderm. Although not direct drivers of cell fate, metabolism is still essential for cell signalling, cellular structure and morphological changes associated with differentiation.

4.5. References

- Agathocleous, M., Love, N. K., Randlett, O., Harris, J. J., Liu, J., Murray, A. J., and Harris, W. A. (2012) 'Metabolic differentiation in the embryonic retina', *Nature Cell Biology*, 14(8), pp. 859–864. doi: 10.1038/ncb2531.
- Agostini, M., Romeo, F., Inoue, S., Niklison-Chirou, M. V., Elia, A. J., Dinsdale, D., Morone, N., Knight, R. A., Mak, T. W., and Melino, G. (2016) 'Metabolic reprogramming during neuronal differentiation', *Cell Death & Differentiation*, 23(9), pp. 1502–1514. doi: 10.1038/cdd.2016.36.
- Aguilar, J. and Reyley, M. (2005) 'The uterine tubal fluid: secretion, composition and biological effects', *Animal Reproduction Science*, 2(2), pp. 91–105.
- Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., Wang, J., Rendl, M., Bernstein, E., Schaniel, C., and Lemischka, I. R. (2011) 'Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network', *Cell*. Elsevier, 145(2), pp. 183–197. doi: 10.1016/j.cell.2011.03.003.
- Aral, B. and Kamoun, P. (1997) 'The proline biosynthesis in living organisms', *Amino Acids*, 13(3–4), pp. 189–217. doi: 10.1007/BF01372588.
- Aranda, S., Mas, G., and Di Croce, L. (2015) 'Regulation of gene transcription by polycomb proteins', *Science Advances*, 1(11), p. e1500737. doi: 10.1126/sciadv.1500737.
- Argelaguet, R., Clark, S. J., Mohammed, H., Stapel, L. C., Krueger, C., Kapourani, C.-A., Imaz-Rosshandler, I., Lohoff, T., Xiang, Y., Hanna, C. W., Smallwood, S., Ibarra-Soria, X., Buettner, F., Sanguinetti, G., Xie, W., Krueger, F., Göttgens, B., Rugg-Gunn, P. J., Kelsey, G., Dean, W., Nichols, J., Stegle, O., Marioni, J. C., and Reik, W. (2019) 'Multi-omics profiling of mouse gastrulation at single-cell resolution', *Nature*, 576(7787), pp. 487–491. doi: 10.1038/s41586-019-1825-8.
- Arnold, S. J. and Robertson, E. J. (2009) 'Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(2), pp. 91–103. doi: 10.1038/nrm2618.

Arrázola, M. S., Andraini, T., Szelechowski, M., Mouledous, L., Arnauné-Pelloquin, L., Davezac, N., Belenguer, P., Rampon, C., and Miquel, M.-C. (2019) 'Mitochondria in developmental and adult neurogenesis', *Neurotoxicity Research*, 36(2), pp. 257–267. doi: 10.1007/s12640-018-9942-y.

Artus, J., Piliszek, A., and Hadjantonakis, A.-K. (2011) 'The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17', *Developmental Biology*, 350(2), pp. 393–404. doi: 10.1016/j.ydbio.2010.12.007.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J. A., Anderson, R. M., May, S. R., McMahon, J. A., McMahon, A. P., Harland, R. M., Rossant, J., and De Robertis, E. M. (2000) 'The organizer factors Chordin and Noggin are required for mouse forebrain development', *Nature*, 403(6770), pp. 658–661. doi: 10.1038/35001072.

Bell, C. C., Amaral, P. P., Kalsbeek, A., Magor, G. W., Gillinder, K. R., Tangermann, P., di Lisio, L., Cheetham, S. W., Gruhl, F., Frith, J., Tallack, M. R., Ru, K.-L., Crawford, J., Mattick, J. S., Dinger, M. E., and Perkins, A. C. (2016) 'The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation', *Scientific Reports*, 6(1), p. 26657. doi: 10.1038/srep26657.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) 'A bivalent chromatin structure marks key developmental genes in embryonic stem cells', *Cell*, 125(2), pp. 315–326. doi: 10.1016/j.cell.2006.02.041.

Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.-A. (2004) 'Differentiation of mouse embryonic stem cells into a defined neuronal lineage', *Nature Neuroscience*, 7(9), pp. 1003–1009. doi: 10.1038/nn1301.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.-A. (2007) 'Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells', *Nature Protocols*, 2(5), pp. 1034–1043. doi: 10.1038/nprot.2007.147.

De Bock, K., Georgiadou, M., Schoors, S., Kuchnio, A., Wong, B. W., Cantelmo, A. R., Quaegebeur, A., Ghesquière, B., Cauwenberghs, S., Eelen, G., Phng, L.-K., Betz, I., Tembuyser, B.,

Brepoels, K., Welte, J., Geudens, I., Segura, I., Cruys, B., Bifari, F., Decimo, I., Blanco, R., Wyns, S., Vangindertael, J., Rocha, S., Collins, R. T., Munck, S., Daelemans, D., Imamura, H., Devlieger, R., Rider, M., Van Veldhoven, P. P., Schuit, F., Bartrons, R., Hofkens, J., Fraisl, P., Telang, S., DeBerardinis, R. J., Schoonjans, L., Vinckier, S., Chesney, J., Gerhardt, H., Dewerchin, M., and Carmeliet, P. (2013) 'Role of PFKFB3-driven glycolysis in vessel sprouting', *Cell*, 154(3), pp. 651–663. doi: <https://doi.org/10.1016/j.cell.2013.06.037>.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell*, 122(6), pp. 947–956. doi: 10.1016/j.cell.2005.08.020.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) 'Polycomb complexes repress developmental regulators in murine embryonic stem cells', *Nature*, 441(7091), pp. 349–353. doi: 10.1038/nature04733.

Brand, M., Nakka, K., Zhu, J., and Dilworth, F. J. (2019) 'Polycomb/Trithorax antagonism: Cellular memory in stem cell fate and function', *Cell Stem Cell*, 24(4), pp. 518–533. doi: <https://doi.org/10.1016/j.stem.2019.03.005>.

Brazil, D. P., Church, R. H., Surae, S., Godson, C., and Martin, F. (2015) 'BMP signalling: Agony and antagonism in the family', *Trends in Cell Biology*, 25(5), pp. 249–264. doi: 10.1016/J.TCB.2014.12.004.

Casalino, L., Comes, S., Lambazzi, G., De Stefano, B., Filosa, S., De Falco, S., De Cesare, D., Minchiotti, G., and Patriarca, E. J. (2011) 'Control of embryonic stem cell metastability by L-Proline catabolism', *Journal of Molecular Cell Biology*, 3(2), pp. 108–122. doi: 10.1093/jmcb/mjr001.

Casanova, E. A., Bürki, K., and Cinelli, P. (2011) 'Molecular mechanisms of pluripotency in murine embryonic stem cells', in Atwood, C. (ed.) *Embryonic Stem Cells: The Hormonal Regulation of Pluripotency and Embryogenesis*. InTech, pp. 27–60. doi: 10.5772/14219.

Chatot, C. L., Ziomek, C. A., Bavister, B. D., Lewis, J. L., and Torres, I. (1989) 'An improved culture medium supports development of random-bred 1-cell mouse embryos in vitro', *Reproduction*. Bristol, UK: Bioscientifica Ltd, 86(2), pp. 679–688. doi: 10.1530/jrf.0.0860679.

Chen, C.-T., Hsu, S.-H., and Wei, Y.-H. (2010) 'Upregulation of mitochondrial function and antioxidant defense in the differentiation of stem cells', *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1800(3), pp. 257–263. doi: 10.1016/j.bbagen.2009.09.001.

Collinson, A., Collier, A. J., Morgan, N. P., Sienerth, A. R., Chandra, T., Andrews, S., and Rugg-Gunn, P. J. (2016) 'Deletion of the Polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells', *Cell Reports*, 17(10), pp. 2700–2714. doi: 10.1016/j.celrep.2016.11.032.

Comes, S., Gagliardi, M., Laprano, N., Fico, A., Cimmino, A., Palamidessi, A., De Cesare, D., De Falco, S., Angelini, C., Scita, G., Patriarca, E. J., Matarazzo, M. R., and Minchiotti, G. (2013) 'L-Proline induces a mesenchymal-like invasive program in embryonic stem cells by remodeling H3K9 and H3K36 methylation', *Stem Cell Reports*, 1(4), pp. 307–321. doi: 10.1016/j.stemcr.2013.09.001.

Di Croce, L. and Helin, K. (2013) 'Transcriptional regulation by polycomb group proteins', *Nature Structural & Molecular Biology*, 20(10), pp. 1147–1155. doi: 10.1038/nsmb.2669.

D'Aniello, C., Fico, A., Casalino, L., Guardiola, O., Di Napoli, G., Cermola, F., De Cesare, D., Tatè, R., Cobellis, G., Patriarca, E. J., and Minchiotti, G. (2015) 'A novel autoregulatory loop between the Gcn2-Atf4 pathway and L-Proline metabolism controls stem cell identity', *Cell Death And Differentiation*, 22, pp. 1094–1105. doi: 10.1038/cdd.2015.24.

D'Aniello, C., Habibi, E., Cermola, F., Paris, D., Russo, F., Fiorenzano, A., Di Napoli, G., Melck, D. J., Cobellis, G., Angelini, C., Fico, A., Blelloch, R., Motta, A., Stunnenberg, H. G., De Cesare, D., Patriarca, E. J., and Minchiotti, G. (2017) 'Vitamin C and L-Proline antagonistic effects capture alternative states in the pluripotency continuum', *Stem Cell Reports*. Elsevier, 8(1), pp. 1–10. doi: 10.1016/j.stemcr.2016.11.011.

Dang, L. T. H. and Tropepe, V. (2010) 'FGF dependent regulation of Zfhx1b gene expression

promotes the formation of definitive neural stem cells in the mouse anterior neurectoderm', *Neural Development*, 5(1), p. 13. doi: 10.1186/1749-8104-5-13.

Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006) 'Regulation of MLL1 H3K4 methyltransferase activity by its core components', *Nature Structural & Molecular Biology*, 13(8), pp. 713–719. doi: 10.1038/nsmb1128.

Doucet-Beaupré, H., Ang, S.-L., and Lévesque, M. (2015) 'Cell fate determination, neuronal maintenance and disease state: The emerging role of transcription factors Lmx1a and Lmx1b', *FEBS Letters*, 589(24A), pp. 3727–3738. doi: 10.1016/j.febslet.2015.10.020.

Engberg, N., Kahn, M., Petersen, D. R., Hansson, M., and Serup, P. (2010) 'Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent Nodal signaling', *Stem Cells*, 28(9), pp. 1498–1509. doi: 10.1002/stem.479.

Erbach, G. T., Lawitts, J. A., Papaioannou, V. E., and Biggers, J. D. (1994) 'Differential growth of the mouse preimplantation embryo in chemically defined media', *Biology of Reproduction*, 50(5), pp. 1027–1033. doi: 10.1095/biolreprod50.5.1027.

Ernst, P. and Vakoc, C. R. (2012) 'WRAD: Enabler of the SET1-family of H3K4 methyltransferases', *Briefings in Functional Genomics*, 11(3), pp. 217–226. doi: 10.1093/bfpg/els017.

Fiorenza, M. T. and Rava, A. (2019) 'The TCL1 function revisited focusing on metabolic requirements of stemness', *Cell Cycle*, 18(22), pp. 3055–3063. doi: 10.1080/15384101.2019.1672465.

Forristal, C. E., Wright, K. L., Hanley, N. A., Oreffo, R. O. C., and Houghton, F. D. (2010) 'Hypoxia inducible factors regulate pluripotency and proliferation in human embryonic stem cells cultured at reduced oxygen tensions', *Reproduction*, 139(1), pp. 85–97. doi: 10.1530/REP-09-0300.

Gangloff, Y., Mueller, M., Dann, S. G., Svoboda, P., Sticker, M., Spetz, J., Um, H., Brown, E. J., Cereghini, S., Thomas, G., Kozma, S. C., and Um, S. H. (2004) 'Disruption of the mouse mTOR gene leads to early postimplantation lethality and prohibits embryonic stem cell development', *Molecular and Cellular Biology*, 24(21), pp. 9508–9516. doi: 10.1128/MCB.24.21.9508.

Gao, X., Tate, P., Hu, P., Tjian, R., Skarnes, W. C., and Wang, Z. (2008) 'ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a', *Proceedings of the National Academy of Sciences*, 105(18), pp. 6656–6661. doi: 10.1073/pnas.0801802105.

Ge, T., Yang, J., Zhou, S., Wang, Y., Li, Y., and Tong, X. (2020) 'The role of the pentose phosphate pathway in diabetes and cancer', *Frontiers in Endocrinology*, 11. doi: 10.3389/fendo.2020.00365.

Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.

Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) 'L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell', in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.

Guillemot, F. (2005) 'Cellular and molecular control of neurogenesis in the mammalian telencephalon', *Current Opinion in Cell Biology*. Elsevier Current Trends, 17(6), pp. 639–647. doi: 10.1016/J.CEB.2005.09.006.

Halestrap, A. P. (2013) 'The SLC16 gene family – Structure, role and regulation in health and disease', *Molecular Aspects of Medicine*, 34(2), pp. 337–349. doi: <https://doi.org/10.1016/j.mam.2012.05.003>.

Hancock, C. N., Liu, W., Alvord, W. G., and Phang, J. M. (2016) 'Co-regulation of mitochondrial respiration by proline dehydrogenase/oxidase and succinate', *Amino Acids*, 48(3), pp. 859–872. doi: 10.1007/s00726-015-2134-7.

Hardy, M. L. M. (2023) *Mechanisms for the beneficial effect of L-Proline on preimplantation embryo development*. University of Sydney.

Harvey, A., Caretti, G., Moresi, V., Renzini, A., and Adamo, S. (2019) 'Interplay between metabolites and the epigenome in regulating embryonic and adult stem cell potency and maintenance',

Stem Cell Reports, 13(4), pp. 573–589. doi: 10.1016/j.stemcr.2019.09.003.

Harvey, A. J., Rathjen, J., and Gardner, D. K. (2016) ‘Metaboloepigenetic regulation of pluripotent stem cells’, *Stem Cells International*, 2016, pp. 1–15. doi: 10.1155/2016/1816525.

Harvey, N. T., Hughes, J. N., Lonic, A., Yap, C., Long, C., Rathjen, P. D., and Rathjen, J. (2010) ‘Response to BMP4 signalling during ES cell differentiation defines intermediates of the ectoderm lineage’, *Journal of Cell Science*, 123(10), pp. 1796–1804. doi: 10.1242/jcs.047530.

Herz, H. M., Garruss, A., and Shilatifard, A. (2013) ‘SET for life: Biochemical activities and biological functions of SET domain-containing proteins’, *Trends in Biochemical Sciences*, 38(12), pp. 621–639. doi: 10.1016/j.tibs.2013.09.004.

Hirai, H., Karian, P., and Kikyo, N. (2011) ‘Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor’, *Biochemical Journal*, 438(1), pp. 11–23. doi: 10.1042/BJ20102152.

Ho, L., Jothi, R., Ronan, J. L., Cui, K., Zhao, K., and Crabtree, G. R. (2009) ‘An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network’, *Proceedings of the National Academy of Sciences*, 106(13), pp. 5187–5191. doi: 10.1073/pnas.0812888106.

Hu, C., Fan, L., Cen, P., Chen, E., Jiang, Z., and Li, L. (2016) ‘Energy metabolism plays a critical role in stem cell maintenance and differentiation’, *International Journal of Molecular Sciences*. Edited by W. Bloch. MDPI, 17(2), p. 253. doi: 10.3390/ijms17020253.

Hunt, P. (1992) ‘Hox codes and positional specification in vertebrate embryonic axes’, *Annual Review of Cell and Developmental Biology*, 8(1), pp. 227–256. doi: 10.1146/annurev.cellbio.8.1.227.

Inman, G. J., Nicolás, F. J., Callahan, J. F., Harling, J. D., Gaster, L. M., Reith, A. D., Laping, N. J., and Hill, C. S. (2002) ‘SB-431542 is a potent and specific inhibitor of Transforming Growth Factor- β Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7’, *Molecular Pharmacology*, 62(1), pp. 65–74. doi: 10.1124/mol.62.1.65.

Ishitobi, H., Wakamatsu, A., Liu, F., Azami, T., Hamada, M., Matsumoto, K., Kataoka, H.,

Kobayashi, M., Choi, K., Nishikawa, S., Takahashi, S., and Ema, M. (2011) 'Molecular basis for Flk1 expression in hemato-cardiovascular progenitors in the mouse', *Development*, 138(24), pp. 5357 LP – 5368. doi: 10.1242/dev.065565.

Jiang, H., Shukla, A., Wang, X., Chen, W., Bernstein, B. E., and Roeder, R. G. (2011) 'Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains', *Cell*, 144(4), pp. 513–525. doi: 10.1016/j.cell.2011.01.020.

Juárez-Morales, J. L., Schulte, C. J., Pezoa, S. A., Vallejo, G. K., Hilinski, W. C., England, S. J., de Jager, S., and Lewis, K. E. (2016) 'Evx1 and Evx2 specify excitatory neurotransmitter fates and suppress inhibitory fates through a Pax2-independent mechanism', *Neural Development*. England, 11(1), p. 5. doi: 10.1186/s13064-016-0059-9.

Ke, Q. and Costa, M. (2006) 'Hypoxia-Inducible Factor-1 (HIF-1)', *Molecular Pharmacology*, 70(5), pp. 1469–1480. doi: 10.1124/mol.106.027029.

Kilberg, M. S., Terada, N., and Shan, J. (2016) 'Influence of amino acid metabolism on embryonic stem cell function and differentiation', *Advances in Nutrition: An International Review Journal*, 7(4), pp. 780S–789S. doi: 10.3945/an.115.011031.

Kim, E. J., Hori, K., Wyckoff, A., Dickel, L. K., Koundakjian, E. J., Goodrich, L. V., and Johnson, J. E. (2011) 'Spatiotemporal fate map of neurogenin1 (Neurog1) lineages in the mouse central nervous system', *The Journal of Comparative Neurology*. United States, 519(7), pp. 1355–1370. doi: 10.1002/cne.22574.

Kim, J. W., Tchernyshyov, I., Semenza, G. L., and Dang, C. V. (2006) 'HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia', *Cell Metabolism*, 3(3), pp. 177–185. doi: 10.1016/j.cmet.2006.02.002.

Kingston, R. E. and Tamkun, J. W. (2014) 'Transcriptional regulation by Trithorax-group proteins', *Cold Spring Harbor Perspectives in Biology*, 6(10), p. a019349. doi: 10.1101/cshperspect.a019349.

Knobloch, M. and Jessberger, S. (2017) 'Metabolism and neurogenesis', *Current Opinion in*

Neurobiology, 42, pp. 45–52. doi: 10.1016/j.conb.2016.11.006.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., and Winter, D. R. (2018) ‘A beginner’s guide to analysis of RNA sequencing data’, *American Journal of Respiratory Cell and Molecular Biology*, 59(2), pp. 145–157. doi: 10.1165/rcmb.2017-0430TR.

Krupinski, P., Chickarmane, V., and Peterson, C. (2011) ‘Simulating the mammalian blastocyst - molecular and mechanical interactions pattern the embryo’, *PLoS Computational Biology*, 7(5), pp. 1–11. doi: 10.1371/journal.pcbi.1001128.

Kumar, P., Tan, Y., and Cahan, P. (2017) ‘Understanding development and stem cells using single cell-based analyses of gene expression’, *Development*, 144, pp. 17–32. doi: 10.1242/dev.133058.

Lancaster, M. a, Renner, M., Martin, C., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., and Knoblich, J. A. (2013) ‘Cerebral organoids model human brain development and microcephaly’, *Nature*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 501(7467), pp. 373–379. doi: 10.1038/nature12517.

Lane, M. and Gardner, D. K. (1997) ‘Nonessential amino acids and glutamine decrease the time of the first three cleavage divisions and increase compaction of mouse zygotes in vitro’, *Journal of Assisted Reproduction and Genetics*, 14(7), pp. 398–403. doi: 10.1007/BF02766148.

Lang, K. J. D., Rathjen, J., Vassilieva, S., and Rathjen, P. D. (2004) ‘Differentiation of embryonic stem cells to a neural fate: A route to re-building the nervous system?’, *Journal of Neuroscience Research*. John Wiley & Sons, Ltd, 76(2), pp. 184–192. doi: 10.1002/jnr.20036.

Lau, P. N. I. and So, C. W. E. (2015) ‘Polycomb and Trithorax factors in transcriptional and epigenetic regulation’, in Huang, S., Litt, M. D., and Blakey, C. A. (eds) *Epigenetic Gene Expression and Regulation*. Elsevier Science & Technology, pp. 63–94. doi: 10.1016/B978-0-12-799958-6.00004-4.

Li, L., Liu, C., Biechele, S., Zhu, Q., Song, L., Lanner, F., Jing, N., and Rossant, J. (2013) ‘Location of transient ectodermal progenitor potential in mouse development’, *Development*, 140(22), pp. 4533–4543. doi: 10.1242/dev.092866.

Lonic, A. (2007) *Molecular mechanism of L-proline induced EPL-cell formation*. University of Adelaide.

Lu, R., Markowetz, F., Unwin, R. D., Leek, J. T., Airoidi, E. M., MacArthur, B. D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L. A., Troyanskaya, O. G., Whetton, A. D., and Lemischka, I. R. (2009) 'Systems-level dynamic analyses of fate change in murine embryonic stem cells', *Nature*, 462(7271), pp. 358–362. doi: 10.1038/nature08575.

Luo, X., van der Veer, B. K., Sun, L., Bartoccetti, M., Boretto, M., Vankelecom, H., Khoueiry, R., and Koh, K. P. (2020) 'Coordination of germ layer lineage choice by TET1 during primed pluripotency', *Genes and Development*, 34(7–8), pp. 598–618. doi: 10.1101/gad.329474.119.

Ma, Q., Chen, Z., Barrantes, I. del B., Luis de la Pompa, J., and Anderson, D. J. (1998) 'Neurogenin1 is essential for the determination of neuronal precursors for proximal cranial sensory ganglia', *Neuron*, 20(3), pp. 469–482. doi: [https://doi.org/10.1016/S0896-6273\(00\)80988-5](https://doi.org/10.1016/S0896-6273(00)80988-5).

MacArthur, B. D., Ma'ayan, A., and Lemischka, I. R. (2009) 'Systems biology of stem cell fate and cellular reprogramming', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(10), pp. 672–681. doi: 10.1038/nrm2766.

Martano, G., Delmotte, N., Kiefer, P., Christen, P., Kentner, D., Bumann, D., and Vorholt, J. A. (2015) 'Fast sampling method for mammalian cell metabolic analyses using liquid chromatography–mass spectrometry', *Nature Protocols*, 10(1), pp. 1–11. doi: 10.1038/nprot.2014.198.

Martano, G., Borroni, E. M., Lopci, E., Cattaneo, M. G., Mattioli, M., Bachi, A., Decimo, I., and Bifari, F. (2019) 'Metabolism of stem and progenitor cells: Proper methods to answer specific questions', *Frontiers in Molecular Neuroscience*, 12, p. 151. doi: 10.3389/fnmol.2019.00151.

Martynoga, B., Drechsel, D., and Guillemot, F. (2012) 'Molecular control of neurogenesis: A view from the mammalian cerebral cortex', *Cold Spring Harbor Perspectives in Biology*, 4(10), p. a008359. doi: 10.1101/cshperspect.a008359.

Maye, P., Becker, S., Siemen, H., Thorne, J., Byrd, N., Carpentino, J., and Grabel, L. (2004) 'Hedgehog signaling is required for the differentiation of ES cells into neurectoderm', *Developmental*

Biology, 265(1), pp. 276–290. doi: <https://doi.org/10.1016/j.ydbio.2003.09.027>.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., and Reik, W. (2017) ‘Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation’, *Cell Reports*, 20(5), pp. 1215–1228. doi: [10.1016/j.celrep.2017.07.009](https://doi.org/10.1016/j.celrep.2017.07.009).

Morey, L., Pascual, G., Cozzuto, L., Roma, G., Wutz, A., Benitah, S. A., and Di Croce, L. (2012) ‘Nonoverlapping functions of the polycomb group Cbx family of proteins in embryonic stem cells’, *Cell Stem Cell*, 10(1), pp. 47–62. doi: [10.1016/j.stem.2011.12.006](https://doi.org/10.1016/j.stem.2011.12.006).

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E. P., Bruneau, B. G., and Di Croce, L. (2015) ‘Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms’, *Cell Stem Cell*. Elsevier, 17(3), pp. 300–315. doi: [10.1016/j.stem.2015.08.009](https://doi.org/10.1016/j.stem.2015.08.009).

Morey, L. and Helin, K. (2010) ‘Polycomb group protein-mediated repression of transcription’, *Trends in Biochemical Sciences*, 35(6), pp. 323–332. doi: [10.1016/j.tibs.2010.02.009](https://doi.org/10.1016/j.tibs.2010.02.009).

Morris, M. B., Ozsoy, S., Zada, Matthew, Zada, Mark, Zamfirescu, R. C., Todorova, M. G., and Day, M. L. (2020) ‘Selected amino acids promote mouse pre-implantation embryo development in a growth factor-like manner’, *Frontiers in Physiology*, 11, p. 140. doi: [10.3389/fphys.2020.00140](https://doi.org/10.3389/fphys.2020.00140).

Nichols, J. and Smith, A. (2009) ‘Naive and primed pluripotent states’, *Cell Stem Cell*, 4(6), pp. 487–492. doi: [10.1016/j.stem.2009.05.015](https://doi.org/10.1016/j.stem.2009.05.015).

Nishimura, K., Fukuda, A., and Hisatake, K. (2019) ‘Mechanisms of the metabolic shift during somatic cell reprogramming’, *International Journal of Molecular Sciences*, 20(9), p. 2254. doi: [10.3390/ijms20092254](https://doi.org/10.3390/ijms20092254).

O’Brien, L. C., Keeney, P. M., and Bennett, J. P. (2015) ‘Differentiation of human neural stem cells into motor neurons stimulates mitochondrial biogenesis and decreases glycolytic flux’, *Stem Cells and Development*, 24(17), pp. 1984–1994. doi: [10.1089/scd.2015.0076](https://doi.org/10.1089/scd.2015.0076).

Oh, Y. and Jang, J. (2019) ‘Directed differentiation of pluripotent stem cells by transcription

factors', *Molecules and Cells*. The Korean Society for Molecular and Cellular Biology, 42(3), pp. 200–209. doi: 10.14348/molcells.2019.2439.

Ohnuki, Y. and Kurosawa, H. (2013) 'Effects of hanging drop culture conditions on embryoid body formation and neuronal cell differentiation using mouse embryonic stem cells: Optimization of culture conditions for the formation of well-controlled embryoid bodies', *Journal of Bioscience and Bioengineering*, 115(5), pp. 571–574. doi: 10.1016/j.jbiosc.2012.11.016.

Okazaki, K. and Maltepe, E. (2006) 'Oxygen, epigenetics and stem cell fate', *Regenerative Medicine*. England, 1(1), pp. 71–83. doi: 10.2217/17460751.1.1.71.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005) 'Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells', *Journal of Biological Chemistry*, 280(7), pp. 5307–5317. doi: 10.1074/jbc.M410015200.

Onoguchi, M., Hirabayashi, Y., Koseki, H., and Gotoh, Y. (2012) 'A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 109(42), pp. 16939–16944. doi: 10.1073/pnas.1202956109.

Pallag, G., Nazarian, S., Ravasz, D., Bui, D., Komlódi, T., Doerrier, C., Gnaiger, E., Seyfried, T. N., and Chinopoulos, C. (2022) 'Proline oxidation supports mitochondrial ATP production when Complex I is inhibited', *International Journal of Molecular Sciences*. Switzerland, 23(9), p. 5111. doi: 10.3390/ijms23095111.

Pandhare, J., Cooper, S. K., Donald, S. P., and Phang, J. M. (2009) 'Regulation and function of proline oxidase under nutrient stress', *Journal of Cellular biochemistry*, 107(4), pp. 759–768. doi: 10.1002/jcb.22174.

Panopoulos, A. D., Yanes, O., Ruiz, S., Kida, Y. S., Diep, D., Tautenhahn, R., Herreras, A., Batchelder, E. M., Plongthongkum, N., Lutz, M., Berggren, W. T., Zhang, K., Evans, R. M., Siuzdak, G., and Belmonte, J. C. I. (2012) 'The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming', *Cell Research*, 22(1), pp. 168–177. doi:

10.1038/cr.2011.177.

Pasini, D., Bracken, A. P., Hansen, J. B., Capillo, M., and Helin, K. (2007) 'The polycomb group protein Suz12 is required for embryonic stem cell differentiation', *Molecular and Cellular Biology*, 27(10), pp. 3769–3779. doi: 10.1128/MCB.01432-06.

Patriarca, E. J., Cermola, F., D'Aniello, C., Fico, A., Guardiola, O., De Cesare, D., and Minchiotti, G. (2021) 'The multifaceted roles of Proline in cell behavior', *Frontiers in Cell and Developmental Biology*, 9, pp. 1–28. doi: 10.3389/fcell.2021.728576.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., and Rathjen, P. D. (2002) 'Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development', *Journal of Cell Science*, 115(2), pp. 329–339. doi: 10.1073/pnas.95.9.5082.

Phang, J. M., Donald, S. P., Pandhare, J., and Liu, Y. (2008) 'The metabolism of proline, a stress substrate, modulates carcinogenic pathways', *Amino Acids*, 35(4), pp. 681–690. doi: 10.1007/s00726-008-0063-4.

Phang, J. M., Liu, W., Hancock, C. N., and Fischer, J. W. (2015) 'Proline metabolism and cancer: emerging links to glutamine and collagen', *Current Opinion in Clinical Nutrition and Metabolic Care*, 18(1), pp. 71–77. doi: 10.1097/MCO.0000000000000121.

Phang, J. M. (2019) 'Proline metabolism in cell regulation and cancer biology: Recent advances and hypotheses', *Antioxidants & Redox Signaling*. 2017/11/15. Mary Ann Liebert, Inc., publishers, 30(4), pp. 635–649. doi: 10.1089/ars.2017.7350.

Phang, J. M., Liu, W., and Zabinryk, O. (2010) 'Proline metabolism and microenvironmental stress', *Annual Review of Nutrition*, 30(1), pp. 441–463. doi: 10.1146/annurev.nutr.012809.104638.

Phang, J. M., Pandhare, J., and Liu, Y. (2008) 'The metabolism of proline as microenvironmental stress substrate', *The Journal of Nutrition*, 138(10), pp. 2008–2015. doi: 138/10S-I/2008S [pii].

Piunti, A. and Shilatifard, A. (2016) 'Epigenetic balance of gene expression by polycomb and COMPASS families', *Science*, 352(6290), p. aad9780. doi: 10.1126/science.aad9780.

Poudel, P. P., Bhattarai, C., Ghosh, A., and Kalthur, S. G. (2022) 'Role of engrailed homeobox 2 (EN2) gene in the development of the cerebellum and effects of its altered and ectopic expressions', *Egyptian Journal of Medical Human Genetics*, 23(1), p. 95. doi: 10.1186/s43042-022-00307-8.

Prigione, A., Rohwer, N., Hoffmann, S., Mlody, B., Drews, K., Bukowiecki, R., Blümlein, K., Wanker, E. E., Ralser, M., Cramer, T., and Adjaye, J. (2014) 'HIF1 α modulates cell fate reprogramming through early glycolytic shift and upregulation of PDK1–3 and PKM2', *Stem Cells*. John Wiley & Sons, Ltd, 32(2), pp. 364–376. doi: 10.1002/stem.1552.

Purnell, C. (2018) *The role of anaerobic metabolism in the L-Proline mediated differentiation of mouse embryonic stem cells*. University of Sydney.

Qi, X., Li, T.-G., Hao, J., Hu, J., Wang, J., Simmons, H., Miura, S., Mishina, Y., and Zhao, G.-Q. (2004) 'BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 101(16), pp. 6027–6032. doi: 10.1073/pnas.0401367101.

Rao, C., Malaguti, M., Mason, J. O., and Lowell, S. (2020) 'The transcription factor E2A drives neural differentiation in pluripotent cells', *Development*, 147(12), p. dev184093. doi: 10.1242/dev.184093.

Rathjen, J., Lake, J.-A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999) 'Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors', *Journal of Cell Science*, 112(5), pp. 601–612. doi: 10.1242/jcs.004613.

Rathjen, J., Haines, B. P., Hudson, K. M., Nesci, A., Dunn, S., and Rathjen, P. D. (2002) 'Directed differentiation of pluripotent cells to neural lineages: homogeneous formation and differentiation of a neurectoderm population', *Development*, 129(11), pp. 2649–2661. Available at: <http://dev.biologists.org/content/129/11/2649>.

Rathjen, J., Washington, J. M., Bettess, M. D., and Rathjen, P. D. (2003) 'Identification of a biological activity that supports maintenance and proliferation of pluripotent cells from the primitive ectoderm of the mouse', *Biology of Reproduction*, 69(6), pp. 1863–1871. doi:

10.1095/biolreprod.103.017384.

Rehman, J. (2010) 'Empowering self-renewal and differentiation: the role of mitochondria in stem cells', *Journal of Molecular Medicine*, 88(10), pp. 981–986. doi: 10.1007/s00109-010-0678-2.

Ringrose, L. and Paro, R. (2004) 'Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins', *Annual Review of Genetics*. Annual Reviews, 38(1), pp. 413–443. doi: 10.1146/annurev.genet.38.072902.091907.

Ringrose, L. and Paro, R. (2007) 'Polycomb/Trithorax response elements and epigenetic memory of cell identity', *Development*. The Company of Biologists Ltd, 134(2), pp. 223–232. doi: 10.1242/dev.02723.

Rodda, S. J., Kavanagh, S. J., Rathjen, J., and Rathjen, P. D. (2002) 'Embryonic stem cell differentiation and the analysis of mammalian development', *The International Journal of Developmental Biology*, 46(4), pp. 449–458. doi: 10.1387/IJDB.12141431.

Ryu, J. M., Lee, S. H., Seong, J. K., and Han, H. J. (2015) 'Glutamine contributes to maintenance of mouse embryonic stem cell self-renewal through PKC-dependent downregulation of HDAC1 and DNMT1/3a', *Cell Cycle*. Taylor & Francis, 14(20), pp. 3292–3305. doi: 10.1080/15384101.2015.1087620.

Ryu, J. M. and Han, H. J. (2011) 'L-Threonine regulates G1/S phase transition of mouse embryonic stem cells via PI3K/Akt, MAPKs, and mTORC pathways', *Journal of Biological Chemistry*, 286(27), pp. 23667–23678. doi: 10.1074/jbc.M110.216283.

Santanach, A., Blanco, E., Jiang, H., Molloy, K. R., Sansó, M., LaCava, J., Morey, L., and Di Croce, L. (2017) 'The polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity', *Nature Communications*, 8(1), p. 1235. doi: 10.1038/s41467-017-01464-w.

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., Huang, K., Liu, Y., Wang, X., Chen, Q., Zhang, J., Shang, B., Li, S., Shi, X., Liao, B., Zhang, C., Lai, K., Zhong, X., Shu, X., Wang, J., Yao, H., Chen, J., Pei, D., and Pan, G. (2017) 'PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve ESCs', *Nature Communications*,

8, p. 672. doi: 10.1038/s41467-017-00668-4.

Shim, W. J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., Peng, G., Jing, N., Wang, Y., Kellis, M., Tam, P. P. L., Smith, A., Piper, M., Christiaen, L., Nguyen, Q., Bodén, M., and Palpant, N. J. (2020) ‘Conserved epigenetic regulatory logic infers genes governing cell identity’, *Cell Systems*, 11(6), pp. 625–639. doi: <https://doi.org/10.1016/j.cels.2020.11.001>.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019a) ‘Embryoid body differentiation of mouse embryonic stem cells into neurectoderm and neural progenitors’, in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019b) ‘Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells’, *Frontiers in Physiology*, 10, p. 705. doi: 10.3389/fphys.2019.00705.

Shyh-Chang, Ng, Locasale, J. W., Lyssiotis, C. A., Zheng, Y., Teo, R. Y., Ratanasirintrawoot, S., Zhang, J., Onder, T., Unternaehrer, J. J., Zhu, H., Asara, J. M., Daley, G. Q., and Cantley, L. C. (2013) ‘Influence of threonine metabolism on S-adenosylmethionine and histone methylation’, *Science*, 339(6116), p. 222. doi: 10.1126/science.1226603.

Shyh-Chang, N., Daley, G. Q., and Cantley, L. C. (2013) ‘Stem cell metabolism in tissue development and aging’, *Development*, 140(12), pp. 2535–2547. doi: 10.1242/dev.091777.

Southall, S. M., Wong, P.-S., Odho, Z., Roe, S. M., and Wilson, J. R. (2009) ‘Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks’, *Molecular Cell*, 33(2), pp. 181–191. doi: <https://doi.org/10.1016/j.molcel.2008.12.029>.

Spangler, A., Su, E. Y., Craft, A. M., and Cahan, P. (2018) ‘A single cell transcriptional portrait of embryoid body differentiation and comparison to progenitors of the developing embryo’, *Stem Cell*

Research, 31, pp. 201–215. doi: 10.1016/j.scr.2018.07.022.

Stavridis, M. P., Collins, B. J., and Storey, K. G. (2010) ‘Retinoic acid orchestrates fibroblast growth factor signalling to drive embryonic stem cell differentiation’, *Development*, 137(6), pp. 881–890. doi: 10.1242/dev.043117.

Stavridis, M. P. and Smith, A. G. (2003) ‘Neural differentiation of mouse embryonic stem cells’, *Biochemical Society Transactions*, 31(1), pp. 45–49. doi: 10.1042/bst0310045.

Sun, Y., Shim, W. J., Shen, S., Sinniah, E., Pham, D., Su, Z., Mizikovsky, D., White, M. D., Ho, J. W. K., Nguyen, Q., Bodén, M., and Palpant, N. J. (2023) ‘Inferring cell diversity in single cell data using consortium-scale epigenetic data as a biological anchor for cell identity’, *Nucleic Acids Research*, p. gkad307. doi: 10.1093/nar/gkad307.

Suter, D. M., Tirefort, D., Julien, S., and Krause, K.-H. (2009) ‘A Sox1 to Pax6 switch drives neuroectoderm to radial glia progression during differentiation of mouse embryonic stem cells’, *Stem Cells*, 27(1), pp. 49–58. doi: 10.1634/stemcells.2008-0319.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. von (2019) ‘STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets’, *Nucleic Acids Research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Tam, P. P. L. and Behringer, R. R. (1997) ‘Mouse gastrulation: The formation of a mammalian body plan’, *Mechanisms of Development*, 68(1–2), pp. 3–25. doi: 10.1016/S0925-4773(97)00123-8.

Tam, P. P. L. and Ho, J. W. K. (2020) ‘Cellular diversity and lineage trajectory: Insights from mouse single cell transcriptomes’, *Development*, 147(2), pp. 1–7. doi: 10.1242/dev.179788.

Tan, B. S. N., Lonic, A., Morris, M. B., Rathjen, P. D., and Rathjen, J. (2011) ‘The amino acid transporter SNAT2 mediates L-proline-induced differentiation of ES cells’, *American Journal of Physiology-Cell Physiology*, 300(6), pp. 1270–1279. doi: 10.1152/ajpcell.00235.2010.

Telley, L., Govindan, S., Prados, J., Stevant, I., Nef, S., Dermitzakis, E., Dayer, A., and Jabaudon,

D. (2016) ‘Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex’, *Science*, 351(6280), pp. 1443–1446. doi: 10.1126/science.aad8361.

Tonge, P. D. and Andrews, P. W. (2010) ‘Retinoic acid directs neuronal differentiation of human pluripotent stem cell lines in a non-cell-autonomous manner’, *Differentiation*, 80(1), pp. 20–30. doi: <https://doi.org/10.1016/j.diff.2010.04.001>.

Tsai, P.-H., Chien, Y., Wang, M.-L., Hsu, C.-H., Laurent, B., Chou, S.-J., Chang, W.-C., Chien, C.-S., Li, H.-Y., Lee, H.-C., Huo, T.-I., Hung, J.-H., Chen, C.-H., and Chiou, S.-H. (2019) ‘Ash2l interacts with Oct4-stemness circuitry to promote super-enhancer-driven pluripotency network’, *Nucleic Acids Research*, 47(19), pp. 10115–10133. doi: 10.1093/nar/gkz801.

Tsogtbaatar, E., Landin, C., Minter-Dykhouse, K., and Folmes, C. D. L. (2020) ‘Energy metabolism regulates stem cell pluripotency’, *Frontiers in Cell and Developmental Biology*, 8, p. 87. doi: 10.3389/fcell.2020.00087.

Varum, S., Rodrigues, A. S., Moura, M. B., Momcilovic, O., Easley IV, C. A., Ramalho-Santos, J., van Houten, B., and Schatten, G. (2011) ‘Energy metabolism in human pluripotent stem cells and their differentiated counterparts’, *PLoS ONE*, 6(6), pp. 1–15. doi: 10.1371/journal.pone.0020914.

Wade, S. L., Langer, L. F., Ward, J. M., and Archer, T. K. (2015) ‘MiRNA-mediated regulation of the SWI/SNF chromatin remodeling complex controls pluripotency and endodermal differentiation in human eSCs’, *Stem Cells*, 33(10), pp. 2925–2935. doi: 10.1002/stem.2084.

Wan, M., Liang, J., Xiong, Y., Shi, F., Zhang, Y., Lu, W., He, Q., Yang, D., Chen, R., Liu, D., Barton, M., and Songyang, Z. (2013) ‘The Trithorax group protein Ash2l is essential for pluripotency and maintaining open chromatin in embryonic stem cells’, *Journal of Biological Chemistry*, 288(7), pp. 5039–5048. doi: 10.1074/jbc.M112.424515.

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T. S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016) ‘Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition’, *Proceedings of the National Academy of Sciences*, 113(42), pp. 11871–11876. doi: 10.1073/pnas.1606857113.

Wang, J., Alexander, P., Wu, L., Hammer, R., Cleaver, O., and McKnight, S. L. (2009)

‘Dependence of mouse embryonic stem cells on Threonine catabolism’, *Science*, 325(5939), pp. 435–439. doi: 10.1126/science.1173288.

Wang, J., Alexander, P., and McKnight, S. L. (2011) ‘Metabolic specialization of mouse embryonic stem cells’, *Cold Spring Harbor Symposia on Quantitative Biology*, 76, pp. 183–193. doi: 10.1101/sqb.2011.76.010835.

Washington, J. M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M. D., Hamra, N., Semendric, L., Tan, B. S. N., Lake, J. A., Keough, R. A., Morris, M. B., and Rathjen, P. D. (2010) ‘L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture’, *American Journal of Physiology-Cell Physiology*, 298(5), pp. C982–C992. doi: 10.1152/ajpcell.00498.2009.

Wu, M., Wang, P. F., Lee, J. S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008) ‘Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS’, *Molecular and Cellular Biology*, 28(24), pp. 7337–7344. doi: 10.1128/MCB.00976-08.

Xia, Y., Nivet, E., Sancho-Martinez, I., Gallegos, T., Suzuki, K., Okamura, D., Wu, M.-Z., Dubova, I., Esteban, C. R., Montserrat, N., Campistol, J. M., and Belmonte, J. C. I. (2013) ‘Directed differentiation of human pluripotent cells to ureteric bud kidney progenitor-like cells’, *Nature Cell Biology*, 15(12), pp. 1507–1515. doi: 10.1038/ncb2872.

Xia, Y., Sancho-Martinez, I., Nivet, E., Rodriguez Esteban, C., Campistol, J. M., and Izpisua Belmonte, J. C. (2014) ‘The generation of kidney organoids by differentiation of human pluripotent cells to ureteric bud progenitor-like cells’, *Nat. Protocols*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 9(11), pp. 2693–2704. doi: 10.1038/nprot.2014.182.

Xue, K., Ng, J.-H., and Ng, H.-H. (2011) ‘Mapping the networks for pluripotency’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1575), pp. 2238–2246. doi: 10.1098/rstb.2011.0005.

Yao, M., Zhou, X., Zhou, J., Gong, S., Hu, G., Li, J., Huang, K., Lai, P., Shi, G., Hutchins, A. P., Sun, H., Wang, H., and Yao, H. (2018) 'PCGF5 is required for neural differentiation of embryonic stem cells', *Nature Communications*, 9(1), p. 1463. doi: 10.1038/s41467-018-03781-0.

Ying, Q.-L., Nichols, J., Chambers, I., and Smith, A. (2003) 'BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3', *Cell*, 115, pp. 281–292. doi: 10.1016/S0092-8674(03)00847-X.

Ying, Q.-L., Stavridis, M., Griffiths, D., Li, M., and Smith, A. (2003) 'Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture', *Nature Biotechnology*. Nature Publishing Group, 21(2), pp. 183–186. Available at: <http://dx.doi.org/10.1038/nbt780>.

Zeng, Z., Miao, N., and Sun, T. (2018) 'Revealing cellular and molecular complexity of the central nervous system using single cell sequencing', *Stem Cell Research & Therapy*, 9(1), p. 234. doi: 10.1186/s13287-018-0985-z.

Zheng, X., Boyer, L., Jin, M., Mertens, J., Kim, Y., Ma, L., Ma, L., Hamm, M., Gage, F. H., and Hunter, T. (2016) 'Metabolic reprogramming during neuronal differentiation from aerobic glycolysis to neuronal oxidative phosphorylation', *eLife*, 5, p. e13374. doi: 10.7554/eLife.13374.

Chapter 5: Modelling early neural developmental processes to determine contributors to onset of motor neurone disease post-partum

5.1. Introduction

Motor neurone disease (MND) is a family of diseases commonly caused by degeneration of neurons, particularly motor neurons. There are two types of MNDs: Hereditary and sporadic (Xu *et al.*, 2020; McCluskey *et al.*, 2022; Park *et al.*, 2022). Although the onset of hereditary MND has been studied to an extent, the onset of sporadic cases is not as well understood. In other known sporadic diseases, which include coronary heart disease, stroke, diabetes and hypertension, adverse impacts during early development, especially *in utero*, can result in changes in physiology and metabolism, which result in increased susceptibility of disease later in life (De Boo and Harding, 2006; Heindel and Vandenberg, 2015). This is known as the “Developmental Origins of Health and Disease” Hypothesis. Based on this hypothesis, onset of sporadic MND may occur during early development of the nervous system *in utero* (Ferreira *et al.*, 2016). However, the potential links between early neural development and the onset of sporadic MND has not been well established, nor have the mechanisms been studied.

MND covers a wide range of diseases across age groups, from paediatric spinal muscular dystrophy to the late onset of amyotrophic lateral sclerosis (ALS). Thus, the pathophysiology of MND is complex and multifactorial (Sterneckert *et al.*, 2014; Teoh *et al.*, 2017; Karagiannis and Inoue, 2020) and multiple genes are associated with the onset of the disease (Teoh *et al.*, 2017). For instance, *C9orf72*-associated onset of MND is caused by the hexanucleotide repeat expansion in the gene’s non-coding region (Sterneckert *et al.*, 2014; Black *et al.*, 2017; Hao *et al.*, 2019; Karagiannis and Inoue, 2020). The expansion causes a decrease in *C9orf72* expression, and this abnormal expression correlates with age-related neuroinflammation in mouse models, as these models showed defects in lysosomal trafficking and immune responses in microglia (O’Rourke *et al.*, 2016; Karagiannis and Inoue, 2020). However, the mutation in *C9orf72* alone is insufficient to trigger onset of neurodegeneration or defects in motor function in MND (Koppers *et al.*, 2015). In human-derived cell models, *C9ORF72* repeat expansions cause excitotoxicity in motor neurons due to an increase in glutamate receptors and suppression

of autophagy as toxic gain- and loss-of function mechanisms, respectively (Selvaraj *et al.*, 2018; Shi *et al.*, 2018; Boivin *et al.*, 2020; Karagiannis and Inoue, 2020). The *C9orf72* gene also regulates neural development in zebrafish embryos via Cyclin G1 activity (Yeh *et al.*, 2018). The C9ORF72 protein, derived from the *C9ORF72* gene, undergoes changes in expression and localisation during prenatal (E18) and up to 56 days postnatal transgenic mouse development (Atkinson *et al.*, 2015).

Progranulin (Grn) is a growth factor that regulates development, tissue remodelling, wound repair and inflammation (Bateman and Bennett, 2009). In zebrafish embryos, Grn enhances neuronal survival and outgrowth both *in vitro* and *in vivo* (De Muynck and Van Damme, 2011). *Null* (or premature STOP codon reads) and missense mutations of the *Grn* gene are associated with neurodegeneration. However, more cases are seen in dementia rather than ALS (Hardy and Rogaeva, 2014; Sternecker *et al.*, 2014; Karch *et al.*, 2016). The mutation decreases production and secretion of the Grn protein, which increases sensitivity of cells to stress (Almeida *et al.*, 2012; Sternecker *et al.*, 2014). A major source of the Grn protein is found in microglia where, after neuronal injury, activated microglia show an increase in the level of this protein (Bateman and Bennett, 2009; Hu *et al.*, 2010; Hardy and Rogaeva, 2014). This is a biomarker of microglial response with progression of MND (Philips *et al.*, 2010).

Mutations of *Fus* and *Tardbp* are also markers of MND (Liu *et al.*, 2017; Harley *et al.*, 2021). *Fus* and *Tardbp* encode for RNA-binding proteins Fus and Tdp-43, respectively. They are involved in RNA transcription, splicing, RNA localisation, function and decay. Thus, they share processes contributing to RNA metabolism. Mutations in these genes result in the proteins aggregating and mislocalising to the cytoplasm from the nucleus (Chen-Plotkin *et al.*, 2010; Ichiyanagi *et al.*, 2016). RNA processing and long non-coding RNAs (lncRNAs) are crucial for nervous system development. Thus, compromising these processes commonly leads to neuronal cell death (Liu *et al.*, 2017; Biscarini *et al.*, 2018; Harley *et al.*, 2021). In embryos, knockout of either *Fus* or *Tardbp* in mouse models is lethal (Sephton *et al.*, 2010; Scekcic-Zahirovic *et al.*,

2016). Mislocalisation of Fus protein causes perinatal lethality, as well as a toxic gain-of-function trigger to motor neuron death in mice models (Scekic-Zahirovic *et al.*, 2016). Knockout of *Tardbp* in mice results in death between E3.5-8.5, and mouse blastocysts grown *in vitro* display an abnormally large inner cell mass (Kraemer *et al.*, 2010; Sephton *et al.*, 2010). However, partial knockout, or heterozygous mutant *Tardbp* mice, only display motor deficiencies without loss of motor neurons (Kraemer *et al.*, 2010).

Missense mutations in *Vcp* also contributes to onset of ALS (Yeo and Yu, 2016). VCP is a ubiquitously expressed protein that is involved in various cellular functions, including protein degradation via the ubiquitin-proteasome system and cellular homeostasis via apoptosis and autophagy (Yeo and Yu, 2016). However, VCP-mediated ALS is commonly associated with familial cases (Abramzon *et al.*, 2012; Gonzalez-Perez *et al.*, 2012). Occasional sporadic cases that are *VCP* mutation-mediated have occurred, and contribute to the accumulation of cytoplasmic TDP-43 that leads to neuronal degeneration (Yeo and Yu, 2016; Harley *et al.*, 2021). In mouse embryo development, deletion of the *Vcp* gene is embryonically lethal (Müller *et al.*, 2007).

NIMA-related kinase 1 (Nek1) protein is a mitotic checkpoint control protein (Chen *et al.*, 2008; Higelin *et al.*, 2018). Nek1 is required for activation of checkpoint kinases 1/2, which mediate proper cell cycle arrest and efficient DNA repair after DNA damage at the G1/S, S- or G2/M-phases (Chen *et al.*, 2008, 2011; Higelin *et al.*, 2018). In ALS cases, heterozygous loss-of-function mutations in *NEK1* compromise of DNA repair after DNA damage and increases cell death of motor neurons (Higelin *et al.*, 2018). In mouse embryo models, *Nek1* itself is not required during prenatal development of mice (Chen *et al.*, 2011). However, as Nek1 protein is essential for maintenance of chromosome stability and control, cells without functional Nek1 result in tumour generation due to unstable cell cycles. This leads to malignant tumourgenesis in postnatal mice (Chen *et al.*, 2008, 2011).

Smn proteins are involved in protein homeostasis, as well as other housekeeping processes, which include mRNA metabolism, localisation and trafficking, cytoskeletal dynamics, endocytosis, autophagy, mitochondrial and bioenergetic pathways. Compromise in these processes have detrimental effects on cellular physiology (Chaytow *et al.*, 2018). For instance, spinal muscular atrophy, a disease under the MND family, is caused by loss of SMN protein due to loss-of-function mutations in *SMN1* and retention of SMN2 protein (Lefebvre *et al.*, 1995; Dominguez *et al.*, 2017). This leads to degeneration of motor neurons in the spinal cord, causing paralysis with muscular atrophy. In mouse models, complete absence of Smn proteins is embryonically lethal (Schrack *et al.*, 1997). Smn1-deficient mice develop normally up to the morula stage but then degenerate and die prior to implantation (Schrack *et al.*, 1997; Sephton *et al.*, 2010). Thus, the *Smn1* gene and its products are essential for cellular survival and function during preimplantation development and has an essential role in regulating programmed cell death (Schrack *et al.*, 1997). SMN proteins also interact with FUS proteins in pre-mRNA splicing (Mirra *et al.*, 2017; Biscarini *et al.*, 2018). Mutation in *SMN* leads to depletion of SMN protein, resulting in disrupted coordination with FUS proteins for pre-mRNA splicing (Mirra *et al.*, 2017; Biscarini *et al.*, 2018). This has been associated with spinal muscular atrophy and other cases of MND (Mirra *et al.*, 2017).

Mutations in the gene encoding for the antioxidant superoxide dismutase 1 (*Sod1*) gene account for 20% of ALS cases (Ferreira *et al.*, 2016). They cause defective protein aggregation, leading to altered neurofilament levels, increased apoptosis, and altered neuronal and mitochondrial morphology (Ferreira *et al.*, 2016; Hayes *et al.*, 2019). Mitochondrial function, redox and reactive oxygen species homeostasis and Ca^{2+} homeostasis are compromised, which all contribute to oxidative stress in motor neurons in the lumbar and thoracic regions of the spinal cord (Ferreira *et al.*, 2016; Hayes *et al.*, 2019). Sod1 helps control redox potential and homeostatic levels of reactive oxygen species, which are essential for preimplantation embryo

development. Increasing oxidative stress compromises preimplantation development to the blastocyst stage and organogenesis in embryos (Yon *et al.*, 2008; Hardy *et al.*, 2021).

A proline-to-serine substitution at position-56 of vesicle-associated membrane protein-associated protein B (Vapb) contributes to the pathogenesis of sporadic and Sod1-linked MND (Teuling *et al.*, 2007; Kabashi *et al.*, 2013; Qiu *et al.*, 2013; Borgese *et al.*, 2021). *Vapb* encodes for the Vapb protein, and is highly conserved and ubiquitously expressed protein involved in vesicle transport and endoplasmic reticulum anchoring (Teuling *et al.*, 2007). The protein is highly abundant in motor neurons (Teuling *et al.*, 2007). The proline-to-serine substitution results in loss of function and vesicle aggregation in the spinal cord, leading to a reduction of endoplasmic reticulum anchoring of lipid-binding proteins which contributes to motor neuron degeneration (Teuling *et al.*, 2007; Kabashi *et al.*, 2013; Qiu *et al.*, 2013). Compromise of Vapb expression is not embryo lethal: Partial and complete deletion of *Vapb* in animal models still allows them to survive into adulthood (Kabashi *et al.*, 2013; Borgese *et al.*, 2021).

Optineurin (Optn) is also an MND-associated gene (Akizuki *et al.*, 2013; Toth and Atkin, 2018). This gene encodes for the ubiquitously expressed protein Optn, which is involved in neuroinflammation, Golgi maintenance, vesicle trafficking, mitophagy and autophagy (Sundaramoorthy *et al.*, 2015; Toth and Atkin, 2018). Optn interacts with TANK-binding kinase 1 (Tbk1) protein, which is also associated with MND (Niimi *et al.*, 2004; Ying and Yue, 2012). Optn and Tbk1 act as activator and inhibitor, respectively, of the NF- κ B pathway, which is directly involved in CNS function (Bonnard *et al.*, 2000; Delhase *et al.*, 2012; Akizuki *et al.*, 2013; Källstig *et al.*, 2021). Missense mutations of these genes lead to suboptimal functioning in Golgi maintenance, vesicle trafficking, mitophagy and autophagy, which compromise the function of long cell types like motor neurons (Sundaramoorthy *et al.*, 2015; Ito *et al.*, 2016; Harding *et al.*, 2021). Mice deficient in Tbk1 die during embryo development due to liver apoptosis (Bonnard *et al.*, 2000; Delhase *et al.*, 2012). Optn defects in mice develop normally without signs of organ abnormality (Slowicka *et al.*, 2016), but demonstrate suboptimal

autophagy (Sirohi *et al.*, 2013) and mislocalisation of Tdp-43 protein in spinal cord motor neurons (Kurashige *et al.*, 2021).

As stated above, MND is associated with a diverse range of genes and proteins which contribute to onset (Sterneckert *et al.*, 2014; Teoh *et al.*, 2017; Karagiannis and Inoue, 2020). Although late onset is common, onset can be triggered anywhere from birth to adulthood (Teoh *et al.*, 2017). Whether MND-associated genes and proteins are expressed in earlier stages of neuronal development, such as during pluripotent states and neural lineage commitment, has not been examined.

The purpose of this Chapter is to quantify the expression profile of various MND-associated genes and proteins across the pluripotency spectrum of mESCs and during their commitment to neural fate using a combination of qPCR and western blot analysis (Shparberg *et al.*, 2019a; Glover *et al.*, 2022). We also use our RNAseq data sets covering the pluripotency spectrum and microarray data sets covering directed differentiation to neural cell types with MEDII together with TRIAGE analysis to gain further insight (Figure 3.2, 4.1A and 4.1B). A published single-cell RNAseq dataset representing early gastrulation (Mohammed *et al.*, 2017) will also be interrogated for the expression of MND-associated genes, together with TRIAGE analysis.

5.2. Hypotheses and aims

In this Chapter, the following hypotheses will be tested:

1. MND-associated genes will be expressed during states of pluripotency, early gastrulation *in vivo*, and neural-fate commitment of mESCs, as shown by bioinformatics analysis of RNAseq and microarray data.
2. TRIAGE analysis of MND-associated genes will provide candidates for those genes involved in cell identity and cell-fate decisions.
3. L-Proline-mediated differentiation of mESCs will result in changes in expression of MND-associated genes and proteins, as shown by qPCR and Western blot analysis, respectively.

5.3. Results

5.3.1. *Relative expression and TRIAGE analysis of MND-associated genes across the pluripotency spectrum*

Cells must transition across the pluripotency spectrum to prime themselves for the appropriate lineage fate. mESCs transition to EPLCs in the presence of L-Proline, which are primed towards neural fate for the basic set up of the nervous system (Glover, 2018; Shparberg, 2018). Consistent with this, transcriptional/gene expression profiles and GO terms show expression of markers of neurectoderm at the EPLC stages and beyond, such as *Sox1*, *Pax6* and *Nestin* (Chapter 3 and 4) (Glover, 2018; Shparberg, 2018) and GO terms associated with nervous system development (Chapter 4) (Glover, 2018). In contrast, EpiSCs are primed towards mesendoderm (Bernemann *et al.*, 2011) and express definitive mesoderm markers such as *Brachyury/T* and *Mixl1*, definitive endoderm markers such as *Gata6* and *Sox17* and have GO terms associated with mesendoderm lineage commitment such as skeletal and digestive system development (Brons *et al.*, 2007; Tesar *et al.*, 2007; Bernemann *et al.*, 2011; Kojima *et al.*, 2014; Glover, 2018). What is the expression of MND-associated genes in these scenarios?

For this Chapter, we used differential gene expression analysis to identify transcriptional profiles of MND-associated genes across the pluripotency spectrum. To limit FDRs and restrict analysis of genes by just their relative changes, we denoted significance as having a Benjamini-Hochberg adjusted $P < 0.05$ regardless of $\log_2$ fold change (Koch *et al.*, 2018). Concurrently, we TRIAGE-transformed the same genes to produce DS and ranked them in descending order of DS across the entire mouse genome; i.e., highest DS is ranked first. A DS rank ≤ 1359 indicates a gene likely associated with cell fate and cell identity (Shim *et al.*, 2020). Beyond this rank a gene is likely associated with housekeeping and structural maintenance (Shim *et al.*, 2020).

Figure 5.1 shows changes in MND-associated gene expression across the pluripotency spectrum from ground-state mESCs to primed-state EpiSCs in the D3 and A2 cell lines. All MND-associated genes are expressed even at these early stages of development. The expression profile of many of them also change over the course of pluripotency. E.g., the expression of *Grn* is higher in the ground state, EPLCs and EpiSCs than in the naïve state, and *Nek1* expression increases across the pluripotency spectrum.

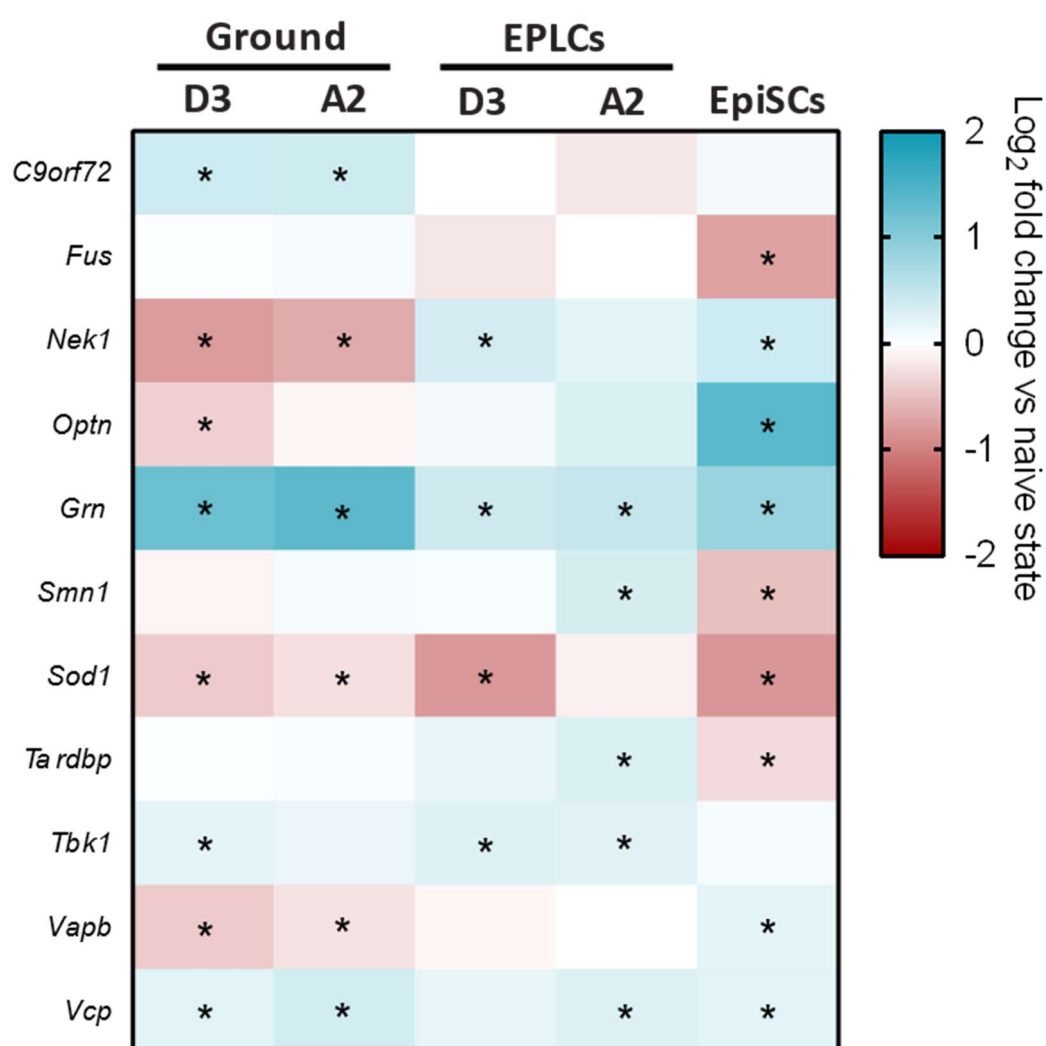


Figure 5.1 Relative expression of MND-associated genes across the pluripotency spectrum. D3 and A2 mESCs were used to produce naïve-state mESCs, ground-state mESCs and EPLCs, as per Figure 3.2. A2 primed-state EpiSCs were maintained, as per Figure 3.2. Samples ($n = 3$) were subjected to RNAseq and raw counts were acquired. MND-associated genes were selected for differential expression analysis relative to naïve-state mESCs within the same cell line. * Indicates significant change in expression based on a Benjamini-Hochberg adjusted $P < 0.05$.

We then used TRIAGE to transform raw counts from the RNAseq dataset to generate DS. Figure 5.2 shows DS ranked across the entire mouse genome. All MND-associated genes have a DS rank >1359 and therefore more likely to be associated with housekeeping and structural maintenance (Shim *et al.*, 2020).

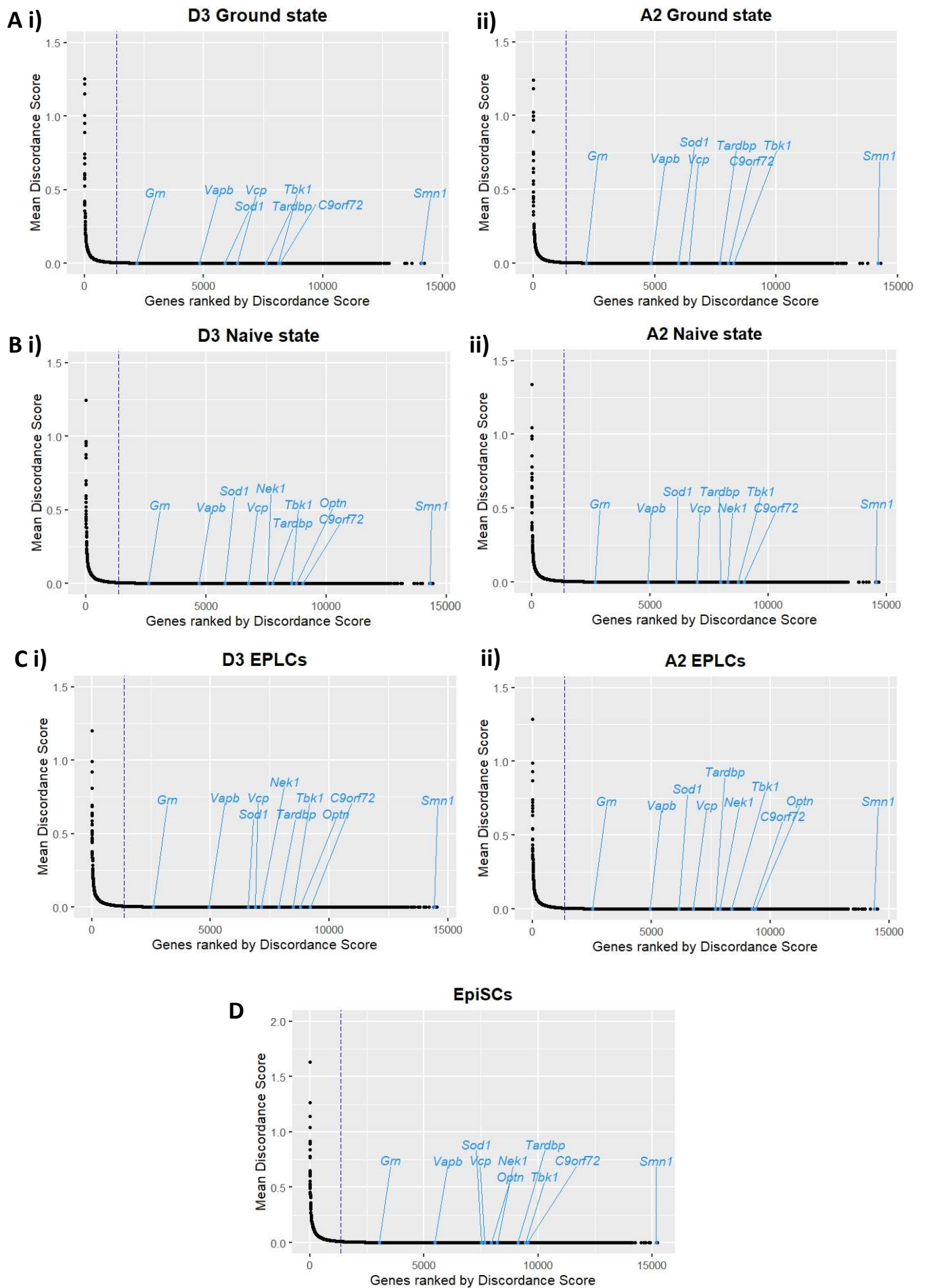


Figure 5.2 TRIAGE transformation of MND-associated genes across the pluripotency spectrum. Mean DS from the TRIAGE transformation were calculated from the results obtained for the three samples of **A.** ground state, **B.** naïve state and **C.** EPLCs from both **i)** D3 and **ii)** A2 cell lines, and **D.** A2 EpiSCs subjected to RNAseq analysis. Mean DS were plotted against the ranked DS values for all genes expressed in each sample, with DS rank threshold (1359) shown by the dotted blue line. MND-associated genes are labelled, with blue genes indicating a DS rank > 1359.

5.3.2. Interrogation of MND-associated genes in relevant cell populations across early gastrulation

Primitive ectoderm is the cellular substrate for gastrulation in which pluripotency is lost and the 3 multipotent germ layers are formed (Rathjen *et al.*, 1999). Therefore, we analysed the expression of MND-associated genes across relevant cell populations during early gastrulation of the mouse embryo, using a published single-cell RNAseq dataset (accession number GEO: GSE100597) (Mohammed *et al.*, 2017).

Figure 5.3A shows the cell clusters we have identified using Uniform Manifold Approximation and Projection (UMAP) clustering analysis in mouse embryos. The cell types consist of a range of pluripotent populations (E3.5 ICM, E4.5, 5.5, and 6.5 Epiblast, and E6.5 primitive streak) and two extraembryonic populations (E4.5 primitive endoderm and E6.5 visceral endoderm). UMAP analysis is used here instead of PCA due to larger number of samples and cell types. The markers used to identify each cell type are those used by (Mohammed *et al.*, 2017).

Figure 5.3B shows the same UMAP plots mapped to the expression of MND-associated genes. *Sod1*, *Smn1*, *Tardbp*, *Tbk1*, *Vapb* and *Vcp* are expressed in all cell populations, although the level of expression can vary. E.g., *Sod1* is expressed more highly in E4.4 and E6.5 epiblast cells and extraembryonic E6.5 visceral endoderm than in other cell types. Some genes are expressed only in certain cell types. E.g., *Optn* is highly expressed in E4.5 epiblast, more lowly expressed in E5.5 epiblast and E3.5 ICM, and not expressed in the other cell types. *C9orf72* is not expressed abundantly in any of the cell populations (Figure 5.3B).

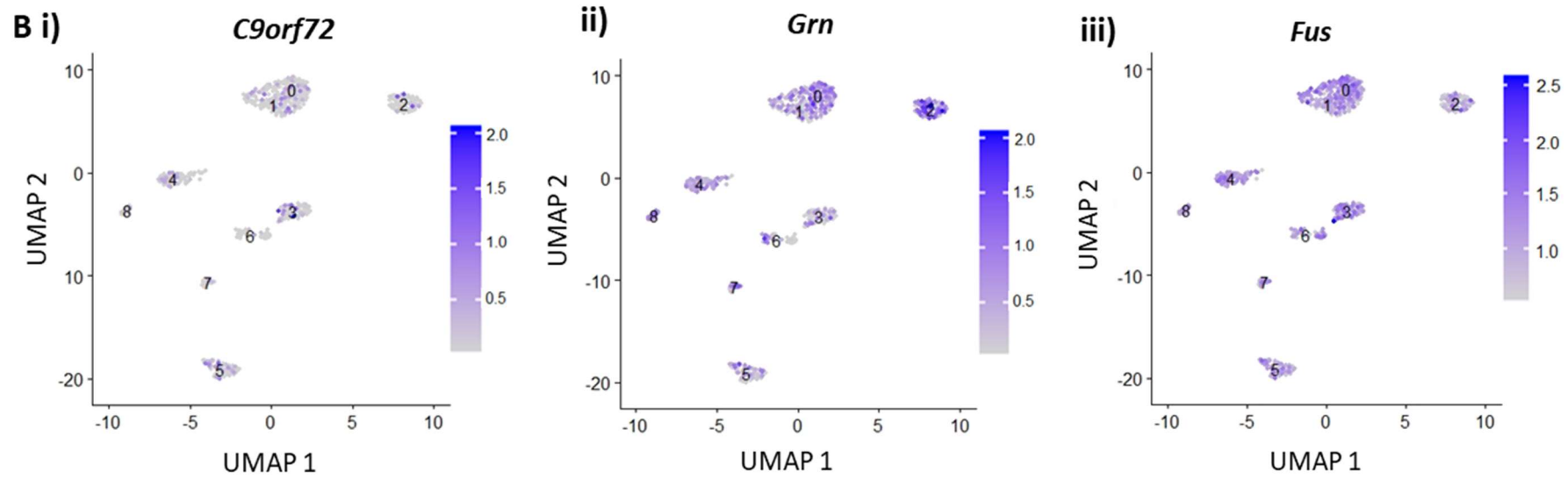
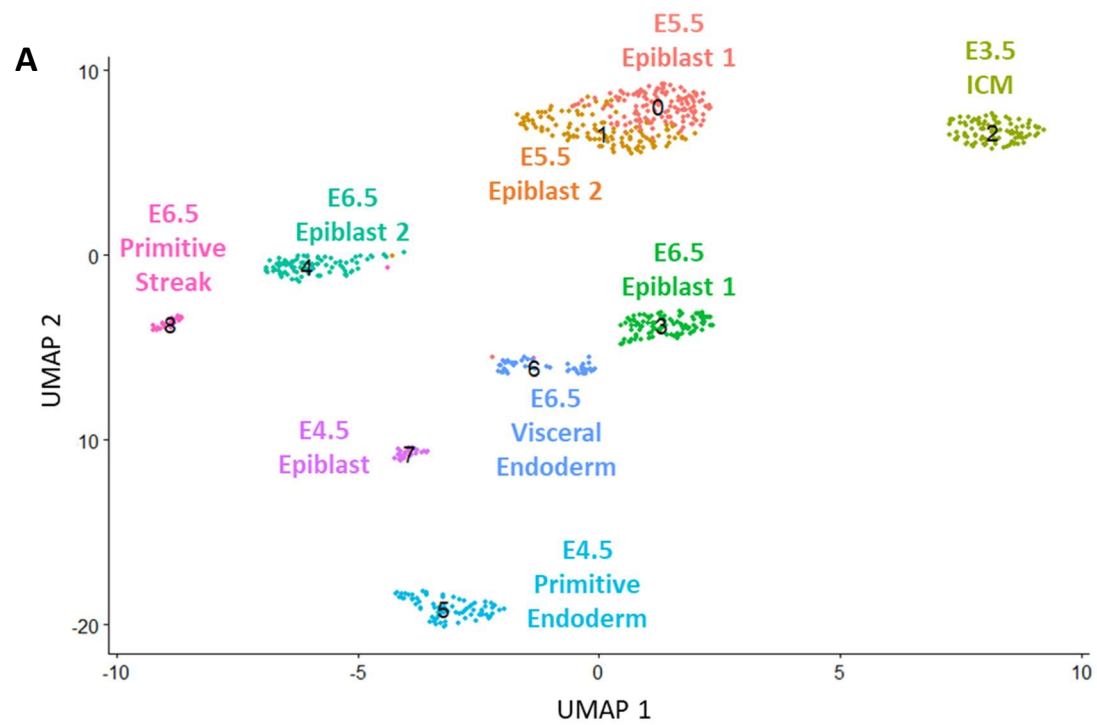


Figure 5.3B continued

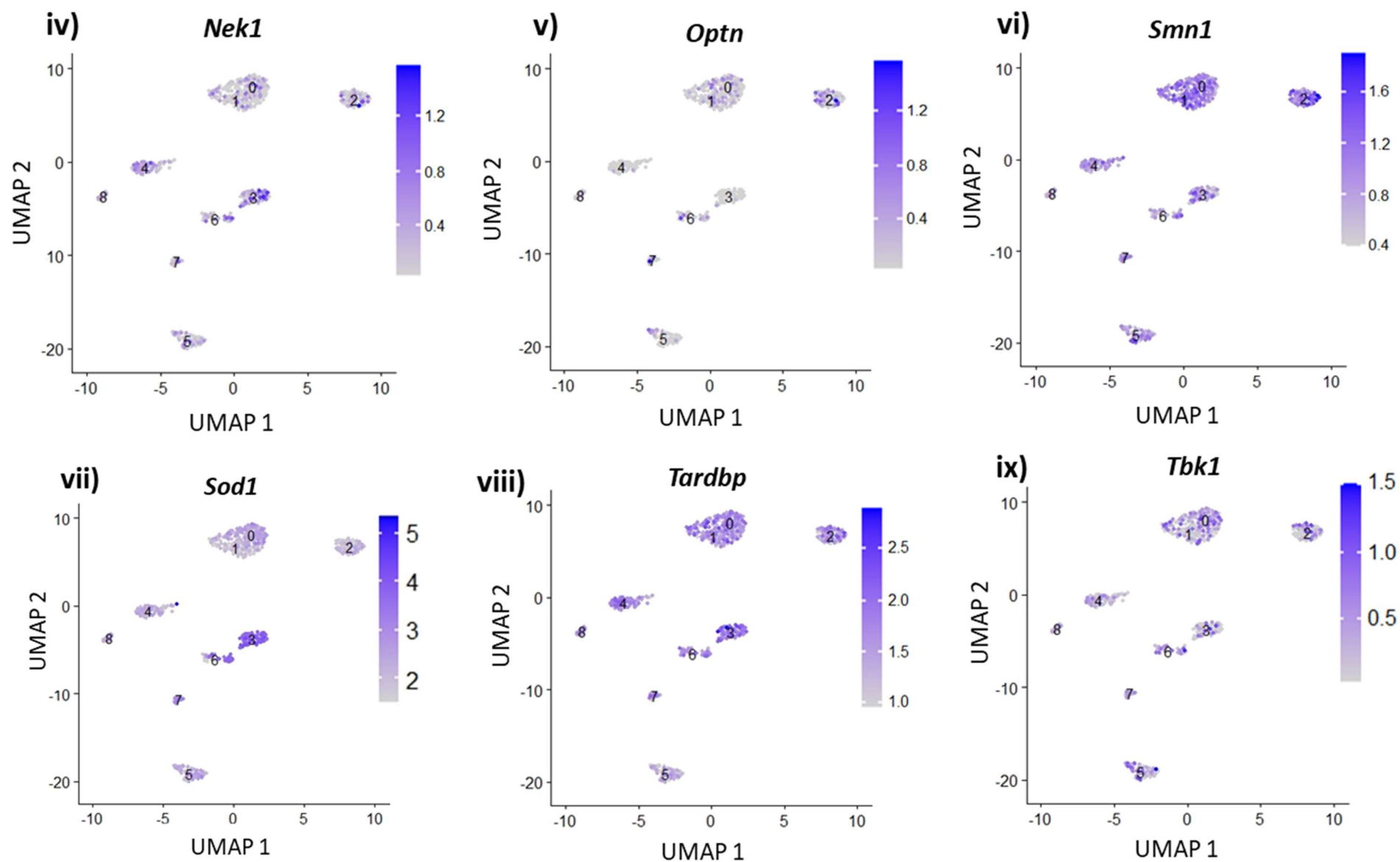


Figure 5.3B continued

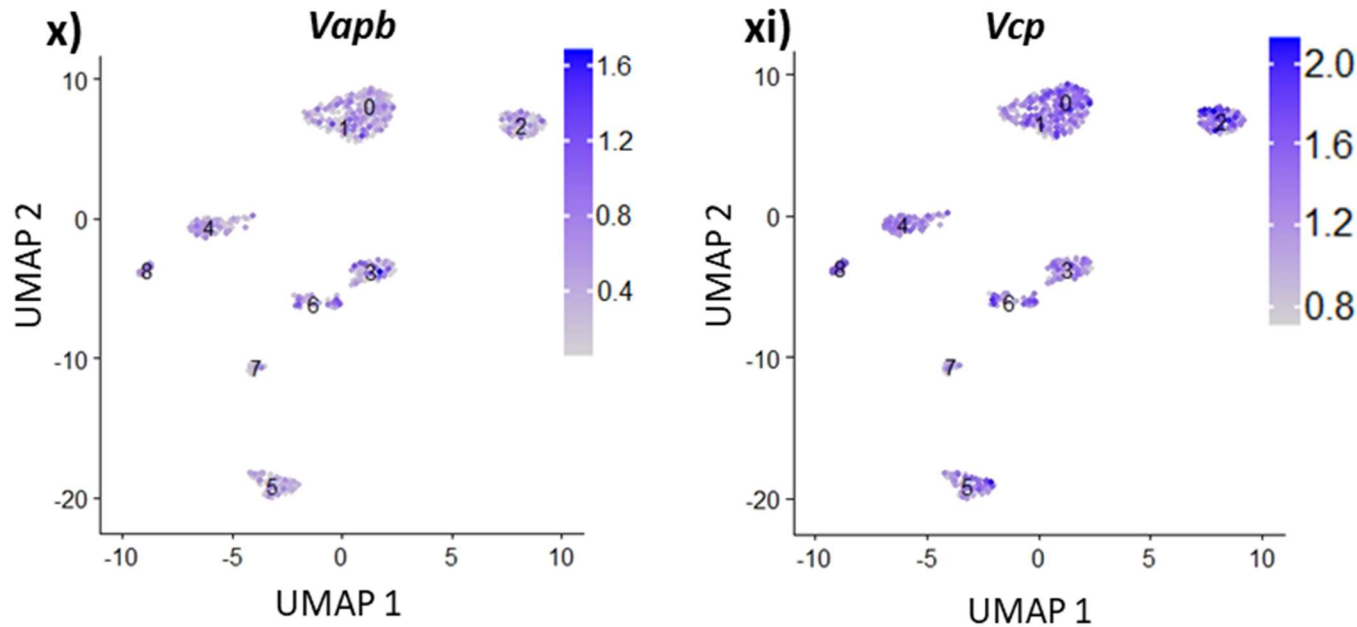


Figure 5.3 Expression of MND-associated genes in different cell types of the mouse embryo during early gastrulation. Early embryo samples were prepared for single cell collection and raw counts from single-cell RNA sequencing (scRNAseq) were initially done by Mohammed *et al.* 2017. In summary, blastocysts (E3.5 and 4.5) and embryos at E5.5 and E6.5 were collected from C57BL/6Bab mice. Each sample was prepared for scRNAseq. Raw scRNAseq count data was made publicly available and obtained from the Gene Expression Omnibus (GEO) website (<https://www.ncbi.nlm.nih.gov/geo/>; accession number GEO: GSE100597). We then clustered data with the Single-Cell Consensus Clustering tool and used lineage marker genes (as identified by Mohammed *et al.* 2017) to identify the different cell types in samples. **A.** Uniform Manifold Approximation and Projection (UMAP) clustering analysis of the scRNAseq dataset. UMAP is used to visualise dimensionality reduction, by taking information from the top principle components to arrange cells in a multidimensional space. 9 clusters are identified (0-8) along with the cell type they represent. **B.** The same UMAP plot was used to show the expression of MND-associated genes (**i-xi**), with an arbitrary scale of expression in each cluster, as produced using the Seurat package on Rstudio.

5.3.3. *Relative expression and TRIAGE analysis of MND-associated genes across neural lineage commitment*

We used the microarray data set of neural lineage directed (MEDII treated) and spontaneously differentiated mESCs to quantify changes in gene expression of NMD-associated genes.

Figure 5.4A shows some of these genes change expression (up or down) as differentiation proceeds, with the changes for *Fus*, *Grn*, *Sod1* and *Vapb* occurring in MEDII-treated cells but not spontaneously differentiating cells (Figure 5.4B), suggesting they are specific to neural lineage commitment. *Optn* was not analysed in this model due to technical limitations of the 22k Affymetrix platform (data not shown).

Figure 5.5 shows that, once TRIAGE transformed, none of the discordance scores (DS) for these MND-associated genes rank in the top 1359 genes, suggesting they don't play a leading role in cell fate. Due to limitations of the 22k Affymetrix platform, some MND-associated genes couldn't be TRIAGE transformed, and thus were not ranked (data not shown).

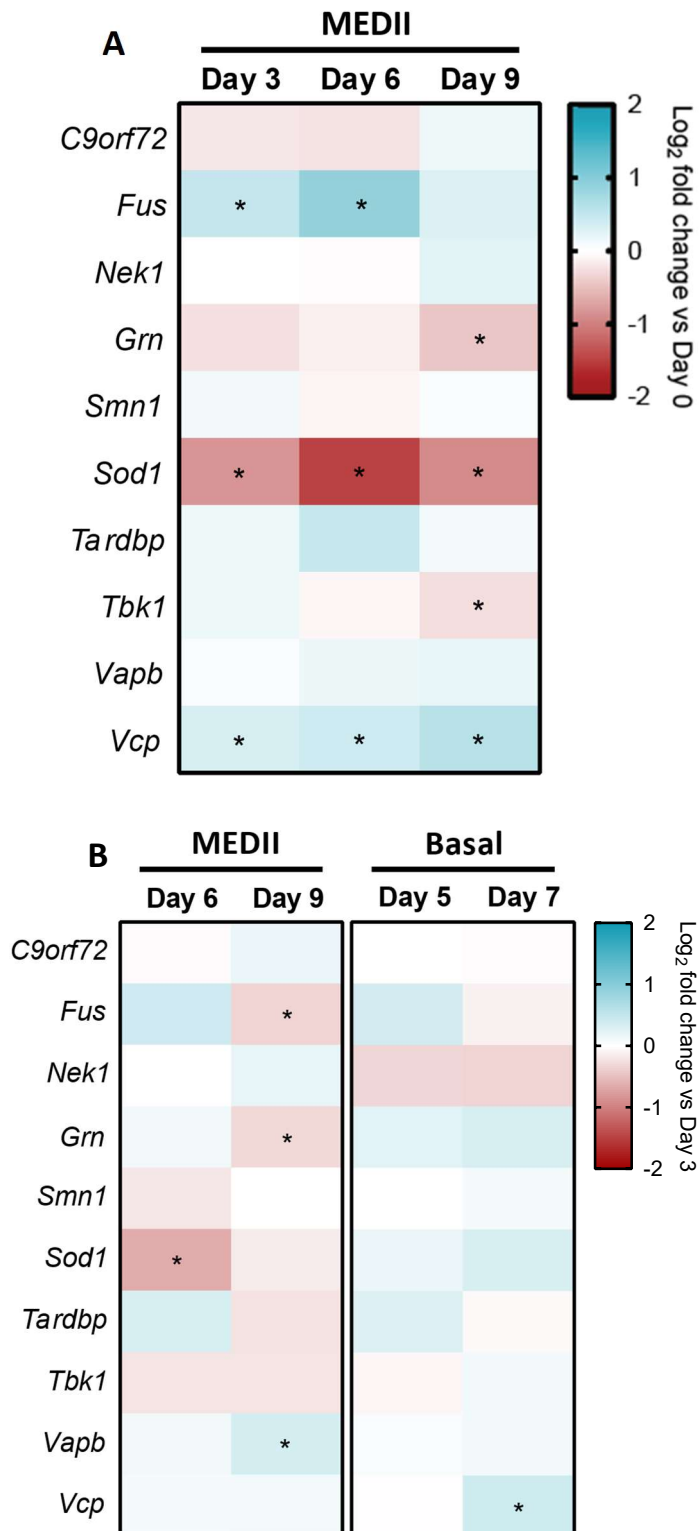


Figure 5.4 Relative expression of MND-associated genes across neural lineage commitment. D3 mESCs were cultured in suspension as embryoid bodies (EBs) in MEDII to Day 3 (EPLCs). These EBs continued to be cultured in MEDII out to Day 6 (definitive ectoderm) and Day 9 (neurectoderm) or switched (with a different set of Day 3 EBs) to Basal medium to spontaneously differentiate out to Day 5 and Day 7. cDNA samples ($n = 4$) (produced by Dr J. Rathjen) were hybridized onto the *Mus musculus* 22k Affymetrix platform. Microarray data was generated and normalized to take into account differences in cDNA concentration across the experiments and days. Raw data was extracted on Rstudio by Dr C. Wilkinson. We selected the MND-associated genes for this analysis and generated the heatmaps using Graphpad Prism. Relative expression was calculated and compared in **A.** to Day 0 for MEDII treatment or in **B.** to Day 3 (EPLCs) for both MEDII treatment and spontaneous differentiation. * Indicates significant change in expression based on a Benjamini-Hochberg adjusted $P < 0.05$.

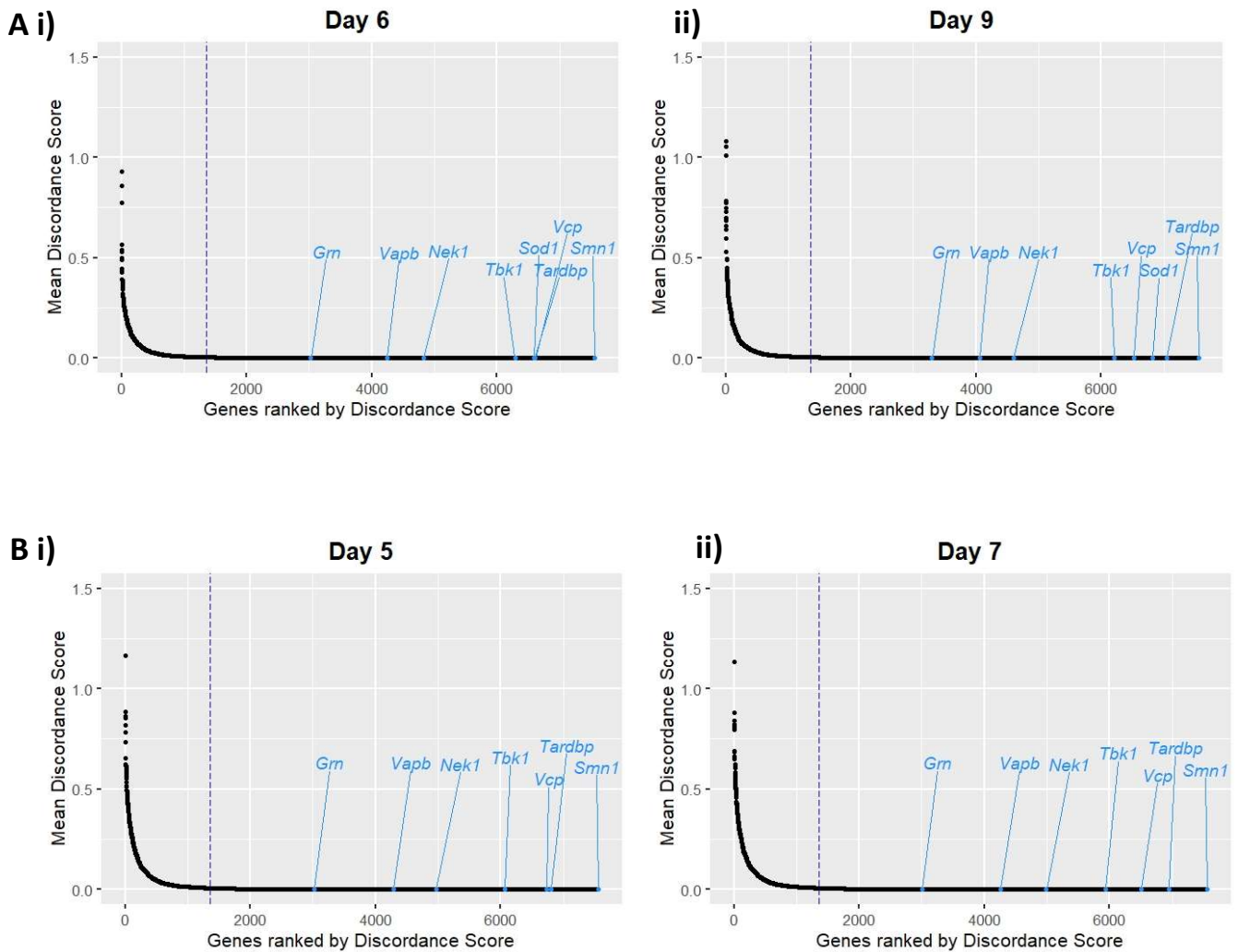


Figure 5.5 TRIAGE transformation of MND-associated genes across neural lineage commitment. D3 mESCs were cultured in suspension as EBs in MEDII to Day 3 (EPLCs). Culture was continued in **A**. MEDII out to **i)** Day 6 (definitive ectoderm) and **ii)** Day 9 (neurectoderm) or switched (with a different set of Day 3 EBs) to **B**. Basal medium to allow spontaneous differentiation out to **i)** Day 5 and **ii)** Day 7, as per Figure 4.1A and 4.1B. cDNA samples ($n = 4$) were used to produce microarray data, as per Figure 4.1A and 4.1B. Raw data was extracted on Rstudio, then fold changes relative to Day 3 were TRIAGE transformed to produce discordance scores (DS). Mean DS were plotted against DS rank, with DS rank threshold (1359) shown by the dotted blue line. MND-associated genes are labelled, with blue genes indicating a DS rank > 1359 .

5.3.4. qPCR analysis of the expression of MND-associated genes across the pluripotency spectrum and neural lineage commitment

To validate discoveries from our bioinformatics analysis, qPCR was used to quantify gene expression for 5 MND-associated genes – *Fus*, *Optn*, *Smn1*, *Sod1* and *Tardbp* – across different stages of pluripotency and neural lineage commitment.

All 5 genes were expressed at all stages of pluripotency and neural lineage commitment but Figure 5.6 shows that only 2 of these genes showed significant changes in expression: Both were downregulated during neural lineage commitment.

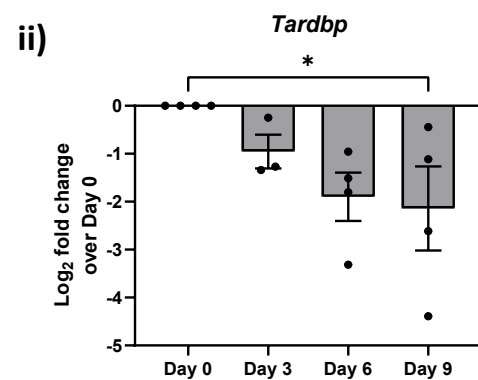
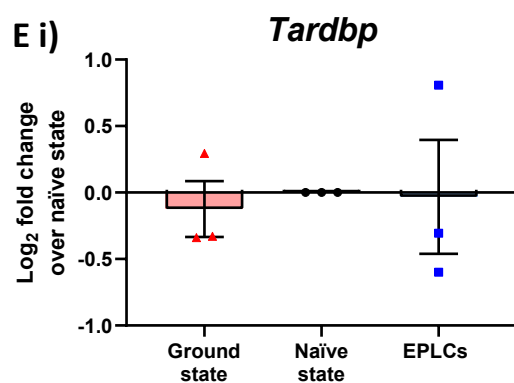
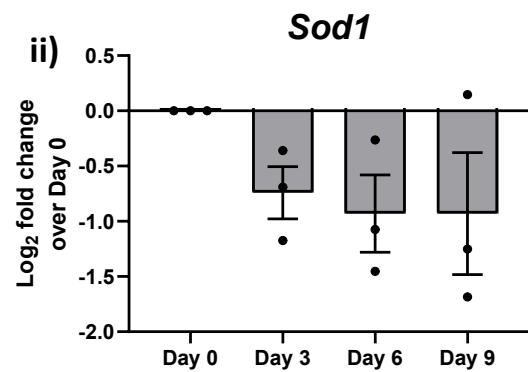
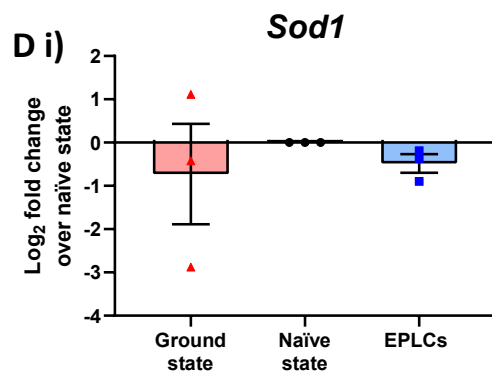
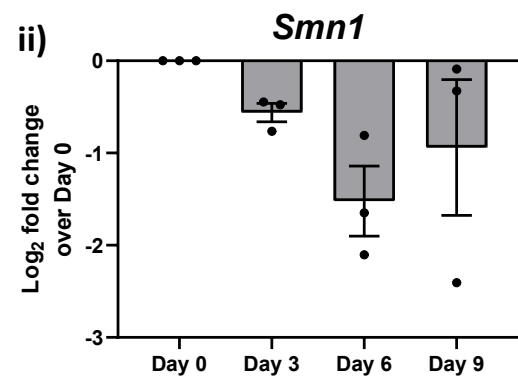
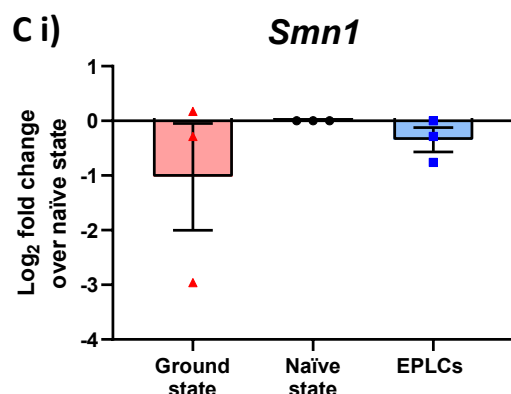
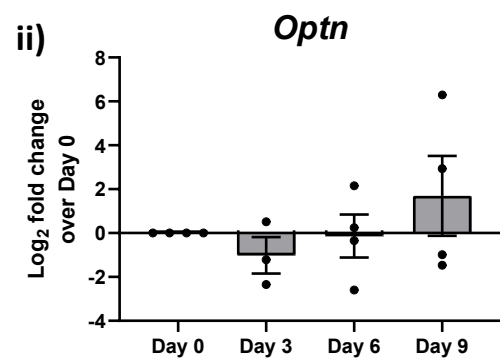
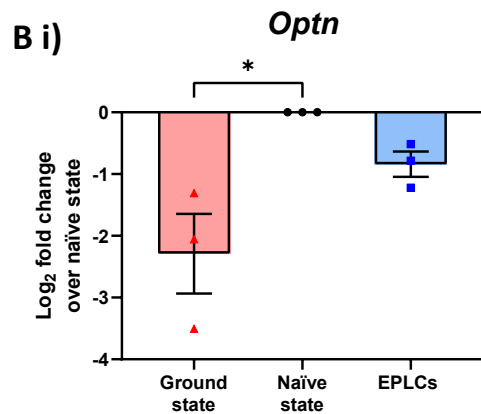
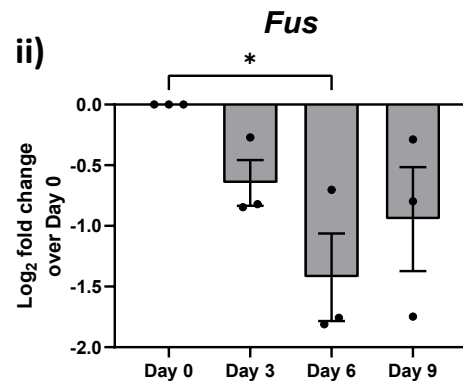
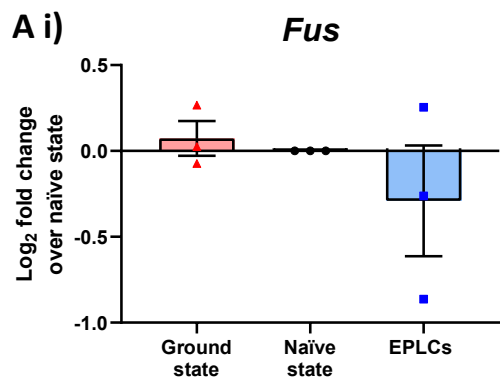


Figure 5.6 Expression of 5 MND-associated genes (A-E) in i) cells across the pluripotency spectrum and ii) EBs cultured by hanging drop method in L-Proline for neural lineage commitment. i) D3 mESCs were used to produce naïve-state mESCs, ground-state mESCs and EPLCs, as per Figure 3.2. **ii)** EBs derived from naïve-state mESCs (Day 0) were produced using hanging drop method using the L-Proline-based protocol for 9 days, as per Figure 4.1C. Samples from ground- and naïve-state mESCs, and EPLC and EB samples at Days 0, 3, 6 and 9 were collected, processed and subjected to qPCR. Results for each sample were normalized to the expression of *18s* as the reference gene, then shown relative to expression of **i)** the naïve state or **ii)** Day 0 as a log₂ fold change. Statistical analysis is by one-way ANOVA with Dunnett's *post hoc* test relative to **i)** naïve state or **ii)** Day 0. Bars represent mean ± SEM, n ≥3, **P* <0.05.

5.3.5. *MND-associated protein expression analysis across the pluripotency spectrum*

Figure 5.7 shows the Western blot analysis for 3 MND-associated proteins across pluripotency: Fus (encoded by the *Fus* gene), Tdp-43 (encoded by the *Tardbp* gene) and Vapb (encoded by the *Vapb* gene). All 3 proteins are expressed across all pluripotency states. The expression of Fus is upregulated in EPLCs compared to the naïve state and that for Vapb is downregulated in the ground state compared to the naïve state.

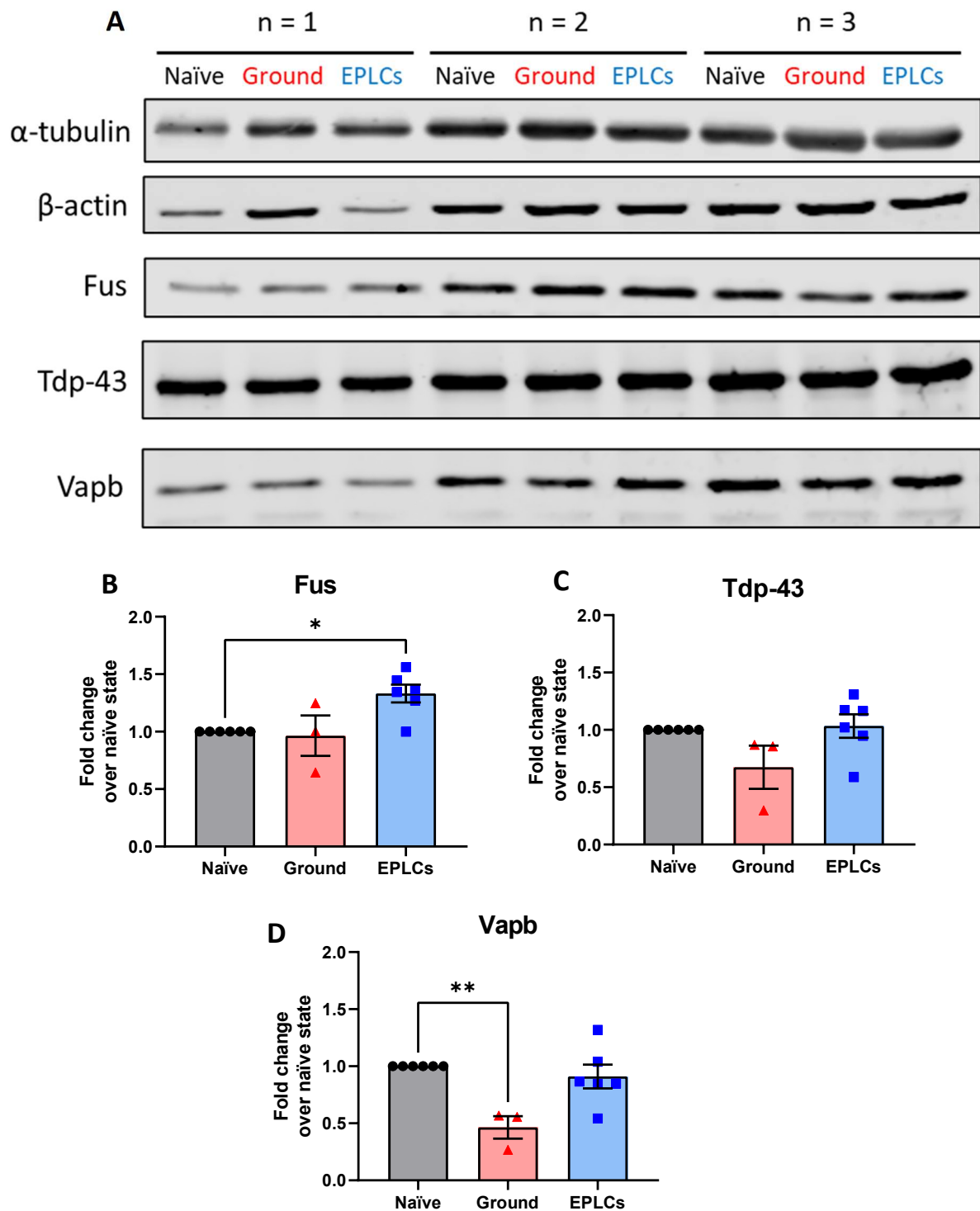


Figure 5.7 Western blot analysis of the expression of 3 MND-associated proteins across the pluripotency spectrum. D3 mESCs were used to produce ground- and naïve-state mESCs and EPLCs, as per Figure 3.2. Cell lysates were collected, processed and subjected to Western blot analysis. **A.** Representative blots. These blots were imaged and quantified on the LI-COR® Odyssey Infrared scanner for **B.** Fus, **C.** Tdp-43 and **D.** Vapb proteins. Fold changes are relative to the naïve state and normalised to loading control α -tubulin (8% (v/v) acrylamide/bis-acrylamide gel for Fus) or β -actin (10% (v/v) acrylamide/bis-acrylamide gel for Tdp-43 and Vapb). Data represents mean $\pm$ SEM, $n \geq 3$ independent experiments. Statistical analysis was conducted using a one-way ANOVA with Dunnett's *post hoc* test relative to naïve state. * $P < 0.05$, ** $P < 0.01$.

5.3.6. MND-associated protein expression analysis across neural lineage commitment

Figure 5.8 shows the Western blots were for the same 3 MND-associated proteins (Fus, Tdp-43 and Vapb) during neural lineage commitment. Again, all 3 proteins are expressed at all stages. The only change is the increased expression of Fus at Day 3. The inclusion of SB431542, the Nodal inhibitor used to suppress off-pathway differentiation to mesendoderm, did not alter expression of any of the proteins.

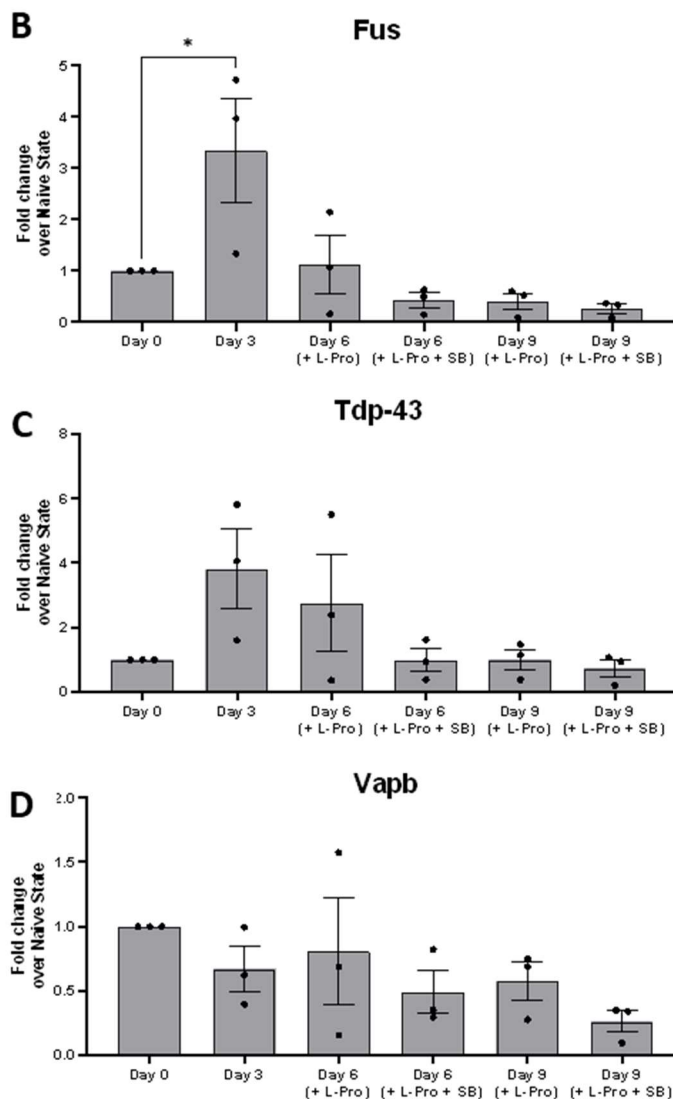
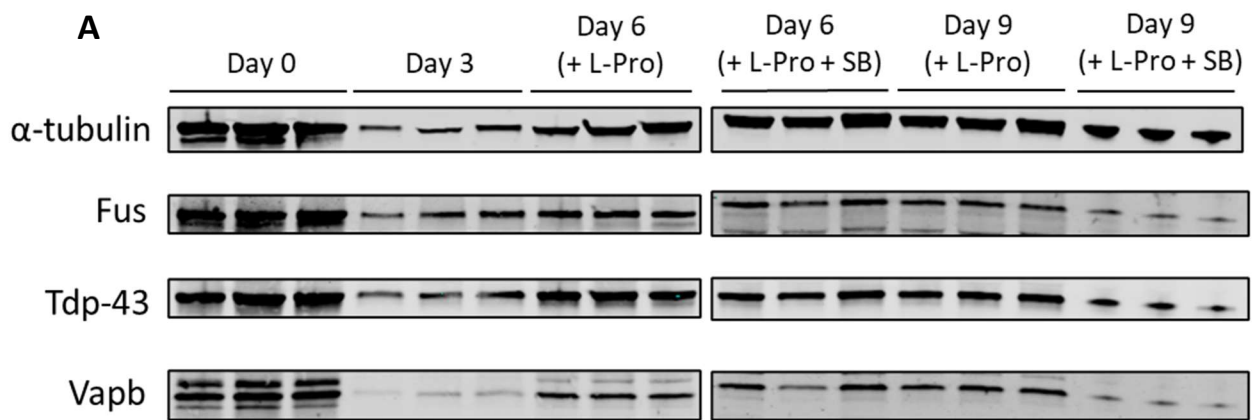


Figure 5.8 Western blot analysis of the expression of 3 MND-associated proteins across neural lineage commitment. Naive-state mESCs (Day 0) were cultured as EBs by the hanging drop method in L-Proline to promote neural lineage commitment over 9 days, as per Figure 4.1C. Samples from Days 0, 3, 6 and 9 were collected, processed and subjected to Western blot analysis. **A.** Blots showing all independent experiments. These blots were imaged and quantified on the LICOR® Odyssey Infrared scanner for **B.** Fus, **C.** Tdp-43 and **D.** Vapb proteins. Fold changes are relative to the Day 0 and normalised to loading control α -tubulin on the same Mini-PROTEAN® TGX™ Precast Gel (4-20%). Data represents mean $\pm$ SEM, n=3. Statistical analysis was conducted using an ordinary one-way ANOVA with Sidak's *post hoc* test, * P <0.05. SB: Nodal inhibitor SB431542.

5.4. Discussion

Results from this Chapter show that a range of MND-associated genes and proteins are expressed in cells across the pluripotency spectrum, early gastrulation, and neural lineage commitment. Various neural differentiation models, such as neuroblastoma cells, induced pluripotent stem cells and adult neural stem cell models, demonstrate the mechanisms of motor neuron differentiation, pathogenesis and potential mechanisms of onset of MND later in life (Nizzardo *et al.*, 2010; Wang *et al.*, 2013; Sterneckert *et al.*, 2014; Arrázola *et al.*, 2019; Bax *et al.*, 2019). However, the potential contributors to late onset of MND have generally not been looked at in the very early stages of development up to and including establishment of the nascent nervous system. Errors in expression, or the localisation of expression, of one of more of these genes at these early stages may not show phenotypically at this time but may, together with environmental stressors *in utero* or later in life, contribute to susceptibility to MND later in life per the Developmental Origins of Health and Disease hypothesis (De Boo and Harding, 2006; Ferreira *et al.*, 2016; Goyal *et al.*, 2019).

In vitro models recapitulating early development, from establishment of pluripotency to neural lineage commitment, allow the production and interrogation of various time-dependent, transient cell populations that are otherwise difficult to study *in vivo* (Rathjen *et al.*, 2002, 2003; Shparberg *et al.*, 2019a, 2019b; Glover *et al.*, 2022). Although we show that the expression of nearly all these MND-associated genes is established very early in development, TRIAGE transformation shows that none appears to contribute to cell fate decisions or cell identity at these times. This does not mean that errors in expression at this time do not play important roles in disease onset later in life.

5.4.1. *MND-associated genes are expressed across the pluripotency spectrum and during neural fate commitment but do not appear to contribute directly to cell-fate decisions*

Several MND-associated genes, including *Fus*, *Smn1*, *Tardbp*, *Tbkl* and *Vcp* are embryonically lethal when partially or fully knocked out (Schrank *et al.*, 1997; Bonnard *et al.*, 2000; Müller *et al.*, 2007; Kraemer *et al.*, 2010; Sephton *et al.*, 2010; Delhase *et al.*, 2012; Scekic-Zahirovic *et al.*, 2016). The requirement for their expression at these early stages may involve cell survival and/or proliferation rather than as drivers of cell fate, given their low DS. The effect of knockout or knockdown of these genes at these early stages using our *in vitro* model would shed light on this distinction and the mechanisms of action of the protein products of these genes, with particular reference to neural lineage commitment. This would include the use of ESC lines expressing mutant forms of these genes known to be associated with MND.

Other MND-associated genes tested in this Chapter are important for embryo development but are not lethal when knocked out (Rezaie *et al.*, 2007; Yon *et al.*, 2008; Bateman and Bennett, 2009; Chen *et al.*, 2011; Ying and Yue, 2012; Kabashi *et al.*, 2013; Amin *et al.*, 2014; Atkinson *et al.*, 2015; Borgese *et al.*, 2021). The expression patterns cover a wide range: That for *C9orf72* appears to be low across pluripotency, early gastrulation and during neural lineage commitment, (Figure 5.1, 5.3Bi and 5.4). It's possible that inappropriate expression during later stages of neural development than tested in our model are associated with later onset of MND.

On the other hand, *Grn* progressively decreases in expression as neural lineage commitment takes place (Figure 5.1, 5.3Bii and 5.4) while that for *Sod1* is low across pluripotency, high across early gastrulation, and low again during neural lineage commitment (Figure 5.1, 5.3Bvii and 5.4). *Vapb* progressively increases during early gastrulation and is maintained across neural lineage commitment (Figure 5.1, 5.3Bx and 5.4). However, their low

DS rank suggests that, like *Fus*, *Smn1*, *Tardbp*, *Tbkl* and *Vcp*, they don't play key roles in cell fate decisions, at least at these stages (Figure 5.2 and 5.5).

5.4.2. Expression of MND-associated genes and proteins do not always follow the same trends as bioinformatics interrogation

Because of various layers of regulatory control in cells, protein expression does not necessarily match gene expression. The 3 MND-associated proteins examined here by Western blot show that they are robustly expressed across various stages of development (Figure 5.7 and 5.8) but protein levels do not necessarily follow changes in gene expression. For example, the *Fus* gene maintains expression across pluripotency and is downregulated at the definitive ectoderm stage (Day 6) (Figure 5.6A). However, the *Fus* protein is upregulated only at the EPLC stage (Figure 5.7 and 5.8). This may indicate protein level changes are lagging behind gene expression changes, and could follow the decline in gene expression patterns as both turn over.

Quantification of erroneous expression, mislocalisation, and aberrant function of mutant forms of these proteins in our system may show their involvement in adversely affecting neural-cell fate. Stressors to the system can also be simultaneously applied. For example, the addition of reactive oxygen species in *Sod1*-compromised cells undergoing neural lineage commitment.

Nonetheless, expression and TRIAGE transformation of these MND-associated genes confirm their roles in housekeeping and cell structure maintenance across pluripotency and neural lineage commitment.

5.4.3. Expression of MND-associated proteins play their roles in housekeeping and cell structure across pluripotency and neural lineage commitment

Fus, *Tardbp* and *Vapb* were the chosen MND-associated proteins for Western blot analysis (Figure 5.7 and 5.8) because of the following rationale: *Fus* and *Tardbp* are proteins associated with RNA metabolism, while *Vapb* proteins are associated with Golgi apparatus maintenance and endoplasmic reticulum anchoring (Teuling *et al.*, 2007; Harley *et al.*, 2021). As *Fus*, *Tardbp*

and *Vapb* are known to be expressed in the embryo, their respective proteins are expected to be expressed during pluripotency and neural lineage commitment (Figure 5.7 and 5.8) (Sephton *et al.*, 2010; Scekcic-Zahirovic *et al.*, 2016).

Inhibiting the Nodal pathway using SB431542 was done to prevent off-pathway differentiation during the neural fate commitment process (Figure 4.1C) (Shparberg *et al.*, 2019a). However, this did not impact MND-associated protein expression throughout neural fate commitment (Figure 5.8), indicating these proteins have no impact on the Nodal pathway. However, given the genes' low DS ranks (Figure 5.5) it remains to be seen if these protein profiles have other roles beyond housekeeping and cellular structure.

5.4.4. Limitations and future directions

One of the limitations of this study is that TRIAGE transformation is only applicable to protein-coding genes (Shim *et al.*, 2020). Involvement of other non-protein-coding genes such as microRNAs and long non-coding RNAs (lncRNAs) are known contributors to cell fate (Dinger *et al.*, 2008; Melton *et al.*, 2010; Pauli *et al.*, 2011; Hu *et al.*, 2012; Fico *et al.*, 2019). For instance, microRNAs accelerate the de-adenylation and clearance of maternal mRNAs to promote developmental transitions and maternal-to-zygotic transition during preimplantation embryo development (Pauli *et al.*, 2011; Walser and Lipshitz, 2011). MicroRNAs regulate cell fate specification by modulating various signalling pathways, like participate in bistable loops that stabilize alternative cell fate decisions. For example, *let-7* and *Lin-28* establishes the trigger to switch between pluripotent and differentiated cell fates (Newman *et al.*, 2008; Piskounova *et al.*, 2008). lncRNAs can serve as scaffolds for the assembly of chromatin-modifying complexes and transcriptional regulators. For example, the lncRNA *HOTAIR* establishes a repressive chromatin state at multiple loci to promote epithelial-to-mesenchymal transition (Pauli *et al.*, 2011; Herz *et al.*, 2013; Lau and So, 2015; Jarroux *et al.*, 2021). Thus, a future direction can be to determine the nature of these non-coding RNAs known to drive cell fate with the role of

RNA-metabolism-associated proteins Fus and Tdp-43 and Smn1 across neural lineage commitment within the embryo.

In addition, mislocalisation of Fus and Tdp-43 is reversible via inhibition of the VCP D2 ATPase domain using the potent and selective inhibitor ML240 (Harley *et al.*, 2021). Thus, another future direction can be to determine the localisation of RNA-binding proteins using immunofluorescence imaging of EPLCs, DE and NE cell populations, and determine the effect of VCP D2 ATPase domain inhibition upon progression, using changes in EB morphology and gene expression associated with the transition as indicators. This would align with research demonstrated non-coding RNAs' contribution to onset of MNDs later in life (Vangoor *et al.*, 2021; Carvelli *et al.*, 2022; Pellegrini *et al.*, 2023).

As noted previously (Section 1.5 and Section 5.1), the pathophysiology of MND is multifactorial and complex. Therefore, onset of motor neuron degeneration involves different processes which, due to time constraints, have not been studied in this Chapter. One of these processes involve the role of kinases and their downstream targets associated with (motor) neuron survival (Guo *et al.*, 2020). The NF- κ B pathway is one of the main contributors to onset of MND that involves kinases I κ B kinase (IKK) α and β (Källstig *et al.*, 2021). This pathway is pro-inflammatory and is linked to many MND-associated genes. One of these is Optn, which inhibits the pathway (Akizuki *et al.*, 2013). Problems with expression of Optn protein upregulates the NF- κ B pathway, increasing expression of p38, Bax and Tnf α protein. Inappropriate activation of these proapoptotic molecules induces neuronal cell death as shown in multipotent cellular models (Shen *et al.*, 2011; Akizuki *et al.*, 2013; Toth and Atkin, 2018; Källstig *et al.*, 2021). Whether Optn-mediated inhibition of the NF- κ B pathway is required for neural lineage commitment remains to be determined. Knockdown of *Optn* via siRNA transfection and Western blot analysis of p38, Bax and Tnf α protein expression would be useful approaches to determine the involvement of the NF- κ B pathway in neural lineage commitment.

In the context of studying MND onset and processes involved in DOHaD, TRIAGE transformation is limited to mechanisms that utilise epigenetic repression and transcriptional patterns of protein-coding genes (Shim *et al.*, 2020; Sun *et al.*, 2023). For instance, although *C9orf72* is categorized to be protein-coding (Figure 5.2), the hexanucleotide repeat expansion in the gene's non-coding region is the trigger towards *C9orf72*-associated MND onset (Sterneckert *et al.*, 2014; Black *et al.*, 2017; Hao *et al.*, 2019; Karagiannis and Inoue, 2020). Genes acting as drivers of disease onset are also not limited to cell fate and identity. For instance, spinal muscular atrophy is known to be associated with *Smn* mutations, which are, in part, associated with various housekeeping processes, including mitochondrial activity (Chaytow *et al.*, 2018). Specific energetic demands of neurons and muscles make them susceptible to mitochondrial dysfunction. Due to the ubiquitous nature of mitochondrial activity, it is more likely genes encoding for such functions have very low DS.

There are limitations in the MEDII microarray dataset. These include the technical limitations of the *Mus musculus* 22k Affymetrix platform to detect certain genes. For example, *Optn* is missing (Figure 5.4 and 5.5). Additional probes for the same gene were detected for the *Sod1* gene upon microarray analysis and TRIAGE transformation (data not shown). This indicates other splice variants of the same gene may be expressed and/or contribute to neural lineage commitment, although one splice variant shows low DS, indicating they aren't involved cell-fate decisions, at least at these early stages (Figure 5.5).

Another limitation to note is the use of MEDII instead of L-Proline directly to acquire the microarray dataset. MEDII was the first medium used to culture mESCs to produce EPLCs and consequent promotion of neural differentiation (Rathjen *et al.*, 2002; Rodda *et al.*, 2002; Harvey *et al.*, 2010). However, this conditioned medium is highly complex, chemically undefined and likely to contain other (currently unknown) components which may assist and/or compromise neural commitment. Fractionation of MEDII resulted in identification of L-Proline as at least one of the pro-neural factors (Washington *et al.*, 2010). Retinoic acid, which is pro-neural when

added to ESCs (Bibel et al., 2004, 2007; Lang et al., 2004; Engberg et al., 2010; Stavridis et al., 2010; Tonge and Andrews, 2010), isn't detectable in MEDII (Morris, unpublished data). Thus, the use of L-Proline, instead of MEDII, allows for a more facile approach for identifying mechanisms which drive cell fate (Shparberg *et al.*, 2019a; Glover *et al.*, 2022). This can be exploited further by using technologies superior to microarray analysis, such as single-cell RNAseq.

5.5. References

Abramzon, Y., Johnson, J. O., Scholz, S. W., Taylor, J. P., Brunetti, M., Calvo, A., Mandrioli, J., Benatar, M., Mora, G., Restagno, G., Chiò, A., and Traynor, B. J. (2012) 'Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis', *Neurobiology of Aging*, 33(9), pp. 2231.e1-2231.e6. doi: <https://doi.org/10.1016/j.neurobiolaging.2012.04.005>.

Akizuki, M., Yamashita, H., Uemura, K., Maruyama, H., Kawakami, H., Ito, H., and Takahashi, R. (2013) 'Optineurin suppression causes neuronal cell death via NF- κ B pathway', *Journal of Neurochemistry*. John Wiley & Sons, Ltd, 126(6), pp. 699–704. doi: <https://doi.org/10.1111/jnc.12326>.

Almeida, S., Zhang, Z., Coppola, G., Mao, W., Futai, K., Karydas, A., Geschwind, M. D., Tartaglia, M. C., Gao, F., Gianni, D., Sena-Esteves, M., Geschwind, D. H., Miller, B. L., Farese, R. V., and Gao, F.-B. (2012) 'Induced pluripotent stem cell models of progranulin-deficient frontotemporal dementia uncover specific reversible neuronal defects', *Cell Reports*, 2(4), pp. 789–798. doi: <https://doi.org/10.1016/j.celrep.2012.09.007>.

Amin, A., Gad, A., Salilew-Wondim, D., Prastowo, S., Held, E., Hoelker, M., Rings, F., Tholen, E., Neuheff, C., Looft, C., Schellander, K., and Tesfaye, D. (2014) 'Bovine embryo survival under oxidative-stress conditions is associated with activity of the NRF2-mediated oxidative-stress-response pathway', *Molecular Reproduction and Development*. John Wiley & Sons, Ltd, 81(6), pp. 497–513. doi: <https://doi.org/10.1002/mrd.22316>.

Arrázola, M. S., Andraini, T., Szelechowski, M., Mouledous, L., Arnauné-Pelloquin, L., Davezac, N., Belenguer, P., Rampon, C., and Miquel, M.-C. (2019) 'Mitochondria in developmental and adult neurogenesis', *Neurotoxicity Research*, 36(2), pp. 257–267. doi: [10.1007/s12640-018-9942-y](https://doi.org/10.1007/s12640-018-9942-y).

Atkinson, R. A. K., Fernandez-Martos, C. M., Atkin, J. D., Vickers, J. C., and King, A.

E. (2015) ‘C9ORF72 expression and cellular localization over mouse development’, *Acta Neuropathologica Communications*, 3(1), p. 59. doi: 10.1186/s40478-015-0238-7.

Bateman, A. and Bennett, H. P. J. (2009) ‘The granulin gene family: From cancer to dementia’, *BioEssays*. John Wiley & Sons, Ltd, 31(11), pp. 1245–1254. doi: <https://doi.org/10.1002/bies.200900086>.

Bax, M., McKenna, J., Do-Ha, D., Stevens, C. H., Higginbottom, S., Balez, R., Cabral-da-Silva, M. E., Farrawell, N. E., Engel, M., Poronnik, P., Yerbury, J. J., Saunders, D. N., and Ooi, L. (2019) ‘The Ubiquitin Proteasome System Is a Key Regulator of Pluripotent Stem Cell Survival and Motor Neuron Differentiation’, *Cells* . doi: 10.3390/cells8060581.

Bernemann, C., Greber, B., Ko, K., Sterneckert, J., Han, D. W., Araúzo-Bravo, M. J., and Schöler, H. R. (2011) ‘Distinct developmental ground states of epiblast stem cell lines determine different pluripotency features’, *Stem Cells*. England, 29(10), pp. 1496–1503. doi: 10.1002/stem.709.

Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., and Barde, Y.-A. (2004) ‘Differentiation of mouse embryonic stem cells into a defined neuronal lineage’, *Nature Neuroscience*, 7(9), pp. 1003–1009. doi: 10.1038/nn1301.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.-A. (2007) ‘Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells’, *Nature Protocols*, 2(5), pp. 1034–1043. doi: 10.1038/nprot.2007.147.

Biscarini, S., Caputo, D., Peruzzi, G., Lu, L., Colantoni, A., Santini, T., Shneider, N. A., Caffarelli, E., Laneve, P., and Bozzoni, I. (2018) ‘Characterization of the lncRNA transcriptome in mESC-derived motor neurons: Implications for FUS-ALS’, *Stem Cell Research*, 27, pp. 172–179. doi: <https://doi.org/10.1016/j.scr.2018.01.037>.

Black, H. A., Leighton, D. J., Cleary, E. M., Rose, E., Stephenson, L., Colville, S., Ross,

D., Warner, J., Porteous, M., Gorrie, G. H., Swingle, R., Goldstein, D., Harms, M. B., Connick, P., Pal, S., Aitman, T. J., and Chandran, S. (2017) 'Genetic epidemiology of motor neuron disease-associated variants in the Scottish population', *Neurobiology of Aging*. 2016/12/21. Elsevier, 51, pp. 178.e11-178.e20. doi: 10.1016/j.neurobiolaging.2016.12.013.

Boivin, M., Pfister, V., Gaucherot, A., Ruffenach, F., Negroni, L., Sellier, C., and Charlet-Berguerand, N. (2020) 'Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders', *The EMBO Journal*. John Wiley & Sons, Ltd, 39(4), p. e100574. doi: <https://doi.org/10.15252/emboj.2018100574>.

Bonnard, M., Mirtsos, C., Suzuki, S., Graham, K., Huang, J., Ng, M., Itié, A., Wakeham, A., Shahinian, A., Henzel, W. J., Elia, A. J., Shillinglaw, W., Mak, T. W., Cao, Z., and Yeh, W.-C. (2000) 'Deficiency of T2K leads to apoptotic liver degeneration and impaired NF- κ B-dependent gene transcription', *The EMBO Journal*. John Wiley & Sons, Ltd, 19(18), pp. 4976–4985. doi: <https://doi.org/10.1093/emboj/19.18.4976>.

De Boo, H. A. and Harding, J. E. (2006) 'The developmental origins of adult disease (Barker) hypothesis', *Australian and New Zealand Journal of Obstetrics and Gynaecology*. John Wiley & Sons, Ltd, 46(1), pp. 4–14. doi: <https://doi.org/10.1111/j.1479-828X.2006.00506.x>.

Borgese, N., Iacomino, N., Colombo, S. F., and Navone, F. (2021) 'The link between VAPB loss of function and amyotrophic lateral sclerosis', *Cells*, 10(8), p. 1865. doi: [10.3390/cells10081865](https://doi.org/10.3390/cells10081865).

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S. M., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., and Vallier, L. (2007) 'Derivation of pluripotent epiblast stem cells from mammalian embryos', *Nature*. England, 448(7150), pp. 191–195. doi: [10.1038/nature05950](https://doi.org/10.1038/nature05950).

Carvelli, A., Setti, A., Desideri, F., Galfrè, S. G., Biscarini, S., Santini, T., Colantoni, A.,

Peruzzi, G., Marzi, M. J., Capauto, D., Di Angelantonio, S., Ballarino, M., Nicassio, F., Laneve, P., and Bozzoni, I. (2022) 'A multifunctional locus controls motor neuron differentiation through short and long noncoding RNAs', *The EMBO Journal*. John Wiley & Sons, Ltd, 41(13), p. e108918. doi: <https://doi.org/10.15252/emboj.2021108918>.

Chaytow, H., Huang, Y.-T., Gillingwater, T. H., and Faller, K. M. E. (2018) 'The role of survival motor neuron protein (SMN) in protein homeostasis', *Cellular and Molecular Life Sciences*, 75(21), pp. 3877–3894. doi: 10.1007/s00018-018-2849-1.

Chen-Plotkin, A. S., Lee, V. M.-Y., and Trojanowski, J. Q. (2010) 'TAR DNA-binding protein 43 in neurodegenerative disease', *Nature Reviews Neurology*, 6(4), pp. 211–220. doi: 10.1038/nrneurol.2010.18.

Chen, Y., Chen, P.-L., Chen, C.-F., Jiang, X., and Riley, D. J. (2008) 'Never-in-mitosis related Kinase 1 functions in DNA damage response and checkpoint control', *Cell Cycle*. Taylor & Francis, 7(20), pp. 3194–3201. doi: 10.4161/cc.7.20.6815.

Chen, Y., Chen, C.-F., Chiang, H.-C., Pena, M., Polci, R., Wei, R. L., Edwards, R. A., Hansel, D. E., Chen, P.-L., and Riley, D. J. (2011) 'Mutation of NIMA-related kinase 1 (NEK1) leads to chromosome instability', *Molecular Cancer*, 10(1), p. 5. doi: 10.1186/1476-4598-10-5.

Delhase, M., Kim, S.-Y., Lee, H., Naiki-Ito, A., Chen, Y., Ahn, E.-R., Murata, K., Kim, S.-J., Lautsch, N., Kobayashi, K. S., Shirai, T., Karin, M., and Nakanishi, M. (2012) 'TANK-binding kinase 1 (TBK1) controls cell survival through PAI-2/serpinB2 and transglutaminase 2', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 109(4), pp. E177–E186. doi: 10.1073/pnas.1119296109.

Dinger, M. E., Amaral, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., Askarian-Amiri, M. E., Ru, K., Solda, G., Simons, C., Sunkin, S. M., Crowe, M. L., Grimmond, S. M., Perkins, A. C., and Mattick, J. S. (2008) 'Long noncoding RNAs in mouse embryonic

stem cell pluripotency and differentiation’, *Genome Research*, 18(9), pp. 1433–1445. doi: 10.1101/gr.078378.108.

Dominguez, C. E., Cunningham, D., and Chandler, D. S. (2017) ‘SMN regulation in SMA and in response to stress: New paradigms and therapeutic possibilities’, *Human Genetics*. Germany, 136(9), pp. 1173–1191. doi: 10.1007/s00439-017-1835-2.

Engberg, N., Kahn, M., Petersen, D. R., Hansson, M., and Serup, P. (2010) ‘Retinoic acid synthesis promotes development of neural progenitors from mouse embryonic stem cells by suppressing endogenous, Wnt-dependent Nodal signaling’, *Stem Cells*, 28(9), pp. 1498–1509. doi: 10.1002/stem.479.

Ferreira, D. J. S., Sellitti, D. F., and Lagranha, C. J. (2016) ‘Protein undernutrition during development and oxidative impairment in the central nervous system (CNS): Potential factors in the occurrence of metabolic syndrome and CNS disease’, *Journal of Developmental Origins of Health and Disease*. Cambridge University Press, 7(5), pp. 513–524. doi: 10.1017/S2040174416000246.

Fico, A., Fiorenzano, A., Pascale, E., Patriarca, E. J., and Minchiotti, G. (2019) ‘Long non-coding RNA in stem cell pluripotency and lineage commitment: functions and evolutionary conservation’, *Cellular and Molecular Life Sciences*, 76(8), pp. 1459–1471. doi: 10.1007/s00018-018-3000-z.

Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.

Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) ‘L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell’, in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.

Gonzalez-Perez, P., Cirulli, E. T., Drory, V. E., Dabby, R., Nisipeanu, P., Carasso, R. L., Sadeh, M., Fox, A., Festoff, B. W., Sapp, P. C., McKenna-Yasek, D., Goldstein, D. B., Brown, R. H., and Blumen, S. C. (2012) 'Novel mutation in VCP gene causes atypical amyotrophic lateral sclerosis', *Neurology*. United States, 79(22), pp. 2201–2208. doi: 10.1212/WNL.0b013e318275963b.

Goyal, D., Limesand, S. W., and Goyal, R. (2019) 'Epigenetic responses and the developmental origins of health and disease', *Journal of Endocrinology*. Bristol, UK: Bioscientifica Ltd, 242(1), pp. T105–T119. doi: 10.1530/JOE-19-0009.

Guo, W., Vandoorne, T., Steyaert, J., Staats, K. A., and Van Den Bosch, L. (2020) 'The multifaceted role of kinases in amyotrophic lateral sclerosis: genetic, pathological and therapeutic implications', *Brain*, 143(6), pp. 1651–1673. doi: 10.1093/brain/awaa022.

Hao, Z., Liu, L., Tao, Z., Wang, R., Ren, H., Sun, H., Lin, Z., Zhang, Z., Mu, C., Zhou, J., and Wang, G. (2019) 'Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR', *Nature Communications*, 10(1), p. 2906. doi: 10.1038/s41467-019-10956-w.

Harding, O., Evans, C. S., Ye, J., Cheung, J., Maniatis, T., and Holzbaur, E. L. F. (2021) 'ALS- and FTD-associated missense mutations in TBK1 differentially disrupt mitophagy', *Proceedings of the National Academy of Sciences*. Proceedings of the National Academy of Sciences, 118(24), p. e2025053118. doi: 10.1073/pnas.2025053118.

Hardy, J. and Rogaeva, E. (2014) 'Motor neuron disease and frontotemporal dementia: Sometimes related, sometimes not', *Experimental Neurology*, 262, pp. 75–83. doi: <https://doi.org/10.1016/j.expneurol.2013.11.006>.

Hardy, M. L. M., Day, M. L., and Morris, M. B. (2021) 'Redox Regulation and Oxidative Stress in Mammalian Oocytes and Embryos Developed In Vivo and In Vitro', *International Journal of Environmental Research and Public Health*, 18(21), p. 11374. doi:

Harley, J., Hagemann, C., Serio, A., and Patani, R. (2021) ‘TDP-43 and FUS mislocalization in VCP mutant motor neurons is reversed by pharmacological inhibition of the VCP D2 ATPase domain’, *Brain Communications*. Oxford University Press, 3(3), p. fcab166. doi: 10.1093/braincomms/fcab166.

Harvey, N. T., Hughes, J. N., Lonic, A., Yap, C., Long, C., Rathjen, P. D., and Rathjen, J. (2010) ‘Response to BMP4 signalling during ES cell differentiation defines intermediates of the ectoderm lineage’, *Journal of Cell Science*, 123(10), pp. 1796–1804. doi: 10.1242/jcs.047530.

Hayes, L. R., Asress, S. A., Li, Y., Galkin, A., Stepanova, A., Kawamata, H., Manfredi, G., and Glass, J. D. (2019) ‘Distal denervation in the SOD1 knockout mouse correlates with loss of mitochondria at the motor nerve terminal.’, *Experimental neurology*. United States, 318, pp. 251–257. doi: 10.1016/j.expneurol.2019.05.008.

Heindel, J. J. and Vandenberg, L. N. (2015) ‘Developmental origins of health and disease: A paradigm for understanding disease cause and prevention’, *Current Opinion in Pediatrics*, 27(2), pp. 248–253. doi: 10.1097/MOP.0000000000000191.

Herz, H. M., Garruss, A., and Shilatifard, A. (2013) ‘SET for life: Biochemical activities and biological functions of SET domain-containing proteins’, *Trends in Biochemical Sciences*, 38(12), pp. 621–639. doi: 10.1016/j.tibs.2013.09.004.

Higelin, J., Catanese, A., Semelink-Sedlacek, L. L., Oeztuerk, S., Lutz, A.-K., Bausinger, J., Barbi, G., Speit, G., Andersen, P. M., Ludolph, A. C., Demestre, M., and Boeckers, T. M. (2018) ‘NEK1 loss-of-function mutation induces DNA damage accumulation in ALS patient-derived motoneurons’, *Stem Cell Research*, 30, pp. 150–162. doi: <https://doi.org/10.1016/j.scr.2018.06.005>.

Hu, F., Padukkavidana, T., Vægter, C. B., Brady, O. A., Zheng, Y., Mackenzie, I. R., Feldman, H. H., Nykjaer, A., and Strittmatter, S. M. (2010) ‘Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin’, *Neuron*, 68(4), pp. 654–667. doi: <https://doi.org/10.1016/j.neuron.2010.09.034>.

Hu, W., Alvarez-Dominguez, J. R., and Lodish, H. F. (2012) ‘Regulation of mammalian cell differentiation by long non-coding RNAs’, *EMBO Reports*, 13(11), pp. 971–983. doi: [10.1038/embor.2012.145](https://doi.org/10.1038/embor.2012.145).

Ichihanagi, N., Fujimori, K., Yano, M., Ishihara-Fujisaki, C., Sone, T., Akiyama, T., Okada, Y., Akamatsu, W., Matsumoto, T., Ishikawa, M., Nishimoto, Y., Ishihara, Y., Sakuma, T., Yamamoto, T., Tsuiji, H., Suzuki, N., Warita, H., Aoki, M., and Okano, H. (2016) ‘Establishment of in vitro FUS-associated familial amyotrophic lateral sclerosis model using human induced pluripotent stem cells’, *Stem Cell Reports*, 6(4), pp. 496–510. doi: <https://doi.org/10.1016/j.stemcr.2016.02.011>.

Ito, Y., Ofengeim, D., Najafzadeh, A., Das, S., Saberi, S., Li, Y., Hitomi, J., Zhu, H., Chen, H., Mayo, L., Geng, J., Amin, P., DeWitt, J. P., Mookhtiar, A. K., Florez, M., Ouchida, A. T., Fan, J., Pasparakis, M., Kelliher, M. A., Ravits, J., and Yuan, J. (2016) ‘RIPK1 mediates axonal degeneration by promoting inflammation and necroptosis in ALS’, *Science*. American Association for the Advancement of Science, 353(6299), pp. 603–608. doi: [10.1126/science.aaf6803](https://doi.org/10.1126/science.aaf6803).

Jarroux, J., Foretek, D., Bertrand, C., Gabriel, M., Szachnowski, U., Saci, Z., Guo, S., Londoño-Vallejo, A., Pinskaya, M., and Morillon, A. (2021) ‘HOTAIR lncRNA promotes epithelial–mesenchymal transition by redistributing LSD1 at regulatory chromatin regions’, *EMBO reports*. John Wiley & Sons, Ltd, 22(7), p. e50193. doi: <https://doi.org/10.15252/embr.202050193>.

Kabashi, E., El Oussini, H., Bercier, V., Gros-Louis, F., Valdmanis, P. N., McDearmid,

J., Meijer, I. A., Dion, P. A., Dupre, N., Hollinger, D., Sinniger, J., Dirrig-Grosch, S., Camu, W., Meininger, V., Loeffler, J.-P., René, F., Drapeau, P., Rouleau, G. A., and Dupuis, L. (2013) 'Investigating the contribution of VAPB/ALS8 loss of function in amyotrophic lateral sclerosis', *Human Molecular Genetics*, 22(12), pp. 2350–2360. doi: 10.1093/hmg/ddt080.

Källstig, E., McCabe, B. D., and Schneider, B. L. (2021) 'The links between ALS and NF- κ B', *International Journal of Molecular Sciences*, 22(8), p. 3875. doi: 10.3390/ijms22083875.

Karagiannis, P. and Inoue, H. (2020) 'ALS, a cellular whodunit on motor neuron degeneration', *Molecular and Cellular Neuroscience*, 107, p. 103524. doi: <https://doi.org/10.1016/j.mcn.2020.103524>.

Karch, C. M., Ezerskiy, L., Redaelli, V., Giovagnoli, A. R., Tiraboschi, P., Pelliccioni, G., Pelliccioni, P., Kapetis, D., D'Amato, I., Piccoli, E., Ferretti, M. G., Tagliavini, F., and Rossi, G. (2016) 'Missense mutations in progranulin gene associated with frontotemporal lobar degeneration: Study of pathogenetic features', *Neurobiology of aging*. United States, 38, pp. 215.e1-215.e12. doi: 10.1016/j.neurobiolaging.2015.10.029.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., and Winter, D. R. (2018) 'A beginner's guide to analysis of RNA sequencing data', *American Journal of Respiratory Cell and Molecular Biology*, 59(2), pp. 145–157. doi: 10.1165/rcmb.2017-0430TR.

Kojima, Y., Kaufman-Francis, K., Studdert, J. B., Steiner, K. A., Power, M. D., Loebel, D. A. F., Jones, V., Hor, A., de Alencastro, G., Logan, G. J., Teber, E. T., Tam, O. H., Stutz, M. D., Alexander, I. E., Pickett, H. A., and Tam, P. P. L. (2014) 'The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak', *Cell Stem Cell*, 14(1), pp. 107–120. doi: <https://doi.org/10.1016/j.stem.2013.09.014>.

Koppers, M., Blokhuis, A. M., Westeneng, H.-J., Terpstra, M. L., Zundel, C. A. C., Vieira

de Sá, R., Schellevis, R. D., Waite, A. J., Blake, D. J., Veldink, J. H., van den Berg, L. H., and Pasterkamp, R. J. (2015) 'C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits', *Annals of neurology*. 2015/07/03. John Wiley and Sons Inc., 78(3), pp. 426–438. doi: 10.1002/ana.24453.

Kraemer, B. C., Schuck, T., Wheeler, J. M., Robinson, L. C., Trojanowski, J. Q., Lee, V. M. Y., and Schellenberg, G. D. (2010) 'Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis', *Acta Neuropathologica*, 119(4), pp. 409–419. doi: 10.1007/s00401-010-0659-0.

Kurashige, T., Kuramochi, M., Ohsawa, R., Yamashita, Y., Shioi, G., Morino, H., Kamada, M., Ayaki, T., Ito, H., Sotomaru, Y., Maruyama, H., and Kawakami, H. (2021) 'Optineurin defects cause TDP43-pathology with autophagic vacuolar formation', *Neurobiology of Disease*, 148, p. 105215. doi: <https://doi.org/10.1016/j.nbd.2020.105215>.

Lang, K. J. D., Rathjen, J., Vassilieva, S., and Rathjen, P. D. (2004) 'Differentiation of embryonic stem cells to a neural fate: A route to re-building the nervous system?', *Journal of Neuroscience Research*. John Wiley & Sons, Ltd, 76(2), pp. 184–192. doi: 10.1002/jnr.20036.

Lau, P. N. I. and So, C. W. E. (2015) 'Polycomb and Trithorax factors in transcriptional and epigenetic regulation', in Huang, S., Litt, M. D., and Blakey, C. A. (eds) *Epigenetic Gene Expression and Regulation*. Elsevier Science & Technology, pp. 63–94. doi: 10.1016/B978-0-12-799958-6.00004-4.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A., and Melki, J. (1995) 'Identification and characterization of a spinal muscular atrophy-determining gene', *Cell*, 80(1), pp. 155–165. doi: [https://doi.org/10.1016/0092-8674\(95\)90460-3](https://doi.org/10.1016/0092-8674(95)90460-3).

Liu, E. Y., Cali, C. P., and Lee, E. B. (2017) 'RNA metabolism in neurodegenerative

disease', *Disease Models & Mechanisms*. England, 10(5), pp. 509–518. doi: 10.1242/dmm.028613.

McCluskey, G., Duddy, W., Haffey, S., Morrison, K., Donaghy, C., and Duguez, S. (2022) 'Epidemiology and survival trends of motor neurone disease in Northern Ireland from 2015 to 2019', *European Journal of Neurology*. John Wiley & Sons, Ltd, 29(3), pp. 707–714. doi: 10.1111/ene.15172.

Melton, C., Judson, R. L., and Blelloch, R. (2010) 'Opposing microRNA families regulate self-renewal in mouse embryonic stem cells', *Nature*, 463(7281), pp. 621–626. doi: 10.1038/nature08725.

Mirra, A., Rossi, S., Scaricamazza, S., Di Salvio, M., Salvatori, I., Valle, C., Rusmini, P., Poletti, A., Cestra, G., Carri, M. T., and Cozzolino, M. (2017) 'Functional interaction between FUS and SMN underlies SMA-like splicing changes in wild-type hFUS mice.', *Scientific reports*. England, 7(1), p. 2033. doi: 10.1038/s41598-017-02195-0.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., Marioni, J. C., and Reik, W. (2017) 'Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation', *Cell Reports*, 20(5), pp. 1215–1228. doi: 10.1016/j.celrep.2017.07.009.

Müller, J. M. M., Deinhardt, K., Rosewell, I., Warren, G., and Shima, D. T. (2007) 'Targeted deletion of p97 (VCP/CDC48) in mouse results in early embryonic lethality', *Biochemical and Biophysical Research Communications*, 354(2), pp. 459–465. doi: <https://doi.org/10.1016/j.bbrc.2006.12.206>.

De Muynck, L. and Van Damme, P. (2011) 'Cellular effects of progranulin in health and disease', *Journal of Molecular Neuroscience*, 45(3), pp. 549–560. doi: 10.1007/s12031-011-9553-z.

Newman, M. A., Thomson, J. M., and Hammond, S. M. (2008) 'Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing', *RNA*, 14(8), pp. 1539–1549. doi: 10.1261/rna.1155108.

Niimi, K., Harding, D. R. K., Parshot, R., King, A., Lun, D. J., Decottignies, A., Niimi, M., Lin, S., Cannon, R. D., Goffeau, A., and Monk, B. C. (2004) 'Chemosensitization of Fluconazole Resistance in *Saccharomyces cerevisiae* and Pathogenic Fungi by a D-Octapeptide Derivative', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology, 48(4), pp. 1256–1271. doi: 10.1128/AAC.48.4.1256-1271.2004.

Nizzardo, M., Simone, C., Falcone, M., Locatelli, F., Riboldi, G., Comi, G. P., and Corti, S. (2010) 'Human motor neuron generation from embryonic stem cells and induced pluripotent stem cells', *Cellular and Molecular Life Sciences*, 67(22), pp. 3837–3847. doi: 10.1007/s00018-010-0463-y.

O'Rourke, J. G., Bogdanik, L., Yáñez, A., Lall, D., Wolf, A. J., Muhammad, A. K. M. G., Ho, R., Carmona, S., Vit, J. P., Zarrow, J., Kim, K. J., Bell, S., Harms, M. B., Miller, T. M., Dangler, C. A., Underhill, D. M., Goodridge, H. S., Lutz, C. M., and Baloh, R. H. (2016) 'C9orf72 is required for proper macrophage and microglial function in mice', *Science*. American Association for the Advancement of Science, 351(6279), pp. 1324–1329. doi: 10.1126/science.aaf1064.

Park, J., Kim, J.-E., and Song, T.-J. (2022) 'The global burden of motor neuron disease: An analysis of the 2019 Global Burden of Disease Study', *Frontiers in Neurology*. Switzerland, 13, p. 864339. doi: 10.3389/fneur.2022.864339.

Pauli, A., Rinn, J. L., and Schier, A. F. (2011) 'Non-coding RNAs as regulators of embryogenesis', *Nature Reviews Genetics*, 12(2), pp. 136–149. doi: 10.1038/nrg2904.

Pellegrini, F., Padovano, V., Biscarini, S., Santini, T., Setti, A., Galfrè, S. G., Silenzi, V., Vitiello, E., Mariani, D., Nicoletti, C., Torromino, G., De Leonibus, E., Martone, J., and

Bozzoni, I. (2023) ‘A KO mouse model for the lncRNA Lhx1os produces motor neuron alterations and locomotor impairment’, *iScience*, 26(1), p. 105891. doi: <https://doi.org/10.1016/j.isci.2022.105891>.

Philips, T., De Muynck, L., Nguyen Thi Thu, H., Weynants, B., Vanacker, P., Dhondt, J., Sleegers, K., Schelhaas, H. J., Verbeek, M., Vandenberghe, R., Sciot, R., Van Broeckhoven, C., Lambrechts, D., Van Leuven, F., Van Den Bosch, L., Robberecht, W., and Van Damme, P. (2010) ‘Microglial upregulation of progranulin as a marker of motor neuron degeneration’, *Journal of Neuropathology & Experimental Neurology*, 69(12), pp. 1191–1200. doi: 10.1097/NEN.0b013e3181fc9aea.

Piskounova, E., Viswanathan, S. R., Janas, M., LaPierre, R. J., Daley, G. Q., Sliz, P., and Gregory, R. I. (2008) ‘Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28’, *Journal of Biological Chemistry*, 283(31), pp. 21310–21314. doi: 10.1074/jbc.C800108200.

Qiu, L., Qiao, T., Beers, M., Tan, W., Wang, H., Yang, B., and Xu, Z. (2013) ‘Widespread aggregation of mutant VAPB associated with ALS does not cause motor neuron degeneration or modulate mutant SOD1 aggregation and toxicity in mice’, *Molecular Neurodegeneration*, 8(1), p. 1. doi: 10.1186/1750-1326-8-1.

Rathjen, J., Lake, J.-A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999) ‘Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors’, *Journal of Cell Science*, 112(5), pp. 601–612. doi: 10.1242/jcs.004613.

Rathjen, J., Haines, B. P., Hudson, K. M., Nesci, A., Dunn, S., and Rathjen, P. D. (2002) ‘Directed differentiation of pluripotent cells to neural lineages: homogeneous formation and differentiation of a neurectoderm population’, *Development*, 129(11), pp. 2649–2661. Available at: <http://dev.biologists.org/content/129/11/2649>.

Rathjen, J., Washington, J. M., Bettess, M. D., and Rathjen, P. D. (2003) 'Identification of a biological activity that supports maintenance and proliferation of pluripotent cells from the primitive ectoderm of the mouse', *Biology of Reproduction*, 69(6), pp. 1863–1871. doi: 10.1095/biolreprod.103.017384.

Rezaie, T., Stoilov, I., and Sarfarazi, M. (2007) 'Embryonic expression of the optineurin (glaucoma) gene in different stages of mouse development', *Molecular vision*. United States, 13, pp. 1446–1450. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17893645>.

Rodda, S. J., Kavanagh, S. J., Rathjen, J., and Rathjen, P. D. (2002) 'Embryonic stem cell differentiation and the analysis of mammalian development', *The International Journal of Developmental Biology*, 46(4), pp. 449–458. doi: 10.1387/IJDB.12141431.

Scekic-Zahirovic, J., Sendscheid, O., El Oussini, H., Jambeau, M., Sun, Y., Mersmann, S., Wagner, M., Dieterlé, S., Sinniger, J., Dirrig-Grosch, S., Drenner, K., Birling, M.-C., Qiu, J., Zhou, Y., Li, H., Fu, X.-D., Rouaux, C., Shelkovnikova, T., Witting, A., Ludolph, A. C., Kiefer, F., Storkebaum, E., Lagier-Tourenne, C., and Dupuis, L. (2016) 'Toxic gain of function from mutant FUS protein is crucial to trigger cell autonomous motor neuron loss', *The EMBO journal*. 2016/03/07. John Wiley and Sons Inc., 35(10), pp. 1077–1097. doi: 10.15252/embj.201592559.

Schrank, B., Götz, R., Gunnensen, J. M., Ure, J. M., Toyka, K. V., Smith, A. G., and Sendtner, M. (1997) 'Inactivation of the survival motor neuron gene, a candidate gene for human spinal muscular atrophy, leads to massive cell death in early mouse embryos', *Proceedings of the National Academy of Sciences of the United States of America*. The National Academy of Sciences of the USA, 94(18), pp. 9920–9925. doi: 10.1073/pnas.94.18.9920.

Selvaraj, B. T., Livesey, M. R., Zhao, C., Gregory, J. M., James, O. T., Cleary, E. M., Chouhan, A. K., Gane, A. B., Perkins, E. M., Dando, O., Lillico, S. G., Lee, Y.-B., Nishimura, A. L., Poreci, U., Thankamony, S., Pray, M., Vasistha, N. A., Magnani, D., Borooah, S., Burr,

K., Story, D., McCampbell, A., Shaw, C. E., Kind, P. C., Aitman, T. J., Whitelaw, C. B. A., Wilmut, I., Smith, C., Miles, G. B., Hardingham, G. E., Wyllie, D. J. A., and Chandran, S. (2018) 'C9ORF72 repeat expansion causes vulnerability of motor neurons to Ca²⁺-permeable AMPA receptor-mediated excitotoxicity', *Nature Communications*, 9(1), p. 347. doi: 10.1038/s41467-017-02729-0.

Sephton, C. F., Good, S. K., Atkin, S., Dewey, C. M., Mayer III, P., Herz, J., and Yu, G. (2010) 'TDP-43 is a developmentally regulated protein essential for early embryonic development', *Journal of Biological Chemistry*. Elsevier, 285(9), pp. 6826–6834. doi: 10.1074/jbc.M109.061846.

Shen, X., Ying, H., Qiu, Y., Park, J.-S., Shyam, R., Chi, Z.-L., Iwata, T., and Yue, B. Y. J. T. (2011) 'Processing of optineurin in neuronal cells', *The Journal of biological chemistry*. 2010/11/08. American Society for Biochemistry and Molecular Biology, 286(5), pp. 3618–3629. doi: 10.1074/jbc.M110.175810.

Shi, Y., Lin, S., Staats, K. A., Li, Y., Chang, W.-H., Hung, S.-T., Hendricks, E., Linares, G. R., Wang, Y., Son, E. Y., Wen, X., Kisler, K., Wilkinson, B., Menendez, L., Sugawara, T., Woolwine, P., Huang, M., Cowan, M. J., Ge, B., Koutsodendris, N., Sandor, K. P., Komberg, J., Vangoor, V. R., Senthilkumar, K., Hennes, V., Seah, C., Nelson, A. R., Cheng, T.-Y., Lee, S.-J. J., August, P. R., Chen, J. A., Wisniewski, N., Hanson-Smith, V., Belgard, T. G., Zhang, A., Coba, M., Grunseich, C., Ward, M. E., van den Berg, L. H., Pasterkamp, R. J., Trotti, D., Zlokovic, B. V., and Ichida, J. K. (2018) 'Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons', *Nature Medicine*, 24(3), pp. 313–325. doi: 10.1038/nm.4490.

Shim, W. J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., Peng, G., Jing, N., Wang, Y., Kellis, M., Tam, P. P. L., Smith, A., Piper, M., Christiaen, L., Nguyen, Q., Bodén, M., and Palpant, N. J. (2020) 'Conserved

epigenetic regulatory logic infers genes governing cell identity’, *Cell Systems*, 11(6), pp. 625–639. doi: <https://doi.org/10.1016/j.cels.2020.11.001>.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019a) ‘Embryoid body differentiation of mouse embryonic stem cells into neurectoderm and neural progenitors’, in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019b) ‘Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells’, *Frontiers in Physiology*, 10, p. 705. doi: 10.3389/fphys.2019.00705.

Sirohi, K., Chalasani, M. L. S., Sudhakar, C., Kumari, A., Radha, V., and Swarup, G. (2013) ‘M98K-OPTN induces transferrin receptor degradation and RAB12-mediated autophagic death in retinal ganglion cells’, *Autophagy*. Taylor & Francis, 9(4), pp. 510–527. doi: 10.4161/auto.23458.

Slowicka, K., Vereecke, L., Mc Guire, C., Sze, M., Maelfait, J., Kolpe, A., Saelens, X., Beyaert, R., and van Loo, G. (2016) ‘Optineurin deficiency in mice is associated with increased sensitivity to Salmonella but does not affect proinflammatory NF- κ B signaling’, *European Journal of Immunology*. John Wiley & Sons, Ltd, 46(4), pp. 971–980. doi: <https://doi.org/10.1002/eji.201545863>.

Stavridis, M. P., Collins, B. J., and Storey, K. G. (2010) ‘Retinoic acid orchestrates fibroblast growth factor signalling to drive embryonic stem cell differentiation’, *Development*, 137(6), pp. 881–890. doi: 10.1242/dev.043117.

Sterneckert, J. L., Reinhardt, P., and Schöler, H. R. (2014) ‘Investigating human disease

using stem cell models’, *Nature Reviews Genetics*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 15(9), pp. 625–639. doi: 10.1038/nrg3764.

Sun, Y., Shim, W. J., Shen, S., Sinniah, E., Pham, D., Su, Z., Mizikovsky, D., White, M. D., Ho, J. W. K., Nguyen, Q., Bodén, M., and Palpant, N. J. (2023) ‘Inferring cell diversity in single cell data using consortium-scale epigenetic data as a biological anchor for cell identity’, *Nucleic Acids Research*, p. gkad307. doi: 10.1093/nar/gkad307.

Sundaramoorthy, V., Walker, A. K., Tan, V., Fifita, J. A., Mccann, E. P., Williams, K. L., Blair, I. P., Guillemin, G. J., Farg, M. A., and Atkin, J. D. (2015) ‘Defects in optineurin- and myosin VI-mediated cellular trafficking in amyotrophic lateral sclerosis’, *Human Molecular Genetics*, 24(13), pp. 3830–3846. doi: 10.1093/hmg/ddv126.

Teoh, H. L., Carey, K., Sampaio, H., Mowat, D., Roscioli, T., and Farrar, M. (2017) ‘Inherited paediatric motor neuron disorders: Beyond spinal muscular atrophy’, *Neural Plasticity*. Hindawi, 2017, pp. 1–22. doi: 10.1155/2017/6509493.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., and McKay, R. D. G. (2007) ‘New cell lines from mouse epiblast share defining features with human embryonic stem cells’, *Nature*, 448(7150), pp. 196–199. doi: 10.1038/nature05972.

Teuling, E., Ahmed, S., Haasdijk, E., Demmers, J., Steinmetz, M. O., Akhmanova, A., Jaarsma, D., and Hoogenraad, C. C. (2007) ‘Motor neuron disease-associated mutant vesicle-associated membrane protein-associated protein (VAP) B recruits wild-type VAPs into endoplasmic reticulum-derived tubular aggregates’, *Journal of Neuroscience*, 27(36), pp. 9801–9815. doi: 10.1523/JNEUROSCI.2661-07.2007.

Tonge, P. D. and Andrews, P. W. (2010) ‘Retinoic acid directs neuronal differentiation of human pluripotent stem cell lines in a non-cell-autonomous manner’, *Differentiation*, 80(1),

pp. 20–30. doi: <https://doi.org/10.1016/j.diff.2010.04.001>.

Toth, R. P. and Atkin, J. D. (2018) ‘Dysfunction of optineurin in amyotrophic lateral sclerosis and glaucoma’, *Frontiers in Immunology*, 9, p. 1017. Available at: <https://www.frontiersin.org/article/10.3389/fimmu.2018.01017>.

Vangoor, V. R., Gomes-Duarte, A., and Pasterkamp, R. J. (2021) ‘Long non-coding RNAs in motor neuron development and disease’, *Journal of Neurochemistry*. 111 RIVER ST, HOBOKEN 07030-5774, NJ USA: WILEY, 156(6), pp. 777–801. doi: 10.1111/jnc.15198.

Walser, C. B. and Lipshitz, H. D. (2011) ‘Transcript clearance during the maternal-to-zygotic transition’, *Current Opinion in Genetics & Development*, 21(4), pp. 431–443. doi: <https://doi.org/10.1016/j.gde.2011.03.003>.

Wang, Z.-B., Zhang, X., and Li, X.-J. (2013) ‘Recapitulation of spinal motor neuron-specific disease phenotypes in a human cell model of spinal muscular atrophy’, *Cell research*. 2012/12/04. Nature Publishing Group, 23(3), pp. 378–393. doi: 10.1038/cr.2012.166.

Washington, J. M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M. D., Hamra, N., Semendric, L., Tan, B. S. N., Lake, J. A., Keough, R. A., Morris, M. B., and Rathjen, P. D. (2010) ‘L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture’, *American Journal of Physiology-Cell Physiology*, 298(5), pp. C982–C992. doi: 10.1152/ajpcell.00498.2009.

Xu, L., Liu, T., Liu, L., Yao, X., Chen, L., Fan, D., Zhan, S., and Wang, S. (2020) ‘Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis’, *Journal of Neurology*, 267(4), pp. 944–953. doi: 10.1007/s00415-019-09652-y.

Yeh, T.-H., Liu, H.-F., Li, Y.-W., Lu, C.-S., Shih, H.-Y., Chiu, C.-C., Lin, S.-J., Huang, Y.-C., and Cheng, Y.-C. (2018) ‘C9orf72 is essential for neurodevelopment and motility

mediated by Cyclin G1', *Experimental Neurology*, 304, pp. 114–124. doi: 10.1016/j.expneurol.2018.03.002.

Yeo, B. K. and Yu, S.-W. (2016) 'Valosin-containing protein (VCP): Structure, functions, and implications in neurodegenerative diseases', *Animal Cells and Systems*. Taylor & Francis, 20(6), pp. 303–309. doi: 10.1080/19768354.2016.1259181.

Ying, H. and Yue, B. Y. J. T. (2012) 'Cellular and molecular biology of optineurin', in Jeon, K. W. B. T.-I. R. of C. and M. B. (ed.) *International Review of Cell and Molecular Biology*. Academic Press, pp. 223–258. doi: 10.1016/B978-0-12-394305-7.00005-7.

Yon, J.-M., Baek, I.-J., Lee, S.-R., Jin, Y., Kim, M.-R., Nahm, S.-S., Kim, J.-S., Ahn, B., Lee, B. J., Yun, Y. W., and Nam, S.-Y. (2008) 'The spatio-temporal expression pattern of cytoplasmic Cu/Zn superoxide dismutase (SOD1) mRNA during mouse embryogenesis', *Journal of Molecular Histology*, 39(1), pp. 95–103. doi: 10.1007/s10735-007-9134-1.

Chapter 6: Final Discussion

Maintenance of cell identity and cell-fate decisions are required for ICM cells in the blastocyst to commit to formation of primitive ectoderm, definitive ectoderm, and neurectoderm sequentially, which sets the embryo up to produce the nervous system. The complex molecular mechanisms required must cooperate to ensure the correct cells in the correct number are produced at the appropriate times and locations. Few studies have looked at the pre- and early stages of nervous system development. This thesis investigates these early stages with an *in vitro* model using either MEDII or L-Proline, which recapitulates the production of the embryologically relevant populations required for the basic set up of the nervous system (Figure 6.1A) (Rathjen *et al.*, 1999, 2002; Washington *et al.*, 2010; Glover, 2018; Shparberg, 2018; Glover *et al.*, 2022).

Amino acids have only recently been recognised as growth factors (Washington *et al.*, 2010) and are now known to activate a variety of mechanisms to stimulate development/differentiation in mESCs and pre-implantation embryos (Washington *et al.*, 2010; Tan *et al.*, 2011, 2016; Shparberg *et al.*, 2019b; Morris *et al.*, 2020). In particular, L-Proline stimulates development when added during fertilisation as well as when added at the late 2-to-8-cell stages. It also promotes hatching of the blastocyst from the *zona pellucida* (Morris *et al.*, 2020; Treleaven *et al.*, 2021, 2023). In mESCs, it stimulates progression along the pluripotency spectrum, as well as directing lineage commitment to neural cells (Washington *et al.*, 2010; Casalino *et al.*, 2011; Shparberg *et al.*, 2019a; Glover *et al.*, 2022). Thus, like any growth factor, it acts at multiple stages of development, works in cooperation with other growth factors (such as LIF), and stimulates development by a variety of molecular mechanisms including cell signalling resulting in a variety of cellular responses. These responses include differentiation, changes in cell shape and motility, and changes in proliferation and apoptosis (Lonc, 2007; Casalino *et al.*, 2011; Comes *et al.*, 2013; D’Aniello *et al.*, 2015, 2017; Glover, 2018; Patriarca *et al.*, 2021). One particular interest in this thesis is that L-Proline’s metabolism, or changes in

metabolism that it modulates, also helps drive development; i.e., that changes in metabolism help drive cell fate and are not just a passenger to it.

Therefore, we used this *in vitro* L-Proline/MEDII mESC model to address the following question: What are the factors that drive cell fate and identity decisions for setting up a normal nervous system and can the model shed any light on the possible embryonic origins of the neurodegenerative disease, MND, which manifests later in life? Figure 6.1 provides a coarse-grained summary of some of the findings in this thesis.

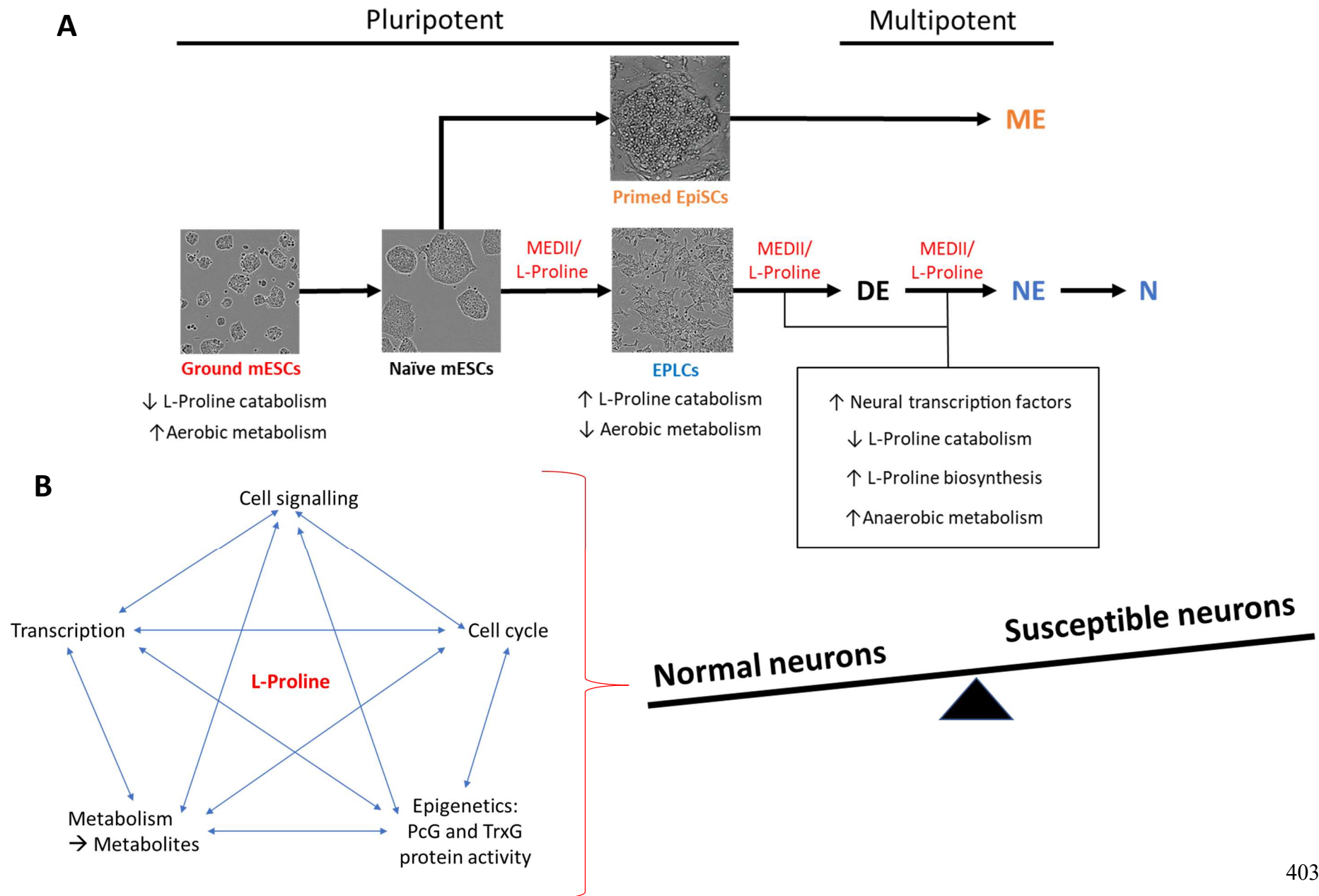


Figure 6.1 Summary of the proposed L-Proline-mediated mechanisms that drive neural fate commitment from mESCs. **A.** Ground-state mESCs represent ICM from pre-implanted blastocysts and the earliest stage of the pluripotency spectrum captured *in vitro*. Upon transition to naïve-state mESCs, addition of L-Proline causes differentiation towards an intermediate/formative stage of pluripotency (EPLCs), which is fated to form neural fate (N) via definitive ectoderm (DE) and neurectoderm (NE). This “priming” process is only required up towards the multipotent NE stage. In contrast, EpiSCs are set for mesendoderm (ME) lineage commitment. Metabolism changes across these various stages, including how L-Proline itself is metabolised. In addition, upregulation of genes encoding transcription factors, including those that drive neural development such as *Neurog1*, contribute to lineage commitment to neurectoderm and subsequent neural-cell production. **B.** Coarse-grained view of the various layers of regulatory control and their interconnection drawn from published work and the results from this thesis. L-Proline-mediated control of these layers of regulatory control favour normal neurons being developed and maintained. Disruption in one or more layers of these regulatory controls (e.g., by mishandling L-Proline) may compromise development, and in the case of MND combine with environmental factors predispose neurons to later dysfunction, thereby contributing to onset in later life.

Cell fate is in part driven by transcription factors whose gene expression is in part controlled epigenetically (MacArthur *et al.*, 2009; Kingston and Tamkun, 2014; Brand *et al.*, 2019; Harvey *et al.*, 2019; Shim *et al.*, 2020). In Chapters 3 and 4, we used TRIAGE transformation, which allows candidate genes to be selected which contribute to cell fate decisions. The advantage of this technique is it can be used on data sets which rely on relative changes in expression between cell types (e.g., microarray data sets) and on cell types without reference to any other cell type (e.g., RNAseq data sets). It effectively eschews genes which are high expressers and involved in cell structural integrity and housekeeping functions and narrows in on low expressers that mediate cell fate via the repressive epigenetic histone mark, H3K27me3 (Shim *et al.*, 2020).

This approach, together with analysis of changes in gene expression, including changes in gene expression mapped onto metabolic pathways, has provided a wealth of ‘intuitively sensible’ candidates for controlling cell fate. Many, as expected, are transcription factors but there are also cell cycle and signalling pathway genes, both of which are known to be modulated by the actions of L-Proline (Figures 3.5-3.9 and 4.6) (Lonc, 2007; Washington *et al.*, 2010; Tan *et al.*, 2011; D’Aniello *et al.*, 2015; Glover, 2018). In addition, some candidates are genes whose enzyme products are involved in metabolic pathway flux (Figures 3.16-3.23, 4.18 and 4.19), again suggest that L-Proline-mediated effects on cell metabolism help control cell fate.

Selecting candidates from any such list for further experiments remains a bottleneck in terms of experimental time and effort and we haven’t directly tested any of these candidates. However, because of the more directed/functional way the TRIAGE lists have been derived we might expect a higher ‘hit rate’ for showing their critical importance in controlling cell fate compared to more traditional approaches such as, for example, ranking candidates based solely on fold-change in expression. In addition, the selection can be narrowed, if necessary, by choosing gene products that lie near the centre of PPI networks, since this makes them more likely to be master regulators of cell fate rather than ‘slaves’.

The search for candidates can be narrowed in useful ways; for example, by focussing on epigenetic control circuits such as PcG and TrxG complexes, which are also involved in establishing lineage-commitment programs in pluripotent cells (Ringrose and Paro, 2004; Boyer *et al.*, 2006; MacArthur *et al.*, 2009; Wang *et al.*, 2016, 2012; Morey *et al.*, 2015; Wade *et al.*, 2015; Collinson *et al.*, 2016; Shan *et al.*, 2017; Sze *et al.*, 2017; Yao *et al.*, 2018; Brand *et al.*, 2019). Our results indicate that PcG genes are more likely to be involved in L-Proline-mediated set up of the nervous system than TrxG genes (Figures 4.13 and 4.14).

Any chance that MND-associated genes are associated with development of the disease later in life because of their expression early (or very early) in development depends, at least at the first pass, in establishing whether or not they are expressed at those times. Chapter 5 showed that many of these MND-associated genes/proteins are indeed expressed very early in development, and during commitment to neural-cell fate. This provides baseline expression patterns for future work, which can include examining any changes in expression when the system is also challenged with environmental stressors (e.g., oxidative stress) or when ESC lines are used which include specific genes with mutations known to be associated with MND, or both.

None of the MND-associated genes appeared to be drivers of development, as measured by TRIAGE transformation (Figures 5.2 and 5.5), but that may change when stressors are applied to the system and mutant cell lines are used. If so, this would further cement links between the role of genes and environment for this poorly understood, multifactorial umbrella of diseases. With this approach, the time and location (i.e., the cell type) of the first signs of the tipping point into disease progression, which normally is only observed several decades later, might be established. These are not limited to disorders associated with neural degeneration but can extend to other neurological disorders associated with neural wiring and connectivity, such as schizophrenia and autism.

6.1. References

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) ‘Polycomb complexes repress developmental regulators in murine embryonic stem cells’, *Nature*, 441(7091), pp. 349–353. doi: 10.1038/nature04733.

Brand, M., Nakka, K., Zhu, J., and Dilworth, F. J. (2019) ‘Polycomb/Trithorax antagonism: Cellular memory in stem cell fate and function’, *Cell Stem Cell*, 24(4), pp. 518–533. doi: <https://doi.org/10.1016/j.stem.2019.03.005>.

Casalino, L., Comes, S., Lambazzi, G., De Stefano, B., Filosa, S., De Falco, S., De Cesare, D., Minchiotti, G., and Patriarca, E. J. (2011) ‘Control of embryonic stem cell metastability by L-Proline catabolism’, *Journal of Molecular Cell Biology*, 3(2), pp. 108–122. doi: 10.1093/jmcb/mjr001.

Collinson, A., Collier, A. J., Morgan, N. P., Sienerth, A. R., Chandra, T., Andrews, S., and Rugg-Gunn, P. J. (2016) ‘Deletion of the Polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells’, *Cell Reports*, 17(10), pp. 2700–2714. doi: 10.1016/j.celrep.2016.11.032.

Comes, S., Gagliardi, M., Laprano, N., Fico, A., Cimmino, A., Palamidessi, A., De Cesare, D., De Falco, S., Angelini, C., Scita, G., Patriarca, E. J., Matarazzo, M. R., and Minchiotti, G. (2013) ‘L-Proline induces a mesenchymal-like invasive program in embryonic stem cells by remodeling H3K9 and H3K36 methylation’, *Stem Cell Reports*, 1(4), pp. 307–321. doi: 10.1016/j.stemcr.2013.09.001.

D’Aniello, C., Fico, A., Casalino, L., Guardiola, O., Di Napoli, G., Cermola, F., De Cesare, D., Tatè, R., Cobellis, G., Patriarca, E. J., and Minchiotti, G. (2015) ‘A novel autoregulatory loop between the Gcn2-Atf4 pathway and L-Proline metabolism controls stem cell identity’, *Cell Death And Differentiation*, 22, pp. 1094–1105. doi: 10.1038/cdd.2015.24.

D’Aniello, C., Habibi, E., Cermola, F., Paris, D., Russo, F., Fiorenzano, A., Di Napoli, G., Melck, D. J., Cobellis, G., Angelini, C., Fico, A., Blelloch, R., Motta, A., Stunnenberg, H. G., De Cesare, D., Patriarca, E. J., and Minchiotti, G. (2017) ‘Vitamin C and L-Proline antagonistic effects capture alternative states in the pluripotency continuum’, *Stem Cell Reports*. Elsevier, 8(1), pp. 1–10. doi: 10.1016/j.stemcr.2016.11.011.

Glover, H. J. (2018) *L-proline-induced transition of mouse ES cells to a spatially distinct primitive ectoderm-like cell population primed for neural differentiation*. University of Sydney.

Glover, H. J., Shparberg, R. A., and Morris, M. B. (2022) ‘L-Proline supplementation drives self-renewing mouse embryonic stem cells to a partially primed pluripotent state: The early primitive ectoderm-like cell’, in Osteil, P. (ed.) *Epiblast Stem Cells. Methods in Molecular Biology*. New York, NY: Springer US, pp. 11–24. doi: 10.1007/978-1-0716-2281-0_2.

Harvey, A., Caretti, G., Moresi, V., Renzini, A., and Adamo, S. (2019) ‘Interplay between metabolites and the epigenome in regulating embryonic and adult stem cell potency and maintenance’, *Stem Cell Reports*, 13(4), pp. 573–589. doi: 10.1016/j.stemcr.2019.09.003.

Kingston, R. E. and Tamkun, J. W. (2014) ‘Transcriptional regulation by Trithorax-group proteins’, *Cold Spring Harbor Perspectives in Biology*, 6(10), p. a019349. doi: 10.1101/cshperspect.a019349.

Lonic, A. (2007) *Molecular mechanism of L-proline induced EPL-cell formation*. University of Adelaide.

MacArthur, B. D., Ma’ayan, A., and Lemischka, I. R. (2009) ‘Systems biology of stem cell fate and cellular reprogramming’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 10(10), pp. 672–681. doi: 10.1038/nrm2766.

Morey, L., Santanach, A., Blanco, E., Aloia, L., Nora, E. P., Bruneau, B. G., and Di Croce, L. (2015) ‘Polycomb regulates mesoderm cell fate-specification in embryonic stem cells

through activation and repression mechanisms’, *Cell Stem Cell*. Elsevier, 17(3), pp. 300–315. doi: 10.1016/j.stem.2015.08.009.

Morris, M. B., Ozsoy, S., Zada, Matthew, Zada, Mark, Zamfirescu, R. C., Todorova, M. G., and Day, M. L. (2020) ‘Selected amino acids promote mouse pre-implantation embryo development in a growth factor-like manner’, *Frontiers in Physiology*, 11, p. 140. doi: 10.3389/fphys.2020.00140.

Patriarca, E. J., Cermola, F., D’Aniello, C., Fico, A., Guardiola, O., De Cesare, D., and Minchiotti, G. (2021) ‘The multifaceted roles of Proline in cell behavior’, *Frontiers in Cell and Developmental Biology*, 9, pp. 1–28. doi: 10.3389/fcell.2021.728576.

Rathjen, J., Lake, J.-A., Bettess, M. D., Washington, J. M., Chapman, G., and Rathjen, P. D. (1999) ‘Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors’, *Journal of Cell Science*, 112(5), pp. 601–612. doi: 10.1242/jcs.004613.

Rathjen, J., Haines, B. P., Hudson, K. M., Nesci, A., Dunn, S., and Rathjen, P. D. (2002) ‘Directed differentiation of pluripotent cells to neural lineages: homogeneous formation and differentiation of a neurectoderm population’, *Development*, 129(11), pp. 2649–2661. Available at: <http://dev.biologists.org/content/129/11/2649>.

Ringrose, L. and Paro, R. (2004) ‘Epigenetic regulation of cellular memory by the polycomb and trithorax group proteins’, *Annual Review of Genetics*. Annual Reviews, 38(1), pp. 413–443. doi: 10.1146/annurev.genet.38.072902.091907.

Shan, Y., Liang, Z., Xing, Q., Zhang, T., Wang, B., Tian, S., Huang, W., Zhang, Y., Yao, J., Zhu, Y., Huang, K., Liu, Y., Wang, X., Chen, Q., Zhang, J., Shang, B., Li, S., Shi, X., Liao, B., Zhang, C., Lai, K., Zhong, X., Shu, X., Wang, J., Yao, H., Chen, J., Pei, D., and Pan, G. (2017) ‘PRC2 specifies ectoderm lineages and maintains pluripotency in primed but not naïve

ESCs’, *Nature Communications*, 8, p. 672. doi: 10.1038/s41467-017-00668-4.

Shim, W. J., Sinniah, E., Xu, J., Vitrinel, B., Alexanian, M., Andreoletti, G., Shen, S., Sun, Y., Balderson, B., Boix, C., Peng, G., Jing, N., Wang, Y., Kellis, M., Tam, P. P. L., Smith, A., Piper, M., Christiaen, L., Nguyen, Q., Bodén, M., and Palpant, N. J. (2020) ‘Conserved epigenetic regulatory logic infers genes governing cell identity’, *Cell Systems*, 11(6), pp. 625–639. doi: <https://doi.org/10.1016/j.cels.2020.11.001>.

Shparberg, R. A. (2018) *L-proline-mediated neural differentiation of mouse embryonic stem cells*. University of Sydney.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019a) ‘Embryoid body differentiation of mouse embryonic stem cells into neurectoderm and neural progenitors’, in Joglekar, M. V. and Hardikar, A. A. (eds) *Progenitor Cells. Methods in Molecular Biology*. New York, NY: Humana, pp. 273–285. doi: 10.1007/978-1-4939-9631-5_21.

Shparberg, R. A., Glover, H. J., and Morris, M. B. (2019b) ‘Modeling mammalian commitment to the neural lineage using embryos and embryonic stem cells’, *Frontiers in Physiology*, 10, p. 705. doi: 10.3389/fphys.2019.00705.

Sze, C. C., Cao, K., Collings, C. K., Marshall, S. A., Rendleman, E. J., Ozark, P. A., Chen, F. X., Morgan, M. A., Wang, L., and Shilatifard, A. (2017) ‘Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation’, *Genes & Development*, 31(17), pp. 1732–1737. doi: 10.1101/gad.303768.117.

Tan, B. S. N., Lonic, A., Morris, M. B., Rathjen, P. D., and Rathjen, J. (2011) ‘The amino acid transporter SNAT2 mediates L-proline-induced differentiation of ES cells’, *American Journal of Physiology-Cell Physiology*, 300(6), pp. 1270–1279. doi: 10.1152/ajpcell.00235.2010.

Tan, B. S. N., Rathjen, P. D., Harvey, A. J., Gardner, D. K., and Rathjen, J. (2016) 'Regulation of amino acid transporters in pluripotent cell populations in the embryo and in culture; novel roles for sodium-coupled neutral amino acid transporters', *Mechanisms of Development*. Elsevier B.V., 141, pp. 32–39. doi: 10.1016/j.mod.2016.06.003.

Treleaven, T., Hardy, M. L. M., Guttman-Jones, M., Morris, M. B., and Day, M. L. (2021) 'In vitro fertilisation of mouse oocytes in L-Proline and L-Pipecolic acid improves subsequent development', *Cells*, p. 1352. doi: 10.3390/cells10061352.

Treleaven, T., Zada, M., Nagarajah, R., Bailey, C. G., Rasko, J. E. J., Morris, M. B., and Day, M. L. (2023) 'Stage-specific L-Proline uptake by amino acid transporter Slc6a19/B0AT1 is required for optimal preimplantation embryo development in mice', *Cells*. doi: 10.3390/cells12010018.

Wade, S. L., Langer, L. F., Ward, J. M., and Archer, T. K. (2015) 'MiRNA-mediated regulation of the SWI/SNF chromatin remodeling complex controls pluripotency and endodermal differentiation in human eSCs', *Stem Cells*, 33(10), pp. 2925–2935. doi: 10.1002/stem.2084.

Wang, C., Lee, J.-E., Cho, Y.-W., Xiao, Y., Jin, Q., Liu, C., and Ge, K. (2012) 'UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity', *Proceedings of the National Academy of Sciences*, 109(38), pp. 15324–15329. doi: 10.1073/pnas.1204166109.

Wang, C., Lee, J.-E., Lai, B., Macfarlan, T. S., Xu, S., Zhuang, L., Liu, C., Peng, W., and Ge, K. (2016) 'Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition', *Proceedings of the National Academy of Sciences*, 113(42), pp. 11871–11876. doi: 10.1073/pnas.1606857113.

Washington, J. M., Rathjen, J., Felquer, F., Lonic, A., Bettess, M. D., Hamra, N.,

Semendric, L., Tan, B. S. N., Lake, J. A., Keough, R. A., Morris, M. B., and Rathjen, P. D. (2010) 'L-Proline induces differentiation of ES cells: a novel role for an amino acid in the regulation of pluripotent cells in culture', *American Journal of Physiology-Cell Physiology*, 298(5), pp. C982–C992. doi: 10.1152/ajpcell.00498.2009.

Yao, M., Zhou, X., Zhou, J., Gong, S., Hu, G., Li, J., Huang, K., Lai, P., Shi, G., Hutchins, A. P., Sun, H., Wang, H., and Yao, H. (2018) 'PCGF5 is required for neural differentiation of embryonic stem cells', *Nature Communications*, 9(1), p. 1463. doi: 10.1038/s41467-018-03781-0.