Generating a proteomic profile of neurogenesis, through the use of human foetal neural stem cells

Shaun Garnett (B.Sc (Hons))

Supervisors:

Prof. Jonathan Blackburn

Prof. Susan Kidson

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

Integrative Biomedical Sciences

Faculty of Health Sciences

University of Cape Town

2019-10-17

Introduction Neurogenesis, the development of new neurons, starts soon after the formation of the neural tube and is largely completed by birth. Development of the brain after birth is mainly reliant on the formation of new connections between surviving neurons. However, adult neurogenesis does continue in the subgranular zone of the hippocampus from quiescent adult neural stem cells. Traditionally neural stem cells were cultured as neurospheres, a heterogeneous agglomeration of neural cells at various stages of differentiation. This heterogeneity prevented accurate quantitative analysis. In 2008 Sun et al produced the first non-immortalised human foetal neural stem (NS) cell line from nine week old human foetal cortex. These cells are cultured as monolayers, have a radial glia like appearance, self renew and form all three neural cell types, neurons, astrocytes and oligodendrocytes upon differentiation. More recently human foetal neuroepithelial like (NES) stem cells have been produced from five week old human foetal hind-brain, they resemble neuroepithelial cells, with characteristic rosettes, upon differentiation they appear to form a pure population of neurons. These homogeneous monolayer cultures enable quantitative proteomic analysis, to increase our understanding of early brain development

Methods Three NES and two NS cell lines were available for analysis. They proliferate by stimulation from FGF and EGF, removal of these growth factors results in spontaneous differentiation. Proliferating NES and NS cells were compared using SILAC labelling. In addition, each cell line was differentiated for 12 days, 6 timepoints were taken and compared using label free quantitation.

Results 4677 proteins were quantitated with 473 differentially expressed, revealing fundamental differences between NES and NS cells. NES cells are less differentiated, expressing SOX2 and LIN28, have active cell cycle processes, DNA elongation, histone modification and miRNA mediated gene silencing. Whereas NS cells are more developmentally defined, express multiple membrane proteins, have activated focal adhesion, thereby increasing their binding and interaction with their environment. NS metabolism is more oxidative, utilises lipid metabolism, the pentose phosphate pathway and produces creatine phosphate. Upon differentiation the cell cycle processes are downregulated and neurogenic and gliogenic processes increased.

Conclusion This work represent a detailed *in vitro* characterisation of non immortalised human foetal neural stem cells, it describes the regulatory, metabolic and structural changes occurring within neural stem cells in early brain development. The information herein points towards de-differentiation potentially through LIN28-*let7*, as a means to produce more neurogenic neural stem cells *in vitro* thus aiding regenerative therapies, as well as provides a wealth of information for better understanding neurological developmental disorders.

Keywords neural stem cells, neuroepithelia, radial glia

I dedicate this Thesis .....

To my family, friends and colleagues who supported me on this journey.

To my supervisor Prof. Jonathan Blackburn who provided me with all the time and resources I needed to do this work.

To Mrs Heather Muller who encouraged me to read when it was most hard, she opened a up a world of possibility, I will always be grateful for.

To the seekers

Live in wonder, live in awe

Life's a blessing, not a core

Your path is not found,

it is bazed through the uncertainty.

Be the change.

DECLARATION OF UNIQUE WORK

I Shaun Garnett hereby grant the University of Cape Town free licence to reproduce the above thesis, in whole or in part, for the purpose of research and declare that the above thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisor

Neither the substance nor any part of the above thesis has been in the past, or is being, or is to be submitted for a degree at this University, or any other University, except the data from Tailor et al. 2013 and Huang et al. 2016 which were both included as supporting evidence and declared as such within the text.

Shaun Garnett

Document Description

The typesetting in this document was done with LaTeX. The pdf document has been set-up with hyper-links to enable easy navigation through the digital document.

The table of contents in the preface, can be used to navigate sections. The table of contents should also be visible in the sidebar in most pdf viewers.

All the reagents, gene symbols, and references are hyper-linked. Hyper-linked to within the document are in blue, while links to external url's are in red.

Only reagent names are used in the text, hyper-links navigate to the full reagent information. Page numbers in the reagent list, navigate to where the reagent was mentioned in the text

Only HUGO gene nomenclature committee (HGNC) gene symbols are used throughout the text, hyper-links navigate to the full gene name information. A hyperlink within the gene symbol list also navigates to the boxplot for the relevant gene.

A link to the full data tables together with interactive plots is made available though an R shiny interface, available at : https://www.shaungarnett.co.za/thesis.html

Contents

List of Figures

List of Tables

Acronyms

σ standard deviation

quad quadrupole

LIN28 abnormal cell lineage 28

let-7 lethal-7

$_{hf}$NES cell human foetal neuroepithelial-like stem cell line

$_{hf}$NES cell differentiation experiment MaxLFQ $_{hf}$NES cell differentiation timecourse experiment

$_{hf}$NS cell human foetal neural stem cell-like cell line

$_{hf}$NS cell differentiation experiment MaxLFQ $_{hf}$NS cell differentiation timecource experiment

^{14}C Carbon-14 isotope

^{3}H-dT ^{3}H-Thymadine

2D gel two dimensional gel electrophoresis

5caC 5caCl5-carboxylcytosine

5fC 5-formylcytosine

5hmC 5-hydroxomethylcytosine

5mC 5-methylcytosine

ACN acetonitrile

AGC automatic gain control

AGO argonaute

AnimalTFDB animal transcription factor database

ANOVA one-way analysis of variance

ASAP complex apoptosis and splicing–associated protein complex

ATP adenosine triphosphate

BAF BRG/BRM-associated factors

BH Benjamini-Hochberg

bHLH basic helix-loop-helix

Blackburn lab Blackburn lab, IDM, University of Cape Town

BMP bone morphogenetic protein

BP biological process

BSA bovine serum albumin

c-trap curved linear ion trap

CC cellular component

CCN cyclin

CCP Cambridge centre for proteomics

CDK cyclin-dependent kinase

CDKN cyclin-dependent kinase inhibitor

centrifuge tube polypropylene centrifuge tube

CHD chromodomain

CID collision induced dissociation

CNS central nervous system

CSF cerebro-spinal fluid

Da Dalton

DDA data dependent acquisition

DIGE differential gel electrophoresis

DNA deoxyribonucleic acid

dNES cell neuroepithelial like stem cells derived from hESC cells *in vitro*

DPYSL dihydropyrimidinase-related protein

ECM extracellular matrix

EGL external granular layer

EJC exon junction complex

esBAF embryonic stem cell BAF

ESC embryonic stem cell

ESI electrospray ionisation

FASP filter aided sample preparation

FDR false discovery rate

FGF fibroblast growth factor

FWHM full width at half maximum

G1 first growth phase of the cell cycle

G2 phase second growth phase of the cell cycle

GE $\frac{\text{NS}}{\text{NES}}$ experiment Illumina bead chip microarrays, gene expression data comparing proliferating $\frac{\text{NS}}{\text{NES}}$

gene symbol human genome organisation gene nomenclature committee (HUGO/HGNC)[1] gene symbol

GFP green fluorescent protein

GO gene ontology

GRN gene regulatory network

H/L heavy/light

HA hyaluronic acid

HCD high-energy collisional dissociation

HES hairy and enhancer of split

hESC human ESC

hESC cell human ESC line

HGNC HUGO gene nomenclature committee

HMG high mobility group

HOX homeobox

HOX2 homeobox cluster B

HPLC high performance liquid chromatography

Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment $\frac{\text{dNES}}{\text{hESC}}$ experiment by Huang et al. 2016 [102]

ICAT isotope coded affinity tag

IEF iso-electric focusing

INO80 INO80 complex

iPSC induced pluripotent stem cell

IRX iroquois

ISWI imitation switch

iTRAQ® isobaric tags for relative and absolute quantitation

KEGG Kyoto encyclopedia of genes and genomes

LC liquid chromatography

LC/MS/MS liquid chromatography coupled to tandem mass spectrometry

LFQ label free quantitation

$\log_{10}$ logarithm base 10

$\log_2$ logarithm base 2

LTD long term depression

LTP long term potentiation

m/z mass/charge ratio

m/z mass/charge ratio

MALDI matrix assisted laser desorption ionization

MaxLFQ MaxQuant label free quantification algorithm

MBD methylated DNA binding domain

MCM complex minichromosome maintenance protein complex

MEIS myeloid ectropic viral integration site

MF molecular function

MHB midbrain-hindbrain boundary

miRNA micro RNA

MMP membrane metalloprotease

MQ MaxQuant

mRNA messenger RNA

MS mass spectrometery

MS mass spectrometer

ms^1 first mass spectrum scan

ms^2 tandem mass spectrum

MudPIT multidimensional protein identification technology

NA not available

nBAF neuron BAF

NCE normalised collision energy

ncRNA non-coding RNA

NMDA N-methyl-D-aspartate

npBAF neural progenitor BAF

NR2 nuclear receptor subfamily 2 or COUP transcription factors I/II

p-value calculated probability

PAX paired box

PBX pre-B-cell leukemia homeobox

PcG polycomb group

PDGF platelet-derived growth factor

PEP posterior error probability

pH potential hydrogen

PKNOX PBX/knotted homeobox

POU Pit-Oct-Unc domain transcription factor

PRM parallel reaction monitoring

PSAP complex Pinin containing ASAP complex

PSM peptide spectral matching

Q Exactive Thermo scientific^{TM}Q Exactive^{TM}hybrid quadrupole-orbitrap mass spectrometer

R R[171]: a language and environment for statistical computing

R-SMAD receptor SMAD

Ran Ras-related nuclear protein

RF radio frequency

RG radial glia

RNA ribonucleic acid

RP reverse phase

rRNA ribosomal RNA

RT retention time

S lens stacked-ring ion guide

S phase DNA synthesis phase of the cell cycle

SCX strong cation exchange

SDE significantly differentially expressed

SDS-PAGE SDS polyacrylamide gel electrophoresis

SGZ subgranular zone of the dentate gyrus in the hippocampus

SH-SY5Y cells differentiation experiment SH-SY5Y cells differentiation experiment

SILAC stable isotope labeling by amino acids in cell culture

SILAC 1 SILAC experiment 1

SILAC 2 SILAC experiment 2

SILAC 3 SILAC experiment 3

SILAC $\frac{\text{NS}}{\text{NES}}$ experiment SILAC experiment for the comparison of proliferating $\frac{\text{NS}}{\text{NES}}$

Smith lab Smith lab at the Wellcome trust-medical research council stem cell institute in Cambridge UK

snRNA small nuclear RNA

SOX sex determining region Y box

SPE solid phase extraction

SRM selected reaction monitoring

STKR serine/threonine kinase receptors

STRINGdb STRINGdb[73] package in R

SVZ subventricular zone of the lateral ventricles

SWI/SNF switching/sucrose non-fermenting subclass of TrxG proteins

t test two-sided Student's t test

TAD topological associated domain

TALE three amino acid loop extension

TBI traumatic brain injury

TC tissue culture

TET ten-eleven translocation

TF transcription factor

TFA trifluoroacetic acid

TIC total ion chromatogram

TMT Thermo scientific^{TM}tandem mass tag

TOF time of flight

topGO topGO[14]: enrichment analysis for gene ontology package in R

tRNA transport RNA

TrxG trithorax group

TSS transcription start site

UCT University of Cape Town

Uniprot the universal protein resource

UTR untranslated region

UV ultraviolet

UV-Vis ultraviolet-visible spectroscopy

weight01 weight01 algorithm in the topGO package[13] in R

Wilkinson lab Wilkinson lab

WNT wingless-type MMTV integration site family

Gene Symbols

ACIN1 apoptotic chromatin condensation inducer 1 : (ACINU)

ACTL6A actin like 6A : BAF53a (ACL6A)

ACTL6B actin like 6B : BAF53b (ACL6B)

ADNP activity-dependent neuroprotector homeobox

ALPL alkaline phosphatase, liver/bone/kidney : (PPBT)

ARID1A AT-rich interaction domain 1A : BAF250a (ARI1A)

ARID1B AT-rich interaction domain 1B : BAF250b (ARI1B)

ASCL1 achaete-scute family bHLH transcription factor 1

ATF3 activating transcription factor 3

BDNF brain-derived neurotrophic factor

BMP1 bone morphogenetic protein 1

BMP2 bone morphogenetic protein 2

BMP3 bone morphogenetic protein 3

BMP4 bone morphogenetic protein 4

BMP5 bone morphogenetic protein 5

BMPR2 bone morphogenetic protein receptor type 2

CADM1 cell adhesion molecule 1

CALB1 calbindin 1

CBX2 chromobox 2

CBX6 chromobox 6

CCNA1 cyclin A1

CCNA2 cyclin A2

CCNB1 cyclin B1

CCND1 cyclin D1

CCND2 cyclin D2

CCNJ cyclin J

CD44 CD44 molecule (Indian blood group)

CDC73 cell division cycle 73

CDH1 cadherin 1 : (CADH1)

CDH2 cadherin 2 : (CADH2)

CDK1 cyclin-dependent kinase 1

CDK2 cyclin-dependent kinase 2

CDK6 cyclin-dependent kinase 6

CER1 cerberus 1, DAN family BMP antagonist

CHAF1B chromatin assembly factor 1 subunit B : (CAF1B)

CHD7 chromodomain helicase DNA binding protein 7

CHRD chordin

CHURC1 churchill domain containing 1 : (CHUR)

COL6A1 collagen type VI alpha 1 : (CO6A1)

CTR9 CTR9 homolog, Paf1/RNA polymerase II complex component

CXCR4 C-X-C motif chemokine receptor 4

DCX doublecortin

DDX21 DEAD-box helicase 21

DGCR8 DGCR8 microprocessor complex subunit

DICER1 dicer 1, ribonuclease III : (DICER)

DNMT1 DNA (cytosine-5-)-methyltransferase 1

DNMT3A DNA (cytosine-5-)-methyltransferase 3 alpha : (DNM3A)

DNMT3B DNA (cytosine-5-)-methyltransferase 3 beta : (DNM3B)

DPF1 double PHD fingers 1 : BAF45b

DPF2 double PHD fingers 2 : BAF45d (REQU)

DPF3 double PHD fingers 3 : BAF45c

DPYSL2 dihydropyrimidinase like 2 : (DPYL2)

DPYSL3 dihydropyrimidinase like 3 : (DPYL3)

DPYSL4 dihydropyrimidinase like 4 : (DPYL4)

DPYSL5 dihydropyrimidinase like 5 : (DPYL5)

DROSHA drosha ribonuclease III : (RNC)

EBNA1BP2 EBNA1 binding protein 2 : (EBP2)

EGFR epidermal growth factor receptor

EGF epidermal growth factor

EGR2 early growth response 2

EN2 engrailed homeobox 2 : (HME2)

EXO1 exonuclease 1

EZH2 enhancer of zeste 2 polycomb repressive complex 2 subunit

FABP7 fatty acid binding protein 7

FEN1 flap structure-specific endonuclease 1

FGF2 fibroblast growth factor 2

FGF8 fibroblast growth factor 8

FST follistatin

GAP43 growth associated protein 43 : (NEUM)

GBX2 gastrulation brain homeobox 2

GDF10 growth differentiation factor 10

GDNF glial cell derived neurotrophic factor

GEMIN4 gem nuclear organelle associated protein 4 : (GEMI4)

GFAP glial fibrillary acidic protein

GSK3B glycogen synthase kinase 3 beta

H1F0 H1 histone family member 0 : (H10)

H1FX H1 histone family member X : (H1X)

H2AFX H2A histone family member X : (H2AX)

H2AFY2 H2A histone family member Y2 : (H2AW)

H2AFY H2A histone family member Y : (H2AY)

HDAC2 histone deacetylase 2

HELLS helicase, lymphoid-specific

HES1 hes family bHLH transcription factor 1

HES3 hes family bHLH transcription factor 3

HES4 hes family bHLH transcription factor 4

HES5 hes family bHLH transcription factor 5

HIST1H1A histone cluster 1, H1a : (H11)

HIST1H1C histone cluster 1, H1c : (H12)

HIST1H2BH histone cluster 1, H2bh : (H2B1H)

HIST1H3C histone cluster 1, H3c

HIST1H3J histone cluster 1, H3j

HIST1H4A histone cluster 1, H4a

HIST1H4J histone cluster 1, H4j

HIST1H4K histone cluster 1, H4k

HIST2H2AC histone cluster 2, H2ac : (H2A2C)

HIST2H2BE histone cluster 2, H2be : (H2B2E)

HMGA1 high mobility group AT-hook 1

HMGA2 high mobility group AT-hook 2

HMGB2 high mobility group box 2

HOXA2 homeobox A2 : (HXA2)

HOXA4 homeobox A4 : (HXA4)

HOXA5 homeobox A5 : (HXA5)

HOXB2 homeobox B2 : (HXB2)

HOXB3 homeobox B3 : (HXB3)

HOXB5 homeobox B5 : (HXB5)

HSPD1 heat shock protein family D (Hsp60) member 1 : (CH60)

IGF1R insulin like growth factor 1 receptor

IGF2BP1 insulin like growth factor 2 mRNA binding protein 1 : (IF2B1)

INHBB inhibin beta B

IRS2 insulin receptor substrate 2

IRX3 iroquois homeobox 3

ITGAV integrin subunit alpha V : (ITAV)

ITGB3 integrin subunit beta 3 : (ITB3)

ITGB4 integrin subunit beta 4 : (ITB4)

ITGB5 integrin subunit beta 5 : (ITB5)

ITGB8 integrin subunit beta 8 : (ITB8)

JARID2 jumonji and AT-rich interaction domain containing 2 : (JARD2)

KDM3B lysine demethylase 3B

KDM4D lysine demethylase 4D

KLF2 Kruppel-like factor 2

KLF4 Kruppel-like factor 4 (gut)

LAMA1 laminin subunit alpha 1

LAMA4 laminin subunit alpha 4

LAMA5 laminin subunit alpha 5

LAMB2 laminin subunit beta 2

LAMC3 laminin subunit gamma 3

LIF leukemia inhibitory factor

LIG1 DNA ligase 1 : (DNLI1)

LIN28A lin-28 homolog A : (LN28A)

LIN28B lin-28 homolog B : (LN28B)

MAGOHB mago homolog B, exon junction complex core component : (MGN2)

MAGOH mago homolog, exon junction complex core component : (MGN)

MAP2 microtubule associated protein 2 : (MTAP2)

MBD2 methyl-CpG binding domain protein 2

MBP myelin basic protein

MCM2 minichromosome maintenance complex component 2

MCM4 minichromosome maintenance complex component 4

MCM6 minichromosome maintenance complex component 6

MDK midkine (neurite growth-promoting factor 2) : (MK)

MECP2 methyl-CpG binding protein 2

MEIS2 Meis homeobox 2

MKI67 marker of proliferation Ki-67 : (KI67)

MLH1 mutL homolog 1

MSH2 mutS homolog 2

MSH3 mutS homolog 3

MSH6 mutS homolog 6

MTF2 metal response element binding transcription factor 2

MYCN v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog

MYC v-myc avian myelocytomatosis viral oncogene homolog

MYT1L myelin transcription factor 1 like

NANOG Nanog homeobox

NCAM1 neural cell adhesion molecule 1

NEFH neurofilament, heavy polypeptide : (NFH)

NEFL neurofilament, light polypeptide : (NFL)

NEFM neurofilament, medium polypeptide : (NFM)

NES nestin : NES

NEUROD2 neuronal differentiation 2 : (NDF2)

NEUROG2 neurogenin 2 : (NGN2)

NFIA nuclear factor I A

NFIB nuclear factor I B

NOG noggin : (G)

NR2F1 nuclear receptor subfamily 2 group F member 1 : (COT1)

NR2F2 nuclear receptor subfamily 2 group F member 2 : (COT2)

NUP153 nucleoporin 153kDa : (NU153)

NUP35 nucleoporin 35kDa : (NUP53)

NUP37 nucleoporin 37kDa

NUP50 nucleoporin 50kDa

OTX2 orthodenticle homeobox 2

PAX3 paired box 3

PAX6 paired box 6

PBX1 PBX homeobox 1

PCNA proliferating cell nuclear antigen

PDGFA platelet derived growth factor subunit A

PDGFB platelet derived growth factor subunit B

PDGFD platelet derived growth factor D

PDGFRB platelet derived growth factor receptor beta : (PGFRB)

PHF10 PHD finger protein 10 : BAF45a

PMS2 PMS1 homolog 2, mismatch repair system component

PNN pinin, desmosome associated protein : (PININ)

POLD1 polymerase (DNA) delta 1, catalytic subunit : (DPOD1)

POLD3 polymerase (DNA) delta 3, accessory subunit : (DPOD3)

POLD4 polymerase (DNA) delta 4, accessory subunit : (DPOD4)

POLE2 polymerase (DNA) epsilon 2, accessory subunit : (DPOE2)

POLE4 polymerase (DNA) epsilon 4, accessory subunit : (DPOE4)

POLE polymerase (DNA) epsilon, catalytic subunit : (DPOE1)

POLQ polymerase (DNA) theta : (D)

POU2F1 POU class 2 homeobox 1 : OCT1 (PO2F1)

POU3F1 POU class 3 homeobox 1 : OCT6/SCIP (PO3F1)

POU3F2 POU class 3 homeobox 2 : BRN2/OCT7 (PO3F2)

POU3F4 POU class 3 homeobox 4 : BRN4/OTF9 (PO3F4)

POU4F1 POU class 4 homeobox 1 : RDC-1 (PO4F1)

POU4F2 POU class 4 homeobox 2 : BRN3B (PO4F2)

POU5F1 POU class 5 homeobox 1 : OCT3/OCT4 (PO5F1)

PRTG protogenin

PXN paxillin : (PAXI)

RBBP4 retinoblastoma binding protein 4

RBFOX3 RNA binding protein, fox-1 homolog 3 : (RFOX3)

RCOR2 REST corepressor 2

RIF1 replication timing regulatory factor 1

RNPS1 RNA binding protein with serine rich domain 1

RUNX2 runt related transcription factor 2

SALL4 spalt-like transcription factor 4

SAP18 Sin3A associated protein 18kDa

SCML2 sex comb on midleg-like 2 (Drosophila)

SFRP2 secreted frizzled related protein 2

SHH sonic hedgehog

SIN3A SIN3 transcription regulator family member A

SLC1A1 solute carrier family 1 member 1 : (EAA3)

SLC1A2 solute carrier family 1 member 2 : (EAA2)

SLC1A3 solute carrier family 1 member 3 : (EAA1)

SMARCA2 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2 : BRM (SMCA2)

SMARCA4 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 : BRG (SMCA4)

SMARCC1 SWI/SNF related, matrix associated, actin dependent regulator of chromatin subfamily c member 1 : BAF155 (SMRC1)

SMARCC2 SWI/SNF related, matrix associated, actin dependent regulator of chromatin

subfamily c member 2 : BAF170 (SMRC2)

SMARCD1 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1 : BAF60b (SMRD1)

SMARCD2 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2 : BAF60a (SMRD2)

SMARCD3 SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3 : BAF60c (SMRD3)

SOX11 SRY-box 11

SOX18 SRY-box 18

SOX2 SRY-box 2

SOX3 SRY-box 3

SOX8 SRY-box 8

SOX9 SRY-box 9

SPP1 secreted phosphoprotein 1 : (OSTP)

SRC SRC proto-oncogene, non-receptor tyrosine kinase

SRSF2 serine/arginine-rich splicing factor 2

SS18L1 SS18L1, nBAF chromatin remodeling complex subunit : (CREST)

SS18 SS18, nBAF chromatin remodeling complex subunit : (SSXT)

STMN1 stathmin 1

SUPT16H SPT16 homolog, facilitates chromatin remodeling subunit : (SP16H)

TAF5 TATA-box binding protein associated factor 5

TERF1 telomeric repeat binding factor 1

TET1 tet methylcytosine dioxygenase 1

TGFB1 transforming growth factor beta 1

TGFBR2 transforming growth factor beta receptor 2 : (TGFR2)

TGFBR3 transforming growth factor beta receptor 3 : (TGBR3)

TLN1 talin 1

TLX3 T-cell leukemia homeobox 3

TNC tenascin C : (TENA)

TRIM71 tripartite motif containing 71, E3 ubiquitin protein ligase : (LIN41)

TUBB3 tubulin beta 3 class III : (TBB3)

VCL vinculin : (VINC)

VIM vimentin : (E)

WIBG PYM homolog 1, exon junction complex associated factor : WIBG

WNT1 Wnt family member 1

XPO5 exportin 5

YAP1 Yes associated protein 1

ZEB2 zinc finger E-box binding homeobox 2

ZFP42 ZFP42 zinc finger protein

ZYX zyxin

Part I

Introduction

Chapter 1

Nervous system development

1.1 Background

Neurogenesis, the development of new neurons, starts soon after the formation of the neural tube and in humans is largely completed by birth.[169] Development of the brain after birth is mainly reliant on the formation of new connections between surviving neurons, rather than the formation of new neurons. In 1965 it was discovered that neurogenesis continues into adult life.[16] Even though neural stem cell are found throughout the adult brain, they develop primarily into glial cells due to strict neurogenic inhibition. Adult neurogenesis is restricted to defined neurogenic niches,[194] within which neurons are allowed to form from neural stem cells.[99,194] Elsewhere in the brain neural stem cells are only able to differentiate into glial cells. In addition to taking place continuously in the neurogenic niches, neurogenesis may also be initiated in response to traumatic brain injury.[130,246] The discovery of adult neurogenesis has initiated much research into the neurogenic niches and the cells living in them.

Recently two homogeneous non-immortalised neural stem cells have been produced by the Smith lab at the Wellcome trust-medical research council stem cell institute in Cambridge UK (Smith lab) from five and nine week old human foetal hind-brain. These cells self renew and can be differentiated into neurons, astrocytes or oligodendrocytes.[206,211] These cells offer the ability to quantitatively study the developing brain in an *in vitro* model.

Proteins are the building blocks and functional components of all cells. Proteomics, the study of global protein expression and state, allows us to quantitatively compare the proteins changes occurring within a biological system. Thus enabling an assessment of the critical components within neural stem cells and the changes they undergo as they differentiate.

An increased understanding of neurogenesis is hoped to aid in the development of therapies for brain degenerative disorders.[137] Currently there are two distinct approaches to neurogenesis related therapies: through the activation of endogenous neural stem cells[239] by infusion of growth factors into damaged regions of the brain, alternatively, neural stem cells produced *in vitro* may be transplanted into damaged areas of the brain.[126,153] In both approaches it is hoped that integration of new neurons into damaged areas of the brain would return some of the functionality to those areas.

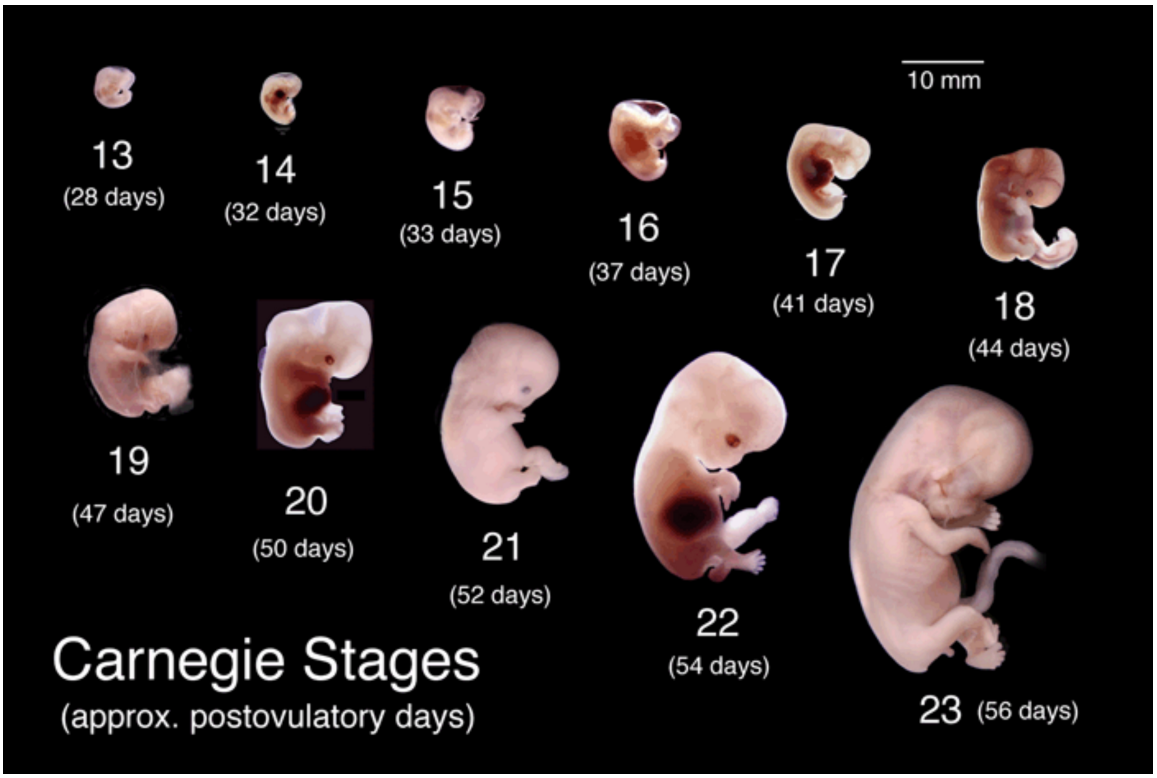
The focus of this Thesis is to use proteomics to contribute to the understanding of the intracellular mechanisms controlling neurogenesis.

1.2 Development

The eight totipotent stem cells of the morula have the potential to form an entire human body. Their DNA methylation and histone modification have largely been stripped besides for a few imprinted genes.[91,175,191] Their DNA does not hold any epigenetic marks specifying them towards any predetermined fate. The morula undergoes compaction, where the 8 cells are drawn close together, increasing cell surface contact.[105] This compaction is the start of cell distinction. The cells inside the morula are specified to form the inner cell mass, while those on the outside go on to form the trophectoderm. The morula continues to divide maintaining the polarity of the cells. Liquid is pumped into the blastocoel, creating a cavity, the blastocyst then contracts and hatches from its glycoprotein outer layer, the zona pellucida. The blastocyst then implants, and the embryo begins to develop, encapsulated within the blastocyst. A body of cells, the inner cell mass will go on to form the epiblast and then the embryo, while the main portion of the blastocyst the trophectoderm develops in the uterine sac.

Soon after implantation the embryo undergoes gastrulation, the development of the three tissue layers; the endoderm, mesoderm and ectoderm. At this point cells are becoming increasingly specified, and begin expressing markers specific for their tissue layers. They lose their pluripotency and become multipotent, restricted to form only the tissue later found within their respective layer. Cells in the embryo do not grow in isolation, as in tissue culture, but rather in intimate contact with the cells around them. Their position within the embryo is specified by their direct interactions with the cells around them and by the diffusion of growth factors from distant cells in the embryo. The embryo is largely symmetric along sagittal plane, left and right. The dorso ventral planes are specified by the secretion and diffusion gradients of *SHH* and bone morphogenetic proteins (BMPs).

The first eight weeks of human development is divided into 23 Carnegie stages. Carnegie stages describe the maturity of an embryo based on its external features. The stages are determined by multiple physical properties, not by age or size of the embryo and are therefore applicable across species (Figure 1.1). Within species post-ovulatory age is used to describe the age of the embryo, since the time of ovulation can be calculated and fertilisation should have occurred soon after ovulation.[96,190]



Species	Stage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Human	Days	1	2-3	4-5	5-6	7-12	13-15	15-17	17-19	20	22	24	28	30	33	36	40	42	44	48	52	54	55	58
Mouse	Days	1	2	3	4	5	6	7.0	8.0	9.0	9.5	E10	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16
Rat	Days	1	3.5	4-5	5	6	7.5	8.5	9	10.5	11	11.5	12	12.5	13	13.5	14	14.5	15	15.5	16	16.5	17	17.5

Figure 1.1: *Carnegie Stages are morphologically characterised embryonic developmental stages used to make comparisons between the embryonic development of different species. These stages are correlated to the embryonic ages within species, since embyonic age, calculated from the time of ovulation, commonly used to track development within a single specie. The cells used in this work, human foetal neuroepithelial-like stem cell lines ($_{hf}$NES cells) and human foetal neural stem cell-like cell lines ($_{hf}$NS cells), originated from Carnegie Stages 15 and 22 respectively.* [96]

1.3 Developmental neurogenesis

The focus of this Thesis was on neurogenesis, so the description of development will continue in the direction of the nervous tissue, disregarding the rest of the body. In gastrulation, the epiblast separates into three layers, through a process of tissue and cell movement. The ectoderm has two potential fates. *BMP4* induces the ectoderm to form epidermis. If *BMP4* signalling is inhibited the ectoderm defaults to nervous tissue. The notocord is the most dorsal layer of the mesoderm. It produces BMP inhibitors, *NOG*, *CHRD* and *FST* thereby causing the formation of neuro-ectoderm. Nervous tissue is the default state for ectodermal cells. In

fact, if embryonic stem cell (ESC) are dissociated *in vitro*, without growth factor stimulation, for several hours they begin forming neural tissue.[95]

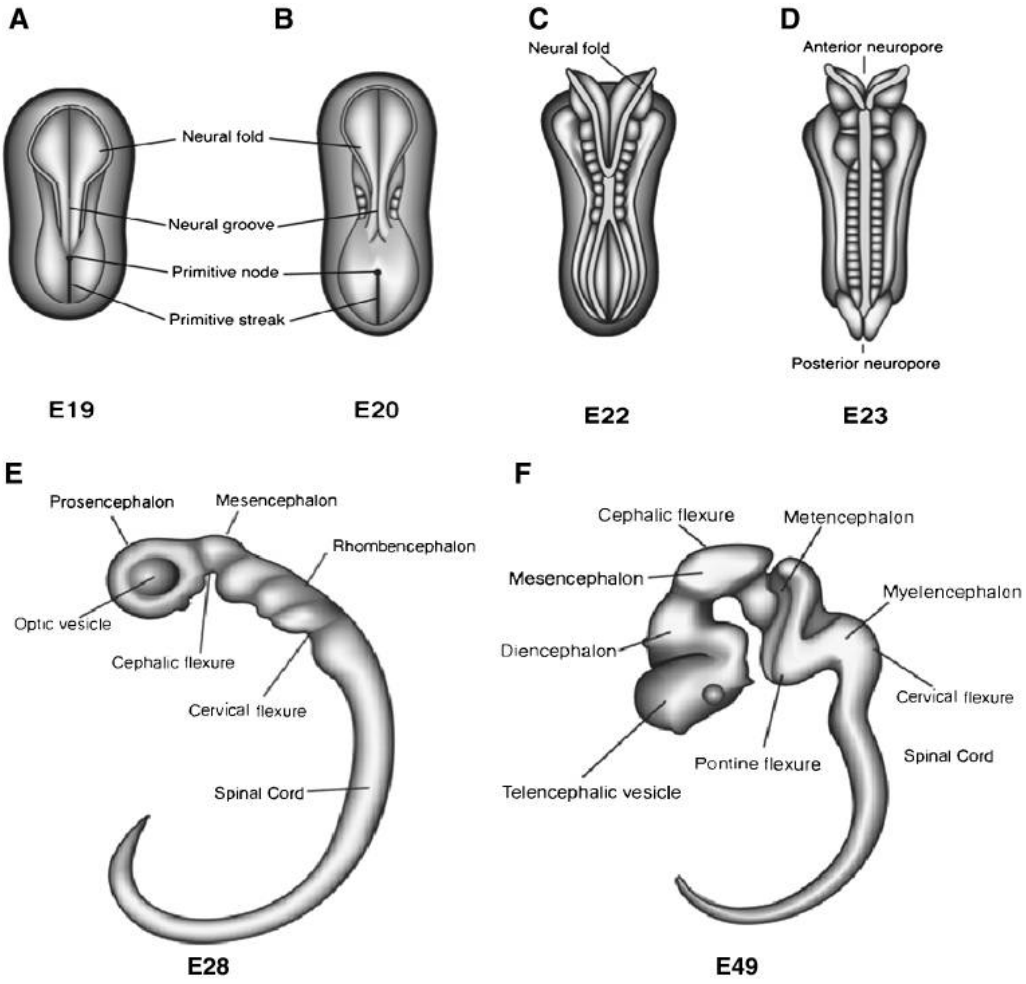


Figure 1.2: *Overview of the central nervous system (CNS) development taken from Stiles and Jernigan, 2010*[202] *(A,B) Initial development of the neural tube, (C,D) closure of the neural tube, (E) formation of the three main divisions of the brain, and the (F) subdivisions within these regions.*

At three weeks (Carnegie stage 9) the neuroectoderm folds along the antero-posterior axis, the sides of the folds connect and fuse forming the neural tube (Figure 1.2 A). The neural tube zips closed from the centre rostrally and caudally, eventually forming the brain and spinal cord respectively. The anterior neuropore (caudal) and the posterior neuropore (rostral) close at embryonic day 25 and 27 respectively,[183] (Carnegie stage 11). The cells connecting the neural tube to the epidermis become neural crest cells, which go on to form the peripheral nervous system and the melanocytes in the skin. Nervous system patterning occurs through a complex

interaction of growth factor gradients and patterning genes.[168]

The neuroepithelial cells, which make-up the neural tube initially divide symmetrically, producing identical neuroepithelial cells thus expanding the pool of cells prior to neurogenesis. They are the only proliferating cells in the early brain and they are connected to the apical lumen of the ventricles forming the ventricular zone thus stimulated by growth factors from the cerebro-spinal fluid (CSF). During the cell cycle the nucleus of neuroepithelial cells migrates between the apical to basal membranes, called interkinetic movement,[75] During the DNA synthesis phase of the cell cycle (S phase) the nucleus is furthest way from the neural tube lumen and the cells appear columnar. During the second growth phase of the cell cycle (G2 phase) the nucleus migrates towards the lumen. Changes in cell adherence during interkinetic movement allows the cells to become rounded before mitosis. Initially the columnar epithelial cells in S phase and the rounded cells undergoing mitosis were thought to be two different cells. Through ^{3}H-Thymadine (^{3}H-dT) labelling it was shown that there is only one homogeneous population of neuroepithelial cells in the early neural tube.[76,77]

Cells incorporate ^{3}H-dT during S phase, after a couple of hours the rounded mitotic cells also appear labelled. These labelled cells later loose contact with the lumen, migrate and differentiate as neurons. Varying the time of labelling made it possible to identify and study neurons born at different times.

Neuroepithelial cells initially divide symmetrically producing identical daughter neuroepithelial cell, but later go on to divide asymmetrically forming the first neurons of the preplate and deep brain layers.[183]

As development progresses a sequence of different neurons are produced in the cortex to make up the six layers of the brain. Cell fate switching appears to trigger intrinsic changes, as early neural progenitors are able to form superficial layer neurons, but late progenitors are unable to form deep brain layers.[170]

During development neuroepithelial cells gradually transform into radial glia (RG), this process most likely happens through "major differentiation" processes.[74,78,87] RG are initially neurogenic, but later become gliogenic.

RG start off as apical RG that undergo interkinetic movement during their cell cycle. If mitosis occurs when the nucleus is at the apical membrane, apical RG divide symmetrically producing additional apical RG. If the nucleus is at the basal membrane during mitosis, basal RG are formed, which loose their connection to the ventricle. These basal RG continue to proliferate forming secondary germinal zones, (*i.e* the subventricular zone of the lateral ventricles (SVZ), lateral to the lateral ventricles). The secondary germinal layers are responsible for gyrification or folding of the brain beginning in the first six months of gestation in humans. The surface of the brain is smooth in most animals, but gyrification occurs in some other primates, dolphins, elephants and pigs.[37,115,216]

Basal RG and apical RG also divide asymmetrically, producing daughter cells that differ from each other. Asymmetric division can be either proliferative, producing one identical daughter cell and another lineage restricted progenitor cell, or consumptively where only lineage restricted progenitor cells are produced. Lineage restricted progenitor cells may divide once or twice more but ultimately go on to terminally differentiate.[219]

Differentiating neurons migrate away from the germinal zone in three ways. Early in brain development differentiating cells migrate by stromal migration. They form a connection to the pial surface of the brain, the nucleus then migrates towards the pial surface and loses it's connection to the apical ventricle. Once the brain has become sufficiently thick, differentiating cells migrate radially along the radial glia which maintain contact between the ventricle or germinal zone and the pial surface. Differentiating neurons usually migrate to the superficial layers of the brain. The brain is therefore formed from the inside out. Neurons can also migrate tangentially, across the brain layers, rather than through them, directed by locally secreted trophic factors.[183].

1.3.1 Cerebellum development

The rostral part of the neural tube is initially divided into three sections the prosencephalon (fore-brain), the mesencephalon (mid-brain) and the rhombencephalon (hind-brain) (Figure 1.2 E). These section are compartmentalised by constrictions in the neural tube. Fluid pressure

build up due to the constrictions causes swelling of these areas. As development continues the three main brain regions are further subdivided into functional areas, specification of these areas is largely controlled by *homeobox (HOX) genes*. HOX genes specify the anterio-posterior axis in zones from the anterior to posterior of the embryo, and are numbered accordingly.[84]

Since the cells studied in this work originate from the hind-brain, the discussion of development will move in that direction. The hind-brain is separated from the mid-brain by the isthmus or the midbrain-hindbrain boundary (MHB). *FGF8* is expressed caudally of the MHB and *WNT1* is expressed rostrally.

The areas of the brain are separated by gene expression, *OTX2* is required for fore-brain and mid-brain development and forms the rostral boundary of the hind-brain. *OTX2* negative and *GBX2* positive areas develop into hind-brain, with homeobox cluster B (HOX2) forming the caudal boundary.

GBX2 appears to play a repressive role on *OTX2* rather than a inductive role on gene expression.[39] FGF8 expressed rostrally in the hind-brain is important for progenitor expansion, but its expression needs to be reduced in order to initiate maturation.[39]

The hind-brain divides into 8 transverse swellings or rhombomeres (R1 to R8). R1-R3 metencephalons forms the pons and the cerebellum, while R4-R8 forms the meyencephalon which forms the medulla oblongata (Figure 1.2 F).

Cerebellar cellular development occurs in waves, with each new wave inducing changes which brings about the induction of the next wave. There is no known molecular clock and the exact mechanism of this timing is still unknown. The ventricular zone, a germinal layer surrounding the ventricles is the first germinal layer. The first neurons to emerge from the ventricular zone are nuclear neurons of the cerebellar cortex.[86] Subsequent cells migrate through existing layers of differentiated cells. After the nuclear and Purkinje cells have finished dividing, a second germinal layer, the external granular layer (EGL) forms from the rhombic lip which extends from the roof of the forth ventricle.

1.3.2 The mature brain

Upon completion of the formation of the central nervous system, the brain consists of 100 billion neurons and one trillion glial cells. The survival of a neuron is dependent on its integration into a functional neural network. Survival is controlled by neurotrophins, which act on electrically active neurons and are released by the target cells of neurons. An activated target cell therefore ensures the survival of the cell that activated it.[151] Once the brain has matured few new neurons are formed. Additional complexity and learning mainly occurs through the the formation of new connections between the surviving neurons and adjustment to the excitability of these connections.

Long term potentiation (LTP) increases the excitability of neurons, and through long term depression (LTD) the excitability of neurons is decreased. LTP is triggered by rapid firing of the neurons, causing the release of excess glutamate, which hyper-depolarises the post synaptic membrane, causing opening of N-methyl-D-aspartate (NMDA) receptors by lifting the inhibition of magnesium. Open NMDA receptors allow calcium into the post synaptic synapse, which induces the insertion of more glutamate receptors into the post synaptic membrane. Long term gene expression changes due to LTP cause morphological changes, most markedly the growth of dendritic spines.[234] The synapses become more excitable, and thereby potentially store memory. This has been shown in *vivo* in a mouse model using high frequency stimulation. When neurons were excited with low frequency the process was reversed.[142]

1.4 Adult neurogenesis

Until recently it was thought that once the brain had matured no new neurons could be produced. This inhibition of neurogenesis is due to strict inhibitory signals within the adult brain. Adult neural stem cells are found throughout the brain and they have the potential to form neurons but are forced to become glial cells.

Through BrdU labeling experiments of rat brains, it was discovered that new neurons were being formed in the adult brains of rats.[16] These neurons were shown to survive for long pe-

riods of time,[110] receive synaptic input,[109] extend axonal projections[200] and enhance brain function.[152] Subsequently adult neurogenesis has been discovered in other vertebrates, mammals, primates and humans.[67] Furthermore is has been shown that human adult neurogenesis declines with increased age[35,66] and even more so in brain degenerative conditions[35] such as Alzheimer's disease.[138]

Neurogenesis has been shown to continue throughout life in two defined regions of the brain, the SVZ lateral to the lateral ventricles and the subgranular zone of the dentate gyrus in the hippocampus (SGZ).[137,194] These neurogenic niches consist of quiescent neural stem cells, often called RG, regulated by niche astrocytes specific to these regions.[99,194]

Niche astrocytes maintain the neurogenic environment through growth factors[48,107,122] and the cell to cell contacts.[106,123]

Rat studies have shown that neurogenesis in the SGZ produces new hippocampal granule cells. Which are involved in memory and learning.[11,167,187,189] Newly formed hippocampal granule cells have been shown to be electrically active, form synapses and integrate into existing networks, thus potentially aid in the formation of new memories.[63] New neurons formed in the SVZ migrate down the rostral migratory stream and become functionally integrated into the olfactory bulb.[53,63,89] These neurons survive for long periods of time, and may enhance the acuity of smell in rodents. Maternal SVZ neurogenesis has been shown to increase in anticipation of birth, enhancing a mothers ability to imprint on her young's smell.[186] Neurogenesis in the SVZ is also increased due to a number of different stimuli such as exercise in rats, caffeine[231] and anti-depressants.[228]

In animals neurogenesis can be induced outside of the neurogenic niches through traumatic brain injury or ischemia,[130,246] *LIF*, released by astrocytes surrounding a site of injury in mice, has been shown to induce proliferation of adult neural stem cells.[21] Some of these newly formed neural stem cells form new neurons[18,159] but most form glial cells. The adult brain's strict inhibition of neurogenesis[11] prevents large numbers of new neurons being formed.[26] The result of traumatic brain injury is thus the formation of a glial scar.

New neurons being formed after injury to the brain[18,131,144,159], shows that the brain has

the potential to partially regenerate itself, even if this capacity is greatly attenuated.

Animal studies involving cell transplantation, have increased understanding into the induction of neurogenesis and the role of the neurogenic niches. Neural stem cells transplanted out of the neurogenic niches are inhibited from forming neurons. Neural stem cells from non neurogenic brain areas such as the cortex, transplanted into neurogenic niches are induced to form neurons.[204] Niche astrocytes transplanted into non neurogenic niches induce neural stem cell from these areas to form new neurons.[107] Transfusion of growth factors such as *SHH*[10,107] and *BDNF*[162] have also been shown to induce neurogenesis in non neurogenic brain regions. Transplantation of neural stem cells derived from ESC grown in culture and later transplanted into non neurogenic regions of rat brains have been shown to differentiate, integrate and survive within existing neural networks.[136, 211]

Human brain tissue is only accessible post mortem, so most work on neurogenesis is either done *in vitro* or in animals, due the ease of access to these materials. Studies on the human brain therefore needs to be done in indirect ways. In the 1950s large quantities of Carbon-14 isotopes (^{14}Cs) were produced during extensive above ground testing of atomic bombs. The levels of atmospheric ^{14}C have been decreasing ever since the suspension of testing in 1963. This ^{14}C has a half life of 11 years.[196] ^{14}C is absorbed by plants consumed by humans and incorporated into DNA when cells divide. Comparison of atmospheric and cellular ^{14}C levels can therefore enable the birth dating of cells. Extensive study of cell nuclei isolated from brains of cadavers has allowed the determination of the rate and location of neurogenesis in the adult human brain.[196] In animals there is clear evidence for neurogenesis both in the SVZ and SGZ. Using ^{14}C birth dating young neurons were observed in the SGZ in humans,[29] but not in the SVZ.[195] All neurons in the SVZ and olfactory bulb originate at birth. No new neurons have been observed in other areas of the human brain including after ischaemic stroke.

Neurogenesis is thought to be inhibited in the adult brain due to its potentially disruptive effect on cognitive function.[203] Despite inhibition, induced neurogenesis is still a promising treatment for neuro-degenerative disorders. These new neurons may be produced from endogenous or transplanted neural stem cells. It is hoped that induced neurogenesis may aid

healing by increasing the number of neurons in damaged areas of the brain.[18] In order to accomplish neurogenic healing a detailed understanding of neural stem cells and the effects of the factors required to induce neurogenesis is necessary. In a review, Conti and Cattaneo stated with regards to *in vivo* neural stem cells "no direct and thorough comparison of their biological and molecular properties or of their physiological relevance and possible relationship to endogenous neural stem cells has yet been carried out", thus highlighting the importance for detailed study of these cells.[50]

1.5 Stem cells

1.5.1 Embryonic stem cells

ESCs are derived from the inner cell mass and are pluripotent, able to form all the cells of the body, but not a whole blastocyst. These cells can be cultured virtually indefinitely. Mouse ESC differ from human ESC in that human ESC require *FGF2* and *TGFB1*/Activin to proliferate while mouse ESC can be cultured with only inhibitors for *GSK3B* and *TGFB1*.

Human ESC are similar to epiblast cells while mouse ESC are similar to the developmentally earlier inner cell mass cells. Human ESC and can be developmentally reverted to resemble inner cell mass cells by a short induced expression of *NANOG* and *KLF2*, taking them one step back in the differentiation process.[215]

ESCs are pluripotent, therefore able to self renew and differentiate into the three primary germ cell layers. They proliferate on mouse fibroblast feeder layers. They can be induced to differentiate spontaneously through the removal of the pluripotent-dependent factors that stimulate self renewal,[81] forming embryoid bodies, which contain ESCs partially differentiated into all three germ cell layers and a central cavity of dead cells. They can also be induced to differentiate directly, through the addition of exogenous chemicals and factors. Neural progenitors are easily formed from ESC through their removal from feeder layers and sparse culture in the presence of fibroblast growth factor (FGF).[95]

1.5.2 Induced pluripotent stem cells

Recent advances in inducing pluripotency in adult cells has resulted in the development of robust techniques for turning adult fibroblasts into induced pluripotent stem cells (iPSCs), such as the insertion of the four main pluripotency factors *POU5F1*, *SOX2*, *MYC* and *KLF4* by adeno-virus. Expression of these exogenous genes induces expression of their endogenous counterparts. These genes induce pluripotency and embryoid bodies form from fibroblasts, which are harvested for expansion and validation. iPSC are initially slightly different to ESCs,[249] but the longer they are cultured, the more closely they begin to resemble ESCs, morphologically, transcriptionally and epi-genetically. Like ESCs, iPSCs can be directed to differentiate into a variety of different cells. Certain cell types, like neuronal tissue is fairly easy to produce. Cardiac myocytes also appear readily, and begin rhythmically contracting within cell cultures.Unfortunately most differentiation results in a heterogeneous mixture of cells. Without the highly specific 3D environment found within the developing embryo it is hard to specify cells perfectly. Differentiation of pluripotent stem cells into specific fates is thus an active field of research. iPSCs offer the opportunity to produce pluripotent cells from any adult patient and they are currently being used to study diseases. Cell models are produced for specific genetic disorders, and are used to increase understanding of the biochemistry involved. Thus enabling the identification and characterisation of the consequences of a faulty gene. Treatments can also be tested on these cells before going into clinical trials. In the near future iPSCs are hoped to be used to produce patient derived cells, or tissues, required to regenerate diseased, or damaged areas of the body.

1.5.3 Neural stem cells

Neural stem cell are proliferating cells of the nervous system which are able to self renew and can form all three nervous tissue cells; neurons, astrocytes and oligodendrocytes.[9] They are routinely cultured *in vitro*. Whether or not the neural stem cells cultured *in vitro* are artefacts of cell culture or are representative of endogenous neural stem cell in the brain is still under

some debate.[9,206]

Neurospheres

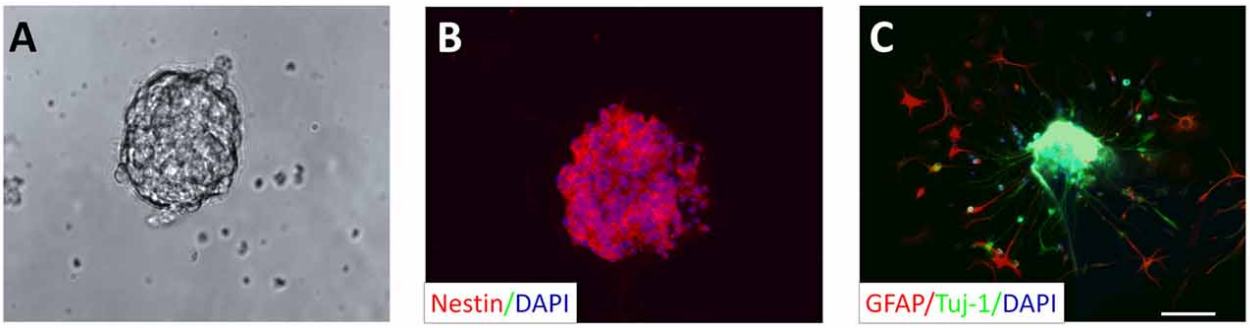


Figure 1.3: *(A) Neurospheres generated from adult rat hippocampus. (B) Immunoreactive for nestin (red) and nuclei stained with DAPI (blue). (C) After being differentiated for two days they display positive immunoreactivity for TUBB3 (green) and GFAP (red).*[205]

Neural stem cells are routinely cultured *in vitro* in the presence of *FGF2* and *EGF* as floating neurospheres (Figure 1.3).[50,145,146,209] Neurospheres are spherical agglomerations of nervous tissue cells at various stages of differentiation. Due to the spherical nature of neurospheres, a declining growth factor gradient is produced towards the interior of neurospheres.[36] Cells on the interior of the neurosphere are more differentiated than the peripheral cells. Removal of growth factors and transferral to adherent culture conditions induces differentiation and migration of the cells out of neurospheres.[8]

A neurosphere can be formed from either a neural stem cell or a neural progenitor cell.[207] Neural progenitors are neural stem cells that have become more restricted in their differentiation potential becoming bi-potent or uni-potent. Neurospheres cultured for long periods of time gradually lose the ability to generate neurons upon differentiation.[235] This may be due to a gradual transition of neurospheres containing neural stem cells to those containing glial progenitors.

One of the difficulties in studying neurospheres is that the heterogeneity of the neurospheres prevents accurate analysis of the discrete cells forming the neurospheres.[206] Quantitative changes occurring in neural stem cell within neurospheres are obscured by the hetero-

geneity of cells within neurospheres and the heterogeneity across neurospheres within the same culture. In order to quantitatively measure changes in neural stem cells a homogeneous culture of neural stem cells that maintains long term tri-potency is required.[206]

Additional source of neural stem cells

A homogeneous adherent population of neural stem cells can be produced through either immortalisation of neural stem cells[40,55] or by deriving neural stem cells from pluripotent stem cells.[58,185] Immortalisation requires the insertion and constant expressions of genes such as *MYC* into neural stem cells. These constantly expressed immortalisation genes may alter both regulation of the cell cycle and the cells differentiation potential. Neural stem cells can also be derived from ESCs or iPSCs, through directed or spontaneous differentiation, but will potentially be affected by the artificial *in vitro* environment in which they are derived. Both immortalisation and generation of neural stem cells from pluripotent stem cells runs the risk of creating a cell which mimics, but is not a true neural stem cell.

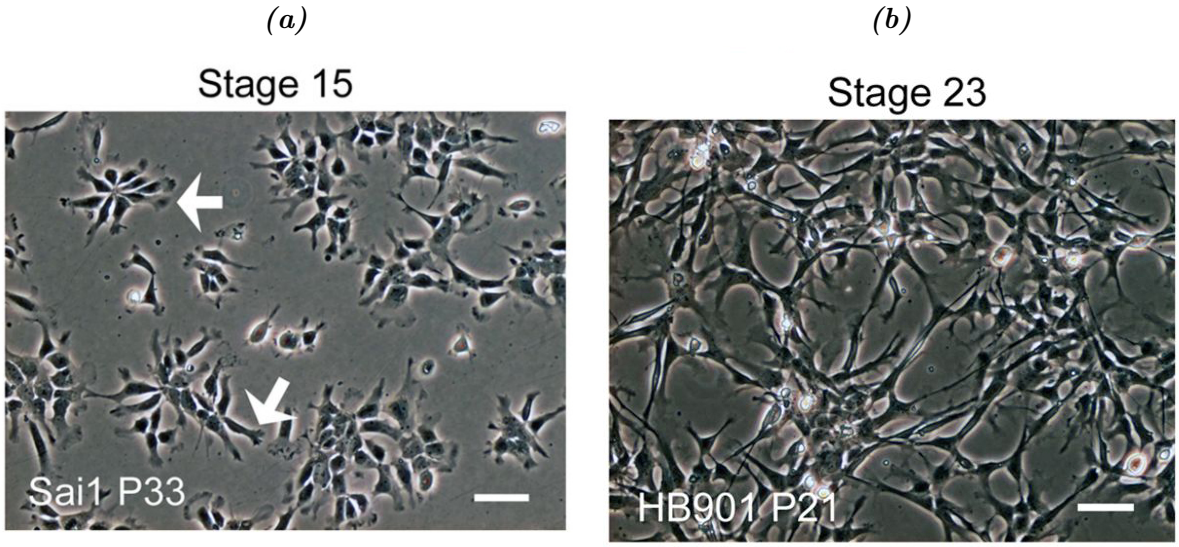


Figure 1.4: *Human foetal hind-brain neural stem cells cultured as monolayers from (a) Carnegie Stage 15 and (b) 22.*[211]

Homogeneous non-immortalised human neural stem cell In 2008, Sun et al. from the Smith lab produced the first non immortalised human neural stem cell line from nine week old

human foetal cortex (Carnegie Stage 22). The group used a serum free medium containing *FGF2* and *EGF* and alternated the use of laminin and gelatine as the substrate upon which the human foetal neural stem cells proliferated. These culture conditions allowed only the survival of neural stem cells, thus effectively separating them from differentiated progenitor cells. These $_{hf}$NS cells self renew and are genetically stable for up to 100 generations. They are tri-potent, thus able to form all three neural tissue cells. This homogeneous population of $_{hf}$NS cells may provide an accessible population of cells which can be used to model neurogenesis *in vitro*. $_{hf}$NS cells resemble *in vivo* RG cells in morphology, cell marker expression and differentiation potential. Upon differentiation $_{hf}$NS cells form a heterogeneous mixture of neurons and astrocytes with few oligodendrocytes, they can however be induced to form greater proportions of these cells through stimulation by *BDNF*, *GDNF* and platelet-derived growth factor (PDGF) respectively.

Neuroepithelial Stem Cells In 2013, Tailor et al. also from the Smith lab, produced another neural stem cell line, this time from five week old human foetal hind-brain (Carnegie Stage 15) using the same procedure as Sun et al. These cells have a epithelial cell like morphology and produce neural tube characteristic rosettes . They were thus named, $_{hf}$NES cells. These cells self renew, proliferating more rapidly than $_{hf}$NS cells. They are karyotypically stable for multiple generations and can be clonally cultured with conditioned media. Upon differentiation they form an almost pure population of neurons, both GABAergic and glutaminergic neurons. *In vitro* differentiated neurons from $_{hf}$NES cells were shown to mature over two months, but became synaptically active within two weeks, as measured by electrophysiology. Proliferating $_{hf}$NES cells are easily transformed to express exogenous proteins such as green fluorescent protein (GFP). Tailor et al. transplanted $_{hf}$NES cells into adult and embryonic mouse brains, and showed they were able to functionally integrate. $_{hf}$NES cells like cells called lt-NES have also been produced directly from ESC[64,65,83,247] and iPSC.[68] .

Chapter 2

Proteomics

2.1 Introduction

Genomics is the of study the of genomes, the entire genetic sequence of organisms. Sequencing of the human genome has enabled the identification of 20258 protein transcribing genes (Uniprot). Transcriptomics is the study of the rates at which these genes are expressed or transcribed into messenger RNA (mRNA). The mRNA is then translated into proteins, the main functional components of cells. Knowing the sequence of the genes has allowed accurate high throughput identification and quantification of the proteins produced by these genes, spawning proteomics, which is the study of all proteins produced by an cell, tissue or organism's genome.[225] A cell's transcriptome and proteome is under constant flux as it responds to environmental changes. Genes within the genome are identified *in vitro*, where there is physical evidence for their existence; mRNA or proteins produced by the genes have been identified. Genes are also identified *in silico*, where their sequence have characteristics similar to other known genes. Many of the genes identified *in silico* have been verified, some may not exist at all. The latest protein level count is 17694.[155]

Gene expression levels do not always correlate with protein abundances resulting in the proteome often differing from the transcriptome. The difference is due at least in part to post transcriptional regulation of protein synthesis and degradation. Protein abundance may

however be more conserved across species than gene expression levels.[116] This makes sense since proteins are the functional components and RNA is a messenger. Even though the message may differ slightly, the end result and goal may be maintained.

Measuring global changes in proteins, i.e. proteomics, is however more complex than measuring transcriptomic changes. mRNA has a simple chemistry, four RNA base-pairs with similar chemical properties forming long polymers. mRNA is easily enriched and can be amplified, thus increasing the overall quantity, without changing the differences between individual mRNA molecules too significantly.

Proteins are produced as linear chains of twenty different amino acids, each with a different chemical property. They can be acidic, basic, polar or non-polar. As proteins are synthesised the chemical properties of the amino acids cause them to fold into specific 3D conformations. The chemical environment of the cytoplasm in which most proteins are produced, affects their folding. Specific environments created within the endoplasmic reticulum, the Golgi apparatus and by protein chaperones induce additional protein conformations. The final conformation is an energetic folded minima, thus stabilising the final structure. The amino acids exposed on the surface of a protein molecule determine how it interacts with its surroundings. The 3D structure and surface chemistry produce the diverse functions of proteins required in a cell.

A protein's affect is influenced by both it's abundance and post-translational state. Phosphorylation for instance usually turns phosphoprotein activity on. There are hundreds of protein post-translation modifications which change the conformation, surface chemistry and activity of proteins. The diversity in protein chemistries influences the extraction, purification and fractionation of proteins. No single extraction and purification is sufficient for all the proteins in the proteome.

Unlike mRNA, proteins cannot be amplified *in vitro*. Protein quantity is therefore often a limiting factor in many studies. Cells are usually lysed in detergents or chaotropic agent to maintain protein solubility. Proteins can then be enriched through precipitation or filter aided sample preparation (FASP). Differential detergent and differential centrifugation allow access to different cellular compartments. Nano-scale liquid chromatography (LC) allows the

proteome to be separated into minute fractions. Mass spectrometer (MS) allows detection of proteins at femtomolar concentrations within these fractions.

The 20258 protein transcribing human genes produce roughly 70 000 protein isoforms and up to 1 million different post-translationally modified proteins. Proteins are expressed across 5 to 8 orders of magnitude in a single cell.[51] A few high abundant proteins constitute the bulk of cellular protein and obscure the identification of numerous low abundant proteins.[135] Currently half of the proteins, not including post-translational modifications, of simple multicellular organisms as can be covered by proteomic analysis[27] and this proteome coverage is constantly being increased. Many post-translational modifications such as phosphorylation can be detected by MS. An pre-enrichment of the modified proteins or peptides is usually required. Metabolic and chemical labelling and recent advances in label free quantification allows absolute and relative quantification between samples.

The are a considerable number of challenges in accessing the complete proteomic. Proteomics however offers a more detailed picture of the functioning of cells than transcriptomics. mRNA merely hints at the potential abundance of proteins, whereas proteomics is able to measure protein abundance and state.

History

Traditionally, proteins have been studied by a variety of molecular biology techniques. Usually these techniques study only a few proteins at a time. Proteins can be individually visualised using antibodies conjugated to fluorophores. Individual proteins can be quantitated using western blotting or localised through immunohistochemistry. Proteins can be isolated and their activity kinetics defined in a test tube. Although this is hugely important, the process is time and labour intensive. Individual proteins need to be selected based on pre-existing knowledge. Proteomics enables the study of proteins without any pre-defined selection. The hypotheses produced by proteomics then need to be validated using traditional molecular biology techniques.

The origin of proteomics is in two dimensional gel electrophoresis (2D gel). Proteins were

separated by iso-electric focusing (IEF) and then by size on SDS polyacrylamide gel electrophoresis (SDS-PAGE). The gels were then silver stained, to visualise the proteins as dark spots. Initially gels were run separately and then visually overlaid. Differences in abundance of proteins could be deduced from the intensity of the spots. Spots were excised and analysed by MS to identify them. Each spot usually contained multiple proteins, the most abundant contributing the majority of the spot. 2D gels provided information on differences in protein abundance. Post translational modification causes changes in protein mass and charge, thus shifting the spots. Due to large variability in the gels, numerous replicates were required for accurate measurements.

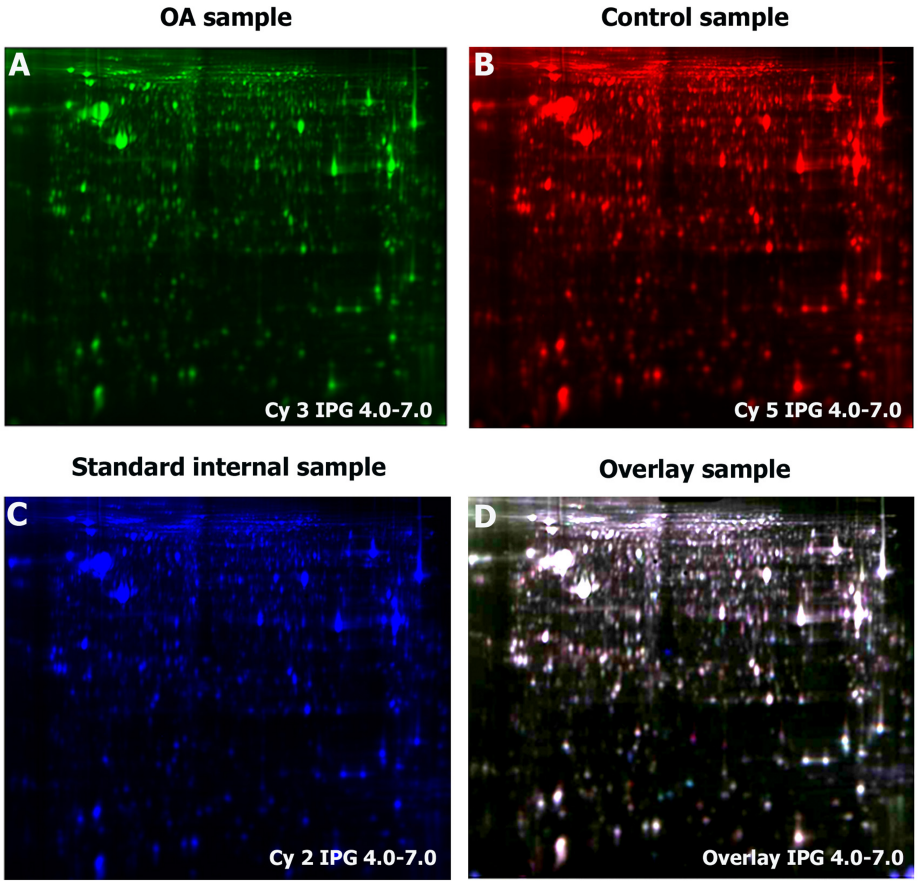


Figure 2.1: *Differential gel electrophoresis (DIGE).*[177]

In differential gel electrophoresis (DIGE) samples are labelled with up to three size and charge matched fluorescent dyes (Figure 2.1). This allows different conditions to be run on

the same gel. A reference sample can be run as the 3rd sample enabling better comparisons between gel replicates. 2D-DIGE is still routinely used but suffers from low resolution with the potential of only being able to identify 10-100s of proteins compared to thousands with liquid chromatography coupled to tandem mass spectrometry (LC/MS/MS). It however is still the only proteomics technique from which protein quantitation can be directly deduced.

2.2 Tandem mass spectrometry

2.2.1 Liquid chromatography

LC/MS/MS is now the cornerstone of proteomics. In LC a sample suspended in a mobile phase flows through an absorbent material called the stationary phase. The difference in interaction of the analytes within the mobile phase to the stationary phase affects their flow rate. The interaction can further be manipulated by changing the solvent composition of the mobile phase. Analytes are therefore separated as fractions in time, and can thus be analysed sequentially by MS.

In proteomics high performance liquid chromatography (HPLC) is generally coupled directly to a MS. HPLC columns are generally produced by filling narrow borosilicate glass capillaries with stationary phase silica beads. The most commonly used stationary phases in proteomics are reverse phase (RP), normal phase and strong cation exchange (SCX). They separate analytes based on hydrophobicity, hydrophilicity and positive charge respectively. The length and diameter of the column together with the size of the beads affects the resolving power of the column.

Sample is loaded onto the column, in the equilibration mobile phase, most of the analytes will bind to the stationary phase. Mobile phase continually flows through the column. Reverse phase separates analytes based on hydrophobicity, so increasing the percentage of organic solvent in the mobile phase causes elution of the analytes. The gradient describes the rate of this increased. As the percentage of organic solvent in the mobile phase is increased, analytes with low hydrophobicity become more loosely bound to the stationary phase. They flow down the

column, with further increases in organic solvent, analytes with greater hydrophobicity elute. Sample can be visualised eluting off the column using in-line ultraviolet-visible spectroscopy (UV-Vis). Sample can either be collected as a number of fractions or analysed continuously as they elute by electrospray ionisation (ESI) coupled MS. If conditions are the same (ie stationary phase, mobile phase, column length, temperature and gradient) analytes will have reproducible retention times.

More than one stationary phase can be used in series, if two types of separations are sufficiently different they are said to be orthogonal. No two stationary phases are truly orthogonal, but the term is still used if their separations are sufficiently different. For example, if a sample is first separated on SCX, peptide eluting together from SCX should not elute together when separated by RP. Different modes of separation can therefore be used one after the other to increase separation.

2.2.2 Mass spectrometry

All MS's consist of three components; an ion source, an analyser and a detector.

Ion Source The purpose of an ion source is to produce gas phase ions from analytes. Initial hard ionisation techniques fragmented the analytes when producing ions. More recent soft ionization techniques do little damage to the analytes during ionisation. The charges on ions allow them to be electromagnetically manipulated within the mass analyser. There are two basic soft ionisation techniques routinely used in proteomics: matrix assisted laser desorption ionization (MALDI) and ESI. The 2002 Nobel prize in chemistry was awarded to Koichi Tanaka and John B. Fenn for applying these techniques to macromolecular ionisation. thereby enabling their use in biology.

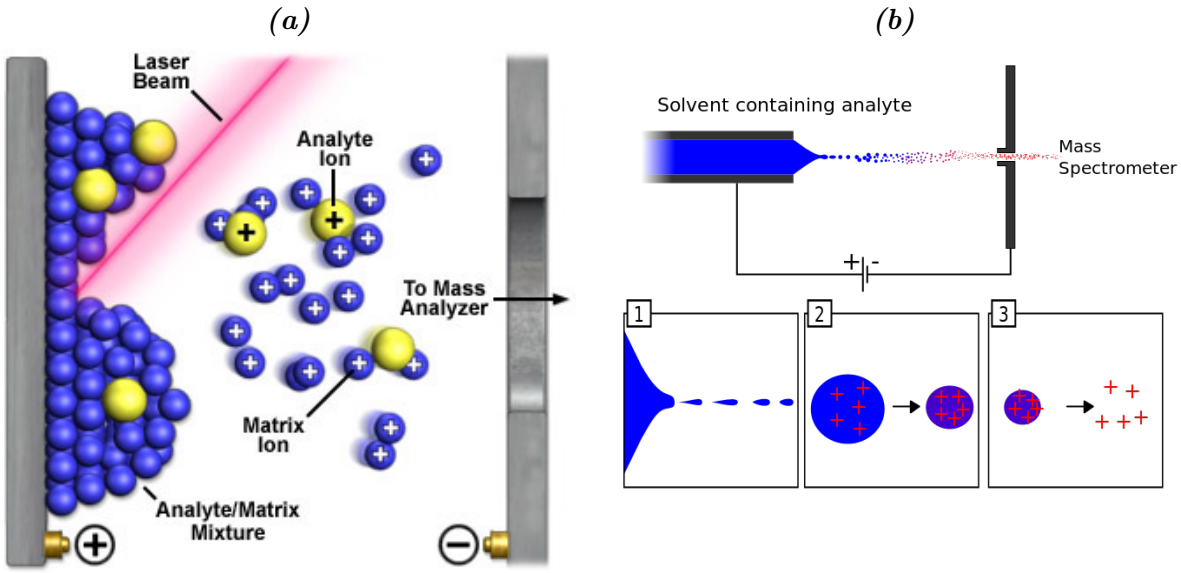


Figure 2.2: *(a) Matrix assisted laser desorption ionization.*[221] *and (b) Electrospray ionisation.*[54]

MALDI In MALDI analytes are deposited within an aromatic matrix. A laser is fired at the matrix triggering ablation and desorption of the matrix. During desorption charges are passed from the aromatic matrix to the analytes, ionising them. These ions are focuses electromagnetically and then accelerated into the mass analyser (Figure 2.2). MALDI was first coined by Karas et al. in 1985, while studying the ionisation of amino acids they discovered an increased desorption of aromatic amino acids. In 1988 Tanaka et al. applied it to marcomolecules making it amenable to biological work. Today organic matrices with nitrogen laser's are routinely used. MALDI works well for simple samples, such as peptide mass fingerprinting, where a purified protein is enzymatically digested and spotted on a MALDI plate. The masses of the peptides produced by digestion enables identification of the protein. MALDI can also be combined with LC, with drop sized fractions mixed with matrix spotted onto a MALDI plate for subsequent analysis. In addition MALDI is now being applied to tissue imaging, where tissue slices are coated in matrix and different areas fired with a laser, thereby identifying proteins occurring in those areas.

Electrospray ionisation ESI was first reported in 1984[237] and in 1989 John Bennet[70] applied it to the study of macromolecules. ESI is a soft ionisation technique that produces multiply charged ions. In ESI a sharp metal or metal coated emitter is coupled directly to the end of a HPLC column. A charge ≈2kV is applied to the emitter producing a Tailor Cone. Charged liquid droplets are sprayed towards the MS ion source which is heated (Figure 2.2). As droplets approach the MS the solvent begins to evaporate and the droplets become smaller. As the droplets shrink the coulomb repulsion forces of the like charges increase. The Rayleigh limit is a point at which the coulomb repulsion forces exceed the surface tension of the droplet. Two models have been proposed as to how gas phase ions are produced.

In 1968 Dole proposed the ion evaporation model, whereby the droplets become deformed and burst into smaller droplets, thus increasing the surface area, called Coulomb fission. This continues to happen until each droplet contains a single ion, from which the solvent then evaporates resulting in gas phase ions

An alternative charge residue model was proposed by Iribarne and Thomson in 1976. As the droplets decrease in size and the field strength increases, individual ions are pushed out of the droplets directly into the gas phase. Both mechanisms are possible depending on the sizes and properties of the analytes. There is still much debate in literature to the exact mechanism in peptides.[147]

Mass Analyser Mass Analysers control the trajectory of ions in a vacuum using electromagnetism. The ions reaction to the magnetic field is affected by their mass/charge ratio (m/z). The simplest mass analyser measures the time of flight (TOF) of ions electromagnetically accelerated through a vacuum. Ions with larger m/z fly slower than ions with smaller m/z. A transmission quadrupole (quad) utilises electromagnetic radio frequency (RF) to direct the flight of ions through parallel electromagnetic rods. RF changes to the charges on the rods stabilised or destabilised the path of the ions. The quad allows a range of m/z's through, narrowing this range enables selection of specific ions. Ion traps utilise similar technology, ions are trapped rather than transferred and can then be selectively ejected by altering the RF

frequency. An orbitrap consists of an outer barrel like electrode and an inner spindle shaped electrode. Ions are trapped around the spindle shaped electrode, the ions rotate around the spindle and oscillate along the axis. The axial frequency is independent of energy and position of the ions and can be used to deduce its m/z.[101,132]

Detectors Detectors record a charge from colliding or closely passing ions. The state of the mass analyser when the charge was produced determines the m/z of the ions detected (i.e the RF frequency that allowed the detected ions to pass through the quad). Typically the number of ions colliding with a detector is small so the signals are increased using electron multipliers. In an orbitrap the ions move close to the the detector, thus producing a charge. The detected wave form signal is converted through fourier transform into a spectrum of m/z peaks.

Mass Spectrometers There are many different types of MSs, with different configurations of ion sources, analysers and detectors. The different configurations enable a large range of applications. The simplest is a MALDI-TOF, a MALDI source is placed upfront of a TOF mass analyser, ions produced by laser ablation are drawn into and focussed by an electromagnet. The ions are then accelerated and their time of flight measured from time of acceleration to time of impact with the detector. This produces the first mass spectrum scan (ms^1), a series of m/z peaks for each ion. MALDI produces singly charged ions, so their time of flight is directly proportional to their mass. A single mass is usually insufficient to definitively identify an analyte. In order to increase their identifiability, analytes can be fragmented within the MS. In a MALDI-TOF/TOF, two TOFs are connected in parallel. In ms^1 mode parent masses are detected and the most intense peaks selected for fragmentation. The selected ions are isolated based on their time of flight and fragmented through collision with gas particles, called collision induced dissociation (CID). In CID peptide amide bonds reproducibly fragment between amino acids. Changing the fragmentation energy or conditions, changes the position of fragmentation (Figure 2.3). The fragments detected in the second TOF make-up the tandem mass spectrum (ms^2). The parent ion mass from the ms^1 scan can the be used in conjunction with the ms^2

scan to more definitively identify the detected analyte.

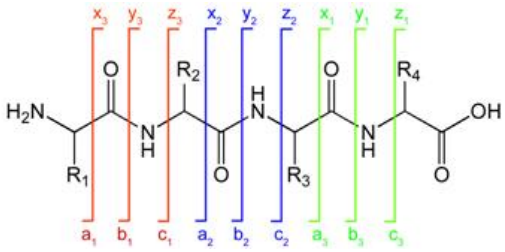


Figure 2.3: *Collision induced dissociation*[176]

Another common MS is a triple quad, where three quads are lined up in series: Q1, Q2 and Q3. A full scan produces ms^1. The Q1 can then be used to isolate selected ions, Q2 as a collision cell, and the Q3 to scan ms^2. Hybrid MSs use combinations of mass analysers, A Q-TOF combines a quad and a TOF. A Q-trap uses a quad and a ion trap. The LTQ orbitrap velos^{TM} uses an ion trap and an orbitrap. The Thermo scientific^{TM}Q Exactive^{TM}hybrid quadrupole-orbitrap mass spectrometer (Q Exactive) uses a quad and an orbitrap. The Thermo Scientific^{TM}Orbitrap Fusion^{TM} Tribrid^{TM}mass spectrometer uses a quad an ion trap and an orbitrap and can do parallel detection

2.2.3 Discovery proteomics

The aim of proteomics is the detection and quantification of proteins and their post translational modifications. In top-down proteomics, proteins are detected directly by LC-MS. MS mass accuracy is calculated as a percentage of the mass measured, so the larger the mass, the greater the degree of error. Proteins are therefore poor candidates for analysis because of their large masses. Large numbers of heavy carbon and nitrogen isotope containing amino acids produce a broad isotopic mass range further complicating identification. In bottom-up proteomics, proteins are enzymatically cleaved into peptides of 6 to 20 amino acids. These peptides have small enough masses to enable high mass accuracy. Their small isotopic series means the mono-isotopic peak can be detected, further increasing the accuracy. Peptides are fragmented within the MS to give a fragmentation spectra from which the peptide sequence can be deduced and the protein identity inferred.

Human cells contain more than 10000 proteins each. Digesting these into short peptides drastically increases the complexity of the samples to be analysed. Peptides in proteomic samples are therefore always separated by LC upstream of MS.

Currently most MS operate in a serial fashion, only able to do one measurement at a time. In discovery proteomics or data dependent acquisition (DDA), a top 10 or top 20 approach is usually used. 10-20 ms^2 scans are performed for every ms^1 scan. During a ms^2 scan only one peptide is being analysed, all other ions entering the MS are lost. The retention time of a peptide usually ranges from 10-60s. It would therefore need to be detected within this time or it would be lost. In a complex sample there is usually not sufficient time to analyse all the peptides present, therefore only the most intense peptides are selected for ms^2 scans. Not detecting a protein or peptide does not mean that is was not in the sample, only that the levels were to low to see in the time available.

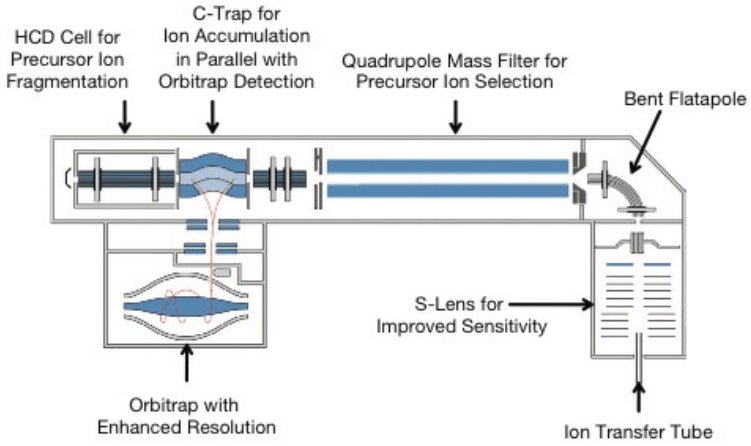


Figure 2.4: *Components of the Thermo scientific^{TM} Q Exactive^{TM} hybrid quadrupole-orbitrap mass spectrometer (Q Exactive)[92], Ions enter the ion source are focussed by the stacked-ring ion guide (S lens) and then passed through a bent flatapole, eliminating unchanged species. A select range of ions is allowed to pass through the quadrapole, they are then collected in the curved linear ion trap (c-trap). The c-trap focusses the ions and passes them either to the Orbitrap of detection m/z or it can pass them to the high-energy collisional dissociation (HCD) cell for fragmentation, after which the ions are returned to the c-trap and then passed to the orbitrap for detection*

Q Exactive The Q Exactive used in this study, is a hybrid MS, with a quad, c-trap, high-energy collisional dissociation (HCD) cell and an orbitrap

In a general top ten discovery experiment, peptide ions are continually injected into the

ion source throughout the LC gradient. In ms^1 mode the quad allows all ions within 300 m/z to 1750 m/z into the c-trap. The c-trap collects ions for a defined fill time or until it reaches a set automatic gain control (AGC) threshold. The AGC is determined by the number of charges filling the c-trap. The c-trap focuses the ions which are then passed to the orbitrap for detection. Ions spin around the spindle of the orbitrap oscillating from side to side, the frequency of oscillation is measured. Using fourier transform an accurate m/z is determined for the ions. The time of flight between the c-trap and the orbitrap is also taken into consideration to increase accuracy. In ms^1 mode the the Q Exactive is operated at a resolution of 70 000 full width at half maximum (FWHM), at 200 m/z and takes 256ms for each scan[166] (Figure 2.4).

Ten peptide ions with the greatest intensity are selected for a ms^2 mode scan. Each of the peptides is isolated and collected individually in the c-trap. This is achieved by reducing the mass range of the quad to 2 Dalton (Da) on either side of the selected peptide m/z. The AGC and fill time threshold on the c-trap is reduced, since with a narrow m/z range less ions are required. The isolated peptide ions are collected, transferred to the HCD cell, where the peptide ions are fragment through collision with nitrogen gas. The fragments are transferred back to the c-trap and then to the orbitrap for detection.

Peptide ions fragmented in the Q Exactive HCD preferentially produce a b/y ion fragmentation series (Figure 2.3). This fragmentation series is later used to generate an amino acid sequence for the isolated peptide through peptide spectral matching (PSM). All ten selected peptide ions are scanned one after the other and placed on a dynamic exclusion list. If the dynamic exclusion time is set to 30s, these peptide ions will not be selected for ms^2 for the next 30 seconds. The scan resolution in ms^2 mode is 35 000, resulting in a scan time of 64ms, with an additional 10ms for transfer and HCD time. The top ten cycle time is therefore approximately 1 second. Once all ten ms^2 scans are complete another ms^1 scan is performed, the ten peptide ions with the highest intensity not on the dynamic exclusion list are selected for another round of ms^2 scans.

Using upstream fractionation processes it is possible to increase peptide separation. With less peptides being eluted simultaneously, a greater percentage of the total number eluting

at any one time can be identified. It is therefore possible to dig deeper into the proteome. Measuring time and cost is however increased. Measuring time and sample quantity are the two main limiting factors on the depth of analysis.

2.2.4 Quantitative proteomics

DDA focuses on the detection of peptides and generates a qualitative list of proteins. Quantitative analysis is more useful since it enables us to determine differential expression of proteins within cells. In discovery proteomics greater than 3000 proteins can be routinely detected in a single experiment with only 1D fractionation, and 80% of these can generally be quantitated.

Traditionally quantitation was only performed through metabolic or chemical labelling of peptides. Despite labelled samples being multiplexed thereby reducing measuring time, the high purity required for metabolic and chemical labels often makes their use prohibitively expensive.

HPLC has increased LC resolution and retention time reproducibility. While mass accuracy and speed have increased in modern MSs. These advances have enabled the development of new algorithms for label free quantitation (LFQ). LFQ is becoming more and more prevalent. LFQ does not require expensive labels, but does however require each sample to be run at least once. Measuring time on MSs is expensive, and is often the limiting factor on the number of samples run in an experiment.

Labelling

Metabolic Labelling Peptides can be labelled in a number of different ways, the most robust form of labelling being metabolic labelling. Cells or whole organisms can be labelled. Cells are grown in tissue culture medium containing either heavy or light isotopes of the amino acids arginine and lysine. These amino acids contain isotopes of carbon or nitrogen with an additional neutron, which increases their mass by one Da, but does not alter the chemistry of heavy peptides. Heavy arginine and lysine are used because trypsin, the most common used protease in proteomics, cleaves peptides after arginine or lysine. This ensures that every

peptide always has a heavy labelled amino acid. Almost all protein's in the heavy labelled cells need to be replaced for full incorporation to occur. Five cell doublings, results in more than 95% of all arginine and lysine amino acids being heavy labelled. Heavy and light labelled cells are harvested counted and combined in equal ratios. All downstream processing is done on this combined sample, so sample processing bias does not affect quantitation. Upon MS analysis peptides from the heavy and light labelled cells, elute at the same retention times, since they have the same chemical properties. The heavy labelled peptides masses have been altered by the number of neutrons added to the heavy amino acid. This can range from 6-10 Da depending on the number of carbons and nitrogens in the amino acid. These mass shifts are easily distinguished by modern MSs. The difference in intensity between the two peaks is used to generate a measure of relative abundance for the peptides, which is later conferred to the proteins.

Only one of the peptide peaks needs to be sequenced for identification. The MS detection software be can programmed to ignore peaks of a specific mass shift. Stable isotope labeling by amino acids in cell culture (SILAC) is generally used in 2 plex experiments but can also be scaled up to 3 plex or more. This increase in complexity of the ms^1 spectra can reduce the depth of analysis.

Chemical Labelling Chemical labeling involves the addition of reporter groups onto proteins or peptides. In order to achieve complete enzymatic peptide cleavage, or digestion, it is important to first reduce disulphide bonds and block the cysteines with iodoacetamide. This process was easily amenable to labelling, through the use of isotope coded affinity tag (ICAT) used in place of iodoacetamide. Samples are combined after labelling and digested, with trypsin. The tags contain biotin, which is used to enrich the peptides containing ICAT. The tags fragment during ms^2. The relative intensities of the reporter fragments are used to deduce relative abundance of peptides and proteins from the respective samples. One of the drawbacks of the ICAT process is the small number of cysteine residues in proteins, so only a few peptides can be quantitated. However this is also an advantage since sample complexity

was also reduced in the process. In theory as long as a protein contained a cystein it could be quantitated.

The most commonly used forms of chemical labelling are isobaric tags for relative and absolute quantitation (iTRAQ®) or Thermo scientific™tandem mass tag (TMT). Cells are lysed, proteins extracted and peptides digested. Labels are then covalently bound to the amino groups of the amino acids, either to the terminal amino group or other free amino groups.

The labels consist of a reported group and a balance group. The labels have been designed to fragment between the reported group and the balance group during CID. The resultant reported groups can then be found in the lower end of the spectrum, below the masses of any amino acids. Each reporter group's mass is separated by 1 Da. Their relative abundance is determined from the ratio between their peaks in the ms^2 spectrum and relates to the relative abundances of proteins from the originating samples. iTRAQ® comes in either four or eight plex, while TMT has up to 12 labels. This multiplexing allows more samples to be compared simultaneously than with SILAC. Sample processing bias however may be introduced into the quantitation since the labelling only occurs after digestion. This variance can be measured by duplicating one of the samples or a pool of the samples, and using the deviation within this sample as a technical variance of the experiment, making any variance observed beyond this level biological variance.[28]

Label Free Quantitation

In SILAC quantitation is done by comparing the integrated area under the curve of the light vs heavy labelled peptides within the same ms^1 scan. In a label free discovery experiment a similar integration is done for every peptide and translated to an intensity value. These intensities values for peptides are then compared across different LC/MS/MS runs. The accuracy of LFQ quantitation is less than for labelled quantitation, due to increased variance caused by the measurements being taken across different LC/MS/MS runs. Some of the variance is accounted for by normalisation with algorithms like MaxQuant label free quantification algorithm (MaxLFQ).

Here the assumption is made that most of the proteins within a cell state are responsible for homoeostatic control and therefore do not vary greatly. Most peptides are then expected to be expressed in a 1:1 manner. MaxLFQ compares individual peptides in a pairwise manner and determines which peptides should have 1:1 ratios. It uses the offset in the ratios of these peptides to normalise the remaining peptides. A more reasonable distribution with a mean of zero is then achieved. This enables the peptides and proteins that are truly differentially expressed to be more clearly discriminated.

Accuracy can be improved by technical replicates, multiple LC/MS/MS runs of the same sample, which however increases measuring time and cost.

2.2.5 Targeted proteomics

If information about specific proteins is required, rather than acquiring all possible spectra, as in a DDA. The MS can be programmed to scan for specific masses at specific retention times, called targeted proteomics. Targeting peptides in this manner increases the sensitivity of detection 5-10 fold.[59] The accuracy of quantitation is also increased by ensuring that a sufficient number of points across the integration curve are detected per peptide. A list of m/z values for the peptides to be targeted is required. A fragmentation spectrum generated on the same MS aids identification. The simplest form of targeted proteomics simply adds an inclusion list to a DDA experiment. Ms^1 scans are performed, if the specified peptides m/z's are detected by ms^1 they are triggered for ms^2. In this mode selected peptides are given preference but other peptides not on the list can also be detected like in a discovery experiment. Quantitation is based on the ms^1 peaks.

Selected reaction monitoring (SRM) usually done in a triple-quad uses only ms^2 fragments ions called transitions. An isolation list of transitions is cycled through from top to bottom. The 1^{st} quad is set to select each peptide individually, the width of detection can be set between 0.5 and 4 Da on either side of the peptide. All ions making it through the Q1 are fragmented in Q2. Q3 isolates then the specified transitions for detection. Downstream software is used to determine if detected transitions originate from the proteins being targeted. Usually three

peptides and three transitions per peptide are required for a protein identification. Integration of ms^2 transitions is used for quantitation. The cycle time is the sum of the time it takes for each transition in the isolation list to be scanned. In order to achieve accurate quantitation, 3-10 scans of each transition are required during its elution time, thus limiting the total number of transitions that can be scanned per cycle.

In parallel reaction monitoring (PRM) using a Q Exactive, up to ten peptides are collected in the c-trap before before fragmentation. All the fragments are analysed simultaneously thus reducing cycle time and allowing more peptides to be targeted per run. ms^2 transitions are used for quantitation.

In SRMs and PRMs, if the elution time of a specific peptide is robust enough, an elution time range can be set in the inclusion list. A peptide or its transitions will only be targeted during this elution time range. This reduces the scan time in all other time ranges and allows more peptides to be added to the inclusion list.

A high accuracy peptide mass together with a robust elution time is sufficient for ms^1 only identification. Peptides need to be identified in by ms^2 fragments and then the elution time shown to be robust. Peptides can then be identified without fragmentation. It speeds up analysis, and improves quantitation accuracy.

2.3 Neural stem cell proteomic studies

Zizkova et al. 2015 wrote a review of the need and importance of proteomic research on neural stem cell development and differentiation. In it they describe a number of fundamental studies that have advanced the field, some of those are mentioned here.

Chaerkady and Kerr 2009 differentiated human ESCs (hESCs) into neural progenitors cells, motor neurons and astrocytes. The digested peptides from each population was labelled with iTRAQ® and analysed by LC/MS/MS. Their work may have been the first comprehensive proteomics analysis of human neural stem cells derived from ESC. 1185 proteins were quantifiable. The pluripotent markers *ALPL*, abnormal cell lineage 28 (*LIN28*) as well as two transcription

factors *HMGA1* and *SIN3A* were shown to be down-regulated during differentiation. They also identified a number of proteins that could be used as potential novel markers of neural stem cell differentiation.

In subsequent work by Chaerkady et al. 2011, hESC were differentiated to neural progenitors, glial restricted precursors and oligodendrocyte precursors. The digested peptides were labelled with iTRAQ® and analysed by LC/MS/MS. They identified 2956 protein groups. Their focus was the analysis of proteins showing differential expression during oligodendrocyte differentiation. They were able to verify many of the genes already discovered by transcriptomic analysis, highlighting *CD44* as important in directing oligodendrocyte fate.

Maltman et al. 2011 use a top-down proteomics approach to study the peptidome. Immortalised human neural progenitors were allowed to differentiate spontaneously by removing the growth factors from the media. All cells stained positive for *TUBB3* and *GFAP* indicating a mixed population of neurons and astrocytes. Protein was extracted from lysates using C_8 columns and analysed by MALDI-TOF/TOF. A number of intermediate filaments and cytoskeletal proteins were found up or down-regulated during differentiation indicating re-arrangement of the cytoskeleton during differentiation, together with *STMN1* a regulator of the cytoskeleton.

Cao et al. 2012 analysed membrane proteins from immortalised mouse neural stem cells differentiated into astroglial cells. The cells had been labelled with SILAC and were analysed by LC/MS/MS on a LTQ-Orbitrap. They identified more than 700 proteins and found that the differentiated cells membranes were far more complex that that of the neural stem cells. They also observed an activation of oxidative metabolism and concluded that this may be a signature of neural stem cell differentiation.

Abraham et al. 2013 analysed changes in neural stem cell using a mouse neurosphere assay. Second passage neurospheres were compared to eighth passage neurospheres by LC/MS/MS. Neurospheres loose their neurogenic potential with successive passages, so this would help identify potential gliogenic markers. Candidates were then further assessed by immunohistochemistry. They observed changes in high mobility group (HMG) proteins in neural progeni-

tors over time and during differentiation. In particular *HMGB2* was shown to decrease during development and may be controlled by lethal-7 (*let-7*).

Singec et al. 2016 differentiated hESC to a pure population of neural stem cell using a small molecule cocktail, to suppress non-neural differentiation. The neural stem cells produced were tri-potent able to be differentiated in the neurons, astrocytes and oligodendrocytes, indicating that were probably RG like cells. A phosphoproteomic comparison was made between the hESC and neural stem cells. They found that suppression of WNT signalling and the up-regulation of *MDK* was necessary for differentiation of hESC into neural stem cells.

A lot of research has been performed on neural stem cell development using proteomic tools. The work in this Thesis is the first to be conduction on neural stem cells derived from human foetal hindbrain and more particularly comparing $_{\text{hf}}$NES cells to $_{\text{hf}}$NS cells.

Chapter 3

Bioinformatics

3.1 PSM identification

LC/MS/MS data comes as ms^1 and ms^2 scans. Both scans are associated with a retention time (RT), the time the sample was eluted from the LC and detected by the MS. The ms^2 scans are associated with the m/z of the isolated ms^1 precursor, called the parent ion. Each scan consists of two data columns m/z values and intensities. The ms^1 scans are used for quantitation discussed in section 2.2.4, while the ms^2 scans are used to identify the peptides sequences. The software MaxQuant (MQ)[49,52] was used in this work for identification and quantitation. A protein sequence database was downloaded from the universal protein resource (Uniprot),[24] with the search criteria taxonomy number 9606 for *homo sapiens*. This protein sequence fasta file contains amino acids sequences for all known human proteins including their isoforms. It is curated and constantly updated with new information.

MQ generates theoretical spectra through an *in silico* digest of the proteins sequences within the fasta file. Trypsin is usually used because it produces the most appropriately sized peptides. Only peptides between 6 and 20 amino acids in length are used for the database search. Smaller peptides are two ambiguous to identify. Larger peptide suffer from mass accuracy and sequencing problems. Trypsin cleaves polypeptides after arginine and lysine residues unless the the next amino acid is a proline. Two missed cleavages are usually allowed

since the enzymes are never perfectly efficient *in vitro*.

A theoretical spectrum is generated for each of the peptides, with peaks representing a b/y ion fragmentation. MQ then uses the Andromeda search engine which is based on a similar algorithm as Mascot,[163] to match the detected ms^2 scans to the theoretical peptides. Using the isotopic mass shifts of the parent ion it's charge is determined and thereby its detected mass from the mass/charge ratio (m/z). Only theoretical spectra with masses matching that of the ms^2 scan's parent ion mass are compared. The peaks of the acquired ms^2 scans are correlated to the peaks in the theoretical spectrum, and the match is scored.

Due to the large search space, the number of theoretical peptides in the database, compared to the number of ms^2 scans, there is a high likely hood of some spectra matching theoretical spectra by random chance. To determine the number of random matches, a second round of matching is done against a decoy database. The decoy databases is generated by reversing the sequences of the original protein database. Any ms^2 scan matching the theoretical spectra from the decoy database is an incorrect match. A PSM match score cut-off is selected where 1% of matches originate from the decoy database. It is then surmised that up to 1% of the experimental data may be false positives, called the false discovery rate (FDR). It is not possible to determine which of the matches might be false positives. Only proteins identified by more than one peptide are used, so a few incorrect matches should not affect the data significantly.

In addition to the global FDR, a local FDR can be calculated on a peptide by peptide basis. This calculation uses the percentage of decoy hits with the same score as the peptide and is referred to as a posterior error probability (PEP). Global FDR is use when a list of proteins in required for downstream functional analysis. PEP is used when a confidence score is important for individual proteins identification.

MQ produces a table for every level of identification allowing users to track the steps the software went through in identifying proteins. The tables are structured like a relational database, with row numbers linking the various files. The MScan.txt and MSMScan.txt tables contain the ms^1 and ms^2 scan information respectively. All PSM's are in the evidence.txt table.

The peptide.txt table contains single entries for identified peptides together with quantitation information represented by a columns for each of the .raw files. The proteins identified need to be inferred from the peptides identified. Often proteins will be identified that share peptides with other proteins, especially if the database consists of many isoforms. When a protein or isoform cannot be uniquely identified it forms part of a protein group in the proteinGroups.txt table. Peptides shared by more than one protein are called razor peptides. Protein group quantitation is collated across multiple .raw files, by assigning the an experiment name to multiple .raw files. This can even be done across 2D fractionated samples.

3.2 Statistics

Once quantitative information is available for samples, the next step is to determine which of the proteins is significantly differentially expressed. In this work ratios were generate between conditions tested. The ratios were logarithm base 2 ($\log_2$) transformed. Plotting the $\log_2$ transformed data by histogram reveals a normal distribution, distributed evenly across zero. A standard deviation (σ) was then calculated. The σ of the replicate data represents the degree of technical variance in the data. According to the Empirical Rule, 95% of data points fall within 2σ. This was selected as the extent of technical variance. Unfortunately only using the mean $\log_2$ ratio is not sufficient to determine significant differential expression. In some cases the variance between replicates is so great that their mean is misleading. In order to test for significance the two-sided Student's t test (t test) was used. The t test determines if the means of two sets of data points are significantly different, while taking the variance into account. Data with too great a variance are negatively scored. The t test returns a calculated probability (p-value), below 0.05 is accepted as significantly differentially expressed.

3.2.1 Multiple testing correction

Using a p-value of 0.05 means that 5% of the tests done may be incorrect by random chance[150]. The likelihood of getting false positives due to random chance is greatly increase by the number

of tests being done. In order to counteract this multiple-testing correction is done. It takes the number of tests being done into consideration and adjusts the p-value accordingly.

Bonferroni adjustment The Bonferroni adjustment is the most widely used and simplest correction method, values remain significant if the p-value divided by the number of tests is still significant.[150] This minimises to the family-wise error, but it often too strict. Bonferroni adjustment will often result in no data being significant, so less conservative approaches need to be applied.

False discovery rate (FDR) FDR correction allows a percentage of positive results to be false. This can be done by comparing real values against values randomly generated, or through the use of the Benjamini-Hochberg (BH) procedure which uses the p-value distribution. The p-values need to be uniform exponentially decaying distribution.[150] The p-values are sorted in ascending order and then divided by the percentile mark, to give the estimated FDR.[150]

In proteomics multiple-testing correction is often problematic due to a number of different factors. Due to the high cost, the analysis power is often low. Often only a small percentage of the data will show an effective, change and the effect size might be small.[160] If the p-value value distribution is exponentially decaying, this is usually a good sign that there is a measurable change. Multiple-testing correction will often still however render all the resulting p-values not significant. It is a blunt tool, but can still be used to reduce false positives if no downstream targeted experiments are planned.

Removing some of values before performing t test can reduce some of the losses due to multiple-testing correction. It is however important not to remove values that would inadvertently bias the t test.

Chapter 4

Aim

Proliferating $_{hf}$NES cells have a neuroepithelial cell like morphology and appear to differentiate into an almost pure population of neurons. $_{hf}$NS cells have a RG like morphology and differentiate into a mixed population of neurons and glia. It is hypothesised that there are specific control mechanisms responsible for producing this difference and that through proteomic analysis it may be possible to identify these mechanisms. It is further hypothesised that identifying the specific mechanisms responsible for maintaining $_{hf}$NES cells cellular state, may produce candidate molecules for the development or brain regenerative therapies.

The project is divided into two experimental phases.

Neural Stem Cell Proliferation Comparison Quantitative comparison of the proteomes of $_{hf}$NES cells and $_{hf}$NS cells.

Neural Stem Cell Differentiation Analysis Quantitative assessment of the proteomic changes occurring in $_{hf}$NES cells and $_{hf}$NS cells as they undergo differentiation.

Part II

Methods

Reagents

100μm fused silica capillary

fused silica capillary, 100 μm ID (Next advance CAP-100) 61

10cm TC dish

10cm TPP tissue culture dish (Sigma Z707686) 44, 45, 58

15ml falcon tube

15ml falcon tube 44

24 well TC plate

Corning® Costar® cell culture plate, 24 well, flat bottom (Sigma CLS3527) 64

30kDa MWCO filters

molecular weight cut off filters 47

5μm SCX beads

strong cation exchange bulk packing material (Phenomenex 04A-4398) 61

5μm luna C18 beads

5μm luna C18(2) bulk packing material (Phenomenex 04A-4398) 49, 61, 66

6 well TC plate

Corning® Costar® cell culture plate, 6 well, flat bottom (Sigma CLS3516) 58

6cm TC dish

6cm TPP tissue culture dish (Sigma Z707678) 58

96 well flat bottom plate

Greiner® polystyrene 96 well plates, flat bottom (Sigma M3061) 46

ABC

0.05M ammonium bicarbonate (Sigma A6141) 47

Alexa 448 (Green) anti-mouse IgG2A

goat anti-mouse IgG2A Alexa Fluor® 488 65

Alexa 555 (Red) anti-mouse IgG1A

goat anti-mouse IgG1A Alexa Fluor® 555 65

Alexa 555 (Red) anti-sheep

donkey anti-sheep IgG Alexa Fluor® 555 65

ammonium acetate

ammonium acetate (Merck SAAR1122220EM) 61, 62

ammonium formate

ammonium formate 59

anti-GFAP

sheep anti-GFAP (R&D Systems, 1:50) 65, 81, 82, 92–95

anti-Nestin

mouse IgG1A anti-Nestin (R&D Systems, 1:200) 64, 65, 81, 82, 92–94

anti-TuJ-1

mouse IgG2A anti-β-III-Tubulin (R&D Systems, 1:200) 64, 65, 81, 82, 92–95

B

ACN 0.1% FA 49, 61, 62

BCA assay

Pierce bicinchoninic acid protein quantitation assay (Pierce 23227) 46, 59

Benzonaze®

Benzonase® Nuclease(Sigma E1014) 46

Bio-Rad® microplate reader

Bio-Rad® iMark® microplate absorbance reader 46

blocking solution

0.3% Triton-X100 and 5% goat serum in PBS 64

C18

octadecyl hydrocarbon bonded silica stationary phase 47, 48, 59

C18 SPE tube

discovery® DSC-18 SPE tube (Supelco 52606-U) 48

cryo vials

Corning® cryogenic vials, external thread (Sigma CLS430661) 45

DAPI

4',6-diamidino-2-phenylindole 64, 65, 93–95

Dionex

Dionex Ultimate 3000 UHPLC 49, 66

DMSO

dimethyl sulphoxide (Sigma D2650) 45, 58

DTT

dithiothreitol (MERCK 3869) 47

Dundee control SILAC

control SILAC DMEM-F12 R0K0 (Dundee cell products DMEF-38) 59

Dundee heavy SILAC

SILAC DMEM-F12 K6R6 media (Dundee cell products DMEF-39) 59

Easy-nLC

Thermo scientific™Easy-nLC II 61

EGF

epidermal growth factor (Gibco PHG0311) 80, 92, 176

FAST READ 102

FAST READ 102 disposable haemocytometer (Davies diagnostic HYC-MISC-BVS100) 45

FCS

foetal calf serum (Sigma F7524) 44, 58

FGF

fibroblast growth factor (Gibco PHG0026) 80, 92, 176

gel loader tip

GT-10-6 gell-well pipette tip (Mettler toledo 17002380) 47

glass insert

glass insert shell 0.35ml (SUPELCO 24715) 48

HB1025

NS cell line 57, 59–62, 64, 85, 86, 95

HB901

NS cell line 57, 60, 61, 64, 85, 86, 95

heavy SILAC medium

neural stem cell medium, with the DMEM:F12 replaced by either Dundee heavy SILAC or Thermo SILAC DMEM:F12 with K6 and R10 58, 59, 61

IAA

UA with 0.05M iodoacetamide 47

isopropanol

isopropanol (Sigma 34965) 45

K0

L-lysine-2HCl (Thermo scientific PI-89987) 61

K6

L-lysine-2HCl, ^{13}C (Thermo scientific PI-89988) 61

laminin

laminin (Sigma L2020) 57, 64

Leica application suite advanced fluorescence

Leica application suite advanced fluorescence 65

Leica inverted fluorescent microscope

Leika inverted fluorescent microscope (Leica DMI 4000) 65

light SILAC medium

neural stem cell medium with the DMEM:F12 replaced by either Dundee control SILAC or Thermo SILAC DMEM:F12 with K0 and R0 58, 59, 61

LTQ orbitrap velos™

Thermo scientific™LTQ orbitrap velos™ 27, 60, 61

modified RIPA buffer

modified radioimmunoprecipitation assay buffer (150mM NaCl, 50mM TEAB pH 8.5, 0.5% DOC, 1% SDS) 46, 59

Mr Frosty®

freezing container, Nalgene® Mr Frosty (C1562 Sigma) 45

neural stem cell differentiation medium

neural stem cell medium without FGF and EGF 64, 65

neural stem cell medium

DMEM:F12, 1x N2, 1x B27, 10ng/ml EGF, 10ng/ml FGF and PenStrep 57, 58, 64, 80

neuroblastoma freezing medium

DMEM, 20% FCS, 10% DMSO, PenStrep 45

neuroblastoma medium

DMEM, 10% FCS, 1x PenStrep 44, 69

PBS

phosphate saline buffer (Sigma D8537) 44, 45, 58, 64, 65

poly-L-ornithine

poly-L-ornithine (Sigma P3655) 57, 64

porous potassium silicate resin

Next advance frit kit 49

pressure injection cell

pressure injection cell (Next advance PC77) 49

protease inhibitor cocktail

protease inhibitor cocktail (Roche 11836170001) 45

R10

L-arginine-HCl $^{13}C^{15}N$ (Thermo scientific PI-89990) 61

retinoic acid

retinoic acid (Sigma R2625) 45, 69, 108

RO

L-arginine-HCl (Thermo scientific PI-8998) 61

Sai1

NES cell line 57, 59–61, 64, 85, 86, 95

Sai2

NES cell line 57, 60, 64, 86, 95

Sai3

NES cell line 57, 61, 64, 85, 86, 95

SDS

sodium dodecyl sulphate (Sigma 05030) 46

SH-SY5Y cell

SH-SY5Y neuroblastoma cell line 45

SH-SY5Y cell

SH-SY5Y neuroblastoma cell line 45, 49–52, 69–71, 74

SPE elution solution

80% ACN 0.1% TFA 48

SPE tube adapter

SPE tube adapter (Supelco 57020-U) 48

SPE wash solution

2% ACN, 0.1% TFA 48

T-150 flask

T-150 flask, canted neck, surface area 150 cm^2 (Sigma CLS3291) 65

T-25 flask

T-25 flask, canted neck, surface area 25 cm^2 (Sigma C6481) 64

T-75 flask

T-75 flask, canted neck, surface area 75 cm^2 (Sigma CLS3290) 58, 64

Thermo SILAC DMEM:F12

DMEM:F12 (1:1) media for SILAC (Thermo) 61

triton

triton 64

trypan blue

Trypan blue (Sigma T8154) 45

TrypLE

TrypLE^{TM}express (Gibco 12605028) 57, 58

trypsin

sequencing grade modified trypsin (Promega PRV5111) 47

Trypsin-EDTA

Trypsin-EDTA (Sigma T4049) 44

UA

8M urea, 0.1M Tris/HCl pH 8.5 47

Waters acquity ultra performance LC

Waters acquity ultra performance LC 59

Waters nano acquity

Waters nano acquity 60

Chapter 5

General methods

5.1 Tissue culture

5.1.1 Neuroblastoma

SH-SY5Y proliferation The SH-SY5Y cell line was kindly donated by the Wilkinson lab (Wilkinson lab) at the University of Cape Town (UCT). The original cells received where labelled passage 1, the actual passage number of the original cells in unknown. Cells were expanded and stored in liquid nitrogen as the working stock. Cells for critical experiments were always taken from this working stock, either passage 2 or 3.

1×10^6 SH-SY5Y cells were plated on a 10cm TC dish, at a density of 1.2×10^4 cells/cm^2 in neuroblastoma medium. The cells reached 80% confluency after five days.

Lifting Cells Cells were cultured in 10cm TC dishes until 80% confluent. Culture medium was removed from the 10cm TC dish and cells were briefly washed with 3ml PBS. 1ml 0.05% Trypsin-EDTA in PBS was added to the cells. They were incubated for 1 min at 37 °C Cells were visually inspected to see if they had rounding up, insuring they had lifted. 3ml of neuroblastoma medium was added, thereby inhibiting trypsin with FCS. The cells were washed off of the dish and transferred to a sterile 15ml falcon tube. The 10cm TC dish was washed again with neuroblastoma medium to ensure all cells were recovered. The cells were pelleted

by 600g centrifugation and then re-suspended in 1ml PBS.

Counting Cells Cells were counted using the trypan blue method. 10μl of cells were added to 10μl of trypan blue and placed in a FAST READ 102 disposable haemocytometer. Trypan blue is actively pumped out of cells with intact cell membranes. Therefore, cells staining blue are not viable and only clear cells are counted. An average of 1×10^7 SH-SY5Y cells were lifted from a 10cm TC dish.

Plating, Freezing or Proteomic Analysing of Cells Cells to be re-plated were plated $^1/_{10}$ on a new 10cm TC dish. Cells for freezing were re-suspended in neuroblastoma freezing medium at 1×10^6 cells/ml in cryo vials. Cells were frozen to -80 °C in a Mr Frosty® containing isopropanol. Isopropanol reduced the cooling rate to 1 °C per minute, allowing the cells to shrink slowly as water moved out of the pores created by DMSO, resulting in frozen dehydrated cells. The cells were then place in liquid nitrogen, which turns their cytoplasm to glass, for long term storage.[71] Cells for proteomic analysis were washed twice with PBS, counted, pelleted and flash frozen in liquid nitrogen, then stored at -80 °C until required.

5.1.2 SH-SY5Y differentiation experiment

1×10^6 SH-SY5Y cell were plated on six 10cm TC dishes. Cell were cultured to 80% confluency. Three of the plates were lifted and the cell pellets flash frozen for MS analysis. The other three cell plates were treated with retinoic acid for 6 days and then lifted for MS analysis.

5.2 Sample processing

5.2.1 Cell lysis

Sample were processed in parallel within individual experiments. Samples were individually labelled, mixed up and treated randomly, to prevent processing bias. The frozen cell pellets were re-suspended in 60μl PBS containing a protease inhibitor cocktail. The cells were tritu-

rated vigorously to ensure a single cell suspension, thus giving greater access of the lysis buffer to the cells. 60μl of the cell suspension were transferred to a new tube and 30μl of 3× modified RIPA buffer was added to obtain a 1× modified RIPA buffer final concentration.

Samples were vortexed briefly to mix and then boiled at 95 °C for 5 minutes. Samples were sonicated for 15 minutes, then cooled to 4 °C and 1μl Benzonaze® was added. Samples were incubate in Benzonaze® for 1 hour at 4 °C , rotating by inversion. After the Benzonaze® incubation, samples were boiled at 95 °C for 5 minutes, to dissolve any precipitated SDS.

Sample were centrifuged at max speed (15 000g) for 5 minutes, no visible insoluble pellet was generally present in any of the samples, if a pellet was observed, the vortexing, boiling and sonication steps were repeated. If the pellets persisted, the concentration of protein was thought to be too high, the volume of 1× modified RIPA buffer in all samples was incrementally increased, until all the pellets had been dissolved. All alterations to the protocol were always mirrored across all samples in the batch.

5.2.2 Protein quantitation

A bovine serum albumin (BSA) protein standard, 2mg/ml supplied as a stock in the BCA assay kit, was used to generate a standard concentration curve. Initially 100ul of 1mg/ml BSA in 1× modified RIPA buffer, was made-up to use for serial dilutions. 50μl of the following concentrations were produced, 1, 0.5, 0.25, 0.125, 0.0625 mg/ml and 1× modified RIPA buffer buffer as a blank.

Samples were diluted 5× and 10× with 1× modified RIPA buffer to fall within the range of the quantitation assay. 10μl of each sample dilution and protein standard was added to separate wells of a 96 well flat bottom plate. 200μl of BCA assay reagent was then added to each well in quick succession using a multi-channel pipette. The 96 well flat bottom plate was incubated for 30 min at room temperature. A visible colour change was observed and measured by UV-Vis using a Bio-Rad® microplate reader at 562nm. The Bio-Rad® microplate reader generated a standard curve, as well as output the concentration of sample in mg/ml. The total quantity of sample and the yield per million cells was calculated.

5.2.3 Tryptic digest

After protein quantitation, 200μg from each sample was transferred to a fresh tube and incubated in 0.1M DTT for 30 minutes. The remaining protein sample was stored at -80 °C. 30kDa MWCO filters were used for FASP. At each FASP step the filters were centrifuged at 14000g. After each centrifugation filters were checked to ensure the liquid had been spun through to the dead volume of 40μl, if not, additional centrifugation steps were added. UA, IAA and ABC solutions for FASP were prepared fresh and used the same day.

The filters were first washed by adding 500μl UA and centrifuging them for 10 minutes. The flow through was discarded. 200μg of DTT reduced protein samples was added to separately labelled filters. 200μl of UA was then added and filters which were centrifuged for 15 minutes, this was repeated twice. The flow through from the filters was discarded. 100μl of IAA was added and the filters were vortex gently at 600 rpm for 1 min. Filters were incubated in the dark for 20 minutes without mixing, followed by centrifugation for 10 minutes. 100μl of UA was added and filters were centrifuged for 10 minutes, repeated 3 times. 100μl ABC was then added, and centrifuged for 10 minutes, repeated twice.

The collection vial on the 30kDa MWCO filters was replaced with a fresh vial. Trypsin was re-suspended at 1ug/μl in ABC and added at a protein to trypsin ratio of 50:1. The filters were then vortexed gently at 600 rmp for 1 minute. The potential hydrogen (pH) of each sample was tested by placing 1μl from each filter onto a pH strip, to ensure it was above a pH of 7. The 30kDa MWCO filters we then place in a wet chamber at 37 °C . After 16 hours of incubation with trypsin, the peptides were eluted from the filter using 50μl 0.1% trifluoroacetic acid (TFA), centrifuged for 10 minutes, repeated three times. The pH of the eluted peptides were tested, to ensure a pH below 3. Peptides were either desalted directly or stored at -20 °C.

5.2.4 Peptide desalting

Custom C18 solid phase extraction (SPE) micro columns were produced in-house from gel loader tips. The last 5mm of a gel loader tip was crimped at 45° using a serrated forceps. The

forceps was then rotated by 90°. Holes were punched through the lids of 1.5ml polypropylene centrifuge tubes (centrifuge tubes) so that the micro columns could be stored upright.

SPE elution solution was added to the micro columns in 50μl volumes and forced through using air pressure. Air pressure was supplied by a 10ml syringe attached through a SPE tube adapter to the micro column. Effective crimping produced back-pressure, but still allowed solvent to pass through slowly.

A C18 SPE tube was broken open and the C18 beads added to a 1.5ml centrifuge tube. A slurry was made by triturating the beads in 500ml of SPE elution solution. 20μl from the bottom of the slurry equalled roughly 4mg of C18 beads. 20μl of C18 beads were added to the SPE elution solution in the micro column. The beads slowly sank to the bottom of the micro column. The beads were then tightly packed by applying air pressure. When the SPE elution solution reached 2mm above the level of the beads, 50μl of SPE elution solution was added to the column and it was re-pressurised, this was repeated three times. Throughout the process the C18 beads were always kept wet.

The beads were then equilibrated by passing three 50μl volumes of SPE wash solution over them. Peptide produced by FASP were centrifuged at 14000g for 10 minutes, to pellet any particulates. The peptides were passed through the beads in 50μl volumes, as slowly as possible. Peptides bound to the the beads were then washed with two 50μl volumes of SPE wash solution. The micro column was transferred to a fresh centrifuge tube containing a 300μl glass insert. The bound peptides were eluted from the micro columns using three 50μl volumes of SPE elution solution and collected in the glass inserts.

Peptide were then dried down in a vacuum centrifuge, and stored dry at -80 °C

5.3 Sample analysis

5.3.1 Liquid chromatography

5.3.2 Preparing columns

Nano Columns for the LC were produced in-house using a pressure injection cell. Fused silica capillaries were cut to length and fritted on one end, using porous potassium silicate resin. 30μl Kasil 1624, 10 μl Kasil 1 and 10μl formamide were mixed together. Fused silica columns were dipped into the solution for 3s, allowing 2cm of the porous potassium silicate resin solution to be draw up by capillary action. The columns were then baked for 8 hours at 150 °C . Frits were trimmed to 1cm. Bulk packing material was suspended in 100% B and forced into columns at 1000psi using a pressure injection cell. Columns were prepared for use by running 80% B at 1μl/min for 1 hour. They were then trimmed to length.

5.3.3 Dionex

All SH-SY5Y cell samples were run on a Dionex with a 2cm 100μm trap column with 5μm luna C18 beads and a 40cm 75μm analytical column with 5μm luna C18 beads. 300μg of sample was injected. Sample loading was normalised to ensure an average total ion chromatogram (TIC) peak height of 1×10^{10}. Peptides were separated by 120 minute gradient from 5%B to 25%B. Nano ESI set a 2.1kV ionised the eluting peptides, for analysis by a Q Exactive.

5.3.4 QE

The Q Exactive was operated in full ms^1 to data dependent ms^2 mode. Peptide ms^1s were initially scanned between 300-1750 m/z at 70 000 resolution, with a c-trap AGC of 3×10^6. Data dependent ms^2 was triggered for ions with charges 2-4 and with intensity greater than 1.7×10^4 using an underfil ratio of 1%. The isolation width was set to 4 m/z. Selected ions were collected until the AGC of 2×10^5 or maximum injection time of 120 ms. Peptide were fragment with a normalised collision energy (NCE) of 25 and then scanned with a resolution

Table 5.1: *Q Exactive setting for SILAC mudpit samples*

Run time	157 min
ms^1 Resolution	70 000
ms^1 AGC Target	3×10^6
Scan Range	300-1750 m/z
ms^2 Resolution	17 500
ms^2 AGC Target	2×10^5
Loop Count	10
Isolation Width	4 m/z
NCE	25
dd Underfil Ration	1%
dd Intensity threshold	1.7×10^4
Dynamic Exclusion	30s

of 17 500. Ten ms^2s were collected for every ms^1, with the dynamic exclusion duration set to 30s.

5.3.5 MaxQuant analysis

MQ version 1.3.0.5 was used to for the SH-SY5Y cell peptide spectral matching and quantitation. Thermo .raw files were searched against the human complete proteome database downloaded from uniprot on 9th January 2014. Reverse sequences of this database were used as the decoy database. Trypsin/P was set as the enzyme, and 2 missed cleavages were allowed. Carbamidomethyl of cystein's was set at a fixed modification. Acetylation of the protein N-terminus and Oxidation of Methionine we set as variables modifications. All FDRs were set to 1%. The match between runs and dependent peptides features were turned on.

Table 5.2: *MQ settings for SH-SY5Y cells*

Experiment	SH
Version	1.5.0.0
Enzyme	Trypsin/P
Max. missed cleavages	2
Min. peptide Length	7
Min. peptides	1
Min. ratio count	2
Min. razor peptides	1
Min. unique peptides	0
Peptides used for protein quantification	Razor
Fixed modifications	Carbamidomethyl (C)
Variable modifications	Acetyl (Protein N-term);Oxidation (M)
PSM FDR	0.01
Protein FDR	0.01
Site FDR	0.01
Match between runs	False
Re-quantify	False
Labels	0

5.4 Bioinformatics

5.4.1 Statistics

All statistical and data manipulation was done in R[171]: a language and environment for statistical computing (R), using a graphical user interface program written in-house using shiny[45] called CRUNCH. CRUNCH can be used to view the data and the plots generated from the data at `https://www.shaungarnett.co.za/thesis.html`. MQ data was first prepared by removing all protein groups that were reverse hits or contaminants marked by MQ. The FDR was set at 1% on the peptide and protein level. No imputation of missing data was done. Where there were no intensity values for proteins, the zero values were converted to not available (NA). Missing values were then ignored in the statistical calculations in R. MQ protein groups were collapsed down to the the first protein in the protein group, to enable ease of comparison between different experiments and downstream meta-analysis. The Uniprot protein accession numbers were further converted to human genome organisation gene nomenclature committee (HUGO/HGNC)[1] gene symbols (gene symbols).

Label free quantitation statistics

In the SH-SY5Y cells pairwise LFQ experiment samples were divided into two groups: control and treated. Ratios were generated between the replicates (control vs control and treated vs treated) and the comparisons (control vs treated), providing information of technical and biological variance respectively. The mean values of each protein replicate ratios was $\log_2$ transformed. Histograms of the $\log_2$ ratios generally showed normal distributions with a mean of 0. These $\log_2$ ratios were then used to calculate the σ. 2σ was taken to represent the extent of the technical variance within the experiment. The t test was applied to the comparisons, LFQ intensity values per protein, producing a p-value indicating the degree of significance. Due to the large number of t test performed multiple-testing correction needed to be applied. The multiple-testing correction test BH was applied to all proteins.. All proteins with a $\log_2$ mean comparison ratio greater than the technical variance 2σ and a BH p-value less than

0.05 were accepted as being significantly differentially expressed. The lists of proteins which were significantly differentially expressed were then used in gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment.

5.4.2 Plotting

Boxplots In an attempt to visualise all the data for individual proteins boxplots were generated using ggplot2[233] in R. For the binary comparison experiments data points are represented as a boxplot. The p-values for each experiment were displayed below the boxplot. For the timecourse data the ratio points were plotted in boxplots, together with the mean line and p-values calculated by one-way analysis of variance (ANOVA). The boxplots were filled with either red or green, if the data were significantly up- or down-regulated (Figure 9.32).

Histograms and Density Plots To generate a histogram data is ordered, then a number of bins are created across the data range of $\log_2$ ratios, plotted on the x axis. A frequency is then determined for the number of data points falling within each bin plotted along the y axis. Histograms are barplot of these frequencies (Figure 6.6 (a)). Density plots use the same principle but are viewed as a continuous line, thus enabling multiple datasets to be overlaid (Figure 6.5 (a)).

FDR Plot Once p-values had been generated either by t test or after multiple-testing correction. The distribution of the p-value values was plotted by histogram. A mean frequency of p-values between 0.8 and 1 was calculated. Using this frequency the number of p-values occurring by random chance was determined. This number was divided by the total number of p-values below 0.05 to give an estimated FDR (Figure 6.6 (a)).

Heatmaps Proteins were plotted as a heatmap using the base heatmap plot R, which performs hierarchical clustering on the protein ratios. Colours were set to green as negative and red as positive. The R heatmap function is only able to generate heatmaps on complete data. Every protein would require a intensity value in every sample. This was not feasible so missing

ratio values were changed to zero, indicating no change or not detected, represented as black on the plot. The hierarchical clustering performed by the heatmap provided a means of determining if there were any large batch effects within the data. Using the tile plot in ggplot2 heatmaps were also generated to show significantly differentially expressed (SDE) of multiple proteins at the same time, thereby condensing the display of data. Red and green were used to show up- or down-regulation respectively (Figure 9.14).

5.4.3 Functional Analysis

topGO

The topGO[14]: enrichment analysis for gene ontology package in R (topGO) was used to for GO enrichment of cellular component (CC), molecular function (MF), and biological process (BP). First, all detected proteins were mapped to GO terms using the STRINGdb[73] package in R (STRINGdb). A list of either the up- or down-regulated differentially expressed proteins were then compared to the full list of proteins detected in that experiment. Performing the enrichment against only those detected reduces detection bias, introduced by the sample analysis technique used.

Each of the three GOs, CC, MF, and BP forms a hierarchical network. There is a high degree of relatedness between the terms within each network. Sibling terms contain a subset of proteins from their parent terms. Terms higher up the hierarchy are often more significantly enriched than the lower, more descriptive terms. Multiple-testing correction further compounds this problem since it cannot take the relatedness of the terms into account. The topGO algorithm accounts for this bias using algorithms that either eliminate the higher terms or negatively weights them. Using the weight01 algorithm in the topGO package[13] in R (weight01), if any of the proteins within the term are detected in significantly enriched child terms, the parent terms are negatively weighted. The weighting only works vertically and does not affect sibling terms.[15] Multiple-testing correction is applied during the generating of the weighted p-values

The STRINGdb package in R is also able to do enrichment calculations.[210] Unlike topGO it does not use weighting, the STRINGdb *get_enrichment* was used to calculate enrichment p-values using the Fisher exact test adjusted by the Bonferroni correction. Enrichment was performed separately on the lists of SDE proteins showing up- or down-regulation.

The resultant corrected enrichment p-values from topGO and STRINGdb were logarithm base 10 ($\log_{10}$) transformed to return negative values. The enrichment scores from the up-regulated proteins were then inverted to positive values. These enrichment scores were then plotted on horizontal barplots. All the experiments in this study were all collated together into one plot. This highlighted the differences in enrichment between those showing up- or down-regulation, within experiments and across them. Blue horizontal lines represent the a p-value threshold of 0.01 (Figure 12.1).

5.4.4 STRINGdb

Each enriched GO term consists of a list of proteins. The proteins for a significantly enriched term, found to be significantly differentially expressed were extracted. This protein list was then plotted as a STRINGdb network using the STRINGdb[210] package in R. A STRINGdb network is comprised of nodes and edges. Nodes are individual proteins, which are visualised as orbs containing their crystal structures. Coloured halo's around the nodes indicate up or down regulation. Edges are the connections between the nodes. The edges are coloured and represent the information used to determine the connection between the nodes (Figure 5.1). Edge information can range from high to low confidence. High confidence interactions are known interactions, extracted from curated databases. Less confident interactions are deduced from protein homology and text mining. A combined score of 400 from STRINGdb version 10 was used and takes all of these evidences into account. Hubs are proteins with many connections within a network, while a clique is a number of highly connected proteins (Figure 12.7).

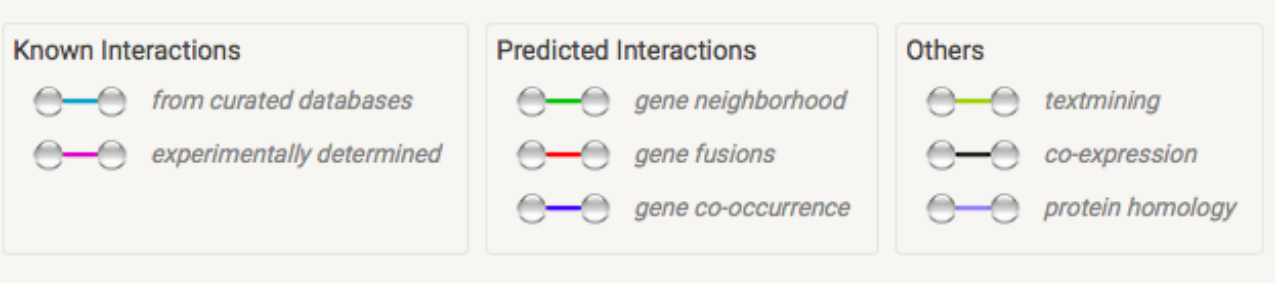


Figure 5.1: *Colour codes on STRINGdb network, from* `string-db.org`

5.4.5 AnimalTFBD

Lists of transcription factor (TF), transcription co-factors and chromatin remodelling factors were downloaded from the AnimalTFDB.[241] The TFs were identified by Hidden Markov Model of DNA-binding domains. They defined transcription factors as having "a sequence specific DNA-binding domain and regulating target gene expression".[241] These list were compared to the SDE genes and proteins in this study, enrichment analysis was performed using the Fisher exact test corrected by BH.

Chapter 6

Human foetal neural stem cells

6.0.1 SILAC proliferation study

6.0.2 Neural stem cells

Cell lines

Table 6.1: *Human foetal neural stem cell cell lines*

key	Cell Line	Cell Type
S1	Sai1	$_{hf}$NES cells
S2	Sai2	$_{hf}$NES cells
S3	Sai3	$_{hf}$NES cells
H9	HB901	$_{hf}$NS cells
H98	HB985	$_{hf}$NS cells
H10	HB1025	$_{hf}$NS cells

Proliferation

Three $_{hf}$NES cells (Sai1, Sai2, Sai3) as well as two $_{hf}$NS cells, (HB901 and HB1025) were kindly donated by the Smith lab. These cells were cultured on tissue culture (TC) plates coated with 100μg/ml poly-L-ornithine for 20-30min at 37 °C and 2μg/ml laminin for 4-12h at 37 °C. $_{hf}$NES cells and $_{hf}$NS cells were plated at high density 4×10^4 cell/cm^2 in neural stem cell medium. Continuous contact between the cells is necessary to maintain stable cultures. Neural stem cell medium was changed daily. Cells were passaged every three days using TrypLE, which

does not require FCS for inhibition. The old medium was first removed. The cells were then washed with PBS and incubated with 1ml TrypLE for three minutes at 37 °C. The TrypLE was then diluted 3× with neural stem cell medium. Cells were pelleted, counted and re-plated $^1/_3$ thereby maintaining a high cell density. For long term storage cells were frozen in neural stem cell medium with 10% DMSO.

Quantitative labelling

SILAC Labelling $_{hf}$NES cells and $_{hf}$NS cells were each plated in a single well of a 6 well TC plate in either light SILAC medium or heavy SILAC medium respectively. The cells were expanded to a 6cm TC dish and then a 10cm TC dish in either light SILAC medium or heavy SILAC medium. This expansion comprised more than five cell doublings. This resulted in full incorporation of light arginine and lysine in one set of cells and heavy arginine and lysine in the other. The cells were referred to as light of heavy from this point on. Each 10cm TC dish was split and $^1/_3$ cryo-preserved, $^1/_3$ tested for incorporation and $^1/_3$ re-plated.

Heavy amino acid incorporation was tested on a sample of all cells cultured in heavy SILAC medium. Peptides from the heavy cells were analysed by MS. The spectra were processed using MQ with variable modifications for heavy arginine, lysine and proline. The total number of heavy arginine and lysine containing peptides were compared to the total number of light arginine and lysine containing peptides. Using formula 6.1 the incorporation rate was determined. An incorporation rate greater than 95% was required for full incorporation. Peptides containing heavy and light prolines were analysed in the same manner to determine the conversion rate of heavy arginine to heavy proline,

$$\text{Incorporation rate } (\%) = (1 - \frac{\#\text{ light Peptides}}{\#\text{ light Peptides} + \#\text{ heavy Peptides}}) * 100 \tag{6.1}$$

The re-plated cells were further expanded to a T-75 flask. The cells were then lifted, flash frozen in liquid nitrogen and stored until required.

SILAC labelled cells are usually combined based on cell number. Due to the difference in cell morphology and size between, $_{hf}$NES cells and $_{hf}$NS cells they were combined based on protein quantity rather than cell number. Cells were thawed and resuspended in modified RIPA buffer. Proteins were quantitated using the BCA assay. 100μg heavy SILAC $_{hf}$NS cells protein was combined with 100μg light SILAC $_{hf}$NES cells proteins. The SILAC labelling was reversed in some cases, light $_{hf}$NS cells combined with heavy $_{hf}$NES cells, to account for labelling bias,

In order to test the accuracy of mixing MA plots, with M (log ratios) and A (Intensity) were generated, showing a distribution over zero across all intensity values

6.0.3 SILAC experiments

SILAC was done in three separate experiments, the exact sample compositions can be seen in table 8.4

SILAC experiment 1 (SILAC 1) The initial pilot SILAC experiment 1 (SILAC 1) was done in the Smith lab where Sai1 and HB1025 where grown in Dundee control SILAC as K0R0 light SILAC medium and Dundee heavy SILAC as K6R6 heavy SILAC medium. Incorporation was tested on the heavy HB1025 cells. In order to test technical variance, 25μg of heavy and light HB1025 were combined, processed by FASP and compared to each other using a single LC/MS/MS run.

High pH reverse phase In the comparison experiment 100μg of light Sai1 and 100μg of heavy HB1025 were combined and processed with FASP. Desalted peptides from FASP were dried down and resuspended in 20mM ammonium formate (pH 10). The entire sample was loaded onto a Waters acquity ultra performance LC. Peptides were separated on a C18 column compatible with high pH, using 20mM ammonium formate in water as Solvent A and 20mM ammonium formate in 80% acetonitrile (ACN) as Solvent B. The gradient started at 0% B for 10 minutes followed by a 50 minute linear ramp up to 70% B. The column was washed at

100%B for 7 minutes before coming down to 0%B. Peptide elution was detected and plotted using an inline ultraviolet (UV) detector. Fractions were taken every 2 minutes from minute 12 to 42, using a automated fraction collector. Fractions from 32-36 minutes and 36-42 minutes were pooled.

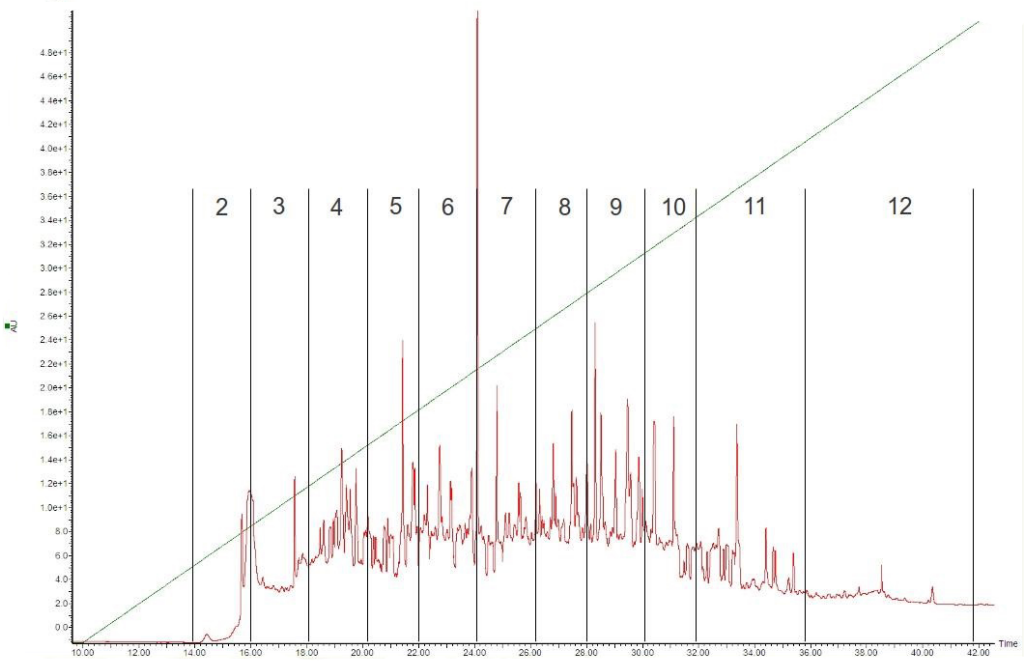


Figure 6.1: *High pH reverse phase UV chromatogram and fractionation times*

Each fraction was desalted, dried down and resuspended in 0.1% TFA. The 13 peptide fractions were separated on a Waters nano acquity. The gradient ramped from 5-40%B in 100 minutes. The column was washed at 85%B for 5 minutes and then equilibrated for 15 minutes at 5%B.

Eluted peptides were analysed by a LTQ orbitrap velos^{TM} at the Cambridge centre for proteomics (CCP) using the settings in table 6.2

SILAC experiment 2 Sai1, Sai2, HB901 and HB1025 for SILAC experiment 2 (SILAC 2) were cultured in the Smith lab in heavy and light SILAC medium as described in SILAC 1. The frozen cells were then brought back to the Blackburn lab, IDM, University of Cape Town

Table 6.2: *LTQ orbitrap velos^{TM} settings for SILAC experiment 1*

Run time	120 min
Mass Range	380-1500 m/z
Minimal Signal Required	500
Isolation width	2 m/z
Normalised Collision Energy	30
Dynamic Exclusion Repeat Duration	20 s
Dynamic Exclusion Duration	300 s

(Blackburn lab) for processing and analysis. Heavy HB901, HB1025 and Sai1 peptides were tested for incorporation. Sai1 and HB901 were used to test technical variance. One reverse label comparison was done to account for labelling bias.

SILAC experiment 3 In SILAC experiment 3 (SILAC 3), Sai3 and HB1025 cells were cultured in duplicate in either Thermo SILAC DMEM:F12 with K0 and RO as light SILAC medium or Thermo SILAC DMEM:F12 with K6 and R10 as heavy SILAC medium in the Blackburn lab. When combined the labels were reversed in one of the duplicates. All heavy labelled cells were tested for incorporation. Heavy and light HB1025 was combined to test the technical variance.

Mudpit SILAC 2 and SILAC 3 peptides generated by FASP were desalted and then fractionated in-line by multidimensional protein identification technology (MudPIT) in the Blackburn lab. A MudPIT column was produced in-house using a 100μm fused silica capillary. The capillary was fritted and packed with 2cm 5μm luna C18 beads, 3cm 5μm SCX beads and 3cm 5μm luna C18 beads. The MudPIT column was used instead of a trap column on an Easy-nLC in the two-column configuration. The analytical column used was a 20cm 75μm with 5μm luna C18 beads.

1ug of sample was loaded onto the MudPIT column in 2% ACN. Peptides bound to the first 3cm of 5μm luna C18 beads. A 70 minute gradient from 2 to 35% B, separated all peptides unable to bind SCX. Peptide bound to the SCX were eluted in fractions by a series of ammonium acetate injections, called salt plugs. The salt plug series consisted of increasing concentrations of ammonium acetate 25, 50, 75, 200, 400 and 500mM, in 5% B. An additional

two fractions increased the percentage of B to 13.5% and 22% while keeping the ammonium acetate concentration at 500mM. Peptides eluted by the salt plugs from the MudPIT column were separated on the analytical column by a 70 minute gradient from 2% to 35%B and analysed by a Q Exactive

Analysis on the Q Exactive was done using the settings in the table 6.3

Table 6.3: *Q Exactive acquisition settings for SILAC experiment 2 and SILAC experiment 3*

Run time	100 min
ms^1 Resolution	70 000
ms^1 AGC Target	3e6
Scan Range	300-1750 m/z
ms^2 Resolution	17 500
ms^2 AGC Target	2e5
Loop Count	10
Isolation Width	4 m/z
NCE	25
dd Underfil Ration	1%
dd Intensity threshold	1.7e4
Dynamic Exclusion	30s

Combined data

The SILAC experiment for the comparison of proliferating $\frac{\text{NS}}{\text{NES}}$ (SILAC $\frac{\text{NS}}{\text{NES}}$ experiment) consists of multiple samples from three separate experiments, that were processed and analysed independently. The SILAC experiments were individually processed in MQ using the setting in Table 6.5. Data from the three MQ proteinGroups.txt files was collated into a single dataset. Rows were combined using the first protein accession in the protein group converted to gene symbol. The SILAC normalised heavy/light (H/L) ratio for each sample was $\log_2$ transformed. The reverse comparison ratios were inverted were so that all ratios were $_{\text{hf}}$NS cells/$_{\text{hf}}$NES cells. Using the SILAC technical replicates, (i.e Heavy HB1025 vs light HB1025), a σ of 0.29 was calculated for the replicate samples (see Table 8.4), from this the assumption is made that any difference greater than 2σ or 0.58 is a biological change and therefore constitutes differential expression. Using a t test all SILAC $\frac{\text{NS}}{\text{NES}}$ experiment comparison H/L ratios were compared

against a mean of 0. BH multiple-testing correction from R was then applied against all proteins present in at least 2 samples.

The final list of significant proteins is generated from samples with a mean ratio greater than 2σ (0.58) with a BH p-value of less than 0.05 and only if the protein is detected in at least 2 samples

6.0.4 Differentiation experiment

Differentiation

$_{Hf}$NS cells and $_{Hf}$NES cells differentiate spontaneously in neural stem cell differentiation medium on plates coated in laminin and poly-L-ornithine. Neural stem cell differentiation medium was only replaced every three days. Upon media change cells stopped proliferating. Many cells died, became rounded and lifted off of the plate. The surviving cells began extending processes. During media change the dead cells were washed off the plate as carefully as possible, by swirling PBS over the cells twice before applying fresh medium. After 6 days few cells were still dying and at this time most cells had developed extensive processes. In one trial $_{hf}$NES cells were cultured for 8 weeks, illustrating the potential of long term culture of these cells.

In the neural stem cell differentiation experiment five cell lines were cultured three $_{hf}$NES cells (Sai1, Sai2 and Sai3) and two $_{hf}$NS cells (HB901, HB1025). Cells were thawed from liquid nitrogen stocks and plated on T-25 flasks, then expanded to three T-75 flask, two of which were frozen for long term storage. The remaining plate was expanded for immunohistochemistry and proteomic analysis.

Immunohistochemistry A single T-75 flask was split to fill all wells of two 24 well TC plates. Cells were then allowed to proliferate in neural stem cell medium for a day, to let them settle. Each plate was differentiated by changing to neural stem cell differentiation medium and fixed on days 2, 4, 8 or 12 of differentiation for immunohistochemistry.

For cell fixing, media was removed from each of the wells and the cells washed with PBS. The cells were fixed with 4% paraformaldehyde for 10 minutes and then washed with PBS. The wells were blocked for 1 hour with blocking solution at room temperature, then washed with PBS. The fixed cells were incubated with primary antibodies in 0.1% triton and 1% BSA for 12 hours. Wells were washed three times with PBS before applying secondary antibodies, including DAPI, for 1 hour.

The primary antibodies used were anti-Nestin an early neuronal stem cell marker, anti-

TuJ-1 a neuronal marker and anti-GFAP a glial marker. Primary antibodies were visualised by secondary antibodies conjugated to fluorophores Alexa 448 (Green) anti-mouse IgG2A for anti-TuJ-1, Alexa 555 (Red) anti-mouse IgG1A for anti-GFAP and Alexa 555 (Red) anti-sheep for anti-Nestin. DAPI was used to stain the nuclei blue. Images were taken on a Leica inverted fluorescent microscope and processed with the Leica application suite advanced fluorescence software.

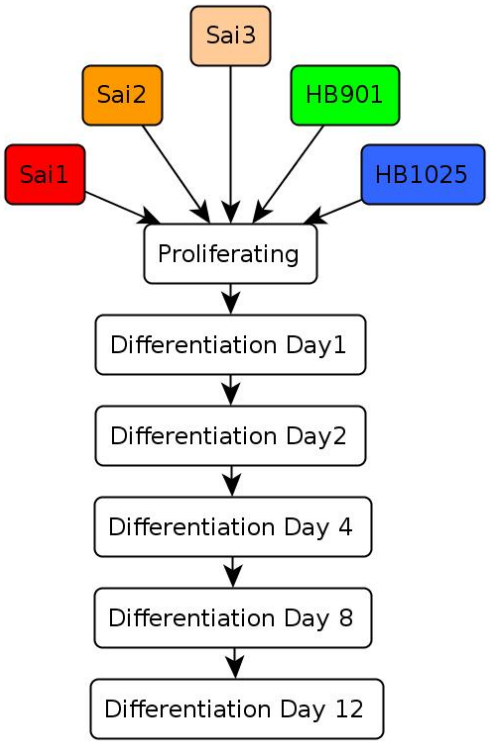


Figure 6.2: *Neural stem cell differentiation experiment design*

Proteomics analysis Cell culture was scaled up to nine T-150 flasks per cell line, Each T-150 flask was expanded until 80% confluent. The first time-point Day 0 was taken by lifting one of the plates for each cell line. The remaining cells were differentiated by changing media to neural stem cell differentiation medium. One T-150 flask was lifted per cell line on days 1,2,4,8 and 12. Lifted cells were spun down and washed twice with PBS before being flash frozen on dry ice and then stored at -80 °C, until all samples could be collected and processed in parallel.

Cell pellets were thawed, lysed and quantitated. 200μg from each sample was processed by FASP, as described previously. Peptides were desalted and 1μg separated by Dionex, using 2cm 5μm luna C18 beads trap and 40cm 5μm luna C18 beads analytical column. The gradient separated peptides for 120 minutes from 2%B to 35%B. Each sample was analysed in three separate injections on a Q Exactive using the settings in table 6.4. All the differentiation samples were analysed as one run in MQ using the settings in column "Diff" of Table 6.5.

Table 6.4: *Q Exactive setting for LFQ Samples*

Run time	190 min
ms^1 Resolution	70 000
ms^1 AGC Target	3e6
Scan Range	300-1750 m/z
ms^2 Resolution	17 500
ms^2 AGC Target	5e4
Loop Count	20
Isolation Width	1.6 m/z
NCE	28
dd Underfil Ration	5%
dd Intensity threshold	4.3e4
Dynamic Exclusion	45s

6.0.5 Differentiation ratios

The data was ordered from Day 0 to Day 12. The data was then transformed into $\log_2$ ratio by dividing all values in the dataset by the first detected sample in a protein row. This first measurement was then zero and all other measurement positive or negative, indicating whether the change was upwards or downward. The data was then separated into groups based on differentiation days and an ANOVA in R base applied. To account for multiple testing, the ANOVA p-values were corrected by BH. The slope of the linear regression used in the ANOVA was used to determine if the change was up or down regulation. All proteins with corrected p-values of less than 0.05 and a positive slope was assumed to be upregulated. These up and down regulated lists were then used to do GO, KEGG enrichment as described before. The list of differentially expressed proteins was always compared to the entire list of protein detected in the entire differentiation experiment. An alternative approach may have been to use only

Table 6.5: *MQ settings across all experiments*

Experiment	Diff	SH	SIL1	SIL2	SIL3
Version	1.5.2.8	1.5.0.0	1.3.0.3	1.3.0.5	1.3.0.5
Protease	Trypsin/P	Trypsin/P	Trypsin	Trypsin/P	Trypsin/P
Max. missed cleavages	2	2	2	2	2
Min. peptide Length	7	7	7	7	7
Min. peptides	1	1	1	1	1
Min. ratio count	2	2	2	2	2
Min. razor peptides	1	1	1	1	1
Min. unique peptides	0	0	1	1	1
Peptides used for protein quantification	Razor	Razor	Unique	Unique	Razor
Fixed modifications	Carbamidomethyl (C)	Carbamidomethyl (C)	Carbamidomethyl (C)	Carbamidomethyl (C)	Carbamidomethyl (C)
Variable modifications	Acetyl (Protein N-term);Oxidation (M)	Acetyl (Protein N-term);Oxidation (M)	Oxidation (M);Acetyl (Protein N-term)	Oxidation (M);Acetyl (Protein N-term)	Oxidation (M);Acetyl (Protein N-term)
PSM FDR	0.01	0.01			
Peptide FDR			0.01	0.01	0.01
Protein FDR	0.01	0.01	0.01	0.01	0.01
Site FDR	0.01	0.01	0.01	0.01	0.01
Dependent peptide FDR	0.01				
Dependent peptide FDR	0.01				
Match between runs	True	False	False	False	False
Matching time window [min]	0.7				
Alignment time window [min]	20				
Re-quantify	False	False	True	True	True
Labels0					
Labels1			Arg6;Lys6	Arg6;Lys6	Arg10;Lys6

those detected in the cell type, but using a more inclusive list was more stringent.

Part III

Results

6.1 SH-SY5Y cell differentiation results

Proliferating SH-SY5Y cells were compared to SH-SY5Y cells differentiated through treatment with retinoic acid. Cells displayed distinct differences in morphology: proliferating cells have a bi- or tripolar morphology, while the differentiated cells extended distinct neuronal processes (Figure 6.3). The cells were lifted, their protein content was extracted, digested, and analysed by LC/MS/MS. The data were analysed using MaxLFQ. Intensity values were processed and statistically validated using CRUNCH. A total of 22701 peptides were detected, with approximately 12000 peptides per sample, resulting in over 2000 proteins per sample (Figure 6.4) and a total of 3700 proteins.

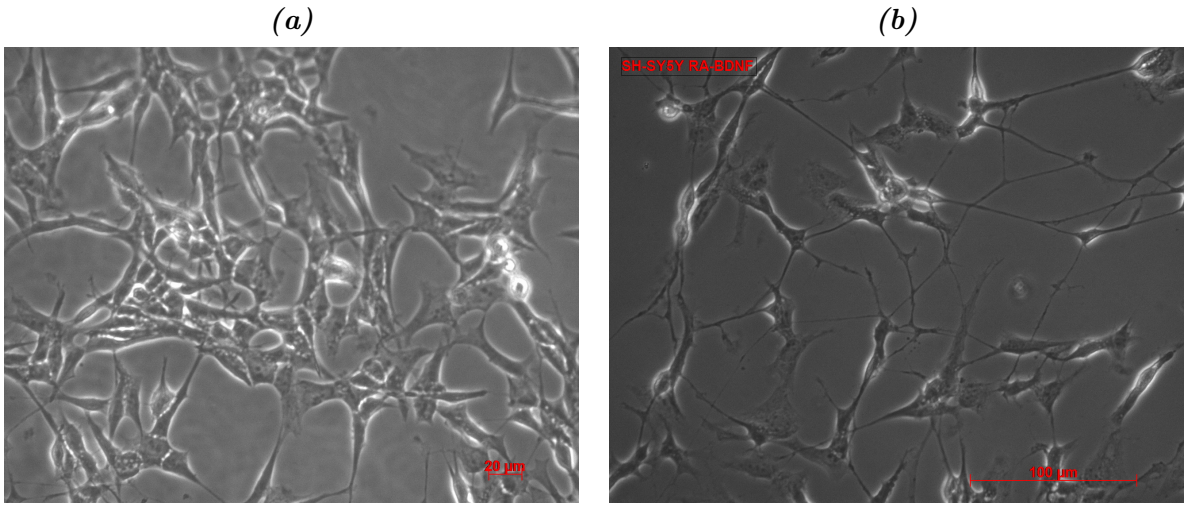


Figure 6.3: *Bright-field microscope images of SH-SY5Y cells. (a) SH-SY5Y cells cultured in neuroblastoma medium. (b) SH-SY5Y cells treated with retinoic acid for 6 days and BDNF for 4 days.*

Table 6.6: *Overview of the number of mass spectra detected and peptides identified per sample in the SH-SY5Y cells differentiation experiment (SH-SY5Y cells differentiation experiment)*

Samples	MS	MS/MS	MS/MS Id	Peptide Id	Peptide Sequences Identified	Percentage MS/MS Id	Percentage Repeated
SH7_10-01	6332	29352	14306	12238	12238	48.74	23.27
SH7_10-02	6125	30191	14964	12543	12543	49.56	24.39
SH7_10-03	6301	30000	15518	13359	13359	51.73	25.10
SH7_R-01	6717	27735	14075	12265	12265	50.75	22.62
SH7_R-02	7106	25529	11120	9918	9918	43.56	21.92
SH7_R-03	6505	28754	15043	12951	12951	52.32	22.93

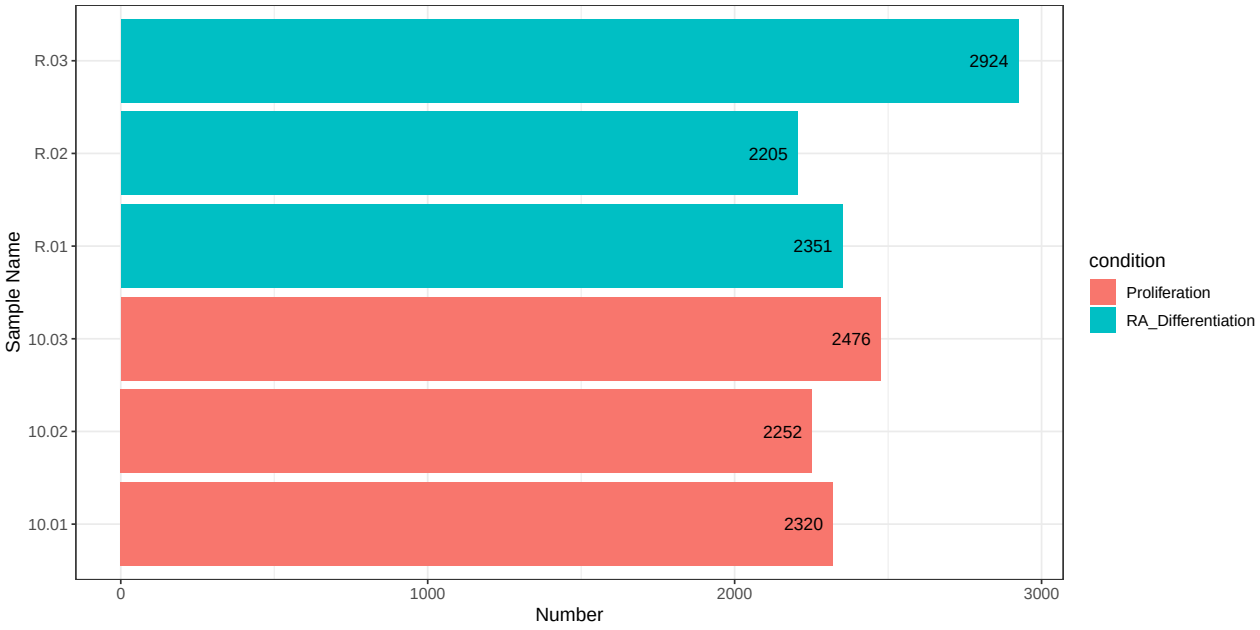


Figure 6.4: *Protein numbers for the SH-SY5Y cells differentiation experiment samples.*

There were a total of 20.55% missed cleavages. Log_2 ratios were generated for replicates of proliferating or differentiating SH-SY5Y cells, and for comparisons between proliferating and differentiating SH-SY5Y cells. Density plots for the replicate and comparison ratios showed that comparisons between replicating and differentiating cells displayed a broader distribu-

tion than comparison between replicates (Figure 6.5(a)). The σ was between 0.27 and 0.46 for replicates, while it was between 0.62 and 0.71 for comparisons between proliferating and differentiating cells (Figure 6.5(b)). This indicates a distinct difference between the two conditions. The maximum σ for replicates was 0.46 and 2σ (0.92, fold change of 1.9) was selected as the threshold for substantial differential expression.

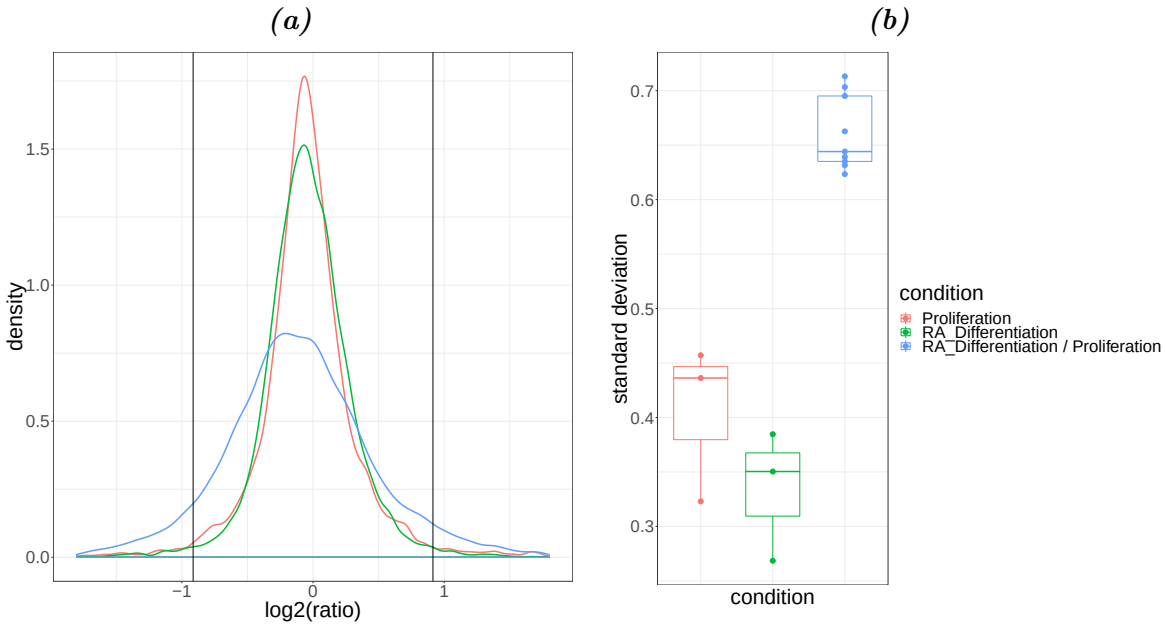


Figure 6.5: *(a) A density plot of the frequency of log_2 ratios for proteins from replicates and comparisons, for the SH-SY5Y cells differentiation experiment samples. The vertical lines indicate 2σ for replicates. (b) Boxplot of σ for SH-SY5Y cell replicates and comparisons.*

An expression change threshold is not sufficient to determine differential expression, so the intensities were further evaluated by t test. From the t test results, the FDR was determined by dividing the total number of proteins with a p-value less than 0.05 by those randomly generated. The frequency of p-values randomly generated is assumed to be the same as the average number of proteins with a p-value between 0.8 and 1. The FDR plot revealed, a FDR of 7.14% (Figure 6.6(a)). After multiple-testing correction by BH there remained 10 proteins up-regulated and 38 proteins down-regulated.

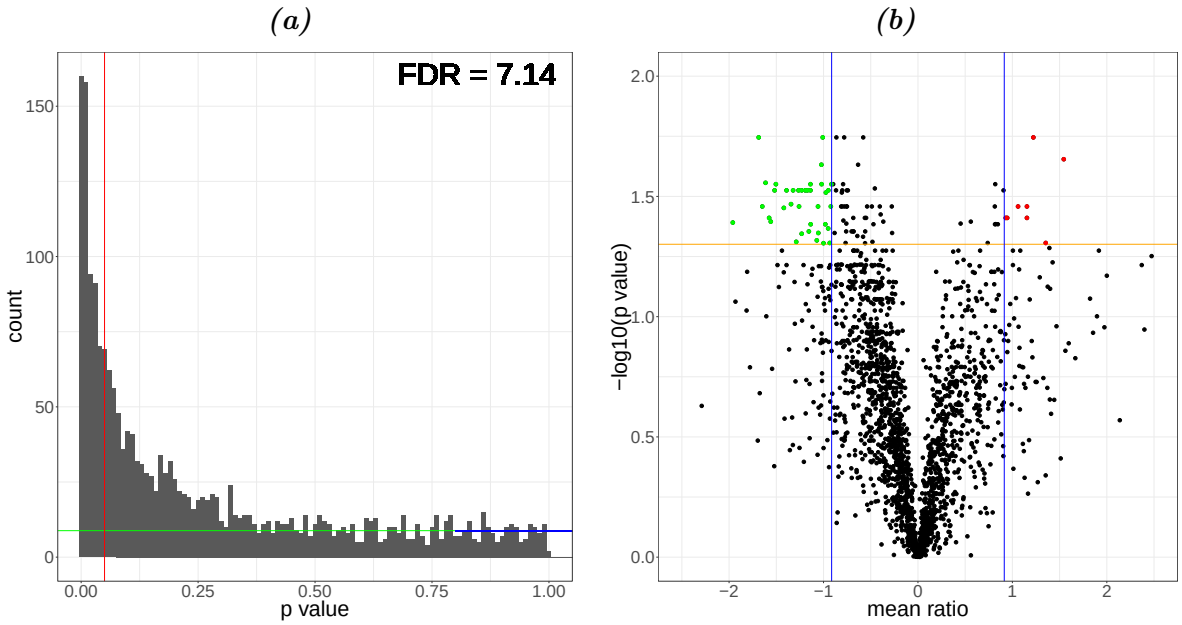


Figure 6.6: *(a) FDR plot for the t test p-values. The mean frequency between 0.8 and 1 was extrapolated to represent p-values generated by random chance (blue line). The FDR is the percentage of all proteins with a p-value smaller than 0.05 (red line) and those with a p-value smaller than 0.05 generated by random chance. (b) The volcano plot plots the log_2 mean ratio of the comparison vs the -log_{10} of the p-value. The vertical blue lines depict the 2σ of the replicates (0.92). The horizontal red line marks a p-value of 0.05. All values above the red line and outside of the blue line represent significantly differentially expressed (SDE) proteins*

There were few enough proteins SDE in the SH-SY5Y cells differentiation experiment, that they could all be displayed as a STRINGdb network (Figure 6.7). The down-regulated proteins formed three clusters. The cluster with *MCM4* as a hub, contained proteins involved in the DNA replication and cell proliferation. The cluster with *EGFR* as a hub, forms part of CC ***focal adhesion***, the ***basolateral plasma membrane***, and the ***nuclear membrane***. The third cluster containing *NUP50*, *NUP37* and, *NUP153* forms part of the CC ***nuclear pore***. The up-regulated proteins did not form any distinct clusters. There was an up-regulation of the histones *HIST1H4A* and *H1F0*, a replication independent histone, this indicating a change in chromatin compaction with cessation of proliferation, discussed in section 13.2.2.

There were not sufficient number of proteins SDE in this work for GO of KEGG enrichment analysis to return details results on the mechanism at play here. The analysis would need to be on protein by proteins basis, but appears to indicate as expected a move from proliferation to differentiation.

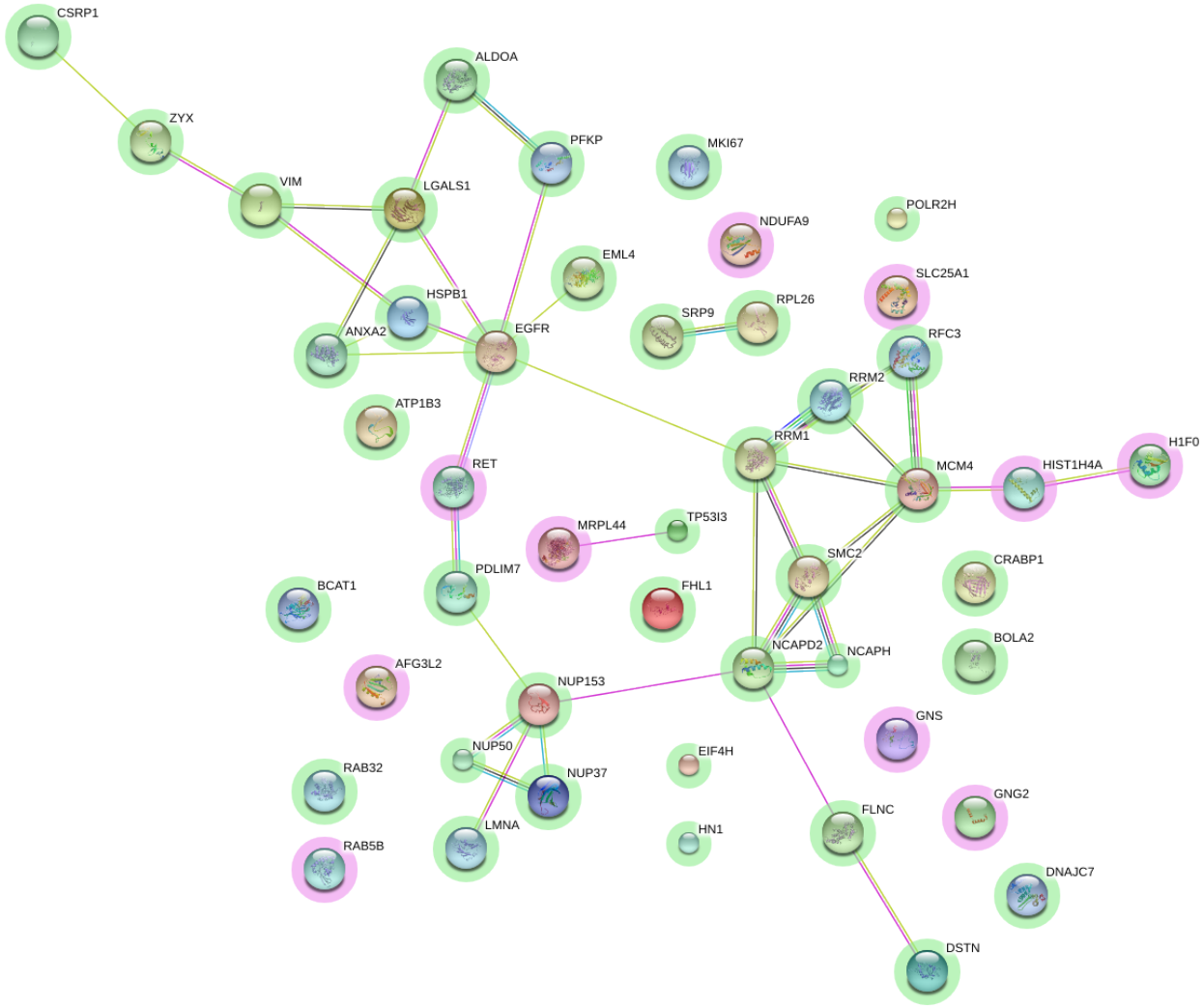


Figure 6.7: *STRINGdb network of the SDE proteins in the SH-SY5Y cells differentiation experiment. Red and green halos indicate up- and down-regulation respectively.*

6.1.1 Gene ontology

Despite the low number of SDE proteins GO enrichment p-values were calculated for all SDE proteins using the weight01. Up- and down-regulated proteins were enriched independently and then combined in boxplots to illustrate the differences between the two conditions (Figure 6.8).

Only three CCs terms were significantly enriched by the weight01 algorithm from topGO, ***condesin complex***, ***focal adhesion***, and ***basolateral plasma membrane***. All three

were enriched in the proteins that were down-regulated during differentiation and indicate a reduction in proliferation and cell-cell contact interactions.

MF terms enriched in proteins downregulated during differentiation of SH-SY5Y cells include the ***ribonucleoside-diphosphate reductase activity*** term proteins which are responsible for conversion of ribonucleotides to deoxyribonucleotide and are important for deoxyribonucleic acid (DNA) synthesis during cell replication. Proteins from ***histone demethylase activity (H4-K20 specific)*** cause de-methylation of histones may affect chromatin during differentiation.

The enriched BPs include terms that would indicate a reduction in proliferation and changes in chromatin structure.

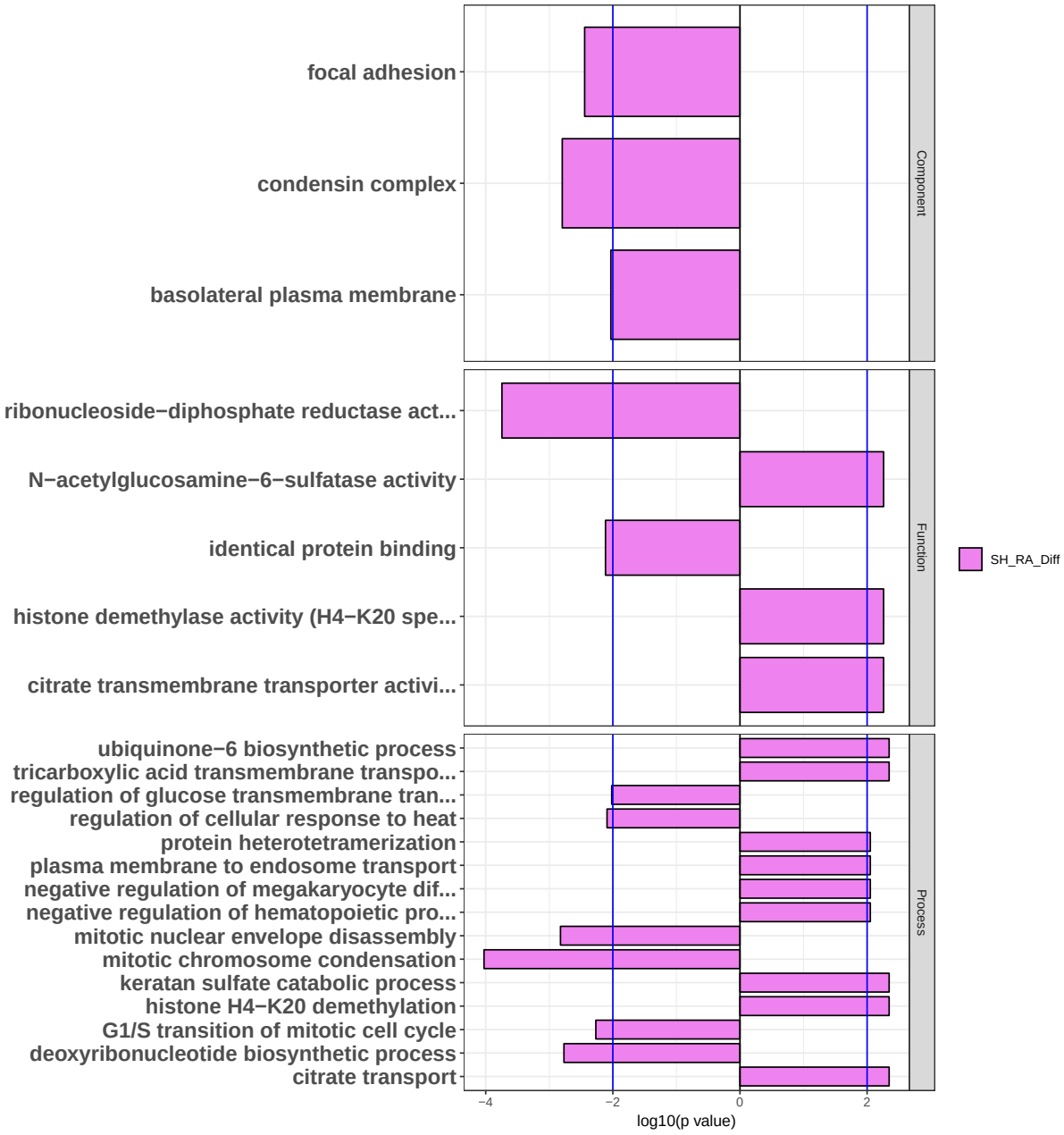


Figure 6.8: *GO for the SH-SY5Y cells differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

KEGG enrichment

KEGG enrichment indicated changes in metabolism and p53 signalling, both expected during differentiation of a cancer cell line (Figure 6.9).

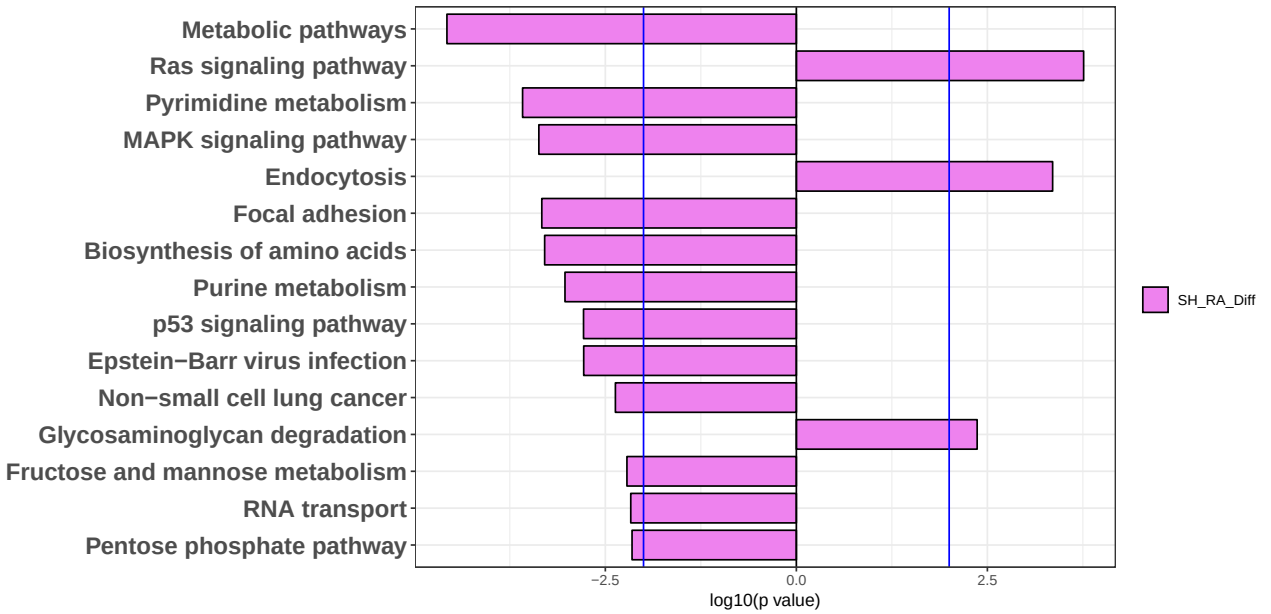


Figure 6.9: *KEGG enrichment for the SH-SY5Y cells differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

Due to the low number of proteins SDE in the SH-SY5Y cells differentiation experiment, not much useful information was gained from this experiment. The low number of SDE proteins could be due to a large degree of variability in the cells, especially the differentiating cells. Additional cell replicates together with a longer timecourse differentiation protocol, could yield better results.

Chapter 7

Results for Human ESC cells / derived NES cells

Huang et al. 2016 [102] published Affymetrix HG-U133 Plus 2.0 Genechip microarray data, comparing proliferating hESC cultured on mouse fibroblast feeder cells to neuroepithelial like stem cells derived from hESC cells *in vitro* (dNES cell). They found involvement of TGF-β and WNT/β-catenin signalling as well as DNA methylation and histone modification begin involved in the differentiation of neuroepithelial stem cells. Since this data focused on a detailed analysis of differentiation to neuroepithelial like stem cells it was incorporated into this study to put the $_{\text{hf}}$NES cell and $_{\text{hf}}$NS cell cell data into perspective.

7.0.1 Statistical re-evaluation

The data were made available by uploading it to a public repository. Upon request, the authors also provided a spreadsheet with the normalised expression values for the experiment. These data were incorporated into this study and are referred to as the $\frac{\text{dNES}}{\text{hESC}}$ experiment by Huang et al. 2016 [102] (Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment). Adding data from ESC to $_{\text{hf}}$NES cell fills an important gap in understanding the subsequent neural stem cell data. In this work the authors used three timepoints; hESC, differentiation at day six as an intermediate timepoint and differentiation at day ten as complete dNES cell differentiation. Huang et al. used ANOVA without multiple-

testing correction.[102] To make the data more comparable, using CRUNCH, the data were statistically re-evaluated using the same statistical approach that was used as with all other data in this study. Due to the increased stringency applied here, far fewer genes were SDE than in the study by Huang et al..

The dataset was divided into two groups: three hESC replicates and three replicates from the ten day dNES cell timepoints, as these are the most relevant timepoints for comparison to this work. The σ was determined at the probe level, as described before, from $\log_2$ dNES cell/hESC ratios. The σ values for this dataset were twice as high as for the Illumina bead chip microarrays, gene expression data comparing proliferating $\frac{\text{NS}}{\text{NES}}$ (GE $\frac{\text{NS}}{\text{NES}}$ experiment), described later and σ for comparison ratios was only marginally greater than for replicate ratios. This indicates a greater degree of experimental variation, but also a smaller difference between cell types than was observed in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. A 2σ cutoff of 0.3 (fold change of 1.2) was selected as a threshold for differential expression.

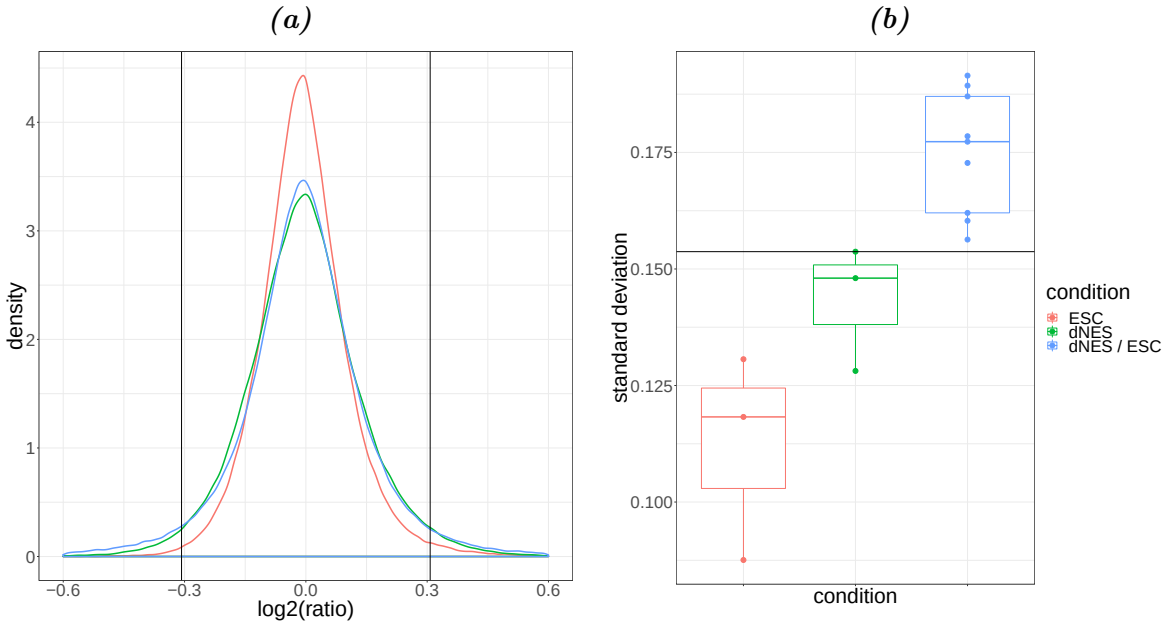


Figure 7.1: *(a) A density plot of the frequency of log_2 ratios for genes from replicates and comparisons, for the Huang $\frac{dNES}{hESC}$ experiment samples. The vertical lines indicate 2σ for replicates. (b) Boxplot of σ for Huang $\frac{dNES}{hESC}$ experiment replicates and comparisons.*

Affymetrix probe ID's were converted to gene symbols. There is a many-to-one ratio between the probes and the gene symbols. Using a t test all probe values for hESC, and

dNES cell were compared for each gene symbol. From the p-values, the FDR was calculated as 51.17%, further supporting the notion that there was a large degree of variability in this data. Using BH corrected p-values smaller than 0.05 and mean $\log_2$ ratios greater than 0.3 (fold change 1.2), 214 genes were considered SDE. Of these 129 and 85 genes were down- and up-regulated, respectively during the transition from hESC to dNES cells. These lists were used for KEGG and GO enrichment, mapped to STRINGdb networks, and compared to the other data in this work.

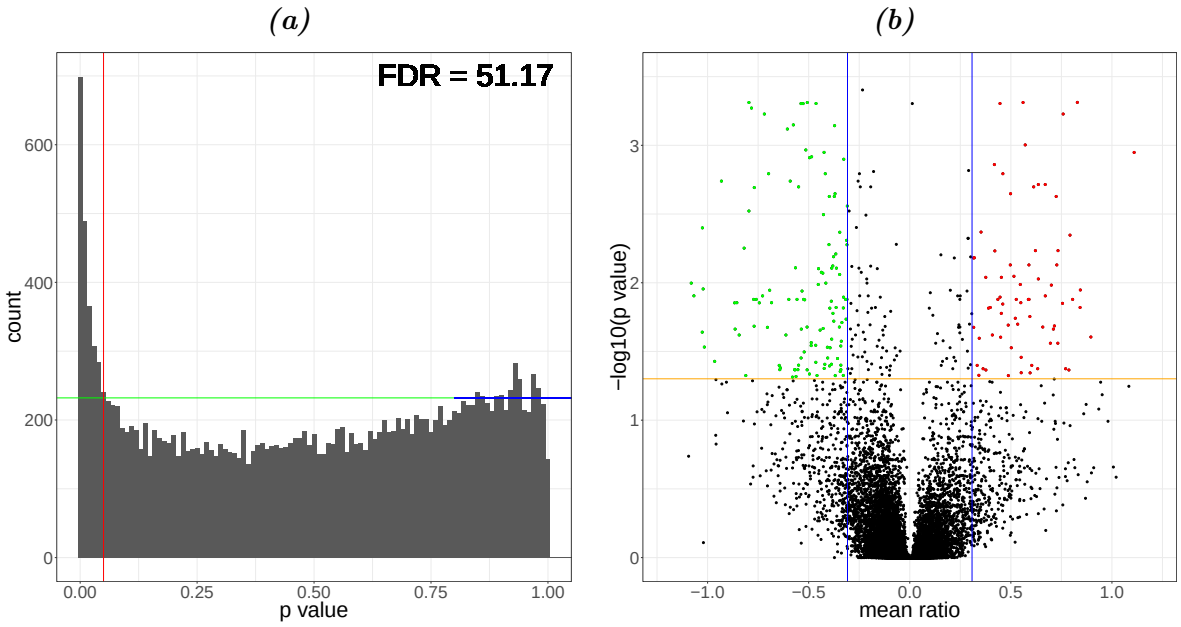


Figure 7.2: *(a) FDR plot for the Huang $\frac{dNES}{hESC}$ experiment t test p-values. The mean frequency between 0.8 and 1 was extrapolated to represent p-values generated by random chance (blue line). The FDR is the percentage of all proteins with a p-value smaller than 0.05 (red line) and those with a p-value smaller than 0.05 generated by random chance. (b) The volcano plot plots the log2 mean ratio of the comparison vs the -log10 of the p-value. The vertical blue lines depict the 2σ of the replicate (0.3). The horizontal red line marks a p-value of 0.05. All values above the red line and outside of the green lines represent significantly differentially expressed (SDE) genes*

Although the dataset differer, most significantly in the variances within the datasets, it appears that there is less difference between the hESC and dNES cells than between the $_{hf}$NES cells and $_{hf}$NS cells.

Chapter 8

Neural stem cell results

8.1 Proliferating $_{hf}$NES cells vs Proliferating $_{hf}$NS cells

8.1.1 Microscope images

$_{Hf}$NES cells were compared to $_{hf}$NS cells visually. Both cell lines were cultured in neural stem cell medium containing EGF and FGF. The cells were plated on flat-bottomed cell culture dishes and visualised using bright-field microscopy and fluorescent microscopy. $_{hf}$NES cells have an epithelial like morphology with neural tube characteristic rosettes (Figure 8.1). The $_{hf}$NS cells have a RG-like appearance with prominent cellular processes (Figure 8.3). $_{hf}$NES cells express *TUBB3* and *NES*, but not *GFAP*, (Figure 8.2). $_{hf}$NS cells express *TUBB3* and *GFAP*, and some *NES* (Figure 8.4). The immunohistochemical staining for *NES* does not appear to be homogeneous, a sub-population of cells stain positive for *NES*. This may indicate heterogeneity in the cell populations.

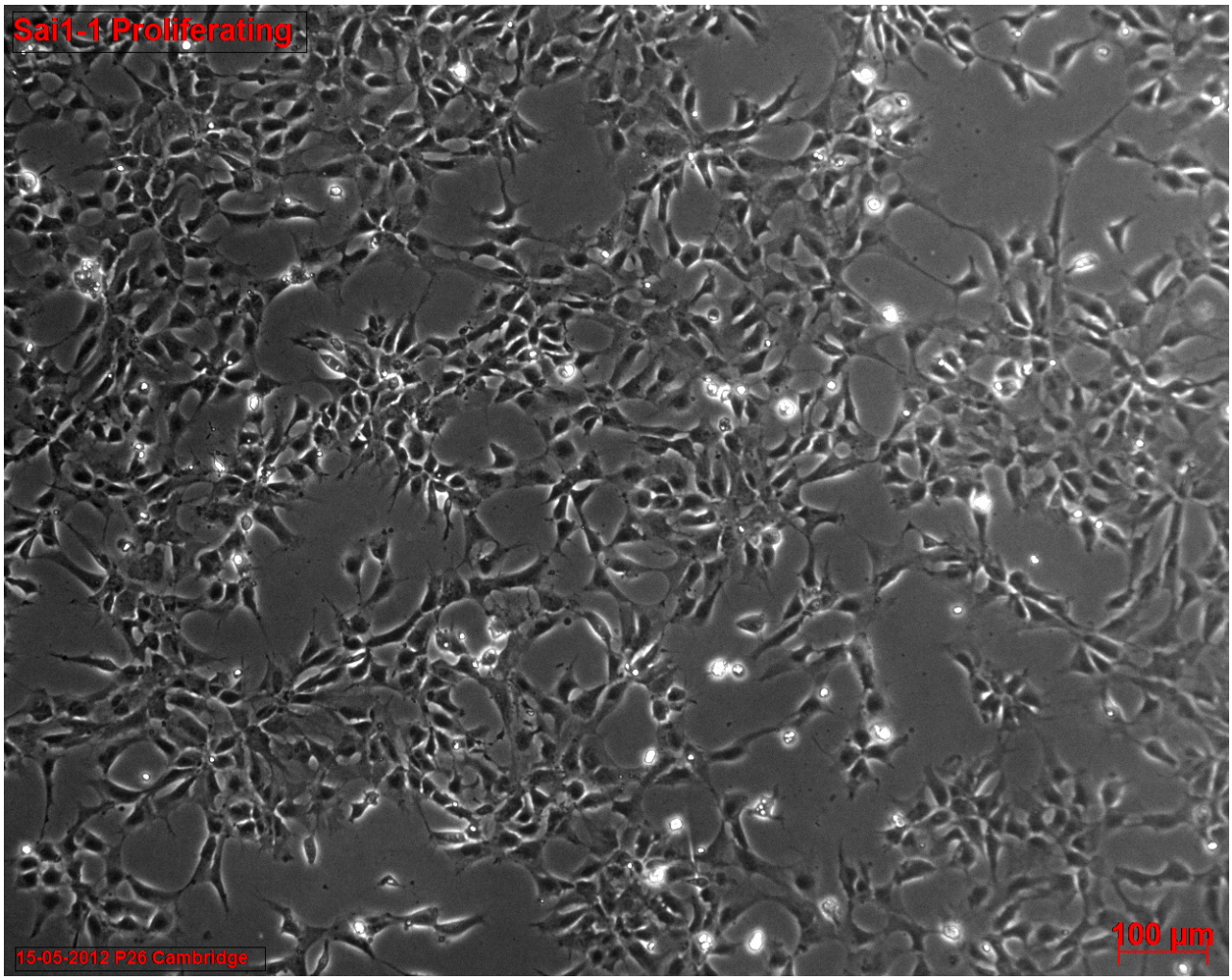


Figure 8.1: *Brigh-field image of proliferating $_{hf}$NES cells*

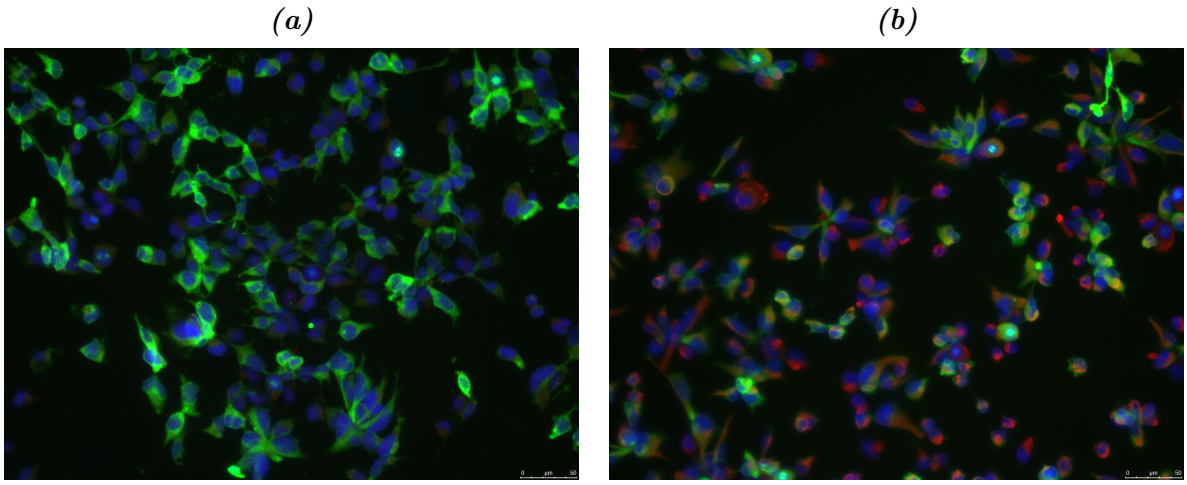


Figure 8.2: *Proliferating $_{hf}$NES cells stain positive for anti-TuJ-1 in green (a and b), negative for anti-GFAP in red (a), Some cells stain positive for anti-Nestin in red (b)*

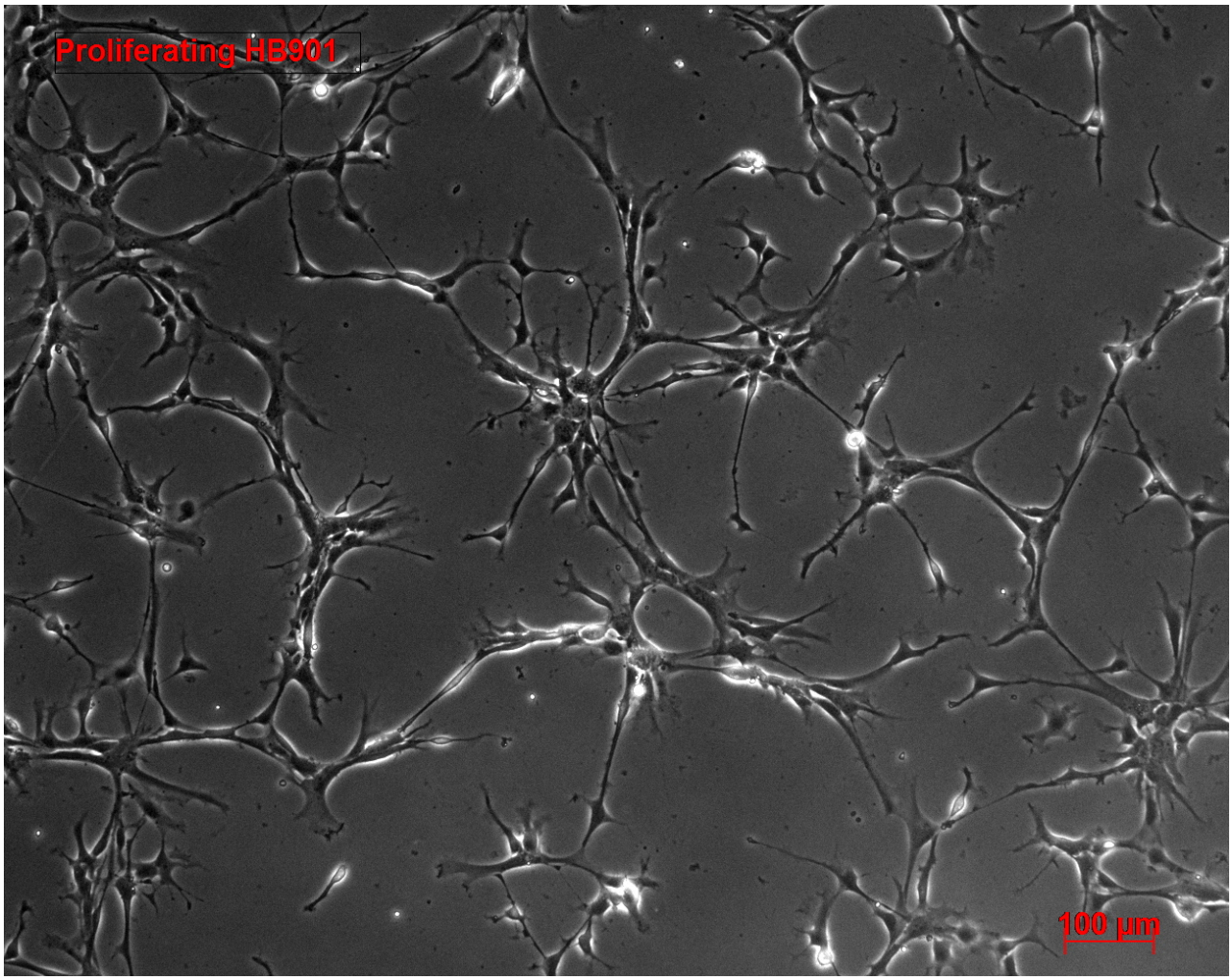


Figure 8.3: *Brigh-field image of proliferating $_{hf}$NS cells*

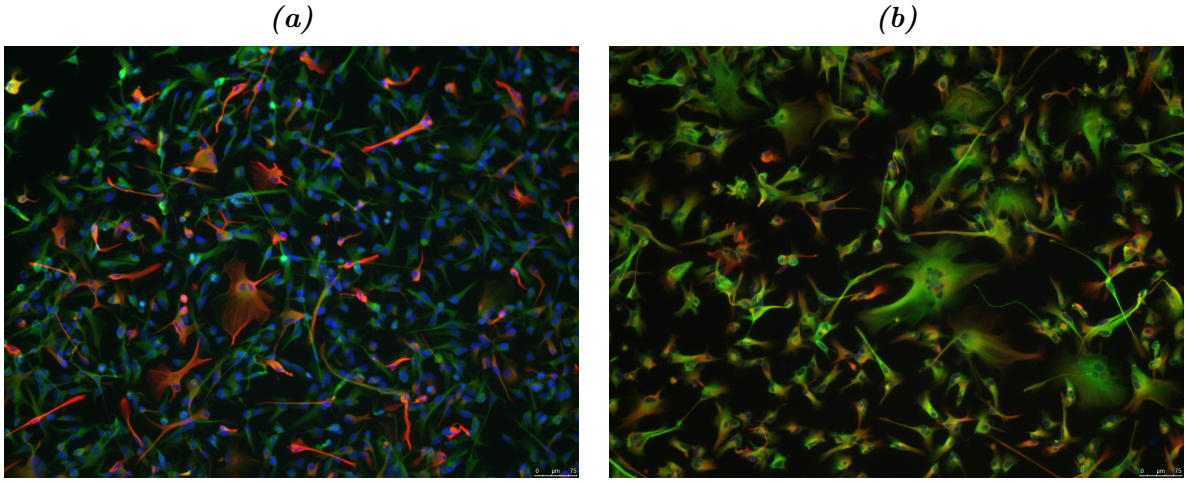


Figure 8.4: *Proliferating $_{hf}$NS cells stain positive for anti-TuJ-1 in green (a and b). Positive for anti-GFAP in red (a). Some cells stain positive for anti-Nestin in red (b).*

8.1.2 Gene expression data

GE $\frac{\text{NS}}{\text{NES}}$ experiment were provided by the Smith lab, as previously described by Tailor et al. 2013. Within this dataset, only the $_{\text{hf}}$NES cell and $_{\text{hf}}$NS cell samples were used. Ratios were generated for replicates ($^{\text{hf}}{}^{NScells}/_{\text{hf}}{}_{NScells}$ and $^{\text{hf}}{}^{NEScells}/_{\text{hf}}{}_{NEScells}$), and for the comparisons $^{\text{hf}}{}^{NScells}/_{\text{hf}}{}_{NEScells}$. The $\log_2$ ratios were plotted using a density plot. Replicates had a narrow distribution and small σ compared to the comparisons (Figure 8.5 (a)), indicating a greater difference between comparisons than replicates. When σ's were calculated separately between the $\log_2$ ratios for the replicates, they had a maximum σ of 0.06 and 0.05 respectively (Figure 8.5 (b)). Comparison ratios yielded a maximum σ of 0.11. From this data a 2σ technical variance of 0.12, which is equal to a fold change of 1.087, was used as a measure of differential expression for the comparison of samples $^{\text{hf}}{}^{NScells}/_{\text{hf}}{}_{NEScells}$.

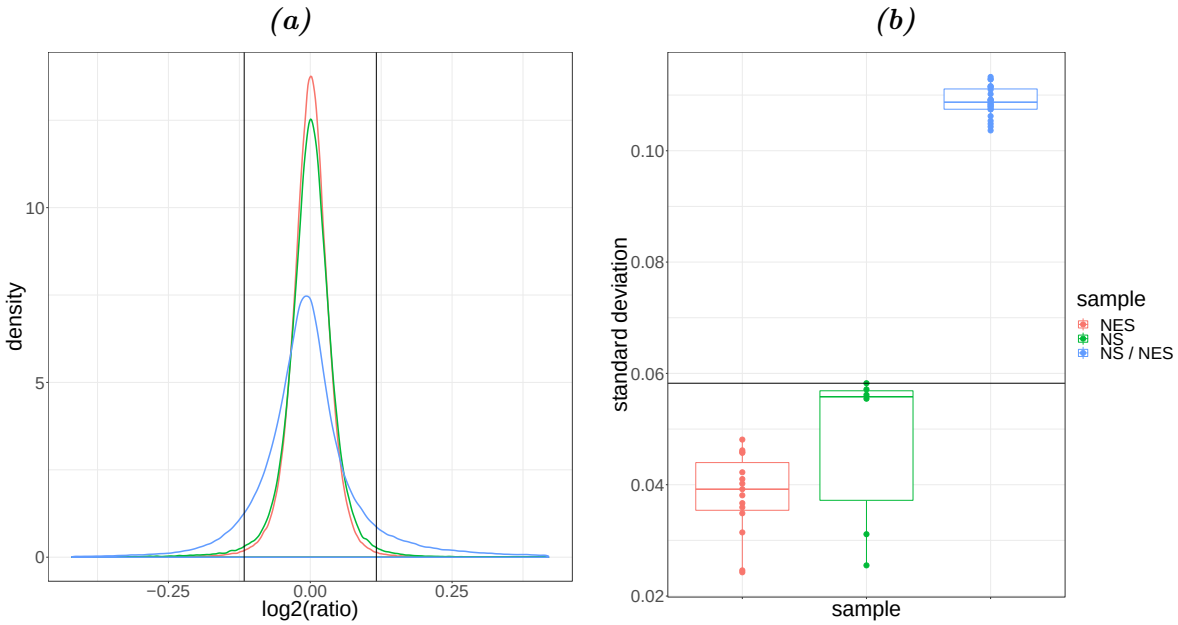


Figure 8.5: *(a) A density plot indicating the frequency of $\log_2$ ratios. Blue curves represent the replicates, which have a tighter distribution than the more broad comparisons ratios in red. Green and orange horizontal lines indicate 2σ for the replicates and comparisons, respectively. (b) A boxplot of the σ for replicates and comparisons.*

A t test was applied comparing $_{\text{hf}}$NES cells expression values to $_{\text{hf}}$NS cells expression values. A histogram of p-value shows an exponentially decaying distribution. FDR for t test p-values was calculated as 4.63% (Figure 8.6 (a)). The p-values were corrected using BH multiple-

testing correction. Comparison $\log_2$ mean ratios and -$\log_{10}$ p-values were plotted on a volcano plot to illustrate the spread of datapoints (Figure 8.6 (b)). After applying a 2σ threshold of 0.12 (fold change 1.087) to the mean $\log_2$ ratios together with BH-corrected p-values of less than 0.05, 1817 genes were SDE, with 1024 upregulated, or more highly expressed in $_{\text{hf}}$NS cells and 793 downregulated, or more highly expressed in $_{\text{hf}}$NES cells

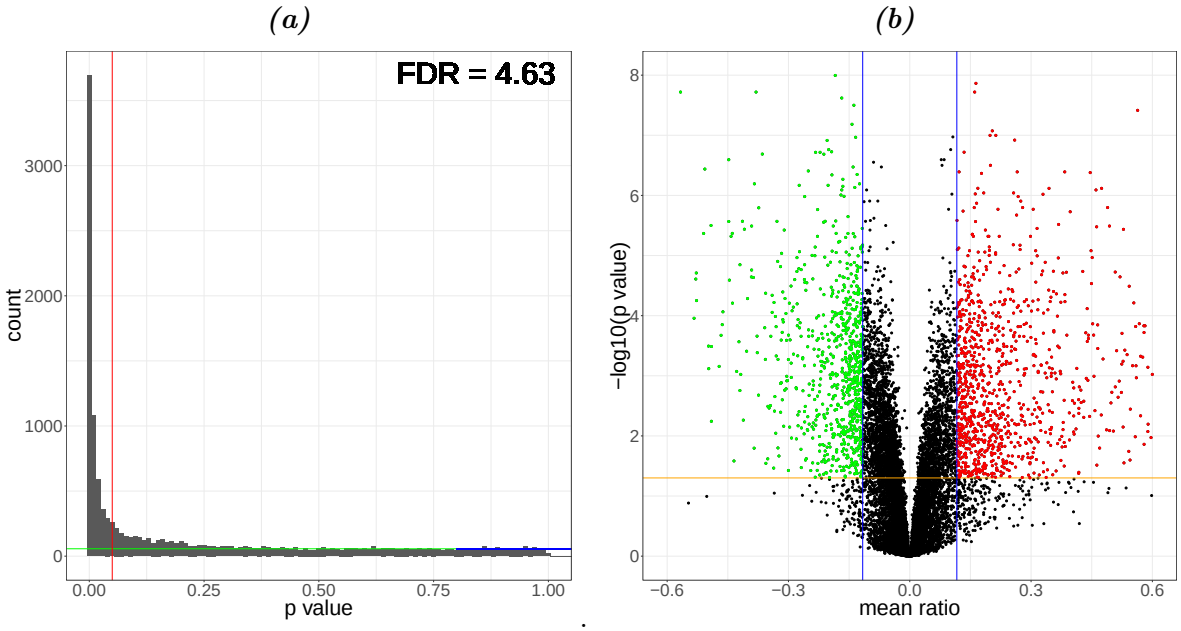


Figure 8.6: t test FDR and volcano plots for the GE. *(a) Histogram of the t test p-values for all GE $\frac{NS}{NES}$ experiment log_2 ratios. (b) Volcano plot, log2 $^{NS}/_{NES}$ ratios on x axis vs -log10(p-values) on the y axis. The green horizontal lines indicate the technical replicate 2σ, while the blue horizontal lines indicate the comparison 2σ. The red line indicates a p-value of 0.05. All values above the red line and outside of the green lines were considered SDE.*

8.1.3 SILAC experiment

MaxQuant summary SILAC labelling was done in three separate experiments: SILAC 1, SILAC 2, and SILAC 3. (Table 8.2). In each experiment, SILAC incorporations were tested and technical replicates were analysed to determine the experimental variance. The exact sample compositions of all experiments can be seen in Table 8.4. The Q Exactive generated raw files were processed via MQ using the parameters listed in Table 6.5. FDR was set to 0.1% at all levels (PSM, peptide, and protein). Less than 1.2% contaminants were identified and there were fewer than 30% missed cleavages in all samples (Table 8.2). SILAC heavy label

incorporation was shown to be greater than 95% in all cases. (Table 8.3).

Table 8.1: *Summay of the mass spectra collected in the SILAC $\frac{NS}{NES}$ experiment*

Raw file	**MS**	**MS/MS**	**MS/MS Identified**	**MS/MS Identified [%]**	**Peptide Sequences Identified**	**Isotope Patterns Repeatedly Sequenced [%]**
SILAC_1	98856	158198	65191	38.03	32870	5.99
SILAC_2	331531	810618	241308	24.04	22448	15.4
SILAC_3	651028	1143790	336942	22.2	24411	17.6

Table 8.2: *SILAC $\frac{NS}{NES}$ experiment experimental overview table*

Experiment	Peptides	Proteins	FDR	Contaminants	Missed Cleavages
SILAC 1 replicates	6085	921	0.9%	1.1%	6.96%
SILAC 1	32870	4406	1.2%	0.5%	7.75%
SILAC 2 replicates	5138	115	1.0%	1.2%	24.38
SILAC 2	22150	3406	1.1%	0.8%	29.36%
SILAC 3 replicates	10017	1658	1.1%	0.4%	16.75%
SILAC 3	24146	3524	1.1%	0.4%	16.74%

Table 8.3: *SILAC heavy label incorporation test results*

Experiment	Sample	# light Peptides	# heavy Peptides	Incorporation rate
SILAC 1	HB1025	51	1170	95.8%
SILAC 2	HB901	41	4582	99.1%
SILAC 2	HB1025	108	4582	97.7%
SILAC 2	Sai1	58	4183	98.6%
SILAC 3	HB901 1	84	3351	99.6%
SILAC 3	HB901 2	68	3326	98.0%
SILAC 3	Sai3 1	49	2746	97.2%
SILAC 3	Sai3 2	68	3039	97.8%

Samples from all experiments were combined and statistically assessed as a single experiment. A total of 4391 proteins were detected from sample SILAC_1_S2L_H10H in SILAC 1,

Table 8.4: *SILAC : Full protein number of all samples*

Sampe Name	Experiment	type	light	heavy	# proteins
SILAC_1_H10	SILAC 1	technical replicate	HB1025	HB1025	926
SILAC_1_S2L_H10H	SILAC 1	comparison	Sai2	HB1025	4391
SILAC_2_HS1.LH9	SILAC 2	reverse comparison	Sai1	HB901	1696
SILAC_2_LH9_HH9	SILAC 2	technical replicate	HB901	HB901	397
SILAC_2_LS1_HH10	SILAC 2	comparison	Sai1	HB1025	1811
SILAC_2_LS1.HH9_07	SILAC 2	comparison	Sai1	HB901	1728
SILAC_2_LS1_HS1	SILAC 2	technical replicate	Sai1	Sai1	593
SILAC_2_LS2_HH10	SILAC 2	comparison	Sai2	HB1025	1506
SILAC_2_LS2.HH9	SILAC 2	comparison	Sai2	HB901	1563
SILAC_3_S3L1H1	SILAC 3	technical replicate	Sai3	Sai3	561
SILAC_3_S3L2H2	SILAC 3	technical replicate	Sai3	Sai3	516
SILAC_3_S3L1.H9H1_06	SILAC 3	comparison	Sai3	HB901	1807
SILAC_3_S3L2.H9H2_03	SILAC 3	comparison	Sai3	HB901	1542
SILAC_3_S3H1.H9L1_06	SILAC 3	reverse comparison	HB901	Sai3	2070

while fewer than 2000 were detected in all other samples. This is a result of high pH RP fractionation being applied in SILAC 1 while MudPIT was used in SILAC 2 and SILAC 3, (Figure 8.7). There is however a high degree of overlap across the samples. A total of 1571 quantitated protein were unique to SILAC_1_S2L_H10H and 2901 and 2370 protein groups were detected in at least 2 and 3 samples, respectively.

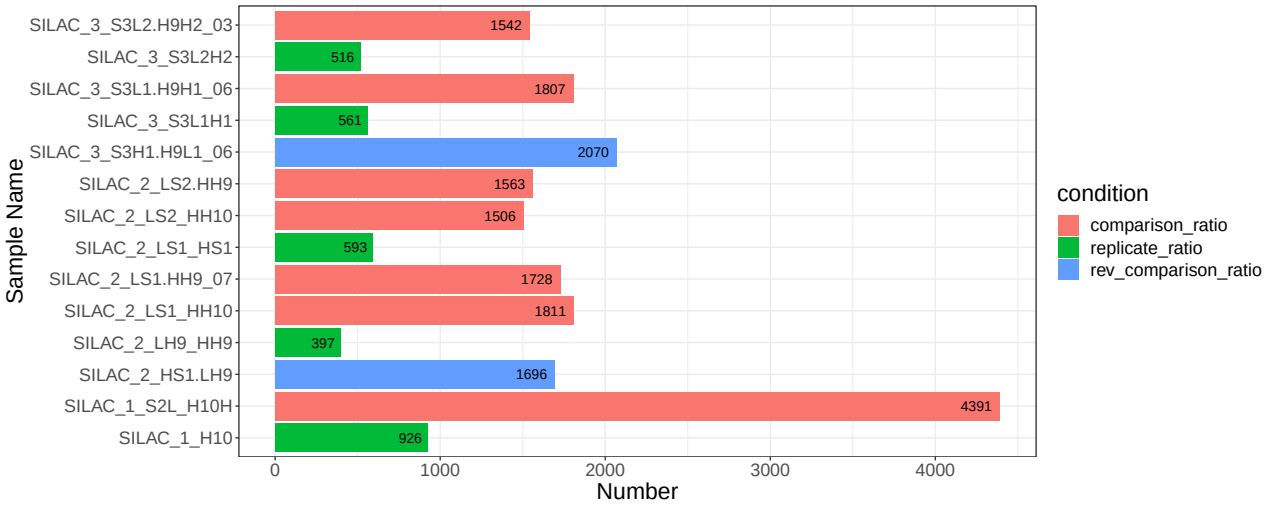


Figure 8.7: *Barplot of the number of proteins identified in the SILAC $\frac{NS}{NES}$ experiment.*

Quality control SILAC comparison samples comprised 100 μg of heavy $_{hf}$NS cells and 100 μg of light $_{hf}$NES cells samples combined. After analysis, quantitation was reflected as H/L ratios. In order to asses the accuracy of mixing, MA plots were generated for each of the comparison samples. PSM, peptide, and proteins intensities were compared to their H/L ratios. Accurate mixing resulted in the H/L ratios being distributed across zero at all intensities. In Figure 8.8, reverse labelled samples from SILAC 3 were compared. MA plots from these samples are mirror images of each other, indicating there is no sample labelling bias.

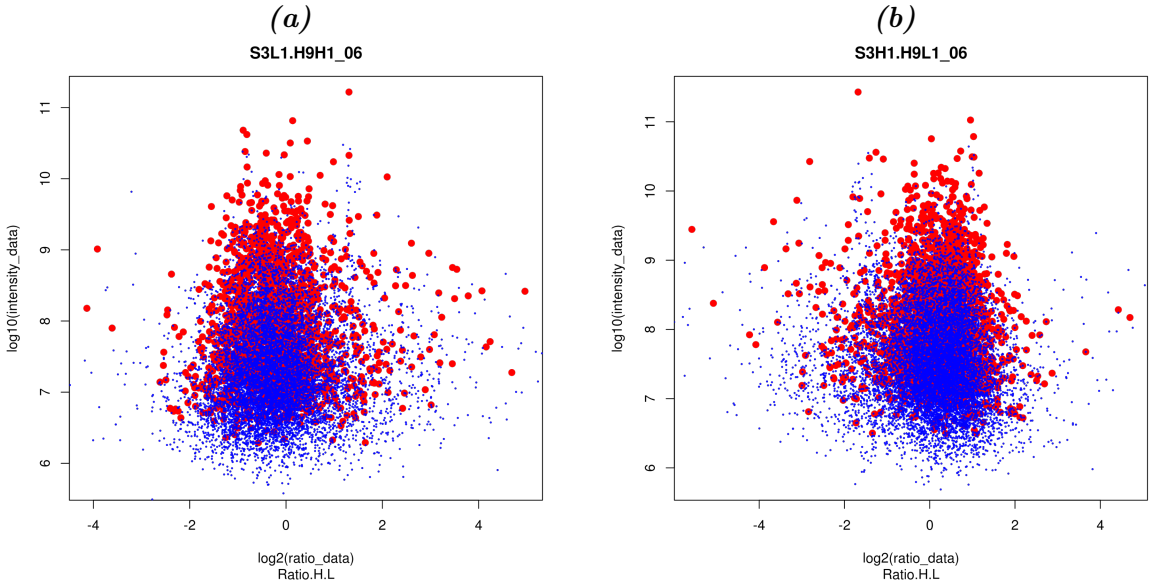


Figure 8.8: *MA plot of two oppositely labelled SILAC samples from SILAC 3 (a) S3L1_H9H1_SILAC_3 and (b) S3H1_H9L1_SILAC_3. The x-axis is the H/L ratio, while the y-axis is the log10 of the intensity data. Black dots represent PSMs, blue represent peptides, and red represent proteins*

In addition to MA plots, the $\log_2$ ratios were plotted on density plots (Figure 8.9a). Like the GE $\frac{\text{NS}}{\text{NES}}$ experiment, there was less variance amongst the replicates than the comparisons. Two of the comparison samples, SILAC_2_LS1_HH10 and SILAC_2_LS1.HH9_07 did not show as much variance as the other samples, but they were still substantially different from the replicates, so these were included. The technical replicates have a maximum σ of 0.29, thus a 2σ of 0.58 (fold change of 1.49) was used as threshold for substantial differential expression (Figure 8.9b).

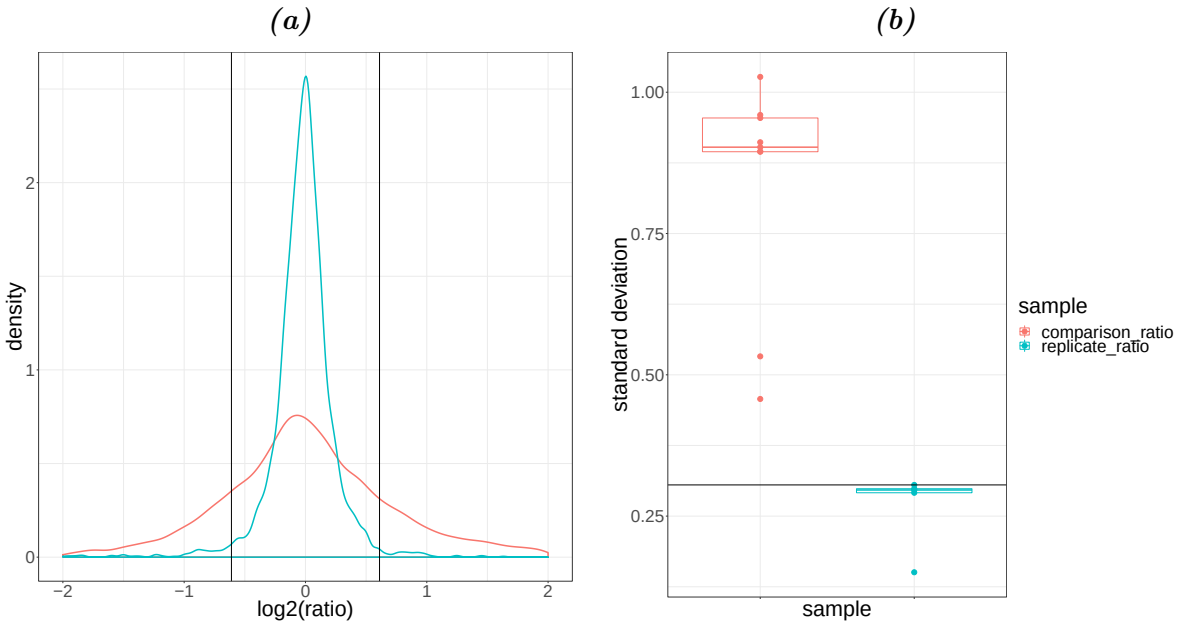


Figure 8.9: *(a) Density plot for the SILAC $\frac{NS}{NES}$ experiment. Blue and red curves represent the replicates and comparisons respectively. The green horizontal line indicates the 2σ of the replicates. (b) Boxplot of the σ for replicates and comparisons*

Upon hierarchical clustering (Figure 8.10), data from different experiments, were evenly distributed across clusters. Samples where labels were reversed, clustered with the normal samples after the $\log_2$ ratios had been inverted, another indication that there was minimal labelling bias (Figure 8.10).

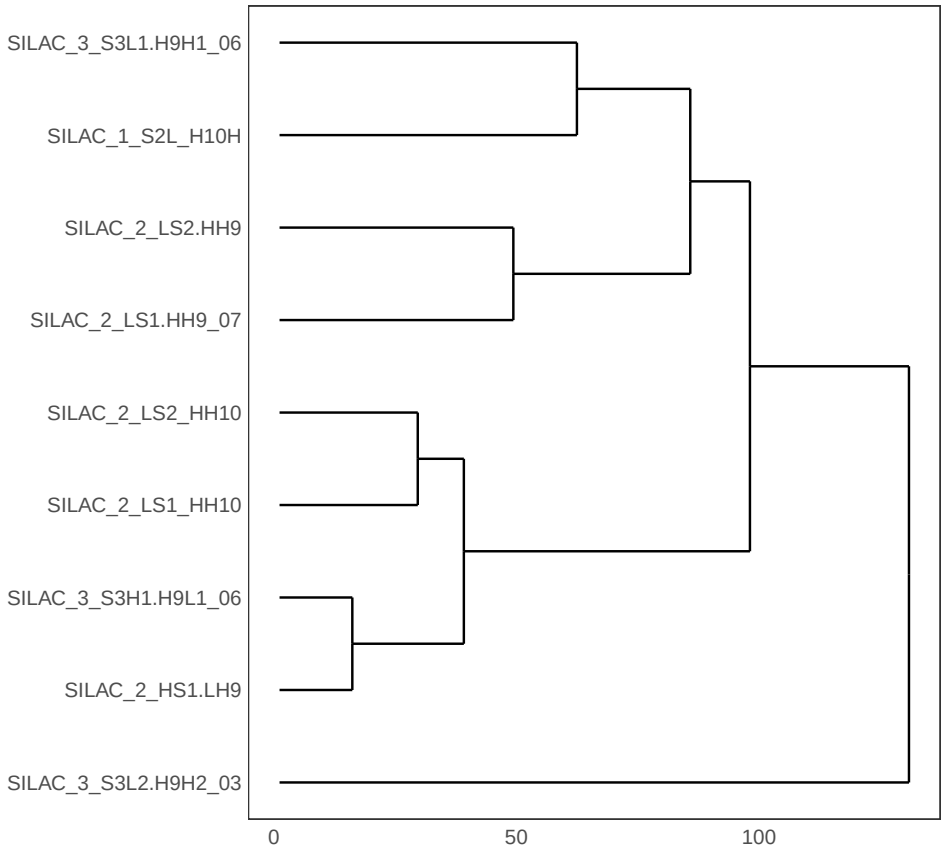


Figure 8.10: *SILAC $\frac{NS}{NES}$ experiment samples clustered by hierarchical clustering*

Statistics T tests were applied to all comparison H/L $\log_2$ ratios against a mean of zero for all proteins detected in two or more samples. The resultant FDR was 3.19% (Figure 8.11). The p-values were corrected using BH multiple-testing correction. The H/L ratios and BH-corrected p-value were plotted on a volcano plot (Figure 8.11). After selection of proteins with a BH-corrected p-value of smaller than 0.05 and mean $\log_2$ ratio greater than the replicate 2σ threshold of 0.61, 473 proteins were considered SDE. A total of 177 and 296 proteins were down or up-regulated respectively.

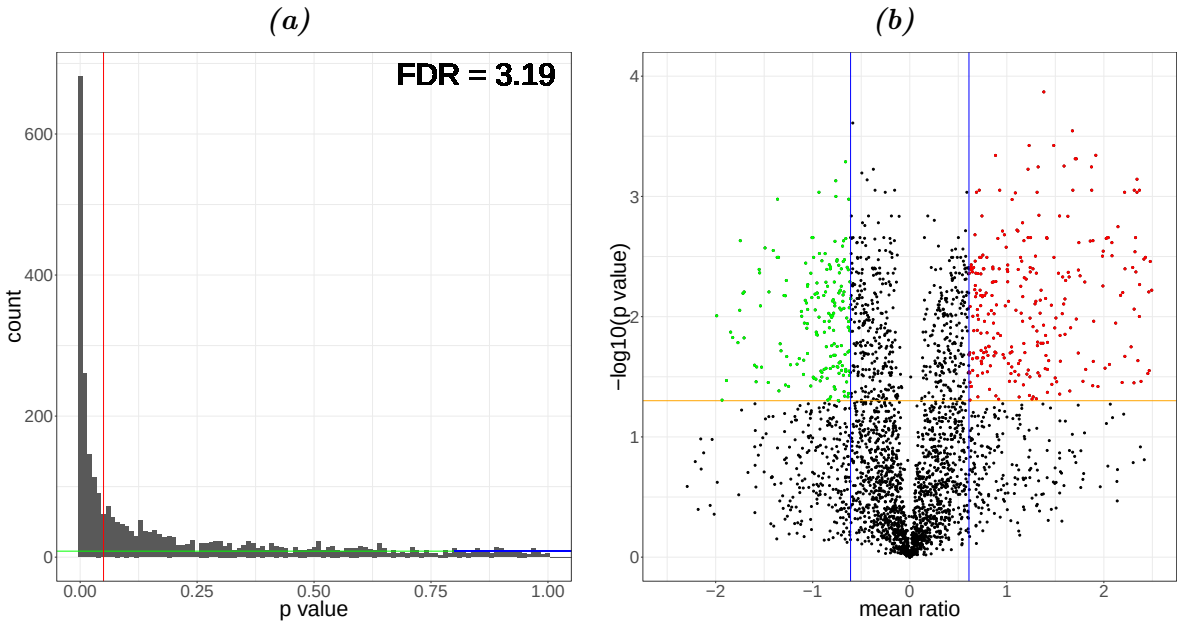


Figure 8.11: *(a) FDR plot for BH-corrected p-values. The mean of frequency between 0.8 and 1 was extrapolated to represent p-values generated by random chance (blue line). The FDR is a percentage of the number of p-values below 0.05 (red line) and those with a p-value below 0.05 generated by random chance. (b) The volcano plot, shows the log_2 mean ratio of the comparison vs the -log_{10} of the p-value. The horizontal green and blue lines depict the 2σ of the replicate (0.92) or comparisons (1.32), respectively. The vertical red line marks a p-value of 0.05. All values above the red line and outside of the green line were considered SDE.*

8.1.4 Collating GE $\frac{\text{NS}}{\text{NES}}$ experiment and SILAC $\frac{\text{NS}}{\text{NES}}$ experiment data

Of the 473 proteins significantly differentially expressed in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, 166 showed similar SDE to genes in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. Furthermore, 307 proteins were SDE in the the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment but not in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. An additional 1651 genes were differentially expressed in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 8.12). Approximately one third of the proteins SDE in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment were also SDE in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. There were not conflicts between in SDE between the datasets. Those not showing overlapping SDE were either dysregulated but did not pass the criteria for SDE or there are post-translational regulation mechanisms at play here.

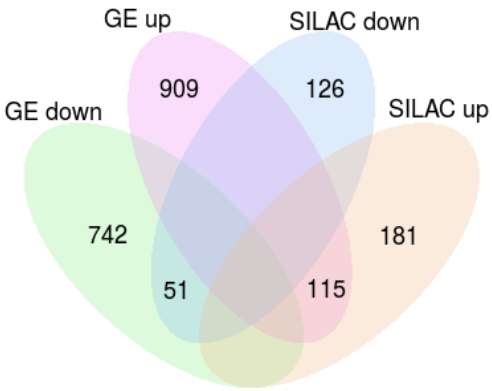


Figure 8.12: *Venn diagram comparing proteins and genes SDE in the SILAC* $\frac{NS}{NES}$ *experiment and GE* $\frac{NS}{NES}$ *experiment.*

8.2 NSC differentiation experiment

8.2.1 Immunohistochemistry

Human foetal derived neural stem cells are differentiated by removing the growth factors EGF and FGF from their culture medium. The cells spontaneously differentiate, $_{hf}$NES cells differentiate into an almost pure population of neurons,[211] whereas $_{hf}$NS cells differentiate into a mixed population of neurons and astrocytes and have the potential to differentiate into oligodendrocytes.[206] Immunohistochemical staining showed all the cells staining positive for *TUBB3* by anti-TuJ-1. *TUBB3* was up-regulated in the SILAC $\frac{NS}{NES}$ experiment (Figure 9.35). The $_{hf}$NES cells and $_{hf}$NS cell stained positive for anti-Nestin. *NES* was not SDE in the SILAC $\frac{NS}{NES}$ experiment (Figure 9.25). The $_{hf}$NS cell stained more positive for anti-GFAP (Figure 8.14), confirmed by the SILAC $\frac{NS}{NES}$ experiment (Figure 9.26).

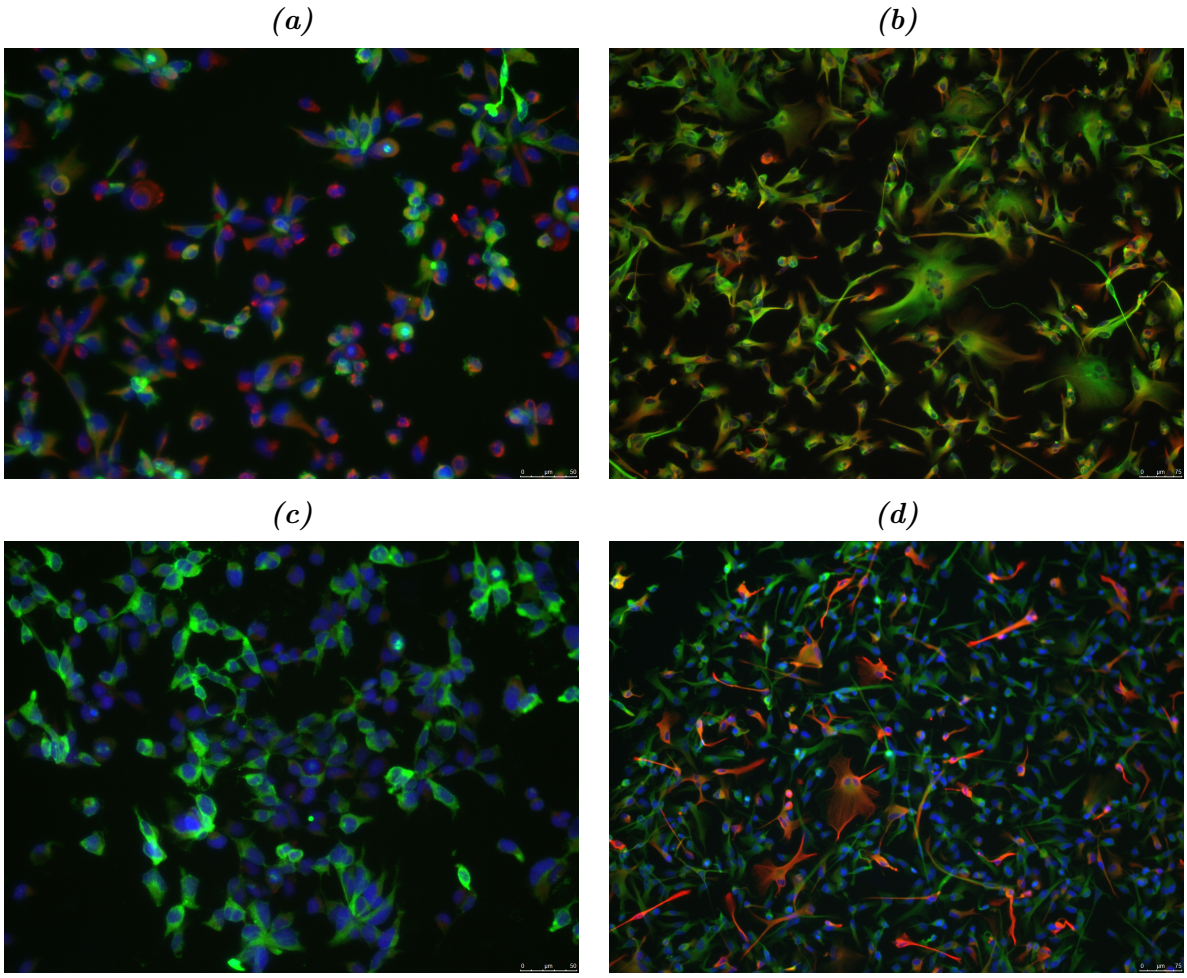


Figure 8.13: *Immunohistochemistry of differentiating human foetal neural stem cells. (a) $_{hf}$NES cells stained with anti-Nestin (red), (b) $_{hf}$NS cells stained with anti-Nestin (red), (c) $_{hf}$NES cells stained with anti-GFAP (red), and (d) $_{hf}$NS cells stained with anti-GFAP (red). All cells were stained with anti-TuJ-1 (green) and nuclei stained with DAPI (blue).*

Upon differentiation the $_{hf}$NES cells still stained positive for anti-Nestin, despite it being down-regulated in the MaxLFQ $_{hf}$NES cell differentiation timecourse experiment ($_{hf}$NES cell differentiation experiment) (Figure 9.25). The cells took on a neuronal morphology forming interconnected networks (Figure 8.14 (a)). The $_{hf}$NS cells were drawn together forming high density tracts, leaving other areas virtually empty. These cell tracts become more defined as differentiation progressed. This may have been due to glial cell attachment creating more complex tissue like structures (Figure 8.14 (b)). Upon differentiation *TUBB3* was up-regulated and *NES* was down-regulated in the $_{hf}$NES cells, while *GFAP* was upregulated in the $_{hf}$NS cells (Figure 8.14).

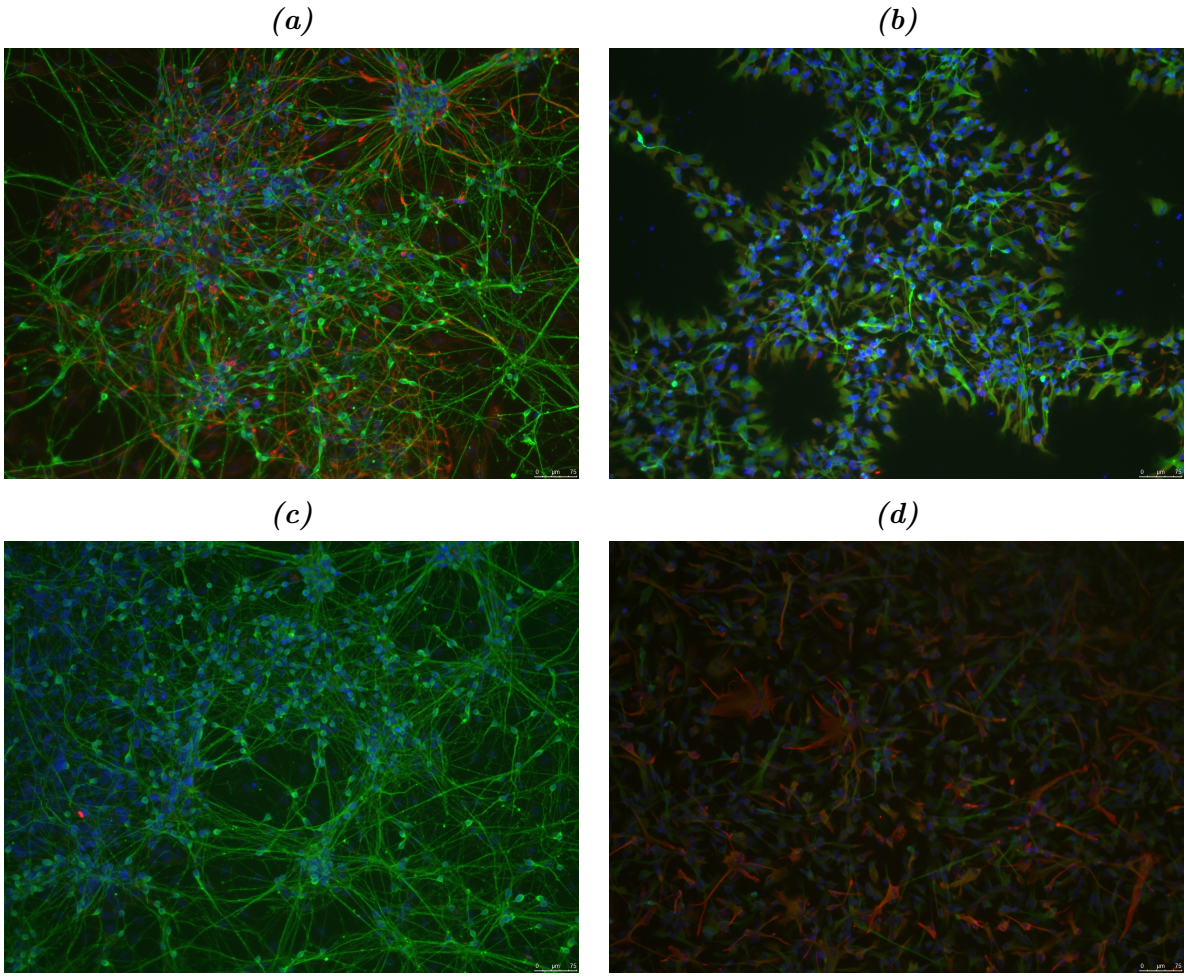


Figure 8.14: *Immunohistochemistry of differentiating human foetal neural stem cells. (a) $_{hf}$NES cells stained with anti-Nestin (red), and (b) $_{hf}$NS cells stained with anti-Nestin (red), (c) $_{hf}$NES cells stained with anti-GFAP (red), and (d) $_{hf}$NS cells stained with anti-GFAP (red), All cells were stained with anti-TuJ-1 (green) and nuclei stained with DAPI (blue).*

In areas where $_{hf}$NES cells were cultured at high density the nuclei group together forming a concentrated cluster out of which neuronal tracks extended into the surrounding cells (Figure 8.15 (a)). In areas where the $_{hf}$NS cells were sparsely cultured, individual cells formed astrocytes that stained strongly for anti-GFAP (Figure 8.15 (b)).

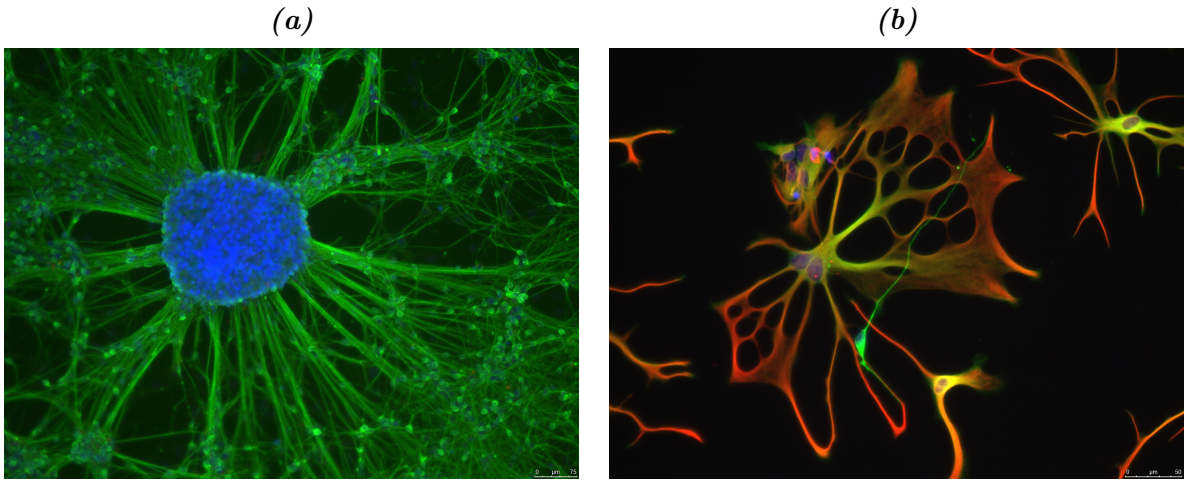


Figure 8.15: *Immunohistochemistry of differentiating human foetal neural stem cells. (a) $_{hf}$NES cells stained with anti-GFAP (red), and (b) $_{hf}$NS cells stained with anti-GFAP (red). All cells were stained with anti-TuJ-1 (green) and nuclei stained with DAPI (blue).*

In the neural stem cell differentiation experiment, six timepoints day 0,1,2,4,8, and 12 were taken across five cell lines, (three $_{hf}$NES cell and two $_{hf}$NS cell lines). No cell culture technical replicates were generated but the cell lines serve as biological replicates, since they have unique biological origins. Each of the samples was analysed in triplicate by a Q Exactive, and these mass spectrometery (MS) technical replicates were combined as a single sample during MQ analysis. The samples are labelled cell line (S1, S2, and S3) for the three $_{hf}$NES cell lines, Sai1, Sai2 and Sai3, respectively. The two $_{hf}$NS cell lines were labelled H9 and H10 for HB901 and HB1025 cell lines, respectively. A total of 5022 proteins were detected with an average of 2631 proteins per sample. Of the 5022 proteins, 872 were found in only one sample while 1036 were detected across all samples (Figure 8.16). Due to low TIC intensities, some samples were excluded from the experiment. There was variability in the total number of detected proteins per sample, probably due to variation in sample preparation.

Ratios were generated by dividing all LFQ intensities by the first measurement made for that protein. The resulting ratios showed clear upwards or downwards trends as differentiation progressed. Density plots showed increased spreading with successive days of differentiation (Figure 8.17). Indicating greater degrees of change as differentiation progressed.

An ANOVA was performed on a per-protein basis. The histogram of the ANOVA p-values

for the $_{hf}$NES cells and $_{hf}$NS cells showed exponentially decaying distributions for both datasets. Their respective FDRs were 16.71% and 14.12%, for $_{hf}$NES cells and $_{hf}$NS cells, respectively (Figure 8.18). Histograms of the linear regression slope from the ANOVA data showed normal distributions for both datasets and volcano plots indicated clear separation of slopes with decreasing BH-corrected p-values (Figure 8.19). After statistical analysis of the data, there were 243 $_{hf}$NES cells proteins with a BH-corrected p-values smaller than 0.05. Of these, 134 were downregulated and 109 were upregulated. In the $_{hf}$NS cells, 177 $_{hf}$NS cells proteins had a BH-corrected p-values smaller than 0.05, with 69 of these upregulated and 108 downregulated.

Table 8.5: *Summary of the number of mass spectra detected in the*

Raw file	**MS**	**MS/MS**	**MS/MS Identified**	**MS/MS Identified [%]**	**Peptide Sequences Identified**	**Isotope Patterns Repeatedly Sequenced [%]**
Total	784195	3458467	1322619	32.14	48371	25.25

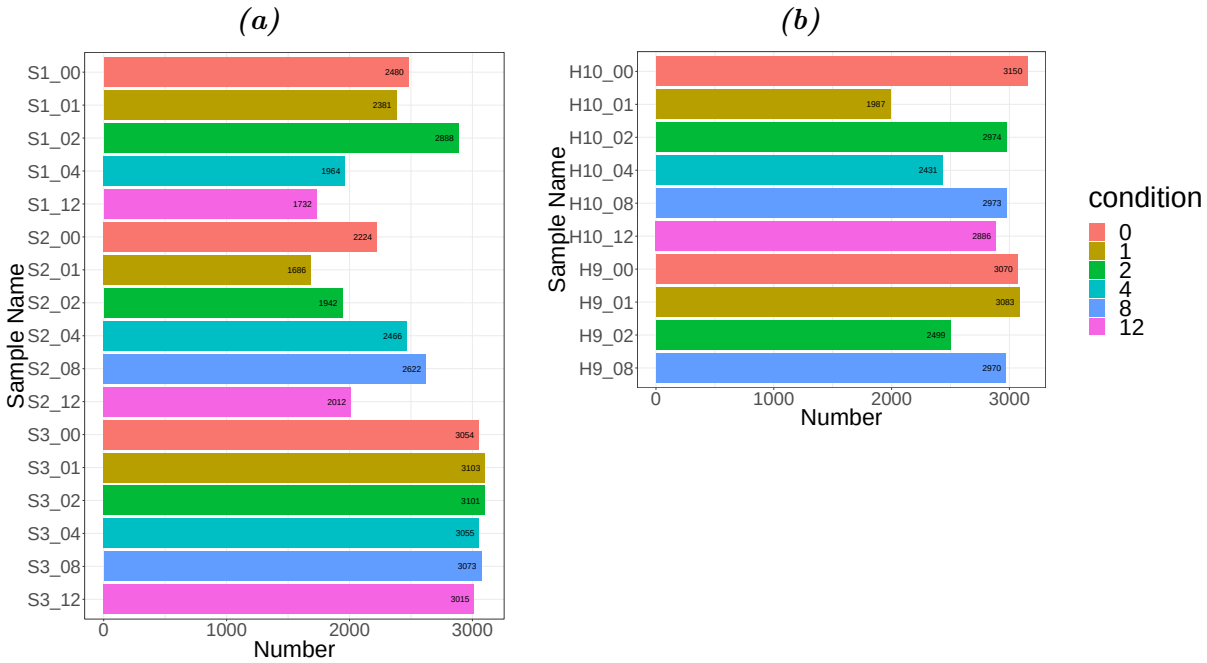


Figure 8.16: *Protein numbers for the (a) $_{hf}$NES cell differentiation experiment and (b) the MaxLFQ $_{hf}$NS cell differentiation timecource experiment ($_{hf}$NS cell differentiation experiment).*

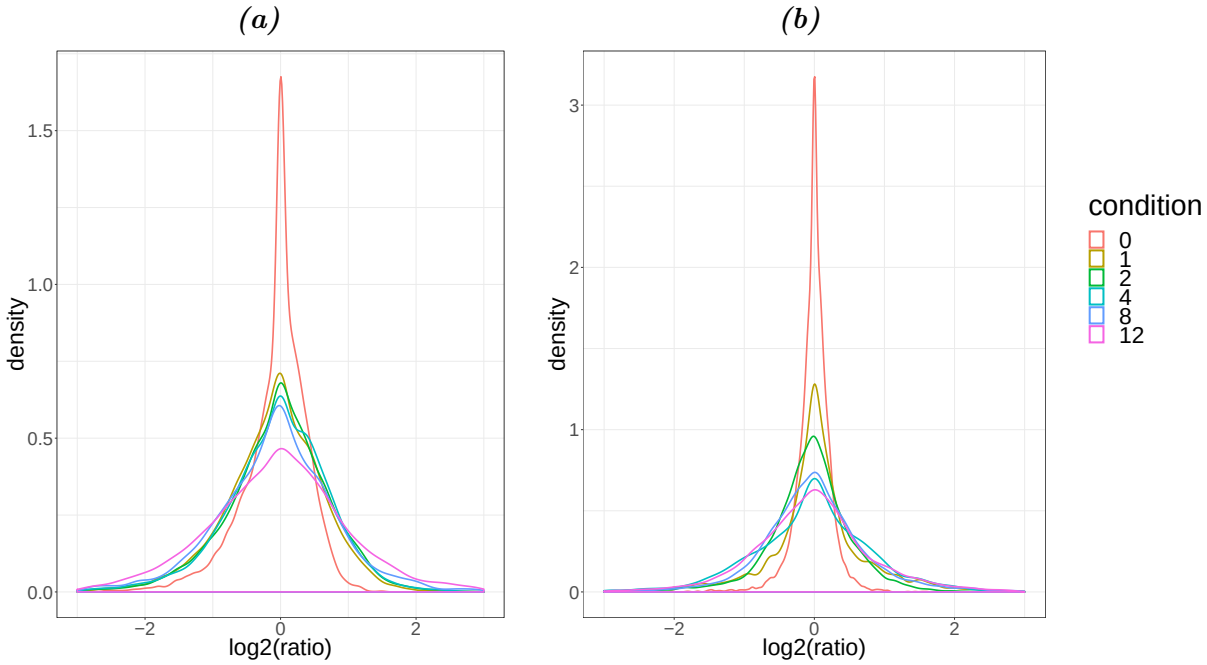


Figure 8.17: *Density plots for the the (a) $_{hf}$NES cell differentiation experiment and (b) the $_{hf}$NS cell differentiation experiment. The density plot plots become more spread out as the cells differentiate, indicating a greater difference in the cells as differentiation progresses.*

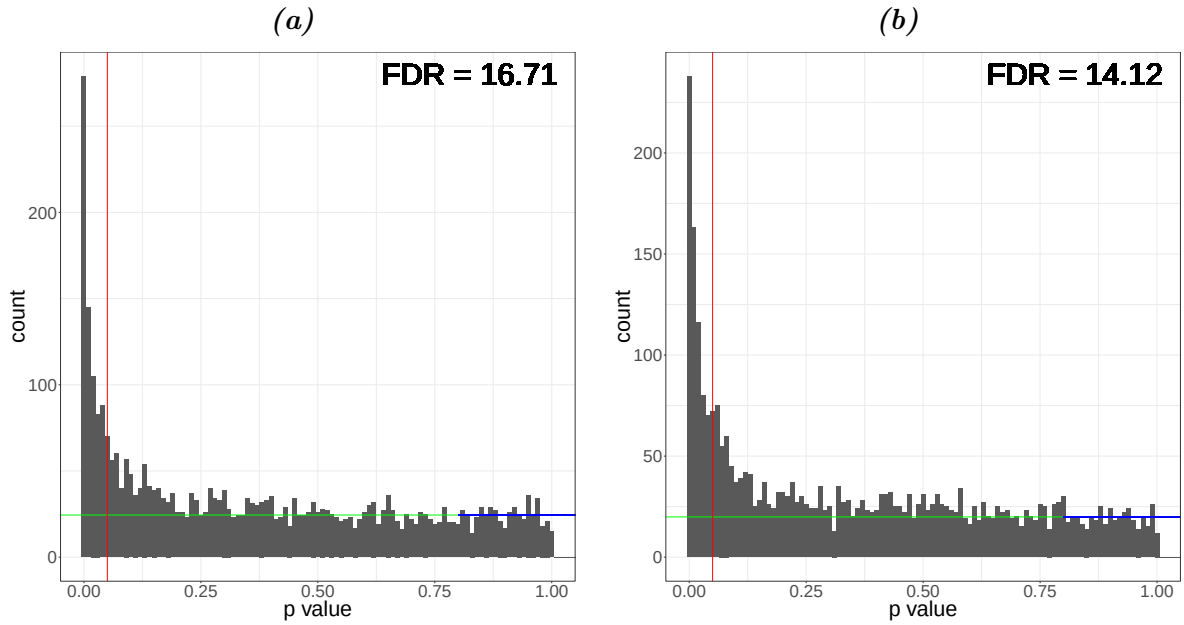


Figure 8.18: *FDR plots for the the (a) $_{hf}$NES cell differentiation experiment and (b) the $_{hf}$NS cell differentiation experiment*

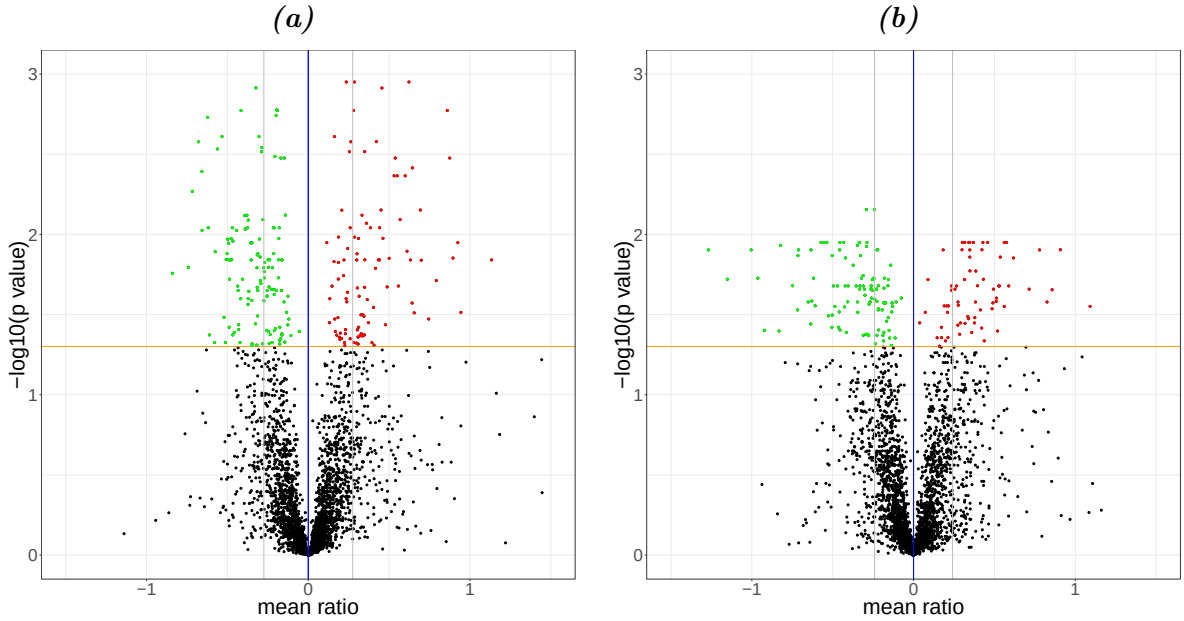


Figure 8.19: *Volcano plots for the (a) $_{hf}$NES cell differentiation experiment and (b) the $_{hf}$NS cell differentiation experiment*

Using the R base heatmap function, samples were hierarchically clustered. The $_{\text{hf}}$NS cells cluster separately from the $_{\text{hf}}$NES cells. Within these two clusters, the more differentiated samples cluster separately from the early proliferation and early differentiated samples. However,

this distinction is not so clear in the $_{hf}$NES cells.

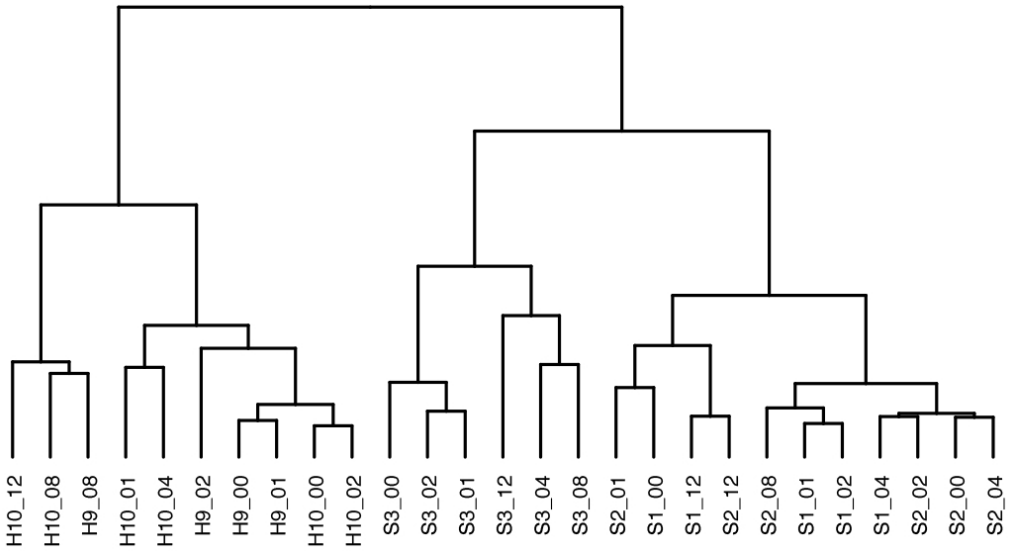


Figure 8.20: *Heirarchical clustering of the MaxLFQ intensity values for the NSC differentiation experiment using the heatmap function in R.*

8.3 Combining of Experiments

The data from the Huang $\frac{\text{dNES}}{\text{hESC}}$, GE $\frac{\text{NS}}{\text{NES}}$, and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and the $_{hf}$NES and $_{hf}$NS cell differentiation experiments were all analysed together in an attempt to generate a unifying description of the differences between the $_{hf}$NES cells and $_{hf}$NS cells. The Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment was used to determine the trajectory the $_{hf}$NES cells were on before differentiating into $_{hf}$NS cells. Most of the GO and KEGG enrichment analysis was based on the GE $\frac{\text{NS}}{\text{NES}}$ experiment due to the more complete nature of transcriptome analysis. The SILAC $\frac{\text{NS}}{\text{NES}}$ experiment served to extend and confirm the conclusions derived from the GE $\frac{\text{NS}}{\text{NES}}$ experiment. The differentiation data was used to further validate observed trajectories of the $_{hf}$NES cells and $_{hf}$NS cells.

Part IV

Analysis

Chapter 9

Markers of cellular identity

Cell markers are molecules, used as monograms to identify cells. These can be proteins, lipids, or cell surface carbohydrates. They are expressed either by specific cells, or in a specific cellular context. Antibodies are readily available from many suppliers, allowing identification of cells using immunohistochemistry. Abcam[3] provides a list of antibodies specifically for the identification of different types of neural cells (Figure 9.1). This list is accompanied by references for the markers.[5] Transcription factors and proteins produced through the activation of specific signalling pathways can also be used to identify cells. This discussion will focus on proteins, as they can be detected by proteomics.

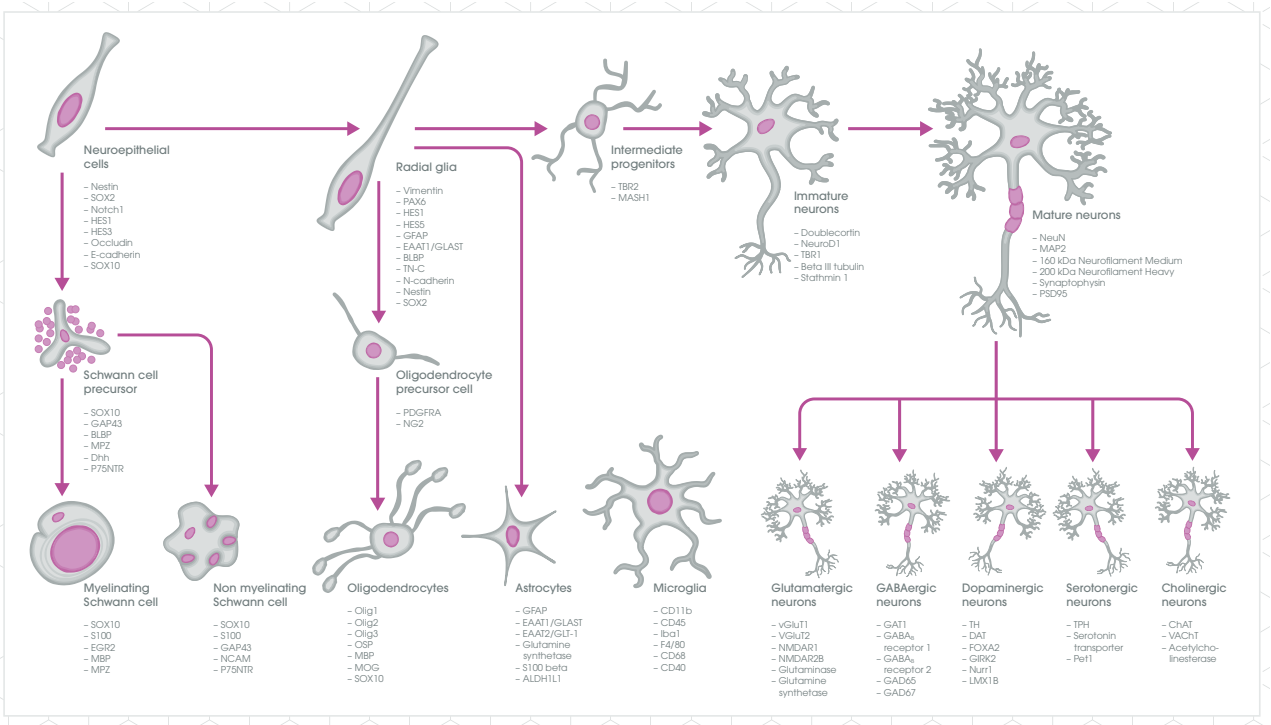


Figure 9.1: *Abcam[3] list of antibodies available for markers of nervous system cells.*

9.1 Pluripotency factors

Many pluripotent and proliferation markers are up-regulated in the $_{\text{hf}}$NES cells.

KL4, POU5F1, and NANOG The pluripotency markers *KLF4*, *POU5F1*, and *NANOG* are known to be highly expressed in ESCs. Together with *SOX2*, they are commonly used to transform adult somatic cells into iPSCs. *In vivo*, their downregulation would coincide with differentiation from ESCs into neuroepithelial cells. *NANOG* and *KLF4* were down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, although not statistically significantly, probably because of the large variance in the data (Figure 9.2).

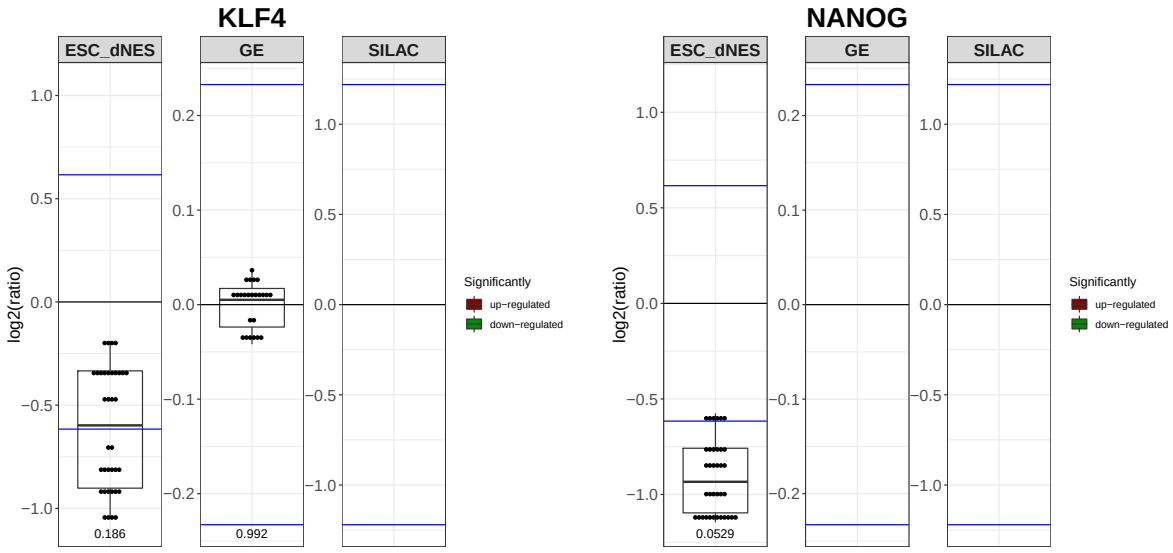


Figure 9.2: *Boxplots comparing the log2 ratios for KLF4 and NANOG. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Expression of *POU5F1* is known to be restricted to pluripotent and germ line cells, and it was down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.3). *SALL4*, an antagonist of *POU5F1*, was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, thereby possibly mitigating any effect *POU5F1* may have been inducing in the $_{\text{hf}}$NES cells. *POU5F1* is likely not necessary once differentiation progresses to the $_{\text{hf}}$NS cell stage. The SILAC 1 sample S2L_H10H supported the GE $\frac{\text{NS}}{\text{NES}}$ experiment evidence for the down-regulation of *SALL4* (Figure 9.3).

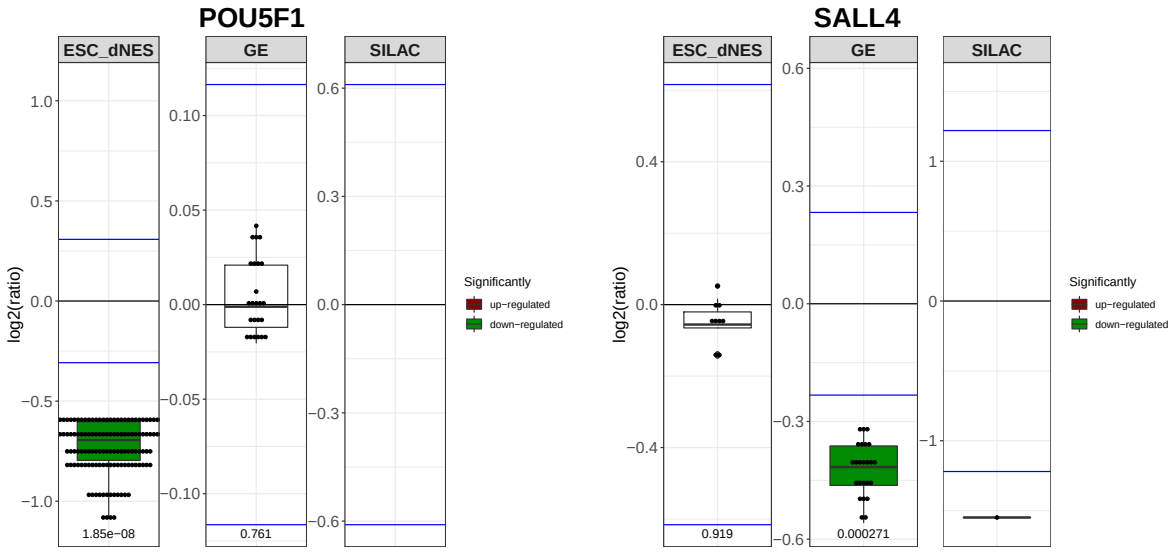


Figure 9.3: *Boxplots comparing the log2 ratios for POU5F1 and SALL4. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.1.1 SOX2

The sex determining region Y box (SOX) family of transcription factors play critical roles in development, influencing the formation of tissues and organs. They form heterodimers, thereby allowing them to have different functions in different cellular contexts. *SOX2* is responsible for maintaining pluripotency during neural stem cell self-renewal by regulating genes involved in stem cell maintenance and neural differentiation.[149,181] In vivo its expression, during gastrulation, is confined to the neural plate and its expression is extinguished in most adult tissues.[212] In ESC, it forms a heterodimer with *POU5F1*, enforcing ESC pluripotency. In neural stem cells, *SOX2* forms a heterodimer with *POU3F2*, thereby conferring neural identity.[129] *SOX2* expression was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.4). *POU5F1* was down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.3). *POU3F2* showed an upwards trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and a downwards trend in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, although neither was statistically significant (Figure 9.5). This fits well with what has been previously observed in literature. Up-regulation of *SOX2* and *POU3F2* in

dNES cells would confer neural identity to the $_{hf}$NES cells and then their down-regulation in $_{hf}$NS cells would allow further differentiation. Furthermore, *SOX2* appears to have decreased steadily in the $_{hf}$NES cell differentiation experiment. However because of a lack of sufficient numbers of data points, the change was not statistically significant. There are a number of additional SOX transcription factors that were SDE in this study, some of which will be discussed in section 14.2 (Figure 14.4)

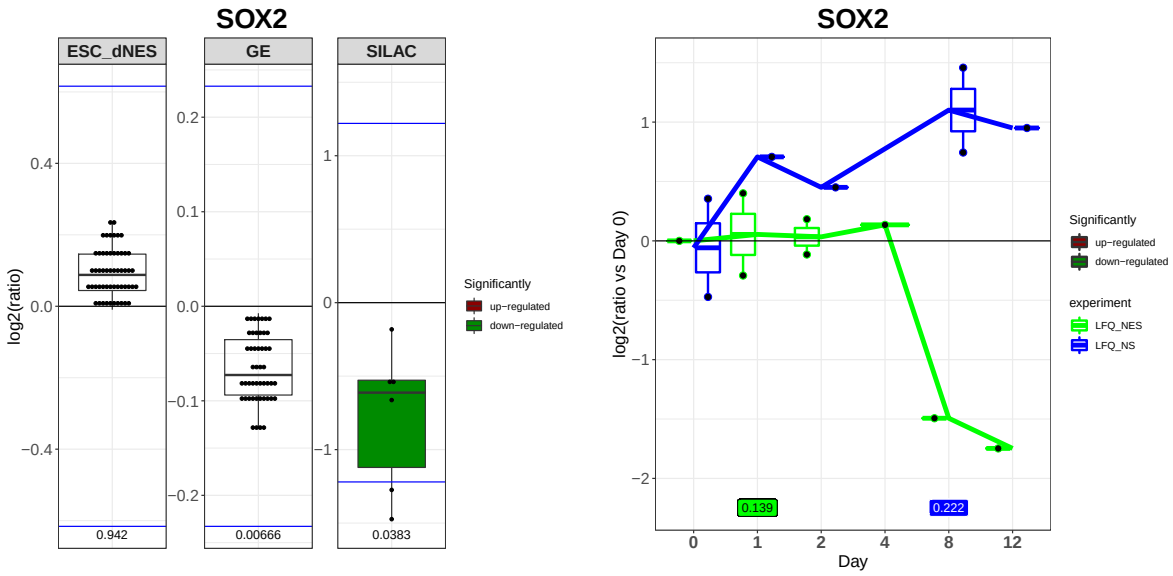


Figure 9.4: *Boxplots comparing the log2 ratios for SOX2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

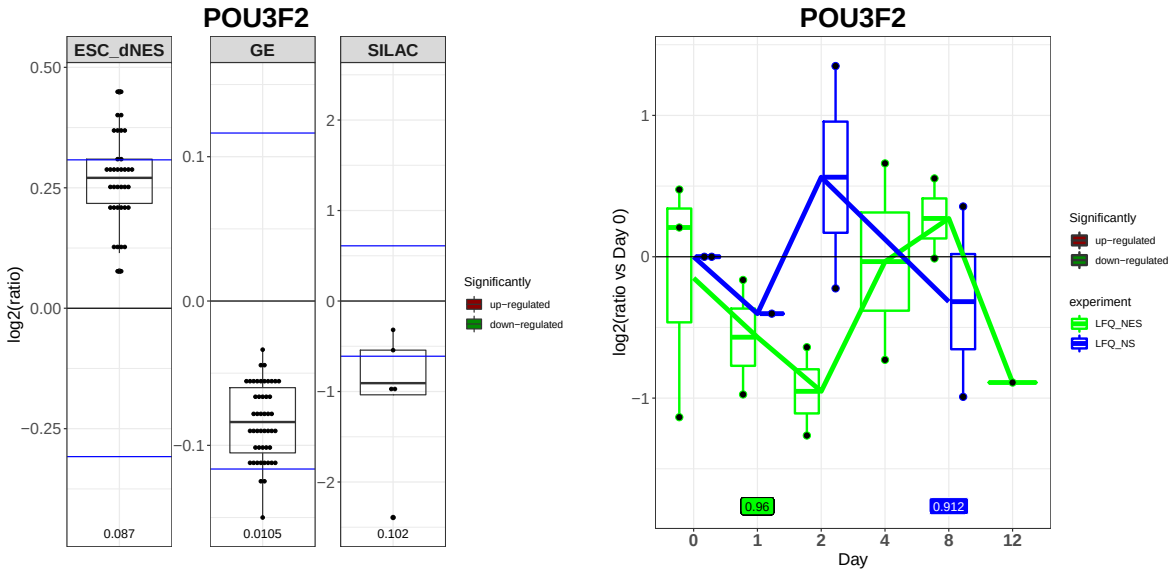


Figure 9.5: *Boxplots comparing the log2 ratios for POU3F2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.1.2 LIN28

LIN28 maintains stemness by binding and blocking the processing of *let-7* microRNA, which antagonises proliferation and self-renewal of stem cells.[238] *LIN28A* and *LIN28B* showed no difference in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment but were significantly down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. *LIN28A* was also significantly down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and $_{\text{hf}}$NES cells differentiation experiment. *LIN28B* showed similar but non-significant trends (Figure 9.6 and 9.7). *LIN28* is discussed further in section 13.2.3.

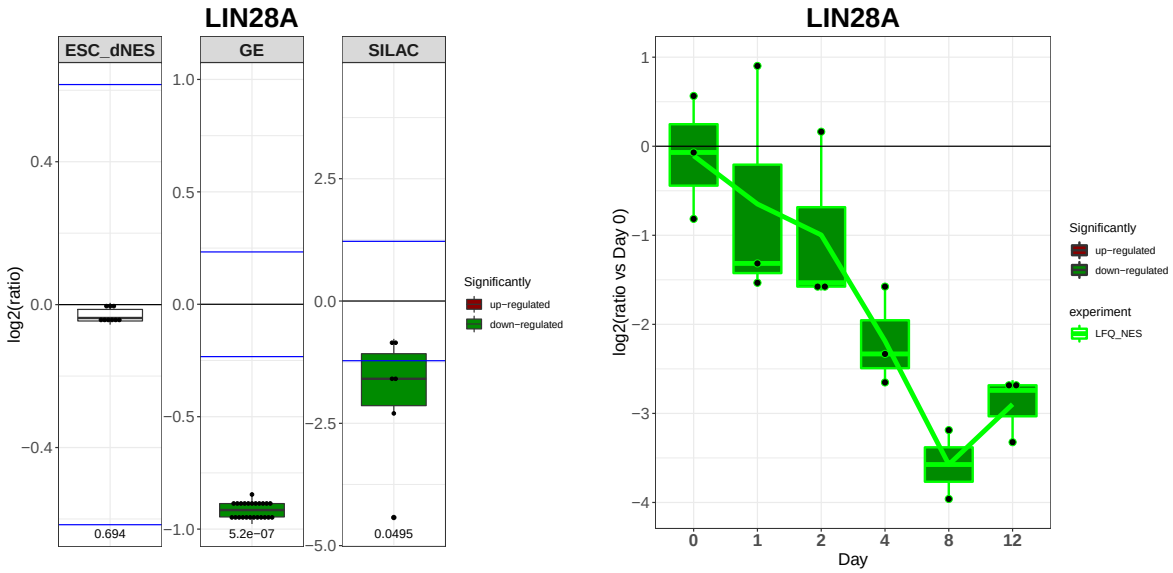


Figure 9.6: *Boxplots comparing the log2 ratios for LIN28A. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

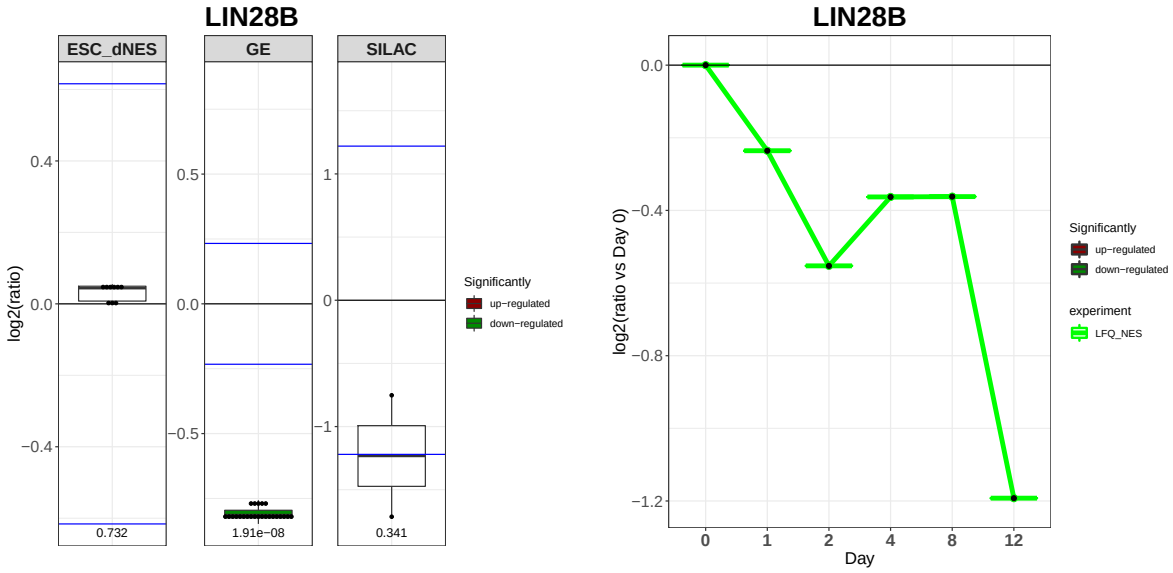


Figure 9.7: *Boxplots comparing the log2 ratios for LIN28B. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.1.3 ZFP42

ZFP42, a transcription factor marker of ESCs, is known to be down-regulated when ESCs are treated with retinoic acid to induce differentiation,[248] although its molecular function remains uncertain. It was down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, but not detected in the other experiments (Figure 9.8).

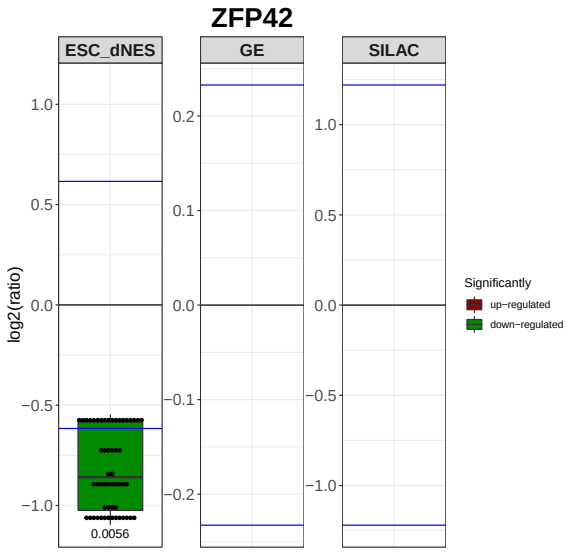


Figure 9.8: *Boxplots comparing the log2 ratios for ZFP42. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.2 Proliferation markers

9.2.1 MYC

The *MYC* proteins induce cell proliferation and cell growth. *MYC* and *MYCN* have complementary expression patterns, with *MYCN* expressed in early development particularly, in the neuroepithelium. *MYC* is expressed in all adult proliferative cells.[232] Differing expression patterns indicate the possibility of functional differences despite similar biochemical properties.[232] In *MYC*-deficient mice, *MYCN* can replace *MYC* with few developmental deficiencies.[232] The endogenous expression of either *MYC* or *MYCN* is essential for maintaining pluripotency.

MYCN was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.9). However, no statistically significant change was observed in *MYC* expression.

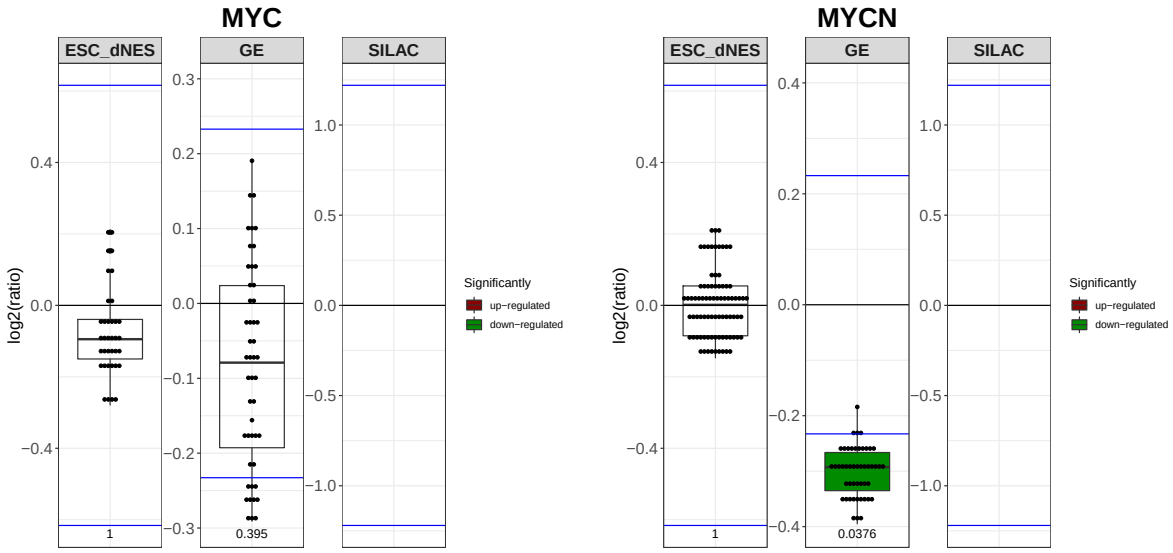


Figure 9.9: *Boxplots comparing the log2 ratios for MYC and MYCN. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.2.2 MKI67 and PCNA

The proliferation markers *MKI67* and *PCNA* were both down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment but not in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. They were also both down-regulated in the $_{\text{hf}}$NES and $_{\text{hf}}$NS cell differentiation experiments (Figure 9.10 and 9.11). *MKI67* covers chromosomes and disperses them by forming a steric and electrostatic charge barrier, thus preventing the chromosomes from collapsing and thereby enabling them to disperse. *PCNA* is essential for DNA replication. It serves as a DNA clamp and localises to pre-replication sites[227] where it encircles the DNA, forming a sliding clamp.[165] It then recruits proteins involved in DNA replication and facilitates their progress along the DNA.

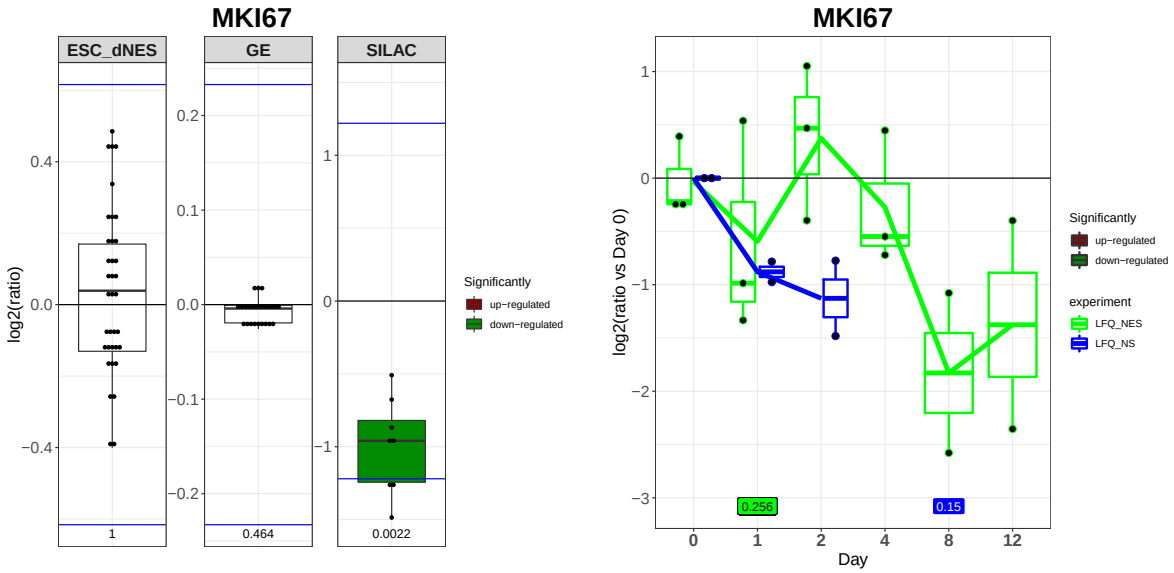


Figure 9.10: *Boxplots comparing the log2 ratios for MKI67. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

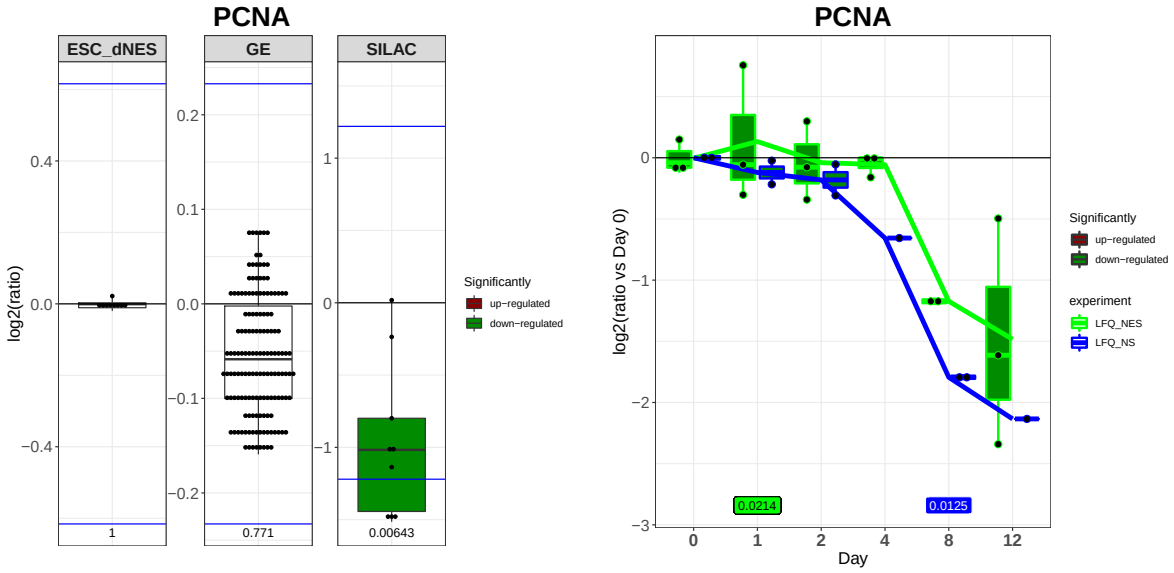


Figure 9.11: *Boxplots comparing the log2 ratios for PCNA. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.2.3 MCM2

MCM2 is a proliferation marker often used as a marker of neural stem cell proliferation. It forms part of the minichromosome maintenance protein complex (MCM complex), which is critical for replisome assembly. It is significantly down-regulated in both the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. It is also significantly down-regulated in the $_{\text{hf}}$NS cell differentiation experiment (Figure 9.12). The MCM complex is discussed in more detail in section 12.1.1.

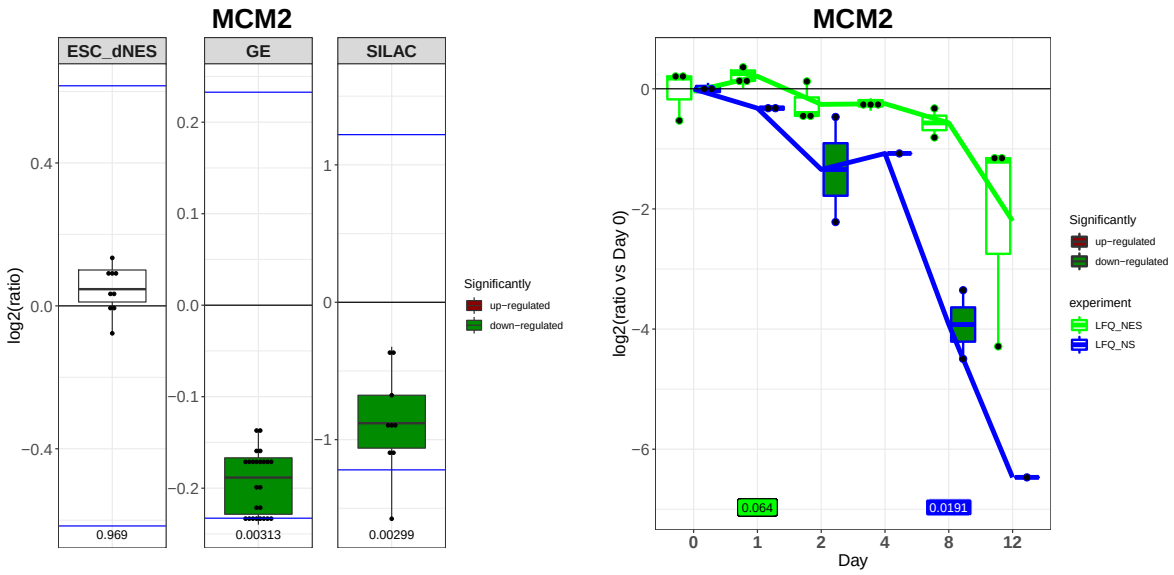


Figure 9.12: *Boxplots comparing the log2 ratios for MCM2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.2.4 HMGA2

The HMG protein *HMGA2* binds DNA and changes its conformation, and then induces transcription through interaction with other transcription factors.[248] It does not have its own transcriptional activation capacity. It increases the cell cycle progression through interaction with *CCNA2*, thereby promoting stem cell self renewal. *HMGA2* was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. *CCNA2* shows a non significant downward trend in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, indicating some degree of co-expression in the $_{\text{hf}}$NES cells (Figure 9.13). There are additional HMG proteins, SDE in this work (Figure 9.14).

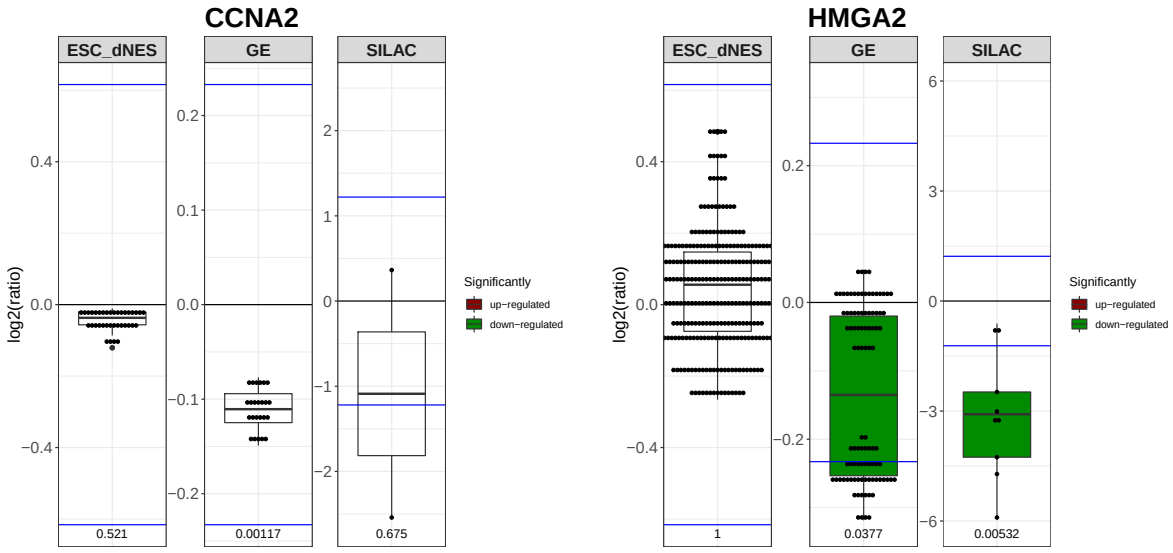


Figure 9.13: *Boxplots comparing the log2 ratios for CCNA2 and HMGA2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

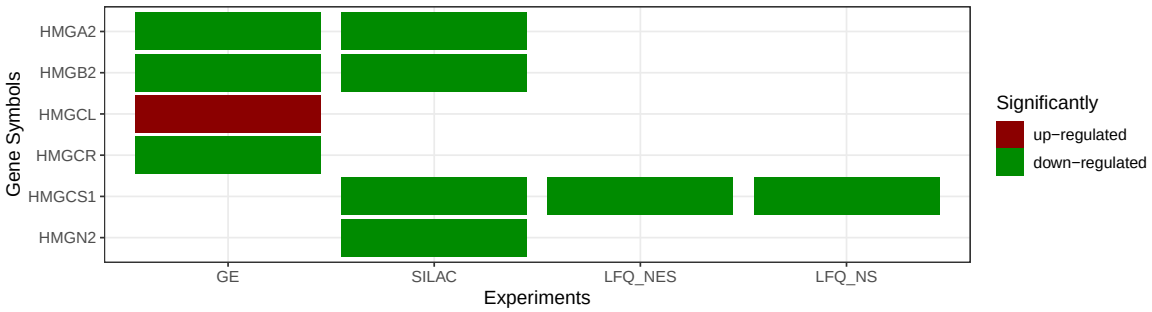


Figure 9.14: *Heatmap of HMG proteins significantly differentially expressed (SDE). Red and green indicates up- and down-regulation respectively.*

9.2.5 PDGFRB

PDGFRB is a tyrosine kinase cell surface receptor, that binds *PDGFD* and *PDGFB* homodimers or *PDGFA* and *PDGFB* heterodimers. None of these ligands were SDE (Figure 9.16). Activation of the receptor results in auto-phosphorylation and phosphorylation of intracellular proteins. Activation usually triggers proliferation and migration, as has been shown in neuroepithelial cells *in vitro*.[72] *In vivo*, its expression in up-regulated in response to traumatic brain injury (TBI), triggering cell proliferation and formation of a glial scar. Its inhibition

disrupts the formation of this glial scar.[161] It was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.15) and may be promoting proliferation of the $_{\text{hf}}$NES cells.

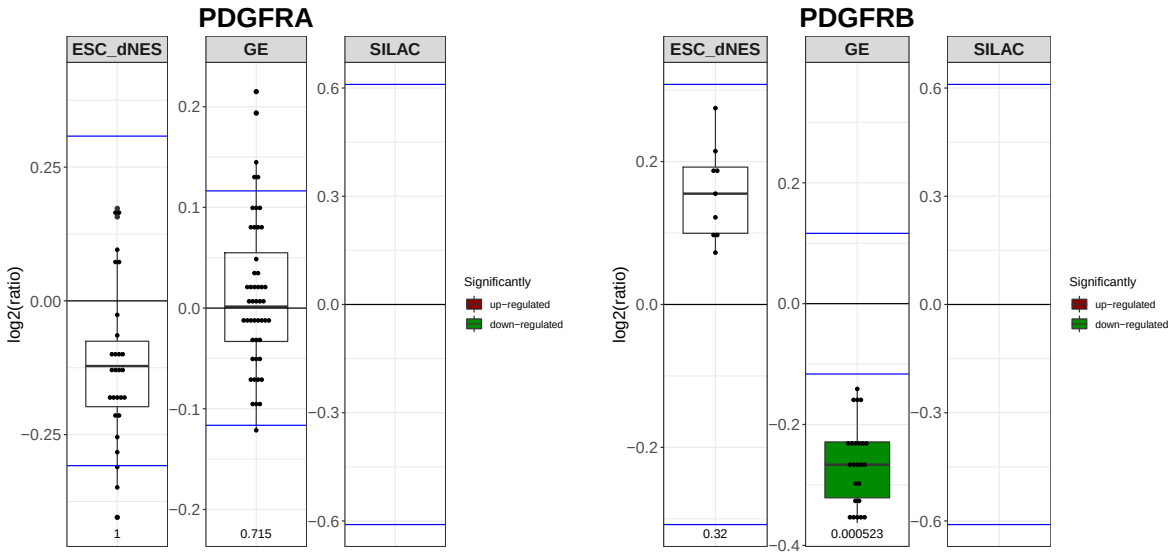


Figure 9.15: *Boxplots comparing the log2 ratios for PDGFRA and PDGFRB. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

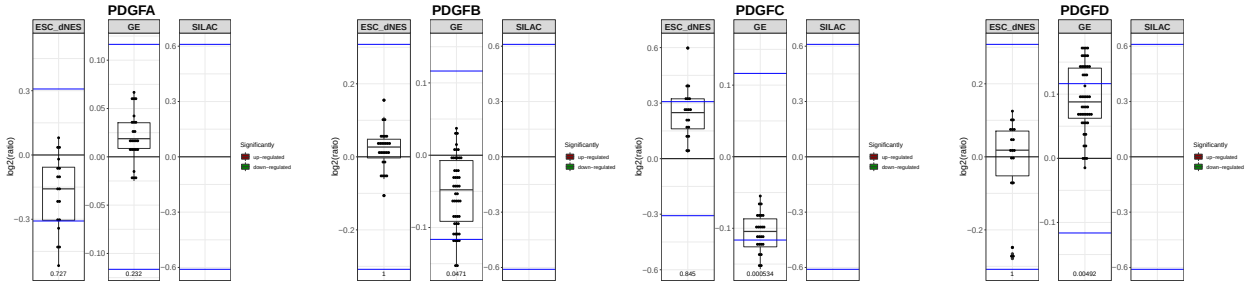


Figure 9.16: *Boxplots comparing the log2 ratios for PDGFA, PDGFB, PDGFC and PDGFD. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.2.6 TERF1 and ALPL

Enzymes can also be used as markers: *TERF1*, a common ESC marker, was down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.17). It negatively regulates telomerase and protects chromosome ends. *ALPL*, an alkaline phosphatase isozyme, is widely used as an ESC pluripotency

marker, and was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, but showed too much variance in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment to deduce any trend (Figure 9.17).

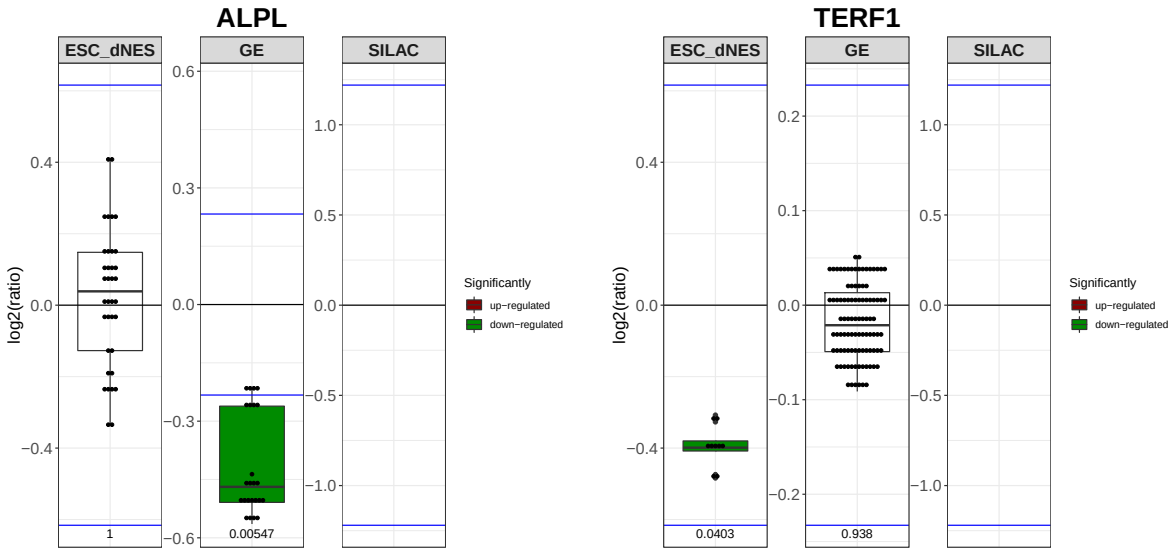


Figure 9.17: *Boxplots comparing the log2 ratios for ALPL and TERF1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3 Neural stem cell markers

9.3.1 Paired box transcripion factors

Zhang and Jiao (2015) published a review on markers of embryonic and adult neurogenesis. Key amongst these is the paired box transcription factor *PAX6*, which plays a pivotal role initiating the neural cell fate. *PAX6* cooperates with *SOX2* in gene regulation during neurogenesis. It activates neuronal genes while repressing those that promote non-neuronal lineages.[220] It was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.18) and may be partially responsible for conferring neural tissue properties upon dNES cells.

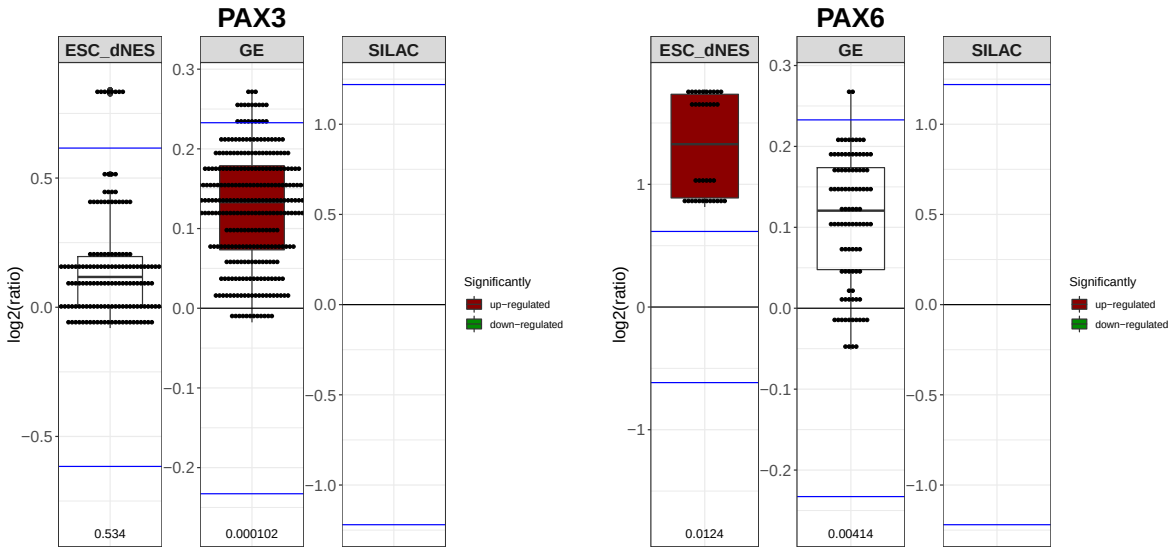


Figure 9.18: *Boxplots comparing the log2 ratios for PAX3 and PAX6. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

PAX3 promotes neural stem cell survival[229] and stemness[127] and is down-regulated during neural stem cell differentiation into neurons[42] or astrocytes.[127] *PAX3* is known to be involved in neural tube development, its hypermethylation may play a role in neural tube defects.[124] *PAX3* was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.18) and may be helping to maintain the stemness of the $_{\text{hf}}$NS cells. None of the other paired box (PAX) genes were SDE (Figure 9.19).

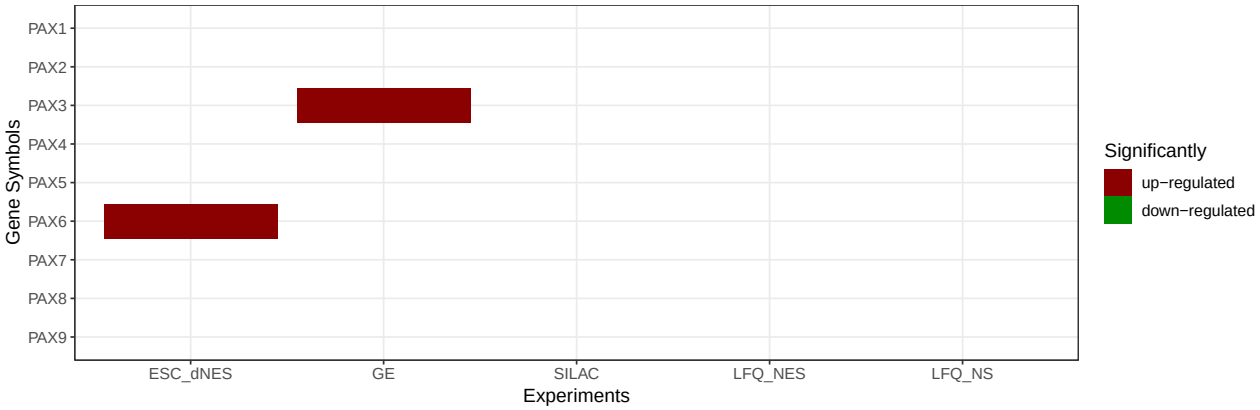


Figure 9.19: *Heatmap for the expression of all PAX genes or proteins across all experiments. Red and green indicates up- and down-regulation respectively.*

9.3.2 Nuclear hormone receptors

The nuclear hormone receptors *NR2F1* and *NR2F2* are key regulators in neural stem cell temporal specification. They were both up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. *NR2F1* was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figures 9.20 and 9.21). They are discussed in further in the transcription factor section 14.4.2.

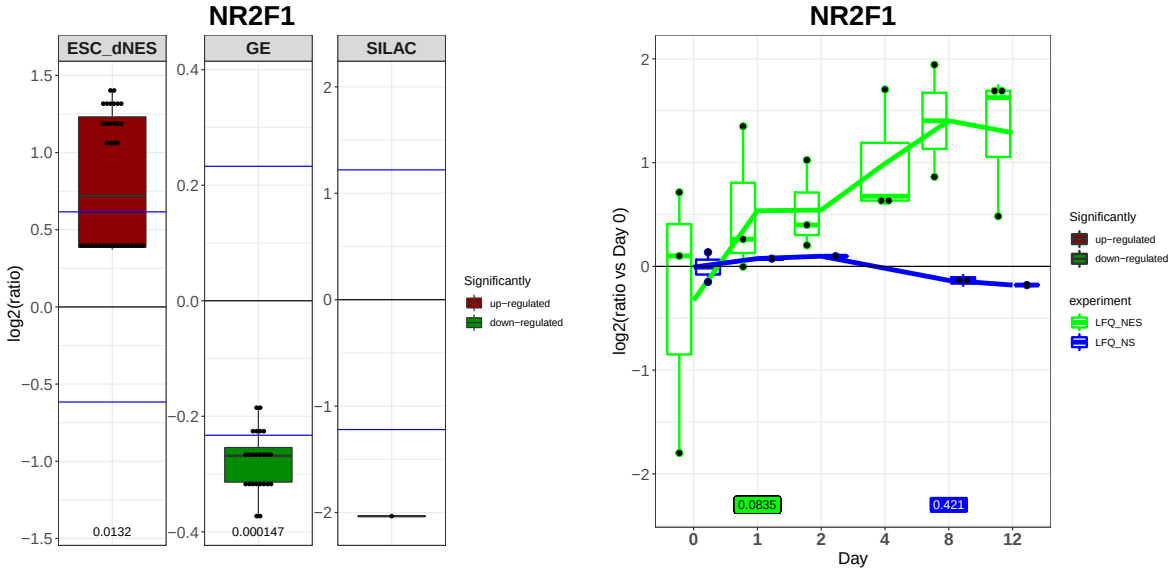


Figure 9.20: *Boxplots comparing the log2 ratios for NR2F1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

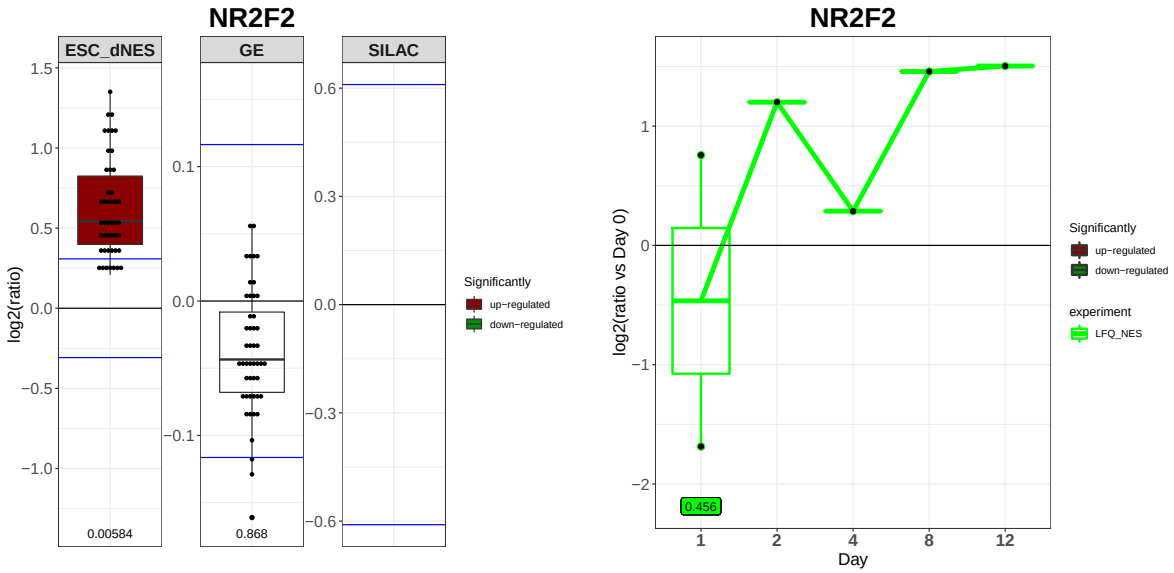


Figure 9.21: *Boxplots comparing the log2 ratios for NR2F2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3.3 EGR2

EGR2 is a transcription regulatory factor important in hind-brain development. Deletion of this gene in mice results in hind-brain defects and an inability to survive after birth.[212] It was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.22), indicating that it is possibly only crucial during the neuroepithelial phase of hind-brain development.

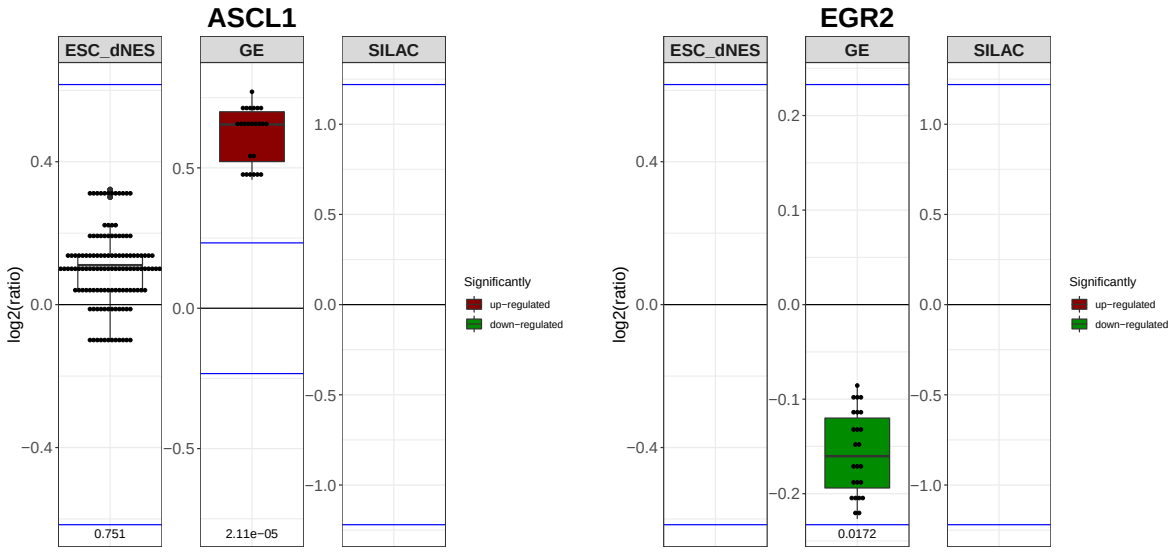


Figure 9.22: *Boxplots comparing the log2 ratios for ASCL1 and EGR2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3.4 ASCL1

ASCL1, a pro-neural basic helix-loop-helix (bHLH) specific transcription factor, is often used as a RG marker, together with *NEUROD2* and *NEUROG2*. It has been shown to be essential for maintaining neural stem cell neurogenesis potential, in late developmental stages, beyond Carnegie Stage 18.[148,223] When it is knocked down in murine neural stem cells that differentiate into neurons, they become glial cells instead.[164] It is induced by BMP-SMAD signalling and is usually expressed after *PAX6*, with *NEUROG2* but before *NEUROD2* and *MYT1L*. *ASCL1* was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment but not detected in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.22). *NEUROG2*, *NEUROD2*, and *MYT1L* were not SDE in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment or detected in the other experiments.

9.3.5 Hairy and enhancer of split transcription factors

The hairy and enhancer of split (HES) family of bHLH transcription factors suppresses transcription, particularly that of pro-neural genes. *HES1*, a common neuroepithelial marker,

showed an non-significant upwards trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.23). *HES3* was not present in this dataset. *HES4* and *HES5* showed non-significant upwards trends in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and were up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.23 and 9.24). They may be responsible for conferring increased gliogenic potential onto the $_{\text{hf}}$NS cells.

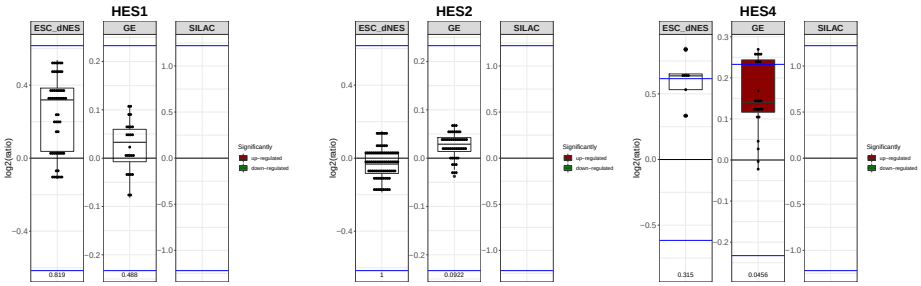


Figure 9.23: *Boxplots comparing the log2 ratios for HES1, HES2, HES3 and HES4. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

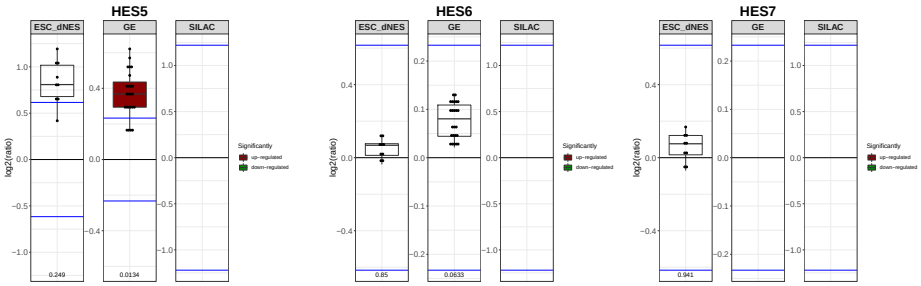


Figure 9.24: *Boxplots comparing the log2 ratios for HES5, HES6 and HES7. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3.6 Intermediate filaments

NES an intermediate filament used as a marker for neural stem cells, it is expressed in neural stem cells and its expression vanishes upon differentiation, so it is absent from nearly all mature CNS cells. In this study, there appeared to be no difference in expression of *NES* in the proliferating cells. Its expression then decreased during differentiation, significantly so in the $_{\text{hf}}$NES cells (Figure 9.25).

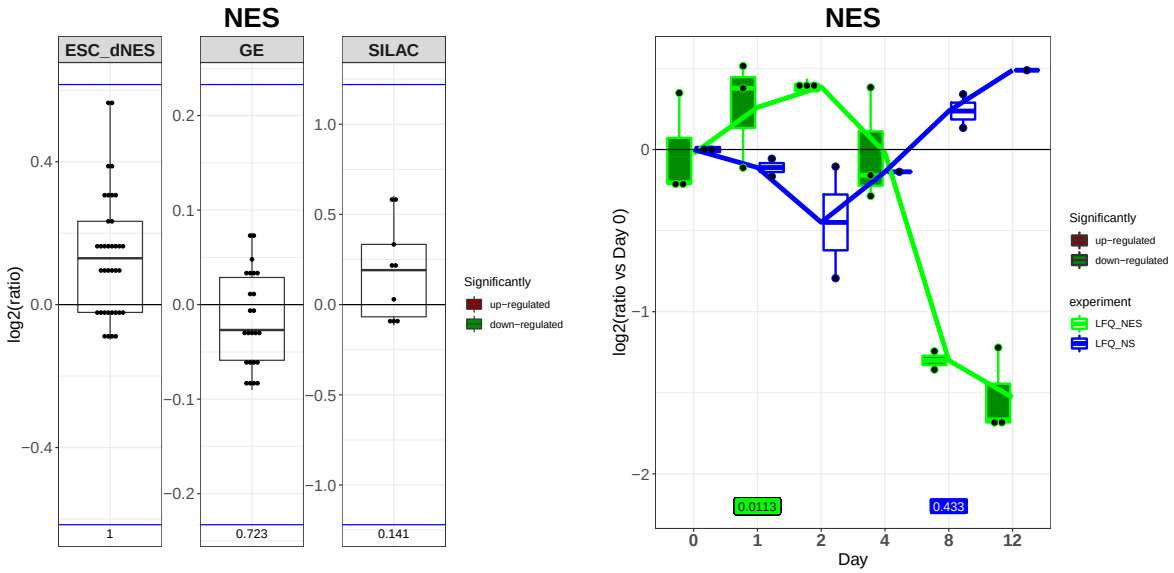


Figure 9.25: *Boxplots comparing the log2 ratios for NES. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

The $_{hf}$NS cells express many RG markers, such as the intermediate filaments *VIM* and *GFAP*. *VIM* is used as a RG marker in the embryo, and vanishes at the termination of gestation, when it is replaced by *GFAP*, which provides mechanical strength to astrocytes. *GFAP*-positive progenitors have been shown to produce all three types of neural cells. *GFAP* and *VIM* were up-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and then *GFAP* was further up-regulated during $_{hf}$NS cell differentiation (Figure 9.26 and 9.27). Multiple isoforms were detected for *GFAP*, P14136 was most abundantly detected and showed the greatest change in expression.

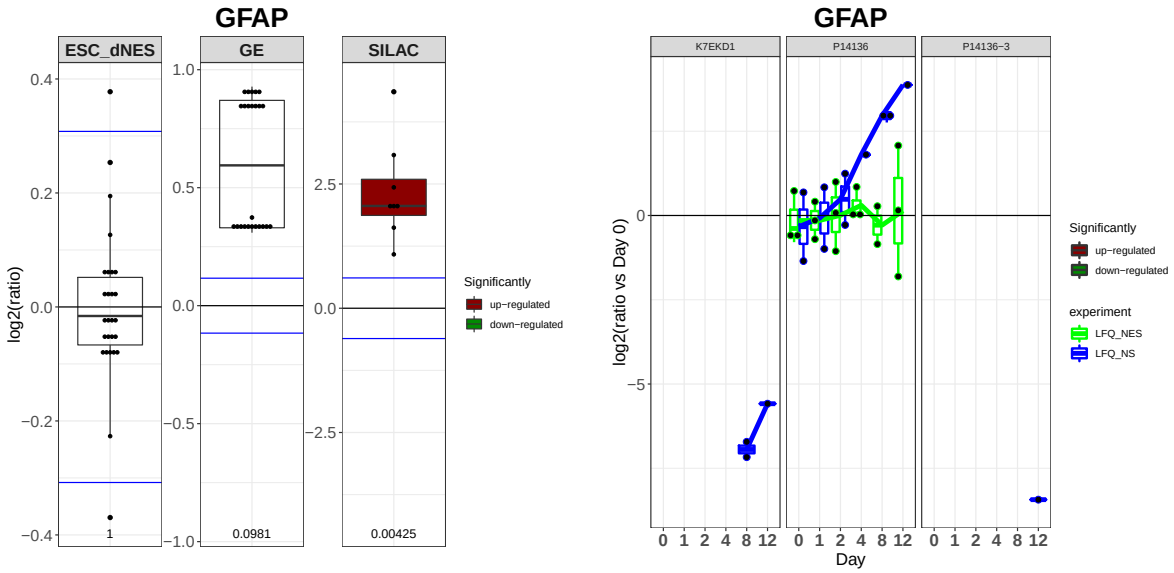


Figure 9.26: *Boxplots comparing the log2 ratios for GFAP. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

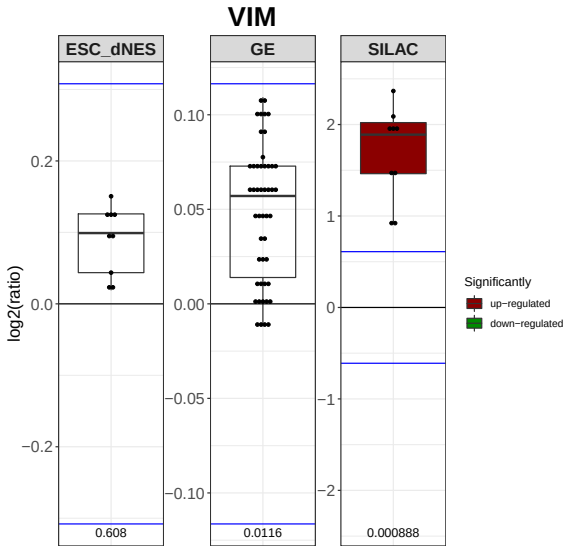


Figure 9.27: *Boxplots comparing the log2 ratios for VIM. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3.7 FABP7

FABP7, another common RG marker, is required for the glial fibre system. This is a network of astrocytes across which neurons migrate. By similarity, *FABP7* has been identified as a fatty acid transporter, but its ligand is as yet unknown. *In vivo* its expression is regulated by *PAX6*. It is not expressed in the neuroepithelium and its expression coincides with RG differentiation.[17] Its expression is maintained in astrocytes. It was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. (Figure 9.28).

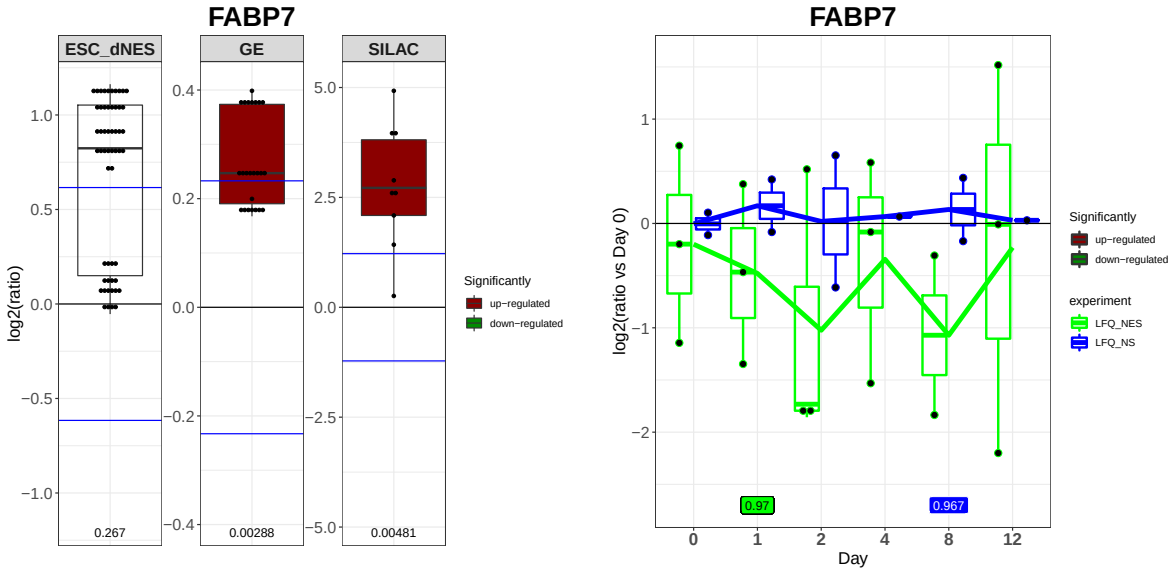


Figure 9.28: *Boxplots comparing the log2 ratios for FABP7. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.3.8 Glutamate transporters

The glutamate transporters *SLC1A1*, *SLC1A2*, and *SLC1A3* have specific expression in different cells and areas of the CNS.[178] *SLC1A3*, commonly known as GLAST-1, was the only glutamate transporter SDE in this study. It is an astrocyte-specific transporter of glutamate and aspartate. It removes glutamate from the synaptic cleft, terminating post-synaptic action. It is highly expressed in Bergmann glia but is also found in neurons.[178] Together with *GFAP*, *FABP7* is used as a marker to determine the transition from neuroepithelial cells to RG.[5]

Here, it was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 9.29). There were many solute carriers SDE in this study (Figure 9.30), and their differential expression across these cells may make them useful as potentially novel markers.

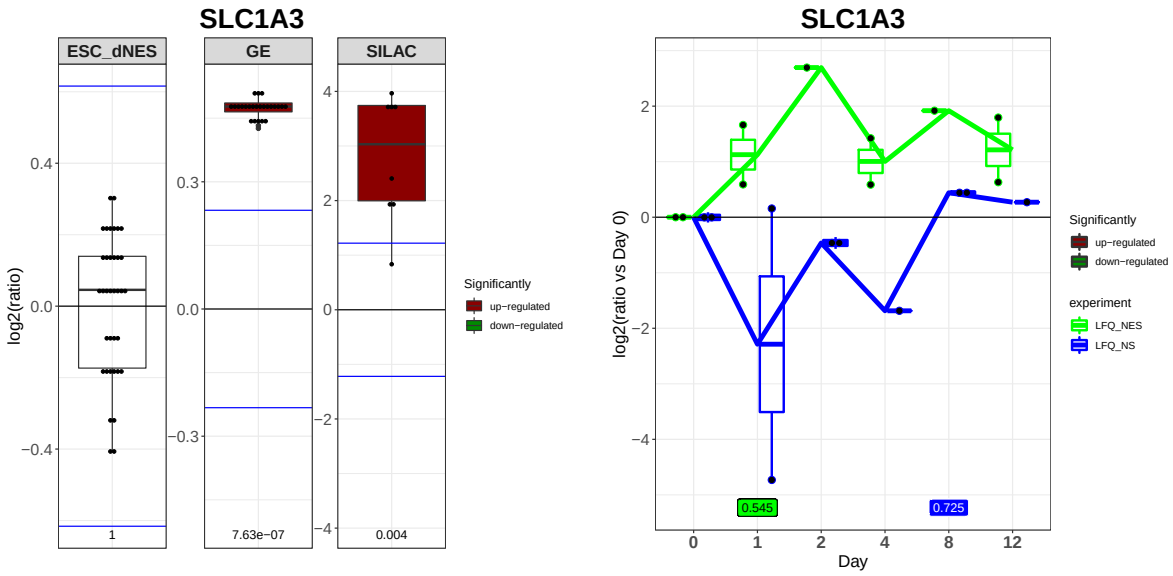


Figure 9.29: *Boxplots comparing the log2 ratios for SLC1A3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

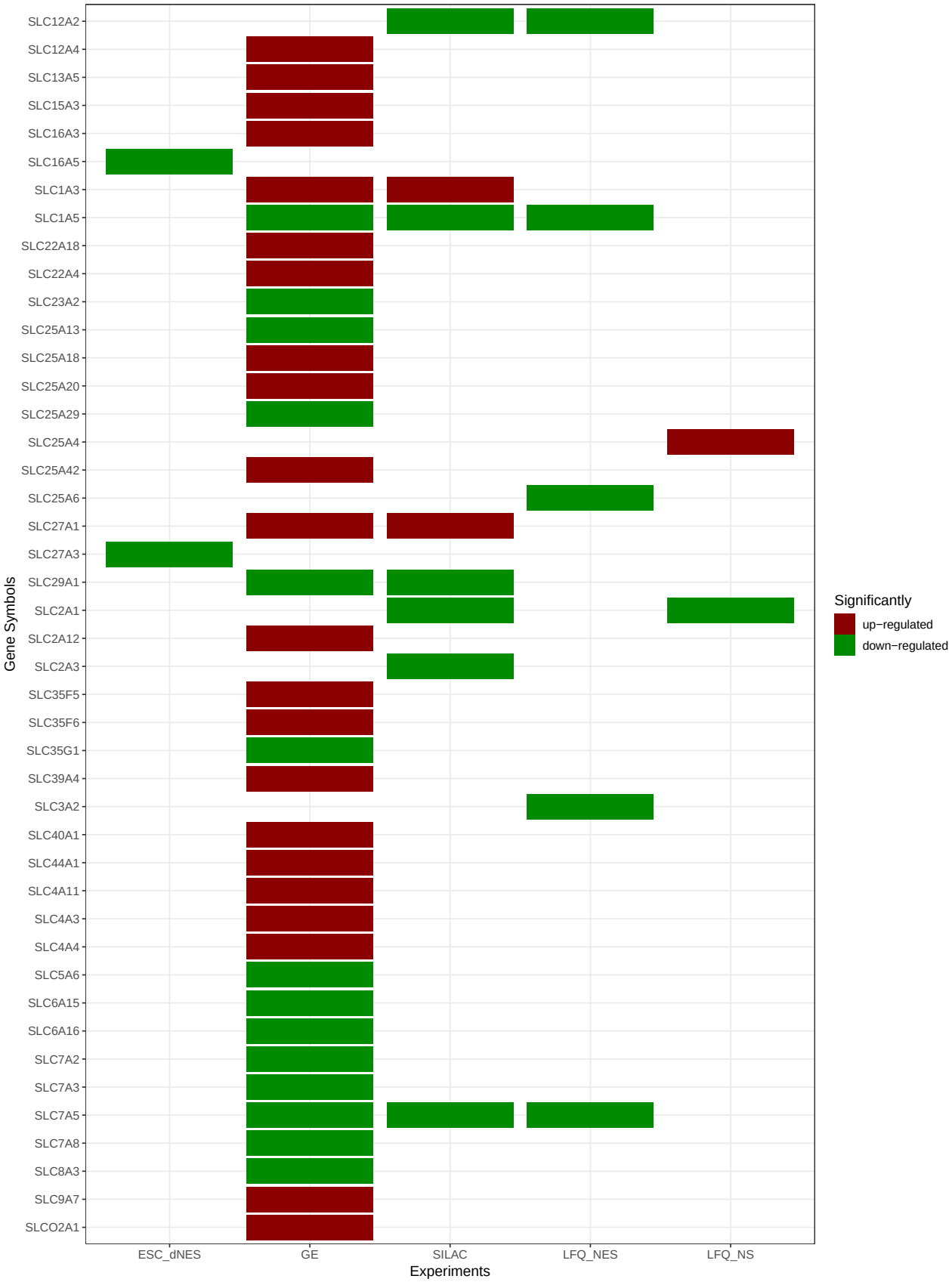


Figure 9.30: *Heatmap of all solute carriers significantly differentially expressed (SDE) across all experiments. Red and green indicates up- and down-regulation respectively.*

9.3.9 MBP

The myelin basic protein (*MBP*) forms and stabilises the myelin membrane structure. It was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment and its expression in $_{\text{hf}}$NS cells indicates the potential for these cells to differentiate into oligodendrocytes.

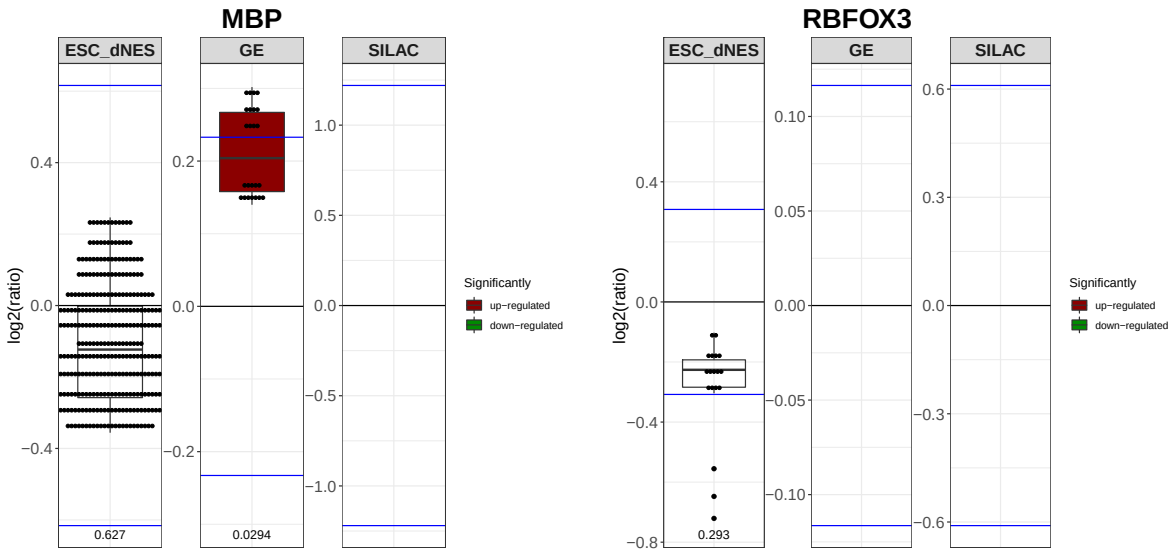


Figure 9.31: *Boxplots comparing the log2 ratios for MBP and RBFOX3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4 Neuronal markers

9.4.1 DCX

DCX is a microtubule-associated protein that facilitates microtubule polymerisation. It is known to become down-regulated as the neurons mature and is replaced by *RBFOX3*. *DCX* was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. The increased expression of *DCX* in both differentiating $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells indicates that the study terminated while the neurons were still immature (Figure 9.32). *RBFOX3* showed a downwards trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.31).

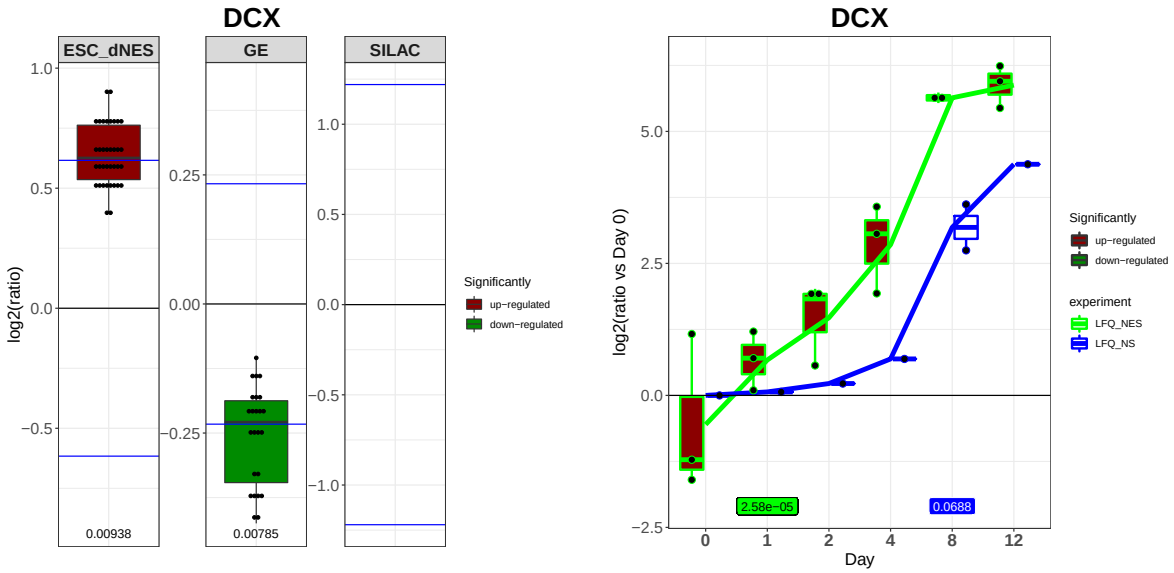


Figure 9.32: *Boxplots comparing the log2 ratios for DCX. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4.2 MAP2

MAP2 is a cytoskeletal protein involved in stabilising microtubules by crosslinking them to other microtubules or to intermediate filaments. It is often used as marker for mature neurons. Interestingly, it was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. The two isoforms detected in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment also show downwards trends. *MAP2* was up-regulated in both the $_{\text{hf}}$NES cell differentiation experiment and the $_{\text{hf}}$NS cell differentiation experiments. Two isoforms were detected, the P11137-3 isoform was the most abundant and showed the greatest increase in expression during differentiation (Figure 9.33).

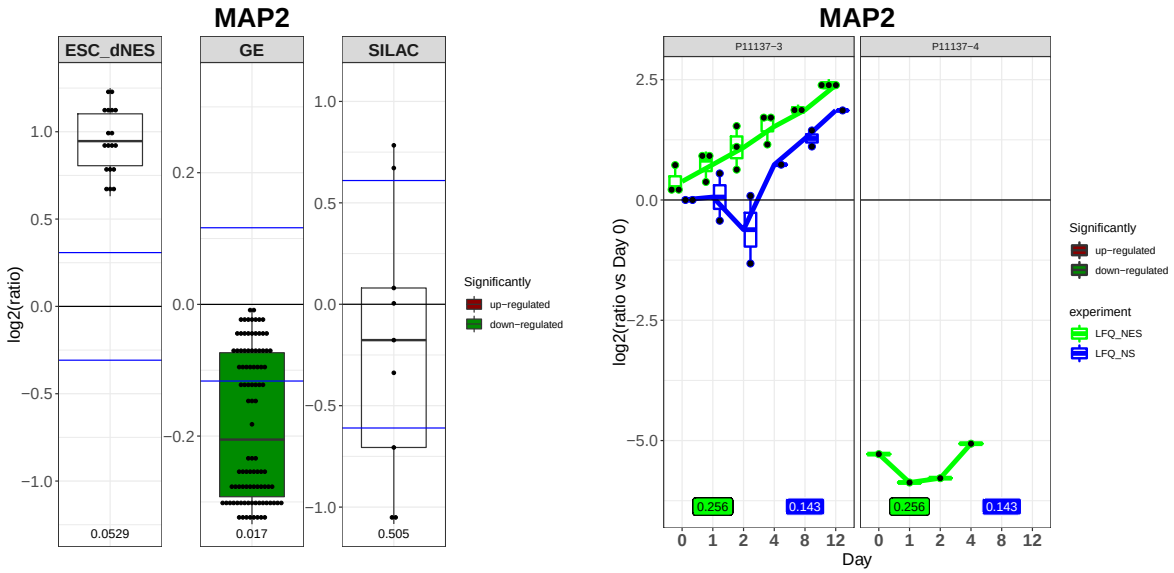


Figure 9.33: *Boxplots comparing the log2 ratios for MAP2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4.3 CALB1

CALB1, a calcium buffer and sensor that is important in regulating pre-synaptic calcium, is down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 9.34). This might indicate that it is playing additional roles in calcium signalling in pluripotent cells.

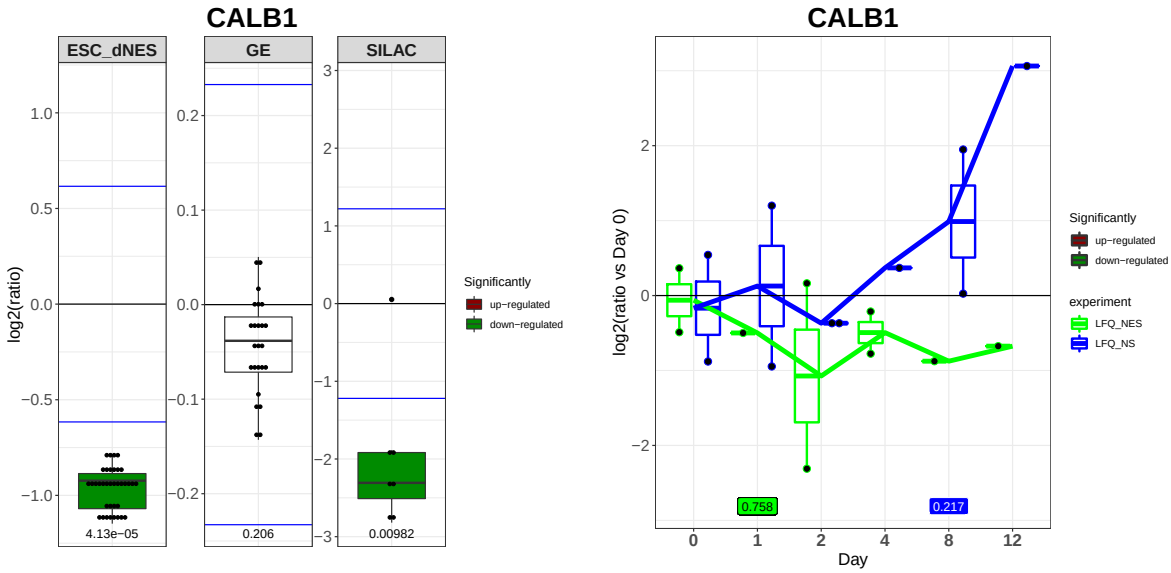


Figure 9.34: *Boxplots comparing the log2 ratios for CALB1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4.4 Tubulins

Tubulins are major constituents of microtubules, their expression is accurately controlled in different neural tissue cells. *TUBB3* is thought to be a neuron-specific marker and is important for proper axon maintenance and guidance. It is more dynamic than other tubulins which is a necessity for the rapid changes that occur during neurite outgrowth. Here it's expression was detected by immunohistochemistry in $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells, and it was up-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and $_{\text{hf}}$NES cells differentiation experiment (Figure 9.35). It may become up-regulated in $_{\text{hf}}$NS cells to produce their processes. Its expression did not appear to change much as the $_{\text{hf}}$NS cells differentiated, but it was up-regulated as the $_{\text{hf}}$NES cells differentiated. There are a number of additional tubulins SDE in this work (Figure 9.36). They may each play specific roles in each of the different cellular contexts.

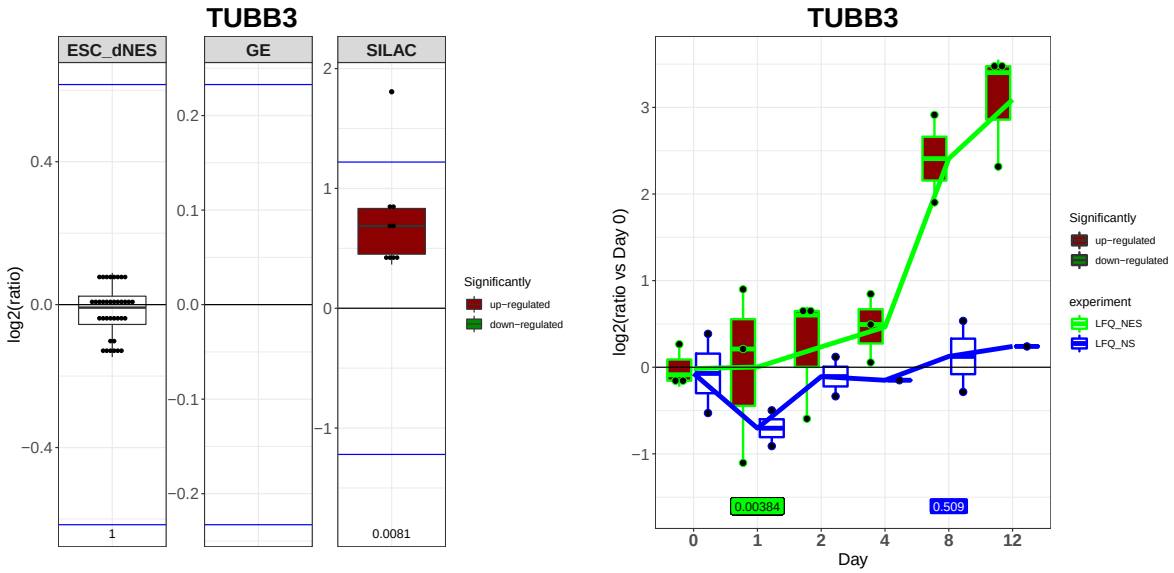


Figure 9.35: *Boxplots comparing the log2 ratios for TUBB3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

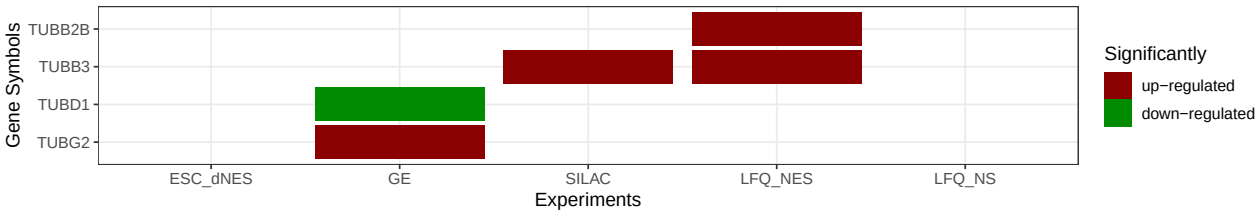


Figure 9.36: *Tubulins significantly differentially expressed (SDE).*

9.4.5 Neurofilaments

Neurofilaments are intermediate filaments that provide structural support for neuronal processes, in particular maintaining the diameter of the axon. *NEFL* and *NEFM* are replaced by *NEFH* as neurons mature. *NEFL* was up-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.37). *NEFM* was down-regulated during $_{\text{hf}}$NS cell differentiation (Figure 9.38). Strangely, *NEFH* was down-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.39), indicating that it was more highly expressed in human ESC lines (hESC cells) than dNES cells, although its

expression should only increase upon neuronal maturation.

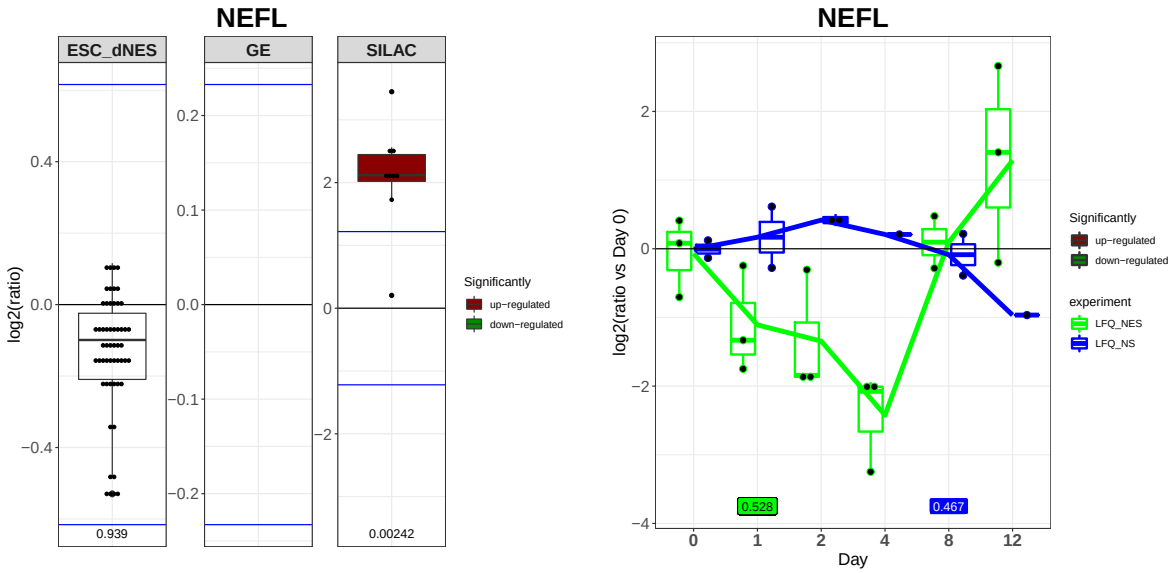


Figure 9.37: *Boxplots comparing the log2 ratios for NEFL. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

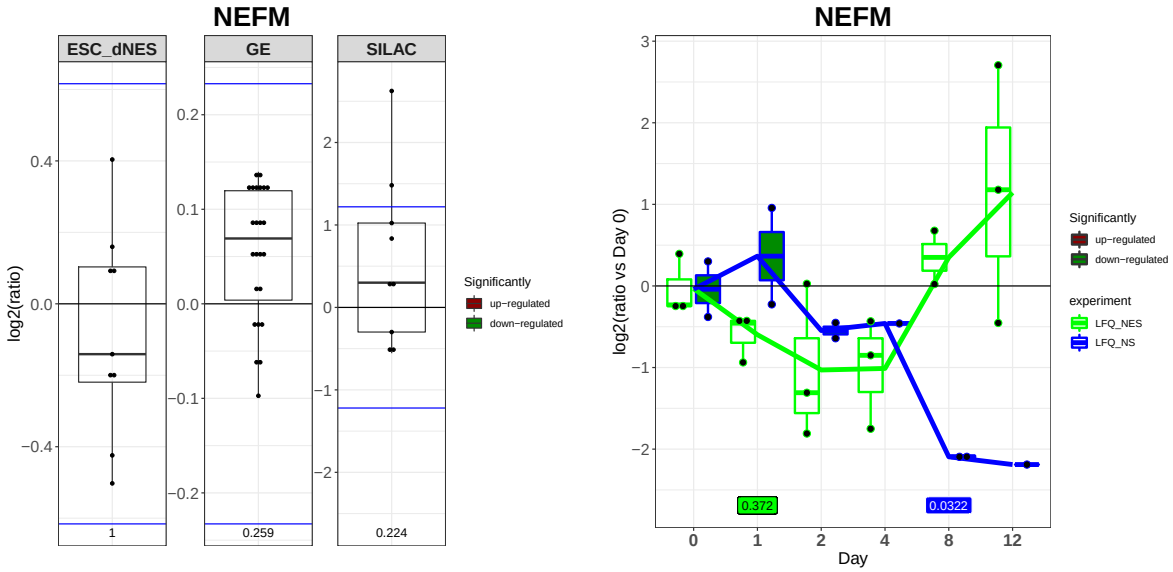


Figure 9.38: *Boxplots comparing the log2 ratios for NEFM. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

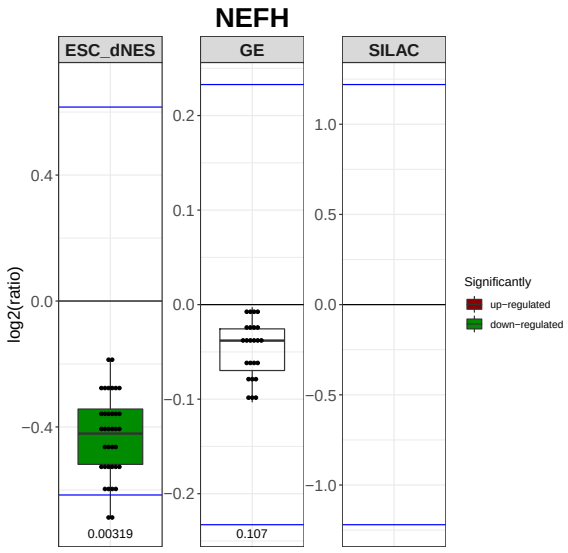


Figure 9.39: *Boxplots comparing the log2 ratios for NEFH. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4.6 NCAM1

NCAM1 is a cell adhesion molecule that creates like-like interactions between neurons. It plays vital roles in neurite outgrowth, synaptic plasticity, memory, and learning. It directs neurite outgrowth through phosphatidylinositol and calcium signalling, and is often used as a neuronal marker. Here, it was up-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and also during differentiation of both $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells (Figure 9.40).

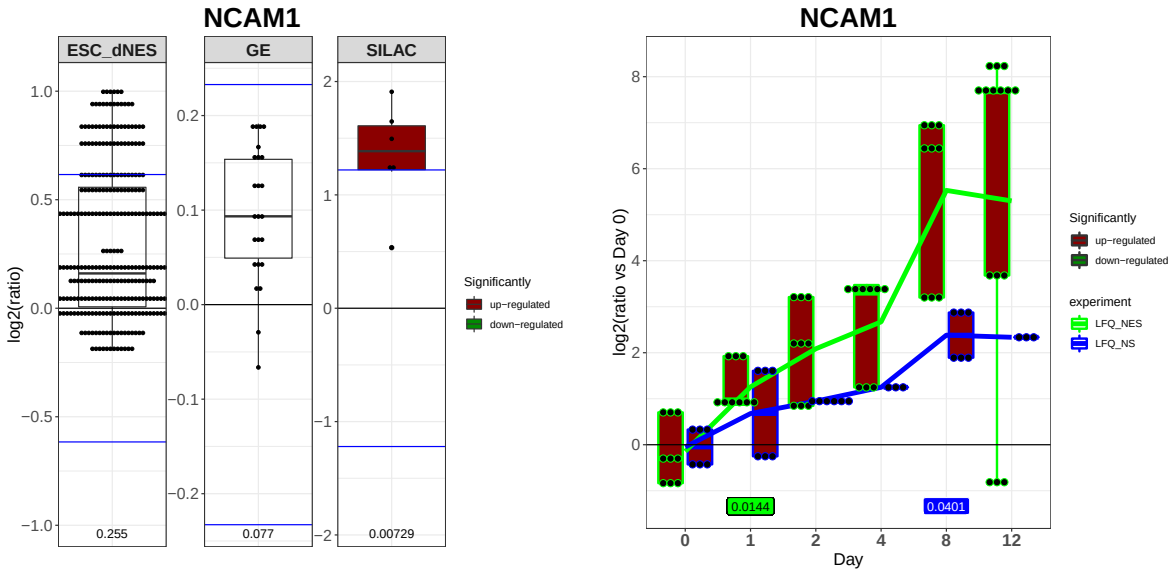


Figure 9.40: *Boxplots comparing the log2 ratios for NCAM1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.4.7 GAP43

GAP43, commonly called neuromodulin, is a cytoplasmic protein that plays a prominent role in brain and particularly cerebellum development. It performs important roles in neurite formation and plasticity. It is expressed at high levels in neuronal growth cones during development. It forms a heterodimer with calmodulin, thereby affecting local calcium signalling. Its influence on calcium signalling may be important for directing growth cone formation. Here, it was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, and the $_{\text{hf}}$NES cell differentiation experiment (Figure 9.41) and may be partly responsible for the more complex morphology of the $_{\text{hf}}$NS cells, as they extend processes that may be immature neurite buds. *GAP43* was up-regulated as the $_{\text{hf}}$NES cells differentiate, but it may be expressed at sufficiently high levels in $_{\text{hf}}$NS cells, for further up-regulation not to be necessary.

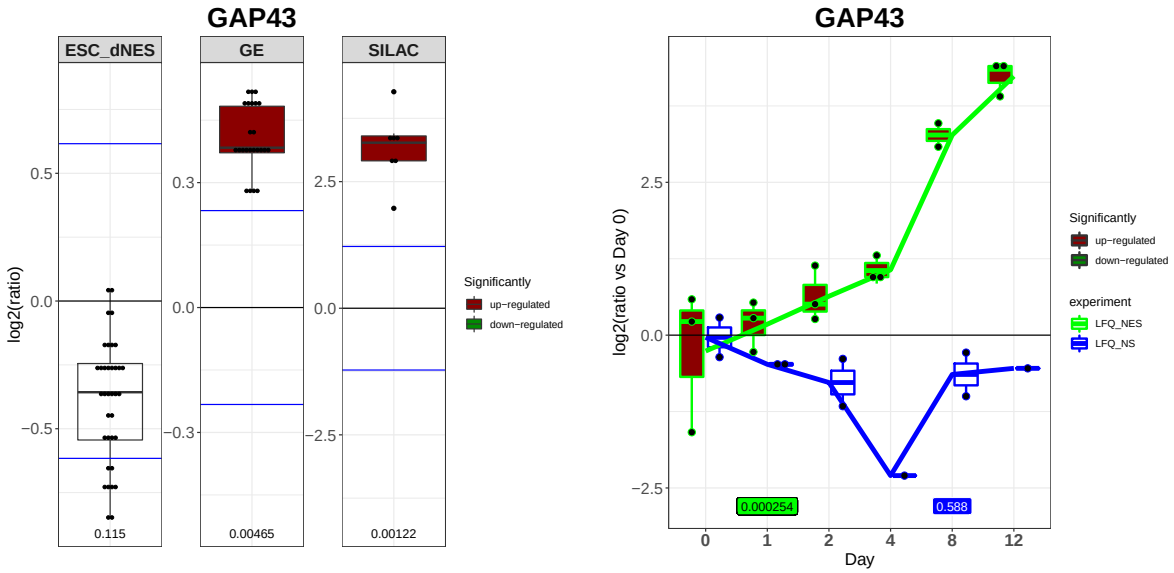


Figure 9.41: *Boxplots comparing the log2 ratios for GAP43. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

9.5 Gliogenic markers

9.5.1 CD44

CD44 is a cell surface glycoprotein receptor for hyaluronic acid (HA), the chief component of extracellular matrix (ECM). It is also a receptor for *SPP1*, membrane metalloproteases (MMPs), and collagens. *CD44* is involved in cell-cell and cell-matrix adhesion, thereby promoting cell aggregation and migration. Here, it was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 9.42). Interestingly, *SPP1* was also up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, which was also corroborated by the SILAC 1 sample (Figure 9.43). *SPP1* may be important for the proliferation and survival of neural stem cells and may even increase neurogenic potential.[172] Its effect appears to be mediated through *CXCR4*, which was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and GE $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 9.43). These interactions explain the importance of the up-regulation of *CD44* in $_{\text{hf}}$NS cells. Many collagens were also SDE in this study, but no MMPs were SDE (Figure 9.44). Information about the differential expression of

collagens may be useful to further optimise cell culture conditions for neural stem cells.

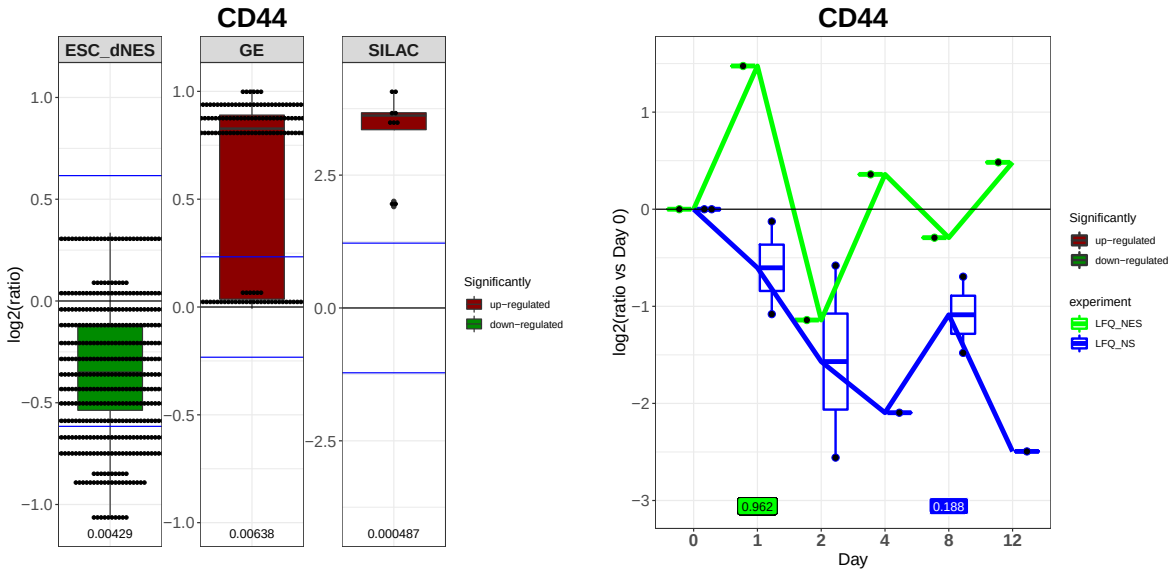


Figure 9.42: *Boxplots comparing the log2 ratios for CD44. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

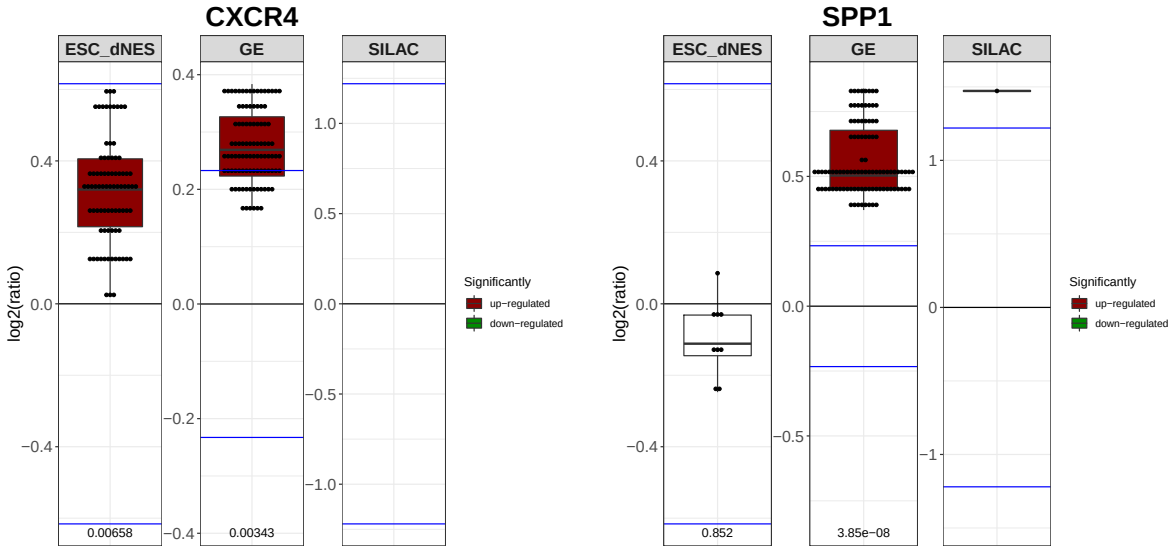


Figure 9.43: *Boxplots comparing the log2 ratios for CXCR4 and SPP1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

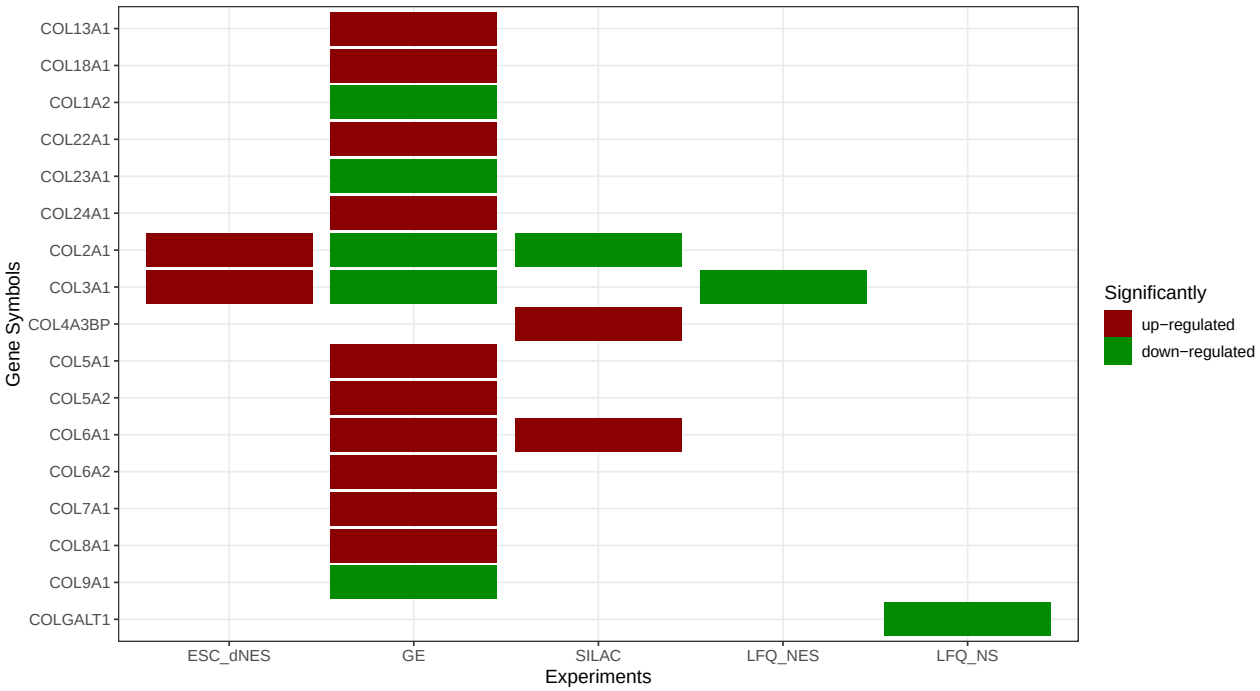


Figure 9.44: *A heatmap of all the collagen molecules significantly differentially expressed (SDE) in all experiment in this work. Red and green indicate up- and down regulation respectively. Red and green indicates up- and down-regulation respectively.*

9.5.2 Cadherins

Epithelial cadherin, *CDH1*, is listed by Abcam[3] as a neuroepithelial cell marker. It showed a non-significant downwards trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and an upwards trend in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.45). It was not present in high enough amounts to be detected in any of the proteomic experiments. Neural cadherin, *CDH2*, is a calcium-dependent cell adhesion protein. It plays important roles in cell attachment and positioning during development. In neural tissue, it is responsible for maintaining attachment between the pre-synaptic and post-synaptic neurons. Here, it showed a gradual upwards trend in all experiments, although it is not SDE in any of these (Figure 9.45). Additional cadherins that are SDE in this work are not discussed (Figure 9.46).

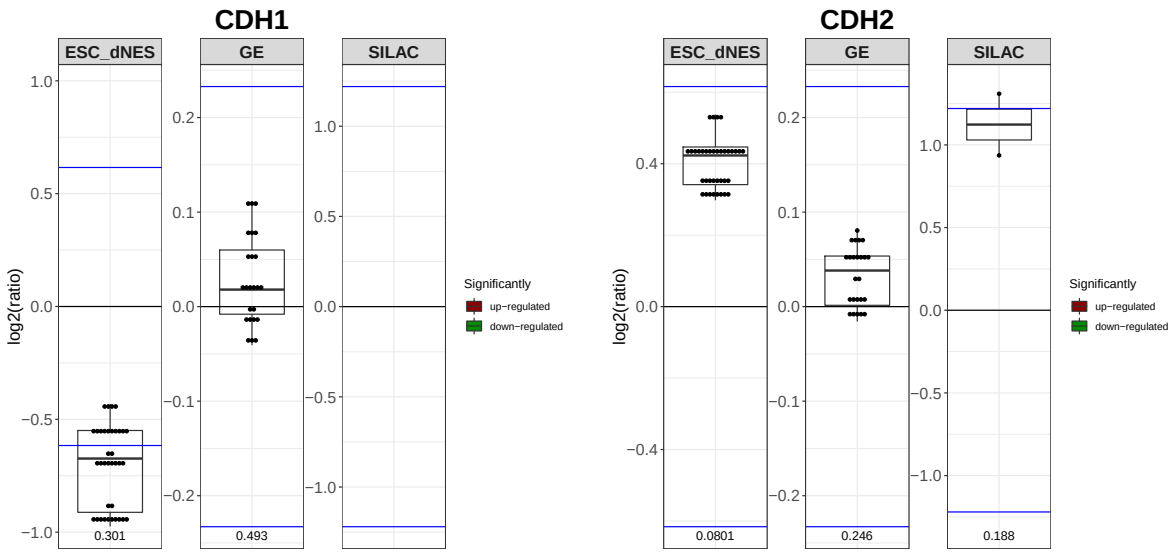


Figure 9.45: *Boxplots comparing the log2 ratios for CDH1 and CDH2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

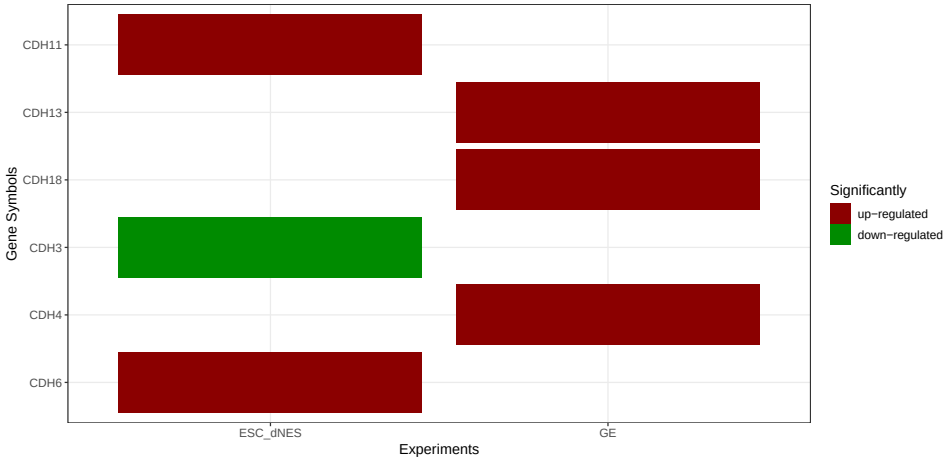


Figure 9.46: *Heatmap of significantly differentially expressed (SDE) cadherin proteins across all experiments, Red and green indicates up- and down-regulation respectively.*

9.5.3 Dihydropyrimidinase-related proteins

During the maturation of the nervous system, members of the dihydropyrimidinase-related protein (DPYSL) family are developmentally regulated.[217] They play roles in neurite outgrowth, axonal guidance, and cell differentiation.[217] *DPYSL2* and *DPYSL3* have been found to be overexpressed in neural stem cells.[249] Interestingly, all the DPYSLs are SDE in this work, and may be useful to characterise the development of neural stem cells. *DPYSL2* has previously been shown to be up-regulated in neural stem cells vs ESC.[249] Here, however, *DPYSL2* only showed a marginal upwards trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, and did not show any significant changes between proliferating $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells. *DPYSL2* was up-regulated in the $_{\text{hf}}$NES cell differentiation experiment and the $_{\text{hf}}$NS cell differentiation experiment, indicating that it may play important roles in differentiated neural tissue cells. *DPYSL3* was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment and in the $_{\text{hf}}$NES cell differentiation experiment. In neuroblastoma cells, its expression was shown to be inhibited by *MYCN*,[217] while here they had inverse expressions patterns. *DPYSL4* was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment and SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, but became up-regulated during differentiation of both cells types. *DPYSL5* was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and up-regulated in the $_{\text{hf}}$NS cell differentiation experiment (Figure 9.48).

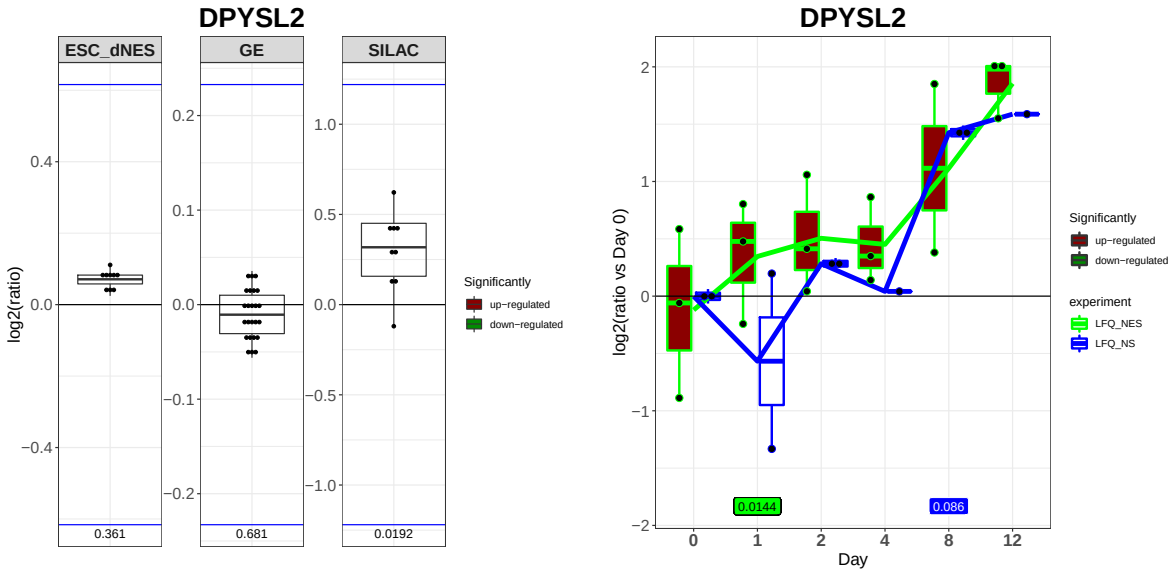


Figure 9.47: *Boxplots comparing the log2 ratios for DPYSL2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

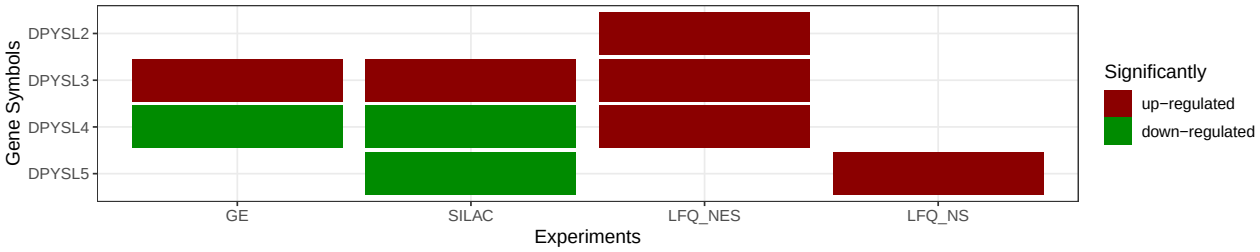


Figure 9.48: *Heatmap of all significantly differentially expressed (SDE) DPSYL proteins across all experiments. Red and green indicates up- and down-regulation respectively.*

9.5.4 Integrins

ITGB4 is a receptor for laminins. Integrins mediate cell-to-cell and cell-to-matrix adhesion. They also transduce signals into cells. *ITGB4* has been shown to be involved in foetal astroglial migration *in vitro*.[230] Here it was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. There are a number of integrins SDE in these data, almost all were up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 9.49).

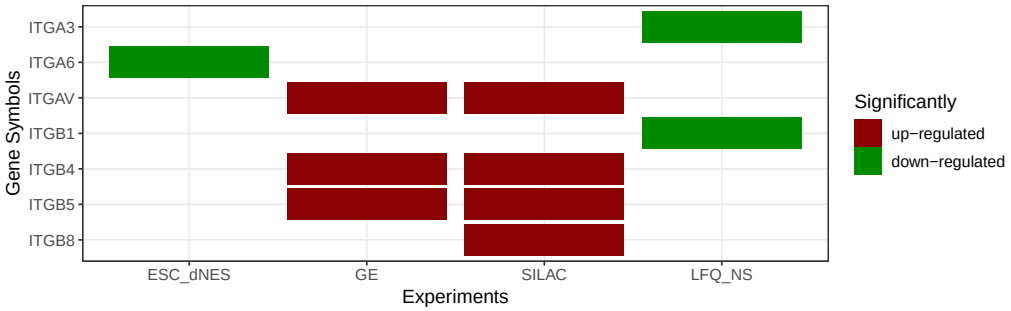


Figure 9.49: *Heatmap of all integrins significantly differentially expressed (SDE) across all experiment in this work*

Chapter 10

Cell and tissue enrichment

10.1 Tissues

Tissue enrichment analysis was performed using the STRINGdb. The SDE genes and proteins from each experiment were enriched relative to that of all genes or proteins detected in that experiment. The p-values were corrected via Bonferroni correction.

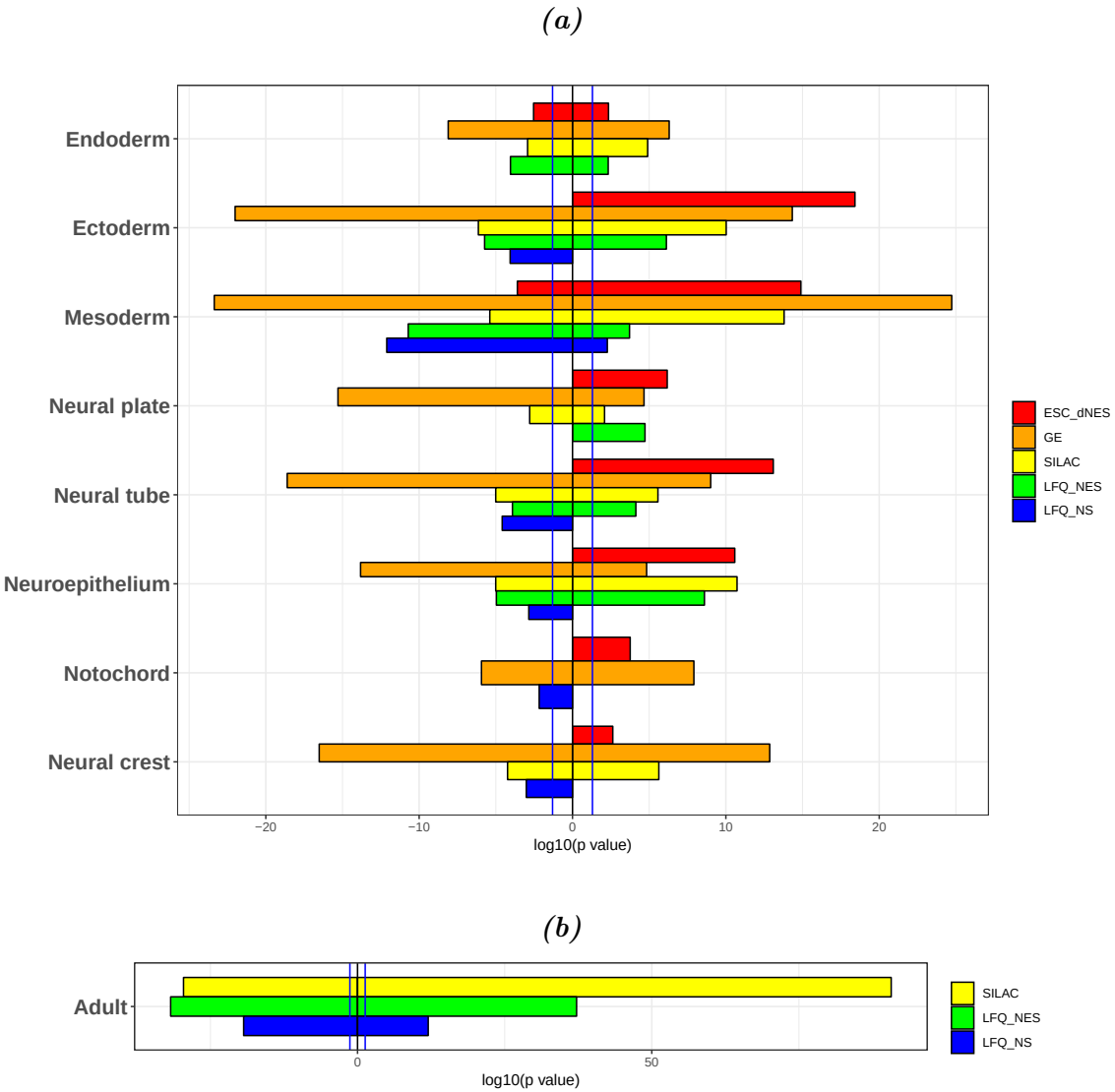


Figure 10.1: *STRINGdb[73] package in R enrichment of (a) germline, embryonic neural tissues and (b) adult tissue. The adult tissues are highly enriched and plotted on a different scale. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the $_{hf}$NES cell differentiation experiment, and LFQ_NS (blue) refers to the $_{hf}$NS cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

Tissue enrichment for the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment revealed the up-regulation of genes associated with the ***mesoderm***, ***ectoderm***, ***neural plate***, ***neural tube***, ***neuroepithelium***, ***notocord***, and ***neural crest*** (Figure 10.1 (a)). This is in concordance with hESCs differentiating into dNES cells. *In vivo*, neuroepithelial cells would be a part of all of these tissues. In

the remaining experiment, the cells down-regulate genes from these tissues as they differentiate into RG and terminally differentiate into neurons and glia. (Figure 10.1 (a)). Interestingly, adult tissue was highly enriched in all the proteomic experiments, but not in the transcriptomic experiments (Figure 10.1 (b)).

10.2 Cells

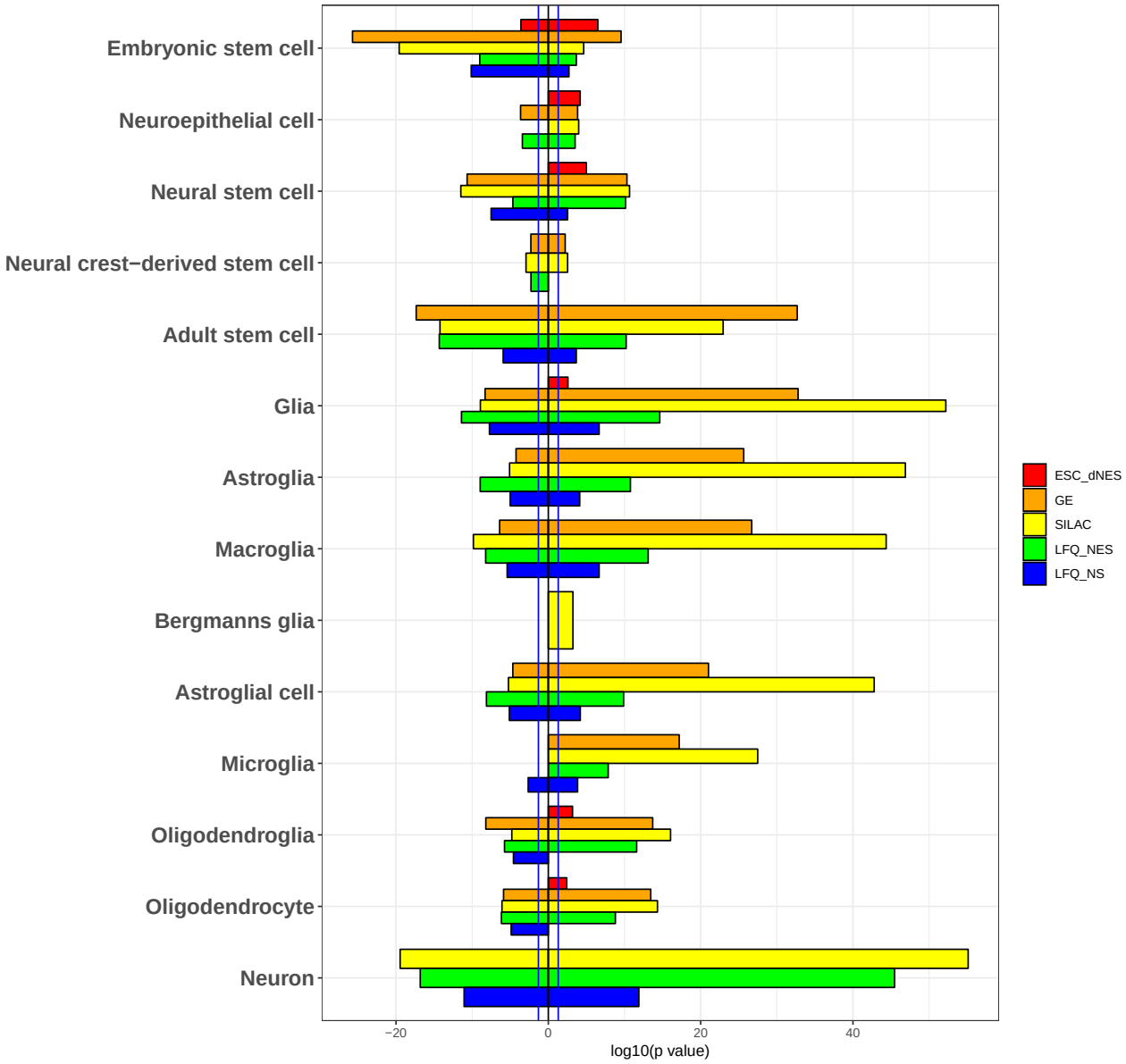


Figure 10.2: *STRINGdb enrichment of stem cells and neural tissue cells. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

Similarly, the genes enriched across cells largely corroborate the cell model (Figure 10.2). ***Embryonic stem cells*** were down-regulated during differentiation. ***Neuroepithelial*** proteins, paradoxically were up-regulated during differentiation. ***Neural stem cells*** were enriched both ways, except in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. ***Adult stem cell*** genes were strongly enriched in all cells except those of the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiments. ***Glial cells*** were more highly enriched in $_{\text{hf}}$NS cells than $_{\text{hf}}$NES cells. Interestingly, ***Bergmanns glia*** found in the cerebellum were only enriched in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment $_{\text{hf}}$NS cells. ***Neurons*** were also enriched in all differentiating cells, but more so in the $_{\text{hf}}$NS cells and differentiating $_{\text{hf}}$NES cell. Paradoxically, there was no ***neuron*** enrichment observed in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 10.2).

Chapter 11

KEGG pathway enrichment

KEGG enrichment analysis was generated using the STRINGdb. The SDE genes or proteins were enriched relative to that of all genes or proteins detected in the experiment. The p-values were corrected via Bonferroni correction.

11.1 KEGG pathways down-regulated in NES

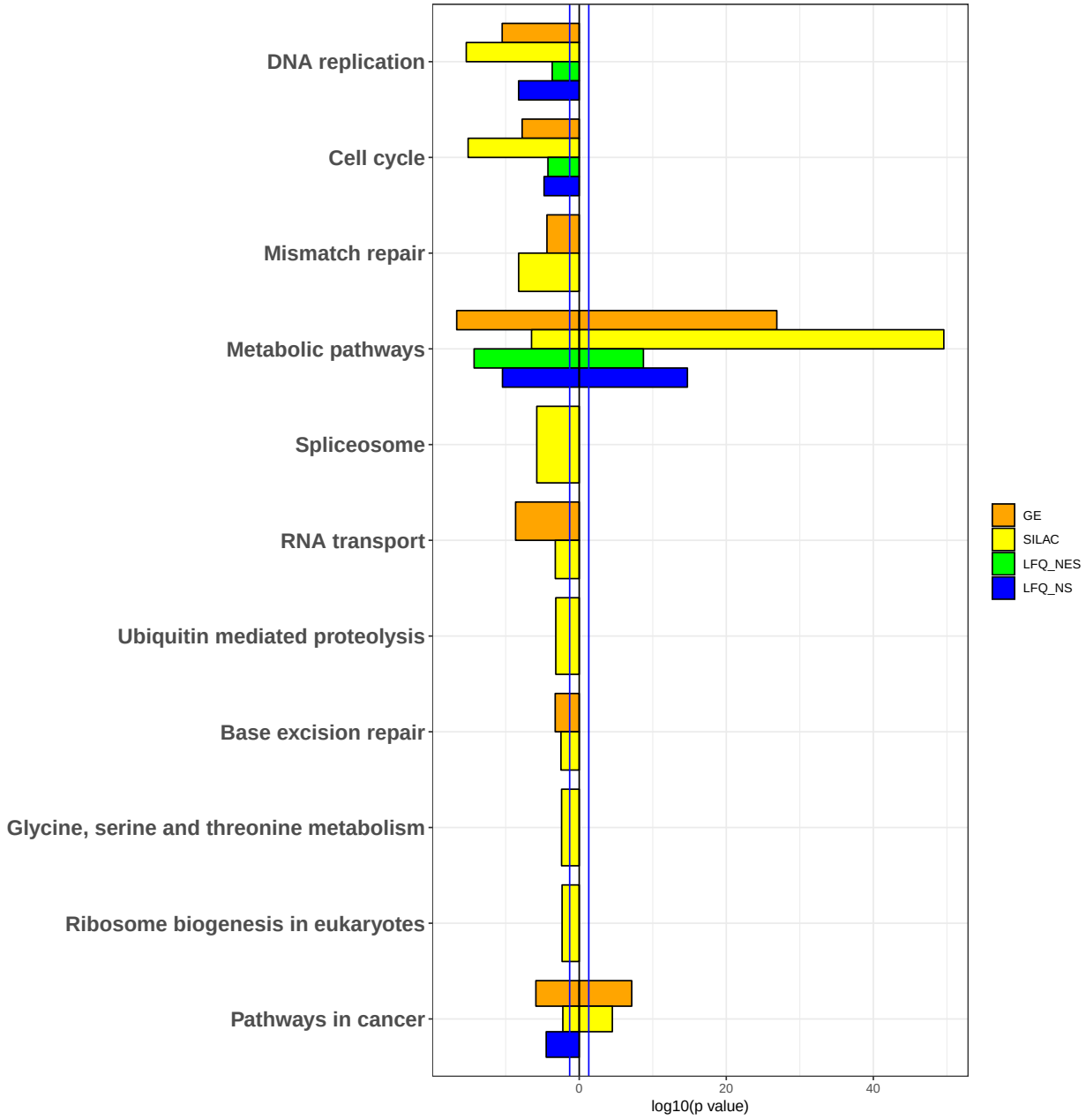


Figure 11.1: *Kyoto encyclopedia of genes and genomes enrichments for SILAC $\frac{NS}{NES}$ experiment down-regulated proteins. All other experiment relative enrichment are overlayed on this plot to illustrate potential differences. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

The pathways downregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment are mainly involved in supporting the cell cycle. Several metabolic pathways were also enriched, and these may also play a role in supporting the cell cycle.

11.1.1 Cell cycle

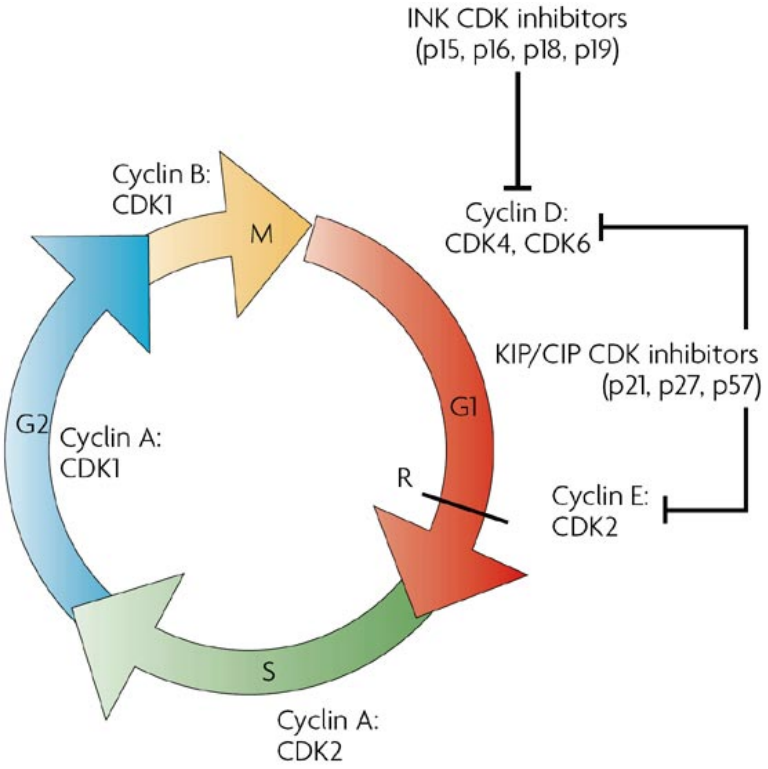


Figure 11.2: *Cell cycle.*[56]

During the progression of stem cells towards lineage restricted progenitors there is a reduction in the rate of the cell cycle. The first growth phase of the cell cycle (G1) phase increases in length allowing cells to be influence more strongly by extracellular signals, and thereby accumulating fate specifying transcription factors.[88,164,180] The cell cycle has a number of different stages (Figure 11.2) controlled by the cycling of the cyclin and their dependent kinases.The cyclin-dependent kinases (CDKs) and cyclins (CCNs) are responsible for progression of the cells cycle, whereas cyclin-dependent kinase inhibitor (CDKN) restrict the cell cycle. The higher expression of *CDK1* and *CDK2* in $_{\text{hf}}$NES cells would push these cells through the growth phases of the cell cycle, thereby facilitating rapid proliferation. Interestingly greater

difference were observed in CDK proteins in the proteomics data than in the genomic data, whereas greater differences were observed for the CCNs in the transcriptomics data than the proteomics data. Most importantly down-regulation of *CCND2*, *CCNJ*, and *CCNB1* and up-regulation of *CCNA1* and *CCND1* (Figure 11.3). During differentiation *CDK1*, *CDK2*, and *CDK6* were down-regulated, this would most likely halt the cell cycle.

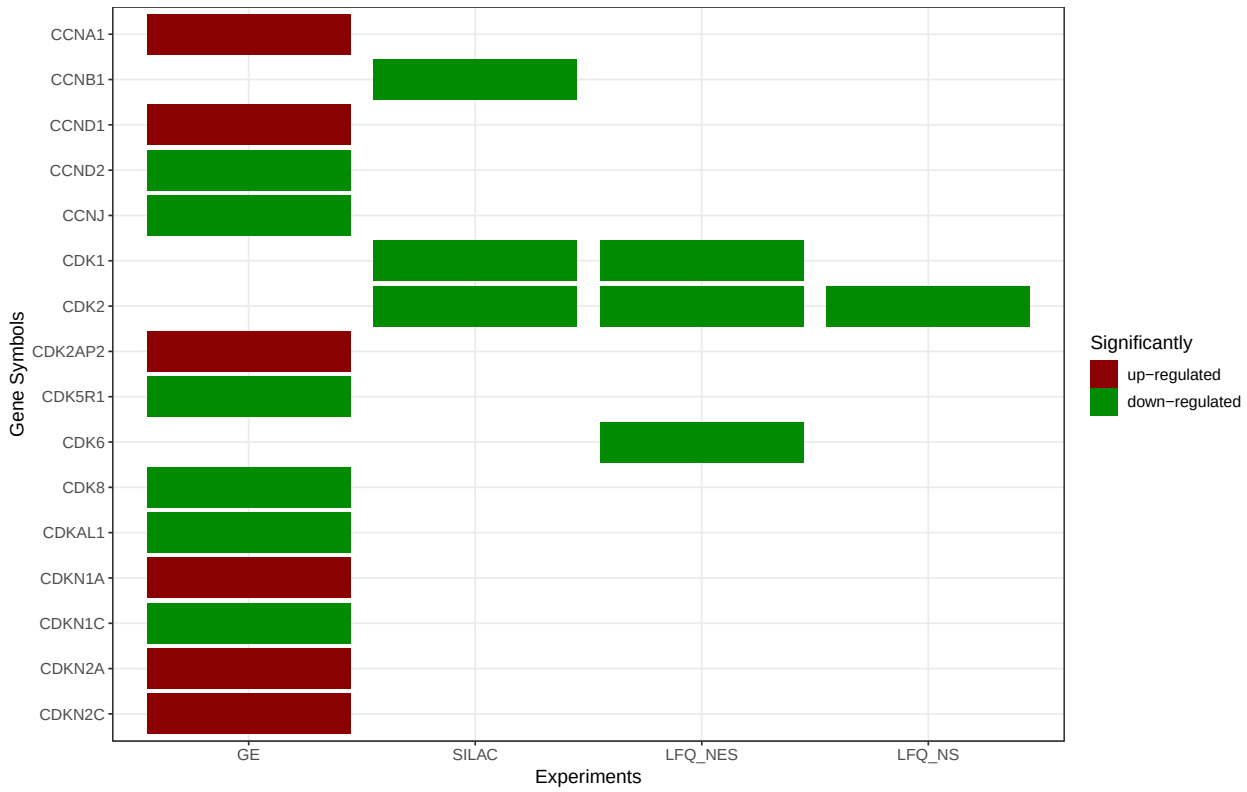


Figure 11.3: *Heatmaps of CDKs and CCNs across all experiments. Red and green indicates up- and down-regulation respectively.*

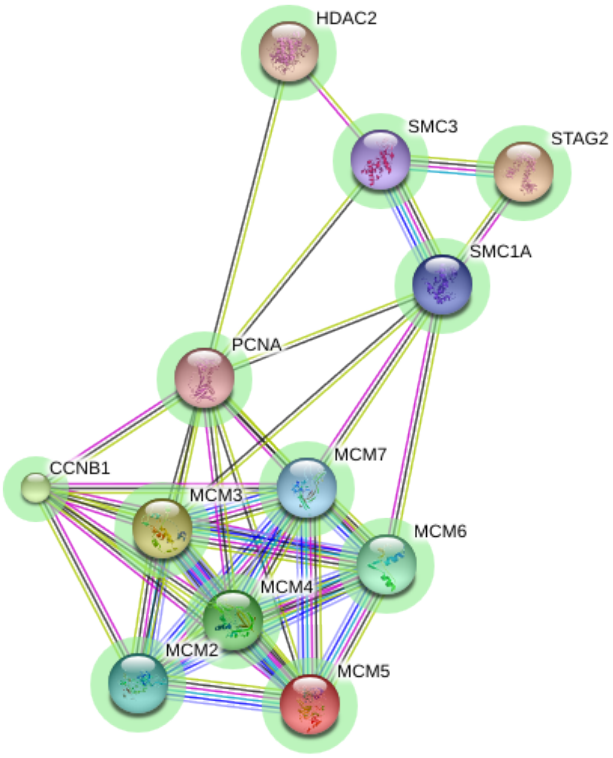


Figure 11.4: *STRINGdb network of KEGG* ***Cell cycle*** *proteins SDE in the SILAC* $\frac{NS}{NES}$ *experiment. Red and green halos indicate up- and down-regulation respectively.*

The KEGG term ***cell cycle*** was highly enriched in $_{hf}$NES cells and therefore down-regulated in the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments and down-regulated in both the $_{hf}$NES and $_{hf}$NS cell differentiation experiments. There were many more genes SDE in the GE $\frac{NS}{NES}$ experiment than in the SILAC $\frac{NS}{NES}$ experiment. Most of those found to be SDE in the SILAC $\frac{NS}{NES}$ experiment were however not SDE in the GE $\frac{NS}{NES}$ experiment (Figure 11.5). Despite the poor overlap, there were no contradictions between the genes and proteins found to be SDE. This indicates the potential post-transcriptional regulation of these proteins. The down-regulated proteins in the SILAC $\frac{NS}{NES}$ experiment mainly form part of the CC ***MCM complex*** (see Figure 11.4 and section 12.1.1). Members of the ***MCM complex*** and *PCNA* and were down-regulated across all proteomic experiments (Figure 11.5).

Figure 11.5: *A heatmap of SDE genes and proteins across all experiments that form part of the KEGG* ***cell cycle*** *pathway. Red and green indicates up- and down-regulation respectively.*

11.1.2 DNA replication

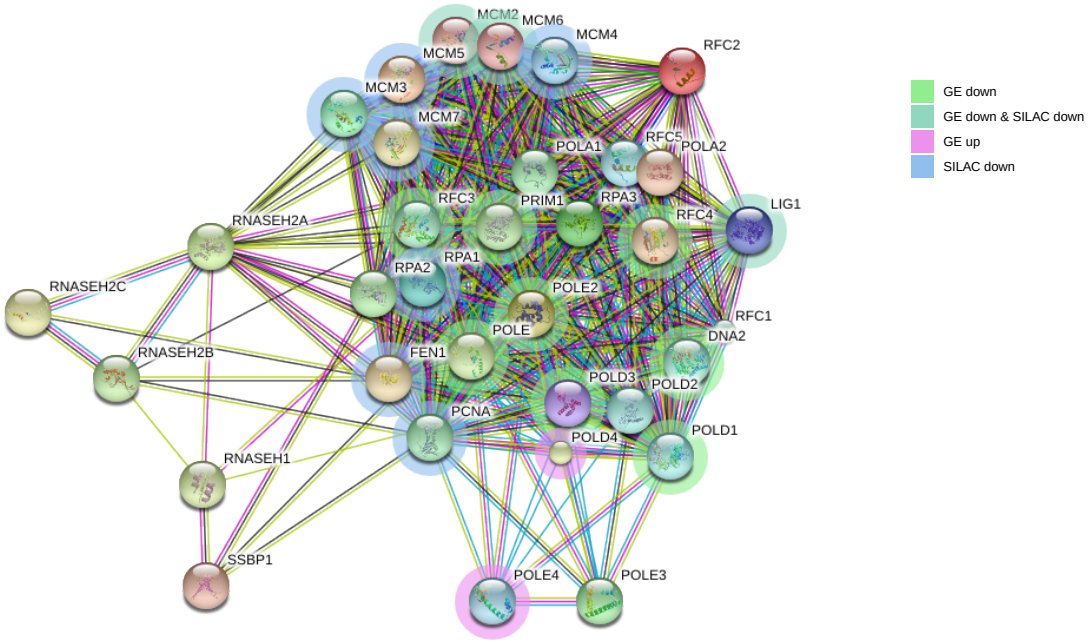


Figure 11.6: *STRINGdb network for the KEGG pathway* ***DNA replication*** *for the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

KEGG ***DNA replication*** was enriched in the down-regulated genes and proteins from all experiments except the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. In the GE $\frac{\text{NS}}{\text{NES}}$ experiment, there was a difference between DNA polymerases expressed in $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells. *POLD1*, *POLD3*, *POLE*, and *POLE2* were down-regulated whereas *POLD4* and *POLE4* was up-regulated (Figure 11.7). *POLD1*, *POLD3*, and *POLE* showed similar trends in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment despite not having sufficient replicates to be SDE (Figure 11.8). These DNA polymerase delta and epsilon, are responsible for extending the leading and lagging strands during DNA replication. Interestingly, *POLQ*, for which there is little information in the literature, is down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. It is involved in the double-stranded break repair pathway alternative end-joining pathway, reviewed in Black et al. 2016.

PCNA was downregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and the $_{\text{hf}}$NES and $_{\text{hf}}$NS cell differentiation experiments, and it is instrumental in DNA replication. It localises to pre-replication

sites before replication begins. It is a DNA clamp, that forms a homodimer encircling the DNA and acts as a scaffold. It tethers the replication complex to the DNA strand. This sliding platform is necessary to increase the processivity of DNA polymerases. *PCNA* is involved in DNA replication, DNA repair and chromatin remodelling. It interacts with *LIG1*, *FEN1*, *DNMT1*, *CHAF1B*, and *MSH6*, which were all down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment.

Figure 11.7: *Heatmap for the KEGG pathway* ***DNA replication*** *for the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experiments. Red and green indicates up- and down-regulation respectively.*

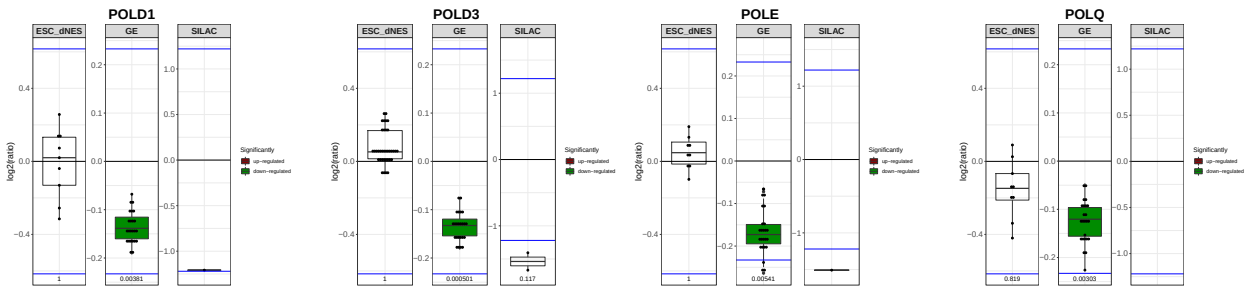


Figure 11.8: *Boxplots comparing the log2 ratios for POLD1, POLD3, POLE and POLQ. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

11.1.3 Mismatch repair

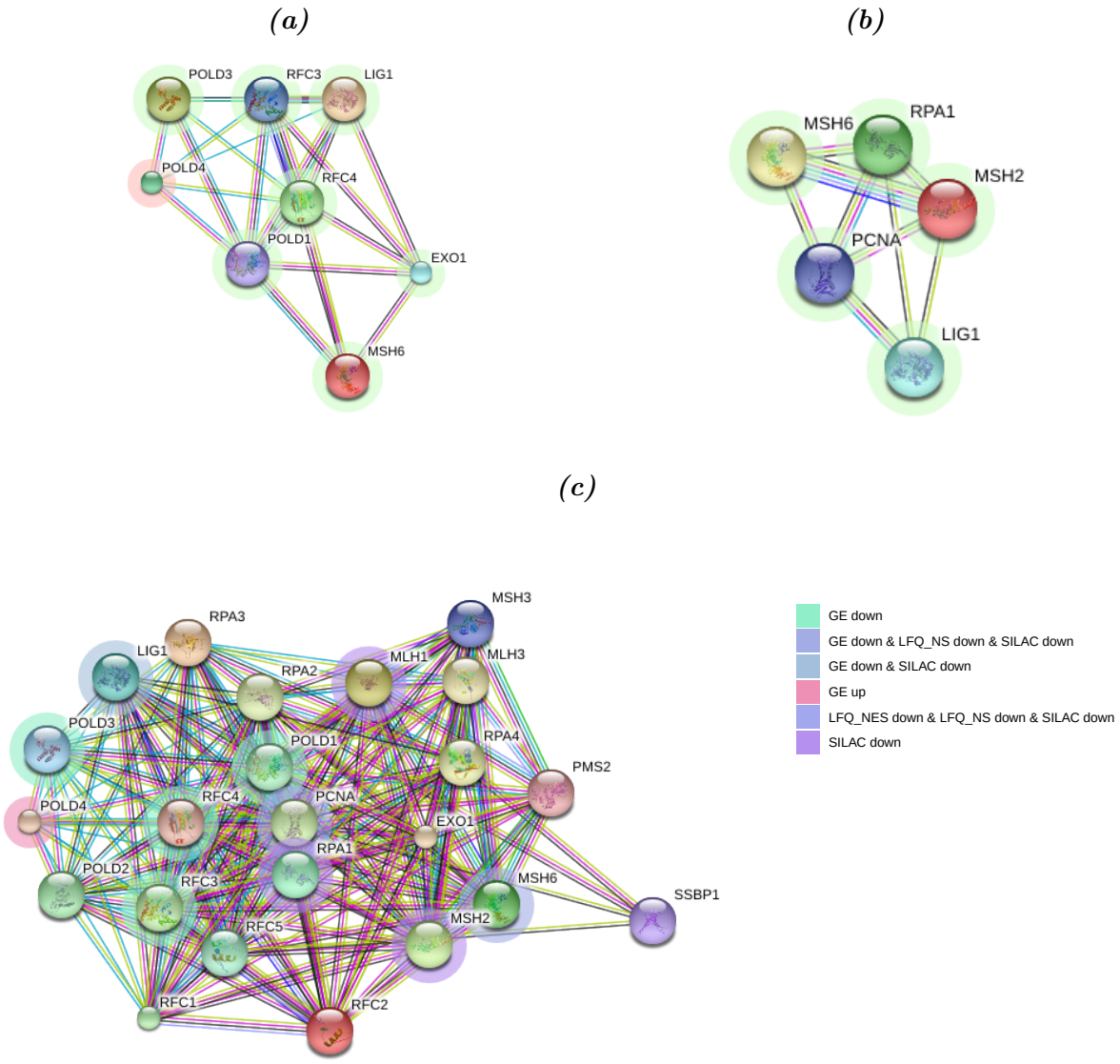


Figure 11.9: *STRINGdb network of KEGG* ***Mismatch repair*** *for SDE genes and proteins from the (a) the GE $\frac{NS}{NES}$ experiment, (b) the SILAC $\frac{NS}{NES}$ experiment, and (c) a combined plot. In plots (a) and (b), red and green halos indicate up- and down-regulation respectively.. In (c) the halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

The KEGG term ***Mismatch repair*** is enriched in down-regulated genes and proteins from the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. Even in cells not exposed to mutagens, DNA damage occurs at a rate of approximately 10000 hits/cell.[165] Mismatch repair recognises erroneously inserted or deleted nucleotides on the newly synthesised daughter DNA strand and repairs them. It is strand-specific, as errors are most likely on the daughter strand. Mismatches are recognised

by the two heterodimers MutS-alpha (*MSH2* and *MSH3*) and MutS-beta (*MSH2* and *MSH3*). *MSH2* and *MSH6* were upregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment. These heterodimers recruit the heterodimer MutL-alpha (*MLH1* and *PMS2*). *PMS2* nicks the DNA spanning the mismatch, and *EXO1* then removes the incorrect DNA fragment. The single-stranded gap is filled by *POLD1* and its cofactors *PCNA* and replication factor C. The remaining nick is sealed by *LIG1*. This pathways is expressed during active cell proliferation.

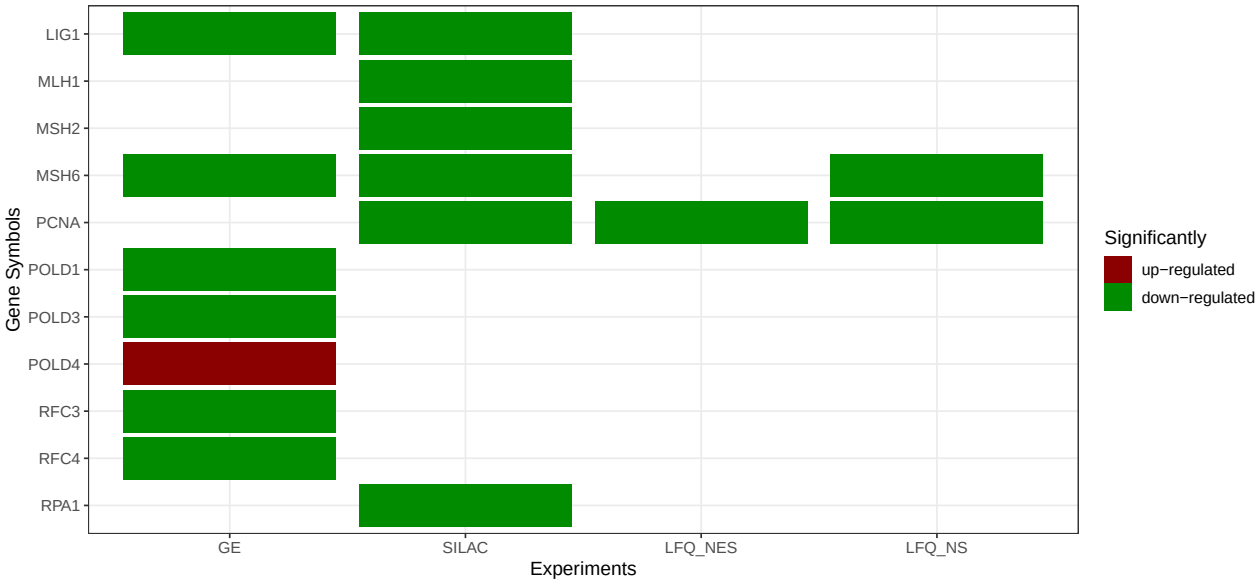


Figure 11.10: *Heatmap of KEGG* ***Mismatch repair*** *for SDE genes and proteins. Red and green indicates up- and down-regulation respectively.*

11.1.4 RNA transport

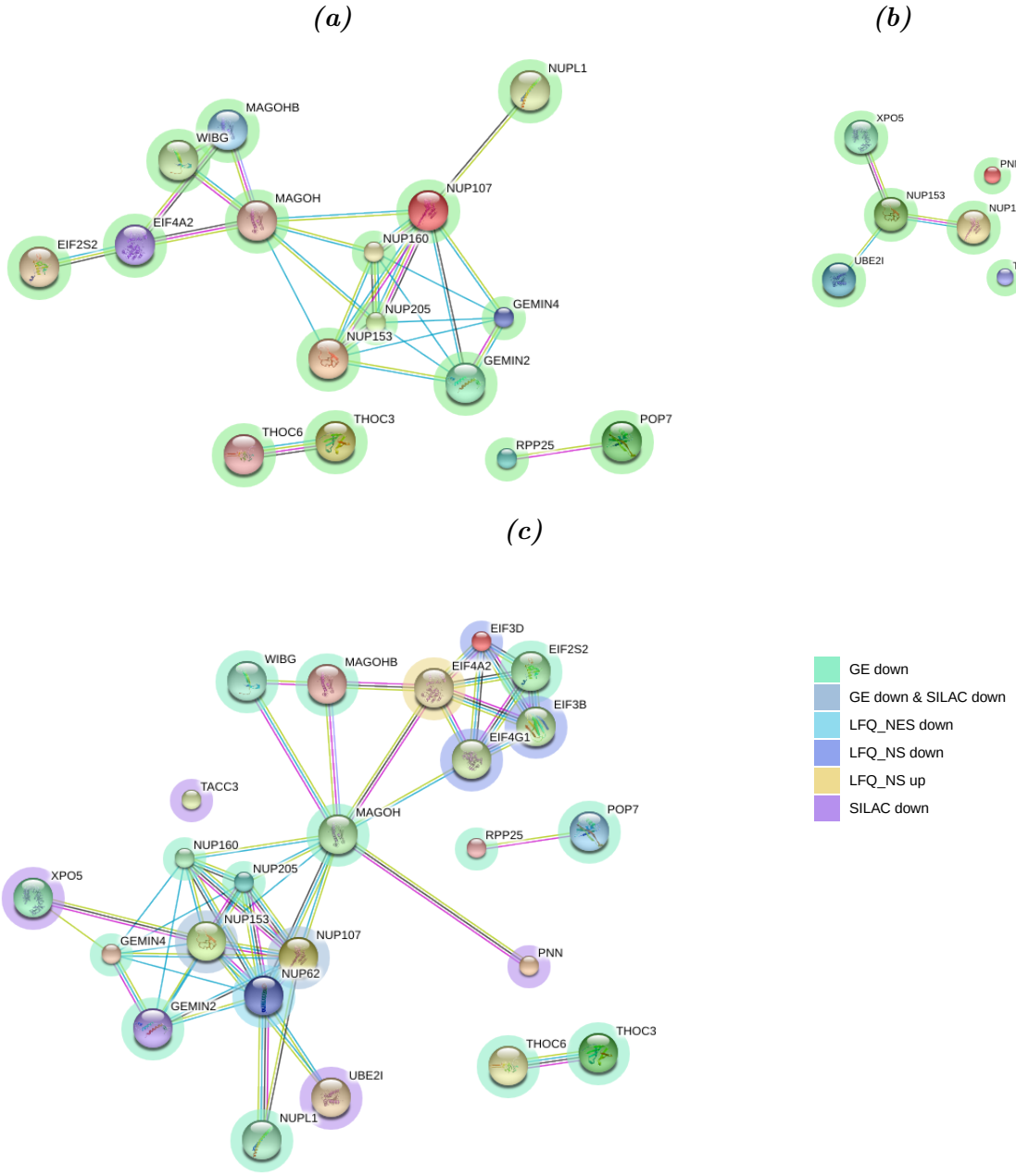


Figure 11.11: *STRINGdb network of KEGG* ***RNA transport*** *for SDE genes and proteins from (a) the GE $\frac{NS}{NES}$ experiment, (b) the SILAC $\frac{NS}{NES}$ experiment, and (c) a combined plot. In plot (a) and (b) red and green halos indicate up- and down-regulation respectively. In (c) the halo colours indicate the overlaps in SDE between experiments, as per the legend*

The KEGG term ***RNA transport*** refers the transport of ribonucleic acid (RNA) from the nucleus to the cytoplasm. This process is fundamental to gene expression. Karyopherin-beta

proteins are involved in the transport of transport RNAs (tRNAs), ribosomal RNAs (rRNAs), and small nuclear RNAs (snRNAs), whereas mRNAs are transported through the nucleopore. *NUP35* and *NUP153*, which were down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, and are components of the nucleopore. A number of additional nucleopore proteins were also down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (see Figure 11.11 and 11.14), together with *GEMIN4*, which is involved in the maturation of snRNAs.

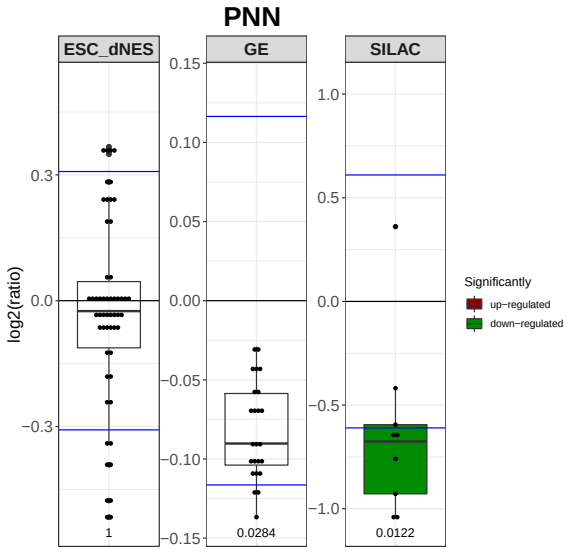


Figure 11.12: *Boxplots comparing the log2 ratios for PNN. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

The apoptosis and splicing–associated protein complex (ASAP complex) is composed of *SAP18*, *RNPS1* and *ACIN1*, none of which were SDE in these experiments (Figure 11.13). *PNN* localises to the desmosome plaque, where it enhances epithelial cell attachment.[100] It is also found in the nuclear speckles, where it replaces *ACIN1* in the ASAP complex to form the Pinin containing ASAP complex (PSAP complex), which is involved in recruiting the exon junction complex (EJC) to splice sites on mRNA. *PNN*, which was downregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 11.12). *MAGOH*, *MAGOHB* and *WIBG*, components of the EJC were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. During development, the expression of *PNN* varies in different tissues and it is highly expressed in the neuroepithelium.[100]

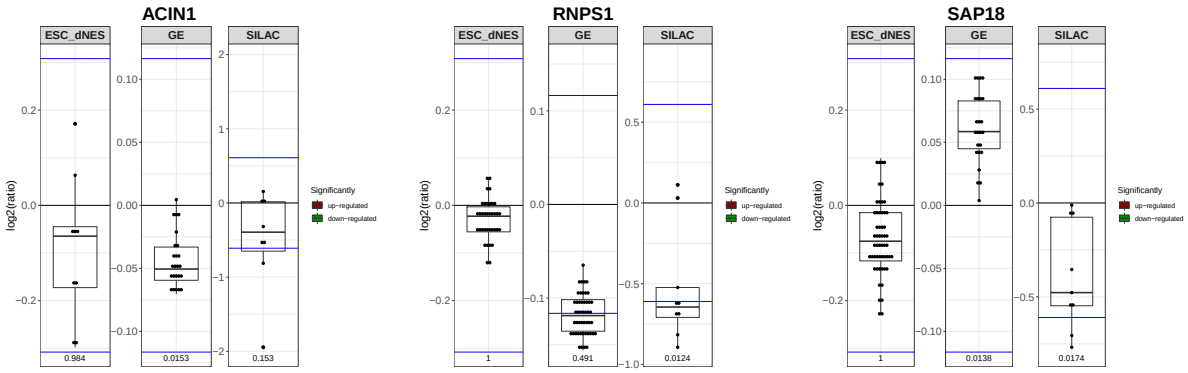


Figure 11.13: *Boxplots comparing the log2 ratios for ACIN1, RNPS1 and SAP18. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

PNN shows a higher affinity for the complex than *ACIN1*, which is unable to displace it. These complexes regulate transcription and splicing. Replacing *ACIN1* with *PNN* would affect splice site selection and thereby potentially affect the identity of cells. *PNN* is also thought to connect the chromatin to splicing events. Splicing is known to take place during transcription and can therefore be affected by the chromatin state. If it is knocked down in the corneal epithelium, it may cause a loss of cell-cell adhesion and changes in cell shape and movement out of the epithelium.[12] Overexpression of *PNN* in HEK293 cells results in more epithelial-like phenotypes.[100] In a pull-down experiment *PNN* was found to associate with *SRSF2* and *MAGOH*[108] both were SDE here.

XPO5, which was downregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 11.14), mediates the transport of proteins containing double-stranded RNA-binding domains and their cargos. These can take the form of short hairpin interfering RNA. Its overexpression can enhance RNA-mediated gene silencing. It requires Ras-related nuclear protein (Ran)-GTP for transport. *XPO5* is responsible for *let-7* export out of the nucleus. Its overexpression would enhance the *LIN28 let-7* pathways, discussed in section 13.2.3.

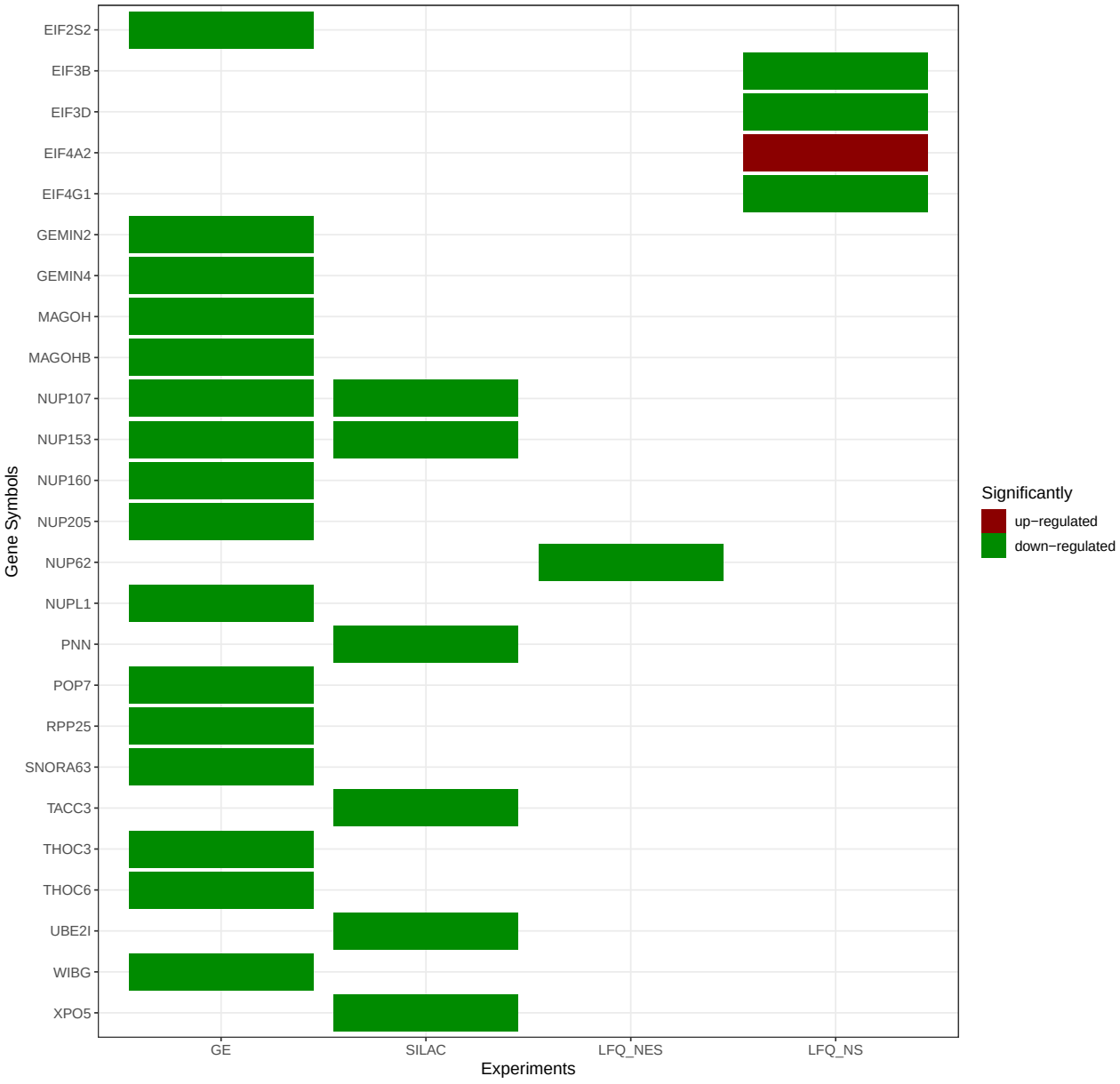


Figure 11.14: *Heatmap of KEGG* ***RNA transport*** *for SDE genes and proteins. Red and green indicates up- and down-regulation respectively.*

11.2 KEGG pathways up-regulated in NS

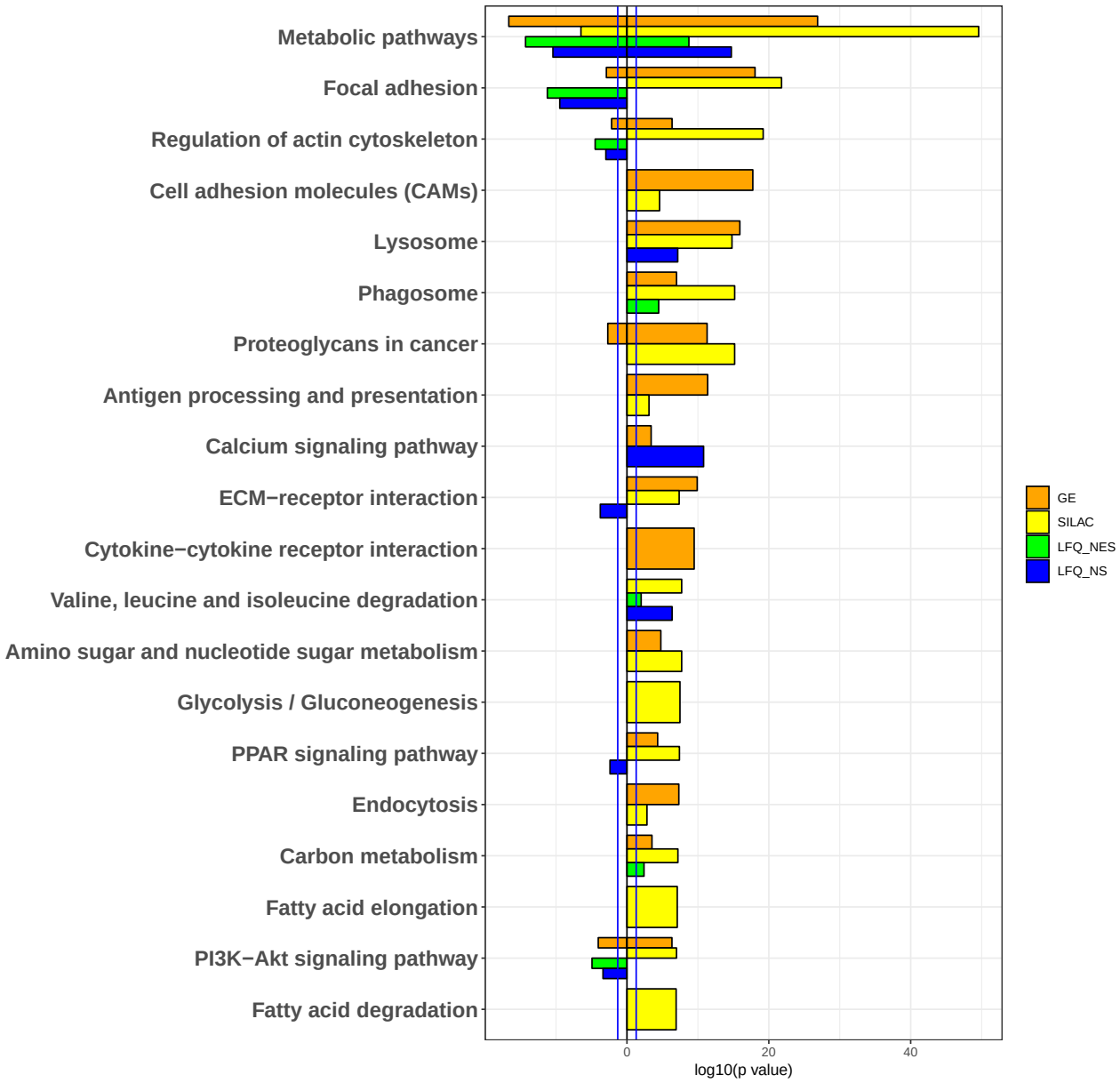


Figure 11.15: *Top twenty KEGG terms enriched in the up-regulated genes and proteins across all experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the $_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the $_{hf}NS$ cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

11.2.1 ECM-receptor interaction

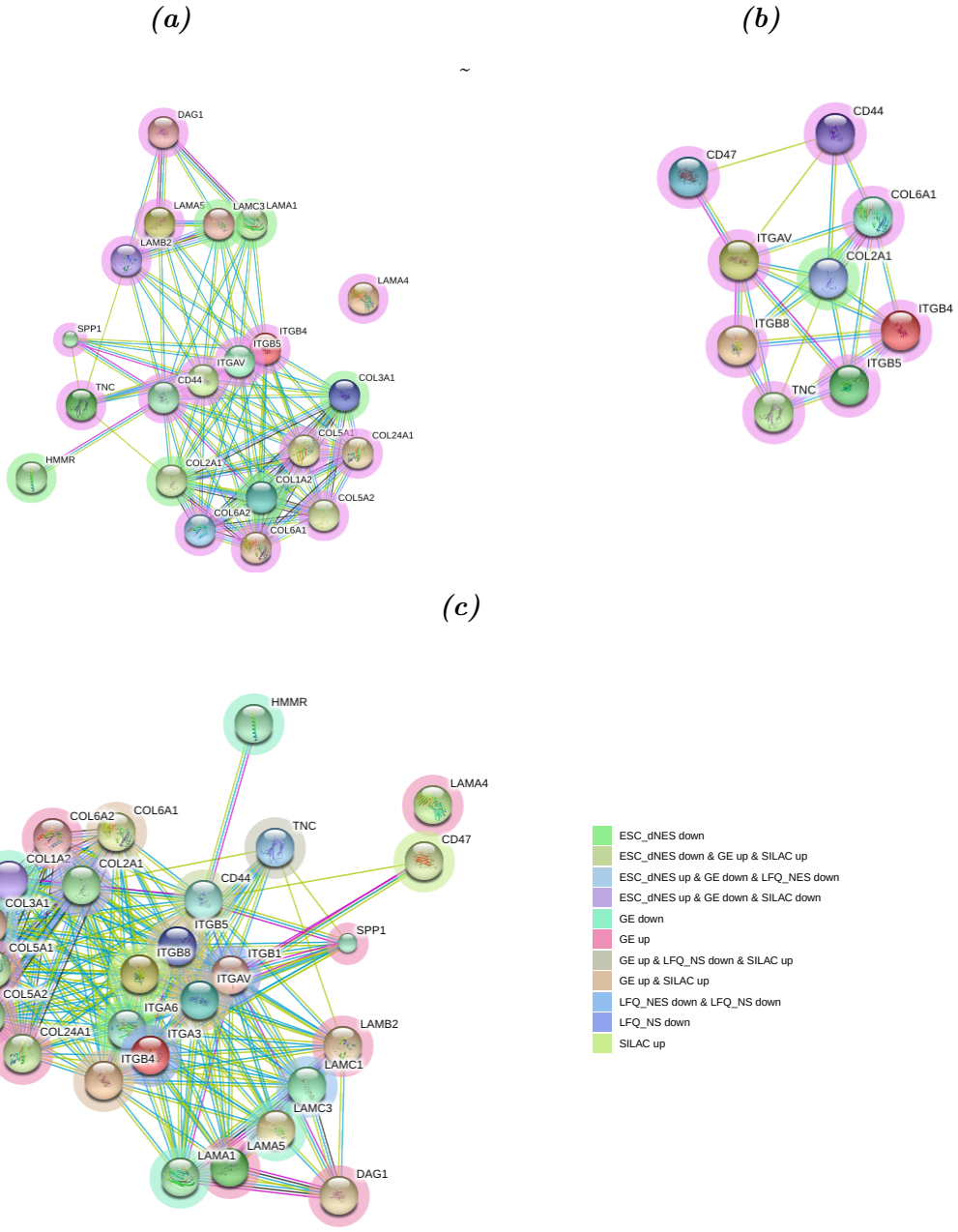


Figure 11.16: *STRINGdb network for KEGG* ***ECM-receptor interaction*** *(a) the GE $\frac{NS}{NES}$ experiment and (b) the SILAC $\frac{NS}{NES}$ experiment, and (c) all the experiments combined. In (a) and (b) red and green halos indicate up- and down-regulation respectively. In (c) the halo colours indicate the overlaps in SDE between experiments, as per the legend*

The KEGG term ***ECM-receptor interaction*** was enriched in up-regulated genes and proteins of the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and in the down-regulated proteins in the

$_{hf}$NS cell differentiation experiment (Figure 11.15). Upon $_{hf}$NES cells differentiation into $_{hf}$NS cells, the protein *PNN*, which plays a vital role in epithelial adherence is down-regulated. This may lead to the up-regulation of ECM receptor proteins such as *CD44*, *COL6A1*, the integrins (*ITGAV*, *ITGB4*, *ITGB5*, *ITGB8*), *LAMB2*, and *TNC* (Figure 11.18).

CD44 is a receptor for hyaluronic acid, one of the chief components of the ECM. *CD44* senses stimuli from the environment and regulates migration and proliferation. It is also involved in cell-cell and cell-substrate adhesion. *CD44* is expressed as a number of different isoforms. The standard isoform is expressed ubiquitously, but some isoforms are selectively expressed in epithelial tissues at specific stages of development. *CD44* also binds *SPP1*, collagens, and MMPs. *SPP1* and certain collagens are up-regulated in $_{hf}$NS cells, but not MMPs. Overexpression of *CD44* is known to attenuate the Hippo pathway.[128]

The integrins *ITGAV*, *ITGB4*, *ITGB5*, and *ITGB8* were also up-regulated. The integrins form heterodimers, usually with one alpha and one beta subunit. Different heterodimers play different roles. They sense the ECM, and are necessary for the proliferation of neural stem cells, particularly *ITGAV* and *ITGB3*. *ITGB3*, was not SDE but did show some degree of up-regulation in $_{hf}$NS cells (Figure 11.17).

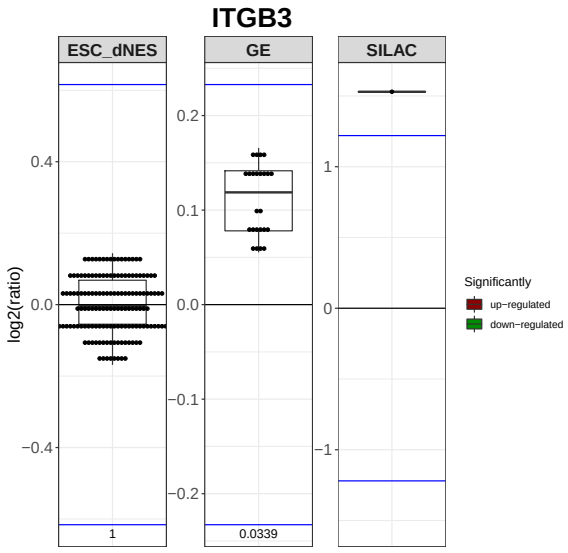


Figure 11.17: *Boxplots comparing the log2 ratios for ITGB3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

In the GE $\frac{\text{NS}}{\text{NES}}$ experiment, laminins *LAMA1* and *LAMC3* were down-regulated while *LAMA4*, *LAMA5*,*LAMB2* were up-regulated. Laminins form part of the ECM basal membrane.

ECM-receptor interaction was enriched in $_{\text{hf}}$NS cells and in the down-regulated proteins of the $_{\text{hf}}$NS cell differentiation experiment. This indicates that the interaction with the ECM is more important in $_{\text{hf}}$NS cells than $_{\text{hf}}$NES cells and is also more important in their stem cell phase than after differentiation.

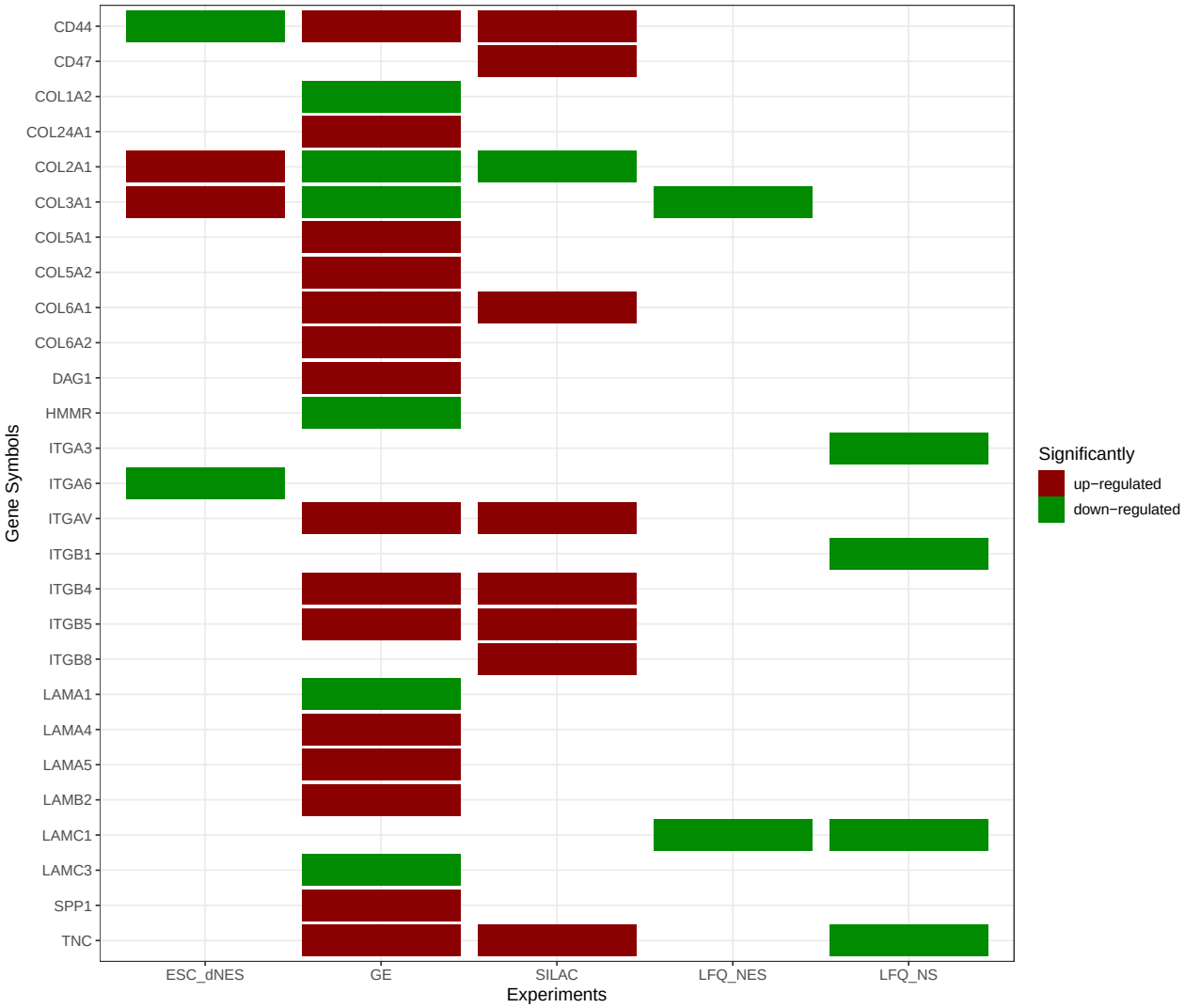


Figure 11.18: *Heatmap of KEGGs* ***ECM-receptor interaction****. Red and green indicates up- and down-regulation respectively.*

11.2.2 Focal adhesion

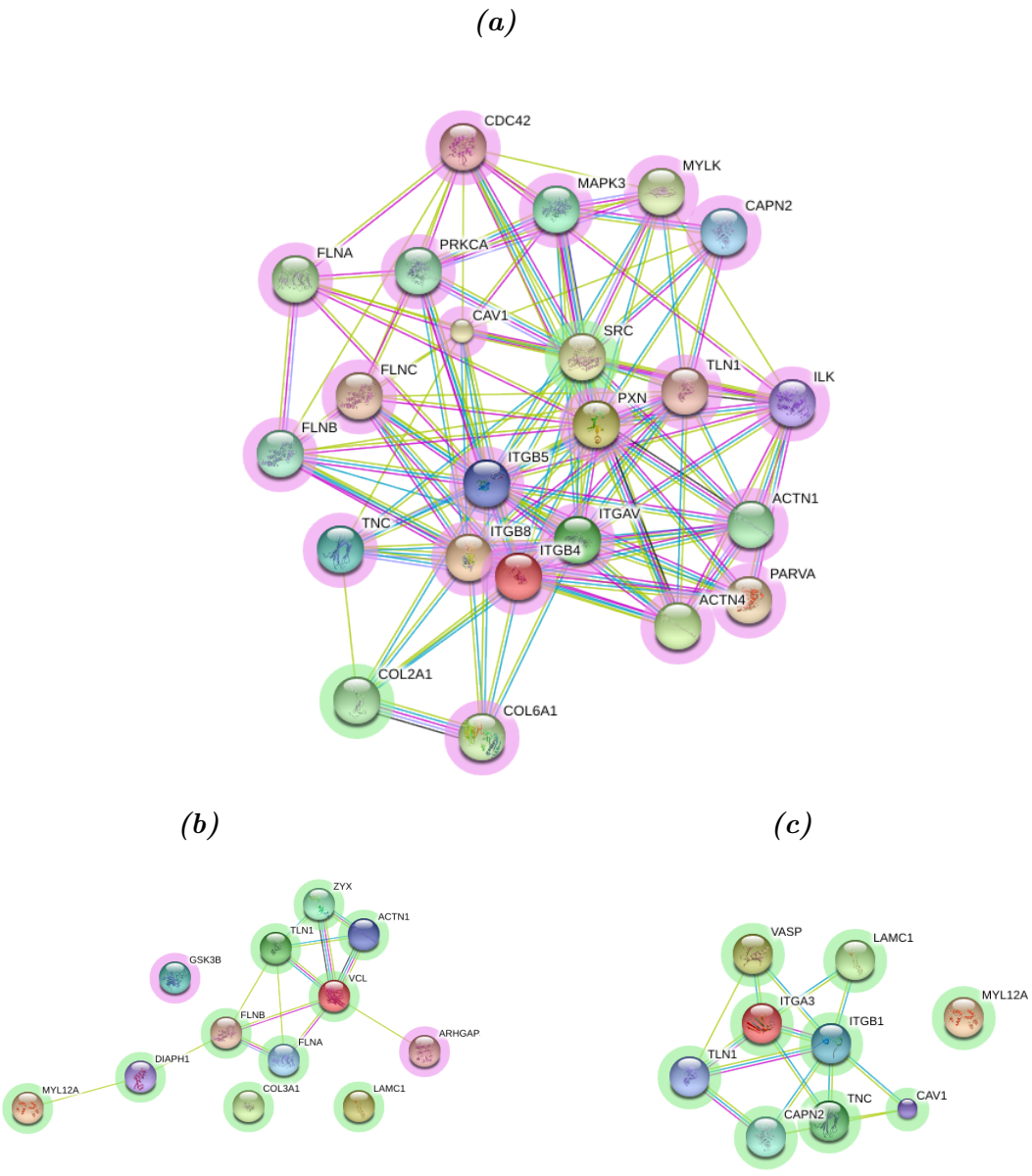


Figure 11.19: *STRINGdb network for KEGG* ***Focal adhesion*** *for (a) the SILAC $\frac{NS}{NES}$ experiment, the (b) $_{hf}NES$ cell differentiation experiment and (c) $_{hf}NS$ cell differentiation experiment. Red and green halos indicate up- and down-regulation respectively.*

Focal adhesion is an adhesive contact between the cell and the ECM, through a large dynamic protein complex. Contact to the ECM is mainly mediated by integrins. Integrins bind laminins, fibronectins, and vibronectins. Scaffold and intercellular signalling molecules, *PXN*, *SRC* *TLN1*, *VCL*, and *ZYX* bind integrins, thereby transferring signalling from the

ECM into the cell, or from the cell to the ECM.[112] Focal adhesion plays a role in cell migration.

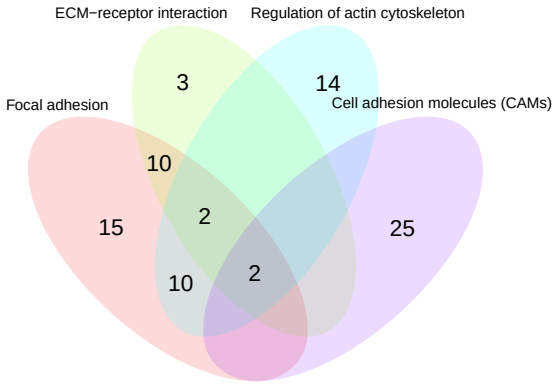


Figure 11.20: *Overlap of the genes and proteins up-regulated in the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experiments that form part of the KEGG enriched terms for cell adhesion and response to ECM interaction*

Many of the proteins forming part of KEGG ***ECM-receptor interaction*** also form part of KEGG ***Focal adhesion*** (Figure 11.20). This term was enriched in both $_{hf}$NES cells and $_{hf}$NS cells in the GE $\frac{NS}{NES}$ experiment, with p-value, of 0.0012 and 8.76×10^{-19}, respectively (Figure 11.15). It was far more highly enriched in $_{hf}$NS cells, in agreement with enrichment in $_{hf}$NS cells in the SILAC $\frac{NS}{NES}$ experiment. ***Focal adhesion*** is also enriched in the down-regulated proteins in both the $_{hf}$NES and $_{hf}$NS cell differentiation experiments, thus indicating that it is important in the pluripotent stages of neural stem cells, more so in $_{hf}$NS cells, but becomes less important after differentiation.

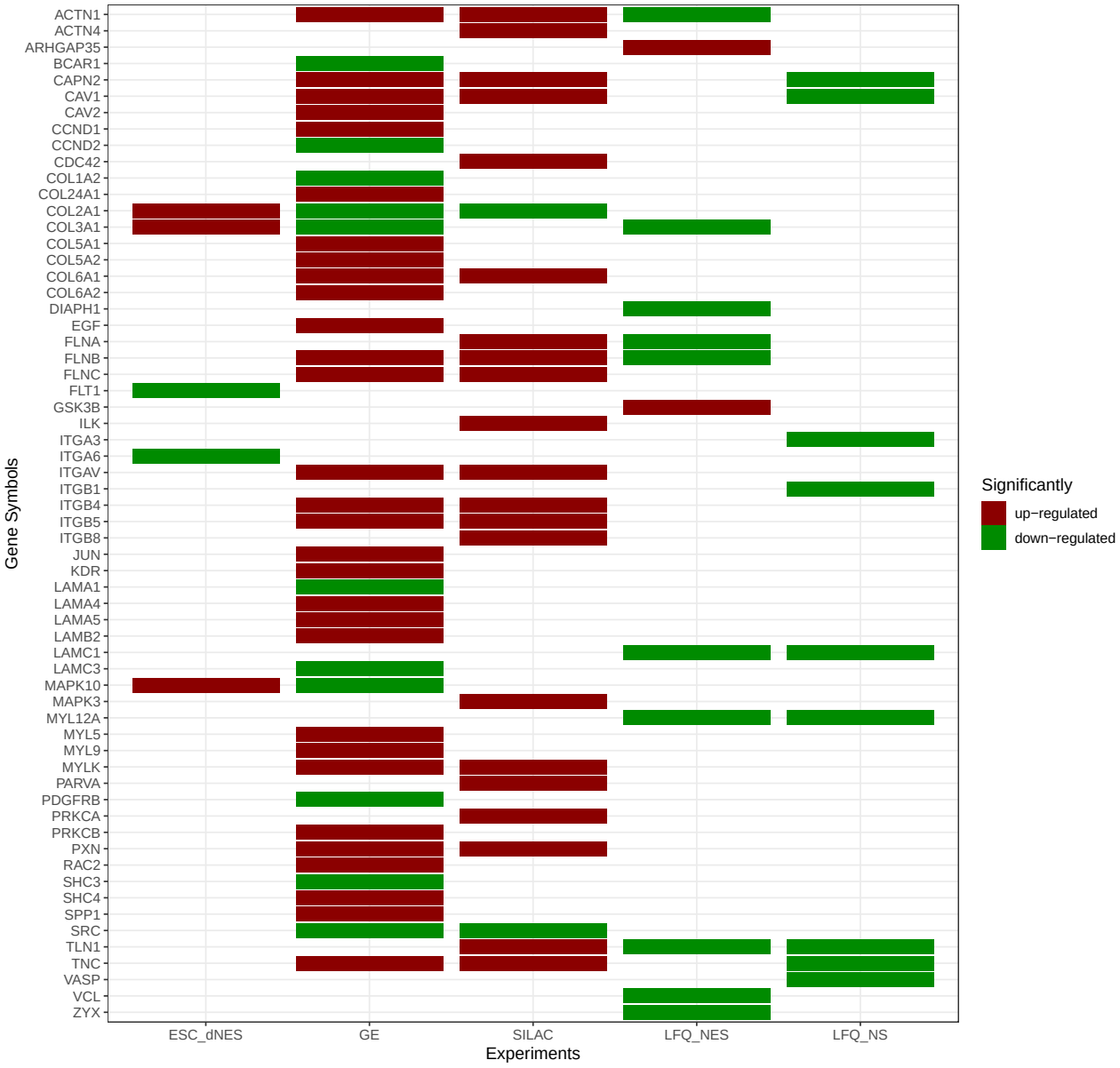


Figure 11.21: *Heatmap of all SDE genes and proteins for KEGG* ***Focal adhesion*** *across all experiments. Red and green indicates up- and down-regulation respectively.*

11.2.3 Cell adhesion molecules

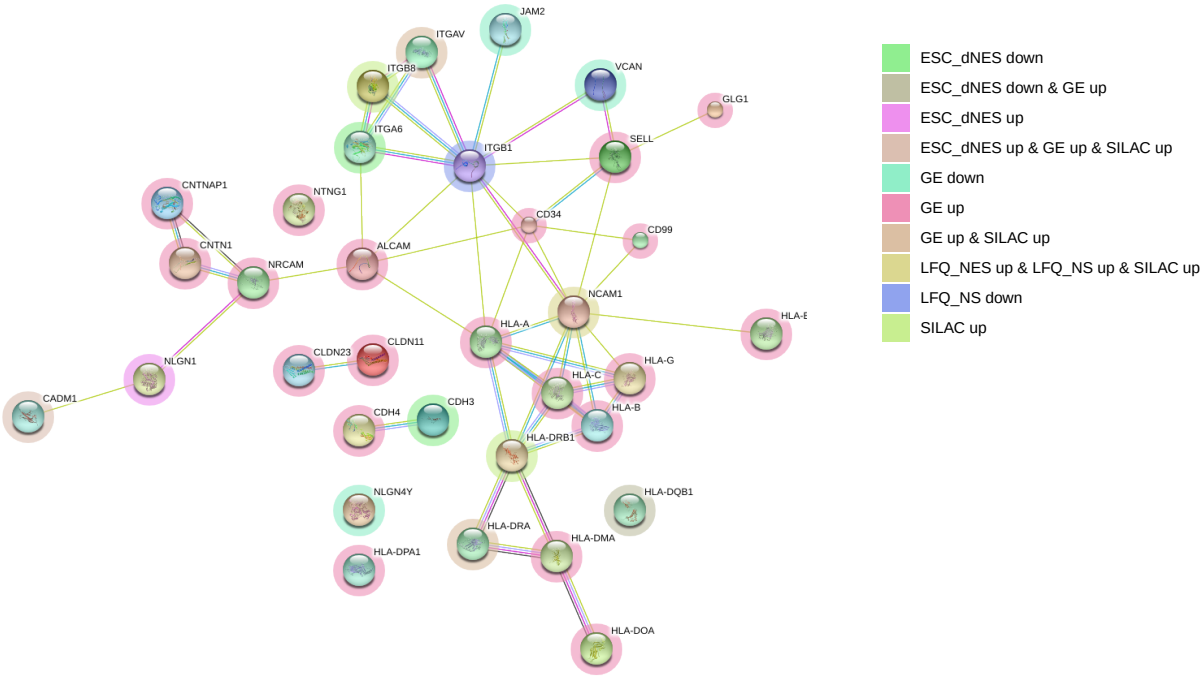


Figure 11.22: *STRINGdb network for KEGG* ***cell adhesion molecules*** *for all experiments. The halo colours indicate the overlaps in SDE between experiments, as per the legend*

Interestingly, the KEGG term ***Focal adhesion*** shares many molecules up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments with the terms ***ECM-receptor interaction*** and ***Regulation of actin cytoskeleton***, although there were many more cell adhesion molecules SDE in $_{\text{hf}}$NS cell not included in ***Focal adhesion*** (Figure 11.20).

Cell adhesion molecules are molecules on the cell surface that allow the cell to bind the ECM or other cells. Cell adhesion is an important part of development as it regulates the structure and formation of the tissues. *CADM1* was upregulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$, GE $\frac{\text{NS}}{\text{NES}}$, and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 11.23). It is critical to neural development and plays many roles in the attachment of neuronal structures, axons, dendritic spines, and synapses. It may also be involved in neuronal migration and axon growth. *NCAM1* was upregulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, and the $_{\text{hf}}$NES and $_{\text{hf}}$NS cell differentiation experiments (Figure 11.23). It is a neural-specific adhesion protein often used as a neuronal marker, involved in nervous system development and neuron-neuron adhesion.

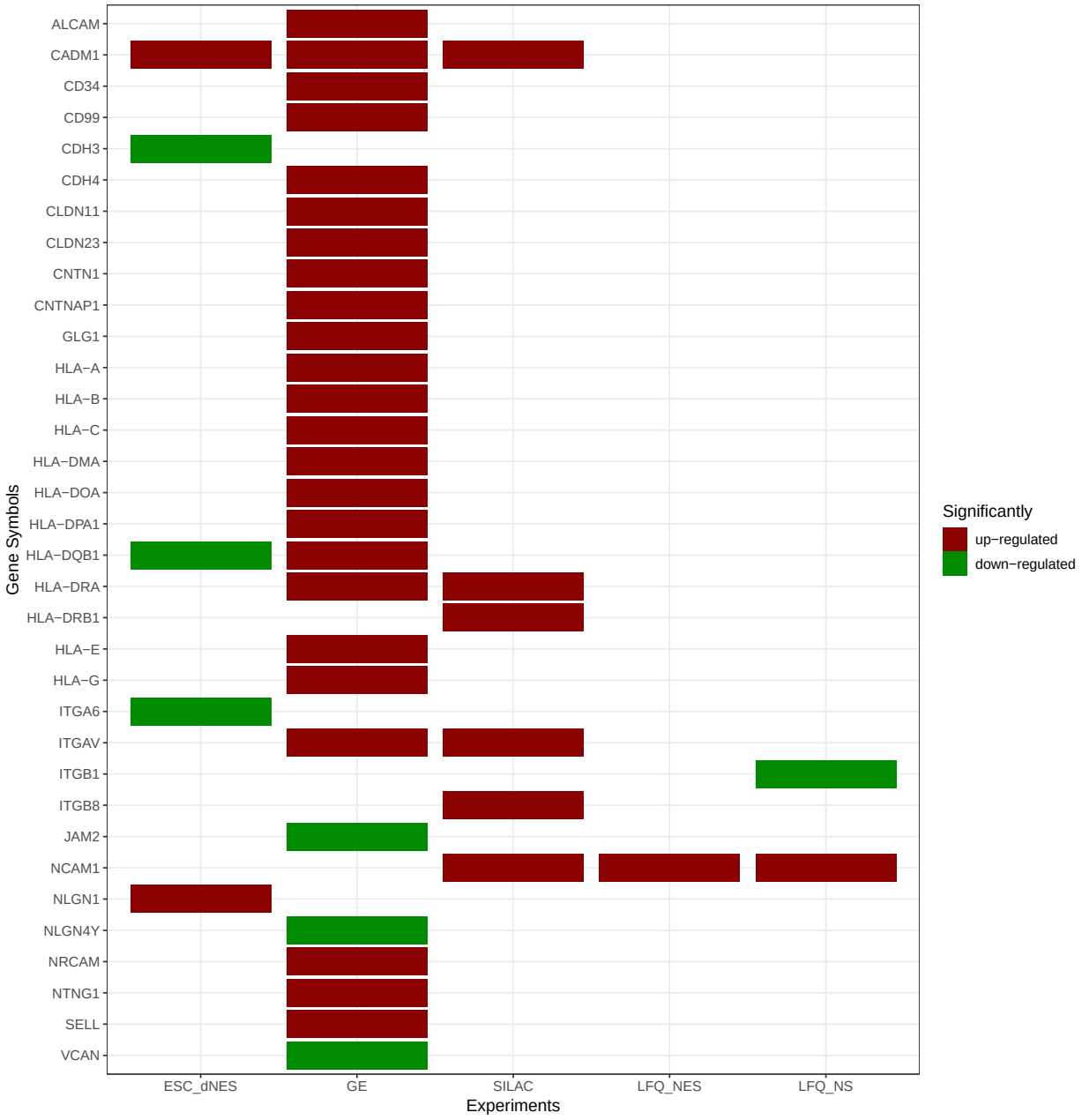


Figure 11.23: *Heatmap of the genes and proteins SDE in all experiments that form part of KEGG* ***Cell adhesion molecules****. Red and green indicates up- and down-regulation respectively.*

11.2.4 Lysosome

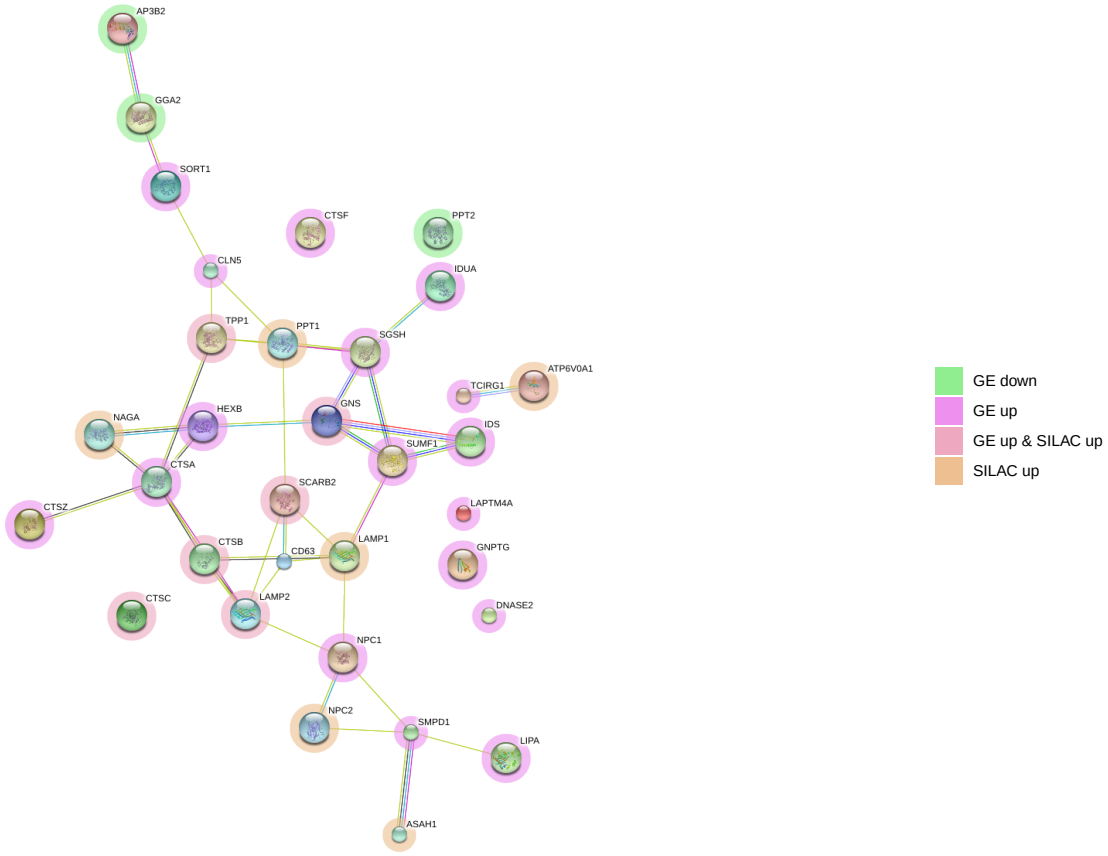


Figure 11.24: *STRINGdb network for KEGG* ***Lysosome*** *for the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experiments. The halo colours indicate the overlaps in SDE between experiments, as per the legend*

The KEGG term ***Lysosome*** was enriched in up-regulated genes and proteins from the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and the $_{\text{hf}}$NS cell differentiation experiment. The lysosome is known to play a role in differentiation by targeting the degradation of signalling molecules, thereby allowing the development of a specific fate. The lysosome may therefore be upregulated in $_{\text{hf}}$NS cells as a means to stabilise their fate determination both as $_{\text{hf}}$NS cells and during $_{\text{hf}}$NS cell differentiation.

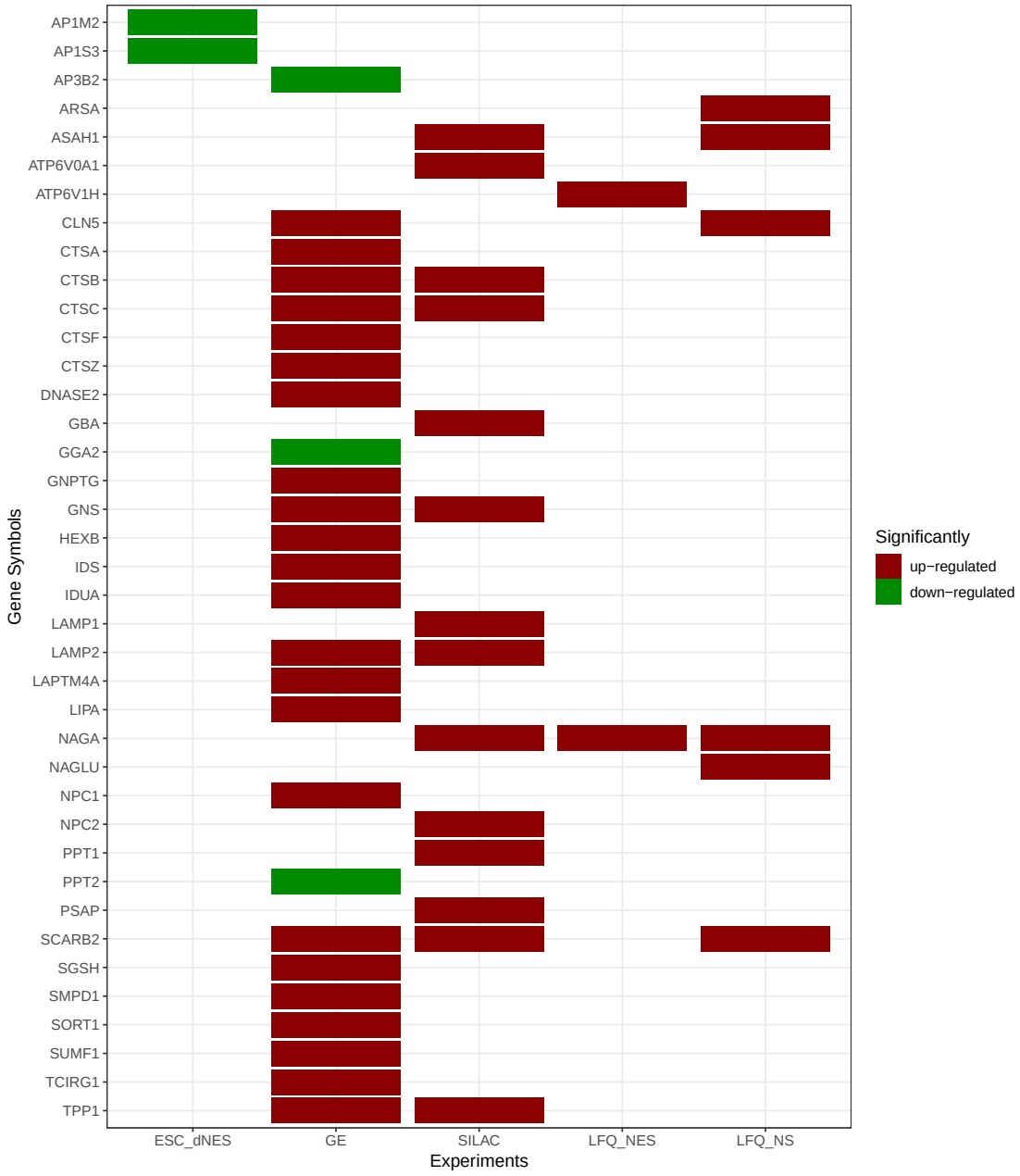


Figure 11.25: *Heatmap for SDE genes and proteins of KEGG* ***Lysosome****. Red and green indicates up- and down-regulation respectively.*

11.3 Signalling pathways

Most of the KEGG signalling pathways significantly enriched in this study are from the up-regulated genes and proteins in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 11.26). This

indicates an increase in complexity of the $_{hf}$NS cells ability to respond to its environment. Thereby enabling these cells to integrate into the developing neuronal tissue and respond to changes as development progresses.

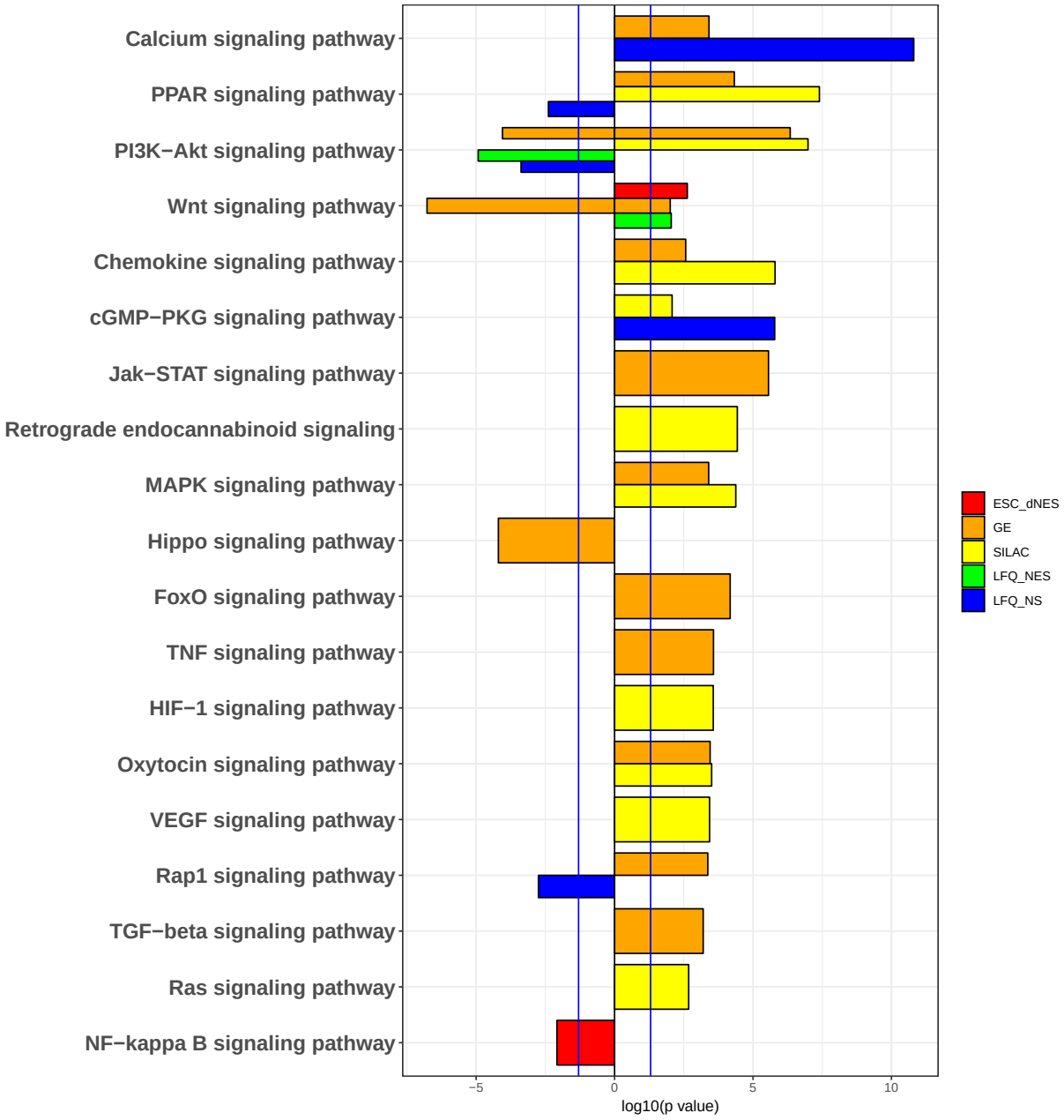


Figure 11.26: *The twenty most highly enriched KEGG Signalling pathways across all the experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

11.3.1 Neural induction

BMP signalling

Repression of BMP signalling is instrumental in the formation of the neural plate. The BMP inhibitors *CHRD*, *FST NOG* and *CER1* prevent the dorsal ectoderm from becoming epidermis, and induce formation of the neural plate. None of the BMP inhibitors were SDE and none of the BMPs proteins were detected. However *BMP1*, *BMP2*, and *GDF10* were up-regulated in GE $\frac{\text{NS}}{\text{NES}}$ experiment whereas *BMP3*, *GDF10*/BMPB3, and *BMP5* were down-regulated in GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 11.27).

The BMPs interact with transmembrane serine/threonine kinase receptors, which transduce the signals into the cells by phosphorylating receptor SMADs (R-SMADs). BMP and their serine/threonine kinase receptors (STKR) are expressed throughout the brain at all stages of development, reviewed in Hegarty et al. 2013. No proteins for BMPs, their receptors or R-SMADs were detected in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment. *BMPR2* and *TGFBR2* were up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. None of the SMAD proteins were detected, and none were differentially expressed in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and GE $\frac{\text{NS}}{\text{NES}}$ experiments.

Figure 11.27: *Heatmap of BMP proteins and their receptors SDE across alll experiments. Red and green indicates up- and down-regulation respectively.*

The SMAD interacting protein *ZEB2* was up-regulated in Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and down-regulated in GE $\frac{\text{NS}}{\text{NES}}$ experiment. It has been shown to induce neural fate by suppressing BMP signalling.[94] In chick embryos *ZEB2* expression is induced by *CHURC1* and induction

of *ZEB2* expression is able to recover phenotype after knock-down of *CHURC1*.[184] *CHURC1* is a downstream effector of FGF signalling and down-regulates early mesodermal genes, while upregulating early neural genes.[119]

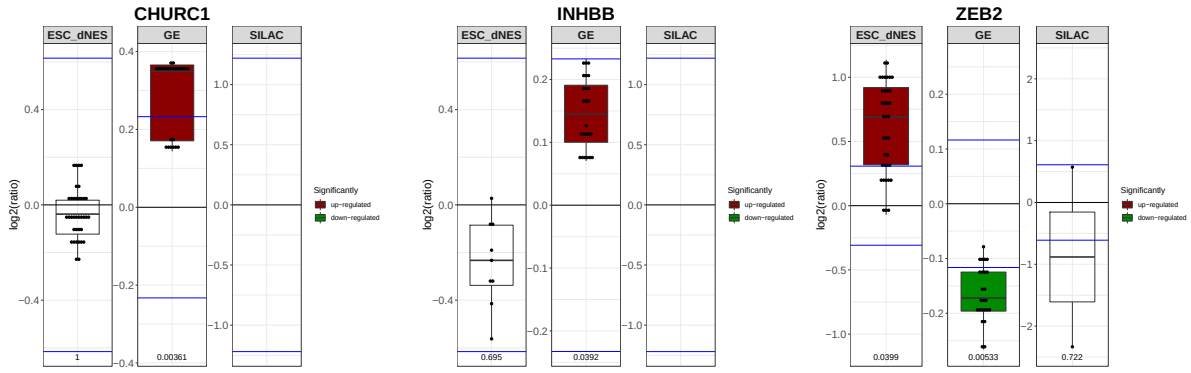


Figure 11.28: *Boxplots comparing the log2 ratios for CHURC1, INHBB and ZEB2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

CHURC1 showed little change in the ESCs but was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. The inverse expression of *ZEB2* and *CHURC1* in the GE $\frac{\text{NS}}{\text{NES}}$ experiment indicates it may be controlled by means other than *CHURC1* induction in human embryos. *CHURC1*'s role in human embryos may be more important in the $_{\text{hf}}$NS cells than the $_{\text{hf}}$NES cells. *ZEB2* does not bind DNA and is therefore hypothesised to mediate this effect through other protein-interactions.[117] Studying its interactors within this cell model may yield interesting information as to its role in foetal brain development.

TGFBR3 was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment and is known to antagonises BMPs,[226] in conjunction with inhibins. The inhibin beta subunit *INHBB* which forms the homodimer activin was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. Activins are known to have neuroprotective effects in the CNS. In the olfactory bulb they have been show to affect *ASCL1* positive progenitors. Activin inhibits proliferation of intermediate neural progenitors, and favours glial over neuronal fate. Both *ASCL1* and *INHBB* were upregulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment so they may be playing a similar roles in $_{\text{hf}}$NS cells.[85]

WNT pathway

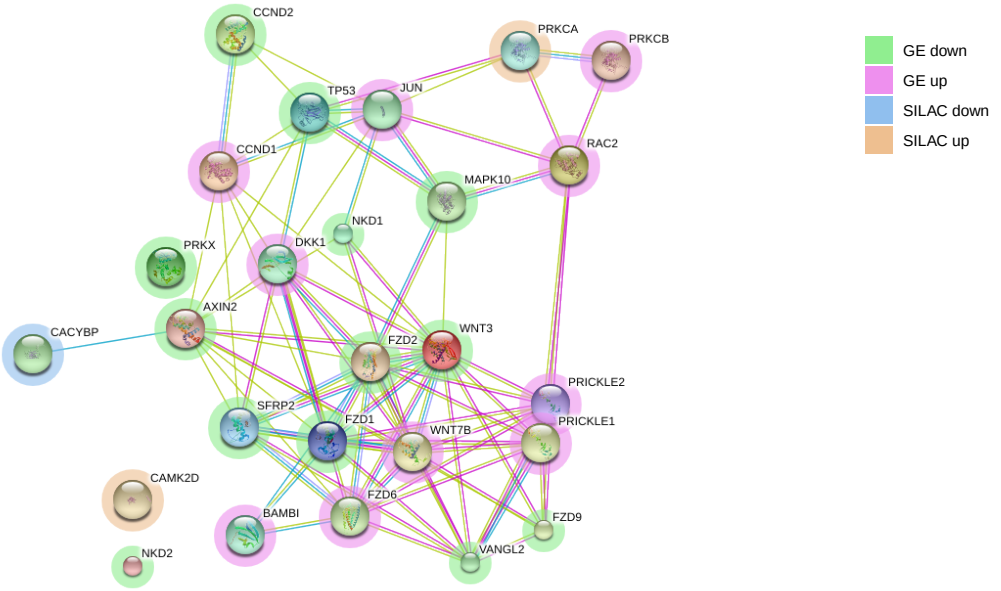


Figure 11.29: *STRINGdb network of KEGG WNT pathway genes and proteins SDE in the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experimentss. the halo colours indicate the overlaps in SDE between experiments, as per the legend*

The proliferating $_{hf}$NES cells and $_{hf}$NS cells were under stimulus of FGF and EGF while proliferating. Upon removal of FGF and EGF stimulation they spontaneously differentiated. There are virtually no FGFs or FGF receptors SDE in this work. The FGF pathway is know to stimulate the wingless-type MMTV integration site family (WNT) pathway. Which was enriched in the down-regulated genes of the GE $\frac{NS}{NES}$ experiment (Figure 11.26), thus potentially more active in the $_{hf}$NES cells. The WNT pathway is known to be involved in cell fate specification, proliferation and migration. *SFRP2* modulates WNT signalling and is involved in many of the pathways and processes enriched in in the $_{hf}$NES cells. It was upregulated in the Huang $\frac{dNES}{hESC}$ experiment and down-regulated in the GE $\frac{NS}{NES}$ experiment (Figure 11.31).

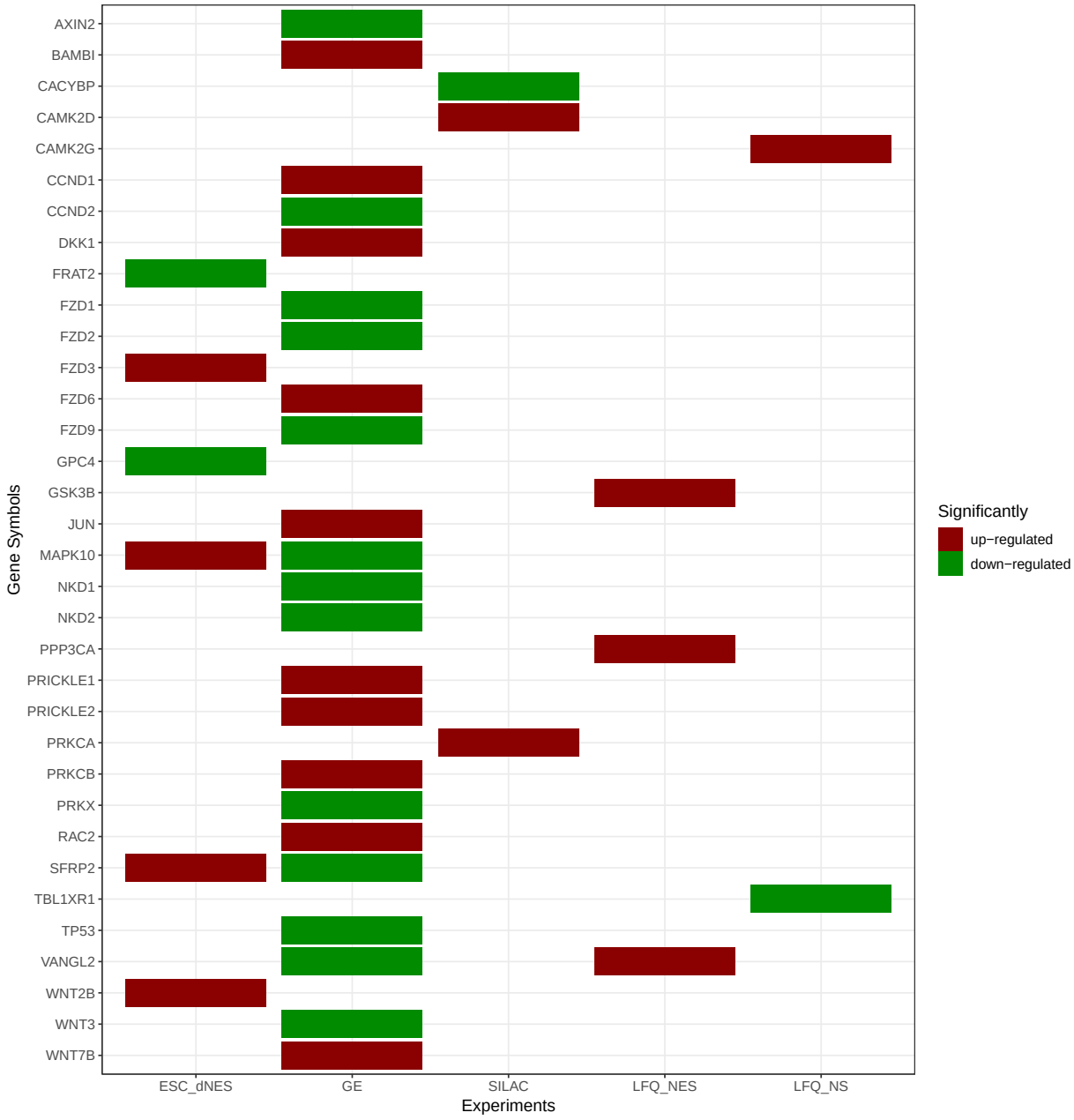


Figure 11.30: *Heatmap of KEGG WNT pathway genes and proteins SDE in the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experimentss. Red and green indicates up- and down-regulation respectively.*

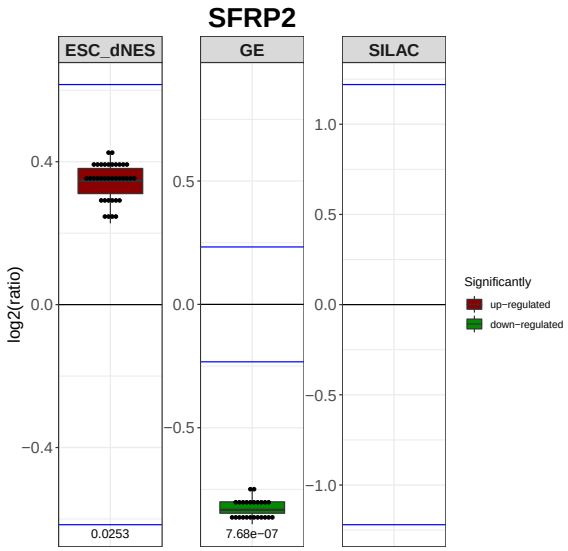


Figure 11.31: *Boxplots comparing the log2 ratios for SFRP2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Hippo signalling pathway

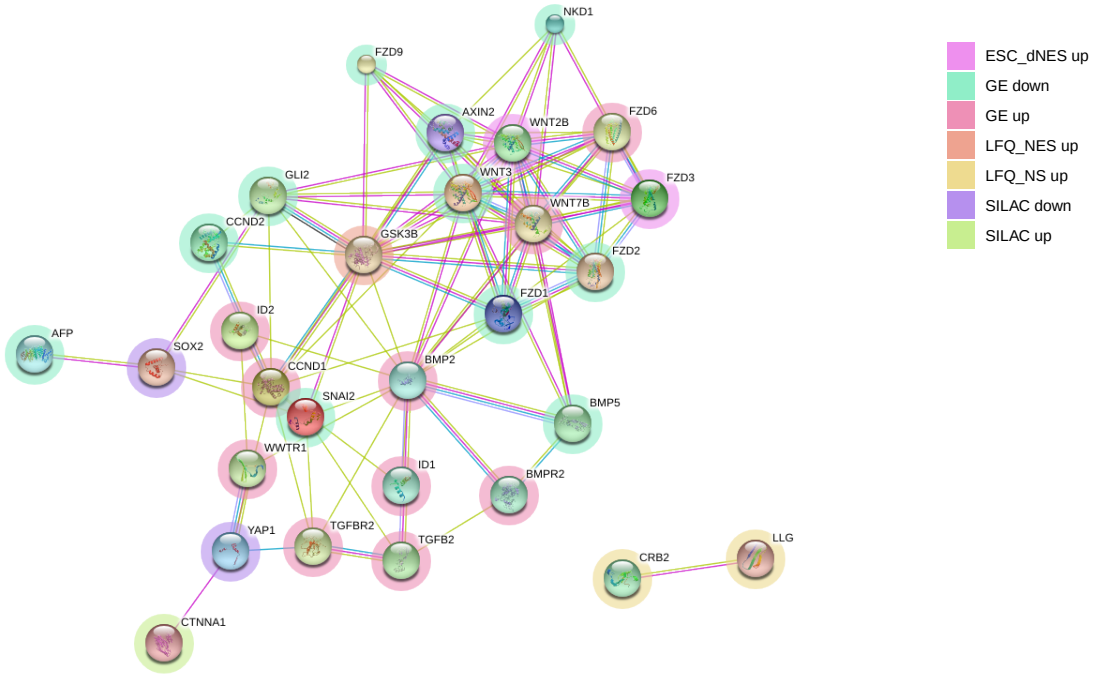


Figure 11.32: *STRINGdb network of KEGG Hippo pathway genes and proteins SDE across all experiments. the halo colours indicate the overlaps in SDE between experiments, as per the legend*

The only other pathway enriched in the down-regulated genes from the GE $\frac{\text{NS}}{\text{NES}}$ experiment is the Hippo signalling pathway. Interestingly it contains many of the effectors molecules that overlap with the WNT pathway and the BMP pathway. It is known to control organ size in animals by regulating the proliferation rate of stem cells during development.[22] The downstream effector of the Hippo pathway is *YAP1*, which is known to be highly expressed in ESCs[121] and neural stem cells.[236,243] Knock-down of *YAP1* in mice results in death before neural differentiation.[243] It is up-regulated by *SOX2*[182] and controls organ size through regulating the proliferation of progenitors. It does this by increasing the expression of CCND1.[236] In this study, *CCND1* and *YAP1* have inverse expression patterns. *YAP1* inhibts the WNT pathway, thereby maintaining stemness,[182] suggesting that the WNT pathways and Hippo pathways are working in opposition in the $_{\text{hf}}$NES cells, the antagonism of the WNT pathway potentially

becomes lifted in the $_{hf}$NS cells allowing WNT to become more influential (Figure 11.26).

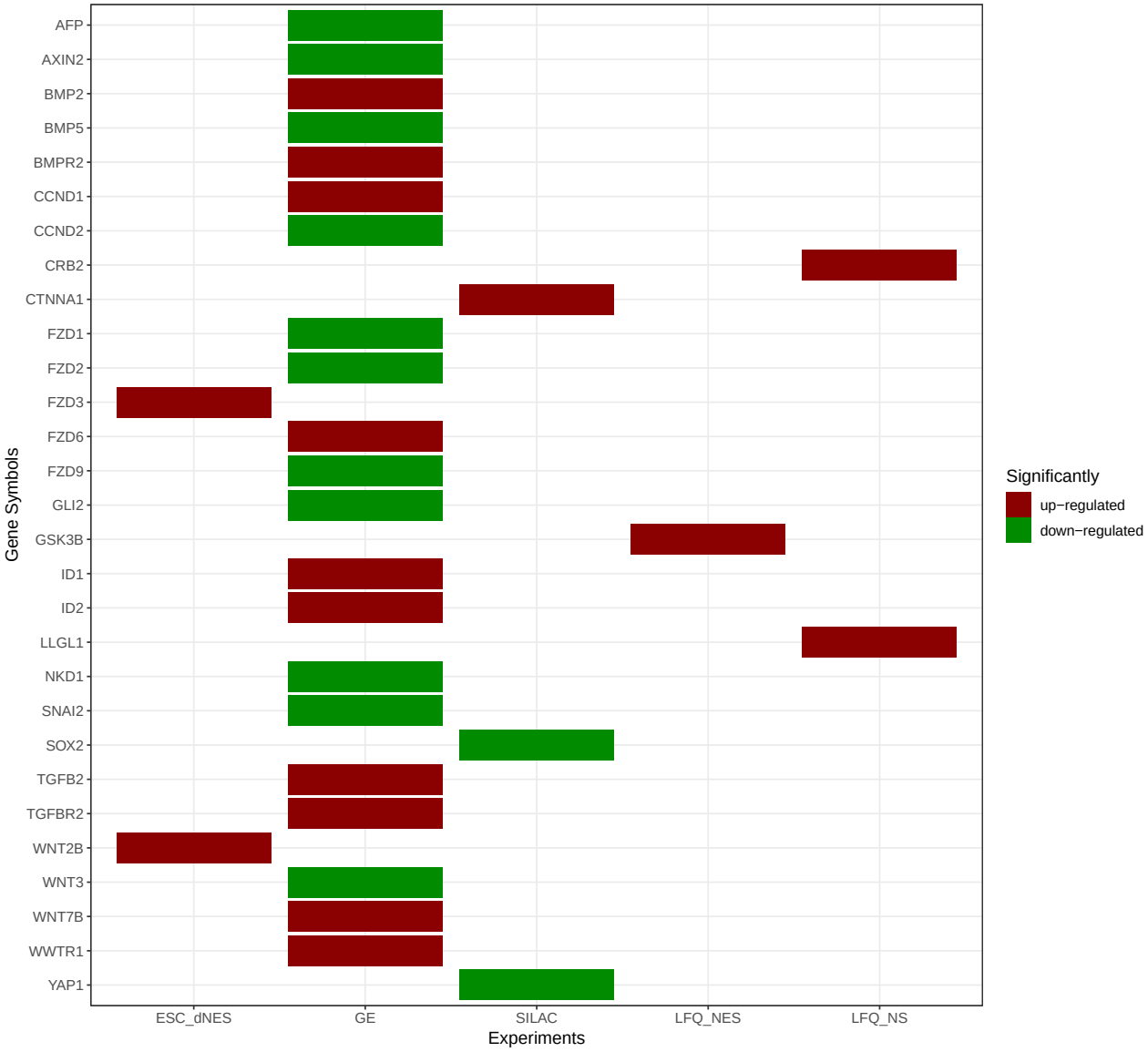


Figure 11.33: *Heatmap of KEGG Hippo pathway genes and proteins SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

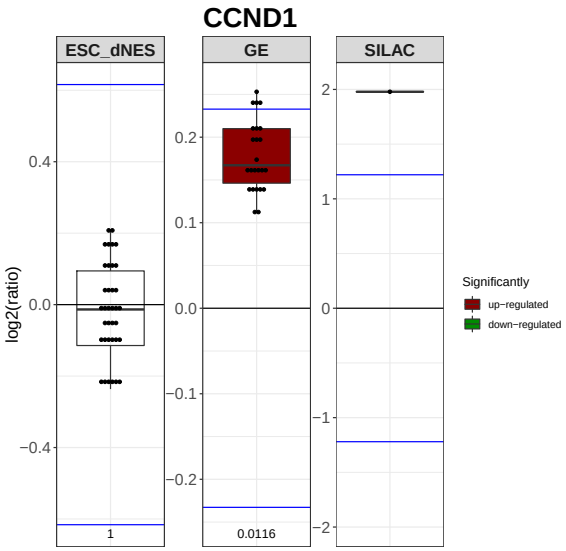


Figure 11.34: *Boxplots comparing the log2 ratios for CCND1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

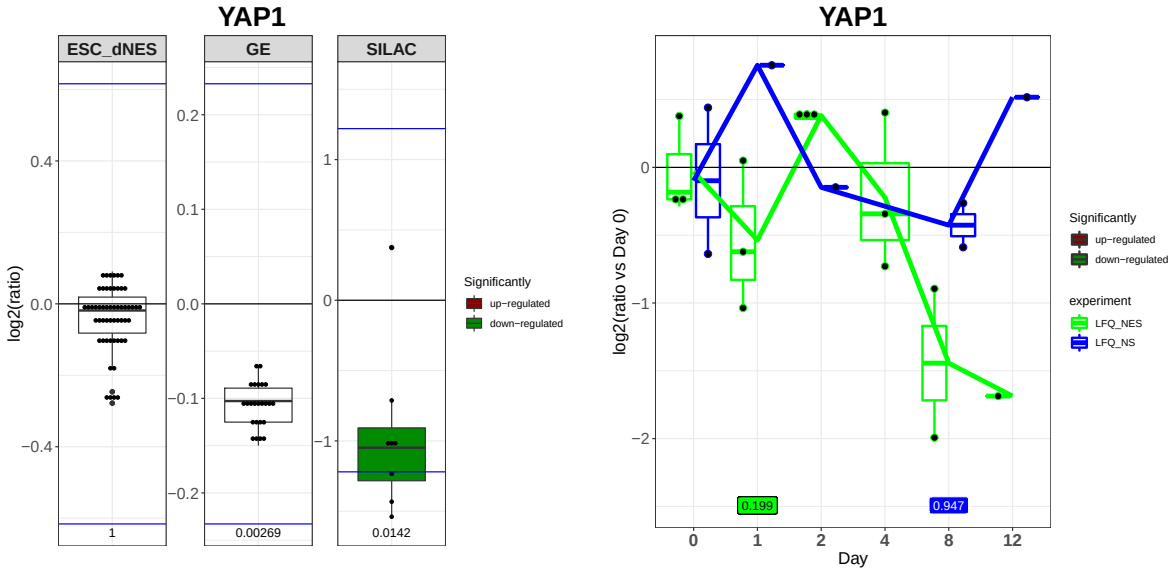


Figure 11.35: *Boxplots comparing the log2 ratios for YAP1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

11.4 Metabolism

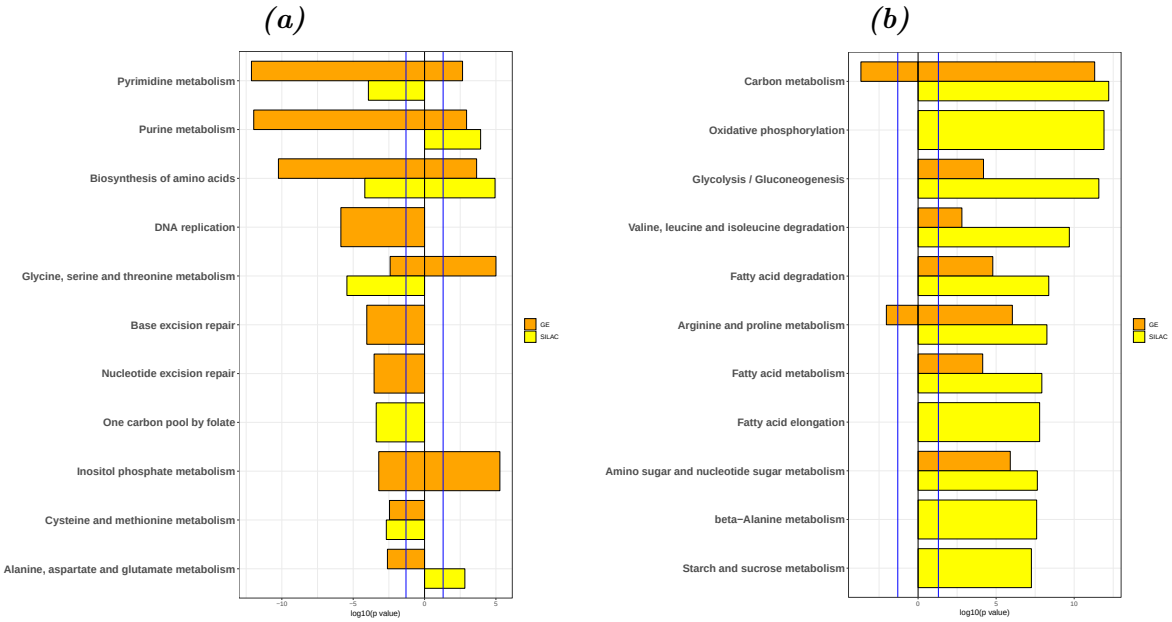


Figure 11.36: *The twenty mostly highly enriched KEGG metabolic pathways for the (a) down-regulated or (b) up-regulated genes and proteins in the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

There were significant changes in the enriched metabolic pathways between the genes and protein SDE in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 11.36). The $_{\text{hf}}$NES cells were more highly enriched for pathways involved in the anabolic processes ***pyrimidine and purine metabolism*** and ***biosynthesis of amino acids***. This indicates that the cells may be producing the building blocks required for cell division and proliferation. The $_{\text{hf}}$NS cells were more highly enriched for catabolic process, perhaps in preparation for cells that are establishing and playing defined roles within the developing neural tissue.

Chapter 12

Gene ontology enrichment

All SDE genes and proteins from the experiments, Huang $\frac{\text{dNES}}{\text{hESC}}$, GE $\frac{\text{NS}}{\text{NES}}$, and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and the $_{\text{hf}}$NES and $_{\text{hf}}$NS cell differentiation experiments were enriched using topGO. The SDE genes and proteins from each experiment was enriched relative to that of all the genes or proteins detected in that experiment, and the p-values were corrected by the weight01. A p-value theshold of 0.01 was used for all enriched terms. Although traditional enrichment multiple-testing correction of GO enrichment removes some of the GO terms enriched. the weight01 algorithm highlights deeper levels of enrichment, by removing the terms higher up in the hierarchy. The deeper levels are more descriptive and the corresponding STRINGdb networks are smaller and therefore easily visualised and annotated.

In the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, down- and up-regulated genes represent the hESCs, and dNES cells respectively, while in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments the down- and up-regulated genes and proteins represent the $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells, respectively. The enrichment p-values have been transformed by log10, down-regulated genes have negative log10(p-value) enrichment scores, whereas up-regulated genes have positive log10(p-values).

12.1 Cellular component

The CCs enriched in $_{\text{hf}}$NES cells were mostly components of the ***nucleus***, whereas those enriched in $_{\text{hf}}$NS cells included components on the periphery of the cell (Figure 12.1).

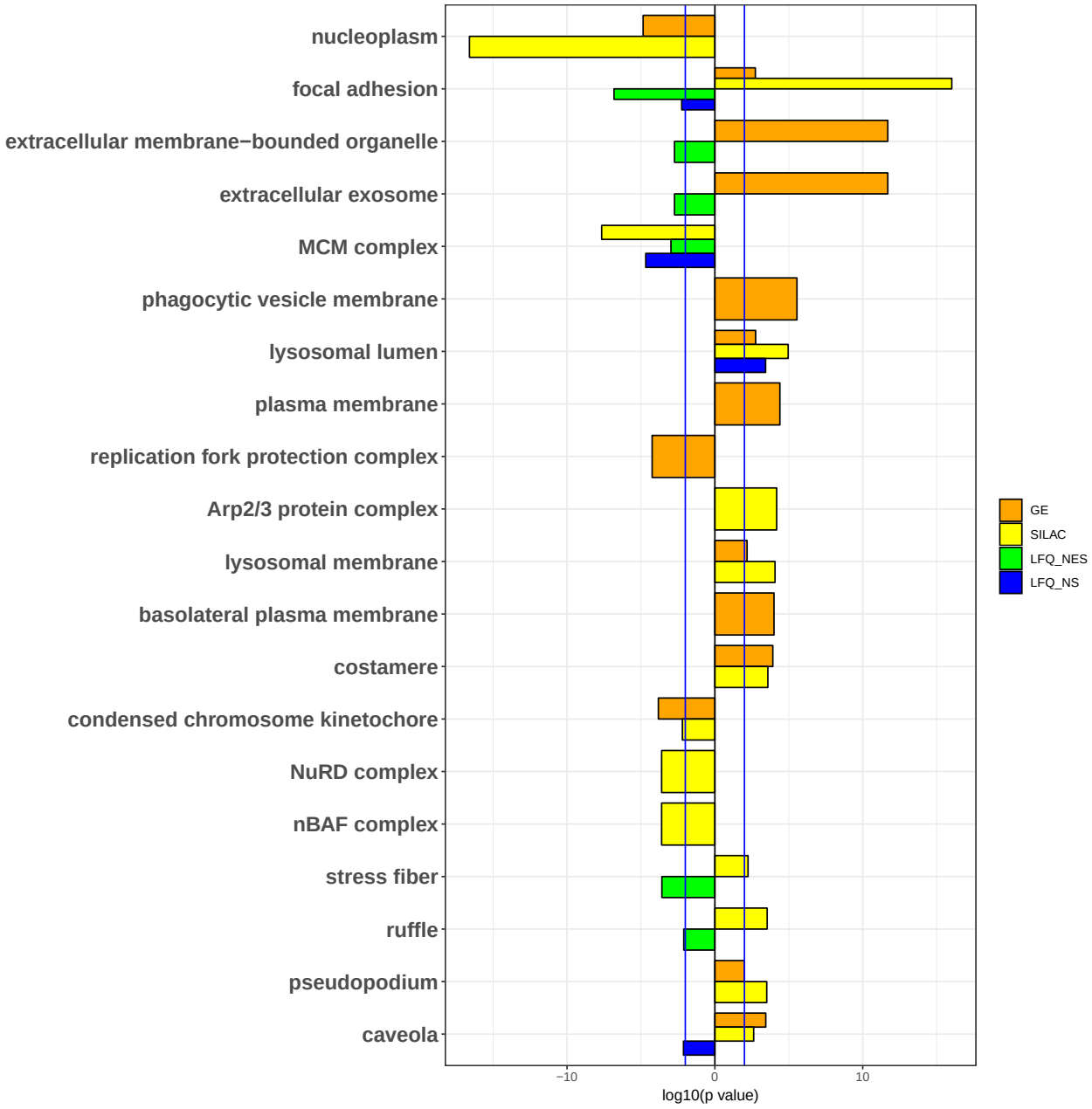


Figure 12.1: *The top twenty most highly enriched cellular component terms enrichments for all experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the $_{hf}$NES cell differentiation experiment, and LFQ_NS (blue) refers to the $_{hf}$NS cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

12.1.1 Cellular components enriched in NES cells

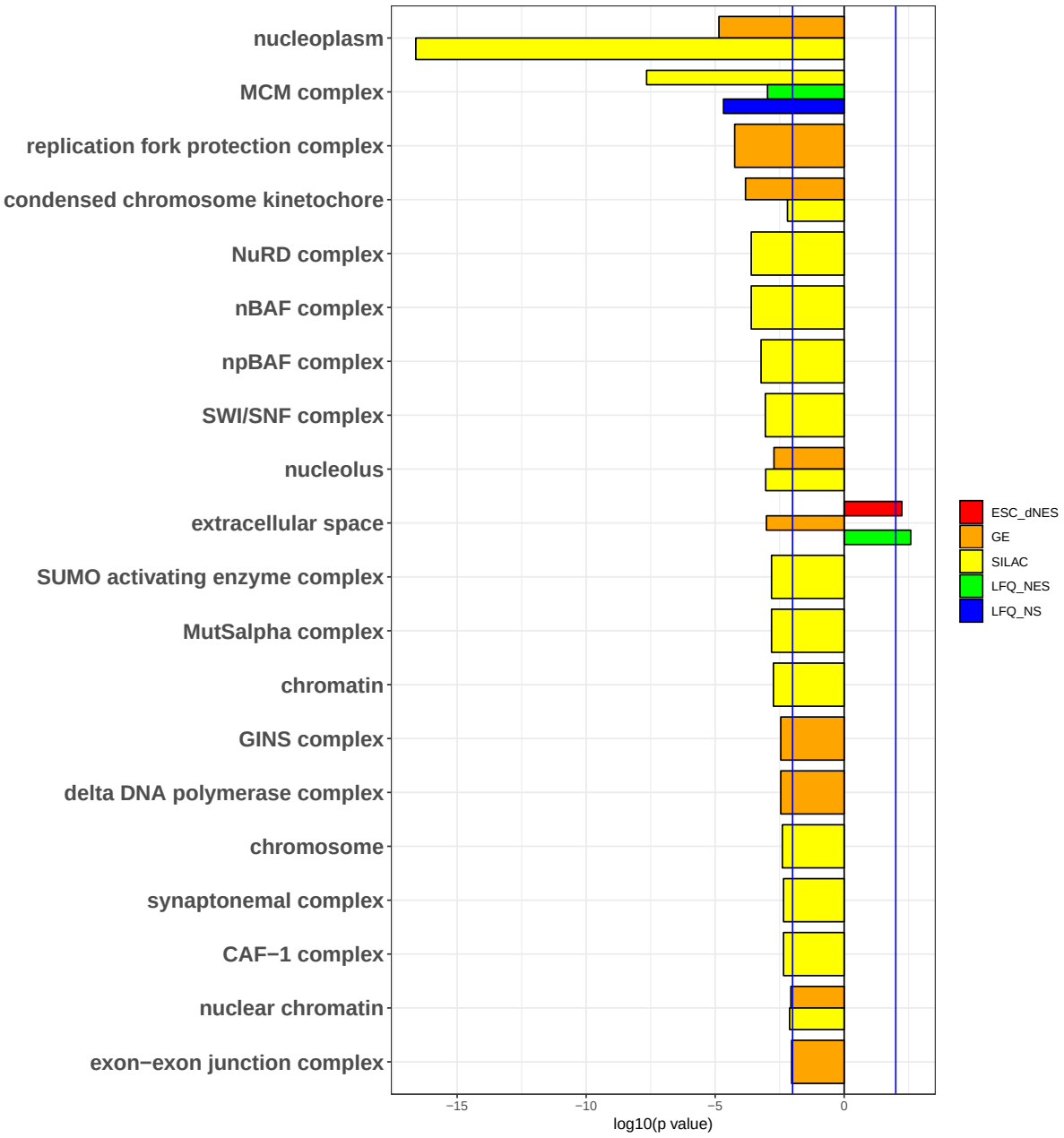


Figure 12.2: *The top twenty most highly enriched CCs terms in the down-regulated genes and proteins in the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

Many of the proteins down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment were parts of the nucleus. The CC terms ***chromosome***, ***chromatin nucleolus***, and the ***nucleoplasm*** were all enriched in the down-regulated proteins from the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 12.2).

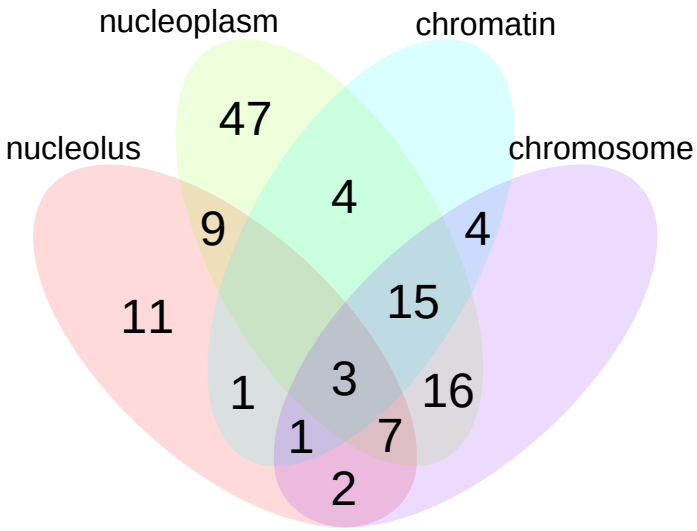


Figure 12.3: *Overlapping proteins down-regulated in the SILAC $\frac{NS}{NES}$ experiment from enriched components of the nucleus*

Nucleoplasm

The nucleoplasm, which is the cytoplasm of the nucleus, was enriched in down-regulated genes and proteins in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, with 179 genes and 100 proteins SDE with p-values of 1.4×10^{-5} and 2.7×10^{-17} respectively. The STRINGdb network for ***nucleoplasm*** was too large to generate a meaningful plots. Most of proteins from the remaining enriched terms formed part of the ***nucleoplasm***.

Nucleolus

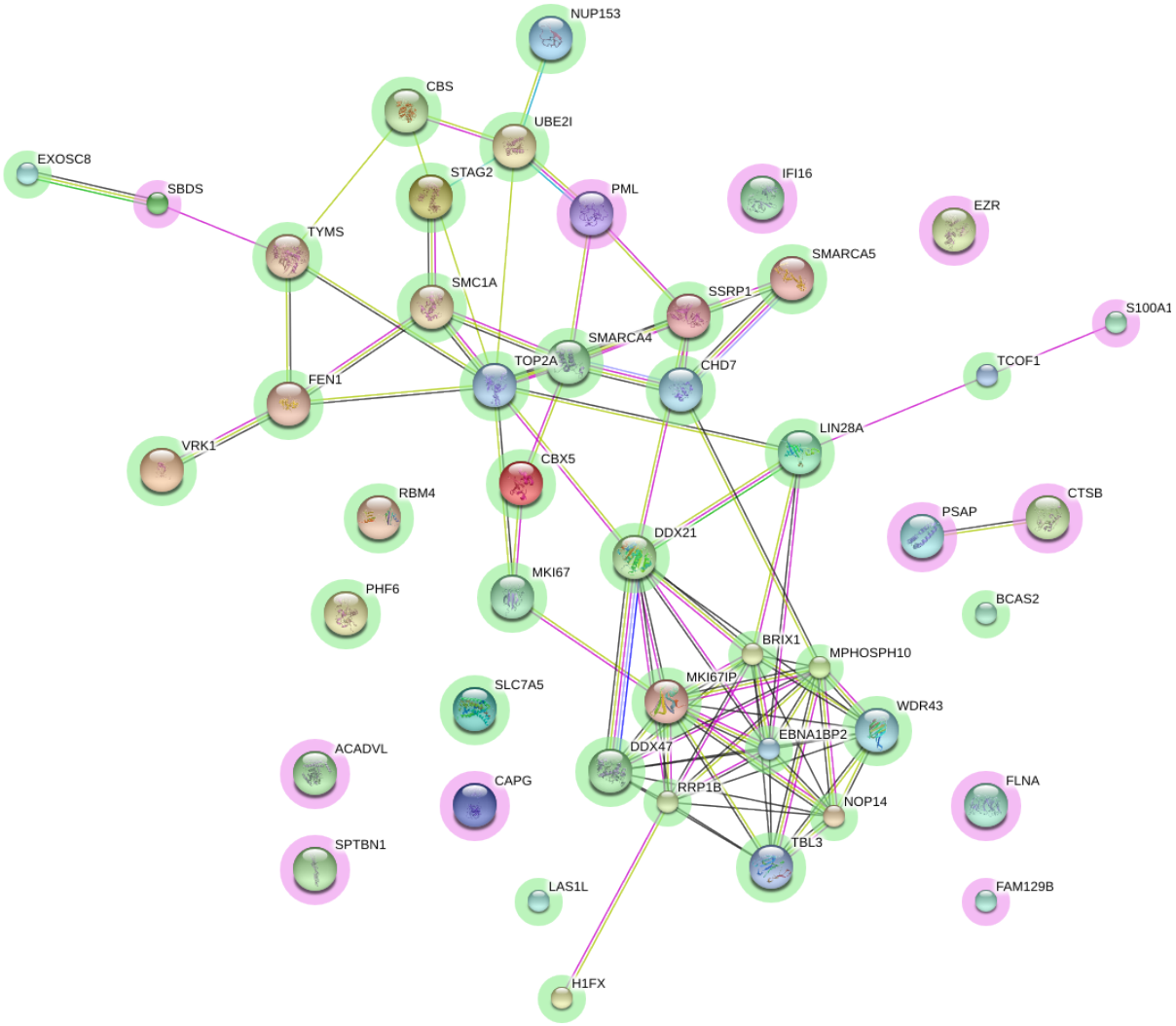


Figure 12.4: *CC* ***Nucleolus*** *proteins SDE in the SILAC* $\frac{NS}{NES}$ *experiment, Red and green halos indicate up- and down-regulation respectively.*

The ***nucleolus*** was enriched in down-regulated genes and proteins in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, with 56 genes and 28 proteins respectively. The nucleolus is the site of ribosome biosynthesis. In studies of human embryos,[154] cells close to the ventricular surface were shown to have a larger number of nucleoli. A second increase in nucleoli occurred in the SVZ, the second germinal layer. The enrichment of nucleolar proteins within $_{\text{hf}}$NES cells corresponds well with the increased numbers of nucleoli observed *in vivo*.

Chromosome

The chromatin condenses to form chromosomes during cell replication. The down-regulated proteins from the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment are enriched for CCs ***chromatin*** and ***chromosomes***. These components are all essential in advancing the cell cycle and allowing rapid proliferation of these cells.

MCM complex

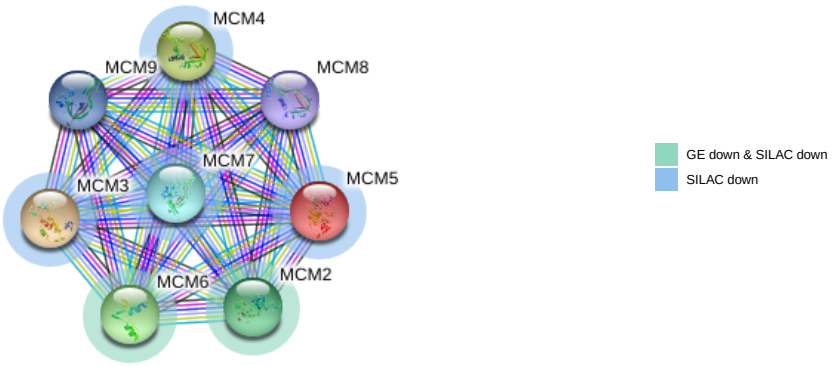


Figure 12.5: *STRINGdb network of the CC* ***MCM complex*** *for the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experiments. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

The MCM complex is a set of eight DNA helicase proteins, required for DNA replication. Six of the eight CC ***MCM complex*** proteins were down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 12.5). The down-regulated protein enrichment p-value for the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment was 2.2×10^{-8}. The complex forms a STRINGdb network clique, with all the proteins being connected to each other (Figure 12.5). *In vivo*, they form a ring structure, which is a component of the pre-replication complex. *MCM2* and *MCM6* were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. The other three MCM complex proteins were only down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, indicating possible post-transcriptional regulation. The ***MCM complex*** is down-regulated as the cells differentiate, and is no longer required when the cells stop proliferating (Figure 12.2 and 12.6).

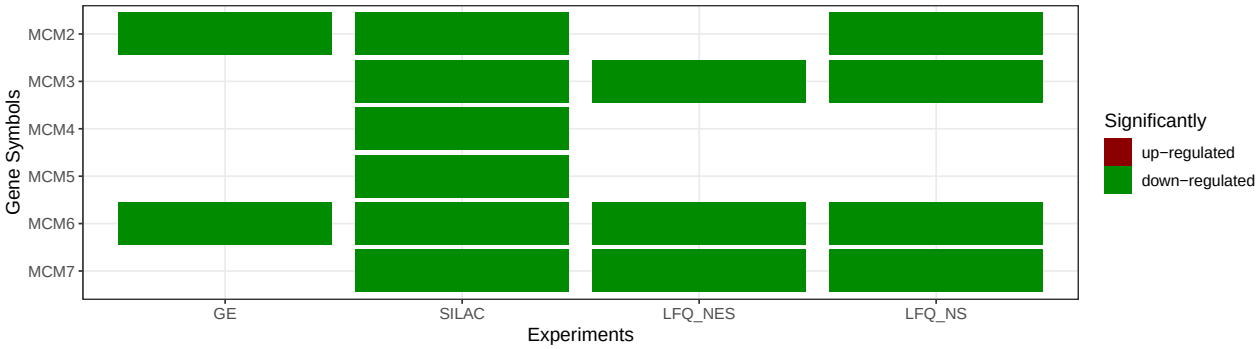


Figure 12.6: *Heatmap of the CC* ***MCM complex*** *for all the experiments. Red and green indicates up- and down-regulation respectively.*

SWI/SNF superfamily type complex

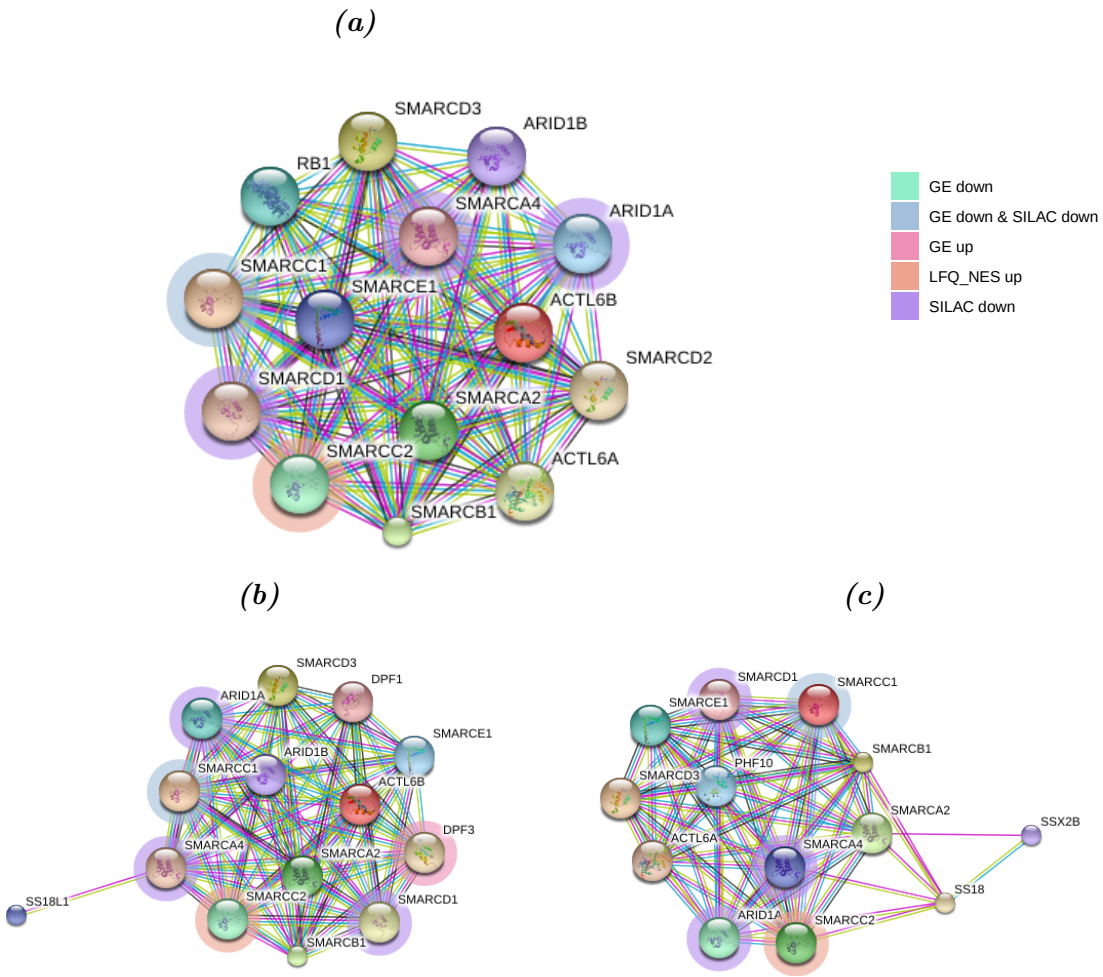


Figure 12.7: *STRINGdb networks for the all genes and proteins from the CC (a)* ***SWI/SNF complex****, (b)* ***nBAF****, and (c)* ***npBAF****. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

The GO terms CC ***SWI/SNF complex***, ***npBAF complex***, and ***nBAF*** are significantly enriched in down-regulated proteins of the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, with p-values of 0.00088, 0.0006, and 0.00025, respectively (Figure 12.2). ***npBAF*** and ***nBAF*** share the same enriched SDE proteins, besides *DPF3*, which was not present in ***nBAF*** (Figure 12.7). As described below, these proteins also form part of the embryonic stem cell BAF (esBAF), for which there is currently no GO annotation. The proteins enriched in the ***npBAF complex*** are involved in the BPs of ***chromatin-mediated maintenance of transcription***, ***nucleosome disassembly***, and ***neural retinal development***. They have MFs related to ***transcription***

co-activation.

The ***SWI/SNF complex*** consists of a number of switching/sucrose non-fermenting subclass of TrxG proteins (SWI/SNF) proteins, also called BRG/BRM-associated factors (BAF) proteins that form a diverse array of complexes. They are ATPase chromatin-remodelling complexes that increase nucleosome mobility by altering DNA-nucleosome contacts via adenosine triphosphate (ATP) hydrolysis, thereby changing the accessibility of DNA to transcription factors. The composition of the complexes specifies the effects on chromatin remodelling. BAF complexes bind close to the transcription start sites (TSSs) of most highly expressed genes.[98] In yeast, the effect is gene activation, but in vertebrates BAF can also repress genes. In higher mammals, BAF has become specialised to control the chromatin state during neurogenesis. *In vivo*, there is a transition from esBAF to neural progenitor BAF (npBAF) and finally neuron BAF (nBAF) during neuronal development.[97] Some of the critical subunits play non-redundant and dosage-sensitive roles in each of these complexes.[120]

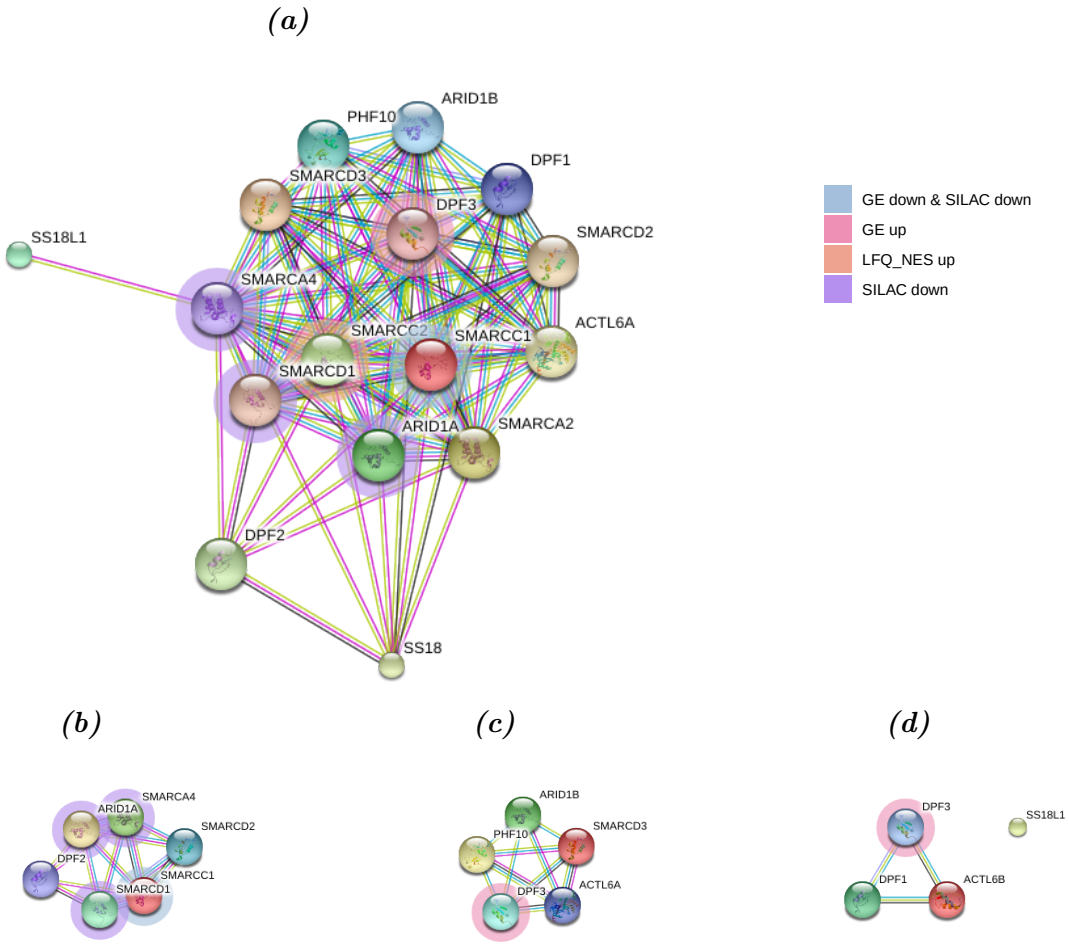


Figure 12.8: *STRINGdb network of BAF genes and proteins characterising specific BAF complexes. (a) All BAF complex molecules, molecules characterising (b) esBAF molecules, (c) nBAF, and (d) npBAF. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

SMARCA4 and *SMARCA2* are used interchangeably with ten other subunits that usually make up a BAF complex. BAF subunits are tightly associated and are not displaced *in vitro*.[120]

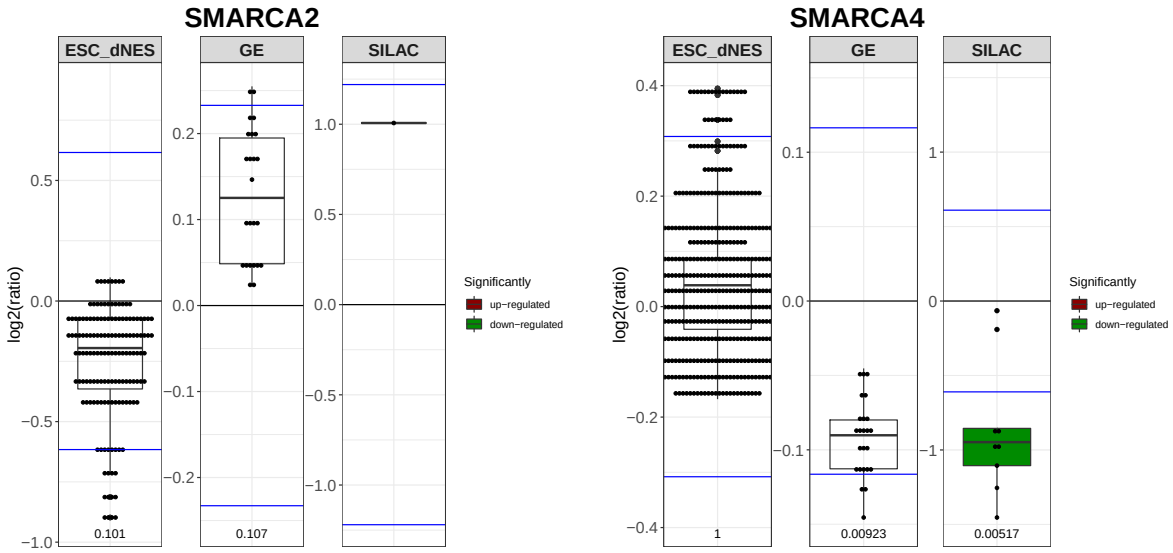


Figure 12.9: *Boxplots comparing the log2 ratios for SMARCA2 and SMARCA4. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

The specific set of subunits *SMARCA4*, *ARID1A*, *SMARCC1*, and *SMARCD2* were previously shown to be expressed in pluripotent cells as part of the esBAF. They are vital to maintaining pluripotency,[97] and showed little difference during differentiation from ESC to dNES cells in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. They were however down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 12.9 and 12.10).

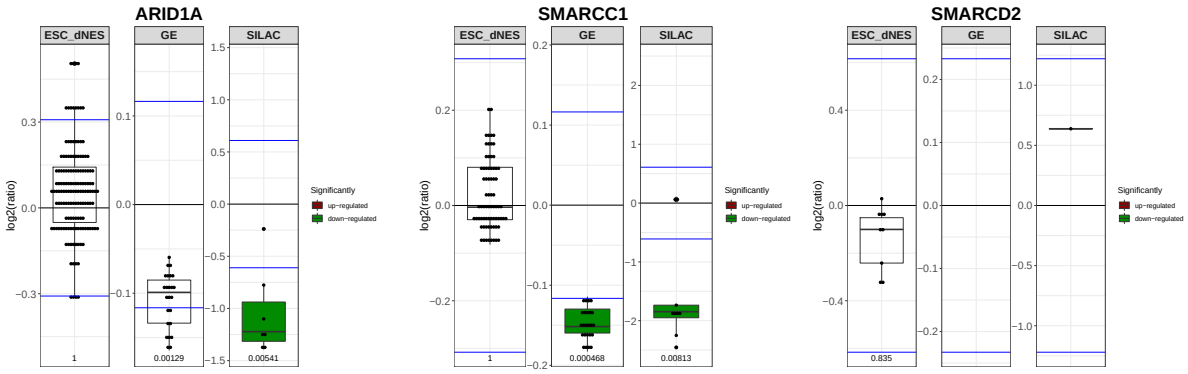


Figure 12.10: *Boxplots comparing the log2 ratios for ARID1A, SMARCC1 and SMARCD2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

esBAF is also characterised by the absence of *SMARCA2*, *ARID1B*, *SMARCC2*, and *SMARCD3*.[97] Of the proteins absent in esBAF, *SMARCA2* was the only one showing up-regulation in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, although this was not statistically significant (Figure 12.9). The SILAC $\frac{\text{NS}}{\text{NES}}$ experiment corroborates the GE $\frac{\text{NS}}{\text{NES}}$ experiment trend .

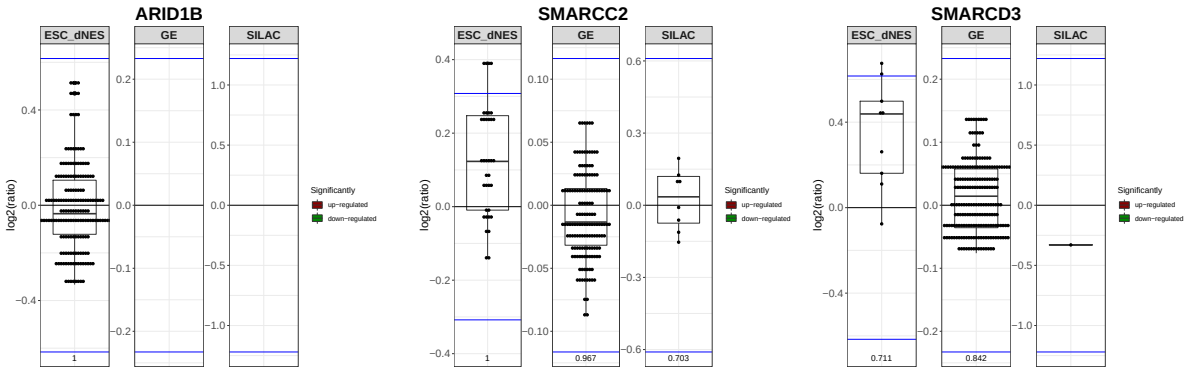


Figure 12.11: *Boxplots comparing the log2 ratios for ARID1B, SMARCC2 and SMARCD3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

As pluripotent cells differentiate into neural stem cells *in vitro*, *SMARCA2*, *ARID1B*, *SMARCD3*, and *DPF3* first appear as npBAF subunits. Here, *DPF3* was significantly up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 12.12), while *SMARCA2* showed a non-significant upwards trend (Figure 12.9).

In vivo, *SMARCC1* in esBAF is replaced by *SMARCC2* in npBAF.[199] There were no significant changes in any of the proliferating datasets for *SMARCC2*, even in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment where *SMARCC2* should be virtually absent.[97] *SMARCC1* was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 12.10). This decrease in *SMARCC1* could result in more npBAF relative to esBAF in $_{\text{hf}}$NS cells. npBAF also contains *DPF1* and *DPF2*, neither of which were SDE in this work.[97] In addition, *SMARCD1*, which is present in esBAF but absent in npBAF, is down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 12.12).

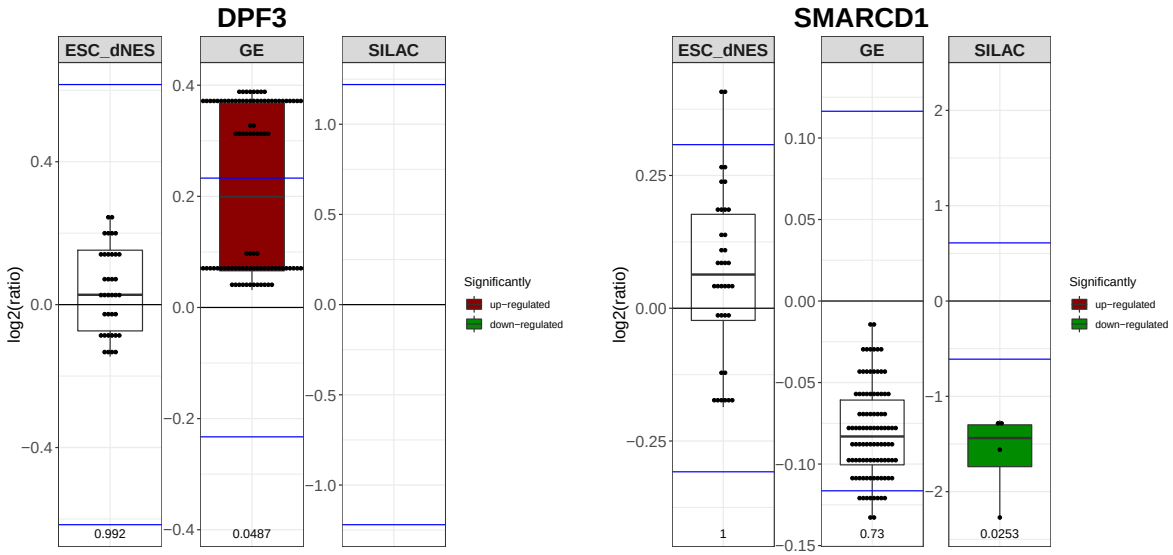


Figure 12.12: *Boxplots comparing the log2 ratios for DPF3 and SMARCD1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

In vivo, there is a rapid and specific switch from npBAF to nBAF during the neural progenitor's last mitotic cycle. This occurs through a sudden reduction in *ACTL6A*, *PHF10*, and *SS18*, and production of *ACTL6B*, *DPF1*, *DPF3*, and *SS18L1*.[120,193,198,199] The npBAF proteins are switched off and the neuronal daughter cells express the nBAF proteins. Here, *ACTL6A* and *PHF10* which are both known to be part of esBAF and npBAF showed no significant downwards trends in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, but were either not detected or unchanged during differentiation (Figure 12.13). *DPF3* which is known to first appear in npBAF was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 12.12).

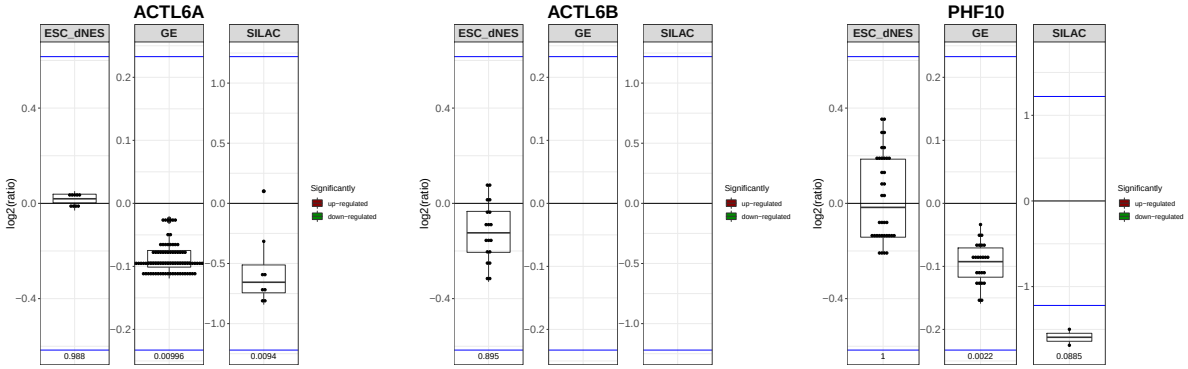


Figure 12.13: *Boxplots comparing the log2 ratios for ACTL6A, ACTL6B and PHF10. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

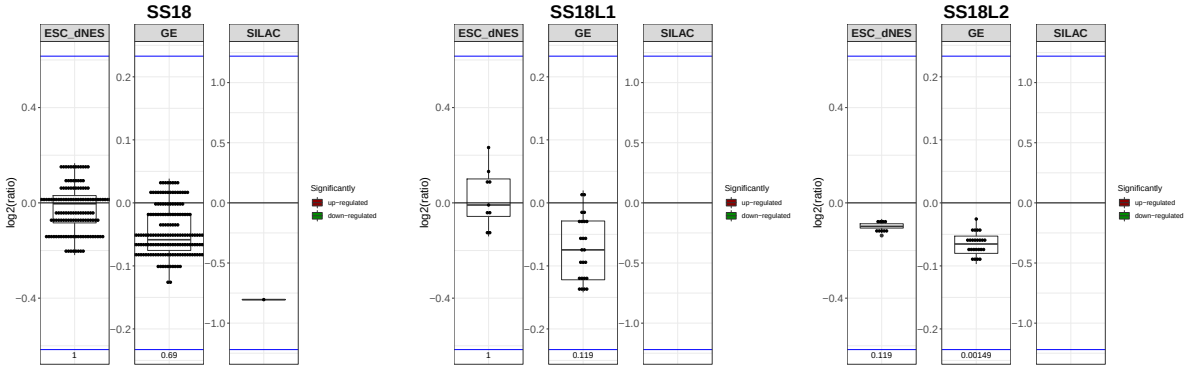


Figure 12.14: *Boxplots comparing the log2 ratios for SS18, SS18L1 and SS18L2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

The reduction of esBAF proteins signifies a transition from esBAF to npBAF in $_{hf}$NS cells. esBAF appears to persists in $_{hf}$NES cells and there is a transition towards npBAF as they transform into $_{hf}$NS cells. Removal of esBAF from hESC has been shown to reduce their proliferation and pluripotency, thereby restricting their self renewal and forcing them to differentiate.[97] The effect on the core pluripotent markers is delayed by many cell divisions, and there is an initial upregulation of pluripotent markers as esBAF has a repressive effect on them. As these genes are being actively induced, their expression is merely modulated by esBAF.[97] esBAF also represses genes involved in differentiation, maintaining ESC pluripotency

and perhaps that of $_{hf}$NES cells. Some of the proteins found to be interacting with esBAF were also upregulated in $_{hf}$NES cells, namely *SOX2*, *SALL4*, *RIF1*, *DNMT3B*, *HELLS*, and *JARID2* (Figure 12.15).

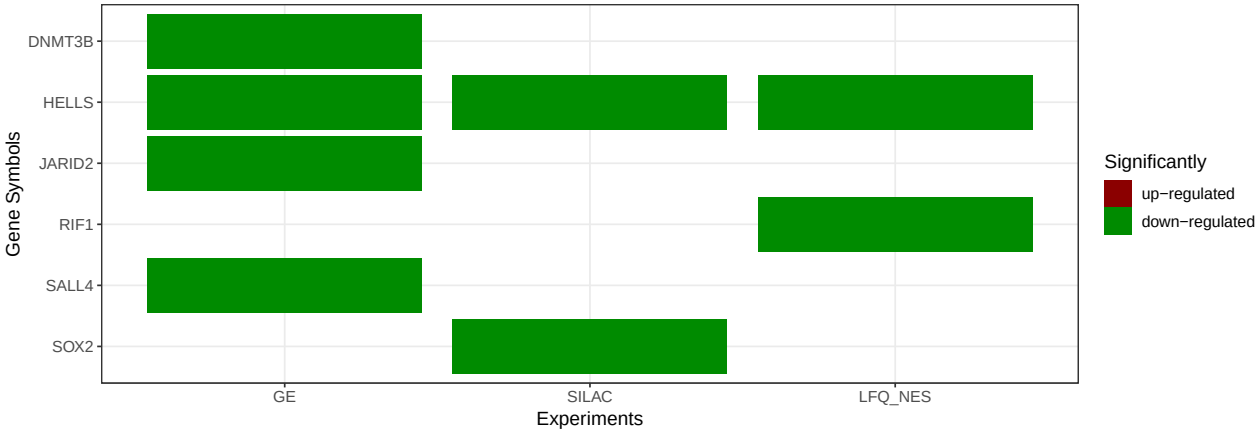


Figure 12.15: *SDE genes and proteins known to interact with esBAF*

12.1.2 Cellular components enriched in $_{hf}$NS cells

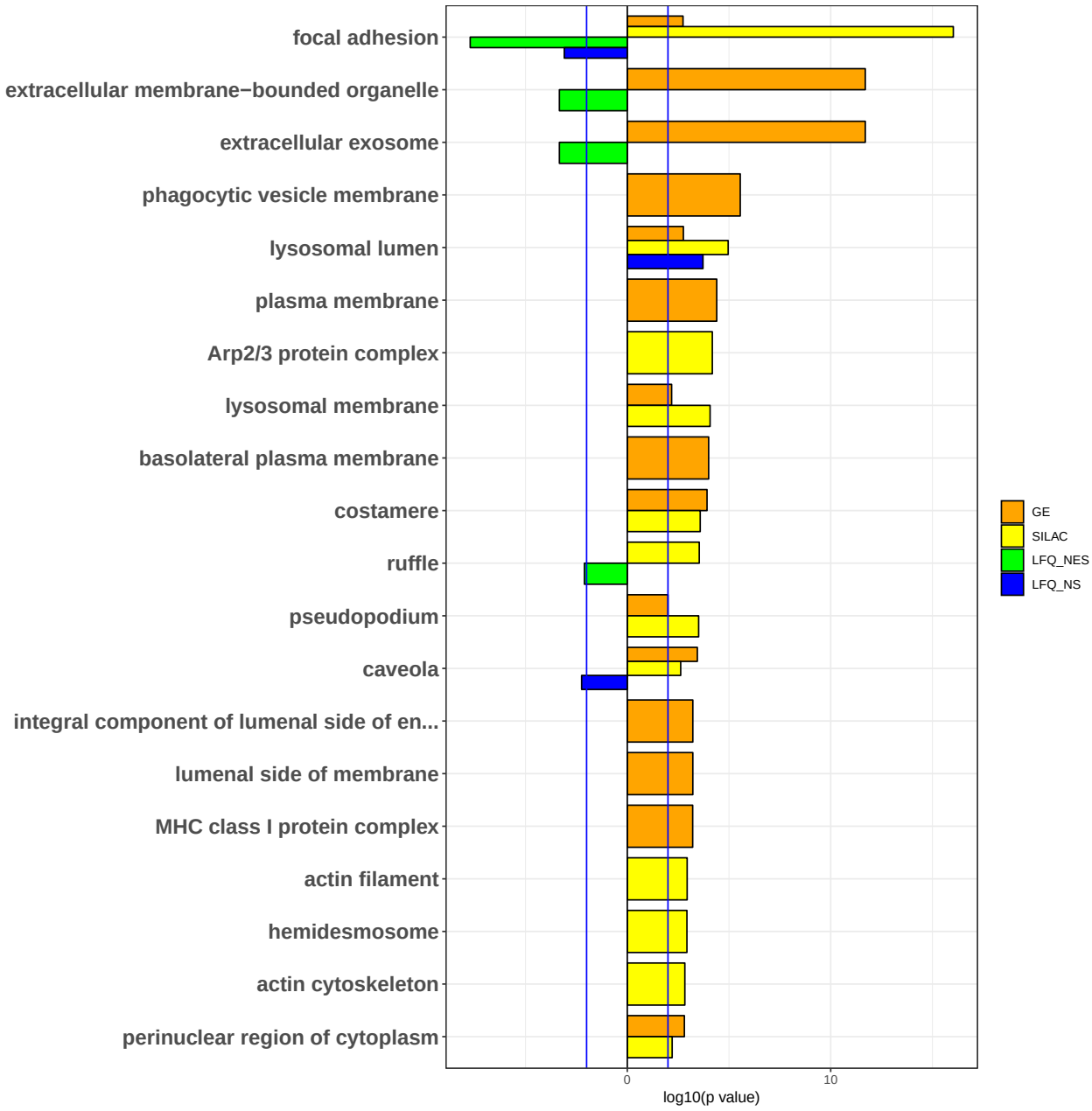


Figure 12.16: *Top twenty most highly enriched CC terms in the upregulated genes or proteins from all the experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the $_{hf}$NES cell differentiation experiment, and LFQ_NS (blue) refers to the $_{hf}$NS cell differentiation experiment. Enrichment p-values were log10 transformed. Down-regulated genes and proteins have negative log10(p-values) while up-regulated genes and proteins have positive log10(p-values), so that they can be plotted together to illustrate the differences in enrichment. Vertical blue lines represent the p-value cutoff of log10(0.01), positive and negative.*

The down-regulated proteins of the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments representing those most highly expressed in $_{\text{hf}}$NES cells and were enriched for CCs of the nucleus responsible for maintaining cell proliferation. The up-regulated proteins represent those most highly expressed in $_{\text{hf}}$NS cells. These proteins were enriched for CC terms from membranes and extracellular regions of the cell, namely the ***plasma membrane***, ***extracellular exosome***, ***lysosome***, ***endosome***, and the ***actin filament*** (Figure 12.16). Enrichment of these terms indicates the cells have moved from intrinsic to extrinsic control mechanisms. Instead of being primarily focused on proliferation, they are now establishing their position within the neural tissue. The neural tissue at this point in becoming more complex, and cells need to adhere to the control mechanisms established in the layering during brain development.

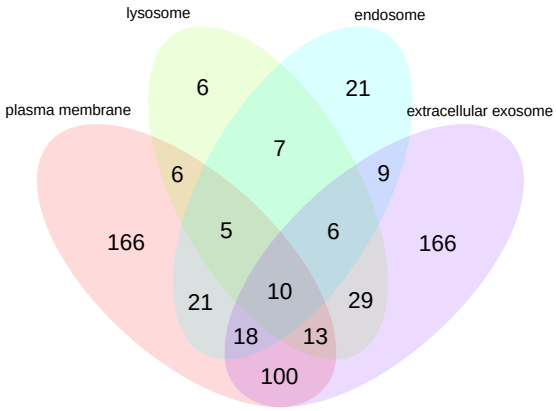


Figure 12.17: *Venn diagram comparing the overlap of genes and proteins SDE in the GE* $\frac{NS}{NES}$ *and SILAC* $\frac{NS}{NES}$ *experiments for the CC terms* ***plasma membrane***, ***lysosome***, ***endosome***, *and* ***extracellular exosome***

Most of these terms contain too many genes and proteins to illustrate their interaction networks or to discuss their components meaningfully. Venn diagrams illustrating the overlap of genes and proteins between terms, and provide insight into the level of uniqueness of the enrichment. The uniqueness of many plasma membrane proteins from those within the vesicles being produced is shown in Figure 12.17. Enrichment of the plasma membrane indicates

changes to the cell surface and the relationship to its environment. Enrichment the ***lysosome***, ***endosome***, and ***extracellular exosome***, may be in preparation for the astrocyte degradation of extracellular compounds accumulating in the brain. ***Extracellular exosome*** was down-regulated in $_{\text{hf}}$NES cell differentiation experiment, perhaps because they are not required by neurons. Exosomes are also used in intracellular communication,[6] further supporting the notion that as the $_{\text{hf}}$NS cells develop they are becoming more responsive and reactive to their environment. ***Focal adhesion*** was already discussed in section 11.2.2, and was enriched in $_{\text{hf}}$NS cells both in KEGG and CC. As discussed, ***Focal adhesion*** interacts with the actin cytoskeleton. ***Actin filaments*** were enriched in the up-regulated proteins in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment $_{\text{hf}}$NS cells.

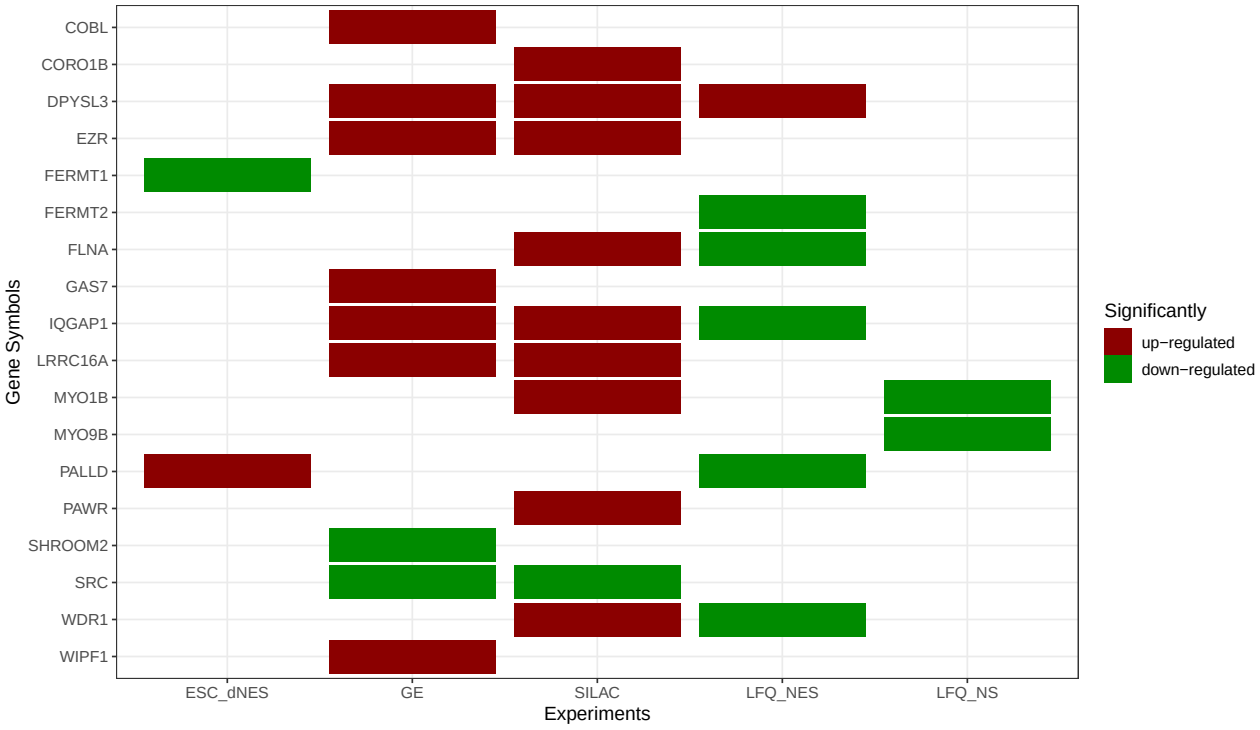


Figure 12.18: *Heatmap of the genes and proteins from CC* ***actin filament*** *SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

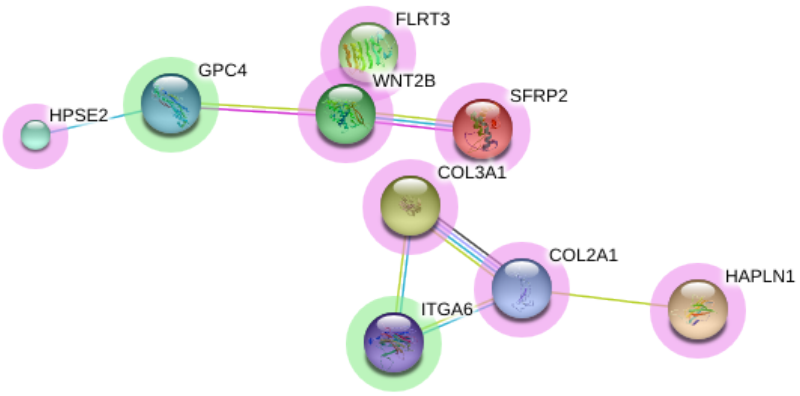


Figure 12.19: *STRINGdb network for Genes SDE in the Huang $\frac{dNES}{hESC}$ experiment from CC* ***extracellular matrix****. Red and green halos indicate up- and down-regulation respectively.*

Interestingly, the CC term ***extracellular matrix*** was upregulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. This up-regulation of ECM components as hESC develop into dNES cells may be a means of producing the correct extracellular substrate for neuroepithelial cells *in vivo* and later for the neural stem cells to develop within (Figure 12.19).

12.2 Molecular Function

The MF terms enriched in down-regulated genes and proteins from the GE $\frac{\text{NS}}{\text{NES}}$ experiment and the SILAC correlate well with the enriched CC. The MFs relate to the cell cycle and cell proliferation, namely ***DNA binding***, ***chromatin binding***, ***damage DNA binding*** and ***helicase activity***. Interestingly two signalling pathways, ***Wnt-protein bindins*** and ***platelet-derived growth factor binding*** are enriched in the up-regulated genes in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment. The up-regulated genes in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment represent proteins being expressed as the hESC differentiate into dNES cell. These same pathways are enriched in down-regulated genes in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. This would mean that the members of this pathways are more highly enriched in the $_{\text{hf}}$NES cells vs the $_{\text{hf}}$NS cells.

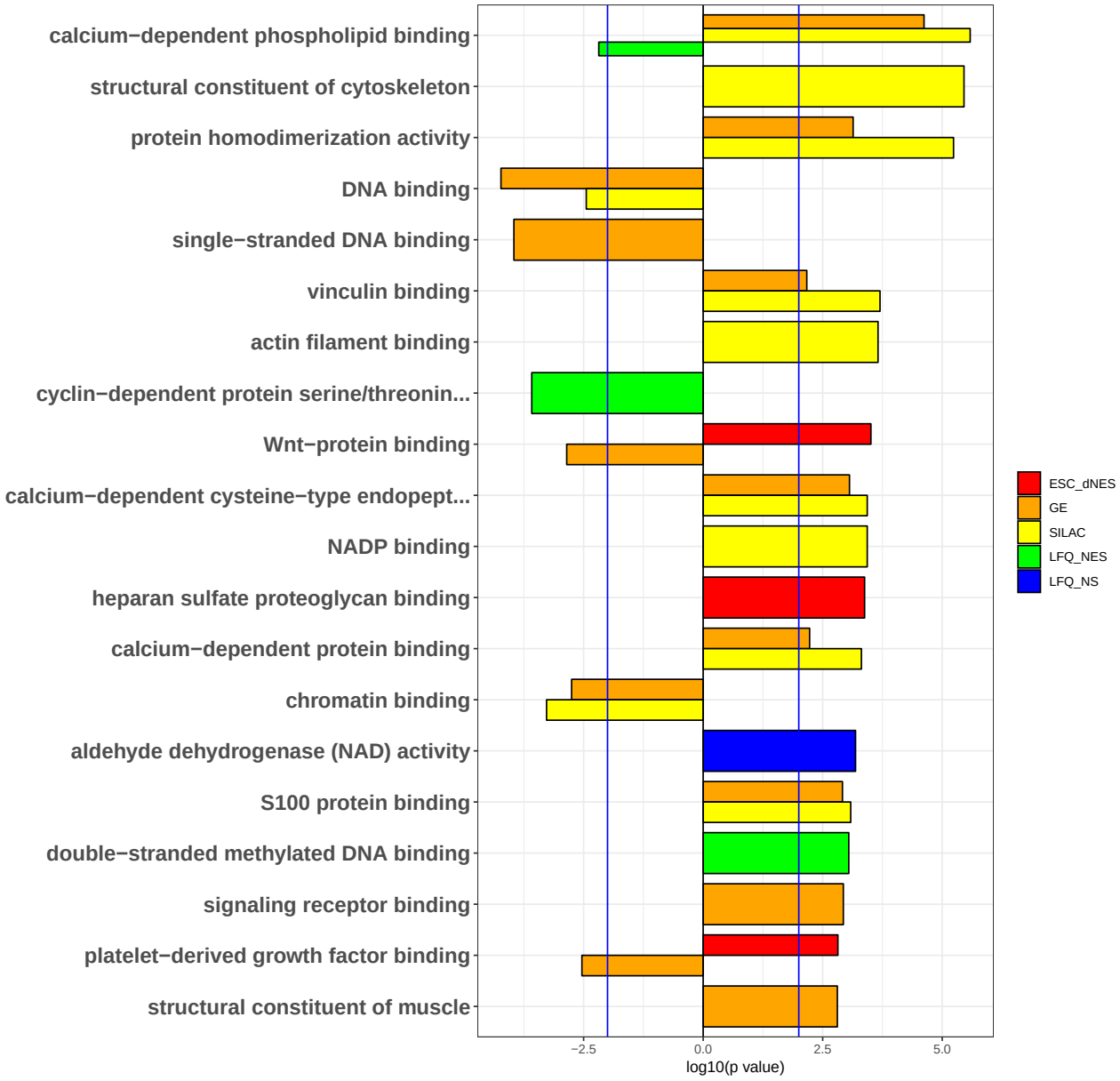


Figure 12.20: *MF Top 20 enriched terms across all experiments. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment.*

Part V

Discussion and Conclusion

Chapter 13

Developmental epigenetics

13.1 Introduction

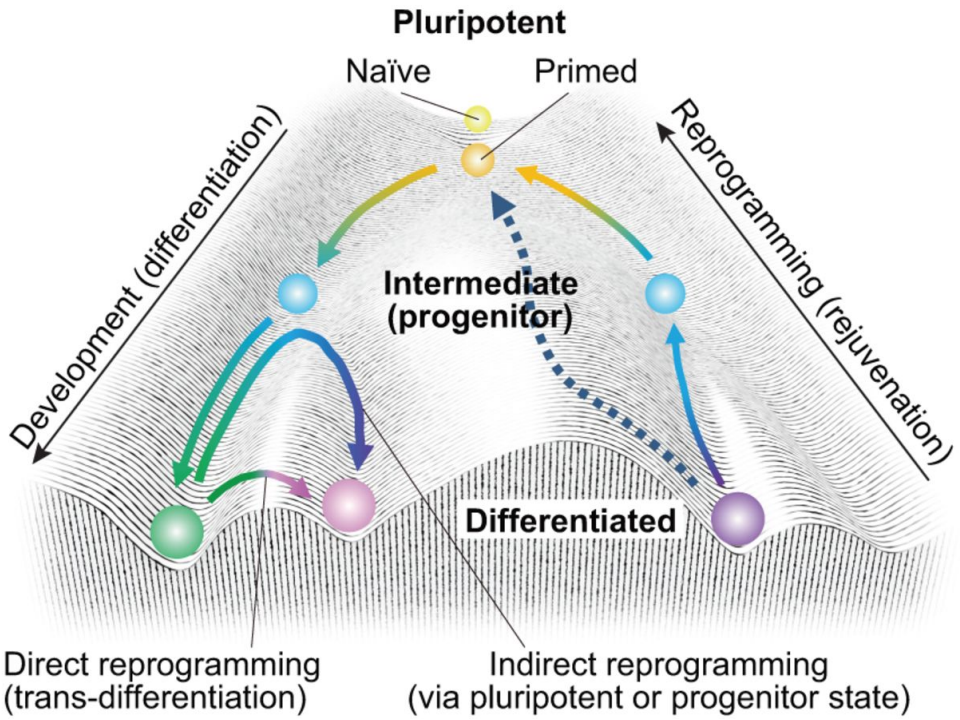


Figure 13.1: *The Waddington landscape of cell fate trajectories (Takahashi and Yamanaka 2015).*

Differentiation according to the Waddington Landscape (Figure 13.1)[213] is described as a cell rolling down a landscape. ESCs would be at the top of a three-faced mountain, with each face representing, one of the three layers formed during gastrulation. Each face consists of

gullies, down which cells are sorted, that empty into valleys. Progenitors cells are formed in the highest valleys, and as these valleys fill up the cells spill over running further down the face until they reach the lowest valley, a stable state and become terminally differentiated.

During gastrulation, the epiblast separates into three layers, the endoderm, mesoderm and ectoderm. BMP produced by the mesoderm, induces the ectoderm to become epithelia, but the notocord or organiser produces BMP inhibitors, *CHRD*, *FST NOG* and *CER1*, thereby enabling the neural plate to spontaneously form in the absence of BMP or any other stimulation. *In vitro*, if ESCs are plated sparsely and in the absence of growth factors, they differentiate into neurons. This may be due to dilution of the BMPs produced by the ESCs. Neuroepithelial like cells can be derived from ESC[65] or iPSC[114] cells *in vitro*, using BMP inhibitors. This indicates that the default differentiation trajectory is towards neurons, known as the default model.[158, 201]

Neuroepithelial cells and $_{\text{hf}}$NES cells would therefore represent cells present early in the development trajectory of the embryo and are therefore possibly closely related to ESCs, as they have been sheltered from the patterning stimulation that the rest of the embryo is subjected to. When comparing the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment to the GE $\frac{\text{NS}}{\text{NES}}$ experiments, it is evident that there are considerably more differences between the $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells than between the hESC and the dNES cells.

In this study, we have used two different cell lines derived at different timepoints from the human foetal hind-brain. The $_{\text{hf}}$NES cells were derived at five weeks, while the $_{\text{hf}}$NS cells were derived at nine weeks. The cells were derived and cultured in identical conditions. The extrinsic environment of the cells was therefore identical. The cells, however, had vastly different morphologies under proliferative conditions and different fates upon differentiation.[206,211] The differences in these cells are therefore likely to be intrinsic.[60] The $_{\text{hf}}$NS cells, however, may be altering their environment though secretion of ECM components and through increased cell-cell interactions.

Here, we show that these $_{\text{hf}}$NES cells still maintain much of the pluripotent characteristics, and they are mainly controlled by intrinsic factors. Only once they develop into $_{\text{hf}}$NS cells do

they become more receptive to their environment.

13.2 Epigenetics

During replication, DNA in chromosomes is tightly packed, but during interphase, it loosens up to form heterogeneous chromatin. Euchromatic areas are open areas of chromatin, within which most transcribed genes are found. Heterochromatin is more densely packed and can be divided into two classes; facultatively and constitutive. Constitutive heterochromatin is condensed in every cell type and usually does not contain genes. Facultative chromatin differs according to cell type and these areas are condensed upon developmental cues. The density of the chromatin affects transcription factor binding and thereby gene expression. Facultative chromating can be further divided into two compartments: A and B.[90] Compartment A consists of accessible DNA that may be transcribed, while compartment B consists of condensed chromatin. As cells differentiate, topological associated domains (TADs) move from one compartment to another. With this rearrangement, there are subtle changes in gene expression. It was previously believed that these changes were irreversible and that once a cell had begun its journey its course could not be altered. Recently with the advent of induced pluripotency and direct programming, this belief has been overturned. Through induced pluripotency, cells are pushed back up to the top of the pyramid and are allowed to start a journey of any trajectory. Using direct reprogramming, pioneer factors enable sufficient plasticity to allow terminal differentiation factors to push cells towards a new valley.

The state of a cell, is a reflection of the genes that are being transcribed, the journey they go on to become mature proteins, and where these proteins find themselves within the cell. Proteins, the building blocks of the cell, are transcribed in the nucleus, assembled by the ribosome, then self-assemble into complexes to perform cellular functions or form the structural components of cells. Each of these building blocks is part of a fractal, finding its perfect place within the whole. If one changes the composition of the parts the nature of the whole is transformed.

The fate determination of a cell is defined by its intrinsic state, the genes that are currently being transcribed, and those that are poised to be transcribed. ESCs have open chromatin, and their transcription is mainly controlled by pluripotent factors that induce proliferation and inhibit differentiation, thereby promoting rapid cell proliferation.[62,82,224] Besides the strong induction of pluripotent factors, transcription is not tightly regulated, and many genes are expressed at low levels but have minimal effect on the cells, due to the dominant effect of pluripotent factors.[61] Upon differentiation, induction of the pluripotent factors is reduced and the effect of differentiation factors can begin to take hold. This effect is usually mediated by extrinsic factors, activating signalling pathways and inducing specific genes that guide the direction of differentiation. Differentiation genes are often controlled by bivalent marks on their chromatin.[30] The genes have both repressive and activation marks but are usually silenced. During differentiation one of these marks is removed, resulting in the genes being either permanently activated or repressed.[180]

The intrinsic regulation of cells is based on their epigenetic and genetic differences, allowing cells to maintain different states in homogenous environments.[60] Epigenetic changes occur as changes in DNA methylation, histone modification, and non-coding RNA. Differences in transcription factor expression and the availability of signalling response elements can also be viewed as intrinsic factors. This internal environment is the state of the cell. Its response to the environment is therefore defined by this state. Changes in the environment need to transverse the internal environment of the cell, to effect changes to the transcription factor expression. Prolonged exposure to a new environment creates a new intrinsic state within the cell, thus changing its nature and how it may respond to the environment in the future.

13.2.1 DNA methylation

DNA is methylated on cytosine mainly in CpG-rich areas except CpG islands. If these areas occur in gene regulatory regions, the result is usually gene silencing.[208] Active genes increase methylation in gene bodies[208] and this is thought to reduce spurious transcription from pseudo TSSs. DNA methylation may have been introduced early in evolution as a means to control

the activation of exogenous DNA inserted into the host genome. It is a effective means of controlling transposable elements. DNA methylation is thought to repress transcription by preventing transcription factors binding to DNA or by recruiting transcription depressors.

DNA methylation is erased during formation of the gametes and again after fertilisation.[156] The genome is then globally methylated, although not all promoters of genes are rich in CpG, and are therefore precluded from silencing. Genes required for housekeeping and possibly those required for pluripotency are therefore able to be expressed ubiquitously. Methylation patterns during development are stage- and tissue-specific, and changes define different cell types. There are highly significant changes in DNA methylation patterns during foetal brain development,[197] particularly site-specific methylation patterns between neurons and glia.[125]

DNMT1 maintains methylation by methylating the newly synthesised DNA strand that is not initially methylated during cell proliferation. *De novo* DNA methylases *DNMT3A* and *DNMT3B* increase methylation as they are active in new areas previously not methylated. *DNMT1* can also function as a *de novo* methyltransferase.[34]

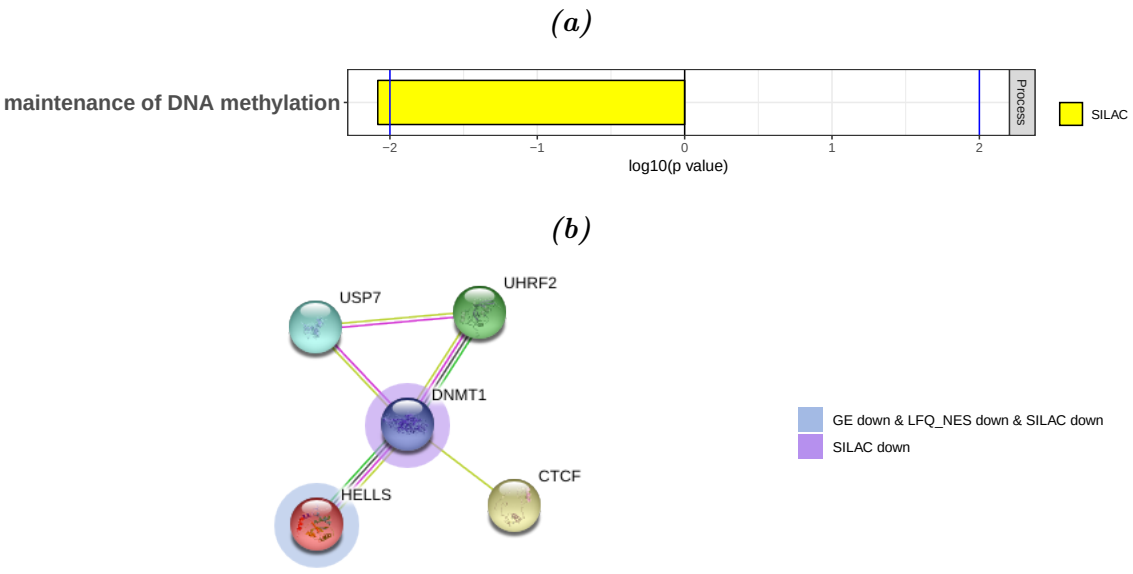


Figure 13.2: *BP* ***Maintenance of DNA methylation****, (a) Log10 enrichment score and (b) STRINGdb network. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

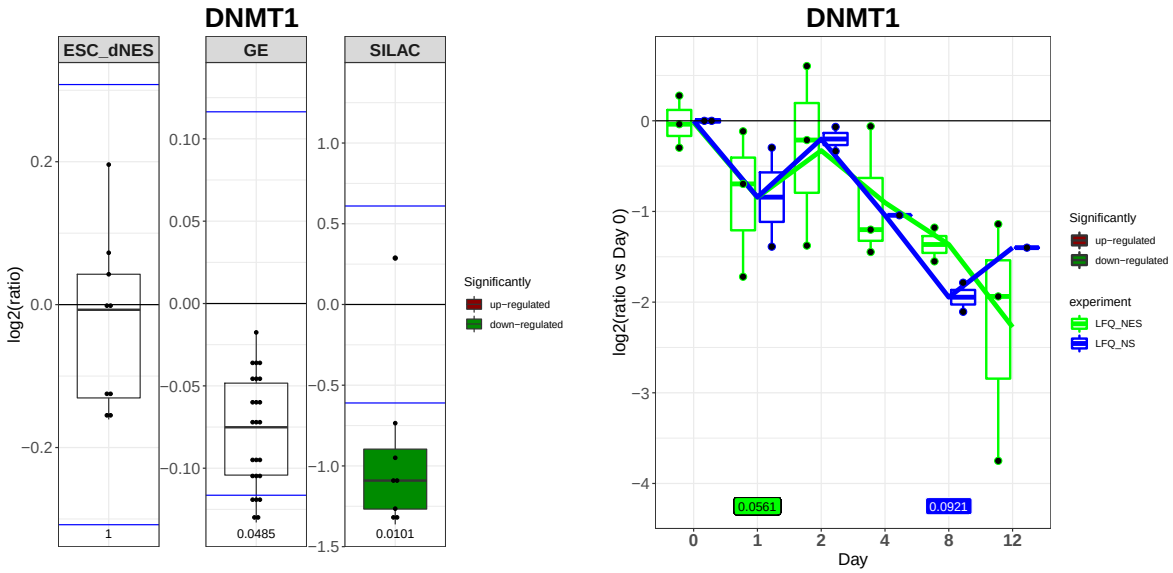


Figure 13.3: *Boxplots comparing the log2 ratios for DNMT1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

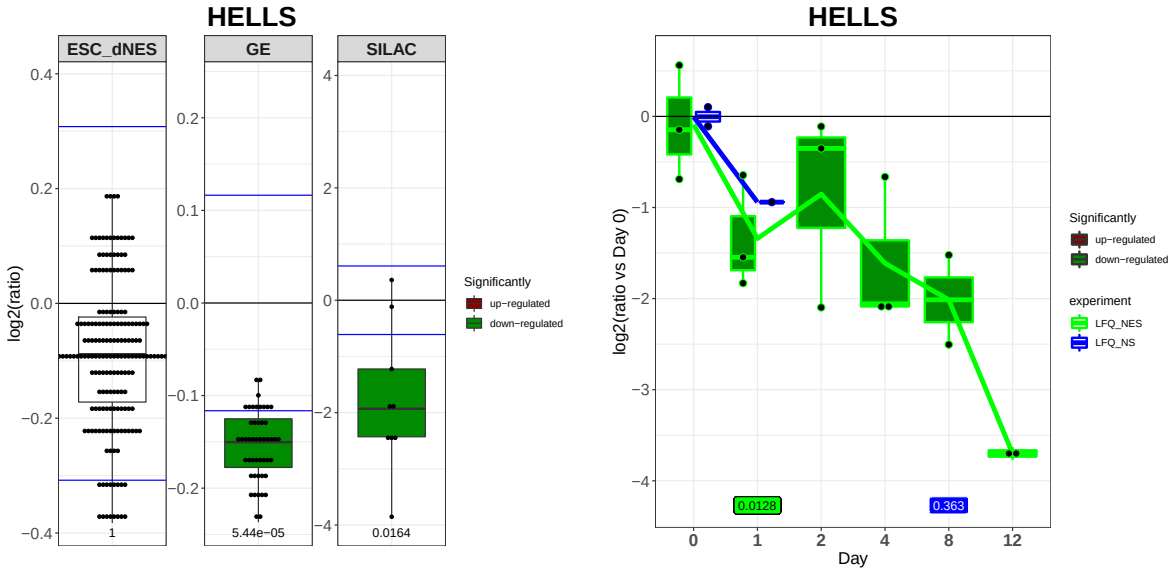


Figure 13.4: *Boxplots comparing the log2 ratios for HELLS. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

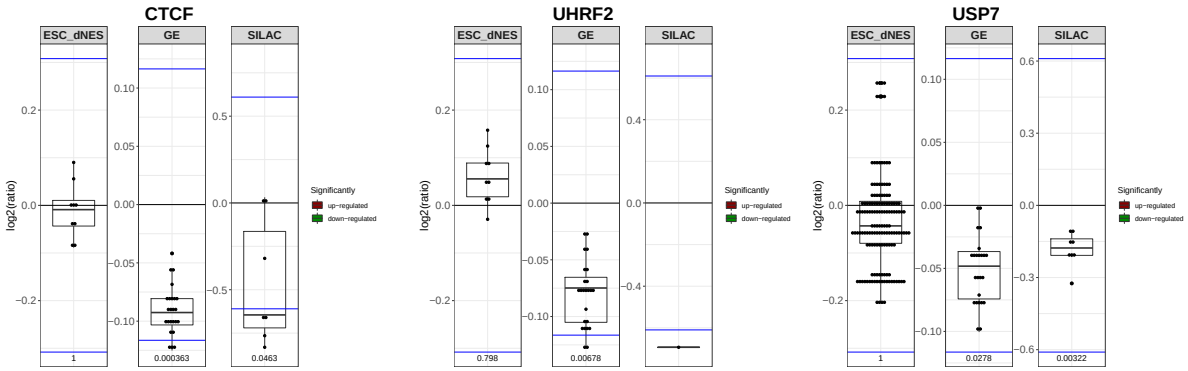


Figure 13.5: *Boxplots comparing the log2 ratios for CTCF, UHRF2 and USP7. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

The BP term ***maintenance of DNA methylation*** was significantly enriched in the down-regulated proteins from the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.2). Within this term *DNMT1* and *HELLS* were SDE. *DNMT1* was more highly expressed in $_{\text{hf}}$NES cells than in $_{\text{hf}}$NS cells in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and then decreased during differentiation of both cell types (Figure 13.3). Knockout of *DNMT1* in the cortex was shown to decrease the number of neurons and induce precocious gliogenesis.[141] *HELLS* was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and the $_{\text{hf}}$NES cell differentiation experiment (Figure 13.4). It is essential in normal development and required for maintenance and *de novo* DNA methylation. All other members of this term showed decreased trends in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments, although none were statistically significant (Figure 13.5).

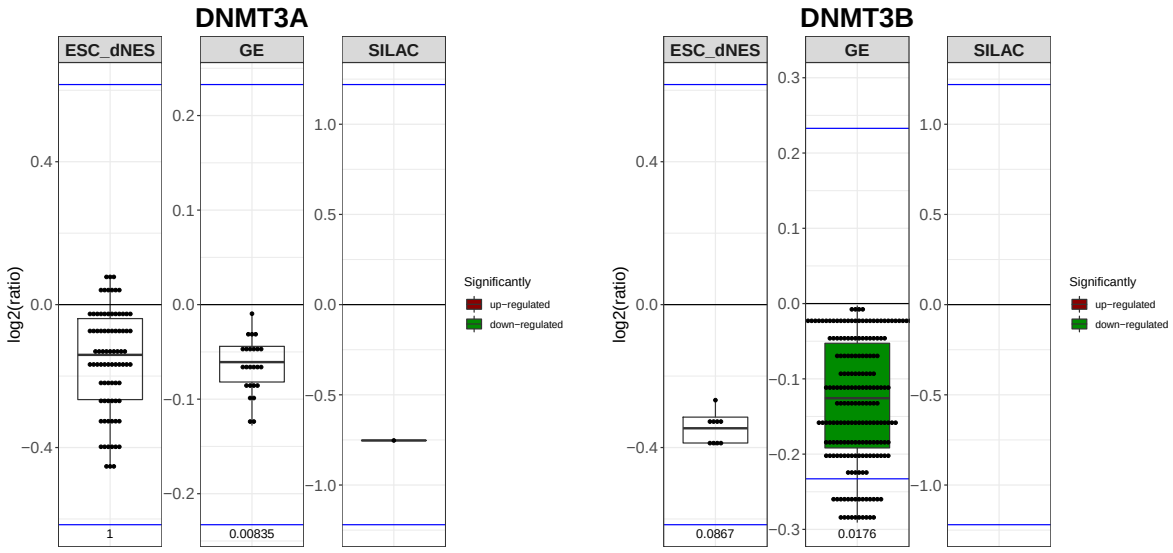


Figure 13.6: *Boxplots comparing the log2 ratios for DNMT3A and DNMT3B. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

In previous studies there was a difference in the time of expression between *DNMT3A* and *DNMT3B* in the developing brain.[197] *DNMT3B* was observed to be expressed during early neurogenesis, while *DNMT3A* is observed in embryonic and postnatal CNS.[69] *DNMT3A* and *DNMT3B* appeared to be more highly expressed in $_{\text{hf}}$NES cells, than in $_{\text{hf}}$NS cells, but only *DNMT3B* was SDE in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. These *de novo* DNA methyltransferases may promote DNA methylation in the $_{\text{hf}}$NES cells, cementing their cellular fates.

There is a correlation between the binding of methylated DNA binding domain (MBD) proteins and the density of DNA methylation.[25] MBD proteins have been found in association with chromatin remodelling complexes and may play roles in alternative splicing and in some cases gene activation.[25,57]

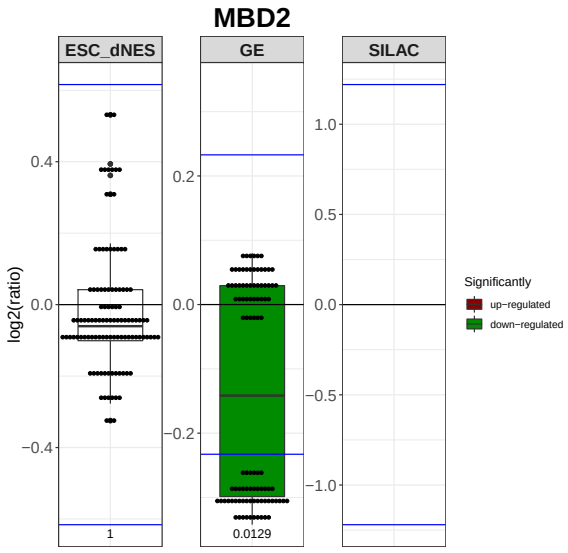


Figure 13.7: *Boxplots comparing the log2 ratios for MBD2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

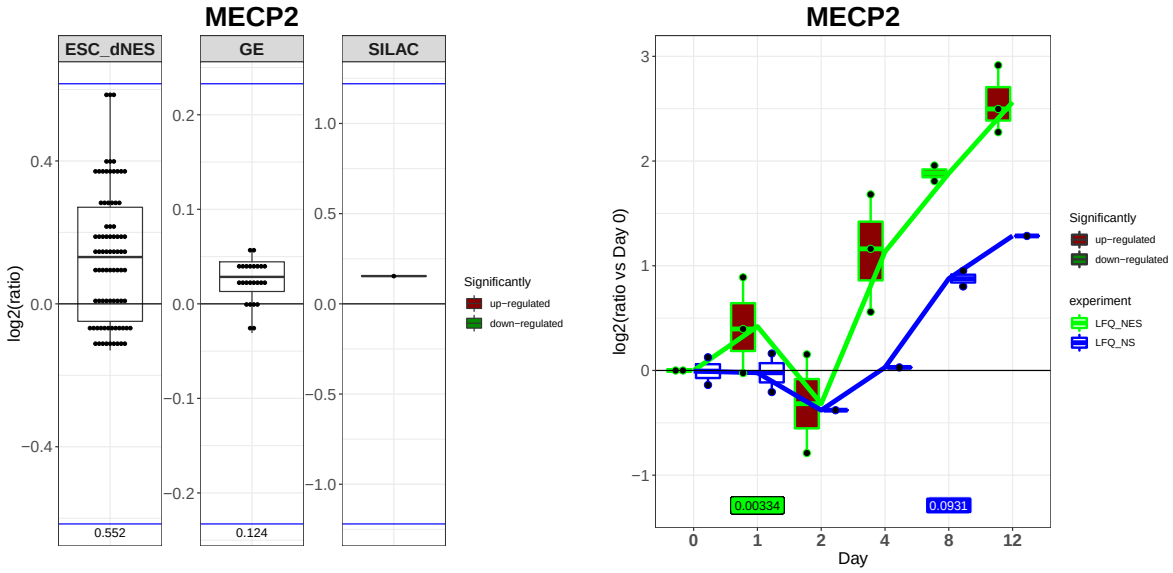


Figure 13.8: *Boxplots comparing the log2 ratios for MECP2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Two MBD proteins *MBD2* and *MECP2* were SDE in this study. *MBD2* was up-regulated in the ${}_{\text{hf}}$NES cells vs ${}_{\text{hf}}$NS cells in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. It appears as two transcripts,

NM_003927.3 and NM_014832.3, although only NM_015832.3 was SDE. *MBD2* has previously been associated with tissue-specific methylated gene repression during differentiation. *MECP2* becomes up-regulated during $_{hf}$NES cell and $_{hf}$NS cell differentiation, and it may play important roles here in the terminal differentiation of neural stem cells.

DNA demethylation may occur in a passive manner during the synthesis of new DNA and through active removal of methylation. The mechanisms of demethylation are as of yet unclear.[113] It is believed that through oxidation, 5-methylcytosine (5mC) is converted to 5-hydroxomethylcytosine (5hmC),[79] 5-formylcytosine (5fC), and 5caCl5-carboxylcytosine (5caC) and then through deamination the oxidised base is converted to thymine or uracil, recognised as an incorrect base-pair or mismatch and then replaced. 5hmC is abundant in ESCs[113] and its levels are critical to neuronal development.[113]

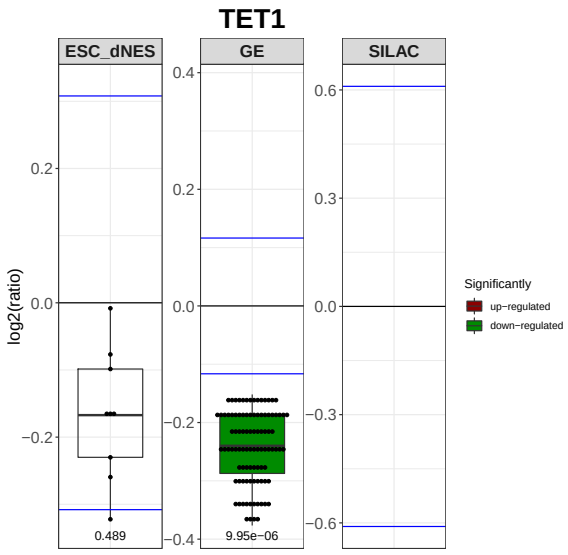


Figure 13.9: *Boxplots comparing the log2 ratios for TET1. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Active DNA demethylation occurs through a series of oxidations of 5mC to 5hmC, 5fC, and 5caC by the ten-eleven translocation (TET) proteins. In the GE $\frac{\text{NS}}{\text{NES}}$ experiment, *TET1* was down-regulated while the others showed no significant changes.

There are a number of key genes and proteins involved in DNA methylation that were

SDE in this study. This indicates that DNA methylation changes may be at play here to direct the identity and fate specification of these cells. No direct DNA methylation data were collected in this study, but future whole-genome DNA methylation analysis of this cell model may provide valuable insights into the changes in DNA methylation and their effects on foetal brain development.

13.2.2 Chromatin

Nucleosomes

A nucleosome consists of four dimers of four core histones, H2A, H2B, H3, and H4. A section of 147bp of DNA is wound tightly around each nucleosome. Higher order structures are formed through linking nucleosomes via linker H1 histones and varying lengths of linker DNA. This enables the almost 2 m of DNA to be condensed enough to fit inside a 6μm nucleus.

Pluripotent cells usually contain more euchromatin. The chromatin becomes progressively replaced by heterochromatin as development progresses. The change is usually permanent and is visible via electron microscopy.

The density of the chromatin is affected by the composition of histones. Histone composition can be changed by exchanging histone variants or covalently modifying the existing histones. There are multiple isoforms of each histones, most notably replication-dependent and -independent isoforms. In yeast, these are synchronised with the cell cycle.[20] In the mouse eye, replication-dependent histones were recently detected in neurons, indicating more complex control in mammals and perhaps the necessity for a more detailed re-classification.[20] Many of the histone isoforms appear to have redundant functionality, but some play distinct roles in proliferation and gene expression. Different expression levels of these isoforms may affect the nature and density of the chromatin.[23]

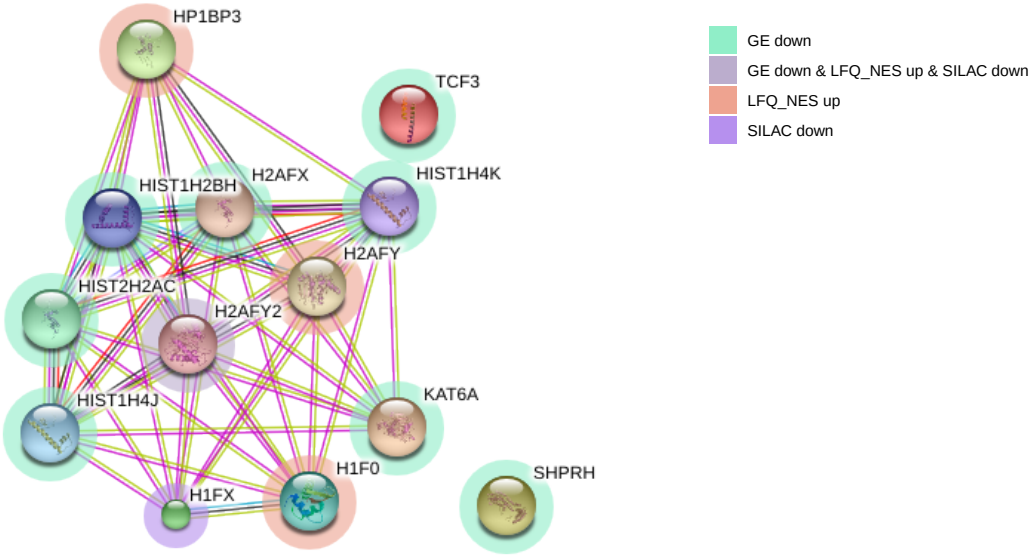


Figure 13.10: *CC* ***nucleosome****. STRINGdb network for the SDE genes and proteins. There are too many members of the nucleosome to illustrate all meaningfully. The colour of the Halo's indicate the SDE in the different experiments.*

The CC term ***nucleosome*** was not enriched in this study, despite many of its components being SDE (Figure 13.10). There are however, BPs related to the nucleosome significantly enriched in the SILAC $\frac{NS}{NES}$ experiment, namely ***nucleosome disassembly***, ***DNA replication-dependent nucleosome assembly***, and ***CENP-A containing nucleosome assembly*** (Figure 13.11).

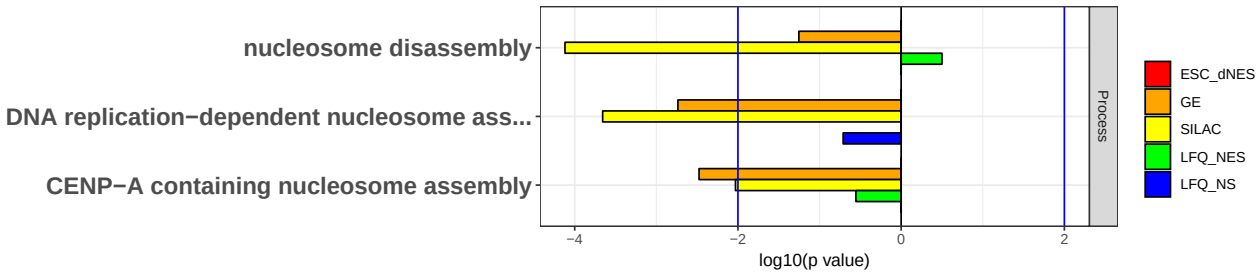


Figure 13.11: *BP nucleosome-related terms, enriched in the SILAC $\frac{NS}{NES}$ experiment. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}$NES cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}$NS cell differentiation experiment.*

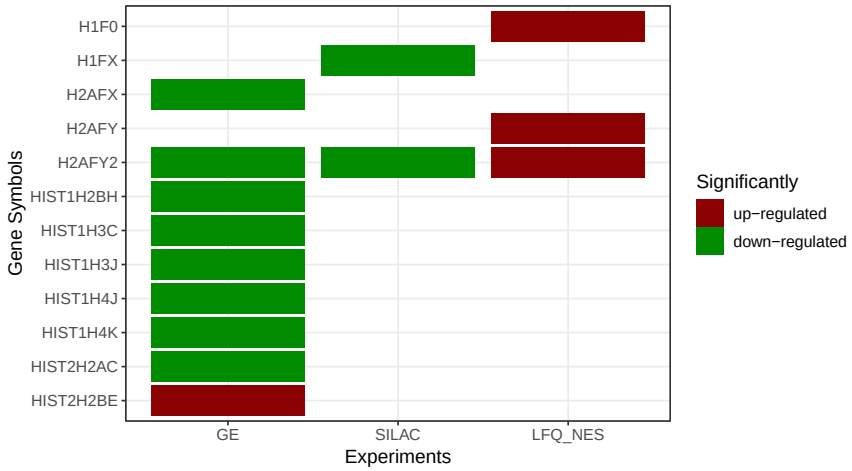


Figure 13.12: *Heatmap of histones SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

The following histones from each core histone class were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment.

H2A : *H2AFX*,*H2AFY2*, *HIST1H2BH*

H2B : *HIST2H2AC*

H3 : *HIST1H3C*, *HIST1H3J*

H4 : *HIST1H4J*, *HIST1H4K*

Interestingly only *HIST2H2BE* was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.12).

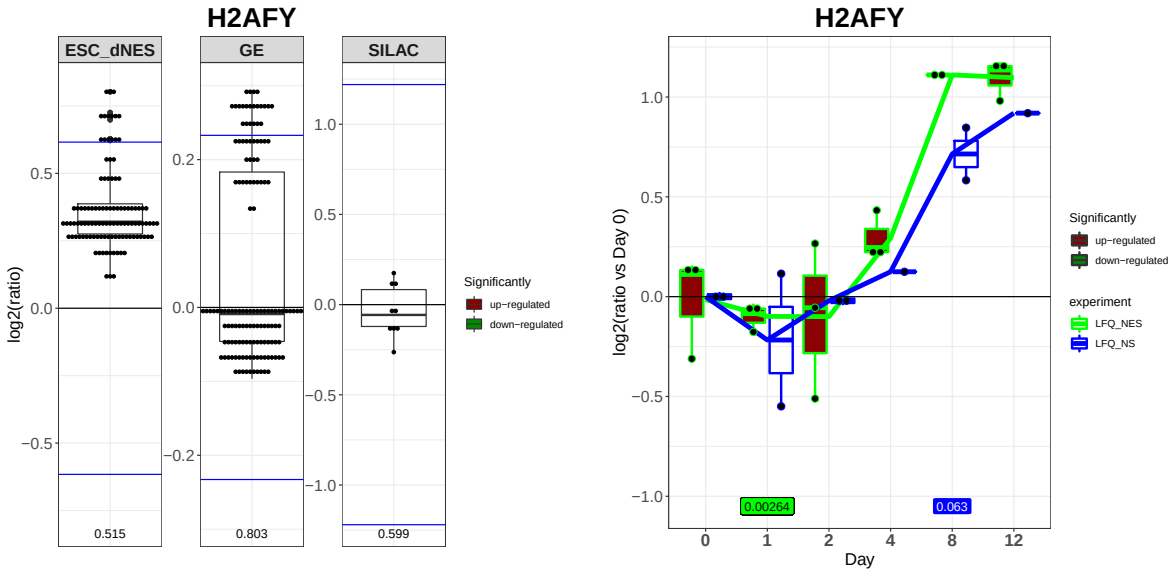


Figure 13.13: *Boxplots comparing the log2 ratios for H2AFY. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

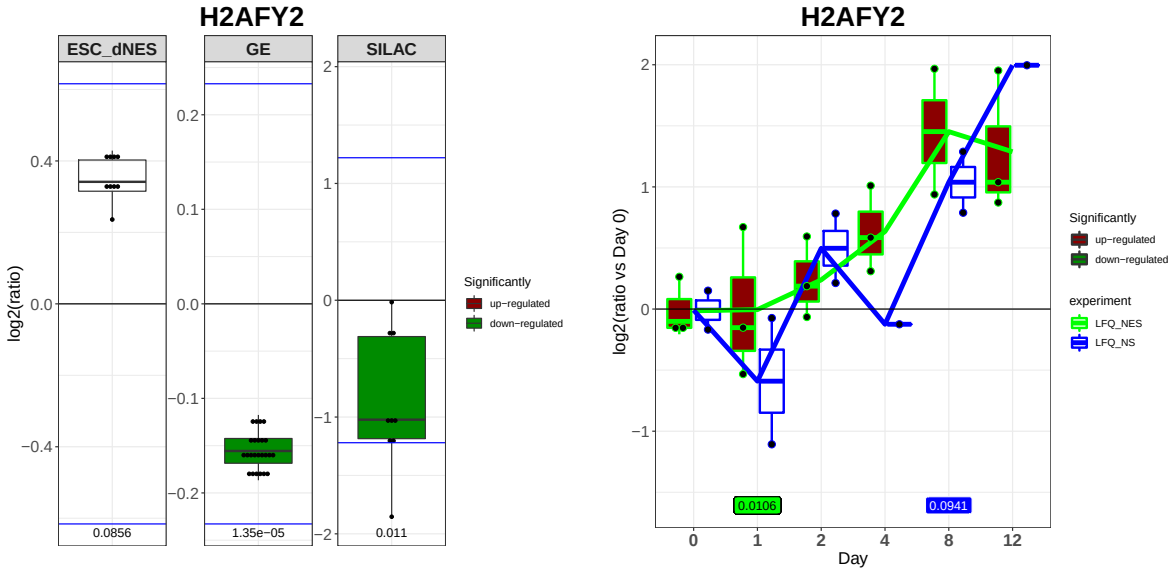


Figure 13.14: *Boxplots comparing the log2 ratios for H2AFY2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Most notably, *H2AFY* and *H2AFY2*, which are more commonly known as macroH2As were SDE. *H2AFY* showed little difference between $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells in the SILAC

$\frac{\text{NS}}{\text{NES}}$ experiment and was then up-regulated during differentiation of both cell lines (Figure 13.13). *H2AFY2* was up-regulated, although not statistically significantly, in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, and was then down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and up-regulated during differentiation (Figure 13.14). The macroH2A histone variants differ from their canonical H2A counterparts. They are known to be epigenetic regulators of cell fate decisions during development,[38] and regulate dynamic chromatin formation possibly in conjunction with the SWI/SNF complex.[38] They are the histones known to be most up-regulated during hESC differentiation, and have been shown to be upregulated in neural precursors which are *SOX9*-positive.[23] *SOX9* was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and GE $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 13.15).

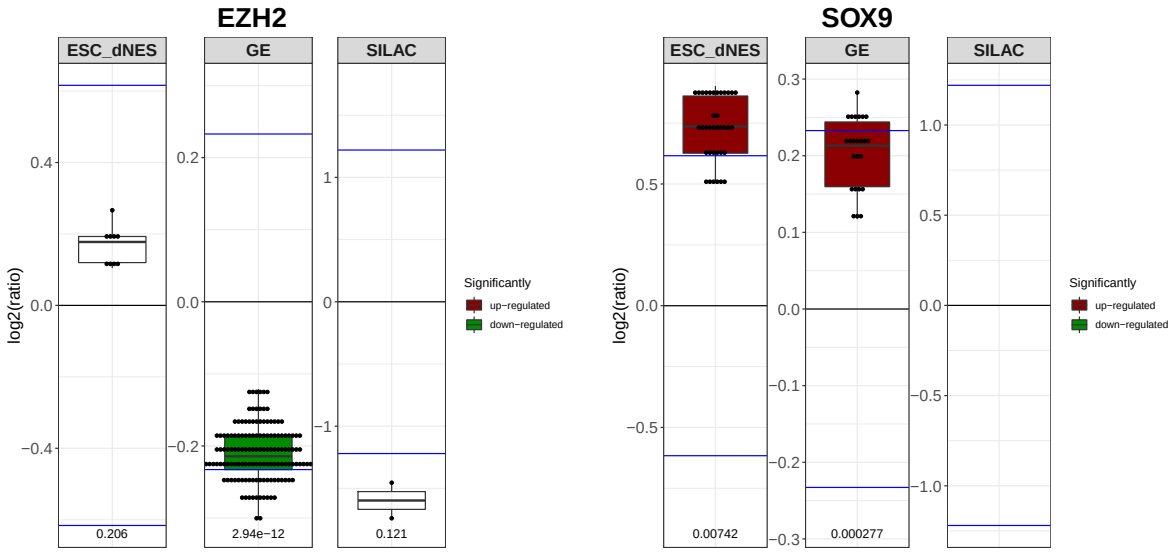


Figure 13.15: *Boxplots comparing the log2 ratios for EZH2 and SOX9. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

MacroH2As have been linked to both passive and active transcription repression, and are found at most transcriptionally inactive genes, most notably those containing the repressive mark H2K27me3. Transcription is passively repressed by macroH2As, by impeding access of transcription factors,[23] SWI/SNF-induced nucleosome sliding, and repressing the initiation of polymerase II transcription.[38] Transcription is actively repressed through interaction of

the macro domain with histone deacetylases,[23] and the polycomb group (PcG).[23] There is a known overlap between *EZH2*, a member of the PcG, and *H2AFY2* localisation.[23] *EZH2* was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.15), and thus co-expressed with *H2AFY2*.

MacroH2A's are found to occupy regulatory regions in genes of transcription factors, particularly those of the pluripotent network and the epigenetic regulators, many of which were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, (*CBX2*, *DNMT3A*, *LIN28A*, and *RCOR2*).[23] MacroH2As are usually linked to differentiation, being expressed at low levels in ESCs and neural stem cells, and then being induced during differentiation, they are most often expressed at higher level in somatic cells.[23] Here, paradoxically, we observed a higher expression in $_{\text{hf}}$NES cells than $_{\text{hf}}$NS cells, followed by an up-regulation during differentiation.

When macroH2As were knocked out in zebrafish, brain malformations were observed, particularly a reduction in the size of the hind-brain.[38] Knock-down in ESC impaired differentiation by delaying the silencing of *SOX2*. Interestingly, the differentiation from ESCs to $_{\text{hf}}$NES cells requires delayed silencing of *SOX2*. MacroH2As may be fundamental in inducing differentiation towards $_{\text{hf}}$NES cells, and then decrease in $_{\text{hf}}$NS cells and are again up-regulated to complete terminal differentiation.

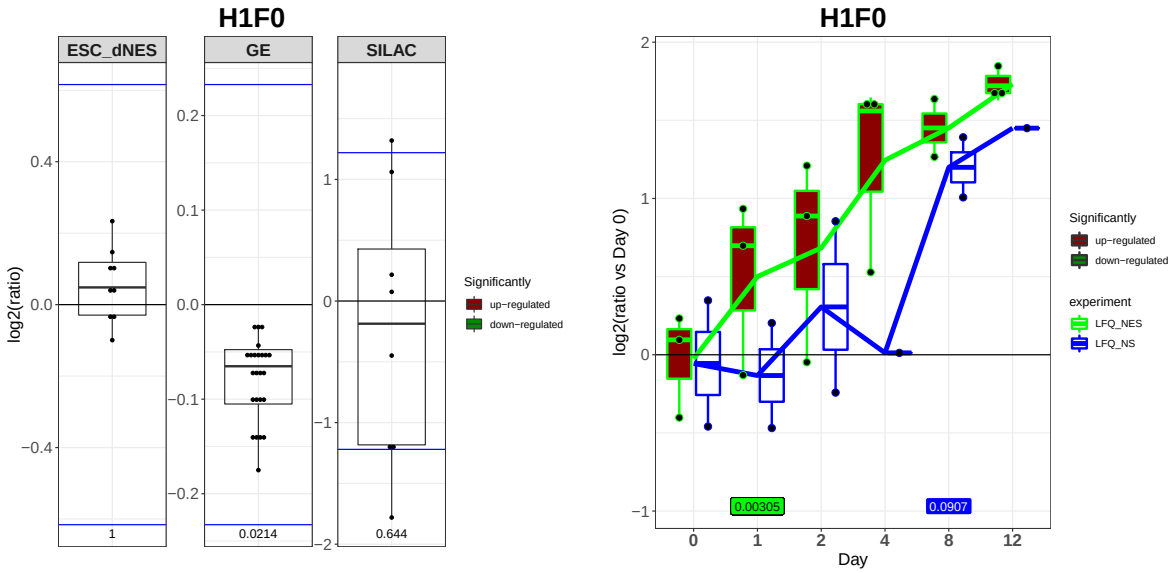


Figure 13.16: *Boxplots comparing the log2 ratios for H1F0. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

None of the replication-dependent H1 histones were SDE in this study, but some of the replication-independent H1 histones were, namely *H1F0* and *H1FX*. The replication-independent H1 histone *H1F0*, known to accumulate in terminally differentiated cells, was up-regulated during differentiation of both $_{\text{hf}}$NES cells and $_{\text{hf}}$NS cells, although not statistically significantly in the $_{\text{hf}}$NS cells (Figure 13.16). *H1FX* was up-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.12)

In a recent study, Zhang et al. 2016 used proteomics to determine the distinct interaction partners of two replication-dependent H1 linker histones (*HIST1H1A* and *HIST1H1C*) and one replication-independent histone *H1FX*. Interestingly many of the proteins interacting with only the replication-dependent histones, were up-regulated in $_{\text{hf}}$NES cells, namely *CDC73*, *CHD7*, *CTR9*, *DDX21*, *EBNA1BP2*, *SUPT16H*, and *TAF5* (Figure 13.17).

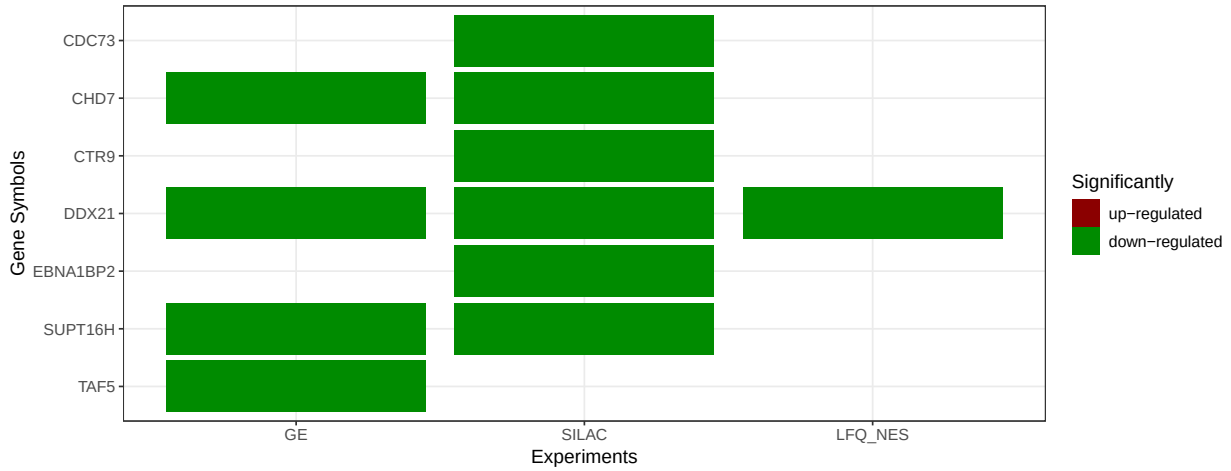


Figure 13.17: *Protein shown to interact with replication-dependent histones in Zhang et al. 2016*

Chromatin-remodelling

In addition to differences in histone composition affecting chromatin structure, changes in histone modification can affect both chromatin structure and gene expression. Histones are subject to a range of modifications, some serve as a means to facilitate chromatin compaction, while others serve as binding sites for proteins, that then affect gene expression. The major histone modifications are acetylation and methylation.

The animal transcription factor database (AnimalTFDB)[242] provides a list of chromatin-remodelling factors. Using the Fisher exact test the SDE genes or proteins from each experiment were analysed against all genes or proteins detected in the same experiment. The list of chromatin remodelling proteins were significantly enriched in the down-regulated proteins from the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (p=0.00015). Using these, a STRINGdb network of chromatin-remodelling factors SDE in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments was generated (Figure 13.18).

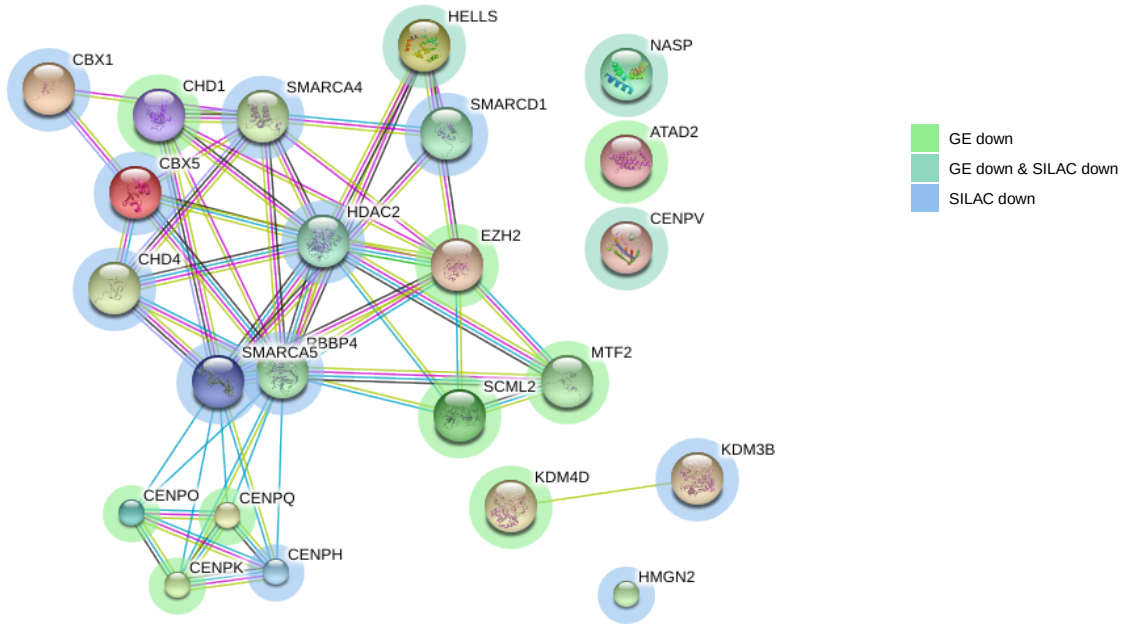


Figure 13.18: *STRINGdb network of chromatin-remodelling genes and proteins from the AnimalTFDB that were or SDE in the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments. The halo colours indicate the overlaps in SDE between experiments, as per the legend*

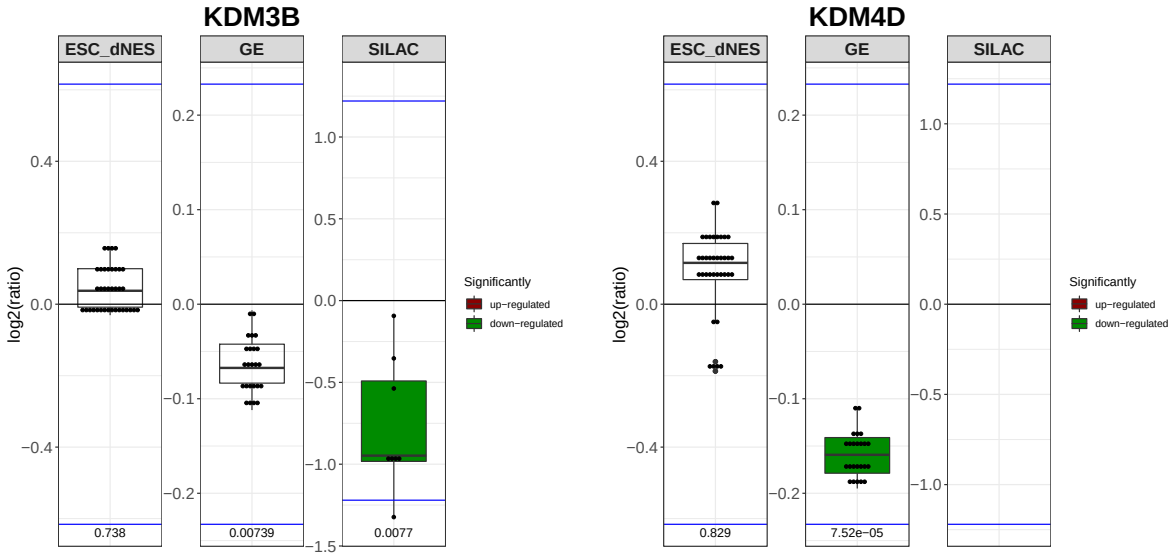


Figure 13.19: *Boxplots comparing the log2 ratios for KDM3B and KDM4D. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Two lysine demethylases, *KDM3B* and *KDM4D*, were down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$

experiment and the GE $\frac{\text{NS}}{\text{NES}}$ experiment respectively (Figure 13.19). Methylation of histones is particularly important during differentiation. Many promoters of differentiation genes are bivalently marked, with activation (H3K4me3) and repressive (H3K27me3) marks. This poises them for transcription, but prevents their transcription in pluripotent cells. As the cells differentiate, one of the marks is removed, thus establishing the genes expression pattern, and thereby the fate of the cell. These demethylases may thus be playing an important role in neural stem cell fate specification

Two major complexes control the regulation of bivalent chromatin marks: the polycomb group PcG and the trithorax group (TrxG). The PcG was not enriched in the GO analysis in this study, as too few of its members were SDE. The majority that were SDE were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 13.20).

Figure 13.20: *STRINGdb network CC* ***PcG protein complex****. the halo colours indicate the overlaps in SDE between experiments, as per the legend*

The PcG and TrxG antagonise each other. The PcG represses transcription while the TrxG activates transcription. PcG is known to overlap with macroH2A already seen to be up-regulated in $_{\text{hf}}$NES cells. Within the PcG, the subunits *EZH2*, *MTF2*, and *SCML2* were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, whereas *HDAC2* and *RBBP4* were down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, and *CBX6* was up-regulated in GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.20).

SWI/SNF forms part of the TrxG. Within the SWI/SNF are esBAF and npBAF. A change from esBAF to npBAF was observed in $_{hf}$NS cells and discussed in detail in section 12.1.1.

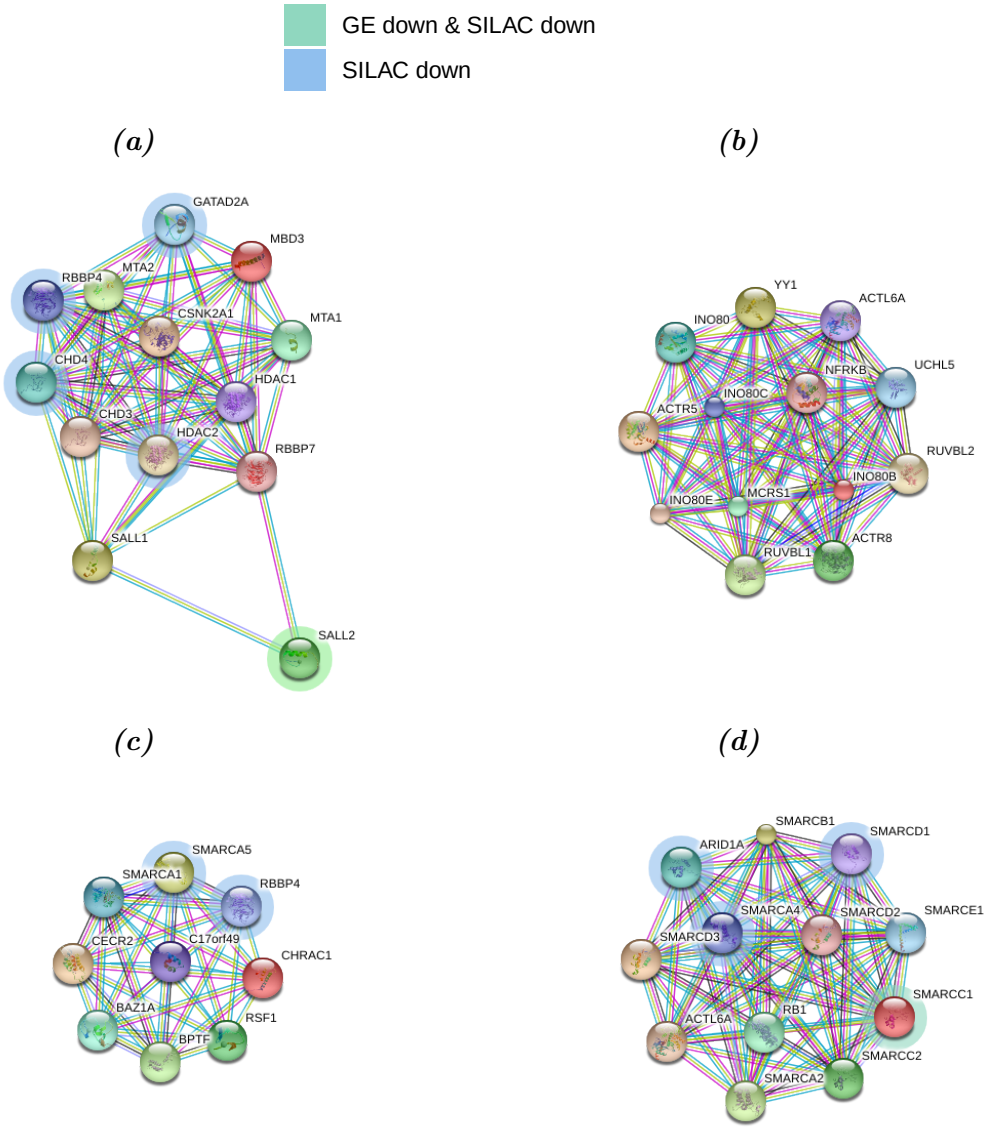


Figure 13.21: *STRINGdb network of the ATP-dependent chromatin remodelling complexes, (a)* ***CHD-type complex****, (b)* ***INO80 complex****, (c)* ***ISWI-type complex****, and (d)* ***SWI SNF complex****. The halo colours indicate the overlaps in SDE between experiments, as per the legend, molecules without halos were not SDE.*

Chromatin structure often requires energy in order to be changed. ATP-dependent chromating-remodelling complexes, change the chromatin structure by sliding nucleosomes, dissociating core histone and re-locating entire nucleosomes. There are four families of ATP-dependent chromatin-remodelling complexes: SWI/SNF, imitation switch (ISWI), chromod-

omain (CHD), and INO80 complex (INO80). Of these only SWI/SNF was significantly enriched according to GO enrichment analysis. However, within some of the others, there were key genes or proteins down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 13.21).

13.2.3 Non-coding RNA

Non-coding RNA (ncRNA) refers to RNAs transcribed from DNA that are not translated into proteins. Recently more and more regulation processes have been discovered for these RNAs. There are multiple different types of ncRNA. Because ncRNA were not investigated in this study, we do not have quantitative evidence of their presence or influence, but many proteins involved in the regulation of ncRNA were however dysregulated here and warrant discussion.

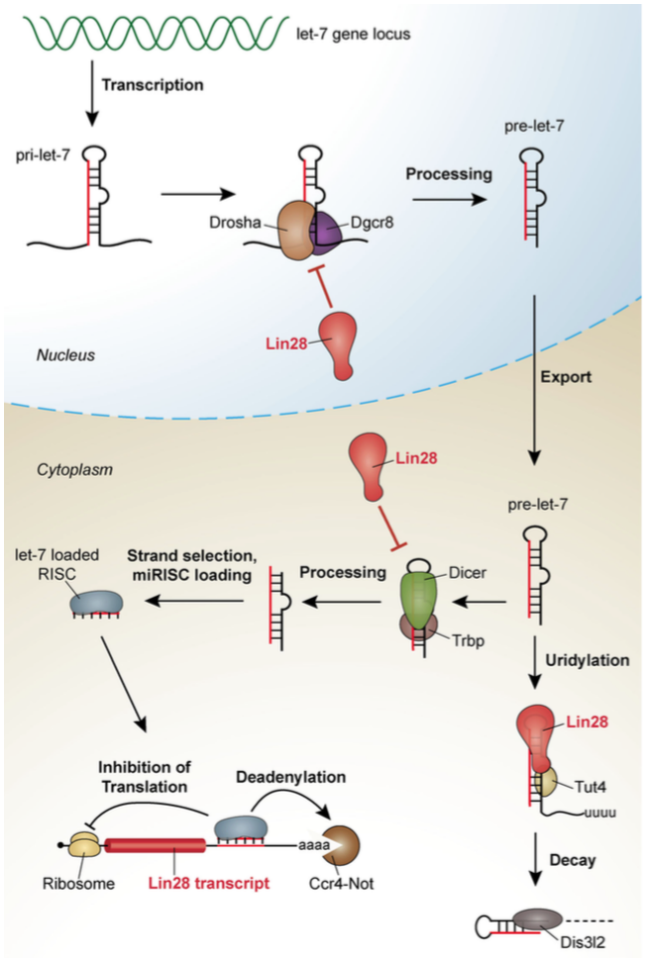


Figure 13.22: *Let-7 bio-genesis from Rehfeld et al. 2015.*

LIN28/*let-7* Micro RNA (miRNA) are produced in the nucleus as long pri-miRNA's. They are processed in the nucleus by *DROSHA* and *DGCR8* to approximately 70 nucleotide hairpin structures and then exported by *XPO5*. In the cytoplasm, *DICER1* cleaves them into the final active 22bp miRNA, after which they are loaded into argonaute (AGO) to induced silencing. miRNAs inhibit protein expression at the post-transcriptional level. miRNA, binds the complementary 3' untranslated region (UTR) regions of mRNA, thereby inhibiting translation and inducing degradation. Binding is often imperfect, allowing them to regulate many targets.

The miRNA *let-7* was one of the first miRNA to be discovered during the study of the heterochronic pathway that controls developmental timing in *Caenorhabditis elegans*.[179] There are 12 isoforms of *let-7* in vertebrates, and it is the most abundant miRNA in most adult tissues, including the brain.[174] In vertebrates levels of *let-7* are low in pluripotent cells and high in differentiated cells. It antagonises proliferation and supresses the self-renewal capacity of stem cells.[118] Levels of *let-7* increase rapidly as neural stem cells differentiated into neurons.[173]

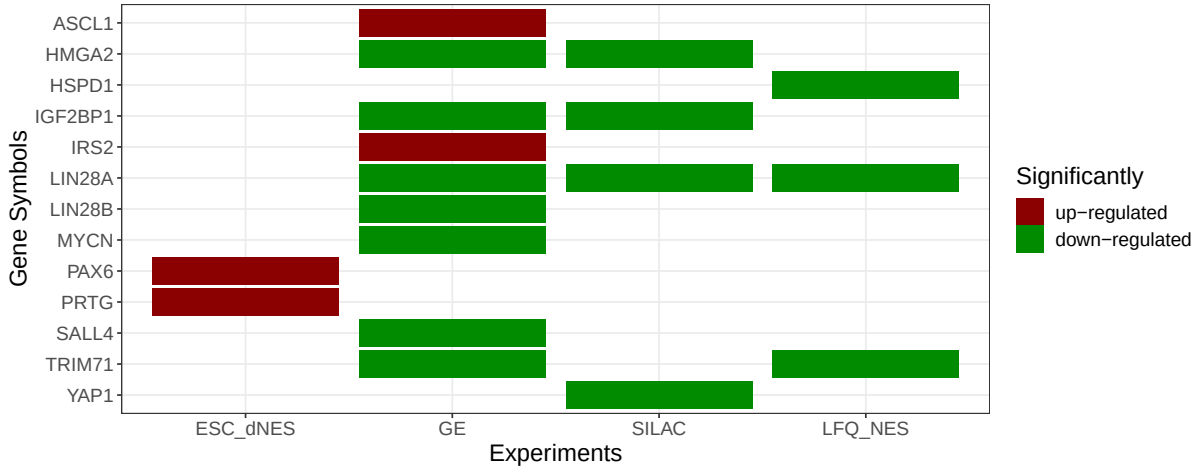


Figure 13.23: *Heatmap of the known let-7 targets from Rehfeld et al. 2014. Red and green indicates up- and down-regulation respectively.*

Among the known targets of *let-7*,[173] (*ASCL1*, *HMGA2*, *HSPD1*, *IGF1R*, *IGF2BP1*, *LIN28A*, *LIN28B*, *TRIM71*, *MYC*, *MYCN*, *PAX6*, *PRTG*, *SALL4*, and *TLX3*), most were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments and some were also down-regulated in the $_{\text{hf}}$NES cell differentiation experiment. Only *ASCL1* and *IRS2* was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 13.23).

Let-7 is regulated by *LIN28* and it in-turn represses *LIN28*. *LIN28* is an RNA-binding protein with two RNA-binding domains. It is highly expressed in the neural tube and co-localises with *SOX2* expression.[238] *LIN28* expression is high in ESCs and decreases during development of the embryo, most significantly in the brain during neurogenesis. *In vivo* expression levels in mice start declining at embryonic day 11.5 (Carnegie stage 13), the timepoint at which the transition to RG begins, and it is virtually absent by embryonic day 14.5 (Carnegie stage 20).[47] RG do not express *LIN28* and its expression has been shown to regulate neurogenesis and gliogenesis balance in early neural progenitors *in vitro*.[19] *LIN28* has also been shown to be down-regulated by miR-145 which is upregulated during neuronal differentiation. Suppression of neuronal differentiation by suppression of miR-145 by anti-miR0145 can also be recovered through silencing of *LIN28*, further supporting *LIN28*'s role in suppressing neural stem cell differentiation.[139] When *LIN28* was overexpressed in mouse primary embryonic cortex at Carnegie stage 23 and then harvested, mainly neurons survived and these showed suppressed neurite outgrowth and increased expression of *HMGA2*.[31] This indicates that *LIN28* impairs neuronal maturation.

LIN28 binds to *pre-let-7*, inhibiting maturation of *let-7* and maintaining pluripotency of cells. There is a direct correlation between the reduction of *LIN28* expression, an increase of *let-7*, and an increase in neurogenesis. In the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, there was a significant reduction in *LIN28A* during transition from $_{\text{hf}}$NES cells to $_{\text{hf}}$NS cells and a subsequent reduction during differentiation of $_{\text{hf}}$NES cells. *LIN28B* showed a similar trend, however it was not statistically significant, because of high variability and limited detection in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment. In the GE $\frac{\text{NS}}{\text{NES}}$ experiment, both *LIN28A* and *LIN28B* were significantly down-regulated (Figure 9.6 and Figure 9.7).

LIN28 expression is induced by *SOX2*, which was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.4). *LIN28* induces expression of *MYCN*, which was down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.9). Proteins associated with *LIN28* and those inhibited by *let-7* were mostly down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments. The most notable *LIN28*-associated proteins that were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ and SILAC $\frac{\text{NS}}{\text{NES}}$ experi-

ments were *CDK1*, *HMGA2* , *IGF2BP1*, *LIN28A*, *LIN28B*, and *YAP1* (Figure 13.24). The co-expression of *LIN28* inducers and its effectors strongly support the likelihood that *LIN28* is a key player in the differences observed in cellular identity between $_{hf}$NES cells and $_{hf}$NS cells.

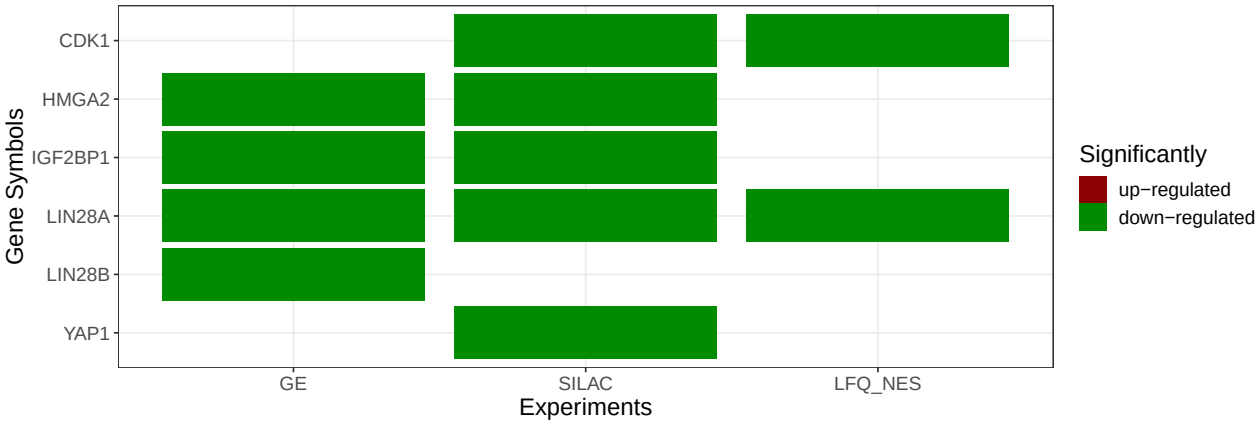


Figure 13.24: *Heatmap of LIN28 associated proteins that were SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

YAP1, the Hippo pathway effector is up-regulated by *SOX2*.[182] It is responsible for maintaining organ size, it is highly expressed in neural stem cells, and suppresses differentiation.[236,243] Knockdown of *YAP1* has been shown to aberrantly increase *let-7*.[46] Here it was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, (Figure 11.35) it may be a significant effector in maintaining the stemness of $_{hf}$NES cells, it and its effectors should be explored further. The Hippo pathway was discussed in section 11.3.1.

13.3 Conclusion

In this study, key components of the three epigenetic mechanisms were shown to be more highly expressed in $_{hf}$NES cells. This would suggest that these mechanisms were more active in $_{hf}$NES cells. The core pluripotent transcription factors were reduced during transition from ESC to $_{hf}$NES cells, except for *SOX2* and *LIN28*. *LIN28* expression maintains inhibition of *let-7*, thereby promoting self-renewal and proliferation of the $_{hf}$NES cells. The *de novo*

DNA methylases were more highly expressed in $_{hf}$NES cells, suggesting that they are silencing specific genes that are not longer required. Extensive chromatin remodelling may be occurring as there were significant changes in histone composition of nucleosomes as well an up-regulation of chromatin-remodelling complexes.

Inhibition of *let-7* may permit the $_{hf}$NES cells to maintain some of their stemness, while the DNA and chromatin is being modified. Once *LIN28* and *SOX2* expression reduces, the self-renewing factors would be reduced and the full effect of the epigenetic changes would come into effect, bringing about the transition from $_{hf}$NES cells to $_{hf}$NS cells.

The mechanisms for altering epigenetic control are broad sweeping mechanisms. The specificity of these mechanisms is controlled by transcription factors. The transcription factors recruit DNA methylation and chromatin-modifying enzymes to specific regions of the genome. These regions can then be marked with activation or repressive marks. Despite the epigenetic complexes making the changes it is really the transcription factors that decide the specificity of where and how these changes, will be implemented. The transcription factors will therefore be discussed next.

Chapter 14

Transcription factors

14.1 Pioneer transcription factors

Pioneer factors are transcription factors that are able to bind to unmarked closed chromatin compacted by H1 linker histones. Closed chromatin is usually inaccessible to transcription factors. Pioneer factors adopt conformations that allow them to bind to DNA wrapped around histones through their affinity for histones. Co-factors are then recruited to open the chromatin and make it accessible for transcription, often through the displacement of H1 linker histones. Pioneer factors make changes to specific areas of compacted DNA to prepare cells for cell fate transition. In development, this is as a precursor step for differentiation.

A pioneer factor will facilitate the opening of chromatin and then additional transcription factors will initiate transcription. These two processes are complimentary. Without the pioneer factors, the chromatin will remain closed and additional transcription factors will have no effect. Without the additional transcription factors, pioneer factors themselves are mostly unable to initiate transcription.[104,140]

During cell reprogramming, a cocktail of factors are used to produce iPSCs or convert one cell type directly to another through direct reprogramming.[90] Such reprogramming cocktails must contain at least one pioneer factor. This pioneer factor gains access to chromatin, which is otherwise inaccessible to other factors. The other factors then drive cell reprogramming

in the required direction. To induce pluripotency, the pioneer factors *POU5F1*, *SOX2*, and *KLF4* are activated, together with *MYC*.

Morris (2016) proposed that genes are organised in gene regulatory networks. These gene regulatory networks are activated in a hierarchical manner. ESCs are sparsely specified. During gastrulation, body plan gene regulatory networks (GRNs) are activated, followed by activation of the body part, body pattern, cell fate, and finally terminal differentiation GRNs. Pioneer factors are responsible for the opening, and in some cases condensing, of chromatin during development. The changes in chromatin can be visually observed as areas of different density in the nucleus. There is a transition of compartments during differentiation. During mouse ESC differentiation to neural progenitors, a 36% change in compartments was observed and this was confirmed in human ESCs.[90]

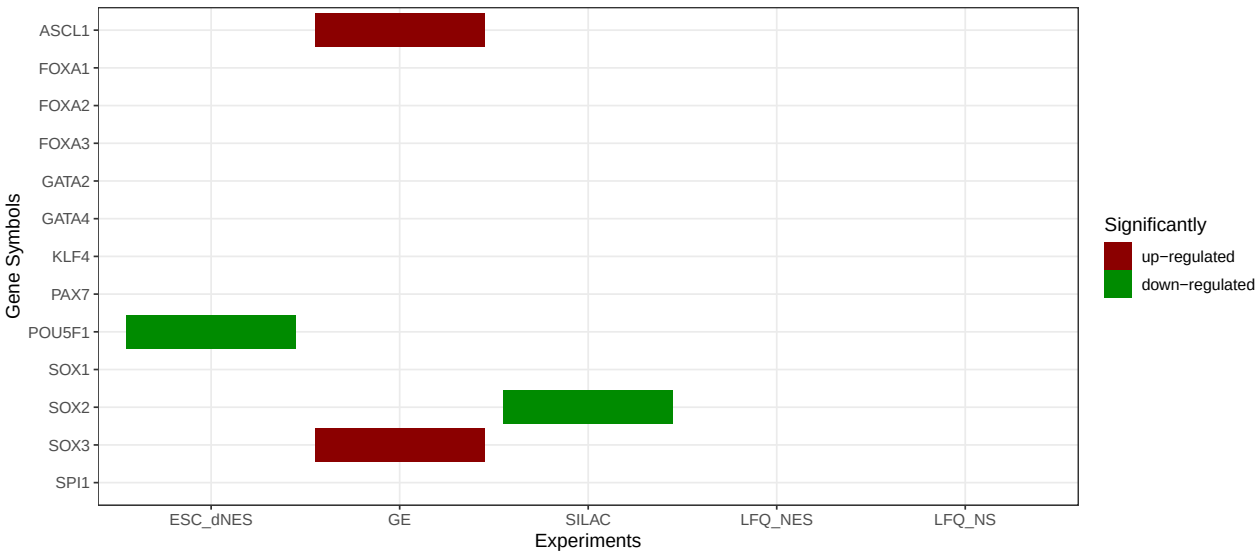


Figure 14.1: *Heatmap of SDE pioneer factors from the literature*[90,103,240]

From the thirteen pioneers factors identified in the literature,[90,103,240] four were SDE across the Huang $\frac{\text{dNES}}{\text{hESC}}$, GE $\frac{\text{NS}}{\text{NES}}$, and SILAC $\frac{\text{NS}}{\text{NES}}$ experiments (*ASCL1*, *POU5F1*, *SOX2*, and *SOX3*) (Figure 14.1).

ASCL1 is used in direct-reprogramming to produce neurons from somatic cells.[140] It was up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 14.1) and may be one of the pivotal driving

forces in turning $_{hf}$NES cells into $_{hf}$NS cells. It has already been discussed in section 9.3.4

14.2 SOX transcription factors

SOX transcription factor's play critical roles during early development, controlling the formation of tissues and organs. Because they do not bind DNA with high enough affinity to initiate transcription, they form dimers with other transcription factors thereby extending their functions.[30] SOX transcription factors are therefore able to play multiple roles in different contexts depending on the expression of their binding partners. Changes in SOX expression have a significant effect on cellular identity.[129]

SOX2 binds genes mainly expressed in ESCs. Genes bound by *SOX2* and *SOX3* are mainly expressed in neural stem cells. Genes only bound by *SOX3* are primarily expressed in neurons. These genes are then activated by *SOX11* binding, which usually displays an expression pattern inverse to that of *SOX3*.[30]

In this study, we observed a distinct difference in expression of SOX transcription factors in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, but because of the low levels of expression, few SOX proteins were detected in the proteomic datasets.

SOX2 was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 9.4) as it is more highly expressed in $_{hf}$NES cells, thereby maintaining their stemness and promoting proliferation. *SOX2* was down-regulated during $_{hf}$NES cell differentiation, although this was not statistically significant.

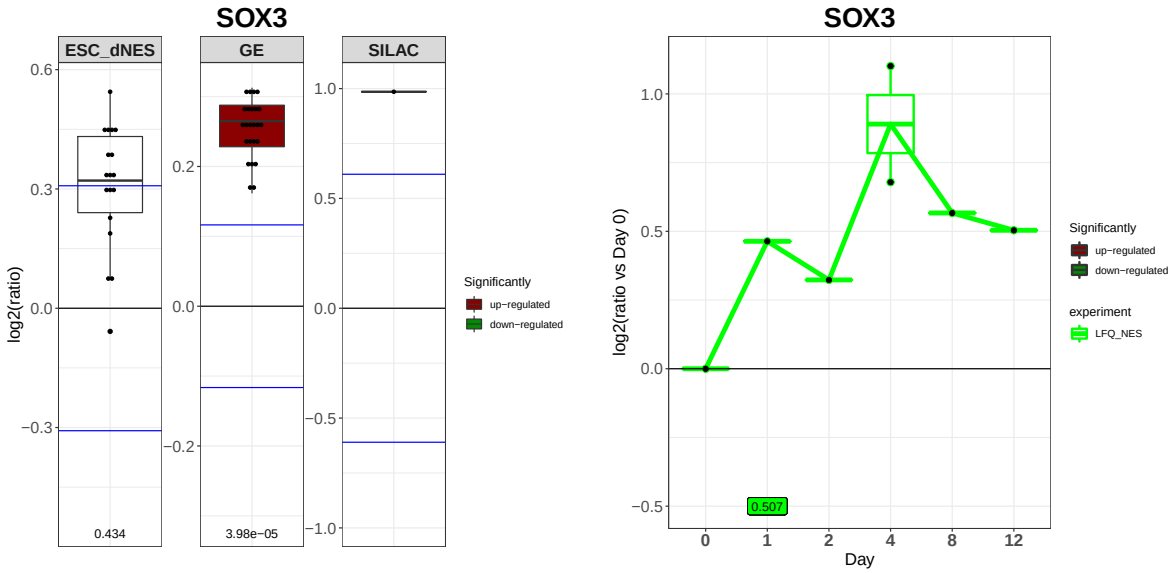


Figure 14.2: *Boxplots comparing the log2 ratios for SOX3. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

SOX3 showed an upward trend in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, although this was not statistically significant, and was then up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment and during $_{\text{hf}}$NES cell differentiation (Figure 14.2). Its up-regulation would confer neuronal identity upon the $_{\text{hf}}$NES cells and even more so upon the $_{\text{hf}}$NS cells.

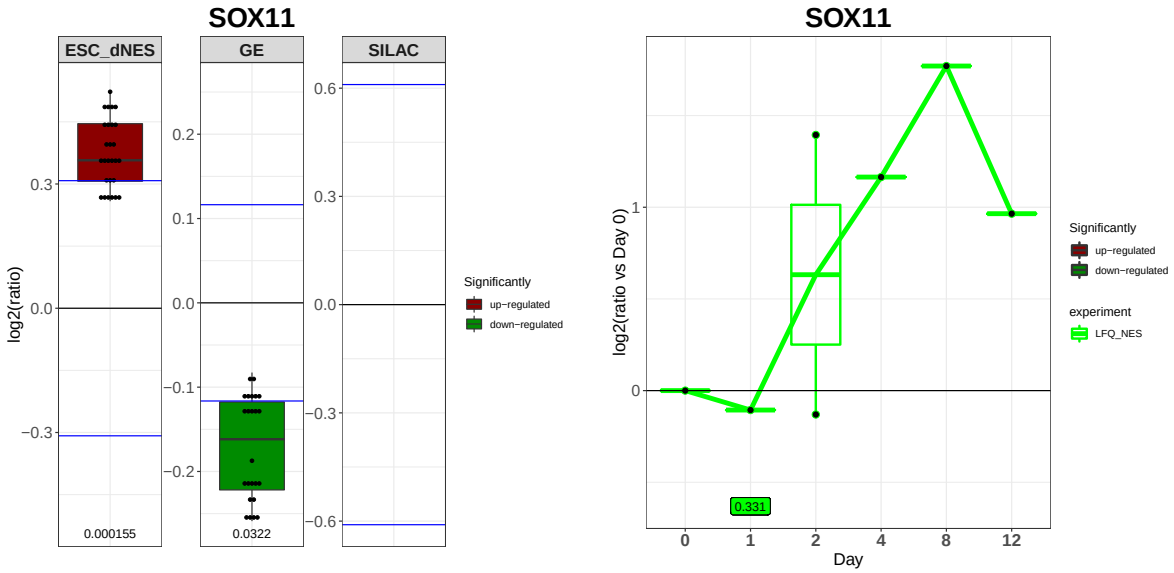


Figure 14.3: *Boxplots comparing the log2 ratios for SOX11. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

SOX11 was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment and down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 14.3). This may mean that it is instrumental in initiating the neuronal program within $_{\text{hf}}$NES cells, but may not be as necessary in the $_{\text{hf}}$NS cells. However, it was once again up-regulated in the $_{\text{hf}}$NES cells differentiation experiment, as it may be required to activate neuronal genes during differentiation.

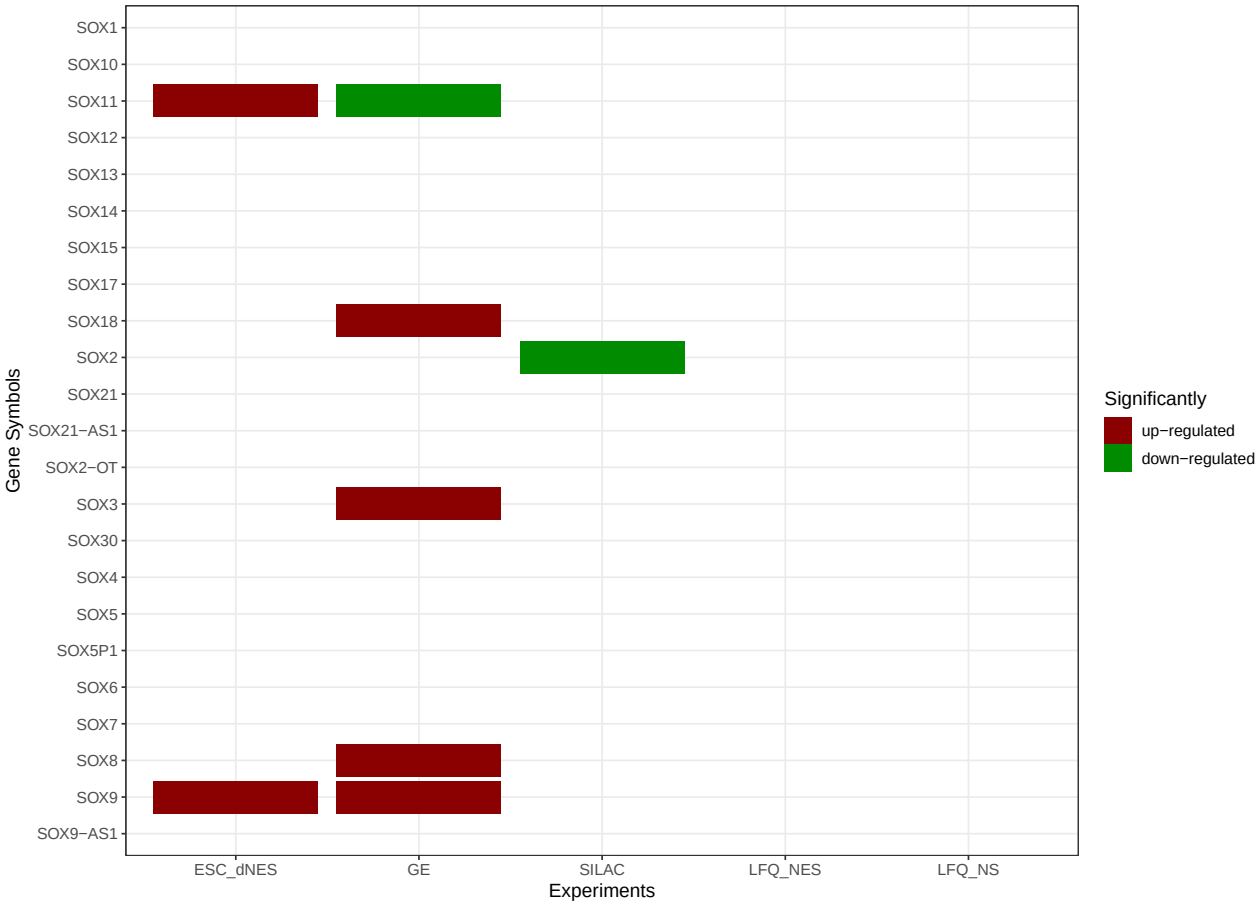


Figure 14.4: *Heatmap of the expression of all SOX2 genes and proteins across all experiments. Red and green indicates up- and down-regulation respectively.*

SOX9 was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and GE $\frac{\text{NS}}{\text{NES}}$ experiments, and *SOX8* and *SOX18* were up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiments (Figure 14.4). The involvement of *SOX3*, *SOX8*, *SOX9*, and *SOX11* in neural stem cell development has previously been reported, but no connection has been made between *SOX18* and neural stem cells as of yet.[4]

SOX2 regulates *LIN28*[47] and *YAP1*[182] both of which show expression patterns that are correlated with that of *SOX2*. *SOX2* forms part of three BPs significantly enriched in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment, namely ***stem cell development***, ***regulation of neurogenesis***, and ***endodermal cell fate commitment*** (Figure 14.5). It is therefore supported by other molecules to effect its function in maintaining development of the neuroepithelium.

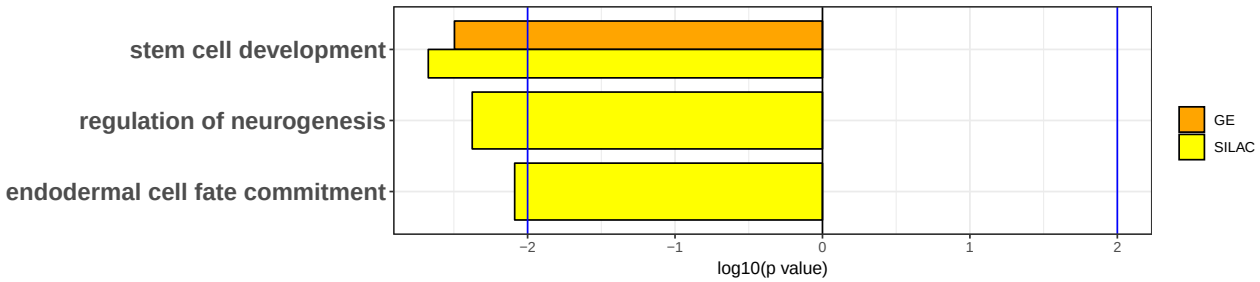


Figure 14.5: *SOX2 was significantly down-regulated in the SILAC $\frac{NS}{NES}$ experiment. These BP terms are significantly enriched and concern SOX2.*

14.3 POU transcription factor family

SOX2 appears to be the principal driver of pluripotency in $_{hf}$NES cells, as it is widely expressed in the neural plate soon after gastrulation, which is the origin of $_{hf}$NES cells. As *SOX2* itself does not bind DNA strongly enough to effect change, it needs to work in complex with other transcriptions factors, and its versatility in association with different transcription factors enables them to perform different roles in different contexts. During nervous system development, it complexes most favourably with the Pit-Oct-Unc domain transcription factors (POUs).

In ESCs, *SOX2* forms a complex with OCT4/*POU5F1*,[129] and in neural progenitors, *SOX2* forms a complex with BRN2/*POU3F2*.[129]

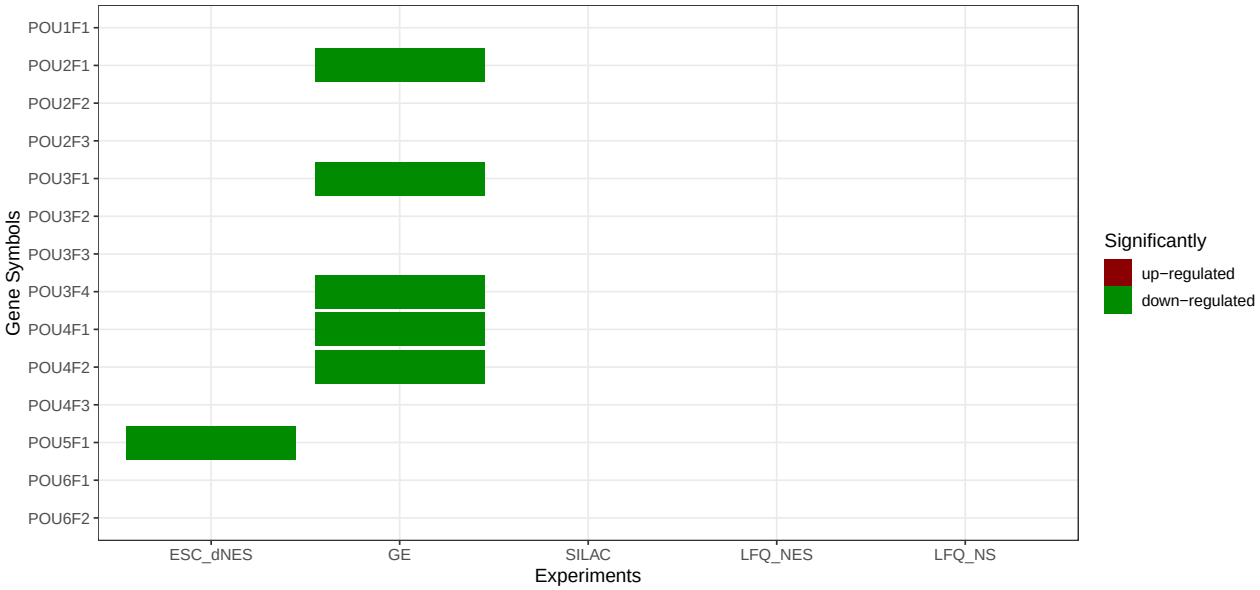


Figure 14.6: *A heatmap of POU transcription factors SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

POU5F1 (OCT4), a core pluripotent factor, was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.3). This would account for a reduction in pluripotency as these cells move towards a more differentiated state. *POU2F1*, *POU3F1*, *POU3F4*, *POU4F1*, and *POU4F2* were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 14.6), and may be important in maintaining $_{\text{hf}}$NES cells.

Despite not being SDE, *POU3F2*, showed substantial differential expression (Figure 9.5). It appeared to be slightly more highly expressed in $_{\text{hf}}$NES cells than in hESC and also slightly more highly expressed in $_{\text{hf}}$NES cells than in $_{\text{hf}}$NS cells. It was also detected in the differentiation experiments, indicating that it was fairly highly expressed in these cells despite not showing much actual change. It is a marker of early neural commitment[129] and co-operates with *SOX2* to promote a neural fate. When its expression was activated in ESCs using light pulses it promoted their differentiation into neural tissue.[192] *POU3F2* also works in conjunction with *ASCL1* and *MYT1L* to reprogram fibroblasts to a neural fate.[213]

14.4 Developmental transcription factor families

14.4.1 Homeobox transcription factor family

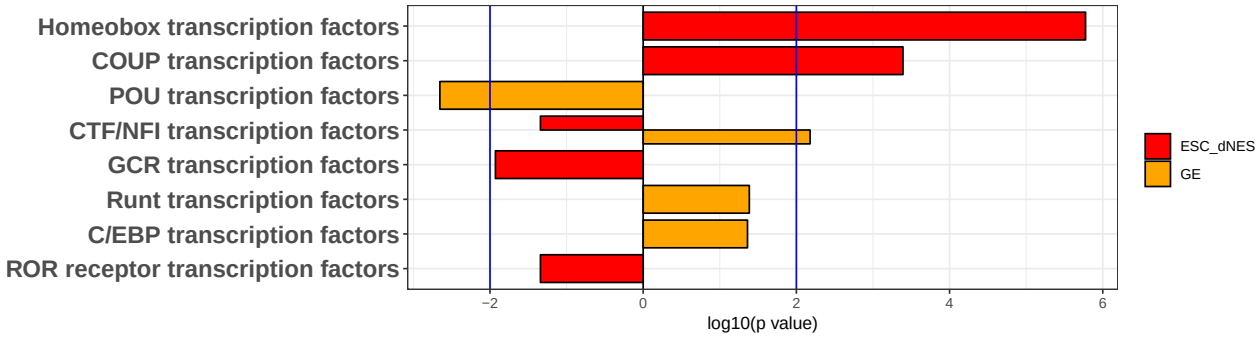


Figure 14.7: *Transcription factor families from the AnimalTFDB significantly enriched by Fisher exact test in the Huang $\frac{dNES}{hESC}$ and GE $\frac{NS}{NES}$ experiments, with a p-value less than 0.05. The horizontal blue lines indicate a p-value threshold of 0.01. ESC_dNES (red) refers to the Huang $\frac{dNES}{hESC}$ experiment, GE (orange) refers to the GE $\frac{NS}{NES}$ experiment, SILAC (yellow) refers to the SILAC $\frac{NS}{NES}$ experiment, LFQ_NES (green) refers to the ${}_{hf}NES$ cell differentiation experiment, and LFQ_NS (blue) refers to the ${}_{hf}NS$ cell differentiation experiment.*

Members of the homeobox transcription factors family are mainly responsible for directing the development of structures during embryo development. The prefix homeo, originated from homeosis, as when these genes are mutated, the results is often the replacement of discrete organs with others. The family is larger than just the commonly known HOX genes that control segmentation patterning in vertebrates. Enrichment analysis was done by Fisher exact test against the transcription factors family lists from the AnimalTFDB. The homeobox transcription factor family was the most significantly enriched transcription factor family (Figure 14.7).

ADNP is a homeobox transcription factor that is vital for brain formation[157] and is highly expressed in the embryonic brain during development and later throughout the nervous system.[134] Deficiency leads to defects in neural tube closures and early embryonic death.[134] *In vitro*, it has been shown to produce neurite outgrowth, as it protects neurons and is involved in microtubule re-organisation, through interaction with tubulins.[157]

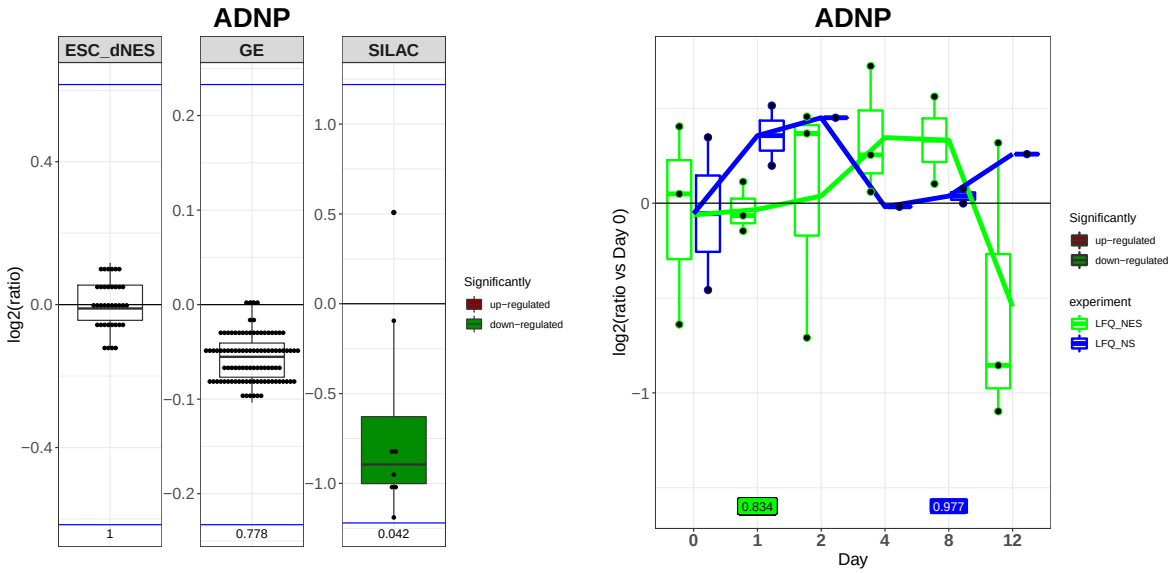


Figure 14.8: *Boxplots comparing the log2 ratios for ADNP. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

ADNP was down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment and its expression was consistent in the differentiation experiments (Figure 14.8). It showed little variance in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment, and this indicates that it is potentially already expressed in ESCs and may be one of key molecules instigating the default model of ESC differentiation into a neural fate. Down-regulation in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment indicates that it may be less necessary once this trajectory is established. However, expression appeared to be maintained during differentiation, perhaps with a slight decrease as the $_{\text{hf}}$NES cells mature into neurons (Figure 14.8).

Figure 14.9: *Heatmap of the genes and proteins form AnimalTFDB that were SDE in the Huang $\frac{dNES}{hESC}$, GE $\frac{NS}{NES}$, and SILAC $\frac{NS}{NES}$ experiments. Red and green indicates up- and down-regulation respectively.*

HOX genes are responsible for patterning in the embryo, and are usually expressed in defined areas along the anterior/posterior axis of the embryo. *HOXA2*, *HOXA4*, *HOXA5*, *HOXB2*, *HOXB3*, and *HOXB5* were all up-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment (Figure 14.9), indicating that their effect on embryo patterning is more important in the $_{\text{hf}}$NS cells than the $_{\text{hf}}$NES cells.

EN2 and *GBX2* are both known to be involved in hind-brain patterning and both were down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment. *GBX2* was also down-regulated in the SILAC $\frac{\text{NS}}{\text{NES}}$ experiment. These expression patterns indicate that their roles in neural development may be more important in $_{\text{hf}}$NES cells than $_{\text{hf}}$NS cells.

TALE transcription factors The three amino acid loop extension (TALE) transcription factors are important co-factors to the HOX proteins.[104] They consist of six transcription families, of which iroquois (IRX), myeloid ectropic viral integration site (MEIS), pre-B-cell leukemia homeobox (PBX), and PBX/knotted homeobox (PKNOX) were SDE in the Huang $\frac{\text{dNES}}{\text{hESC}}$ and GE $\frac{\text{NS}}{\text{NES}}$ experiments. *IRX3* is known to be involved in *SHH*-mediated neural patterning. *MEIS2* works in conjunction with *ASCL1* for direct reprogramming of fibroblasts into neurons. *PBX1* is a potential pioneer factor,[33] involved in chromatin remodelling.[93] Further study of the dynamics of the individual genes detected as SDE may provide important insights into their roles in the fate specification of neural stem cells.

Many of the remaining homeobox transcription factors SDE in this work have previously been shown to be involved in neurogenesis or astrogliogenesis.[222] They have not been discussed in this study, but further investigation into their effect upon this cell model may provide further insight into the control mechanisms at play here.

14.4.2 COUP transcription factors

Nuclear receptor subfamily 2 or COUP transcription factors I/II (NR2) are known to regulate neural stem cell temporal specification. *In vivo*, expression begins at Carnegie stage 16 and diminishes with development. When NR2s are knocked down *in vitro*, the neurogenic potential of neurospheres, which usually declines with successive passages, is maintained.[143]

NR2F1 was up-regulated in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment but down-regulated in the GE $\frac{\text{NS}}{\text{NES}}$ experiment, and this was further verified by one SILAC $\frac{\text{NS}}{\text{NES}}$ experiment sample (Figure 9.20). It also showed a strong upwards trend during $_{\text{hf}}$NES cell differentiation. *NR2F2* was up-regulated only in the Huang $\frac{\text{dNES}}{\text{hESC}}$ experiment (Figure 9.21).

There is evidence for their involvement in changes in *GFAP* promoter methylation from poised to activated.[143] They may in turn affect other gliogenic genes as well. This opens up expression of these genes to extrinsic stimulation, which may enable gliogenic potential.

14.4.3 CTN/NFI

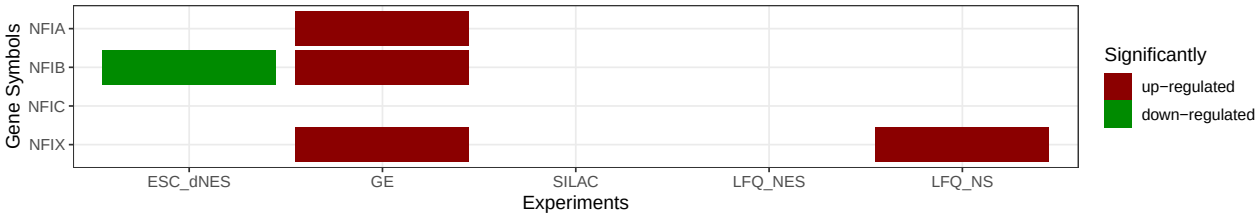


Figure 14.10: *Heatmap of CTN/NFI transcription factor family SDE across all experiments. Red and green indicates up- and down-regulation respectively.*

NFIA and *NFIB* are induced at the onset of gliogenesis.[143] Furthermore Tiwari et al. 2018 showed that *NFIA* and *ATF3* are drivers of astrogliogenesis,[222] by effecting changes to chromatin remodelling, thereby making poised gliogenic genes active for transcription. *RUNX2* was also shown to be involved in the maturation of astrocytes.[222]

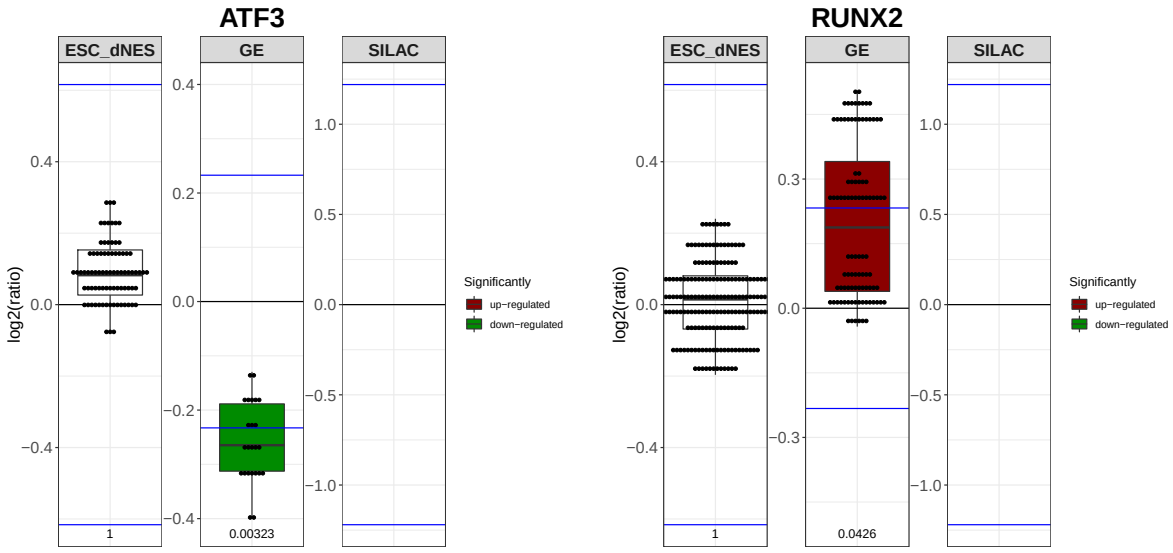


Figure 14.11: *Boxplots comparing the log2 ratios for ATF3 and RUNX2. Each experiment has its own scale and the horizontal blue lines indicate the 2 standard deviation (σ) thresholds for the respective experiment. Red and green filled boxplots indicate significant up- or down-regulation respectively. The p-values are included on the bottom of the plots.*

Chapter 15

Conclusion and future work

During nervous system development neuroepithelial cells proliferate rapidly to increase the cell mass required for the nervous system development. These neuroepithelial cells are able to self renew or differentiate into neurons. As development progresses neuroepithelial cells develop into RG. RG are able to self-renew and further add to the cell mass required for development of the nervous system. Unlike the neuroepithelial cells, RG are tri-potent, thereby able to form neurons, astrocytes and oligodendrocytes. RG also go on to become the adult neural stem cells responsible for maintaining the population of glial cells in the brain. No new neurons are formed in most areas of the brain after birth, except in the SVZ and SGZ.[80,137,194] These neurogenic niches create environments which allow the adult neural stem cells to still take on neuronal fates. The property is not inherent to the cells within these areas, but rather due the micro environment created within these areas.[48,99,106,107,122,123,194] This means that all adult neural stem cells maintain the ability to differentiate into neurons, but the environment of the adult brain suppresses this ability.

In this study we had access to human foetal neural stem cells derived from five week old hindbrain ($_{hf}$NES cells) and from 9 week old hindbrain ($_{hf}$NS cells). The earlier timepoint resembled neuroepithelial cells and differentiated into an almost pure population of neurons, while the later timepoint resembled RG and differentiated into a mixed population of neurons and astrocytes.

A quantitative proteomic study was conducted comparing the relative abundance of proteins being expressed in each of these cell lines while under proliferative conditions. Removal of growth factor stimulation allowed the cells to spontaneously differentiate. Six timepoints of early differentiation were isolated and a quantitative proteomics analysis was performed. The aim of the study was to determine the control mechanisms at play that distinguished $_{hf}$NES and $_{hf}$NS cells, with the hopes of discovering druggable targets, that could be used for brain regenerative therapies.

Due to the subtle nature of the mechanisms at play here, it became evident that the depth of coverage of proteomics analysis was not sufficient to deduce the answers required. Fortunately the Smith lab who had derived the $_{hf}$NES and $_{hf}$NS cells had done a transcriptomic analysis of these cells,[211] which they kindly allowed us to use.

The cells from these two studies came from the same source, but had been cultured at different times and in different labs. Interestingly there was a good overlap between the two datasets. There were numerous genes and proteins found to be uniquely SDE in each of the datasets, but there were no contradictions in those that were SDE. Not being SDE versus being SDE was not seen as a contradiction, because often the trends, despite not being statistically significant, were also correlated. The increased numbers of genes observed to be SDE in the transcriptomic analysis resulted in a more detailed GO and KEGG pathway enrichment. This allowed more detailed conclusions to be derived from the data. The proteomics data was then used as validation of the transcriptomics data, rather than the basis of the analysis as was previously intended.

While trying to make sense of the wealth of information gleaned from the datasets, it also became evident that understanding the differences between $_{hf}$NES and $_{hf}$NS cells was confounded by not knowing anything about the trajectory that the $_{hf}$NES cells were on prior to differentiating into $_{hf}$NS cells. What had essentially been created was a data silo; data standing independently of an origin. Fortunately dNES cells have previously been derived in vitro from hESC, a dataset by Huang et al. 2016 contained a transcriptomic analysis of these cells. This dataset was then added to the analysis as a means to establish a possible trajectory

of $_{hf}$NES cell differentiation.

The same statistical evaluation was applied to all three datasets in order to make them most comparable. The data were analysed using R scripts wrapt in a R Shiny dashboard. R Shiny produces a browser based interactive interface to R code. The data could then be navigated graphically without needing to deal with complex code. A link to the Shiny interface is available at https://www.shaungarnett.co.za/thesis.html.

In addition to the Huang $\frac{dNES}{hESC}$, GE $\frac{NS}{NES}$, and SILAC $\frac{NS}{NES}$ experiments, the $_{hf}$NES and $_{hf}$NS cell differentiation experiments were also added to the Shiny interface. Thus far the differentiation experiment was merely used to substantiate the findings observed in the GE $\frac{NS}{NES}$ and SILAC $\frac{NS}{NES}$ experiments. The differentiation dataset warrants it's own detailed independent assessment, which will be done in a subsequent publication.

$_{hf}$NES cells appear to differentiate into an almost pure population of neurons, while $_{hf}$NS cells differentiate into a mixed population of neurons and glia. The focus of the study was to identify molecules that would direct neural stem cells towards a neuronal fate rather than a glial fate. As a premise to this search, it was assumed that molecules conferring cellular identity on the $_{hf}$NES cells, opposed to the $_{hf}$NS cells, would be good candidates. Therefore, the main focus of the analysis was to understand what maintained the cellular state of $_{hf}$NES cells.

With 1817 SDE genes and 473 SDE proteins, there were too many molecules to try and deduce meaning through individual investigation. In order to identify important molecules warranting further study, the lists of SDE genes and proteins were analysed by gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment. This produced an overarching characterisation of the cells through, cell, tissue, cellular component (CC), molecular function (MF), biological process (BP) and KEGG pathway enrichment. Once major processes were identified, the SDE genes were visualised in STRINGdb networks to identify those with the most interconnectedness within these processes. These molecules were then researched in literature to understand the connections drawn between them. Many of the genes traditionally implicated in literature as inferring neuronal fate were not SDE in the

comparison of $_{hf}$NES and $_{hf}$NS cells, due to both of the cell lines being neural stem cells. Some of these were up-regulated in Huang $\frac{dNES}{hESC}$ experiment, indicating that they were possibly highly expressed in both cells and therefore not SDE between them.

15.0.1 Summary of Findings

During the analysis it became evident that $_{hf}$NES and $_{hf}$NS cells were controlled by vastly different mechanisms. The $_{hf}$NES cells were controlled by the intrinsic default neural differentiation program of ESCs. Key pluripotent factors *SOX2* and *LIN28* were maintaining these cells in a more pluripotent state than the $_{hf}$NS cells. In conjunction with the higher expression of these pluripotent factors, there were also higher expressions of DNA methylating and chromatin modifying proteins; namely *DNMT1*, *DNMT3B*, *HELLS*, *TET1*, *H2AFY2*, *EZH2*, *KDM3B* and *KDM4D*. These could be executing epigenetic changes to the $_{hf}$NES cells genome, while the $_{hf}$NES cells may be maintained in their current state, through the expression of the pluripotent factors *SOX2* and *LIN28*. A number of other transcription factors are also at play to maintain the $_{hf}$NES cells cellular identity, *ATF3* *HMGA2*, *NR2F1*, and *PBX1*, and warrant further study.

The molecules that are likely the most instructive in maintaining the $_{hf}$NES cells cellular identity are *LIN28A* and *LIN28B*. These molecules suppress the miRNA *let-7*, which is instrumental in neural stem cell differentiation. *LIN28* has previously been shown the be reduced during nervous system development. Here we show clear persistence of expression in $_{hf}$NES cells. In addition to this, the Hippo pathway appeared to be active in the $_{hf}$NES cells. It may be the main control mechanism maintaining proliferation of the $_{hf}$NES cells, until sufficient cells mass has been created, and before induction of differentiation into $_{hf}$NS cells. Supporting this, the Hippo effector *YAP1* was more highly expressed in the $_{hf}$NES cells.

Once the expression of the pluripotent factors is reduced, the full effect of the epigenetic changes would take effect, and the cells would then be established as $_{hf}$NS cells. Furthermore the transcription factors up-regulated in $_{hf}$NS cells *PAX3*, *ASCL1*, *HES4*, *HES5*, *INHBB*, *NFIA*, *NFIB*, *RUNX2* and *SOX3* may further induce the transition from $_{hf}$NES cells to $_{hf}$NS

cells. It was also evident from the GO and KEGG analysis that the $_{hf}$NS cells were more complex in their structure, metabolism and control mechanisms than the $_{hf}$NES cells. The plasma membrane of the $_{hf}$NS cells undergoes significant changes as does the cytoskeleton. This indicates a greater ability of these cells to interact with their environment. A number of signalling pathways were enriched in the $_{hf}$NS cells (Figure 11.26). This indicates that they are able to interact more readily with their environment and transduce those signals more readily within the cells, thereby being more responsive to their environment. As the nervous system develops the nervous tissue would become increasingly complex, integration into this complex tissue would require cells to become more responsive to their environment.

15.0.2 Novelty

Much of the information attained for neural tissue has been produced in mice and rats or human neural stem cells derived from ESC. This study makes distinct comparisons between human foetal neuroepithelial-like stem cell lines ($_{hf}$NES cells) and human foetal neural stem cell-like cell lines ($_{hf}$NS cells). This is the most comprehensive comparison of these newly derived cell lines to date. This study validates that discoveries made in animal models are also at play in human cells derived directly from human foetal brain. It also produced novel information for many of these molecules. Previously these molecules had only been observed to change between ESC and neural stem cell. Here we observe more precise changes within two neural stem cells populations, $_{hf}$NES and $_{hf}$NS cells. There are also a number of SDE molecules within the dataset for which there is very little information in literature, so they were not discussed, They may therefore be playing important roles in these cells, but are as of yet uncharacterised, and warrant further study, such as *POLQ*

15.0.3 Future Work

This study has identified a number of candidate molecules that play roles in neural stem cell fate specification. In order to validate and fully understand their effects, detailed biochemical

characterisation of individual molecules would need to be performed.

Manipulating the expression levels of the identified $_{hf}$NES and $_{hf}$NS cell producing transcriptions factors (Section 14) in either of the cell lines and observing the effect on their cellular identity and differentiation potential may yield interesting information as to their roles in neural stem cell development and differentiation. Expression perturbations that would most likely yield the greatest results would be that of *SOX2*, *LIN28A* and *YAP1*.

With the differences in nucleosomes (Figure 13.10), chromatin-remodelling (Figure 13.18) and DNA methylation (Section 13.2.1) observed between $_{hf}$NES cells than $_{hf}$NS cells, this cell model would be a good place to do follow-up epigenetic studies. Understanding the changes in DNA methylation and chromatin compaction in these two cell types together with the corresponding gene expression differences may offer valuable insights into the epigenetic mechanisms at play here. Manipulating expression of key components in these mechanisms may also provide more detailed information on there individual roles within these cells. In particular, altered expression of *DNMT1*, *DNMT3B*, *H2AFY2*, *EZH2* and *KDM3B* would be most interesting.

Using this cell model there is also potential to learn more about the nature of the $_{hf}$NS cells. Detailed analysis of the signalling pathways enriched in the $_{hf}$NS cells (Figure 11.26), may yield information on how these cells decide to become neurons or glia, and the effect that activation of these signalling pathways have on the $_{hf}$NS cells.

The $_{hf}$NS cells up-regulate numerous proteins that form part of the plasma membrane, most notably those for focal adhesion (Figure 11.19). Furthermore understanding the changes to the cell surface could be used to enhance cell culture conditions as well as identify new markers to more clearly distinguish between these cells. The collagens and lamins (Figure 11.18) SDE in this data could be used to design more appropriate tissue culture environments, which would be better suited to directing the differentiation of iPSC. *FABP7*, *SLC1A3* and *DPYSL2* are often used as cell markers. There were a number of fatty acid binding proteins, solute carriers (Figure 9.30) and dihydropyrimidinase like proteins (Figure 9.48) SDE in this data. For many of these there is limited information in literature pertaining to neural stem cells, and they may

potentially be novel markers for these cells.

There were a number of other molecules which were SDE, but whose functionality fell outside of the scope of this study. Within the information held in this study, there may also be information that could increase understanding of neurological developmental disorders. Defective proteins more highly expressed in $_{hf}$NES cells may explain reduced brain size, while defective proteins in $_{hf}$NS cells may explain more complex disturbances in the balance between the number of neurons and glia. It is hoped that by making this data accessible, researchers with defined questions about human foetal brain development may be able to identify particular molecules that warrant further study within the context of their own research.

15.0.4 Final Words

As intended this study has provided a number of candidate molecules for further study. It is hoped that more in-depth analysis on a molecule by molecule basis of these candidate molecules could yield fundamental breakthroughs in the development of brain-regenerative therapies.

Part VI

Appendix

Bibliography

[1] HGNC database of human gene names — HUGO Gene Nomenclature Committee. URL https://www.genenames.org/.

[2] UniProt. URL http://www.uniprot.org/.

[3] Neuronal, neural stem cell and glial cell markers — Abcam. URL http://www.abcam.com/neuroscience/neural-markers-guide.

[4] Home - PubMed - NCBI. URL http://www.ncbi.nlm.nih.gov/pubmed.

[5] Abcam. Neural markers guide. URL http://docs.abcam.com/pdf/neuroscience/neural-markers-guide-web.pdf.

[6] Abels, E. R. and Breakefield, X. O. Introduction to Extracellular Vesicles: Biogenesis, RNA Cargo Selection, Content, Release, and Uptake. *Cellular and Molecular Neurobiology*, 36(3):301–312, 2016. ISSN 15736830. doi: 10.1007/s10571-016-0366-z.

[7] Abraham, A. B., Bronstein, R., Chen, E. I., Koller, A., Ronfani, L., Maletic-Savatic, M., and Tsirka, S. E. Members of the high mobility group B protein family are dynamically expressed in embryonic neural stem cells. *Proteome Science*, 11(1):1, 2013. ISSN 14775956. doi: 10.1186/1477-5956-11-18. URL ProteomeScience.

[8] Ahmed, F. E. Utility of mass spectrometry for proteome analysis: part I. Conceptual and experimental approaches. *Expert review of proteomics*, 5(6):841–64, dec 2008. ISSN 1744-8387. doi: 10.1586/14789450.5.6.841. URL http://www.ncbi.nlm.nih.gov/pubmed/19086863.

[9] Ahmed, S. The culture of neural stem cells. *Journal of cellular biochemistry*, 106(1):1–6, jan 2009. ISSN 1097-4644. doi: 10.1002/jcb.21972. URL http://www.ncbi.nlm.nih.gov/pubmed/19021147.

[10] Ahn, S. and Joyner, A. L. In vivo analysis of quiescent adult neural stem cells responding to Sonic hedgehog. *Nature*, 437(7060):894–7, oct 2005. ISSN 1476-4687. doi: 10.1038/nature03994. URL http://www.ncbi.nlm.nih.gov/pubmed/16208373.

[11] Aimone, J. B., Wiles, J., and Gage, F. H. Potential role for adult neurogenesis in the encoding of time in new memories. *Nature neuroscience*, 9(6):723–7, jun 2006. ISSN 1097-6256. doi: 10.1038/nn1707. URL http://www.ncbi.nlm.nih.gov/pubmed/16732202.

[12] Akin, D., Newman, J. R. B., McIntyre, L. M., and Sugrue, S. P. RNA-seq analysis of impact of PNN on gene expression and alternative splicing in corneal epithelial cells. *Molecular vision*, 22(January):40–60, 2016. ISSN 1090-0535 (Electronic).

[13] Alexa, A. and Rahnenführer, J. Gene set enrichment analysis with topGO. 2013.

[14] Alexa, A. and Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology, 2018.

[15] Alexa, A., Rahnenfuhrer, J., and Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, 22(13):1600–1607, jul 2006. ISSN 1367-4803. doi: 10.1093/bioinformatics/btl140. URL https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btl140.

[16] Altman, J. and Das, G. D. Autoradiographic and histological evidence of postnatal hippocampal neurogenesis in rats. *The Journal of comparative neurology*, 124(3):319–35, jun 1965. ISSN 0021-9967. URL http://www.ncbi.nlm.nih.gov/pubmed/5861717.

[17] Anthony, T. E., Klein, C., Fishell, G., and Heintz, N. Radial glia serve as neural progenitors in all regions of the central nervous system. *Neuron*, 41:881 – 890, 2004.

[18] Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z., and Lindvall, O. Neuronal replacement from endogenous precursors in the adult brain after stroke. *Nature Medicine*, 8(9):963–970, sep 2002. ISSN 1078-8956. doi: 10.1038/nm747. URL http://www.nature.com/articles/nm747.

[19] Balzer, E., Heine, C., Jiang, Q., Lee, V. M., and Moss, E. G. LIN28 alters cell fate succession and acts independently of the let-7 microRNA during neurogliogenesis in vitro. *Development*, 137(6):891–900, mar 2010. ISSN 0950-1991. doi: 10.1242/dev.042895. URL http://dev.biologists.org/cgi/doi/10.1242/dev.042895.

[20] Banday, A. R., Baumgartner, M., Seesi, S. A., Karunakaran, D. K. P., Venkatesh, A., Congdon, S., Lemoine, C., Kilcollins, A. M., Mandoiu, I., Punzo, C., and Kanadia, R. N. Replication-dependent histone genes are actively transcribed in differentiating and aging retinal neurons. *Cell Cycle*, 13(16):2526–2541, 2014. ISSN 15514005. doi: 10.4161/15384101.2015.941757.

[21] Banner, L. R., Moayeri, N. N., and Patterson, P. H. Leukemia inhibitory factor is expressed in astrocytes following cortical brain injury. *Experimental neurology*, 147(1):1–9, sep 1997. ISSN 0014-4886. doi: 10.1006/exnr.1997.6536. URL http://www.ncbi.nlm.nih.gov/pubmed/9294397.

[22] Bao, X.-m., He, Q., Wang, Y., Huang, Z.-h., and Yuan, Z.-q. The roles and mechanisms of the Hippo/YAP signaling pathway in the nervous system. *Yi chuan = Hereditas*, 39(7):630–641, jul 2017. ISSN 0253-9772. doi: 10.16288/j.yczz.17-069. URL http://www.ncbi.nlm.nih.gov/pubmed/28757477.

[23] Barrero, M. J., Sese, B., Martí, M., and Izpisua Belmonte, J. C. Macro Histone Variants Are Critical for the Differentiation of Human Pluripotent Cells. *Journal of Biological Chemistry*, 288(22):16110–16116, may 2013. ISSN 0021-9258. doi: 10.1074/jbc.M113.466144. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668766/pdf/zbc16110.pdfhttp://www.jbc.org/lookup/doi/10.1074/jbc.M113.466144.

[24] Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Apweiler, R., Alpi, E., Antunes, R., Arganiska, J., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Chavali, G., Cibrian-Uhalte, E., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Gane, P., Castro, L. G., Garmiri, P., Hatton-Ellis, E., Hieta, R., Huntley, R., Legge, D., Liu, W., Luo, J., Macdougall, A., Mutowo, P., Nightingale, A., Orchard, S., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Turner, E., Volynkin, V., Wardell, T., Watkins, X., Zellner, H., Cowley, A., Figueira, L., Li, W., McWilliam, H., Lopez, R., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M. C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., De Castro, E., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A. L., Wu, C. H., Arighi, C. N., Arminski, L., Chen, C., Chen, Y., Garavelli, J. S., Huang, H., Laiho, K., McGarvey, P., Natale, D. A., Suzek, B. E., Vinayaka, C. R., Wang, Q., Wang, Y., Yeh, L. S., Yerramalla, M. S., and Zhang, J. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2015. ISSN 13624962. doi: 10.1093/nar/gku989.

[25] Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell*, 153(2):480–492, apr 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.03.011. URL http://dx.doi.org/10.1016/j.cell.2013.03.011http://linkinghub.elsevier.com/retrieve/pii/S0092867413003334.

[26] Bauer, S. and Patterson, P. H. Leukemia inhibitory factor promotes neural stem cell

self-renewal in the adult brain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(46):12089–99, nov 2006. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.3047-06.2006. URL http://www.ncbi.nlm.nih.gov/pubmed/17108182.

[27] Beck, M., Claassen, M., and Aebersold, R. Comprehensive proteomics. *Current opinion in biotechnology*, pages 1–6, sep 2010. ISSN 1879-0429. doi: 10.1016/j.copbio.2010.09.002. URL http://www.ncbi.nlm.nih.gov/pubmed/20888217.

[28] Bereman, M. S., Cilia, M., Cristea, I. M., Dufresne, C., and Pappin, D. J. Proteomics Workshop. In *Cold Spring Harbour Laboratories*, Cold Spring Harbour, 2012. URL https://meetings.cshl.edu/courses.aspx?course=C-PROTEO{&}year=17.

[29] Bergmann, O., Liebl, J., Bernard, S., Alkass, K., Yeung, M., Steier, P., Kutschera, W., Johnson, L., Landén, M., Druid, H., Spalding, K., and Frisén, J. The Age of Olfactory Bulb Neurons in Humans. *Neuron*, 74(4):634–639, 2012. ISSN 08966273. doi: 10.1016/j.neuron.2012.03.030. URL http://linkinghub.elsevier.com/retrieve/pii/S0896627312003418.

[30] Bergsland, M., Ramsköld, D., Zaouter, C., Klum, S., Sandberg, R., and Muhr, J. Sequentially acting Sox transcription factors in neural lineage development. *Genes & development*, 25(23):2453–64, 2011. ISSN 1549-5477. doi: 10.1101/gad.176008.111. URL http://genesdev.cshlp.org/content/25/23/2453.full{#}sec-8.

[31] Bhuiyan, M. I. H., Kim, S. Y., and Cho, K.-O. Lin28 overexpression inhibits neurite outgrowth of primary cortical neurons in vitro. *Acta neurobiologiae experimentalis*, 78 (4):297–304, 2018. ISSN 1689-0035. doi: 10.21307/ane2018028. URL http://www.ncbi.nlm.nih.gov/pubmed/30624428.

[32] Black, S., Kashkina, E., Kent, T., and Pomerantz, R. DNA Polymerase θ: A Unique Multifunctional End-Joining Machine. *Genes*, 7(9):67, sep 2016. ISSN 2073-4425. doi: 10.3390/genes7090067. URL http://www.mdpi.com/2073-4425/7/9/67.

[33] Bobola, N. From DNA binding to transcriptional activation: Is the TALE complete? *The Journal of Cell Biology*, 1:jcb.201706113, aug 2017. ISSN 0021-9525. doi: 10.1083/jcb.201706113. URL https://doi.org/10.1083/jcb.201706113http://www.jcb.org/lookup/doi/10.1083/jcb.201706113.

[34] Bogdanović, O. and Veenstra, G. J. C. DNA methylation and methyl-CpG binding proteins: Developmental requirements and function. *Chromosoma*, 118(5):549–565, 2009. ISSN 00095915. doi: 10.1007/s00412-009-0221-9.

[35] Boldrini, M., Fulmore, C. A., Tartt, A. N., Simeon, L. R., Pavlova, I., Poposka, V., Rosoklija, G. B., Stankov, A., Arango, V., Dwork, A. J., Hen, R., and Mann, J. J. Human Hippocampal Neurogenesis Persists throughout Aging. *Cell Stem Cell*, 22(4):589–599.e5, apr 2018. ISSN 19345909. doi: 10.1016/j.stem.2018.03.015. URL http://www.nature.com/articles/nbt.3301https://linkinghub.elsevier.com/retrieve/pii/S1934590918301218.

[36] Breier, J. M., Gassmann, K., Kayser, R., Stegeman, H., Groot, D. D., Fritsche, E., Shafer, T. J., and De Groot, D. Neural progenitor cells as models for high-throughput screens of developmental neurotoxicity: state of the science. *Neurotoxicology and teratology*, 32 (1):4–15, 2010. ISSN 1872-9738. doi: 10.1016/j.ntt.2009.06.005. URL http://dx.doi.org/10.1016/j.ntt.2009.06.005http://www.ncbi.nlm.nih.gov/pubmed/19559083.

[37] Burrows, L. How, not why, the human brain folds. URL http://news.harvard.edu/gazette/story/2016/02/how-not-why-the-human-brain-folds/.

[38] Buschbeck, M., Uribesalgo, I., Wibowo, I., Rué, P., Martin, D., Gutierrez, A., Morey, L., Guigó, R., López-Schier, H., and Di Croce, L. The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nature Structural & Molecular Biology*, 16(10):1074–1079, oct 2009. ISSN 1545-9993. doi: 10.1038/nsmb.1665. URL http://www.nature.com/nsmb/journal/v16/n10/pdf/nsmb.1665.pdfhttp://www.nature.com/doifinder/10.1038/nsmb.1665.

[39] Butts, T., Green, M. J., and Wingate, R. J. T. Development of the cerebellum: simple steps to make a 'little brain'. *Development*, 141:4031–4041, 2014. ISSN 0950-1991. doi: 10.1242/dev.106559. URL http://dev.biologists.org/cgi/doi/10.1242/dev.106559.

[40] Cacci, E., Villa, a., Parmar, M., Cavallaro, M., Mandahl, N., Lindvall, O., Martinez-Serrano, a., and Kokaia, Z. Generation of human cortical neurons from a new immortal fetal neural stem cell line. *Experimental cell research*, 313(3):588–601, feb 2007. ISSN 0014-4827. doi: 10.1016/j.yexcr.2006.11.001. URL http://www.ncbi.nlm.nih.gov/pubmed/17156776.

[41] Cao, R., Chen, K., Song, Q., Zang, Y., Li, J., Wang, X., Chen, P., and Liang, S. Quantitative proteomic analysis of membrane proteins involved in astroglial differentiation of neural stem cells by SILAC labeling coupled with LC-MS/MS. *Journal of Proteome Research*, 11(2):829–838, 2012. ISSN 15353893. doi: 10.1021/pr200677z.

[42] Cao, S., Du, J., Lv, Y., Lin, H., Mao, Z., Xu, M., Liu, M., and Liu, Y. PAX3 inhibits β-Tubulin-III expression and neuronal differentiation of neural stem cell. *Biochemical and Biophysical Research Communications*, 485(2):307–311, 2017. ISSN 10902104. doi: 10.1016/j.bbrc.2017.02.086. URL http://dx.doi.org/10.1016/j.bbrc.2017.02.086.

[43] Chaerkady, R. and Kerr, C. Temporal Analysis of Neural Differentiation Using Quantitative Proteomics†. *Journal of proteome ...*, 8(3):1315–1326, 2009. doi: 10.1021/pr8006667.Temporal. URL http://pubs.acs.org/doi/abs/10.1021/pr8006667.

[44] Chaerkady, R., Letzen, B., Renuse, S., Sahasrabuddhe, N. A., Kumar, P., All, A. H., Thakor, N. V., Delanghe, B., Gearhart, J. D., Pandey, A., and Kerr, C. L. Quantitative temporal proteomic analysis of human embryonic stem cell differentiation into oligodendrocyte progenitor cells. *Proteomics*, 11(20):4007–4020, 2011. ISSN 16159853. doi: 10.1002/pmic.201100107.

[45] Chang, W., Cheng, J., Allaire, J. J., Xie, Y., and McPherson, J. shiny: Web Application Framework for R, 2018. URL https://cran.r-project.org/package=shiny.

[46] Chaulk, S. G., Lattanzi, V. J., Hiemer, S. E., Fahlman, R. P., and Varelas, X. The hippo pathway effectors TAZ/YAP regulate dicer expression and MicroRNA biogenesis through Let-7. *Journal of Biological Chemistry*, 289(4):1886–1891, 2014. ISSN 00219258. doi: 10.1074/jbc.C113.529362.

[47] Cimadamore, F., Amador-Arjona, A., Chen, C., Huang, C.-T., and Terskikh, A. V. SOX2-LIN28/let-7 pathway regulates proliferation and neurogenesis in neural precursors. *Proceedings of the National Academy of Sciences of the United States of America*, 110(32):E3017–26, 2013. ISSN 1091-6490. doi: 10.1073/pnas.1220176110. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3740872{&}tool=pmcentrez{&}rendertype=abstract.

[48] Colak, D., Mori, T., Brill, M. S., Pfeifer, A., Falk, S., Deng, C., Monteiro, R., Mummery, C., Sommer, L., and Götz, M. Adult neurogenesis requires Smad4-mediated bone morphogenic protein signaling in stem cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28(2):434–46, jan 2008. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4374-07.2008. URL http://www.ncbi.nlm.nih.gov/pubmed/18184786.

[49] Computational Systems Biochemistry. MaxQuant, 2015. URL www.maxquant.org.

[50] Conti, L. and Cattaneo, E. Neural stem cell systems: physiological players or in vitro entities? *Nature reviews. Neuroscience*, 11(3):176–87, mar 2010. ISSN 1471-0048. doi: 10.1038/nrn2761. URL http://www.ncbi.nlm.nih.gov/pubmed/20107441.

[51] Corthals, G. L., Wasinger, V. C., Hochstrasser, D. F., and Sanchez, J. C. The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis*, 21(6):1104–15, apr 2000. ISSN 0173-0835. doi: 10.1002/(SICI)1522-2683(20000401)21:6⟨1104::AID-ELPS1104⟩3.0.CO;2-C. URL http://www3.interscience.wiley.com/journal/71500077/abstracthttp://www.ncbi.nlm.nih.gov/pubmed/10786884.

[52] Cox, J. and Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology*, 26(12):1367–72, dec 2008. ISSN 1546-1696. doi: 10.1038/nbt. URL http://www.ncbi.nlm.nih.gov/pubmed/19029910.

[53] Curtis, M. a., Kam, M., Nannmark, U., Anderson, M. F., Axell, M. Z., Wikkelso, C., Holtås, S., van Roon-Mom, W. M. C., Björk-Eriksson, T., Nordborg, C., Frisén, J., Dragunow, M., Faull, R. L. M., and Eriksson, P. S. Human neuroblasts migrate to the olfactory bulb via a lateral ventricular extension. *Science (New York, N.Y.)*, 315 (5816):1243–9, mar 2007. ISSN 1095-9203. doi: 10.1126/science.1136281. URL http://www.ncbi.nlm.nih.gov/pubmed/17303719.

[54] Dahlin, A. wikipedia, ESI positive mode, 2015. URL https://commons.wikimedia.org/wiki/File:ESI{_}positive{_}mode{_}(21589986840).jpg.

[55] De Filippis, L., Lamorte, G., Snyder, E. Y., Malgaroli, A., and Vescovi, A. L. A novel, immortal, and multipotent human neural stem cell line generating functional neurons and oligodendrocytes. *Stem cells (Dayton, Ohio)*, 25(9):2312–21, sep 2007. ISSN 1549-4918. doi: 10.1634/stemcells.2007-0040. URL http://www.ncbi.nlm.nih.gov/pubmed/17556596.

[56] Dehay, C. and Kennedy, H. Cell-cycle control and cortical development. *Nature Reviews Neuroscience*, 8(6):438–450, jun 2007. ISSN 1471-003X. doi: 10.1038/nrn2097. URL http://www.nature.com/articles/nrn2097.

[57] Devailly, G., Grandin, M., Perriaud, L., Mathot, P., Delcros, J.-G., Bidet, Y., Morel, A.-P., Bignon, J.-Y., Puisieux, A., Mehlen, P., and Dante, R. Dynamics of MBD2 deposition across methylated DNA regions during malignant transformation of human mammary epithelial cells. *Nucleic Acids Research*, 43(12): 5838–5854, jul 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv508. URL https:

//www.ncbi.nlm.nih.gov/pmc/articles/PMC4499136/pdf/gkv508.pdfhttps://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv508.

[58] Dhara, S. K. and Stice, S. L. Neural differentiation of human embryonic stem cells. *Journal of cellular biochemistry*, 105(3):633–40, oct 2008. ISSN 1097-4644. doi: 10.1002/jcb.21891. URL http://www.ncbi.nlm.nih.gov/pubmed/18759328.

[59] Dunkley, T., Costa, V., Friedlein, A., Lugert, S., Aigner, S., Ebeling, M., Miller, M. T., Patsch, C., Piraino, P., Cutler, P., and Jagasia, R. Characterization of a human pluripotent stem cell-derived model of neuronal development using multiplexed targeted proteomics. *Proteomics - Clinical Applications*, 9(7-8):684–694, 2015. ISSN 18628354. doi: 10.1002/prca.201400150.

[60] Durand, C. and Charbord, P. *Stem cell biology and regenerative medicine*. River Publishers, 2015. ISBN 9788793237070.

[61] Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D. G., Buetow, K. H., Gingeras, T. R., Misteli, T., and Meshorer, E. Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, 2(5):437–447, 2008. ISSN 19345909. doi: 10.1016/j.stem.2008.03.021.

[62] Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D. J., Dash, C., Bazett-Jones, D. P., Le Grice, S., McKay, R. D., Buetow, K. H., Gingeras, T. R., Misteli, T., and Meshorer, E. Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell*, 2(5):437–447, may 2008. ISSN 19345909. doi: 10.1016/j.stem.2008.03.021. URL https://linkinghub.elsevier.com/retrieve/pii/S193459090800177Xhttps://linkinghub.elsevier.com/retrieve/pii/S1934590908001616.

[63] Ehninger, D. and Kempermann, G. Neurogenesis in the adult hippocampus. *Cell and tissue research*, 331(1):243–50, jan 2008. ISSN 1432-0878. doi: 10.1007/s00441-007-0478-3. URL http://www.ncbi.nlm.nih.gov/pubmed/17938969.

[64] Elkabetz, Y. and Studer, L. Human ESC-derived Neural Rosettes and Neural Stem Cell Progression Human ESC-derived Neural Rosettes and Neural Stem Cell Progression. LXXIII:377–387, 2009. ISSN 1943-4456. doi: 10.1101/sqb.2008.73.052.

[65] Elkabetz, Y., Panagiotakos, G., Al Shamy, G., Socci, N. D., Tabar, V., and Studer, L. Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes and Development*, 22(2):152–165, 2008. ISSN 08909369. doi: 10.1101/gad.1616208.

[66] Encinas, J. M., Michurina, T. V., Peunova, N., Park, J.-H., Tordo, J., Peterson, D. A., Fishell, G., Koulakov, A., and Enikolopov, G. Division-Coupled Astrocytic Differentiation and Age-Related Depletion of Neural Stem Cells in the Adult Hippocampus. *Cell Stem Cell*, 8(5):566–579, may 2011. ISSN 19345909. doi: 10.1016/j.stem.2011.03.010. URL https://linkinghub.elsevier.com/retrieve/pii/S1934590911001202.

[67] Eriksson, P. S., Perfilieva, E., Björk-Eriksson, T., Alborn, A. M., Nordborg, C., Peterson, D. A., Gage, F. H., and Jolla, L. Neurogenesis in the adult human hippocampus. *Nature medicine*, 4(11):1313–7, nov 1998. ISSN 1078-8956. doi: 10.1038/3305. URL http://www.ncbi.nlm.nih.gov/pubmed/9809557http://www.ncbi.nlm.nih.gov/pubmed/19714567.

[68] Falk, A., Koch, P., Kesavan, J., Takashima, Y., Ladewig, J., Alexander, M., Wiskow, O., Tailor, J., Trotter, M., Pollard, S., Smith, A., and Brüstle, O. Capture of Neuroepithelial-Like Stem Cells from Pluripotent Stem Cells Provides a Versatile System for In Vitro Production of Human Neurons. *PLoS ONE*, 7(1):e29597, jan 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0029597. URL http://dx.plos.org/10.1371/journal.pone.0029597.

[69] Feng, J., Chang, H., Li, E., and Fan, G. Dynamic expression of de novo DNA methyltransferases Dnmt3a and Dnmt3b in the central nervous system. *Journal of Neuroscience*

Research, 79(6):734–746, mar 2005. ISSN 0360-4012. doi: 10.1002/jnr.20404. URL http://doi.wiley.com/10.1002/jnr.20404.

[70] Fenn, J., Mann, M., Meng, C., Wong, S., and Whitehouse, C. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, oct 1989. ISSN 0036-8075. doi: 10.1126/science.2675315. URL http://www.sciencemag.org/content/246/4926/64.

[71] Food and Agriculture Organization of the United Nations (FAO). Chapter 7: Basic principles of cryopreservation. *Cryoconservation of animal genetic resources*, pages 85–94, 2012. URL http://www.fao.org/3/i3017e/i3017e04.pdf.

[72] Forsberg-Nilsson, K., Behar, T. N., Afrakhte, M., Barker, J. L., and McKay, R. D. Platelet-derived growth factor induces chemotaxis of neuroepithelial stem cells. *Journal of neuroscience research*, 53(5):521–30, 1998. ISSN 0360-4012. doi: 10.1002/(SICI)1097-4547(19980901)53:5⟨521::AID-JNR2⟩3.0.CO;2-B. URL http://www.ncbi.nlm.nih.gov/pubmed/9726423.

[73] Franceschini, A. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research (Database issue)*, 41, 2013.

[74] Franco, S. J. and Müller, U. Shaping Our Minds: Stem and Progenitor Cell Diversity in the Mammalian Neocortex. *Neuron*, 77(1):19–34, 2013. ISSN 08966273. doi: 10.1016/j.neuron.2012.12.022.

[75] Fujita, S. Mitotic Pattern and Histogenesis of the Central Nervous System. *Nature*, 185(4714):702–703, mar 1960. ISSN 0028-0836. doi: 10.1038/185702a0. URL http://www.ncbi.nlm.nih.gov/pubmed/13825588http://www.nature.com/doifinder/10.1038/185702a0.

[76] Fujita, S. The matrix cell and cytogenesis in the developing central nervous system. *The Journal of comparative neurology*, 120:37–42, feb 1963. ISSN 0021-9967. URL http://www.ncbi.nlm.nih.gov/pubmed/13960118.

[77] Fujita, S. The Discovery of the Matrix Cell, the Identification of the Multipotent Neural Stem Cell and the Development of the Central Nervous System. *Cell Structure and Function*, 28(4):205–228, 2003. ISSN 0386-7196. doi: 10.1247/csf.28.205. URL http://joi.jlc.jst.go.jp/JST.JSTAGE/csf/28.205?from=CrossRef.

[78] Fujita, S. 50 Years of Research on the Phenomena and Epigenetic Mechanism of Neurogenesis. *Neuroscience Research*, 86:3–13, 2014. ISSN 18728111. doi: 10.1016/j.neures.2014.10.005. URL http://dx.doi.org/10.1016/j.neures.2014.03.007.

[79] Gabel, H. W. and Greenberg, M. E. The Maturing Brain Methylome. *Science*, 341(6146):626–627, aug 2013. ISSN 0036-8075. doi: 10.1126/science.1242671. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4445635/pdf/nihms691722.pdfhttp://www.sciencemag.org/cgi/doi/10.1126/science.1242671.

[80] Gage, F. H. Mammalian Neural Stem Cells. *Science*, 287(5457):1433–1438, feb 2000. ISSN 00368075. doi: 10.1126/science.287.5457.1433. URL http://www.sciencemag.org/cgi/doi/10.1126/science.287.5457.1433.

[81] Gao, L., Syahril, A., and Rozita, R. Neural Commitment of Embryonic Stem Cells through the Formation of Embryoid Bodies (EBs). 21(5):8–16, 2014.

[82] Gaspar-Maia, A., Alajem, A., Meshorer, E., and Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nature Reviews Molecular Cell Biology*, 12(1):36–47, jan 2011. ISSN 1471-0072. doi: 10.1038/nrm3036. URL http://www.nature.com/articles/nrm3036.

[83] Gerrard, L., Rodgers, L., and Cui, W. Differentiation of Human Embryonic Stem Cells to Neural Lineages in Adherent Culture by Blocking Bone Morphogenetic Protein Signaling. *Stem Cells*, 23(9):1234–1241, 2005. ISSN 10665099. doi: 10.1634/stemcells.2005-0110. URL http://doi.wiley.com/10.1634/stemcells.2005-0110.

[84] Gilbert, S. F. and Tyler, M. S. *Developmental Biology*. Sinauer Associates Inc., Sunderland, Masssachusetts, 7th editio edition, 2003.

[85] Gokoffski, K. K., Wu, H.-H., Beites, C. L., Kim, J., Kim, E. J., Matzuk, M. M., Johnson, J. E., Lander, A. D., and Calof, A. L. Activin and GDF11 collaborate in feedback control of neuroepithelial stem cell proliferation and fate. *Development*, 138(19):4131–4142, 2011. ISSN 0950-1991. doi: 10.1242/dev.065870.

[86] Goldowitz, D. and Hamre, K. The cells and molecules that make a cerebellum. *Trends in Neurosciences*, 21:375–382, 1998. ISSN 01662236. doi: 10.1016/S0166-2236(98)01313-7.

[87] Gorris, R., Fischer, J., Erwes, K. L., Kesavan, J., Peterson, D. a., Alexander, M., Nöthen, M. M., Peitz, M., Quandel, T., Karus, M., and Brüstle, O. Pluripotent stem cell-derived radial glia-like cells as stable intermediate for efficient generation of human oligodendrocytes. *Glia*, pages n/a–n/a, 2015. ISSN 08941491. doi: 10.1002/glia.22882. URL http://doi.wiley.com/10.1002/glia.22882.

[88] Götz, M. and Huttner, W. B. The cell biology of neurogenesis. *Nat Rev Mol Cell Biol*, 6(10):777–788, 2005. ISSN 1471-0072. doi: 10.1038/nrm1739. URL http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:The+cell+biology+of+neurogenesis{#}0.

[89] Gould, E., Reeves, A. J., Graziano, M. S., and Gross, C. G. Neurogenesis in the Neocortex of Adult Primates. *Science*, 286(5439):548–552, oct 1999. ISSN 00368075. doi: 10.1126/science.286.5439.548. URL http://www.sciencemag.org/cgi/doi/10.1126/science.286.5439.548http://www.ncbi.nlm.nih.gov/pubmed/10521353.

[90] Guo, C. and Morris, S. A. Engineering cell identity: establishing new gene regulatory and chromatin landscapes. *Current Opinion in Genetics and Development*, 46:50–57, 2017. ISSN 18790380. doi: 10.1016/j.gde.2017.06.011. URL http://dx.doi.org/10.1016/j.gde.2017.06.011.

[91] Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J., Jin, X., Shi, X., Liu, P., Wang, X., Wang, W., Wei, Y., Li, X., Guo, F., Wu, X., Fan, X., Yong, J., Wen, L., Xie, S. X., Tang, F., and Qiao, J. The DNA methylation landscape of

human early embryos. *Nature*, 511(7511):606–610, 2014. ISSN 0028-0836. doi: 10.1038/nature13544. URL http://www.nature.com/doifinder/10.1038/nature13544.

[92] Hao, Z., Zhang, Y., Eliuk, S., and Blethrow, J. A Quadrupole-Orbitrap Hybrid Mass Spectrometer Offers Highest Benchtop Performance for In-Depth Analysis of Complex Proteomes, 2012. URL http://thermofisher.cn/Resources/201301/16164727718.pdf.

[93] Hau, A.-C., Grebbin, B. M., Agoston, Z., Anders-Maurer, M., Müller, T., Groß, A., Kolb, J., Langer, J. D., Döring, C., and Schulte, D. MEIS homeodomain proteins facilitate PARP1/ARTD1-mediated eviction of histone H1. *The Journal of Cell Biology*, page jcb.201701154, jul 2017. ISSN 0021-9525. doi: 10.1083/jcb.201701154. URL http://www.jcb.org/lookup/doi/10.1083/jcb.201701154.

[94] Hegarty, S. V., O'Keeffe, G. W., and Sullivan, A. M. BMP-Smad 1/5/8 signalling in the development of the nervous system. *Progress in Neurobiology*, 109:28–41, 2013. ISSN 03010082. doi: 10.1016/j.pneurobio.2013.07.002. URL http://dx.doi.org/10.1016/j.pneurobio.2013.07.002.

[95] Hemmati-Brivanlou, A. and Melton, D. Vertebrate Embryonic Cells Will Become Nerve Cells Unless Told Otherwise. *Cell*, 88(1):13–17, jan 1997. ISSN 00928674. doi: 10.1016/S0092-8674(00)81853-X. URL http://dx.doi.org/10.1016/S0092-8674(00)81853-Xhttp://linkinghub.elsevier.com/retrieve/pii/S009286740081853X.

[96] Hill, M. Embryology Main Page, 2015. URL https://embryology.med.unsw.edu.au/embryology.

[97] Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A. I., Ranish, J., and Crabtree, G. R. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proceedings of the National Academy of Sciences, USA*, 106(13):5181–5186, 2009. doi: Doi10.1073/Pnas.0812889106.

[98] Ho, L., Ronan, J. L., Wu, J., Staahl, B. T., Chen, L., Kuo, A., Lessard, J., Nesvizhskii, A. I., Ranish, J., and Crabtree, G. R. An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A*, 106(13):5181–5186, 2009. ISSN 1091-6490. doi: 0812889106[pii]\r10.1073/pnas.0812889106. URL http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{_}uids=19279220.

[99] Horner, P. J. and Palmer, T. D. New roles for astrocytes: The nightlife of an 'astrocyte'. La vida loca! *Trends in Neurosciences*, 26(11):597–603, nov 2003. ISSN 01662236. doi: 10.1016/j.tins.2003.09.010. URL http://linkinghub.elsevier.com/retrieve/pii/S0166223603002935http://www.ncbi.nlm.nih.gov/pubmed/14585599.

[100] Hsu, S.-Y., Chen, Y.-J., and Ouyang, P. Pnn and SR family proteins are differentially expressed in mouse central nervous system. *Histochemistry and cell biology*, 135(4):361–73, 2011. ISSN 1432-119X. doi: 10.1007/s00418-011-0795-1. URL http://www.ncbi.nlm.nih.gov/pubmed/21437620.

[101] Hu, Q., Noll, R. J., Li, H., Makarov, A., Hardman, M., and Cooks, R. G. The Orbitrap: A new mass spectrometer. *Journal of Mass Spectrometry*, 40:430–443, 2005. ISSN 10765174. doi: 10.1002/jms.856.

[102] Huang, C. T.-L., Tao, Y., Lu, J., Jones, J. R., Fowler, L., Weick, J. P., and Zhang, S.-C. Time-Course Gene Expression Profiling Reveals a Novel Role of Non-Canonical WNT Signaling During Neural Induction. *Scientific reports*, 6(April):32600, 2016. ISSN 2045-2322. doi: 10.1038/srep32600. URL http://www.ncbi.nlm.nih.gov/pubmed/27600186.

[103] Iwafuchi-Doi, M. and Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes & Development*, 28(24):2679–2692, dec 2014. ISSN 0890-9369. doi: 10.1101/gad.253443.114. URL http://www.ncbi.nlm.nih.gov/pubmed/25512556http:

//www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4265672http://genesdev.cshlp.org/lookup/doi/10.1101/gad.253443.114.

[104] Iwafuchi-Doi, M. and Zaret, K. S. Cell fate control by pioneer transcription factors. *Development*, 143(11):1833–1837, 2016. ISSN 0950-1991. doi: 10.1242/dev.133900. URL http://dev.biologists.org/lookup/doi/10.1242/dev.133900.

[105] Iwata, K., Yumoto, K., Sugishima, M., Mizoguchi, C., Kai, Y., Iba, Y., and Mio, Y. Analysis of compaction initiation in human embryos by using time-lapse cinematography. *Journal of Assisted Reproduction and Genetics*, 31(4):421–426, 2014. ISSN 15737330. doi: 10.1007/s10815-014-0195-2.

[106] Jiao, J.-w., Feldheim, D. A., and Chen, D. F. Ephrins as negative regulators of adult neurogenesis in diverse regions of the central nervous system. *Proceedings of the National Academy of Sciences of the United States of America*, 105(25):8778–83, jun 2008. ISSN 1091-6490. doi: 10.1073/pnas.0708861105. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2438395{&}tool=pmcentrez{&}rendertype=abstract.

[107] Jiao, J. and Chen, D. F. Induction of neurogenesis in nonconventional neurogenic regions of the adult central nervous system by niche astrocyte-produced signals. *Stem cells (Dayton, Ohio)*, 26(5):1221–30, may 2008. ISSN 1549-4918. doi: 10.1634/stemcells.2007-0513. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2683670{&}tool=pmcentrez{&}rendertype=abstract.

[108] Joo, J. H., Ryu, D., Peng, Q., and Sugrue, S. P. Role of Pnn in alternative splicing of a specific subset of lncRNAs of the corneal epithelium. *Molecular Vision*, 20(August):1629–1642, 2014. ISSN 10900535. URL http://www.ncbi.nlm.nih.gov/pubmed/25489234.

[109] Kaplan, M. S. and Bell, D. H. Neuronal proliferation in the 9-month-old rodent-radioautographic study of granule cells in the hippocampus. *Experimental brain research*, 52(1):1–5, jan 1983. ISSN 0014-4819. URL http://www.ncbi.nlm.nih.gov/pubmed/6628589.

[110] Kaplan, M. S. and Hinds, J. W. Neurogenesis in the adult rat: electron microscopic analysis of light radioautographs. *Science (New York, N.Y.)*, 197(4308):1092–4, sep 1977. ISSN 0036-8075. URL http://www.ncbi.nlm.nih.gov/pubmed/887941.

[111] Karas, M., Bachmann, D., and Hillenkamp, F. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Analytical Chemistry*, 57:2935–2939, 1985. ISSN 0003-2700. doi: 10.1021/ac00291a042. URL http://pubs.acs.org/doi/abs/10.1021/ac00291a042.

[112] Kawauchi, T. Cell Adhesion and Its Endocytic Regulation in Cell Migration during Neural Development and Cancer Metastasis. *International Journal of Molecular Sciences*, 13(12):4564–4590, apr 2012. ISSN 1422-0067. doi: 10.3390/ijms13044564. URL http://www.mdpi.com/1422-0067/13/4/4564/.

[113] Kim, H., Jang, W. Y., Kang, M. C., Jeong, J., Choi, M., Sung, Y., Park, S., Kwon, W., Jang, S., Kim, M. O., Kim, S. H., and Ryoo, Z. Y. TET1 contributes to neurogenesis onset time during fetal brain development in mice. *Biochemical and Biophysical Research Communications*, 471(4):437–443, 2016. ISSN 10902104. doi: 10.1016/j.bbrc.2016.02.060. URL http://dx.doi.org/10.1016/j.bbrc.2016.02.060.

[114] Koch, P., Opitz, T., Steinbeck, J. a., Ladewig, J., and Brüstle, O. A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9):3225–30, mar 2009. ISSN 1091-6490. doi: 10.1073/pnas.0808387106. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2651316{&}tool=pmcentrez{&}rendertype=abstract.

[115] Kuhl, E. Biophysics: Unfolding the brain. *Nature Physics*, pages 1–2, 2016. ISSN 1745-2473. doi: 10.1038/nphys3641. URL http://www.nature.com/doifinder/10.1038/nphys3641.

[116] Laurent, J. M., Vogel, C., Kwon, T., Craig, S. a., Boutz, D. R., Huse, H. K., Nozue,

K., Walia, H., Whiteley, M., Ronald, P. C., and Marcotte, E. M. Protein abundances are more conserved than mRNA abundances across diverse taxa. *PROTEOMICS*, 10 (23):4209–4212, dec 2010. ISSN 16159853. doi: 10.1002/pmic.201000327. URL http://doi.wiley.com/10.1002/pmic.201000327.

[117] Lee, B. M., Buck-Koehntop, B. A., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. Embryonic Neural Inducing Factor Churchill Is not a DNA-binding Zinc Finger Protein: Solution Structure Reveals a Solvent-exposed ??-Sheet and Zinc Binuclear Cluster. *Journal of Molecular Biology*, 371(5):1274–1289, 2007. ISSN 00222836. doi: 10.1016/j.jmb.2007.06.021.

[118] Lee, H., Han, S., Kwon, C. S., and Lee, D. Biogenesis and regulation of the let-7 miRNAs and their functional implications. *Protein and Cell*, 7(2), 2015. ISSN 16748018. doi: 10.1007/s13238-015-0212-y.

[119] Lee, H.-K., Lee, H.-s., and Moody, S. A. Neural Transcription Factors: from Embryos to Neural Stem Cells. *Molecules and Cells*, 37(10):705–712, 2014. ISSN 1016-8478. doi: 10.14348/molcells.2014.0227. URL http://www.molcells.org/journal/view.html?doi=10.14348/molcells.2014.0227.

[120] Lessard, J., Wu, J. I., Ranish, J. A., Wan, M., Winslow, M. M., Staahl, B. T., Wu, H., Aebersold, R., Graef, I. A., and Crabtree, G. R. An Essential Switch in Subunit Composition of a Chromatin Remodeling Complex during Neural Development. *Neuron*, 55(2):201–215, jul 2007. ISSN 08966273. doi: 10.1016/j.neuron.2007.06.019. URL http://linkinghub.elsevier.com/retrieve/pii/S0896627307004527https://linkinghub.elsevier.com/retrieve/pii/S0896627307004527.

[121] Lian, I., Kim, J., Okazawa, H., Zhao, J., Zhao, B., Yu, J., Chinnaiyan, A., Israel, M. A., Goldstein, L. S., Abujarour, R., Ding, S., and Guan, K. L. The role of YAP transcription coactivator in regulating stem cell self-renewal and differentiation. *Genes and Development*, 24(11):1106–1118, 2010. ISSN 08909369. doi: 10.1101/gad.1903310.

[122] Lie, D.-C., Colamarino, S. A., Song, H.-J., Désiré, L., Mira, H., Consiglio, A., Lein, E. S., Jessberger, S., Lansford, H., Dearie, A. R., and Gage, F. H. Wnt signalling regulates adult hippocampal neurogenesis. *Nature*, 437(7063):1370–5, oct 2005. ISSN 1476-4687. doi: 10.1038/nature04108. URL http://www.ncbi.nlm.nih.gov/pubmed/16251967.

[123] Lim, D. a. and Alvarez-Buylla, A. Interaction between astrocytes and adult subventricular zone precursors stimulates neurogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 96(13):7526–31, jun 1999. ISSN 0027-8424. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22119{&}tool=pmcentrez{&}rendertype=abstract.

[124] Lin, S., Ren, A., Wang, L., Santos, C., Huang, Y., Jin, L., Li, Z., and Greene, N. D. E. Aberrant methylation of Pax3 gene and neural tube defects in association with exposure to polycyclic aromatic hydrocarbons. *Clinical Epigenetics*, 11(1):13, dec 2019. ISSN 1868-7075. doi: 10.1186/s13148-019-0611-7. URL https://clinicalepigeneticsjournal.biomedcentral.com/articles/10.1186/s13148-019-0611-7.

[125] Lister, R., Mukamel, E. a., Nery, J. R., Urich, M., Puddifoot, C. a., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., Sejnowski, T. J., Behrens, M. M., and Ecker, J. R. Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science*, 341(6146):1237905–1237905, aug 2013. ISSN 0036-8075. doi: 10.1126/science.1237905. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1237905.

[126] Liu, C. Y., Westerlund, U., Svensson, M., Moe, M. C., Varghese, M., Berg-Johnsen, J., Apuzzo, M. L. J., Tirrell, D. A., and Langmoen, I. a. Artificial niches for human adult neural stem cells: possibility for autologous transplantation therapy. *Journal of hematotherapy & stem cell research*, 12(6):689–99, dec 2003. ISSN 1525-8165. doi: 10.1089/15258160360732713. URL http://www.ncbi.nlm.nih.gov/pubmed/14977478.

[127] Liu, Y., Zhu, H., Liu, M., Du, J., Qian, Y., Wang, Y., Ding, F., and Gu, X. Downregulation of Pax3 expression correlates with acquired GFAP expression during NSC differentiation towards astrocytes. *FEBS Letters*, 585(7):1014–1020, 2011. ISSN 00145793. doi: 10.1016/j.febslet.2011.02.034. URL http://dx.doi.org/10.1016/j.febslet.2011.02.034.

[128] Liu, Y. C. and zhou Wang, Y. Role of Yes-associated protein 1 in gliomas: pathologic and therapeutic aspects. *Tumor Biology*, 36(4):2223–2227, 2015. ISSN 14230380. doi: 10.1007/s13277-015-3297-2.

[129] Lodato, M. a., Ng, C. W., Wamstad, J. a., Cheng, A. W., Thai, K. K., Fraenkel, E., Jaenisch, R., and Boyer, L. a. SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. *PLoS Genetics*, 9(2):e1003288, feb 2013. ISSN 1553-7404. doi: 10.1371/journal.pgen.1003288. URL http://dx.plos.org/10.1371/journal.pgen.1003288.

[130] Magavi, S. S., Leavitt, B. R., and Macklis, J. D. Induction of neurogenesis in the neocortex of adult mice. *Nature*, 405(6789):951–5, jun 2000. ISSN 0028-0836. doi: 10.1038/35016083. URL http://www.ncbi.nlm.nih.gov/pubmed/10879536.

[131] Magnusson, J. P., Goritz, C., Tatarishvili, J., Dias, D. O., Smith, E. M. K., Lindvall, O., Kokaia, Z., and Frisen, J. A latent neurogenic program in astrocytes regulated by Notch signaling in the mouse. *Science*, 346(6206):237–241, oct 2014. ISSN 0036-8075. doi: 10.1126/science.346.6206.237. URL http://www.sciencemag.org/cgi/doi/10.1126/science.346.6206.237.

[132] Makarov, A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*, 72(6):1156–1162, 2000. ISSN 00032700. doi: 10.1021/ac991131p.

[133] Maltman, D. J., Brand, S., Belau, E., Paape, R., Suckau, D., and Przyborski, S. a. Top-Down Label-Free LC-MALDI Analysis of the Peptidome During Neural Progenitor

Cell Differentiation Reveals Complexity in Cytoskeletal Protein Dynamics and Identifies Progenitor Cell Markers. *Proteomics*, pages 3992–4006, jul 2011. ISSN 1615-9861. doi: 10.1002/pmic.201100024. URL http://www.ncbi.nlm.nih.gov/pubmed/21761558.

[134] Mandel, S., Rechavi, G., and Gozes, I. Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis. *Developmental Biology*, 303:814–824, 2007. ISSN 00121606. doi: 10.1016/j.ydbio.2006.11.039.

[135] Mann, M. Filter Aided Sample Preparation (FASP) Method. pages 8–11. URL http://www.biochem.mpg.de/mann/approaches/fasp/index.html.

[136] Maroof, A. M., Brown, K., Shi, S.-H., Studer, L., and Anderson, S. a. Prospective isolation of cortical interneuron precursors from mouse embryonic stem cells. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(13):4667–75, mar 2010. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.4255-09.2010. URL http://www.ncbi.nlm.nih.gov/pubmed/20357117.

[137] Ming, G.-l. and Song, H. Adult neurogenesis in the mammalian central nervous system. *Annual review of neuroscience*, 28:223–50, jan 2005. ISSN 0147-006X. doi: 10.1146/annurev.neuro.28.051804.101459. URL http://www.ncbi.nlm.nih.gov/pubmed/16022595.

[138] Moreno-Jiménez, E. P., Flor-García, M., Terreros-Roncal, J., Rábano, A., Cafini, F., Pallas-Bazarra, N., Ávila, J., and Llorens-Martín, M. Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nature Medicine*, 25(4):554–560, 2019. ISSN 1546170X. doi: 10.1038/s41591-019-0375-9. URL http://dx.doi.org/10.1038/s41591-019-0375-9.

[139] Morgado, A. L., Rodrigues, C. M., and Solá, S. MicroRNA-145 Regulates Neural Stem Cell Differentiation Through the Sox2-Lin28/let-7 Signaling Pathway. *STEM CELLS*,

34(5):1386–1395, may 2016. ISSN 10665099. doi: 10.1002/stem.2309. URL http://doi.wiley.com/10.1002/stem.308http://doi.wiley.com/10.1002/stem.2309.

[140] Morris, S. A. Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks. *Development*, 143(15):2696–2705, 2016. ISSN 0950-1991. doi: 10.1242/dev.138263. URL http://dev.biologists.org/content/143/15/2696.abstract.

[141] MuhChyi, C., Juliandi, B., Matsuda, T., and Nakashima, K. Epigenetic regulation of neural stem cell fate during corticogenesis. *International Journal of Developmental Neuroscience*, 31(6):424–433, 2013. ISSN 07365748. doi: 10.1016/j.ijdevneu.2013.02.006. URL http://dx.doi.org/10.1016/j.ijdevneu.2013.02.006.

[142] Nabavi, S., Fox, R., Proulx, C. D., Lin, J. Y., Tsien, R. Y., and Malinow, R. Engineering a memory with LTD and LTP. *Nature*, jun 2014. ISSN 0028-0836. doi: 10.1038/nature13294. URL http://www.nature.com/doifinder/10.1038/nature13294.

[143] Naka, H., Nakamura, S., Shimazaki, T., and Okano, H. Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development. *Nature Neuroscience*, 11(9):1014–1023, sep 2008. ISSN 1097-6256. doi: 10.1038/nn.2168. URL http://www.nature.com/neuro/journal/v11/n9/pdf/nn.2168.pdfhttp://www.nature.com/doifinder/10.1038/nn.2168http://www.ncbi.nlm.nih.gov/pubmed/19160499.

[144] Nato, G., Caramello, A., Trova, S., Avataneo, V., Rolando, C., Taylor, V., Buffo, A., Peretto, P., and Luzzati, F. Striatal astrocytes produce neuroblasts in an excitotoxic model of Huntington's disease. *Development (Cambridge)*, 142(5):840–845, 2015. ISSN 14779129. doi: 10.1242/dev.116657.

[145] Nelson, A. D. and Svendsen, C. N. Low concentrations of extracellular FGF-2 are sufficient but not essential for neurogenesis from human neural progenitor cells. *Molecular and cellular neurosciences*, 33(1):29–35, sep 2006. ISSN 1044-7431. doi: 10.1016/j.mcn.2006.06.003. URL http://www.ncbi.nlm.nih.gov/pubmed/16860999.

[146] Nelson, A. D., Suzuki, M., and Svendsen, C. N. A high concentration of epidermal growth factor increases the growth and survival of neurogenic radial glial cells within human neurosphere cultures. *Stem cells (Dayton, Ohio)*, 26(2):348–55, feb 2008. ISSN 1549-4918. doi: 10.1634/stemcells.2007-0299. URL http://www.ncbi.nlm.nih.gov/pubmed/18032704.

[147] Nguyen, S. and Fenn, J. B. Gas-phase ions of solute species from charged droplets of solutions. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4):1111–1117, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0609969104.

[148] Nieto, M., Schuurmans, C., Britz, O., and Guillemot, F. Neural bHLH Genes Control the Neuronal versus Glial Fate Decision in Cortical Progenitors. *Neuron*, 29(2): 401–413, feb 2001. ISSN 08966273. doi: 10.1016/S0896-6273(01)00214-8. URL http://ac.els-cdn.com/S0896627301002148/1-s2.0-S0896627301002148-main.pdf?{_}tid=09bd252e-535e-11e7-88fa-00000aab0f26{&}acdnat=1497705081{_}5693d55e35aab8364a9230a2cfb4263ahttp://linkinghub.elsevier.com/retrieve/pii/S0896627301002148.

[149] Niwa, H. How is pluripotency determined and maintained? *Development*, 134(4):635–646, 2007. ISSN 0950-1991. doi: 10.1242/dev.02787. URL http://dev.biologists.org/cgi/doi/10.1242/dev.02787.

[150] Noble, W. S. How does multiple testing correction work? *Nature Biotechnology*, 27 (12):1135–1137, dec 2009. ISSN 1087-0156. doi: 10.1038/nbt1209-1135. URL http://www.nature.com/doifinder/10.1038/nbt1209-1135.

[151] Nolte, J. *Elsivier's Integrated Neuroscience*. Mosby, 2007. ISBN 9780323034098.

[152] Nottebohm, F. The road we travelled: discovery, choreography, and significance of brain replaceable neurons. *Annals of the New York Academy of Sciences*, 1016:628–58, jun 2004. ISSN 0077-8923. doi: 10.1196/annals.1298.027. URL http://www.ncbi.nlm.nih.gov/pubmed/15313798.

[153] Okano, H. and Sawamoto, K. Neural stem cells: involvement in adult neurogenesis and CNS repair. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1500):2111–22, jun 2008. ISSN 0962-8436. doi: 10.1098/rstb.2008.2264. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2610183{&}tool=pmcentrez{&}rendertype=abstract.

[154] Omel'chenko, N. V. and Smirnov, E. B. The nucleolar apparatus of neuroepithelial cells and the organization of the ventricular zone in the neocortex rudiment in humans. *Neuroscience and Behavioral Physiology*, 30(6):703–706, 2000. ISSN 00970549. doi: 10.1023/A:1026607001957.

[155] Omenn, G. S., Lane, L., Overall, C. M., Corrales, F. J., Schwenk, J. M., Paik, Y.-K., Van Eyk, J. E., Liu, S., Snyder, M., Baker, M. S., and Deutsch, E. W. Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *Journal of Proteome Research*, 17(12):4031–4041, dec 2018. ISSN 1535-3893. doi: 10.1021/acs.jproteome.8b00441. URL https://pubs.acs.org/doi/10.1021/acs.jproteome.8b00441.

[156] Ooi, S. K. T. and Bestor, T. H. The Colorful History of Active DNA Demethylation. *Cell*, 133(7):1145–1148, 2008. ISSN 00928674. doi: 10.1016/j.cell.2008.06.009.

[157] Oz, S., Ivashko-Pachima, Y., and Gozes, I. The ADNP derived peptide, NAP modulates the tubulin pool: implication for neurotrophic and neuroprotective activities. *PloS one*, 7(12):e51458, 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0051458. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3522725{&}tool=pmcentrez{&}rendertype=abstract.

[158] Ozair, M. Z., Kintner, C., and Brivanlou, A. H. Neural induction and early patterning in vertebrates. *Wiley Interdisciplinary Reviews: Developmental Biology*, 2(4):479–498, 2013. ISSN 17597684. doi: 10.1002/wdev.90.

[159] Parent, J. M., Vexler, Z. S., Gong, C., Derugin, N., and Ferriero, D. M. Rat forebrain

neurogenesis and striatal neuron replacement after focal stroke. *Annals of Neurology*, 52 (6):802–813, 2002. ISSN 03645134. doi: 10.1002/ana.10393.

[160] Pascovici, D., Handler, D., Wu, J. X., and Haynes, P. A. Multiple testing corrections in quantitative proteomics: A useful but blunt tool. *PROTEOMICS*, jul 2016. ISSN 16159853. doi: 10.1002/pmic.201600044. URL http://doi.wiley.com/10.1002/pmic.201600044.

[161] Pei, D., Liu, N., Li, D., Yan, H., bo Wang, Q., Fang, Y., Xie, L., and Li, H. P. Inhibition of platelet-derived growth factor receptor β reduces reactive glia and scar formation after traumatic brain injury in mice. *Brain Research Bulletin*, 134:121–127, 2017. ISSN 18732747. doi: 10.1016/j.brainresbull.2017.06.020. URL http://dx.doi.org/10.1016/j.brainresbull.2017.06.020.

[162] Pencea, V., Bingaman, K. D., Wiegand, S. J., and Luskin, M. B. Infusion of brain-derived neurotrophic factor into the lateral ventricle of the adult rat leads to new neurons in the parenchyma of the striatum, septum, thalamus, and hypothalamus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(17):6706–17, sep 2001. ISSN 1529-2401. URL http://www.ncbi.nlm.nih.gov/pubmed/11517260.

[163] Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, dec 1999. ISSN 0173-0835. doi: 10.1002/(SICI)1522-2683(19991201)20:18⟨3551::AID-ELPS3551⟩3.0.CO;2-2. URL http://doi.wiley.com/10.1002/{%}28SICI{%}291522-2683{%}2819991201{%}2920{%}3A18{%}3C3551{%}3A{%}3AAID-ELPS3550.CO{%}3B2-2.

[164] Pilz, G. A., Shitamukai, A., Reillo, I., Pacary, E., Schwausch, J., Stahl, R., Ninkovic, J., Snippert, H. J., Clevers, H., Godinho, L., Guillemot, F., Borrell, V., Matsuzaki, F., and ouml Tz, M. G. Amplification of progenitors in the mammalian telen-

cephalon includes a new radial glial cell type. *Nature Communications*, 4:2125, 2013. ISSN 2041-1723. doi: 10.1038/ncomms3125. URL http://dx.doi.org/10.1038/ncomms3125{%}5Cnpapers2://publication/doi/10.1038/ncomms3125.

[165] Prives, C. and Gottifredi, V. The p21 and PCNA partnership: A new twist for an old plot. *Cell Cycle*, 7(24):3840–3846, 2008. ISSN 15384101. doi: 10.4161/cc.7.24.7243.

[166] Proteom, C., Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Molecular & cellular proteomics : MCP*, 10(9):M111.011015, sep 2011. ISSN 1535-9484. doi: 10.1074/mcp.M111.011015. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3284220{&}tool=pmcentrez{&}rendertype=abstract.

[167] Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M. Neuroscience, 2001. URL http://www.ncbi.nlm.nih.gov/books/NBK10799/.

[168] Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., and Williams, S. M. *Neuroscience 2nd Edition*. Sinuaer Associates, 2001. ISBN 0-87893-742.0.

[169] Purves, D., Augustine, G. j., David, F., Hall, W. C., LaMantia, A.-S., McNamara, J. O., and Mark, W. S., editors. *NEUROSCIENCE*. Sinauer Associates Inc., Sunderland, MA 01375 U.S.A, 3rd edition, 2004. ISBN 0878937250.

[170] Qian, X., Shen, Q., Goderie, S. K., He, W., Capela, A., Davis, A. A., and Temple, S. Timing of CNS Cell Generation : A Programmed Sequence of Neuron and Glial Cell Production from Isolated Murine Cortical Stem Cells. 28:69–80, 2000.

[171] R Core Team. R: A Language and Environment for Statistical Computing, 2018. URL https://www.r-project.org/.

[172] Rabenstein, M., Hucklenbroich, J., Willuweit, A., Ladwig, A., Fink, G. R., Schroeter, M., Langen, K.-J., and Rueger, M. A. Osteopontin mediates survival, proliferation and migration of neural stem cells through the chemokine receptor CXCR4. *Stem Cell Research & Therapy*, 6(1):99, 2015. ISSN 1757-6512. doi: 10.1186/s13287-015-0098-x. URL http://stemcellres.com/content/6/1/99.

[173] Rehfeld, F., Rohde, A. M., Nguyen, D. T. T., and Wulczyn, F. G. Lin28 and let-7: ancient milestones on the road from pluripotency to neurogenesis. *Cell and tissue research*, may 2014. ISSN 1432-0878. doi: 10.1007/s00441-014-1872-2. URL http://www.ncbi.nlm.nih.gov/pubmed/24825413.

[174] Rehfeld, F., Rohde, A. M., Nguyen, D. T. T., and Wulczyn, F. G. Lin28 and let-7: ancient milestones on the road from pluripotency to neurogenesis. *Cell and Tissue Research*, 359(1):145–160, jan 2015. ISSN 0302-766X. doi: 10.1007/s00441-014-1872-2. URL http://link.springer.com/10.1007/s00441-014-1872-2.

[175] Reik, W. and Kelsey, G. Cellular memory erased in human embryos. *Nature*, pages 4–5, 2014.

[176] Roepstorff, P. and Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical mass spectrometry*, 11(11):601, nov 1984. ISSN 0306-042X. doi: 10.1002/bms.1200111109. URL http://www.ncbi.nlm.nih.gov/pubmed/6525415.

[177] Rollín, R., Tornero, P., Marco, F., Camafeita, E., Calvo, E., López-Durán, L., Jover, J. Á., López, J. A., Lamas, J. R., and Fernández-Gutiérrez, B. Differential Proteome of Articular Chondrocytes From Patients with Osteoarthritis. *Journal of Proteomics & Bioinformatics*, 01(August):267–280, 2008. ISSN 0974276X. doi: 10.4172/jpb.1000034.

[178] Rothstein, J. D., Martin, L., Levey, A. I., Dykes-Hoberg, M., Jin, L., Wu, D., Nash, N., and Kuncl, R. W. Localization of neuronal and glial glutamate transporters. *Neuron*, 13(3):713–725, 1994. ISSN 08966273. doi: 10.1016/0896-6273(94)90038-8.

[179] Rougvie, A. E. Control of developmental timing in animals. *Nature Reviews Genetics*, 2(9):690–701, sep 2001. ISSN 1471-0056. doi: 10.1038/35088566. URL http://www.nature.com/articles/35088566.

[180] Ruijtenberg, S. and van den Heuvel, S. Coordinating cell proliferation and differentiation: Antagonism between cell cycle regulators and cell type-specific gene expression. *Cell Cycle*, 15(2):196–212, 2016. ISSN 1538-4101. doi: 10.1080/15384101.2015.1120925. URL http://www.tandfonline.com/doi/full/10.1080/15384101.2015.1120925.

[181] Seo, E., Basu-Roy, U., Zavadil, J., Basilico, C., and Mansukhani, A. Distinct functions of Sox2 control self-renewal and differentiation in the osteoblast lineage. *Molecular and cellular biology*, 31(22):4593–608, 2011. ISSN 1098-5549. doi: 10.1128/MCB.05798-11. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3209254{&}tool=pmcentrez{&}rendertype=abstract.

[182] Seo, E., Basu-Roy, U., Gunaratne, P. H., Coarfa, C., Lim, D. S., Basilico, C., and Mansukhani, A. SOX2 Regulates YAP1 to Maintain Stemness and Determine Cell Fate in the Osteo-Adipo Lineage. *Cell Reports*, 3(6):2075–2087, 2013. ISSN 22111247. doi: 10.1016/j.celrep.2013.05.029. URL http://dx.doi.org/10.1016/j.celrep.2013.05.029.

[183] Shen, Q., Wang, Y., Dimos, J., and Fasano, C. The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. *Nature*, 9(6):743–751, 2006. ISSN 1097-6256. doi: 10.1038/nn1694. URL http://scholar.google.com/scholar?hl=en{&}btnG=Search{&}q=intitle:The+timing+of+cortical+neurogenesis+is+encoded+within+lineages+of+individual+progenitor+cells{#}0{%}5Cnhttp://www.nature.com/neuro/journal/v9/n6/abs/nn1694.html.

[184] Sheng, G., Dos Reis, M., and Stern, C. D. Churchill, a Zinc Finger Transcriptional Activator, Regulates the Transition between Gastrulation and Neurulation. *Cell*, 115(5):603–613, 2003. ISSN 00928674. doi: 10.1016/S0092-8674(03)00927-9.

[185] Shin, S., Mitalipova, M., Noggle, S., Tibbitts, D., Venable, A., Rao, R., and Stice, S. L. Long-term proliferation of human embryonic stem cell-derived neuroepithelial cells using defined adherent culture conditions. *Stem cells (Dayton, Ohio)*, 24(1):125–38, jan 2006. ISSN 1066-5099. doi: 10.1634/stemcells.2004-0150. URL http://www.ncbi.nlm.nih.gov/pubmed/16100006.

[186] Shingo, T., Gregg, C., Enwere, E., Fujikawa, H., Hassam, R., Geary, C., Cross, J. C., and Weiss, S. Pregnancy-stimulated neurogenesis in the adult female forebrain mediated by prolactin. *Science (New York, N.Y.)*, 299(5603):117–20, jan 2003. ISSN 1095-9203. doi: 10.1126/science.1076647. URL http://www.ncbi.nlm.nih.gov/pubmed/12511652.

[187] Shors, T. J., Miesegaes, G., Beylin, A., Zhao, M., Rydel, T., Gould, E., Costall, A., Textons, B., and E. Neurogenesis in the adult is involved in the formation of trace memories. *Nature*, 410(6826):372–376, mar 2001. ISSN 0028-0836. doi: 10.1038/35066584. URL http://www.nature.com/nature/journal/v410/n6826/abs/410372a0.htmlhttp://www.ncbi.nlm.nih.gov/pubmed/11268214.

[188] Singec, I., Crain, A. M., Hou, J., Tobe, B. T., Talantova, M., Winquist, A. A., Doctor, K. S., Choy, J., Huang, X., La Monaca, E., Horn, D. M., Wolf, D. A., Lipton, S. A., Gutierrez, G. J., Brill, L. M., and Snyder, E. Y. Quantitative Analysis of Human Pluripotency and Neural Specification by In-Depth (Phospho)Proteomic Profiling. *Stem Cell Reports*, 7(3):527–542, sep 2016. ISSN 22136711. doi: 10.1016/j.stemcr.2016.07.019. URL http://www.ncbi.nlm.nih.gov/pubmed/27569059http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5032292https://linkinghub.elsevier.com/retrieve/pii/S2213671116301436.

[189] Sisti, H. M., Glass, A. L., and Shors, T. J. Neurogenesis and the spacing effect: learning over time enhances memory and the survival of new neurons. *Learning & memory*, 14(5):368–75, may 2007. ISSN 1549-5485. doi: 10.1101/lm.488707. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1876761{&}tool=pmcentrez{&}rendertype=abstract.

[190] Smith, B. Multi-dimensional Human Embryo Images, 2015. URL http://embryo.soad.umich.edu/carnStages/carnStages.html.

[191] Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. DNA methylation dynamics of the human preimplantation embryo. *Nature*, 511(7511):611–615, 2014. ISSN 0028-0836. doi: 10.1038/nature13581. URL http://www.nature.com/doifinder/10.1038/nature13581.

[192] Sokolik, C., Liu, Y., Bauer, D., McPherson, J., Broeker, M., Heimberg, G., Qi, L. S., Sivak, D. a., and Thomson, M. Transcription Factor Competition Allows Embryonic Stem Cells to Distinguish Authentic Signals from Noise. *Cell Systems*, 1(2):117–129, 2015. ISSN 24054712. doi: 10.1016/j.cels.2015.08.001. URL http://www.sciencedirect.com/science/article/pii/S2405471215000538http://linkinghub.elsevier.com/retrieve/pii/S2405471215000538.

[193] Son, E. Y. and Crabtree, G. R. The role of BAF (mSWI/SNF) complexes in mammalian neural development. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics*, 166(3):333–349, 2014. ISSN 15524876. doi: 10.1002/ajmg.c.31416.

[194] Song, H., Stevens, C. F. C. C. F., and Gage, F. H. F. Astroglia induce neurogenesis from adult neural stem cells. *Nature*, 417(6884):39–44, may 2002. ISSN 0028-0836. doi: 10.1038/417039a. URL http://www.nature.com/nature/journal/v417/n6884/abs/417039a.htmlhttp://www.ncbi.nlm.nih.gov/pubmed/11986659.

[195] Spalding, K. L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H. B., Bostrom, E., Westerlund, I., Vial, C., Buchholz, B. a., Possnert, G., Mash, D. C., Druid, H., and Frisen, J. Dynamics of hippocampal neurogenesis in adult humans. *Cell*, 153:1219–1227, 2013. ISSN 1097-4172. doi: 10.1016/j.cell.2013.05.002. URL http://www.ncbi.nlm.nih.gov/pubmed/23746839.

[196] Spalding, K. L., Bhardwaj, R. D., Buchholz, B. A., Druid, H., and Frisén, J. Retrospective Birth Dating of Cells in Humans. *Cell*, 122(1):133–143, 2005. ISSN 00928674.

doi: 10.1016/j.cell.2005.04.028. URL http://linkinghub.elsevier.com/retrieve/pii/S0092867405004083.

[197] Spiers, H., Hannon, E., Schalkwyk, L. C., Smith, R., Wong, C. C., O'Donovan, M. C., Bray, N. J., and Mill, J. Methylomic trajectories across human fetal brain development. *Genome Research*, 25(3):338–352, mar 2015. ISSN 1088-9051. doi: 10.1101/gr.180273.114. URL http://genome.cshlp.org/lookup/doi/10.1101/gr.180273.114.

[198] Staahl, B. T. and Crabtree, G. R. Creating a neural specific chromatin landscape by npBAF and nBAF complexes. *Current Opinion in Neurobiology*, 23(6):903–913, dec 2013. ISSN 09594388. doi: 10.1016/j.conb.2013.09.003. URL http://linkinghub.elsevier.com/retrieve/pii/S1074761310004966http://linkinghub.elsevier.com/retrieve/pii/S0959438813001827.

[199] Staahl, B. T., Tang, J., Wu, W., Sun, A., Gitler, A. D., Yoo, A. S., and Crabtree, G. R. Kinetic analysis of npBAF to nBAF switching reveals exchange of SS18 with CREST and integration with neural developmental pathways. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(25):10348–61, 2013. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.1258-13.2013. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3685834{&}tool=pmcentrez{&}rendertype=abstract.

[200] Stanfield, B. B. and Trice, J. E. Evidence that granule cells generated in the dentate gyrus of adult rats extend axonal projections. *Experimental Brain Research*, 72(2):399–406–406, sep 1988. ISSN 0014-4819. doi: 10.1007/BF00250261. URL http://www.ncbi.nlm.nih.gov/pubmed/2465172http://www.springerlink.com/content/j1315rm301q2vum4/.

[201] Stern, C. D. Neural induction: 10 years on since the 'default model'. *Curr Opin Cell Biol*, 18(6):692–697, 2006. ISSN 09550674. doi: 10.1016/j.ceb.2006.09.002. URL http://www.ncbi.nlm.nih.gov/pubmed/17045790.

[202] Stiles, J. and Jernigan, T. L. The basics of brain development. *Neuropsychology Review*, 20(4):327–348, 2010. ISSN 10407308. doi: 10.1007/s11065-010-9148-4.

[203] Suh, H., Deng, W., and Gage, F. H. Signaling in adult neurogenesis. *Annual review of cell and developmental biology*, 25:253–75, jan 2009. ISSN 1530-8995. doi: 10.1146/annurev.cellbio.042308.113256. URL http://www.ncbi.nlm.nih.gov/pubmed/19575663.

[204] Suhonen, J. O., Peterson, D. A., Ray, J., and Gage, F. H. Differentiation of adult hippocampus-derived progenitors into olfactory neurons in vivo. *Nature*, 383(6601):624–7, oct 1996. ISSN 0028-0836. doi: 10.1038/383624a0. URL http://www.nature.com/nature/journal/v383/n6601/pdf/383624a0.pdfhttp://dx.doi.org/10.1038/383624a0.

[205] Sun, P., Ortega, G., Tan, Y., Hua, Q., Riederer, P. F., Deckert, J., and Schmitt-Böhrer, A. G. Streptozotocin Impairs Proliferation and Differentiation of Adult Hippocampal Neural Stem Cells in Vitro-Correlation With Alterations in the Expression of Proteins Associated With the Insulin System. *Frontiers in Aging Neuroscience*, 10(MAY):1–13, may 2018. ISSN 1663-4365. doi: 10.3389/fnagi.2018.00145. URL https://www.frontiersin.org/article/10.3389/fnagi.2018.00145/full.

[206] Sun, Y., Pollard, S., Conti, L., Toselli, M., Biella, G., Parkin, G., Willatt, L., Falk, A., Cattaneo, E., and Smith, A. G. Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. *Molecular and cellular neurosciences*, 38(2):245–58, jun 2008. ISSN 1095-9327. doi: 10.1016/j.mcn.2008.02.014. URL http://www.ncbi.nlm.nih.gov/pubmed/18450476.

[207] Suslov, O. N., Kukekov, V. G., Ignatova, T. N., and Steindler, D. a. Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres. *Proceedings of the National Academy of Sciences of the United States of America*, 99(22):14506–11, oct 2002. ISSN 0027-8424. doi: 10.1073/pnas.

212525299. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137913{&}tool=pmcentrez{&}rendertype=abstract.

[208] Suzuki, M. M. and Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics*, 9(6):465–76, 2008. ISSN 1471-0064. doi: 10.1038/nrg2341. URL http://www.ncbi.nlm.nih.gov/pubmed/18463664.

[209] Svendsen, C. N., Caldwell, M. a., Shen, J., ter Borg, M. G., Rosser, a. E., Tyers, P., Karmiol, S., and Dunnett, S. B. Long-term survival of human central nervous system progenitor cells transplanted into a rat model of Parkinson's disease. *Experimental neurology*, 148(1):135–46, nov 1997. ISSN 0014-4886. doi: 10.1006/exnr.1997.6634. URL http://www.ncbi.nlm.nih.gov/pubmed/9398456.

[210] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., and von Mering, C. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(Database issue):D447–52, 2015. ISSN 1362-4962. doi: 10.1093/nar/gku1003. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4383874{&}tool=pmcentrez{&}rendertype=abstract.

[211] Tailor, J., Kittappa, R., Leto, K., Gates, M., Borel, M., Paulsen, O., Spitzer, S., Karadottir, R. T., Rossi, F., Falk, A., and Smith, A. Stem Cells Expanded from the Human Embryonic Hindbrain Stably Retain Regional Specification and High Neurogenic Potency. *Journal of Neuroscience*, 33(30):12407–12422, jul 2013. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.0130-13.2013. URL http://www.ncbi.nlm.nih.gov/pubmed/23884946http://www.jneurosci.org/cgi/doi/10.1523/JNEUROSCI.0130-13.2013.

[212] Tailor, J. K. *Human neuroepithelial stem cells from the embryonic hindbrain*. PhD thesis, University of Cambridge, 2012.

[213] Takahashi, K. Cellular reprogramming - lowering gravity on Waddington's epigenetic landscape. *Journal of Cell Science*, 125(11):2553–2560, jun 2012. ISSN 0021-9533. doi:

10.1242/jcs.084822. URL http://jcs.biologists.org/content/125/11/2553http://jcs.biologists.org/cgi/doi/10.1242/jcs.084822.

[214] Takahashi, K. and Yamanaka, S. A developmental framework for induced pluripotency. 2015. doi: 10.1242/dev.114249. URL http://dev.biologists.org/content/develop/142/19/3274.full.pdf.

[215] Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., Reik, W., Bertone, P., and Smith, A. Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell*, 158(6):1254–1269, 2014. ISSN 0092-8674. doi: 10.1016/j.cell.2014.08.029. URL http://dx.doi.org/10.1016/j.cell.2014.08.029.

[216] Tallinen, T., Chung, J. Y., Rousseau, F., Girard, N., Lefèvre, J., and Mahadevan, L. On the growth and form of cortical convolutions. *Nature Physics*, 476(7358):57–62, feb 2016. ISSN 1745-2473. doi: 10.1038/nphys3632. URL http://dx.doi.org/10.1038/nature10277http://www.nature.com/doifinder/10.1038/nphys3632.

[217] Tan, F., Wahdan-Alaswad, R., Yan, S., Thiele, C. J., and Li, Z. Dihydropyrimidinase-like protein 3 expression is negatively regulated by MYCN and associated with clinical outcome in neuroblastoma. *Cancer Science*, 104(12):1586–1592, 2013. ISSN 13479032. doi: 10.1111/cas.12278.

[218] Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. Protein and polymer analyses up tom/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, 2(8):151–153, aug 1988. ISSN 0951-4198. doi: 10.1002/rcm.1290020802. URL http://doi.wiley.com/10.1002/rcm.1290020802.

[219] Taverna, E., Götz, M., and Huttner, W. B. *The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex.*, volume 30. 2014. ISBN 1530-8995 (Electronic)\r1081-0706 (Linking). doi: 10.1146/

annurev-cellbio-101011-155801. URL http://www.annualreviews.org/doi/abs/10.1146/annurev-cellbio-101011-155801.

[220] Thakurela, S., Tiwari, N., Schick, S., Garding, A., Ivanek, R., Berninger, B., and Tiwari, V. K. Mapping gene regulatory circuitry of Pax6 during neurogenesis. *Cell Discovery*, 2:15045, feb 2016. ISSN 2056-5968. doi: 10.1038/celldisc.2015.45. URL http://www.nature.com/articles/celldisc201545.

[221] The National High Magnetic Field Laboratory. Matrix-Assisted Laser Desorption Ionization (MALDI). URL https://nationalmaglab.org/user-facilities/icr/techniques/maldi.

[222] Tiwari, N., Pataskar, A., Péron, S., Thakurela, S., Sahu, S. K., Figueres-Oñate, M., Marichal, N., López-Mascaraque, L., Tiwari, V. K., and Berninger, B. Stage-Specific Transcription Factors Drive Astrogliogenesis by Remodeling Gene Regulatory Landscapes. *Cell stem cell*, 23(4):557–571.e8, oct 2018. ISSN 1875-9777. doi: 10.1016/j.stem.2018.09.008. URL http://www.ncbi.nlm.nih.gov/pubmed/30290178http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6179960.

[223] Tomita, K. Mammalian achaete-scute and atonal homologs regulate neuronal versus glial fate determination in the central nervous system. *The EMBO Journal*, 19 (20):5460–5472, oct 2000. ISSN 14602075. doi: 10.1093/emboj/19.20.5460. URL http://www.ncbi.nlm.nih.gov/pubmed/11032813http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC314003http://emboj.embopress.org/cgi/doi/10.1093/emboj/19.20.5460.

[224] Turner, B. M. Open Chromatin and Hypertranscription in Embryonic Stem Cells. *Cell Stem Cell*, 2(5):408–410, 2008. ISSN 19345909. doi: 10.1016/j.stem.2008.04.006.

[225] Twynman, R. *Principles of Proteomics*. BIOS Scientific Publishers, New York, 2004.

[226] Vale, W., Wiater, E., Gray, P., Harrison, C., Bilezikjian, L., and Choe, S. Activins and

inhibins and their signaling. *Annals of the New York Academy of Sciences*, 1038(1):142–7, dec 2004. ISSN 0077-8923. doi: 10.1196/annals.1315.023. URL http://doi.wiley.com/10.1196/annals.1315.023http://www.ncbi.nlm.nih.gov/pubmed/15838109.

[227] Warbrick, E., Lane, D. P., Glover, D. M., and Cox, L. S. A small peptide inhibitor of DNA replication defines the site of interaction between the cyclin-dependent kinase inhibitor p21WAF1 and proliferating cell nuclear antigen. *Current Biology*, 5(3):275–282, 1995. ISSN 09609822. doi: 10.1016/S0960-9822(95)00058-3.

[228] Warner-Schmidt, J. L. and Duman, R. S. Hippocampal neurogenesis: opposing effects of stress and antidepressant treatment. *Hippocampus*, 16(3):239–49, jan 2006. ISSN 1050-9631. doi: 10.1002/hipo.20156. URL http://www.ncbi.nlm.nih.gov/pubmed/16425236.

[229] Wei, C., Ren, L., Li, K., and Lu, Z. The regulation of survival and differentiation of neural stem cells by miR-124 via modulating PAX3. *Neuroscience Letters*, jun 2018. ISSN 03043940. doi: 10.1016/j.neulet.2018.05.051. URL https://doi.org/10.1016/j.neulet.2018.05.051http://linkinghub.elsevier.com/retrieve/pii/S0304394018304026.

[230] Wei, S., Wang, J., Oyinlade, O., Ma, D., Wang, S., Kratz, L., Lal, B., Xu, Q., Liu, S., Shah, S. R., Zhang, H., Li, Y., Quiñones-Hinojosa, A., Zhu, H., Huang, Z.-y., Cheng, L., Qian, J., and Xia, S. Heterozygous IDH1R132H/WT created by "single base editing" inhibits human astroglial cell growth by downregulating YAP. *Oncogene*, page 1, 2018. ISSN 0950-9232. doi: 10.1038/s41388-018-0334-9. URL http://www.nature.com/articles/s41388-018-0334-9.

[231] Wentz, C. T. and Magavi, S. S. P. Caffeine alters proliferation of neuronal precursors in the adult hippocampus. *Neuropharmacology*, 56(6-7):994–1000, 2009. ISSN 1873-7064. doi: 10.1016/j.neuropharm.2009.02.002. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2743873{&}tool=pmcentrez{&}rendertype=abstract.

[232] Westermark, U. K., Wilhelm, M., Frenzel, A., and Henriksson, M. A. The MYCN oncogene and differentiation in neuroblastoma. *Seminars in Cancer Biology*, 21(4):256–266, 2011. ISSN 1044579X. doi: 10.1016/j.semcancer.2011.08.001. URL http://dx.doi.org/10.1016/j.semcancer.2011.08.001.

[233] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL http://ggplot2.org.

[234] Williams, M., Purves, D., Austtime, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., and McNamara, J. O. *Neuroscience*. The Japanese classic collection. Sinauer Associates, Incorporated, Sunderland, Massachusetts U.S.A., third edit edition, 2004. ISBN 9780878937257. URL http://books.google.co.za/books?id=YZGDbwAACAAJ.

[235] Wright, L. S., Prowse, K. R., Wallace, K., Linskens, M. H. K., and Svendsen, C. N. Human progenitor cells isolated from the developing cortex undergo decreased neurogenesis and eventual senescence following expansion in vitro. *Experimental cell research*, 312(11):2107–20, jul 2006. ISSN 0014-4827. doi: 10.1016/j.yexcr.2006.03.012. URL http://www.ncbi.nlm.nih.gov/pubmed/16631163.

[236] Xia, W., Liu, Y., and Jiao, J. GRM7 Regulates Embryonic Neurogenesis via CREB and YAP. *Stem Cell Reports*, 4(5):795–810, 2015. ISSN 22136711. doi: 10.1016/j.stemcr.2015.03.004. URL http://linkinghub.elsevier.com/retrieve/pii/S2213671115000983.

[237] Yamashita, M. and Fenn, J. B. Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry*, 88(20):4451–4459, sep 1984. ISSN 0022-3654. doi: 10.1021/j150664a002. URL http://dx.doi.org/10.1021/j150664a002.

[238] Yang, M., Yang, S.-l., Herrlinger, S., Liang, C., Dzieciatkowska, M., Hansen, K. C., Desai, R., Nagy, A., Niswander, L., Moss, E. G., and Chen, J.-f. Lin28 promotes the proliferative capacity of neural progenitor cells in brain development. *Development*, 142:1616–1627, 2015. doi: 10.1242/dev.120543.

[239] Zaleska, M. M., Mercado, M. L. T., Chavez, J., Feuerstein, G. Z., Pangalos, M. N., and Wood, A. The development of stroke therapeutics: promising mechanisms and translational challenges. *Neuropharmacology*, 56(2):329–41, feb 2009. ISSN 0028-3908. doi: 10.1016/j.neuropharm.2008.10.006. URL http://www.ncbi.nlm.nih.gov/pubmed/19007799.

[240] Zaret, K. S. and Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development*, 37(7503):76–81, apr 2016. ISSN 0959437X. doi: 10.1016/j.gde.2015.12.003. URL http://linkinghub.elsevier.com/retrieve/pii/S0959437X15001446.

[241] Zhang, H.-M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.-Y. AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Research*, 40(D1):D144–D149, jan 2012. ISSN 0305-1048. doi: 10.1093/nar/gkr965. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkr965.

[242] Zhang, H.-M., Liu, T., Liu, C.-J., Song, S., Zhang, X., Liu, W., Jia, H., Xue, Y., and Guo, A.-Y. AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. *Nucleic Acids Research*, 43(D1):D76–D81, jan 2015. ISSN 1362-4962. doi: 10.1093/nar/gku887. URL http://www.ncbi.nlm.nih.gov/pubmed/25262351http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4384004http://academic.oup.com/nar/article/43/D1/D76/2439474/AnimalTFDB-20-a-resource-for-expression-prediction.

[243] Zhang, H., Deo, M., Thompson, R. C., Uhler, M. D., and Turner, D. L. Negative regulation of Yap during neuronal differentiation. *Developmental Biology*, 361(1):103–115, jan 2012. ISSN 00121606. doi: 10.1016/j.ydbio.2011.10.017. URL http://linkinghub.elsevier.com/retrieve/pii/S0012160611013236.

[244] Zhang, J. and Jiao, J. Molecular Biomarkers for Embryonic and Adult Neural Stem Cell

and Neurogenesis. *BioMed Research International*, 2015:1–14, 2015. ISSN 2314-6133. doi: 10.1155/2015/727542. URL http://www.hindawi.com/journals/bmri/2015/727542/.

[245] Zhang, P., Branson, O. E., Freitas, M. A., and Parthun, M. R. Identification of replication-dependent and replication-independent linker histone complexes: Tpr specifically promotes replication-dependent linker histone stability. *BMC Biochemistry*, 17(1):18, dec 2016. ISSN 1471-2091. doi: 10.1186/s12858-016-0074-9. URL https://bmcbiochem.biomedcentral.com/track/pdf/10.1186/s12858-016-0074-9?site=bmcbiochem.biomedcentral.comhttp://bmcbiochem.biomedcentral.com/articles/10.1186/s12858-016-0074-9.

[246] Zhang, R., Zhang, Z., Wang, L., Wang, Y., Gousev, A., Zhang, L., Ho, K.-L., Morshead, C., and Chopp, M. Activated Neural Stem Cells Contribute to Stroke-Induced Neurogenesis and Neuroblast Migration Toward the Infarct Boundary in Adult Rats. *Journal of Cerebral Blood Flow & Metabolism*, 24:441–448, apr 2004. ISSN 0271-678X. doi: 10.1097/00004647-200404000-00009. URL http://www.nature.com/doifinder/10.1097/00004647-200404000-00009.

[247] Zhang, S. C., Wernig, M., Duncan, I. D., Brüstle, O., and Thomson, J. A. In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nature biotechnology*, 19(12):1129–33, dec 2001. ISSN 1087-0156. doi: 10.1038/nbt1201-1129. URL http://dx.doi.org/10.1038/nbt1201-1129.

[248] Zhao, W., Ji, X., Zhang, F., Li, L., and Ma, L. Embryonic Stem Cell Markers. *Molecules*, 17(12):6196–6236, may 2012. ISSN 1420-3049. doi: 10.3390/molecules17066196. URL http://www.mdpi.com/1420-3049/17/6/6196/.

[249] Zizkova, M., Sucha, R., Tyleckova, J., Jarkovska, K., Mairychova, K., Kotrcova, E., Marsala, M., Gadher, S. J., and Kovarova, H. Proteome-wide analysis of neural stem cell differentiation to facilitate transition to cell replacement therapies. *Expert Review of*

Proteomics, 12(1):83–95, feb 2015. ISSN 1478-9450. doi: 10.1586/14789450.2015.977381. URL http://informahealthcare.com/doi/abs/10.1586/14789450.2015.977381.