

NORTHWESTERN UNIVERSITY

Mechanisms of Epigenetic Control of Pluripotency in
Blastula and Neural Crest Stem Cells

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Biological Sciences

By

Anjali Narsing Rao

EVANSTON, ILLINOIS

June 2019

Abstract

Mechanisms of Epigenetic Control of Pluripotency in Blastula and Neural Crest Stem Cells

Anjali Narsing Rao

The embryonic neural crest is a unique vertebrate stem cell population that has the ability to retain its stem attributes while neighboring cells in the embryo undergo lineage restriction. These cells possess multi-germ layer developmental potential and can give rise to a diverse array of derivatives such as components of the craniofacial skeleton, connective tissue and parts of the peripheral nervous system that have contributed to the evolution of vertebrates. However, how neural crest cells came to possess this remarkable potency and the mechanisms utilized by them to retain their stem cell attributes has not yet been well characterized. Studies of the mechanisms of neural crest stem cell maintenance provide a unique opportunity to explore the regulation of stem cell potential during embryonic development and give us insights into molecular players that are necessary for the control of pluripotency.

In this thesis, I explore the role of epigenetic regulation of pluripotency in blastula and neural crest stem cells and investigate the mechanisms through which neural crest cells retain their stem cell attributes during embryonic development. I found that HDAC activity and histone acetylation are critical for the maintenance of pluripotency of these two cell types. Loss of HDAC activity results in a failure to form the neural crest as well as loss of pluripotency in blastula cells. Further, depletion of HDAC activity in pluripotent blastula cells results in aberrant expression of markers of different lineages and failure to commit to a single lineage. Fascinatingly, I identified that low level of histone acetylation is a shared feature of both blastula

and neural crest stem cells suggesting that HDACs are performing a similar role in these two cell types. Using genome scale approaches, I investigated the mechanisms through which HDACs and histone acetylation control the pluripotency of blastula cells. I found the low level of histone acetylation in blastula cells read by BET proteins is also critical for maintaining the stem cell state, and BET proteins regulate pluripotency through distinct mechanisms from HDACs. Further, using mass spectrometry, I elucidated the changes in histone modifications as cells progress from a pluripotent to a lineage restricted state. Finally, through a genome wide transcriptomics study, I characterized the gene expression changes that take place during neural crest formation and identify new factors that might play important roles in the maintenance of pluripotency of these cells. Taken together, the work presented in this thesis enhances our current understanding of the epigenetic mechanisms utilized by neural crest cells to retain their stem cell attributes and provides a framework to explore further the gene regulatory circuitry that controls the pluripotency of these cells.

Dedication

I dedicate this thesis to,

My mentor, Carole,
who helped me achieve my full scientific potential,

My lab mates, and now close friends,
for all the good times working together,

My close family and friends,
for their constant love and encouragement,

My parents, Amma and Appa,
who taught me I could do anything I set my mind to,

My big brother, Arjun,
for always being there for me,

And my husband and my best friend, Kaushik,
for always believing in me and supporting me through everything.

Abstract.....	5
Dedication	4
Chapter 1	19
General Introduction	19
Embryonic Development and the Stem Cell State	20
Figure 1.1 Embryonic development is characterized by progressive lineage restriction .	22
Figure 1.2 Neural crest cells give rise to diverse derivatives	23
Neural Crest: A vertebrate innovation	24
Figure 1.3 Neural crest and vertebrate evolution.....	25
Figure 1.4 Neural crest cell specification, EMT and migration.....	26
The regulation of neural crest development.....	28
Figure 1.5 Neural crest cell derived tumors.....	30
Figure 1.6 Neural crest gene regulatory network.....	31
Neural crest as a stem cell population.....	35
The genesis of the neural crest: A new model	37
Figure 1.7 Traditional and new model for neural crest formation.....	38
Figure 1.8 Schematic representation of neural crest formation in <i>Xenopus</i> embryos	39
Figure 1.9 Control of pluripotency of neural crest cells	42
Figure 1.10 Shared transcriptional landscape of blastula and neural crest cells.....	43
Figure 1.11 Schematic representation of histone post-translational modifications	45
Figure 1.12 Chromatin remodelers: Epigenetic writers, readers and erasers	46
Epigenetic regulation: An overview	47

	6
The epigenetic stem cell state: Poised for activation.....	48
Epigenetic regulation during embryonic development	50
Figure 1.13 Dynamic changes in histone modification during embryonic development .	51
Epigenetic regulation of the neural crest.....	53
Histone acetylation in the maintenance of pluripotency	59
Histone Deacetylases: Erasers of histone acetylation	61
Figure 1.14 Classes of Histone Deacetylases	63
Figure 1.15 Classification of HDAC inhibitors	64
HDACs in embryonic development.....	65
Figure 1.16 Expression of HDACs during <i>Xenopus</i> development.....	66
HDACs in neural crest formation	68
HDACs in stem cell maintenance	70
<i>Xenopus</i> as a model organism	72
Figure 1.17 <i>Xenopus</i> as a model organism	73
Figure 1.18 <i>Xenopus laevis</i> developmental life cycle.....	74
Figure 1.19 Embryological workflow.....	75
Figure 1.20 Schematic of animal cap assay	76
Specified questions to be addressed in this thesis	78
Chapter 2	80
Histone Deacetylases are essential for the formation and maintenance of the vertebrate neural crest	80
Introduction.....	81

	7
Figure 2.1 HDAC activity is necessary for neural crest formation.....	84
Results	85
HDAC activity is essential for neural crest and placode formation.....	85
Figure 2.2 TSA treatment prevents normal embryological development.....	86
Figure 2.3 HDAC activity is required for expression of neural crest markers.	87
Figure 2.4 HDAC activity is required for proper ectodermal patterning.....	89
Figure 2.5 Loss of HDAC activity disrupts neural plate border formation.	90
HDAC activity is required for reprogramming to a neural crest state	91
HDAC1 is necessary for neural crest formation	92
Figure 2.6 HDAC activity is necessary for blastula explants to be reprogrammed to neural crest state using Pax3/Zic1.....	93
Figure 2.7 HDAC activity is necessary for blastula explants to be reprogrammed to neural crest state using Wnt/Chd.	94
Figure 2.8 HDAC1/2 activity is necessary for blastula explants to be reprogrammed to neural crest state.....	95
Figure 2.9 <i>HDAC1</i> and <i>HDAC2</i> expression during <i>Xenopus</i> development.....	96
HDAC activity is essential for proper gene expression in pluripotent blastula cells.....	97
Figure 2.10 HDAC1 activity is essential for neural crest formation.	98
Figure 2.11 <i>HDAC1</i> expression during development.	99
Figure 2.12 HDAC activity is essential for pluripotency gene expression in blastula cells.	100
Figure 2.13 Blocking HDAC activity with VPA lead to loss of pluripotency gene expression in blastula cells.....	101

Figure 2.14 HDAC activity is required pluripotency gene expression in pluripotent animal cap explants.....	102
HDAC activity is essential for the pluripotency of blastula cells.....	103
Figure 2.15 TSA treated explants do not default to an epidermal state.....	104
Figure 2.16 TSA treated explants do not default to an epidermal state.....	105
Figure 2.17 Animal cap explants can be induced to give rise to various fates	107
Figure 2.18 Loss of HDAC activity results in an inability to form neural fate	108
Figure 2.19 Loss of HDAC activity results in an inability to respond to activin and form mesoderm.....	109
Figure 2.20 Loss of HDAC activity results in an inability form endoderm lineage	110
Inhibition of HDAC activity results in precocious expression of multi-lineage markers...	111
Figure 2.21 Loss of HDAC activity results precocious expression of markers of various lineages	112
Figure 2.22 Loss of HDAC activity results precocious expression of markers of various lineages	113
Figure 2.23 TSA treatment causes low level of expression of mesoderm/endoderm markers in comparison to Activin induction.....	114
Histone acetylation increases as cell transit from pluripotency to lineage restriction	115
Figure 2.24 Validation of GFP-Tubulin transgenics.....	116
Figure 2.25 TSA treatment enhances expression of Tubb2-GFP reporter.....	117
Figure 2.26 TSA treatment causes precocious expression of Tubb2-GFP reporter	118
HDAC activity promotes retention of pluripotency.....	119
Figure 2.27 Histone acetylation is increase with lineage restriction	120

Figure 2.28 TSA treatment dramatically increases histone acetylation in animal cap explants	121
Figure 2.29 TSA treatment causes significant increase in histone acetylation in aging animal cap explants.....	122
Figure 2.30 TSA treatment causes significant increase in histone acetylation in aging animal cap explants.....	123
Figure 2.31 HDAC1 activity promotes pluripotency gene expression in animal cap explants	124
Figure 2.32 Increased HDAC1 activity prevents normal lineage restriction in animal cap explants	125
HDAC activity promotes the neural crest state.....	126
Discussion	127
Figure 2.33 HDAC1/2 perform functionally redundant roles for promoting pluripotency gene expression.....	128
Figure 2.34 HDAC1 activity does not cause animal cap explants to default to neural state	129
Figure 3.35 HDAC1 activity enhances the ability of animal explants to be reprogrammed to a neural crest state.....	130
Figure 2.36 HDAC activity promotes neural crest state.....	131
Figure 2.37 Neural crest cells retain low levels of histone acetylation similar to blastula cells	132
Figure 2.38 Histone acetylation levels in neural crest explants is similar to blastula pluripotent explants.....	133

	10
Materials and Methods.....	138
Embryological methods	138
Western blot analysis	139
RNA isolation, cDNA synthesis and qRT-PCR	139
Immunofluorescence Analysis.....	140
DNA Constructs and Inhibitors	140
Animals	141
Chapter 3	142
Mechanisms of epigenetic regulation	142
of stem cell maintenance.....	142
Introduction.....	143
Results	148
Transcriptomic analysis in pluripotent cells depleted for HDAC activity.....	148
Figure 3.1 Animal cap explants RNA-sequencing workflow.....	149
Figure 3.2 Hierarchical clustering of RNAseq biological replicates	150
Figure 3.3 TSA treatment causes equal upregulation and downregulation of gene expression in pluripotent cells	151
Figure 3.4 TSA treatment results in loss of pluripotency gene expression and upregulation of lineage genes	152
HDAC activity controls appropriate gene expression in pluripotent cells	153
Figure 3.5 Functional categories of genes of differentially expressed genes after loss of HDAC activity	154

Figure 3.6 Comparative analysis of upregulated and downregulated genes after TSA treatment	155
Figure 3.7 TSA treatment causes dramatic changes in gene expression	157
Figure 3.8 Global changes in gene expression after TSA treatment.....	158
Figure 3.9 Loss of HDAC activity disrupts normal lineage restriction	159
Loss of HDAC activity leads to failure in normal lineage restriction	160
H3K27Ac enrichment is lower at lineage genes when compared to pluripotency genes...	160
Figure 3.10 ChIP-sequencing with animal cap explants workflow	161
Figure 3.11 Quality control for H3K27Ac ChIP-seq.....	162
Figure 3.12 H3K27Ac enrichment is higher in genes downregulated after TSA treatment than upregulated by TSA	163
Figure 3.13 H3K27Ac enrichment is higher at pluripotency genes vs lineage genes	164
Figure 3.14 H3K9Ac enrichment is higher at pluripotency genes vs lineage genes by ChIP-qPCR	166
Figure 3.15 H3K27Ac enrichment is higher at pluripotency genes vs lineage genes by ChIP-qPCR	167
Figure 3.16 H3K27Ac enrichment at pluripotency and lineage genes after TSA treatment	168
Figure 3.17 Comparing H3K27Ac enrichment at genes up/downregulated after TSA treatment in Stage 13 explants	169
BET protein readers regulate pluripotency through different mechanisms than HDACs ..	170
Figure 3.18 Loss of BET activity leads to inhibition of pluripotency gene expression and loss of neural crest formation.....	171

	12
Figure 3.19 Loss of BET protein activity does not affect histone acetylation.....	172
Figure 3.20 Loss of BET protein activity does not cause upregulation of lineage genes like loss of HDAC activity.....	173
BRD proteins use distinct mechanisms to regulate pluripotency	174
Figure 3.21 HDAC and BET proteins regulate pluripotency through distinct mechanisms	175
Figure 3.22 Comparing gene expression changes after loss of HDAC and BET activity	176
Figure 3.23 Examining genes that are co-regulated by HDAC and BET activity	177
Table 3.1 Genes that are significantly changed by TSA and IBET	178
Figure 3.24 Genes affected specifically after loss of BET protein activity	180
Figure 3.25 BET proteins regulate several genes not affected by TSA treatment.....	181
Figure 3.26 Mechanism of BET activity regulating zygotic gene transcription.....	182
Figure 3.27 Motif analysis of promoters of genes altered by IBET at MBT.....	183
Global changes in histone modifications during lineage restriction.....	184
Figure 3.28 Epiproteomic analysis of early <i>Xenopus</i> development.....	185
Figure 3.29 Dynamic changes in histone modifications over developmental time	186
Figure 3.30 Epigenomic signature of early embryonic development	187
Figure 3.31 Histone modifications that increase during lineage restriction	189
Figure 3.32 Histone modifications that decrease during lineage restriction.....	190
Figure 3.33 Histone modifications that change variably during lineage restriction	191
Discussion	192
Materials and Methods.....	195

Embryological methods	13
Western blot analysis	195
RNA isolation, cDNA synthesis and qRT-PCR	196
RNA-sequencing sample preparation	197
Chromatin Immunoprecipitation (ChIP).....	198
Sample collection for mass spectrometry	199
DNA Constructs and Inhibitors	199
Animals	200
Bioinformatics Analysis:	200
RNA-sequencing Analysis.....	200
ChIP-sequencing Analysis	200
Chapter 4	202
Exploring the circuitry controlling pluripotency of neural crest stem cells.....	202
Introduction.....	203
Results	207
Transcriptomic analysis of neural crest formation.....	207
Differential analysis identifies differences between neural crest explants and control explants	208
Figure 4.1 Animal cap explants RNA-sequencing workflow	209
Figure 4.2 Hierarchical clustering of RNAseq biological replicates	210
Figure 4.3 Global changes in gene expression during lineage choices.....	211
Figure 4.4 Differential expression between control and neural crest explants	213

Figure 4.5 Changes in gene expression between control and neural crest explants at blastula stage.....	214
Figure 4.6 Changes in gene expression between control and neural crest explants at early neurula stage	215
Figure 4.7 Changes in gene expression between control and neural crest explants at late neurula stage	216
Analysis of early neural crest development reveals an early and late neural crest signature	217
Figure 4.8 Changes in gene expression between neural crest explants at blastula and early neurula stages.....	218
Figure 4.9 Changes in gene expression between neural crest explants at early and late neurula stages.....	219
Figure 4.10 Expression of neural crest genes	221
Figure 4.11 Expression of late neural crest genes (specification)	222
Figure 4.12 Expression of early neural crest genes (border)	223
WGCNA identifies genes that highly correlated with an early and late neural crest signature.....	224
Figure 4.13 Weighted Co-expression Network Analysis	225
Figure 4.14 Sample clustering to detect outliers.....	226
Figure 4.15 Determination of power beta value in WGCNA	227
Figure 4.16 WGCNA on top 15000 most varying genes.....	229
Figure 4.17 Relating modules to sample type – developmental time/cell state.....	230
Figure 4.18 Module blue genes are highly correlated with early neural crest state.....	231

	15
Figure 4.19 Module yellow genes are highly correlated with late neural crest state.....	232
Figure 4.20 Schematic representation of principal component analysis and sparse PCA	234
Figure 4.21 PCA identifies three significant components	235
Figure 4.22 Gene expression data projected onto principal component space	236
Principal Component Analysis identifies 3 significant components that explain dynamics of neural crest cell fate decisions	237
Figure 4.23 Biplot of PC1 vs PC2	238
Figure 4.24 Determination of Lambda penalty parameter for sparsePCA	239
Figure 4.25 Sparse principal components from gene expression data.....	240
Figure 4.26 Sparse principal components from gene expression data.....	241
Comparing early and late neural crest state from sparsePCA and WGCNA.....	243
Figure 4.27 Comparing WGCNA and sparsePCA	244
Table 4.1 Early neural crest (border) genes from WGCNA and sparsePCA.....	245
Table 4.2 Late neural crest (progenitor) genes from WGCNA and sparsePCA.....	246
Discussion	247
Materials and Methods.....	250
Embryological methods	250
RNA-sequencing sample preparation	251
Animals	252
Bioinformatics Analysis: RNA-sequencing Analysis.....	252
Chapter 5	254
General Discussion.....	254

	16
HDAC activity is essential for blastula and neural crest cells	256
Figure 5.1 Old and new model of neural crest formation	257
Histone acetylation and stem cell maintenance.....	261
The epigenomic landscape of embryonic development	264
Gene regulatory circuitry controlling neural crest stem cell state	266
Concluding remarks	268
Significance of the thesis work.....	270
References	272
Chapter 6	293
Appendix 1:	293
Snail proteins are necessary for stem cell maintenance	293
Figure 6.1 <i>Snail1</i> expression profile during early <i>Xenopus</i> development	296
Figure 6.2 <i>Snail2</i> expression profile during early <i>Xenopus</i> development.....	297
Figure 6.3 Cartoon depicting domains of Snail1 and Snail2	298
Snail proteins are required for pluripotency of blastula cells	301
Figure 6.4 Snail function is necessary for neural crest formation	302
Figure 6.5 Snail function is necessary for pluripotency gene expression.....	303
Figure 6.6 Snail function is necessary for blastula cells to form mesoderm	304
Figure 6.7 Snail function is necessary for blastula cells to form endoderm	305
Figure 6.8 Snail function is necessary for blastula cells to form epidermis	306
SNAG domain is essential to mediate Snail protein function in blastula cells	307
Figure 6.9 Snail function is necessary for neural crest formation	308

	17
Figure 6.10 Snail1 Δ SNAG results in loss in epidermal formation.....	309
Snail1 interacts with the core pluripotency network.....	310
Figure 6.11 Snail function is necessary for neural crest formation	311
Figure 6.12 Snail1 but not Snail2 interacts with cMyc.....	313
Figure 6.13 <i>Xenopus</i> and <i>Drosophila</i> Snail1 interacts with cMyc	314
Snail1 and Snail2 interact with HDACs in <i>Xenopus</i>.....	315
Figure 6.14 Temporal control of Snail1/2 interaction with HDAC1	316
Figure 6.15 Snail1/2 interaction with HDAC1 is partially dependent on acetylation activity.....	317
Chapter 6	320
Appendix 2:	320
Bioinformatics Analysis scripts	320
RNA-Sequencing Analysis	321
FastQC Analysis – Shell script	321
STAR Alignment – Shell script	321
HTSEQ Count – Shell script.....	322
RSEM – Shell script	322
RSEM – Prepare reference transcriptome.....	322
RSEM – Calculate expression	323
DESeq2 – R script.....	324
WGCNA – R script.....	331
PCA – R script	337

	18
Heatmaps – R script	342
ChIP-Seq Analysis	343
Alignment using Bowtie2 – Shell script	343
Building reference genome	343
Alignment	343
Filter Bam files – Shell script	344
Deeptools – Shell script	345
Convert Bam to BigWig	345
computeMatrix	345
plotHeatmaps	346
Generating Upstream sequences for motif analysis – Rscript	346
Proteomics Analysis – Python Script	347

Chapter 1

General Introduction

Perhaps one of the most fascinating questions in biology is how a multicellular organism develops from a single-celled fertilized egg. The fertilized egg possesses the ability to give rise to all the cell types of an adult organism as well contains the blueprints that dictate the patterning and complexity of the entire body plan. However, this ability is transient and as the egg divides, each individual cell retains only partial capability to give rise to different cell types. Indeed, embryonic development is characterized by a gradual and progressive restriction of developmental potential as the cell progresses from a totipotent fertilized embryo, to pluripotent embryonic stem (ES) cells to multipotent progenitor cells. As cells proceed through development, and travel down Waddington's epigenetic landscape, their ability to give rise to different derivatives is greatly reduced (Goldberg, Allis, & Bernstein, 2007). An exception to this rule, however, is the neural crest, a unique vertebrate cell type, which arises in the early ectoderm, and yet possesses broad multi germ layer developmental potential (Prasad, Sauka-Spengler, & LaBonne, 2012). How neural crest cells retain their stem cell attributes in a cellular environment that promotes lineage restriction remains an area of intense investigation. Here, we explore the epigenetic mechanisms utilized by these unique cells to adopt a suspended state of pluripotency. We identify a novel mechanism for histone acetylation and HDAC activity to maintain the neural crest stem cell state and elucidate transcriptional and epigenetic mechanisms that contribute to the formation of this cell type.

Embryonic Development and the Stem Cell State

Embryonic stem cells (ESCs) are pluripotent cells that have the capacity to self-renew and are derived from the Inner Cell Mass (ICM) of the blastocyst of vertebrate embryos (Niwa, 2007; Young, 2011). The analogous cell population in *Xenopus*, are the deep cells of the blastula

roof (animal pole cells) (Snape, Wylie, Smith, & Heasman, 1987). These cells have the potential to give rise to derivatives of all three germ layers (ectoderm, mesoderm and endoderm) and hence all cell lineages of the embryonic and adult organism (Solter, 2006; Young, 2011). The transcriptional and epigenetic circuitry of these cells allows them to proliferate in the same state (self-renew) and yet remain poised to respond to developmental cues to form various lineages (differentiate) (Young, 2011). A core network of transcription factors namely, *Oct4*, *Nanog* and *Sox2*, have been identified to be essential for controlling and regulating the ES cell state (Chambers, 2004; Chambers & Tomlinson, 2009; X. Chen, Vega, & Ng, 2008; Takahashi et al., 2007). The *Xenopus* orthologs of these factors are *Oct25/60/91*, *Vent1/2*, *Sox2/3*, *cMyc* respectively (King, Roberts, & Eisenman, 1986; G. M. Morrison & Brickman, 2006; Penzel, Oschwald, Chen, Tacke, & Grunz, 1997; Scerbo et al., 2012). These factors are also expressed in the *Xenopus* animal pole cells suggesting conserved roles of these factors between species. It has been suggested that by modulating their protein levels in response to differentiation signals, these factors are able to function both in maintenance of the stem cell state as well as during lineage commitment (Chambers & Tomlinson, 2009; Thomson et al., 2011).

As development proceeds, the cells within the embryo undergo a sequential set of events that results in complete lineage restriction. The pluripotent cells give rise to multipotent progenitor cells of the three germ layers and get successively lineage restricted to terminally differentiated cell types (Figure 1.1) (Berdasco & Esteller, 2011). This coordinated effort is the interaction of various signaling molecules and transcription factors that direct the formation of the complex vertebrate body plan. This paradigm of progressive lineage restriction dictates that cells gradually lose their ability to self-renew and are restricted in their capability to give rise to

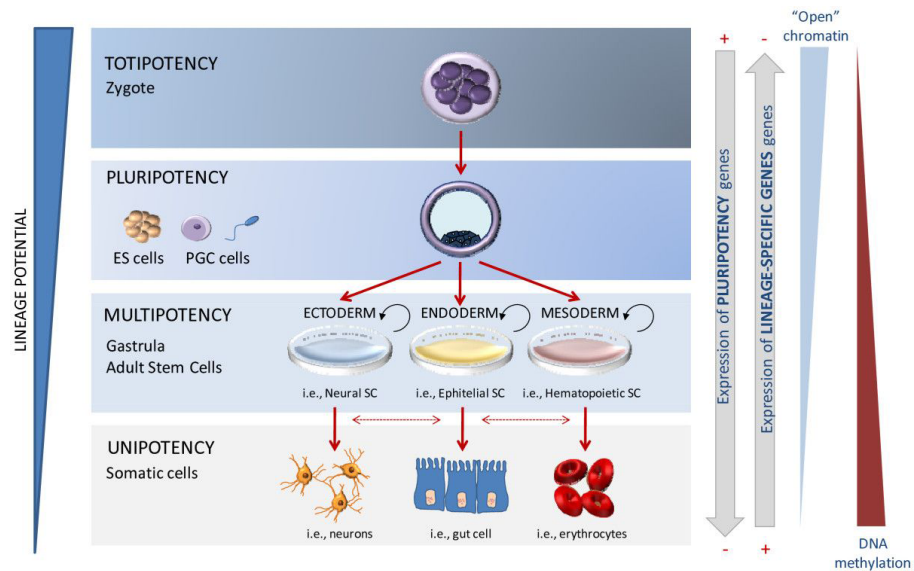


Figure 1.1 Embryonic development is characterized by progressive lineage restriction

The totipotent zygote gets progressively lineage restricted as development proceeds and gives rise to unipotent terminally differentiated cells. This process is molecularly characterized by a loss of pluripotency gene expression and upregulation of lineage specific genes. (Adapted from Berdasco and Estellar,

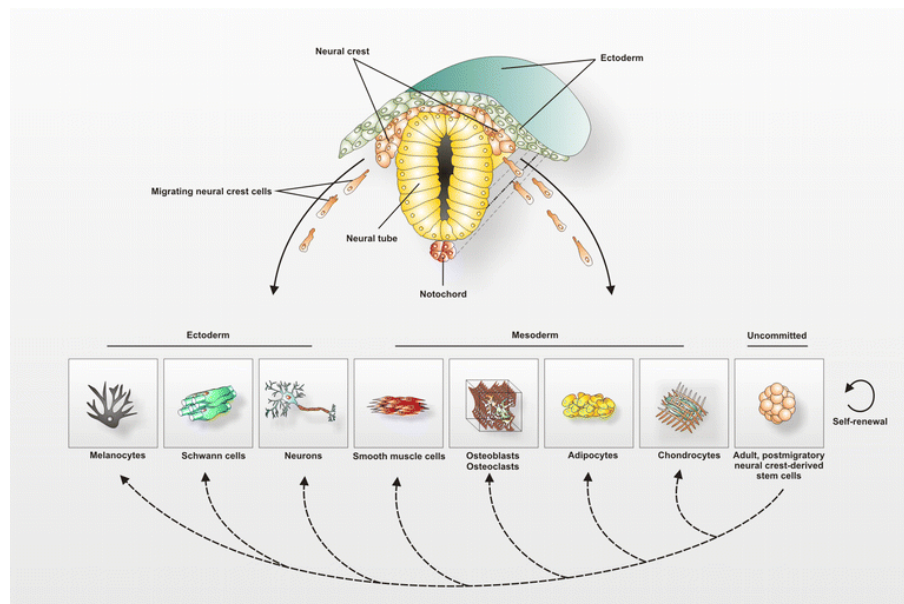


Figure 1.2 Neural crest cells give rise to diverse derivatives

Neural crest cells are a stem cell population that have the ability to give rise to derivatives as well as undergo self-renewal. They have multi-germ layer developmental potential and give rise to ectodermal as well as mesodermal derivatives which contribute to several embryonic and adult structures in the vertebrate body plan. (Adapted from Kaltschmidt B et al, 2012)

other cell lineages. For instance, a cell that has been lineage restricted to an ectodermal state will never be able to give rise to endodermal or mesodermal derivatives. Thus, inductive cues within the embryo instruct the cells to form particular fates and disrupt the gene regulatory network that maintains the pluripotent state.

Neural crest progenitors represent one of the few examples during embryonic development that defy this paradigm of progressive lineage restriction. Neural crest cells are a developmental and evolutionary novelty and have the capability to give rise to mesodermal derivatives *in vivo* in spite of their ectodermal origins (Prasad et al., 2012). They migrate extensively and contribute to numerous diverse cell types of the embryonic and adult organism. These characteristics have allowed for the formation of unique cell types which are characteristic of vertebrates. The embryonic neural crest, thus, presents an excellent model for studying mechanisms for establishment and maintenance of pluripotency, but also differentiation and cell migration, as well as the evolution of vertebrates.

Neural Crest: A vertebrate innovation

The neural crest is an embryonic stem cell population that has the novel property of multi-germ layer developmental potential. The neural crest arises from the prospective ectoderm at the border between the neural ectoderm and non-neural ectoderm and gives rise to derivatives that are considered bona fide ectodermal cell types like melanocytes, glia and sensory neurons but also form mesodermal derivatives like smooth muscle cells, osteoblasts and chondrocytes (Figure 1.2) (Bronner & LeDouarin, 2012; B. Kaltschmidt, Kaltschmidt, & Widera, 2012). Neural crest cells contribute to a variety of structures within the vertebrate body plan, including critical

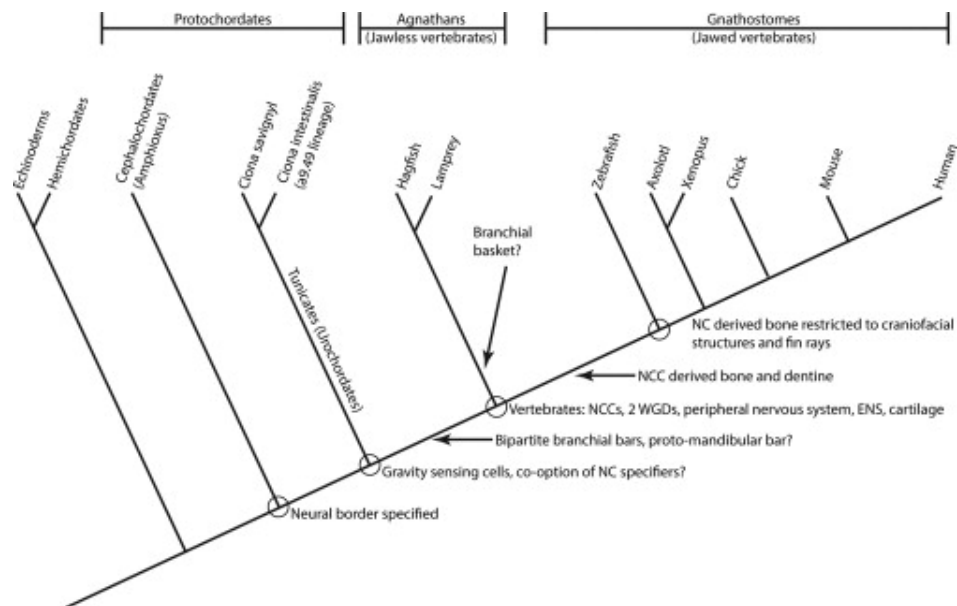


Figure 1.3 Neural crest and vertebrate evolution

Neural crest cells emerged at the base of the vertebrates and gave rise to the “new” head and skeletal tissues that distinguishes the vertebrates from protochordates.

(Adapted from Munoz and Trainor, 2015)

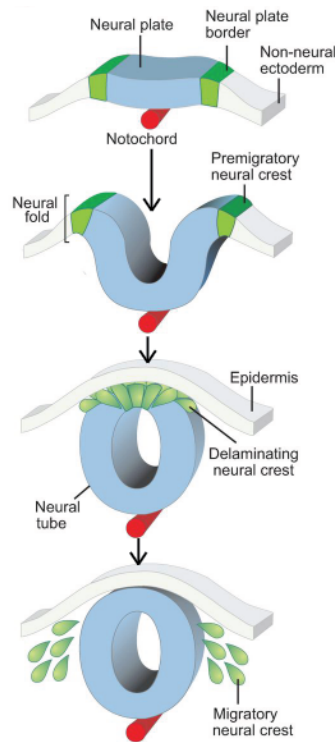


Figure 1.4 Neural crest cell specification, EMT and migration

At the onset of neurulation, neural crest cells are formed at the border of the neural ectoderm and the non-neural ectoderm. As the neural tube closes, these cells undergo epithelial to mesenchymal transition (EMT) and delaminate from their initial position and migrate to various parts of the developing embryo. (Adapted from Simoes-Costa and Bronner, 2015)

components of the craniofacial skeleton such as the jaws and skull, as well other structural cell types like bone, cartilage, connective tissue and the paired sensory organs. The unique property of multi-germ layer developmental potential has led to neural crest cells to be referred to as ectomesenchyme, comprising the “fourth germ layer” and is considered to be the defining feature of the vertebrate phylum (Bronner & LeDouarin, 2012; Hall, 2000). The evolutionary introduction of the neural crest at base of the vertebrates led to the development of a “new head”, with a distinct craniofacial skeleton, sensory organs, and complex brain (Gans & Northcutt, 1983; Le Douarin & Dupin, 2012). This, in turn, led to a change in lifestyle between protochordates and vertebrates, promoting a more active predatory lifestyle and higher cognitive abilities. Indeed, studies in lamprey, the most basal extant jawless vertebrate, identified neural crest-like cells at embryonic stages, and confirmed the contributions of these cells to several larval and adult tissues (Medeiros & Crump, 2012; Muñoz & Trainor, 2015). Thus, the emergence of the neural crest is considered a key innovation driving vertebrate evolution and sets vertebrates apart from all other metazoans (Hall, 2000) (Muñoz & Trainor, 2015) (Figure 1.3).

First identified by Wilhelm His in 1868 and named based on the position they occupied at the crest of the closing neural tube, the neural crest is a strip of cells lying between the neural tube and presumptive epidermis (His, 1868). From this initial position, neural crest cell undergo epithelial to mesenchymal transition (EMT) and delaminate from the neural tube and migrate to various parts of the embryo (Figure 1.4) (Simões-Costa & Bronner, 2015). These cells migrate in distinct streams controlled by their position on the anterior-posterior axis and differentiate in a region-specific manner. For example, the cephalic NC is considered to give rise to the cartilage, bone and connective tissues for the skull and facial structures, while the vagal NC cells

contribute to the aortic outflow tract of the heart and the connective tissue, neuronal and glial cells of the enteric nervous system, and the trunk NC cells contributes to the most posterior region of the dorsal root ganglia, and the adrenal gland and produces melanocytes (Szabó & Mayor, 2018).

Since neural crest cells contribute to multiple tissues, defects in neural crest formation translates into a large number of birth defect and abnormalities, syndromes called neurocristopathies. These include cleft lip and palate, and the CHARGE, Treacher Collins, Waardenburg and Collins syndromes, which are all caused by genetic defects that lead to improper neural crest formation or migration (Vega-Lopez, Cerrizuela, Tribulo, & Aybar, 2018). Neural crest cells also serve as an excellent model for the study of metastasis and malignancies. Malignant cells adopt several programs of neural crest formation, including EMT, migration and tissue invasion. Further, neural crest derived tissues themselves lead to several childhood and adult cancers (Figure 1.5) (Maguire, Thomas, & Goldstein, 2015). This includes melanoma, a tumor of neural crest derived melanocytes, as well as neuroblastoma, a variably aggressive tumor of neuroblasts, and the most prevalent childhood solid tumor (Maguire et al., 2015). Thus, understanding the regulation of neural crest formation is vital for both preventing neurocristopathies and syndromes, as well as for better understanding and treating various cancers.

The regulation of neural crest development

The formation of the neural crest can be considered a highly controlled multi-step process involving the specification, migration and ultimate differentiation of neural crest cells (Sauka-Spengler & Bronner-Fraser, 2008; Simões-Costa & Bronner, 2015). A complex gene regulatory

network orchestrates the precise timing and formation of the neural crest (Figure 1.6) (Bronner & Simões-Costa, 2016). Several signaling molecules and transcription factor networks have been identified to be functionally required for this process. This network is essential for both maintaining the stem cell state and preventing differentiation, but later promoting EMT, migration and differentiation of these cells.

The neural crest is considered to arise from a broad competence domain formed between the neural ectoderm and the non-neural ectoderm. This region called the neural plate border is a site where key patterning events take place distinguishing these cells from the adjacent ectoderm and give rise to two key populations: the neural crest and the cranial placodes (Groves & LaBonne, 2014; Milet & Monsoro-Burq, 2012). The formation of the neural plate border has been shown to require the combination of signals, including Bone Morphogenetic Protein (BMP), Fibroblast Growth Factor (FGF), WNT and Notch signaling and is thought to occur during early/mid gastrulation (LaBonne & Bronner-Fraser, 1999; Milet & Monsoro-Burq, 2012). These signaling molecules are vital to the formation of the neural crest and have been shown to be sufficient in certain cases to set up the neural crest state. Intermediate levels of BMP signaling are necessary but not sufficient for neural crest formation (LaBonne & Bronner-Fraser, 1998; Nordin & LaBonne, 2014). There is evidence for a two-signal model, suggestive that intermediate levels of BMP combined with other signals, would induce the neural crest state. Indeed, WNT or FGF2 combined with an inhibitor of BMP such as chordin/noggin have been shown to be sufficient to set up a neural crest state in *Xenopus* animal caps (LaBonne & Bronner-Fraser, 1998). Interestingly, it has also been shown that WNT8 might be functioning downstream of FGF, suggestive that FGFs might regulate neural crest formation in a WNT dependent manner (Hong, Park, & Saint-Jeannet, 2008; LaBonne & Bronner-Fraser, 1998). Studies in frog, chick

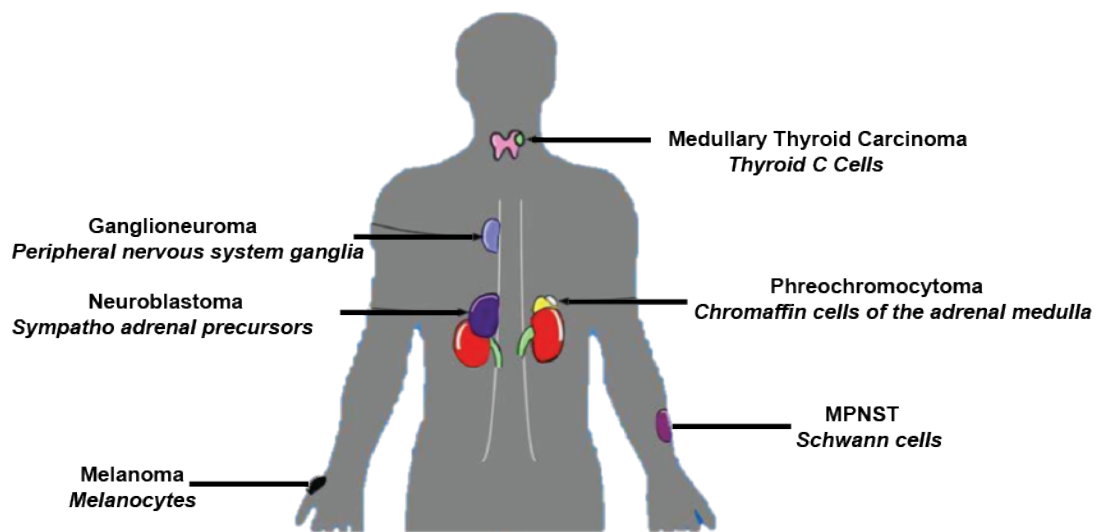


Figure 1.5 Neural crest cell derived tumors

Neural crest cell derived tissues give rise to several adult and pediatric malignancies. These include melanoma, a tumor in the neural crest derived melanocytes, and neuroblastoma of the sympatho-adrenal precursors, and medullary thyroid carcinoma of the Thyroid C cells etc. (Adapted from Maguire et al, 2015)¹¹

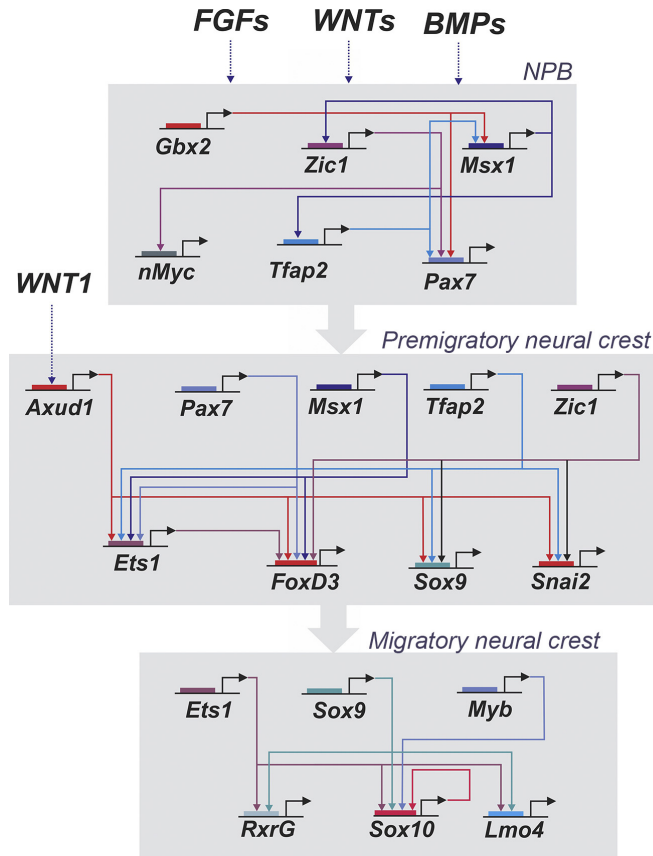


Figure 1.6 Neural crest gene regulatory network

The process of neural crest formation involves tight spatial and temporal control. Firstly, the WNTs, FGFs and BMPs set up a zone of competence and activate the neural plate border specifiers. These factors act in concert to activate a set of neural crest specifier genes which establish the premigratory neural crest state. The reiterative use of neural crest specifier genes as well as other genes controls the EMT and eventual migration of these cells. (Adapted from Bronner and Simoes-Costa, 2016)

and zebrafish embryos have shown that Delta/Notch signaling also plays an important role in neural crest formation (Cornell & Eisen, 2002; Glavic, Silva, Aybar, Bastidas, & Mayor, 2004; Wakamatsu, Maynard, & Weston, 2000). While the individual requirement of these signals seems to vary in a timing and species dependent manner, it is clear that these extracellular environmental signals set into motion a cascade of events and are essential for establishing and maintaining the neural crest stem cell state.

A precise temporal and spatial integration of this complex set of molecular signals results in the expression of a suite of transcription factors called neural plate border specifiers. Studies have shown that these early markers of the neural plate border are directly downstream of these signaling pathways and include genes like *Pax3*, *Pax7*, *TFAp2*, *Zic1*, *Msx1*, *Msx2* etc. The expression of these genes are initiated at early gastrula stages and they are expressed continuously in the neural crest forming region (Milet & Monsoro-Burq, 2012). Interestingly, recent work from our lab has identified that many of these markers including *Pax3*, *Zic1*, *TFAp2* are actually expressed even earlier in the blastula pluripotent cells and then get restricted to the border regions, suggestive that the presumptive neural plate border cells might be specified earlier than was previously thought (Buitrago-Delgado, Nordin, Rao, Geary, & LaBonne, 2015). These border specifiers delineate the NC territory, and separate them from other border derived structures like pre-placodal cells and Rodon-Beard primary neurons as well as prevent the formation of other ectodermal cell types (Milet & Monsoro-Burq, 2012; Sauka-Spengler & Bronner-Fraser, 2008). Perturbation of levels of neural crest border specifiers has shown to be detrimental to neural crest development as well shown to be sufficient to promote neural crest formation confirming the integral role of these specifiers in the neural crest gene regulatory network. When *TFAp2* is depleted, there is a loss of expression of other border specifiers

suggestive that TFAP2 is essential to initiate the border region and might be upstream in the gene-regulatory cascade (de Crozé, Maczkowiak, & Monsoro-Burq, 2011). Indeed, recent studies in human ESCs have shown that TFAP2 allows for the formation of permissive chromatin states and functions like a pioneer factor during neural crest formation (Rada-Iglesias et al., 2012). Further, Pax3 and Zic1 co-expression is sufficient to cause ectopic neural crest formation in whole embryos. (Hong & Saint-Jeannet, 2007; Milet, Maczkowiak, Roche, & Monsoro-Burq, 2013; Plouhinec et al., 2014). Interestingly, over expression of precise levels of Pax3/Zic1 is sufficient to form neural crest cells in *Xenopus* animal cap explants and play a key role in the development of pre-migratory neural crest cells within the embryo. The requirement of these border specifiers is likely to be evolutionarily conserved as studies from lamprey have shown that knockdown of these border specifiers using morpholinos results in abnormal development and loss of neural crest derivatives (Sauka-Spengler, Meulemans, Jones, & Bronner-Fraser, 2007; Simões-Costa & Bronner, 2013). Further, the neural plate border specifiers are also responsible for sharpening the boundary between the neural plate and border regions with inhibitory interactions with neural factors. Mis-expression of various specifiers like TFAP2, Msx1 in the neural plate leads to inhibition of Sox2/3 expression, while knockdown of these genes causes an expansion of the neural domain (Luo, Lee, Saint-Jeannet, & Sargent, 2003). Thus, neural plate border specifiers play a dual role, promoting neural crest cell fate while simultaneously restricting the formation of other cell types.

The synergistic expression of neural plate border specifiers, in turn, induces the expression of neural crest specifier genes which are characteristic of neural crest precursors and are responsible for maintaining the potency as well as the epithelial to mesenchymal transition of these cells. The earliest expressed neural crest specifier genes are *Snail1/2*, *Sox8/9*, *cMyc*,

FoxD3, *Twist*, *Ets1*, *TFAP2* and *Id3* (Prasad et al., 2012; Sauka-Spengler & Bronner-Fraser, 2008). As with the neural plate border specifiers, the timing of expression as well as the individual requirement varies between species. The proto-oncogene *Myc*, is considered to be one of the earliest expressed factors in the neural crest. *Myc* and its downstream effector *Id3* have been shown to be essential for the potency of neural crest cells and are considered to be a bridge between the neural plate border and neural crest state (Bellmeyer, Krase, Lindgren, & LaBonne, 2003; Light, Vernon, Lasorella, Iavarone, & LaBonne, 2005). Recent work in our lab has also identified that neural crest specifiers like *Snail1*, *FoxD3* and *Id3* are expressed much earlier than previously thought, in the blastula pluripotent cells. *Snail2* and *Sox9* expression seems to be turned on much later, coming on at mid-to late gastrulation in *Xenopus* embryos (Buitrago-Delgado et al., 2015) (Buitrago-Delgado, Schock, Nordin, & LaBonne, 2018; LaBonne & Bronner-Fraser, 2000). Ablation of any of these factors has shown to cause a failure in neural crest formation, as seen by loss of expression of all neural crest markers as well as failure to form neural crest derivatives. In *Xenopus* and in chick, *Snail2* has been shown to be essential for neural crest formation. Knockdown of *Snail1/2* function with a dominant negative results in a loss of neural crest formation (LaBonne & Bronner-Fraser, 2000). Similarly, *Foxd3* and *Sox9* have also been shown to be essential for the formation of the neural crest (N. Sasai, Mizuseki, & Sasai, 2001; Taylor & LaBonne, 2005).

After their specification, the neural crest cells that reside along the dorsal neural tube undergo epithelial to mesenchymal transition (EMT). This entails the reiterative use of the same genes that were involved in specifying the neural crest state. Neural crest specifiers like *Snail1/2*, *FoxD3* and *Twist* have been shown to involved in the EMT and cell migration in the neural crest as well as other systems. Interestingly, the *Snail1/2* proteins have been shown to directly repress

E-cadherin expression, and to be involved in EMT in cell culture as well as mesoderm formation (Leptin, 1991; Nieto, 2002; Peinado, Ballestar, Esteller, & Cano, 2004). Snail2 has been shown directly regulate neural crest EMT by mediating transitions in cell junction assembly, motility and adhesion of these cells (Sauka-Spengler & Bronner-Fraser, 2008). Once these cells have delaminated, both cell-cell and cell-environment interactions provide directional cues that target these cells to the correct final destination within the embryo. Interestingly, the timing of neural crest delamination varies between species. While delamination begins while the neural tube is still open in *Xenopus* and mouse, in chick neural crest cells migrate after the neurulation is completed (Szabó & Mayor, 2018). After delamination, neural crest cells invade the tight space between the epidermal and mesodermal layers as a continuous sheet and migrate along segmentally organized paths till they reach their final destination where they differentiate into diverse cell lineages (Szabó & Mayor, 2018).

Neural crest as a stem cell population

Decades of research have shown neural crest cells have been shown to be a multipotent stem cell population (Dupin & Sommer, 2012). Neural crest cells have been shown to have the ability to self-renew as well differentiate to other cell types, the primary characteristics of a stem cell. *In vitro* clonal analysis and *in vivo* labeling experiments have demonstrated that neural crest cells have the ability to self-renew as well differentiate into different cell types (Baroffio, Dupin, & Le Douarin, 1988; Bronner-Fraser & Fraser, 1988; Calloni, Glavieux-Pardanaud, Le Douarin, & Dupin, 2007; Calloni, Le Douarin, & Dupin, 2009; Stemple & Anderson, 1992) (S. J. Morrison, White, Zock, & Anderson, 1999; Trentin, Glavieux-Pardanaud, Le Douarin, & Dupin, 2004). Single cell lineage tracing experiments in chick embryos have shown that neural crest

cells can give rise to various derivatives *in vivo* (Bronner-Fraser & Fraser, 1988). Using primary rat neural crest cells, it was shown that these cells can self-renew *in vitro* for 10 days when grown at clonal density (Stemple & Anderson, 1992). Strikingly, using multi-color labeling of pre-migratory and post migratory neural crest in mouse, it was shown that the majority of neural crest progenitors are multipotent *in vivo* (Baggiolini et al., 2015). This suggests that NC cells are multipotent stem cells and must express several regulatory factors tasked with maintaining their potency and stem cell attributes.

Neural crest cells express several transcription factors known to be associated with the maintenance of stem cell attributes and multipotency. The most striking among these is c-Myc, a factor known for its role in stem cell renewal and iPSC reprogramming (Takahashi et al., 2007). *cMyc* is initially expressed broadly in the neural plate border region in *Xenopus*, and further gets restricted to the NC forming regions as development proceeds (Bellmeyer et al., 2003). Using loss-of-function studies, c-Myc and its downstream mediator, Id3 have both been shown to be essential regulators of neural crest specification in *Xenopus* (Bellmeyer et al., 2003; Light et al., 2005). Apart from *cMyc*, we have recently discovered other factors linked with pluripotency to be expressed broadly in the neural crest forming regions which includes genes like *Oct60*, *Vent2* and *Sox3* (Buitrago-Delgado et al., 2015).

Interestingly, several NC specifiers like *Snail1*, *FoxD3* and *Sox5* are maternally expressed from early cleavage stages to the pluripotent blastula animal pole cells (Buitrago-Delgado et al., 2015). The expression of these factors in the blastula cells is suggestive of a functional role in maintaining the potency of these cells. This is not surprising, as some of these factors have been previously shown to have links with the maintenance of stem cell attributes in embryonic stem cells. *Foxd3*, in particular, is extremely vital for maintaining ES cells in the ICM

during embryonic development. In fact, it has been shown that *FoxD3*^{-/-} mice embryos undergo abnormal development and do not survive till birth (L. A. Hanna, Foreman, Tarasenko, Kessler, & Labosky, 2002). Another study has found through *in vitro* clonal analysis that *FoxD3* is required for maintaining self-renewal and multipotency of NC progenitors (Mundell & Labosky, 2011). More recently, Snail factors have also been linked with promoting stem cell attributes. It has been suggested that *Snail* confers mesenchymal stem cell characteristics to cancerous cells through EMT, as well mediates ES cell maintenance and lineage commitment in mice (Batlle et al., 2013; Y. Lin et al., 2014; Mani et al., 2008). Interestingly, the pluripotency factor *Lin28a* was also recently shown to be necessary for maintaining the neural crest stem cell identity (Bhattacharya, Rothstein, Azambuja, & Simões-Costa, 2018). Using a new technique called Spatial Genomic Analysis and unbiased hierarchical clustering, it was demonstrated that neural crest cells have a unique stem cell niche in the dorsal neural tube characterized by the expression of pluripotency genes and neural crest factors. This study identified that several pluripotency factors like *Oct4*, *Nanog* and *Klf4* are preferentially expressed in the dorsal neural tube (prospective neural crest) in chick embryos and distinguishes these cells from neural stem cells (Lignell, Kerosuo, Streichan, Cai, & Bronner, 2017). Thus, there is a large amount evidence that supports that the neural crest is unique multipotent stem cell population in the embryo and the ability of neural crest cells to give rise to variety of derivatives can be attributed to this stem cell potential. However, there is yet not a clear picture on how neural crest cells came to possess this developmental potential and maintain their stem cell attributes.

The genesis of the neural crest: A new model

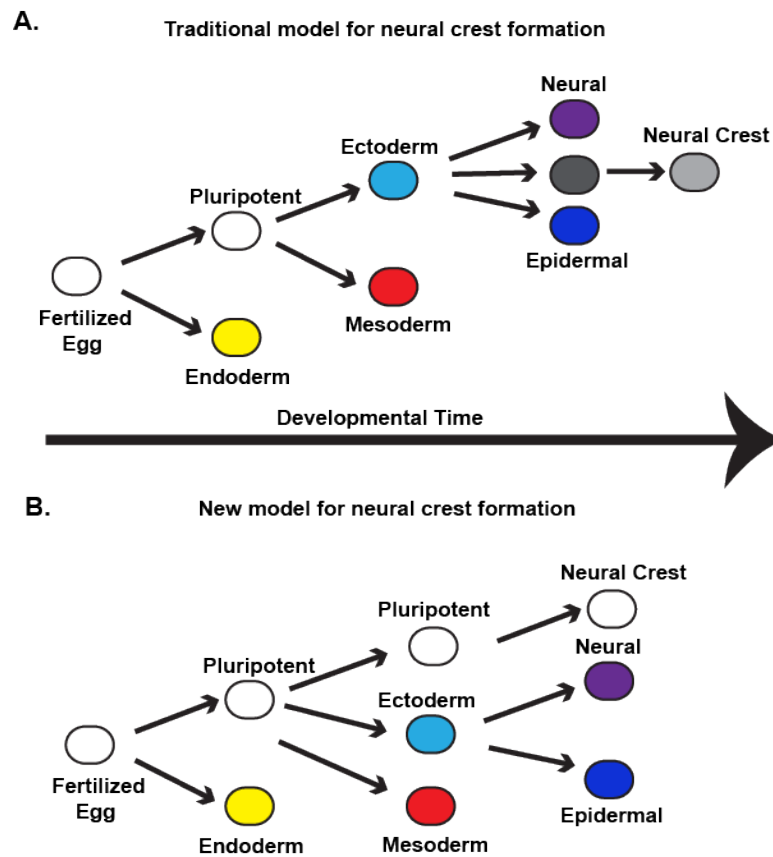


Figure 1.7 Traditional and new model for neural crest formation

Schematic depicting traditional model for neural crest formation wherein a subset of the cells in the ectoderm gain developmental potential through an inductive event and these give rise to neural crest cells. Schematic depicting a new model for neural crest formation wherein a subset of cells in the ectoderm have the ability to retain their stem cell attributes and these give rise to neural crest cells. 11

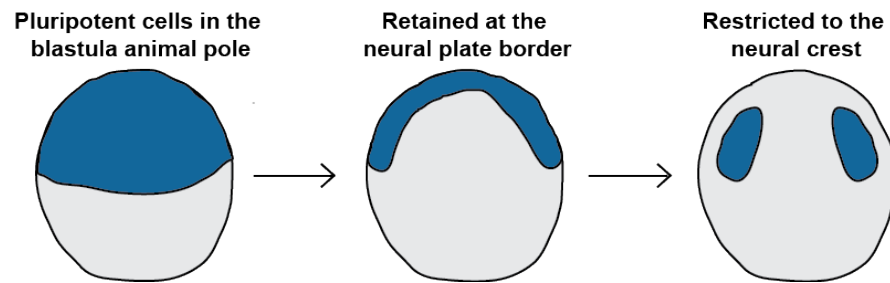


Figure 1.8 Schematic representation of neural crest formation in *Xenopus* embryos

Neural crest cells are formed through a retention of stem cell attributes. A subset of blastula pluripotent cells (blue) have the unique capability to retain their stem attributes, and these cells get restricted to the neural plate border region during gastrulation. At neurulation, these cells give to the neural crest which has the stem cell attributes of their blastula ancestors.

The stem cell potential of neural crest cells has puzzled scientists for decades as they seemingly defy the embryonic paradigm of progressive lineage restriction. Much of the work in the field has been directed into determining the developmental processes that lead to these cells possessing greater developmental potential than the cells from which they were embryologically derived. The classical model that was used to describe the apparent gain of potential of neural crest cells hypothesized that, these cells arise due to an induction of stem cell potential at the neural plate border (Figure 1.7A). This induction process occurs during early gastrulation and involves intermediate levels of BMP and FGF signaling as well as WNT/Notch and results in the expression of early neural crest markers and resetting of the potency of these cells. Thus, the traditional model for neural crest formation postulated that a subset of cells at the neural plate border gain stem cell attributes as the result of an inductive event, leading to the formation of the neural crest stem cells with greater developmental potential than the cells from which they were derived. The induction model, although an elegant solution to the question of neural crest cell potential, was never completely satisfying for biologists as it defied the paradigm of progressive restriction of developmental potential as depicted by Waddington's developmental landscape (Hoppler & Wheeler, 2015). Recent work from our lab identified that neural crest cells share much of the transcriptional circuitry of blastula pluripotent cells. Based on these findings, our lab proposed an alternate and more parsimonious model for neural crest formation which hypothesized that neural crest cells arise due to a retention of potential at the neural plate border (Figure 1.7B). This model suggests that a subset of blastula cells have the capability to retain their stem cell attributes as they proceed through development, and these cells ultimately give rise to the neural crest cells. By delaying the onset of lineage restriction when compared to their cellular neighbors, neural crest cells retain a similar developmental potential to blastula cells and

are hence able to give rise to cell types of a multitude of lineages. Thus, instead of a gain in stem cell attributes, this new model suggests that neural crest cells never lose their stem cell potential and simply delay the onset of lineage restriction (Figure 1.8).

The big question that this model raises is how do neural crest cells evade cues that instruct other cells in the embryo to start lineage restriction? What mechanisms have these cells adopted to maintain their pluripotency? Further, what features are shared between blastula cells and neural cells and what are the dissimilarities between these cell types? And finally, what are the key players that are involved in the maintenance of pluripotency of the neural crest?

The process of maintenance of pluripotency would necessitate the complex interplay of signaling molecules, transcription factors and epigenetic modifiers (Figure 1.9) Previous work in the lab identified that Snail1 and Sox5, two transcription factors that have been previously shown to be essential for neural crest formation are also required for the maintenance of pluripotency of blastula cells (Buitrago-Delgado et al., 2015; LaBonne & Bronner-Fraser, 2000; Nordin & LaBonne, 2014)(Chapter 6). When Snail protein function is blocked using a dominant negative form of the protein or Sox5 function is blocked using a translation blocking morpholino in *Xenopus* embryos, we observe a strong loss of pluripotency gene expression in these cells (Buitrago-Delgado et al., 2015)(Chapter 6). Alongside the loss of gene expression, animal cap explants depleted for Snail or Sox5 function are unable to respond to activin inducing signals to form mesoderm or endoderm. This suggests that the neural crest specifiers Snail and Sox5 are essential for the maintenance of stem cell attributes of blastula cells. Interestingly, other neural crest factors have been shown to play roles in both pluripotent cells and blastula cells. Myc, and its downstream effector Id3 as well as FoxD3 have been previously shown to be essential for

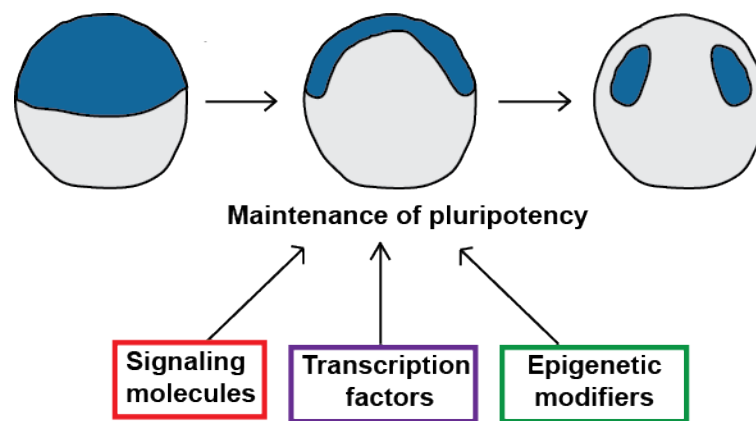


Figure 1.9 Control of pluripotency of neural crest cells

The process of control of pluripotency of the neural crest is the complex interplay of signaling molecules, transcription factors and epigenetic modifiers that are tasked to maintain the pluripotent state of a subset of blastula cells as they proceed through development and give rise to neural crest cells.

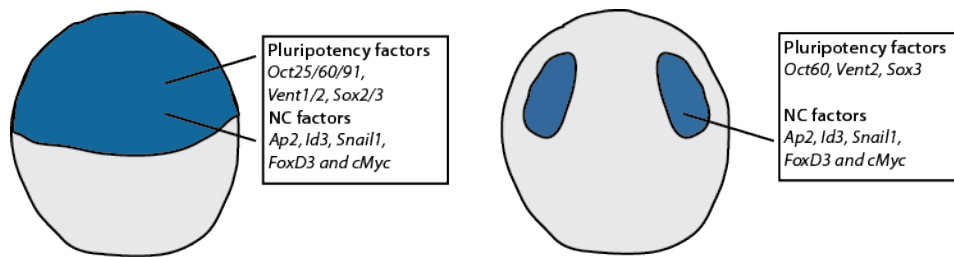


Figure 1.10 Shared transcriptional landscape of blastula and neural crest cells

Recent work from our lab has identified a shared transcriptional program between blastula and neural crest cells suggestive that these cells employ the same factors to regulate their pluripotency.

pluripotency of blastula cells and neural crest cells (Mundell & Labosky, 2011; N. Sasai et al., 2001). Interestingly, while several shared factors exist, a major difference between the two cell types seem to be with respect to Sox factor function. While SoxB1 factors (Sox1/2/3) function predominantly in pluripotent cells, that role is fulfilled by SoxE factors (Sox8/9/10) in neural crest cells (Buitrago-Delgado et al., 2018) (Figure 1.10).

Recent work from our lab has also shown that signaling pathways required for neural crest formation also play essential roles in blastula pluripotent cells. When BMP signaling is blocked using Chordin (BMP antagonist) or Smad7 (repressor Smad), we lose pluripotency gene expression in blastula cells (Nordin & LaBonne, 2014). Our lab also recently showed that FGF signaling is essential for the maintenance of pluripotency of both neural crest cells and blastula cells (Geary & LaBonne, 2018). Indeed, FGF signaling mediated through the MAPK cascade is a vital regulatory process for maintaining the pluripotency of neural crest and blastula cells, and as lineage restriction proceeds there is a switch in cascade activation to the PI3K cascade.

While we know a lot about the signaling molecules and transcription factors involved in neural crest stem cell maintenance, less is known about the epigenetic regulation of this process. The process of neural crest formation requires several levels of regulatory complexity and it is likely that tight epigenetic control would need to be maintained during neural crest specification as well as later during the migration and differentiation of neural crest cells. Indeed, studies have identified the essential requirement of epigenetic regulation in the formation and later differentiation of the neural crest. However, less is known about the role of these factors in the maintenance of pluripotency of neural crest cells. Epigenetic regulation through DNA methylation and histone methylation/acetylation might provide the layer of regulation required to maintain neural crest cells initially in a pluripotent state while the rest of the embryo undergoes

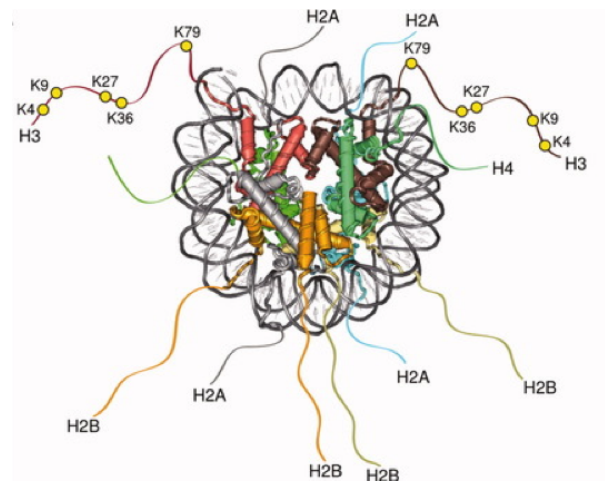


Figure 1.11 Schematic representation of histone post-translational modifications

N-terminal tail of histone undergoes extensive post-translational modifications on specific lysine and arginine residues. Shown above are modified residues that have known roles in development and stem cells. Each histone modification marks functionally distinct element in the genome (Adapted from Bogdanovic et al, 2012)

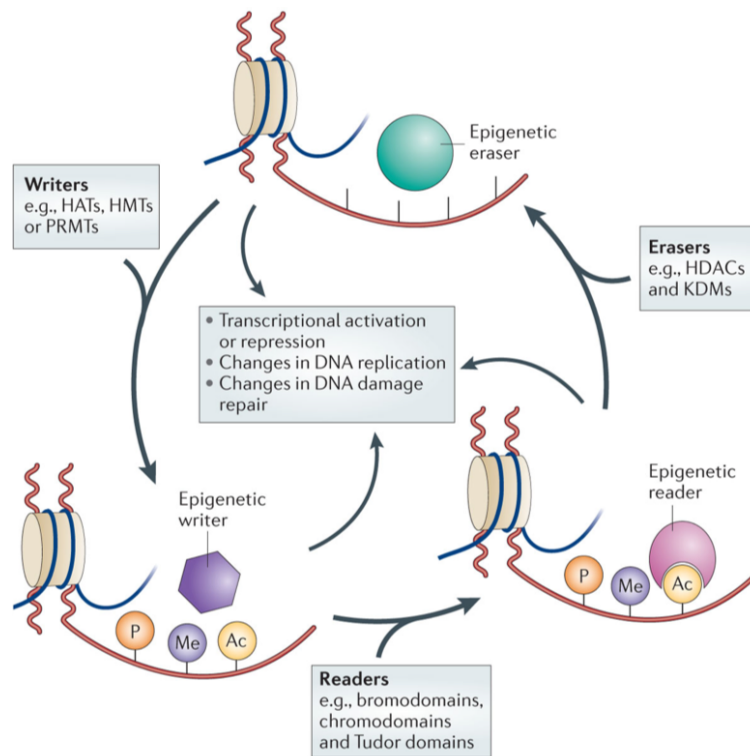


Figure 1.12 Chromatin remodelers: Epigenetic writers, readers and erasers

Chromatin remodeler typically belong to three different categories. Writers are enzymes that post-translationally modify residues on the histone tail and include methyltransferase and acetyltransferases while erasers are proteins responsible for the removal of histone modifications like demethylases and deacetylases. Bromodomains and chromodomain protein belong to a third category of remodelers called readers which specifically recognize acetylation and methylated histones. (Adapted from Falkenberg and Johnstone, 2014)

lineage restriction, and later provide these cells with the competency to respond to various differentiation cues that allow them to give rise to a large number of derivatives (Figure 1.11).

Epigenetic regulation: An overview

‘Epigenetics’, a term coined by Conrad Waddington in 1942, was used to describe changes in the phenotype without changes in the underlying genotype. Since then our understanding of epigenetics has been greatly enhanced and we now understand that epigenetic mechanisms exert precise control over gene activation and repression through modification of the DNA and histones. The core histones H2A, H2B, H3 and H4 are post-translationally modified by acetylation, methylation and phosphorylation at specific residues (Figure 1.11) (Bogdanovic, van Heeringen, & Veenstra, 2012b). This modified DNA and chromatin landscape are characteristic of a stable cell state and dictates the gene expression program of the cell. This regulation is particularly critical during development and differentiation as it is essential for defining the pluripotent stem cell state and controls the ability of a pluripotent cell to respond to inductive cues (Goldberg et al., 2007).

Epigenetic regulation is maintained by several factors tasked with controlling the chromatin landscape of a cell. Epigenetic modifiers are broadly categorized into 3 categories: writers, readers and erasers. Writers are enzymes that add post translational modifications to the histone, which include histone methyltransferases (HMTs) and histone acetyltransferases (HATs). Erasers are proteins that remove specific modifications from the histones such as histone demethylases (KDMs) and histone deacetylases (HDACs). The third category are readers and include proteins which contain Bromodomains or chromodomains which are capable of recognizing post-translational modifications of the histones (Figure 1.12) (Falkenberg &

Johnstone, 2014). Apart from histone modifications, genomic DNA can also be methylated or hydroxy-methylated by DNA methyltransferases and demethylases that offers a further layer of control. This epigenetic machinery functions together with various transcription factors in the cell to tightly control the type, number and location of posttranslational modifications deposited on chromatin.

The epigenetic stem cell state: Poised for activation

Embryonic stem cells (ESCs) are under tight epigenetic control to ensure the maintenance of a pluripotent state while keeping the cells poised for receiving differentiation cues (Goldberg et al., 2007; Spivakov & Fisher, 2007). Pluripotent cells are characterized by a unique expression pattern of histone modifiers and distinct distributions of modified histones. The promoter of pluripotent genes and genes associated with self-renewal in the ESCs have the presence of H3K4me₃, an active mark, a sign of active gene expression. This mark is deposited by the Trithorax complex such as Set/MLL methyltransferases (Watanabe, Yamada, & Yamanaka, 2013). Indeed, WDR5, a component of the Trithorax complex has been shown to selectively bind to H3K4me₂ and convert it to H3K4me₃ during vertebrate development (Wysocka et al., 2005). This activating mark is balanced by the presence of H3K27me₃, a repressive mark that is used to silence/repress developmentally regulated genes in these cells. This mark is preferentially added by the Polycomb repressor complex (PRC2). Interestingly, it has been shown that PRC2 binds to over 300 genes in the ESC genome, many of which are important for differentiation. This suggests that a hallmark of the stem cell epigenetic state is selective activation of pluripotency gene expression and the silencing of lineage restriction markers. LSD1 demethylase has also been shown to balance global levels of H3K4me₃ and H3K27me₃ at the regulatory regions of

several developmental genes in pluripotent cells (Adamo et al., 2011). Strikingly, studies have shown that embryonic stem cell state is characterized by bivalent domains, i.e. the simultaneous and dual presence of H3K4me3 and H3K27me3. Indeed, ESCs have a higher number of poised or bivalent promoters, when compared to differentiated cells (Azuara et al., 2006; B. E. Bernstein, Mikkelsen, Xie, Kamal, Huebert, Cuff, Fry, Meissner, Wernig, Plath, Jaenisch, Wagschal, Feil, Schreiber, & Lander, 2006a). These bivalent domains silence developmental regulators in pluripotent cells while keeping them poised for activation at the onset of differentiation.

Studies have also identified a unique enhancer signature in ESCs. In mouse ESCs, it was found that H3K4me1 and H3K27Ac marks denoted active enhancers, while poised enhancers lacked H3K27Ac and were preferentially associated with developmental regulators (Creyghton et al., 2010). In an independent study in human ESCs, it was shown that pluripotency genes in ESCs are indeed marked with active enhancers containing H3K4me1 and H3K27Ac while 'poised' enhancers are instead marked with H3K4me1 and H3K27me3 (Rada-Iglesias et al., 2011).

There is also a dramatic redistribution and expansion of repressive H3K9me3 and H3K27me3 marks in differentiated cells relative to pluripotent cells (Hawkins et al., 2010). The removal of H3K9me2/3 is essential for self-renewal of ESCs. Indeed, loss of activity of demethylases Jmjd1a and Jmjd2c (selectively function on H3K9 methylation) leads to a decrease in pluripotency gene expression and results in stem cell differentiation (Loh, Zhang, Chen, George, & Ng, 2007). This suggests that control of global levels of H3K9 methylation is critical for the maintenance of the stem cell state. Further, levels of DNA methylation have also been shown to be very important for the pluripotency and differentiation. DNA

methylation/demethylation has been shown to regulate the expression of master developmental regulators in ES cells. Indeed, Dnmt1-deficient ES cells undergo cell death upon induction of differentiation. Interestingly, DNMT3A and DNMT3B have been shown to methylate the promoters of pluripotency genes such as Oct4 and Nanog. Studies have also shown that DNA methylation at CpG rich sequences is very low in ES cells and is deposited de novo on pluripotency genes at the onset of lineage restriction. This would suggest that DNA methylation provides an additional layer of control that is responsible for shutting down the expression of the pluripotency transcriptional program during lineage specification (Farthing et al., 2008; Alexander Meissner et al., 2008; Mohn et al., 2008). Thus, there appears to a distinct epigenetic state that correlates with pluripotency, and dynamic changes in histone modifications occur with lineage restriction. Interestingly, this epigenetic control is true for ESCs grown within in a petri-dish, but also true for stem cells within the developing embryo.

Epigenetic regulation during embryonic development

The chromatin landscape of the cell dynamically changes during embryonic development. Epigenetic modifications are deposited in a cell-type specific manner and determine the developmental potential of the cell, executing lineage specific transcriptional programs. A hierarchical deposition of marks is seen during embryonic development with permissive marks added first in the pluripotent blastula embryos, and the deposition of repressive marks as development proceeds and more cells are lineage restricted. (van Heeringen et al., 2014) (Akkers et al., 2009; Bogdanovic, van Heeringen, & Veenstra, 2012b)(Figure 1.13).

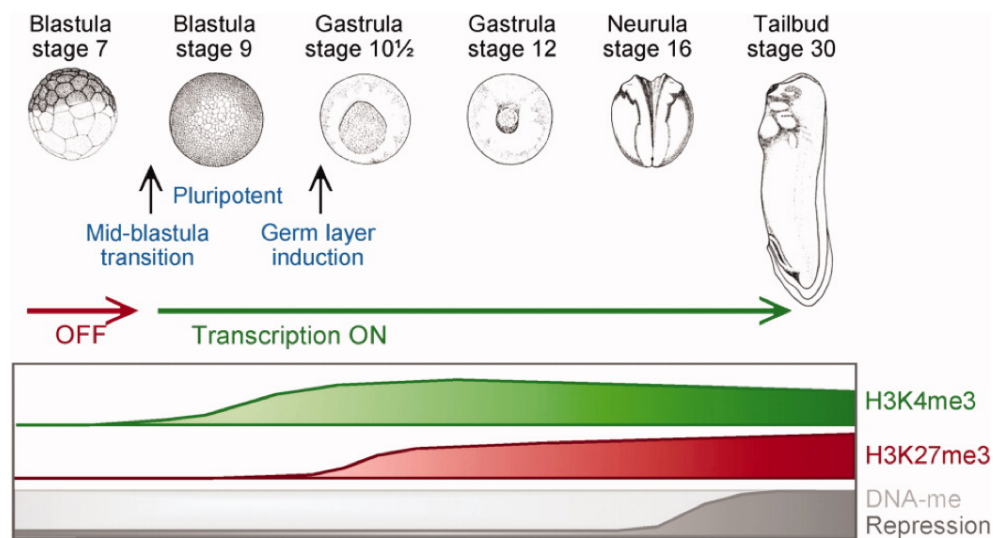


Figure 1.13 Dynamic changes in histone modification during embryonic development

Histone modifications deposited dynamically during embryonic development and lineage restriction. Studies have shown that the activating mark, H3K4me3 increases from blastula to gastrula stages coinciding with the onset of expression of lineage specific genes, while the repressive mark H3K27me3 is deposited during gastrulation as subset of genes are turned off. (Adapted from Bogdanovic et al, 2012)

The zygotic genome is transcriptionally silent after fertilization and is activated after mid blastula transition (MBT) (Bogdanovic, van Heeringen, & Veenstra, 2012b). The embryonic epigenetic landscape develops from an unprogrammed state and the activation of the zygotic genome (ZGA) begins with the appearance of permissive marks such H3K4me3. Indeed, the deposition of H3K4me3 seems to precede the transcriptional activation of developmental regulators (Akkers et al., 2009; Vastenhouw et al., 2010). Interestingly, another study showed that H3K4me3 acquisition prior to MBT is necessary for the onset of transcription, as seen at the earliest expressed genes such as *Nodal3.1(Xnr3)* and *Siamois* (Blythe, Cha, Tadjuidje, Heasman, & Klein, 2010).

In addition, active enhancers need to be established to drive transcriptional programs (Akkers et al., 2009). Such developmental enhancers are considered to be primed or poised when marked with H3K4me1 alongside H3K27me3, and active when marked with H3K4me1 with H3K27Ac (B. E. Bernstein, Mikkelsen, Xie, Kamal, Huebert, Cuff, Fry, Meissner, Wernig, Plath, Jaenisch, Wagschal, Feil, Schreiber, & Lander, 2006b; Creyghton et al., 2010; Gupta, Wills, Ucar, & Baker, 2014). Such enhancers allow for differentiation into diverse cell fates by launching cell-type specific transcriptional programs. Thus, the deposition of permissive chromatin marks is critical for opening up the embryonic epigenetic landscape at the onset of zygotic transcription.

The permissive nature of the chromatin is gradually restricted as development proceeds. During gastrulation, H3K27me3 (repressive mark) is deposited on spatially regulated genes to repress multilineage gene expression. ChIP-sequencing for H3K27me3 in animal and vegetal halves show that there are spatial differences in the differences are predictive of spatially regulation of gene expression. Interestingly, a quantitative mass spectrometry study performed in

Xenopus embryos shows an increase of repressive marks over developmental time, from blastula to tadpole stages (Schneider et al., 2011). Studies in zebrafish have also correlated H3K27Ac enrichment at enhancers that accompany a shift from a pluripotency to tissue specific gene expression (Bogdanovic, Fernandez-Miñán, Tena, la Calle-Mustienes, et al., 2012a). Interestingly, such distal H3K27Ac enhancers have been shown to be predictive of the developmental state of the cell (Creyghton et al., 2010).

A recent study in mammalian pre-implantation embryos identified that large parts of the genome are in an accessible state prior to zygotic genome activation, suggestive of complex mechanism controlling the transcriptional regulation at these stages (J. Wu et al., 2018). While mammalian embryos seem to undergo global demethylation post fertilization, this does not seem to be the case in zebrafish and *Xenopus* embryos. These data may represent species specific differences in the epigenetic regulation of early embryonic development. Thus, tight epigenetic control is critical for maintaining stem cells during embryonic development. It is likely that epigenetic regulation is thus an essential part of the circuitry that facilitates the maintenance of pluripotency of the neural crest. Interestingly, it has previously been shown that epigenetic regulation is necessary for formation and migration of neural crest cells.

Epigenetic regulation of the neural crest

The chromatin architecture of the cell has been shown to be essential for the maintenance, migration and differentiation of neural crest cells. Epi-genomic profiling of induced neural crest stem cells (hNCCs) from human ESCs identified and annotated several putative cell-type specific enhancers that are bivalent with H3K27Ac and H3K4me1 (Rada-Iglesias et al., 2012). For several of these enhancer regions, corresponding regions could be

identified in the chicken genome with increased H3K27Ac enrichment (Rada-Iglesias et al., 2012). Further, a recent study in chick embryos used ATAC-Seq and identified a class of distal regulatory elements that are neural crest specific enhancer regions that are necessary for the establishment of an early bona fide neural crest program. Interestingly, this study identified a class of elements that are accessible in naïve epiblast and pre-migratory neural crest cells and become inaccessible by migratory stages suggestive that the neural crest chromatin landscape is important for the competence of these cells (biorxiv, Sauka-Spengler lab, University of Oxford). Indeed, studies in mouse embryos have identified that the plasticity of premigratory neural crest progenitors is retained through maintaining the promoters and enhancers of positional genes involved in craniofacial morphology poised for activation (Minoux et al., 2017). This regulation allows for post migratory neural crest cells to respond to local cues in a positional dependent manner. Interestingly, this chromatin regulation seems to be a key driver of morphological divergence between species. Comparisons between hNCCs and chimpanzee NCCs identified cis regulatory elements that drive changes in gene expression of neural crest factors essential for the formation of the craniofacial skeleton and morphology and have contributed to the differences between humans and our closest evolutionary relative (Prescott et al., 2015).

Interestingly, several neural crest specifier genes have been identified to be necessary and important for regulating the chromatin landscape of neural crest cells. It has been suggested that TFAP2A is a master lineage specifier that selectively binds to NC- specific enhancer elements and establishes a transcriptionally permissive chromatin state that drives neural crest formation in hNCCs. Analysis for motif over-representation in these elements also identified high motif enrichment for other neural crest factors such as ETS, E-Box and SoxE factors as well as effectors of signaling pathways such as WNT and BMP suggesting that the regulation of neural

crest specifiers and signaling molecules converge on these enhancer elements (Rada-Iglesias et al., 2012). In zebrafish, FoxD3 has been shown to function as pioneer factor in the neural crest and modulate distal enhancers to prime genes for neural crest specification (Lukoseviciute et al., 2018).

Several epigenetic modifiers have also been shown to be essential for the formation of the neural crest. Studies in the chick embryo have identified that Dnmt3A is predominantly expressed in the neural crest forming regions and loss of Dnmt3A function results in the loss of neural crest specifier genes. This study also showed that Dnmt3A essential for repressing Sox2/Sox3 expression in the presumptive neural crest forming regions (N. Hu, Strobl-Mazzulla, Sauka-Spengler, & Bronner, 2012). Interestingly, mutations in Dnmt3B have been implicated in cranio-facial deformities consistent with a role for Dnmt3B in neural crest development. Indeed, studies in zebrafish suggest that Dnmt3B and G9a histone methyltransferase cooperate to promote formation of craniofacial derivatives, brain and retina (Rai et al 2010). Interestingly, in hESCs, knockdown of Dnmt3B results in increased expression of several neural crest specifiers. This would suggest that levels of DNA methylation are critical for normal development of the neural crest.

Histone modifiers have also been shown to be essential for neural crest formation. Histone demethylase, Jmjd2A (member of the Jumonji family) was the first epigenetic modifier shown to be required for neural crest formation by modulation of H3K9me3 at the promoter of neural crest specifier genes like *Sox10* and *Snail2* (Strobl-Mazzulla, Sauka-Spengler, & Bronner-Fraser, 2010). Interestingly, knockdown of Jmjd2A results in loss of expression of these genes, suggesting that Jmjd2A mediated histone demethylation is necessary for normal neural crest formation to occur. Consistent with this finding, studies in *Xenopus*, identified that

overexpression of Prdm12, a histone methyltransferase that methylates at H3K9 causes a loss in neural crest formation as seen by loss of expression of neural crest specifier, reduced head size and no. of melanocytes (Matsukawa, Miwata, Asashima, & Michiue, 2015). Similarly, the histone demethylase PHF8 has been shown to regulate MsxB during craniofacial development in zebrafish (Phillips et al., 2006). Interestingly, PHD12 has been shown to form a complex with Snail2 to epigenetically regulate epithelial to mesenchymal transition (Strobl-Mazzulla & Bronner, 2012). Furthermore, other demethylases like LSD1 have been shown to physically interact with the neural crest specifier, Snail2 and be essential for mediating EMT (T. Lin, Ponn, Hu, Law, & Lu, 2010). Consistent with this role of histone methylation during neural crest development, patients with mutations in histone demethylase PHF8 have craniofacial deformities (Qi et al., 2010). Strikingly, the lysine methyltransferase, NSD3 has been shown to be essential for neural crest formation and have an independent later role in H3K36 demethylation during migration (Jacques-Fricke & Gammill, 2014). This data would suggest that histone methylation plays multifarious roles during neural crest formation.

The polycomb repressor complex (PRC2) has also been shown to be important for neural crest formation. PRC2 methylates histone H3 at lysine 27 (K27) and is considered a repressor of gene expression and to be associated with a closed chromatin conformation. Core proteins of the PRC2 complex such as EED, EZH2 and Suz12 have all been implicated in to have functions in neural crest formation and migration. EZH2, the catalytic component of the PRC2 complex, has been shown to physically interact with Snail2 and modulate gene expression of neural crest specifiers, as well as co-occupy promoter regions of E-cadherin during neural crest migration in *Xenopus* (Tien et al., 2015). In mammalian cells, EZH2 has been shown to interact with neural plate border marker, Msx1 (J. Wang et al., 2011). EED, another core component of this complex,

was found to interact with YY1, a transcription factor shown to be necessary for neural crest formation (Satijn, Hamer, Blaauwen, & Otte, 2001). Loss of EZH2 results in bone and craniofacial deformities in the EZH2-KO mouse model through the depression of Hox genes (Schwarz et al., 2014). Another component of the PRC2 complex, Aebp2 was found to be expressed specifically in cell of neural crest origin and loss of Aebp2 had similar phenotypes as seen by loss of other PRC2 complex components (H. Kim, Kang, Ekram, Roh, & Kim, 2011). Additionally, Aebp2 occupies similar genomic loci of neural crest genes as EZH2 which are simultaneously enriched for H3K27me3 suggesting that Aebp2 might be a neural crest specific PRC2 component (H. Kim et al., 2011).

A sub-class of histone methyl transferases that target arginine residues on the histone tail, called Protein Arginine Methyl Transferases (PRMT) have also been shown to be necessary for cranio-facial bone formation and been implicated in epithelial to mesenchymal transition. A cranial neural crest specific deletion of Prmt1 resulted in defects in craniofacial structures and led to loss of Msx1 expression (Gou et al., 2018). This suggests that Prmt1 is upstream of Msx1 for regulating neural crest development. Further, Prmt5 has been shown to be recruited by the Snail adaptor protein AJUBA to mediate transcriptional repression during EMT (Hou et al., 2008). These studies suggest a potential role for PRMT family proteins in neural crest formation and migration and is an interesting area of future investigation.

Other epigenetic regulators have also been found to play important roles in neural crest formation. ATP dependent chromatin remodeling complexes such as SWI/SNF and CHD have also been shown to be required for neural crest formation. CHD7, an ATP dependent chromatin remodeler has been found to act in concert with PBAF (a component of the SWI/SNF complex) to activate the core transcriptional circuitry including *Sox9* and *Twist* during neural crest

induction from hESCs (Bajpai et al., 2010). Interestingly, mutations in CHD7 have been implicated in CHARGE syndrome and cause craniofacial deformations, peripheral nervous system and heart defects in humans, mice and *Xenopus* (Bajpai et al., 2010). Additionally, another member of the SWI/SNF family of chromatin remodelers, Brg1, has been shown to be required for Snail2 expression in neural crest cells. Further, several studies have identified that the folate receptor, RFC, is required for epigenetic regulation of neural crest. Morpholino mediated knock down of XRFC in *Xenopus* embryos, results in loss of *Zic1*, *Snail2* and *FoxD3* expression by deregulating H3K4me1/3. Interestingly, a recent study identified that knockdown of folate transporters Rfc1 and FolR1 results in reduced DNA methylation at the Sox2 locus, causing an expansion of neural plate formation at the expense of the neural crest. Further, another study identified a novel H2A.Z nucleosome binding protein, that seems to be essential for neural crest differentiation and migration (Pünzeler et al., 2017).

Histone acetylation has also been shown to be essential for neural crest formation and differentiation. Apart from specific histone acetylation marks playing important roles in neural crest formation, histone acetyltransferases have also been implicated to be involved in this process. It has been shown that inhibition of nitric oxide using a small molecule TRIM, attenuates HAT activity and resulted in a decrease in histone H4 acetylation and defects in cranial neural crest migration and chondrocyte lineage differentiation (Kong et al., 2014). Interestingly, it has recently been shown that Kat2a and Kat2b acetyltransferases (HATs) play important roles in craniofacial development. Zebrafish and mouse mutants for Kat2a/2b have several craniofacial defects including have shortened craniofacial cartilage elements and hypoplastic bone and cartilage. Reduced H3K9Ac in these mutants results in altered expression

of cartilage markers Sox9 and Col2a1 (Sen et al., 2018). HDACs have also been shown to be required for neural crest migration and differentiation (detailed in a later section).

Thus, epigenetic regulation is essential for formation, and later migration and differentiation of neural crest cells. However, much less is known about the epigenetic regulation of stem cell maintenance of neural crest cells. In particular, the role of histone acetylation/ HDACs during the process of maintenance of pluripotency during neural crest formation within the embryo is largely unknown.

Histone acetylation in the maintenance of pluripotency

Histone acetylation is a critical histone modification shown to have a vital but complex role in pluripotency and stem cell maintenance. Studies in human ESCs have shown that dynamic changes in levels of histone acetylation occur during differentiation. Interestingly, it has been shown that induction of differentiation in human and mouse ESCs, results in global reduction of glycolysis mediated acetyl-CoA production and hence histone acetylation, and inhibition of glycolysis leads to deacetylation and promotes differentiation of pluripotent cells (Moussaieff et al., 2015). In accordance with this, it has been shown that H3K9Ac gradually decreases during the first few days of *in vitro* neural differentiation, and then increases (P. Liu et al., 2015; Qiao, Wang, Yang, Tang, & Jing, 2015). H3K9ac and H3K14Ac seem to be present not only on the promoters of active genes, but also at bivalent poised promoters of developmentally regulated genes suggestive of histone acetylation playing a vital role in regulating lineage specification in ES cells (Karmodiya, Krebs, Oulad-Abdelghani, Kimura, & Tora, 2012). H3K27Ac has also been shown to mark active enhancers in ESCs (Creyghton et al., 2010). Strikingly, another histone acetylation mark, H3K56Ac has also been shown to be linked

to the core pluripotency network and interact with Oct4 to promote stem cell maintenance (Tan, Xue, Song, & Grunstein, 2013; W. Xie et al., 2009). The histone acetyltransferase, p300/CBP, has been shown to maintain mouse embryonic stem cell identity (Fang et al., 2014). They are recruited to Nanog loci to mediate the formation of enhancer loops and maintain ESCs in an undifferentiated state. Interestingly, Gcn5 has also been shown to be necessary for developmental potential of ESCs. Embryoid bodies from Gcn5^{-/-} mutants are have compromised differentiation potential (L. Wang et al., 2018). This suggests that levels of histone acetylation are critical for the maintenance of the pluripotent state.

Disruption of early histone deacetylation using HDAC inhibition results in a failure to differentiate, suggesting that the levels of histone acetylation are critical for early lineage choices. Consistent with this, HDACi in mouse and human ES cells has been shown to promote self-renewal (Ware et al., 2009). Interestingly, it has also shown that low levels of HDAC inhibition using VPA results in increased pluripotency and causes changes in global levels of H3K9Ac in mESCs. Strikingly, H3K9Ac increase does not occur at bivalent promoters in these cells suggesting that histone acetylation is tightly regulated at developmental genes (Hezroni, Sailaja, & Meshorer, 2011).

Readers of histone acetylation have also been shown to be critical for pluripotency of ES cells, further demonstrating the integral role of histone acetylation in the process. Bromodomain reader, BRD4, has been shown to interact with acetylated H4 to regulate pluripotency in ES cells (Gonzales-Cope, Sidoli, Bhanu, Won, & Garcia, 2016). Interestingly, this study reported global reduction in multiply acetylated histone H4 peptides. It has also been shown that Brd4 interacts with Oct4 to regulate the pluripotency gene network (T. Wu, Pinto, Kamikawa, & Donohoe, 2015). Additionally, inhibition of BRD4 function using a chemical inhibitor also resulted in a

loss of Nanog mediated pluripotency in mouse ESCs (Horne et al., 2015). This suggests that histone acetylation is a vital component of the pluripotency regulatory circuit, and the levels of histone acetylation are tightly regulated. The levels of histone acetylation within the cell is a balance maintained by histone acetyl transferases (HATs) (discussed previously) and histone deacetylases (discussed below).

Histone Deacetylases: Erasers of histone acetylation

Histone Deacetylases (HDACs) are enzymes that catalyze the removal of acetyl groups from modified histones and proteins. First identified in calf thymus extract in 1969, HDACs have since then been shown to be involved in a variety of cellular processes including gene regulation (Seto & Yoshida, 2014). There are 18 different HDACs found in mammals that are broadly classified into four different categories – Zn^{2+} dependent (Class I, ClassII(a,b) and Class IV) and NAD^{+} dependent (Class III or Sirtuins) (Figure 1.14) (Seto & Yoshida, 2014). These factors modulate levels of histone acetylation critical for transcriptional regulation and maintaining appropriate gene expression in the cell. HDACs are hence vital for several processes such as stem cell maintenance and differentiation and have been demonstrated to play important roles during embryogenesis as well as in adult tissues. It is hence not surprising that deregulation and aberrant expression of HDACs has been implicated in various diseases including several solid and hematological malignancies, neurodegenerative diseases, inflammation, immunological diseases, cardiac and pulmonary diseases (Y. Li & Seto, 2016; Seto & Yoshida, 2014).

Class I HDACs, in particular, are of high interest as they are ubiquitously expressed and predominantly found in the nucleus suggestive that histones are their primary substrate. Class I HDACs have short C terminal tail and high sequence similarity to yeast Rpd3, and consist of

HDAC1, HDAC2, HDAC3 and HDAC8 (Adams, Chandru, & Cowley, 2018; Seto & Yoshida, 2014). Interestingly, Class I HDACs have been shown to be vital for the gene regulation in different cell types. The highly related deacetylases HDAC1 and HDAC2 share 82% amino acid identity and form the catalytic core of numerous co-repressor complexes including Sin3A, NuRD, CoREST and MiDAC (Adams et al., 2018). Apart from these complexes, HDAC1/2 have been shown to be recruited by and interact with several transcription factors in different cellular contexts. Since HDACs lack the ability to directly bind to DNA, the specificity of HDAC activity is highly dependent on the cellular context and availability of binding partners.

The function of HDACs has been primarily interrogated with the use of HDAC inhibitors (Figure 1.15). These include several classes of compounds such hydroxamic acids such as Trichostatin A and SAHA, short chain fatty acids such as Valproic Acid and Butyrate, benzamides such as Entinostat, cyclic peptides such as Romidepsin and sirtuin inhibitors such as sirtinol and cambinol (Eckschlager, Plch, Stiborova, & Hrabeta, 2017) and have a broad range of specificity. For instance, hydroxamic acids are pan-HDAC inhibitors and inhibit all classes, while others are more specific e.g. Valproic Acid targets only Class I and Class II while Romidepsin targets HDAC1 and HDAC2 specifically (Göttlicher et al., 2001) (Furumai et al., 2002). Most inhibitors function by disrupting the Zn binding pocket and hence preventing activity of the enzyme. Trichostatin A, a naturally occurring compound, was one of the first HDAC inhibitors to be identified (Yoshida, Kijima, Akita, & Beppu, 1990). Interestingly, valproic acid, another HDAC inhibitor, was originally used for treatment of epilepsy, bipolar disorders and migraines and was later found to have deacetylase inhibition activity. HDAC inhibitors also make very attractive drug candidate for treatment for diseases such as cancer.

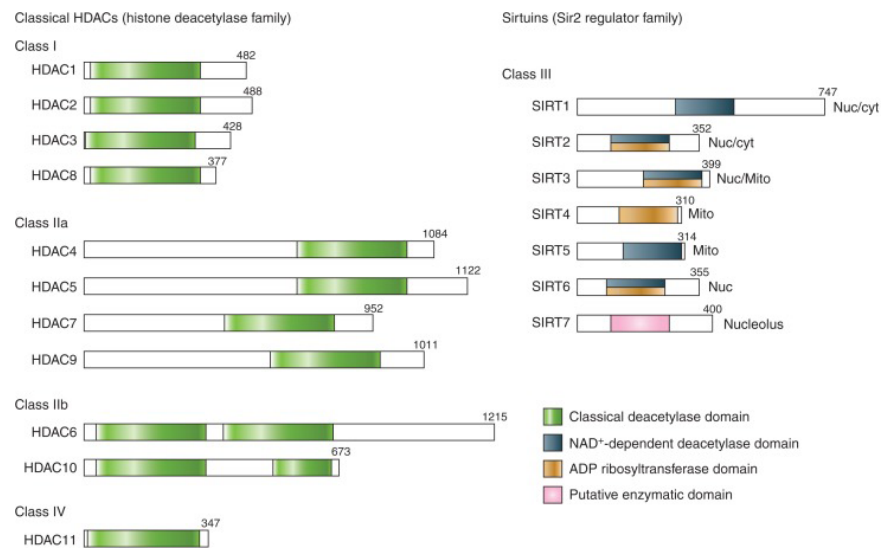


Figure 1.14 Classes of Histone Deacetylases

Domain organization of histone deacetylase families. HDACs are divided into 4 distinct categories based on whether they are Zn²⁺ dependent (Class I, Class IIa and b and Class IV) or NAD⁺ dependent (Class III). Class I HDACs are located primarily in the nucleus and primarily target histones, while other classes may target other proteins as well. (Adapted from Seto and Yoshida, 2014)

Classification	Inhibitors	Specificity
Aliphatic fatty acids	Butyrate Valproic Acid	Class I and IIa Class I and IIa
Hydroxamic Acids	SAHA (Vorinostat) Trichostatin A PXD101 (Belinostat)	Pan inhibitor Class I and II Pan inhibitor
Benzamides	MS-275(Entinostat) MGCD0103 CI-994	Class I Class I Class I
Cyclic peptides	FK228 (Romidepsin) Apicidin	HDAC1/2 Class I

Figure 1.15 Classification of HDAC inhibitors

HDAC function has been primarily investigated through the use of HDAC inhibitors. Different classes of HDAC inhibitors have been found and developed with varying degrees of specificity. Trichostatin A, is a naturally occurring compound that inhibits Class I and II HDACs. 11

Indeed, several HDAC inhibitors have also received FDA approval for treatment of cancers such as T-cell lymphoma (SAHA, Belinostat) and multiple myeloma (Panobinostat) etc. HDACs are typically considered to be canonical repressors of gene expression. Acetylated histones on chromatin is thought to promote an open, transcriptionally active conformation, while the removal of histone acetylation would result in a closed, transcriptionally repressed state. In accordance with this, HDACs have been thought to be involved in gene repression. Interestingly, recent studies have identified that HDACs may be playing more complex roles in gene regulation. Indeed, genome wide mapping of HDAC binding revealed that they are bound to both active and inactive genes (Z. Wang et al., 2009) suggestive of dynamic shuttling of HATs and HDACs at the promoters of active genes. Strikingly, one study has shown that HDAC1 and p300 directly associate with chromatin (DNA and histones) and compete for binding *in vitro* and *in vivo*. Interestingly, this study shows that HDACs associate with p300 through distinct domains and can be acetylated (X. Li, Yang, Huang, & Qiu, 2014). Through the genetic studies as well *in vitro* assays and cell culture, it has become clear that HDACs play an important role in a myriad of different cell types and tissue contexts. In particular, HDAC activity has been shown to be very important for embryonic development.

HDACs in embryonic development

Histone deacetylase activity is critical for normal embryonic development. In mouse, a genetic knock out of HDAC1 is embryonically lethal. Indeed, HDAC1-null mice die by E10.5 and have several proliferation defects and growth retardation (Lagger et al., 2002; Montgomery et al., 2007). This suggests that HDAC1 is important for early developmental decisions and processes. Interestingly, knock of HDAC2 does not result in a similar effect on embryogenesis,

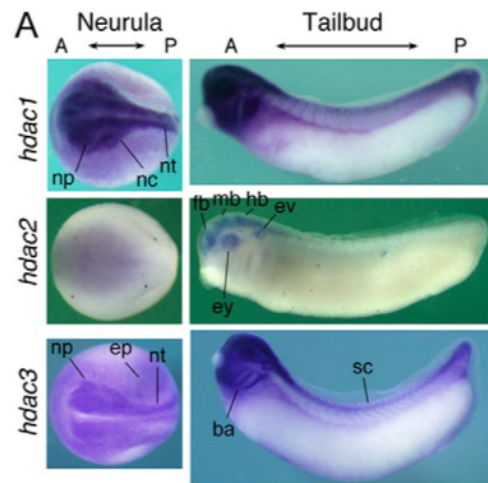


Figure 1.16 Expression of HDACs during *Xenopus* development

HDAC1, HDAC2 and HDAC3 are expressed during *Xenopus* development. HDAC1 seems to be most highly expressed, with a heightened expression in the neural plate and neural crest during neurula stages, and expression in the brain and cranial neural crest at tadpole stages. HDAC2 and HDAC3 are not as highly expressed and have distinctive expression patterns in these embryos. (Adapted from Zhang et al, 2017)

and pups are born but die shortly after birth in most cases. This would be suggestive that HDAC1 and HDAC2 might have functionally distinct roles during early development. A study in mouse oocytes and pre-implantation embryos has shown that HDAC1 and HDAC2 have very different expression profiles in this context (Ma & Schultz, 2016). While HDAC1 seems to be involved in cell cycle regulation and zygotic genomic activation, HDAC2 seems to regulate global DNA methylation in mouse oocytes. Interestingly, conditional null mutants of HDAC1/2 have revealed that they may have redundant roles in certain tissues, and simultaneously deletion has resulted in more severe phenotypes than individual deletion later in embryonic development. Conditional deletion of HDAC1/2 in the cardiac lineage together results in severe cardiac abnormality and neonatal lethality due to cardiac arrhythmias and dilated cardiomyopathy. Loss of HDAC1/2 activity also results in a upregulation of genes related to skeletal muscle specific contractile proteins and calcium channels in the heart (Montgomery et al., 2007). Other than the heart, HDACs also seem to be involved in other embryonic processes. Treatment of gastrula stage *Xenopus* embryos with VPA block HDAC activity, results in axial malformations and neural tube defects (Almouzni, Khochbin, Dimitrov, & Wolffe, 1994). Consistent with this, HDAC1 is expressed broadly during *Xenopus* embryonic development (Figure 1.16) (Z. Zhang et al., 2017). Interestingly, one study has shown that HDAC activity is necessary for establishing left-right patterning during early *Xenopus* development (Carneiro et al., 2011). It has also been shown that HDAC activity is necessary for tail regeneration in *Xenopus*, suggestive of a critical link to maintaining the stem cell state (Tseng, Carneiro, Lemire, & Levin, 2011). In particular, HDACs have been shown to play an important role during neural crest formation, migration and differentiation.

HDACs in neural crest formation

HDACs have been previously shown to be important for various aspects of neural crest formation and differentiation. Infants who were exposed to Valproic acid during pregnancy have higher risk of defects in neural crest derived tissues including cleft lip and palate, and cardiovascular defects suggestive of a critical role of HDAC activity in neural crest formation during early development (Alsdorf & Wyszynski, 2005; Wyszynski et al., 2005). In accordance with this, HDAC1 and HDAC2 conditional knockout in neural crest cells in mice resulted in embryonic lethality at E11.5, reduced proliferation in NC progenitor cells within the neural tube and first pharyngeal arch (Milstone, Lawson, & Trivedi, 2017). Interestingly, loss of HDAC activity using Trichostatin A has been shown to promote trunk neural crest specification in chick embryos. *In ovo* treatment of TSA results in the upregulation of neural crest factors such as BMP4, Pax3, Sox9 and Sox10, and premature loss in epithelial characteristics indicative that HDAC activity is very important for regulating the timing of gene expression in these cells (Murko et al., 2013). This would also suggest that HDACs perform different functions in these cells at different times of development.

Molecularly, HDAC1 has been shown to interact with the neural crest specifier, Ets1, to modulate the expression of Id3 downstream of BMP signaling in *Xenopus*. Interestingly, another study has shown that Snail2 recruits the HDAC-Sin3A complex to repress Cadherin (Cadh6B) during epithelial to mesenchymal transition of neural crest cells in chick embryos (Strobl-Mazzulla & Bronner, 2012). Interestingly, Snail2 has been shown to form a complex with HDACs with the adaptor LMO4 in *Xenopus* (Ochoa, Salvador, & LaBonne, 2012). This suggests HDAC functions in cohorts with neural crest specifier genes in order to promote neural crest specification during development.

Later in development, HDACs play important roles during neural crest cell differentiation into lineages. HDAC1 has been shown to be important for melanophore differentiation in zebrafish. HDAC1 mutant embryos exhibit severe reduction in the melanoblasts expressing MitFa resulting in their inability to differentiate normally into melanophores (Ignatius, Moose, El-Hodiri, & Henion, 2008). Interestingly, HDAC1 has also been shown to be essential for craniofacial morphogenesis as well as peripheral neuron development in temporally specific manner (Ignatius et al., 2013). Early specification defects in HDAC1 mutants results in abnormal posterior branchial arches, while later defects in differentiation are seen in the anterior mandibular and hyoid arches. HDAC1 mutants also have a disruption in the anterior to posterior patterning of neurons and ganglia (Ignatius et al., 2013). HDAC1 activity has also been shown to be necessary for oligodendrocyte formation in the zebrafish central nervous system. This defect is seen specifically in this cell lineage while other cell types within the brain do not seem to require HDAC1 activity (Cunliffe & Casaccia-Bonnel, 2006). HDAC1 and HDAC2 have also been shown to function together to regulate differentiation of neural crest cells into peripheral glia by regulating Pax3 expression (Jacob et al., 2014). This suggests that HDAC1/2 play tissue and cell type specific roles and seems to preferentially regulate neural crest derivatives.

Other HDACs have also been implicated to be important in the neural crest. Loss of HDAC4 activity also results in the craniofacial defects in zebrafish (DeLaurier et al., 2012). Interestingly, in humans, haploinsufficiency in HDAC4 is associated with brachydactyly mental retardation syndrome with craniofacial and skeletal abnormalities (Williams et al., 2010). HDAC3 has also been shown to be required for neural crest derived smooth muscle differentiation and cardiac outflow tract formation in mice (Singh et al., 2011). Strikingly,

HDAC8 has been shown to be functioning to regulate homeobox transcription factors Otx2 and Lhx1 during skull morphogenesis (Haberland, Mokalled, Montgomery, & Olson, 2009a). This suggest that HDACs play integral roles in neural crest development. However, how HDACs may be involved in the maintenance of stem cell potential of these cells is still unknown.

HDACs in stem cell maintenance

HDACs have been shown to play critical roles in stem cell maintenance and lineage restriction in ES cells. Using conditional knockout of HDAC1 and HDAC2 in ES cells, it was shown that HDAC1 and not HDAC2 is necessary for embryonic stem cell differentiation (Dovey, Foster, & Cowley, 2010). In accordance with this, HDAC1/2 knockout seems to result in a loss of pluripotency and defects in cell cycle in ES cells (Jamaladdin et al., 2014). Interestingly, another study has shown that HDAC inhibition promotes ESC self-renewal and push the cells into a state intermediate between ESCs and EpiSCs in mouse and human (Ware et al., 2009). This would suggest that HDACi is pushing cells into a more primed state of pluripotency, leading them to be trapped in a pluripotent state. Interestingly, HDAC1 knockout in mESCs resulted in loss in ability of cell to differentiate into neural lineage and promoted mesendoderm formation (P. Liu et al., 2015). Strikingly, this study found that HDAC1 is necessary for neural fate commitment *in vivo* as injected HDAC1 mutant cells showed reduced incorporation into head neural ectoderm and trunk neural tube in chimeric mice.

HDACs also seem to be critical for maintaining appropriate gene expression in ES cells. Interestingly, loss of HDAC1 in embryoid bodies resulted in increased expression of cardiomyocyte, muscle and neuronal specific markers suggestive of loss of pluripotency and precocious differentiation (Dovey et al., 2010). Interestingly, another study also demonstrated

that inhibition of histone deacetylase activity resulted in acceleration of stem cell differentiation and global changes in gene expression in ES cells (Karantzali et al., 2008). This would suggest that HDACs function to repress the expression of lineage restriction markers in ES cells.

Strikingly, it has been shown that silencing specific isoforms of HDACs can enhance the ability of bovine fibroblasts to be reprogrammed to a pluripotent state by upregulation of Nanog, key driver of reprogramming (Staszkievicz et al., 2013). Indeed, HDACi has shown to enhance reprogramming efficiency various contexts using TSA/VPA, suggesting a complex role of histone acetylation in maintaining the pluripotent state (Huangfu et al., 2008). It has also been recently suggested that HDAC1/2 activity might have broader roles than just regulating histone acetylation in ES cells. Indeed, this study showed that HDAC1/2 are critical for histone crotonylation in these cells (Kelly et al., 2018).

Surprisingly, HDACs have also been shown to have a positive regulatory role in ES cells. The mSin3A-HDAC complex has been shown to be recruited by Sox2 to positively regulate Nanog expression in ES cells (Baltus, Kowalski, Tutter, & Kadam, 2009). Interestingly, knockdown of components of Sin3A complex or loss of HDAC activity results in a loss of Nanog expression and pluripotency. Another study found that the Sin3A-HDAC complex cooperates with Nanog for maintaining ESC pluripotency and promoting somatic cell reprogramming (Saunders et al., 2017). Interestingly, Nanog and Sin3A-HDAC complex co-occupy the promoters of active genes and function to activate pluripotency genes and repress markers of lineage restriction. Strikingly, another HDAC containing complex, NuRD has been shown to play an important role in attenuating gene expression of a number of pluripotency genes allowing for exit from a pluripotent state (Reynolds et al., 2012). HDAC1 has also been found to be bound to the promoters of active genes (pluripotency factors) in ES cells and is

necessary for maintaining the pluripotency of these cells (Kidder & Palmer, 2012). These studies would suggest that HDACs may play multiple roles in stem cell maintenance i.e not only that they regulate/repress the expression of lineage markers but also function to positively regulate pluripotency gene expression in these cells.

While a large number of studies have looked at HDAC activity and histone acetylation in this context, as yet no clear model has emerged regarding the role of these factors in regulating pluripotency and lineage restriction. Additionally, these studies have all been performed *in vitro*, and the role of HDAC activity in stem cell maintenance during embryonic development *in vivo* is largely unknown. Understanding the regulation of HDAC activity during the neural crest stem cell maintenance will provide insights into how HDACs regulate stem cells within the developing embryo and provide us valuable information regarding the mechanisms of epigenetic regulation of stem cell maintenance.

***Xenopus* as a model organism**

Xenopus laevis is a powerful model organism for studying vertebrate embryology and development as well as basic cell and molecular biology, genomics, toxicology, neurobiology and to model human diseases (Figure 1.17). Commonly known as the African clawed frog, *Xenopus* eggs and embryos have several advantageous features that make them an outstanding tool for biomedical research. Frogs are relatively cheap to maintain and can be induced with the human gonadotropin hormone to produce large quantities of eggs. Further, the eggs can be easily fertilized *in vitro* to give rise to hundreds of synchronously dividing embryos that are large and develop externally. In addition to this, extensive fate maps of the cell fate of each

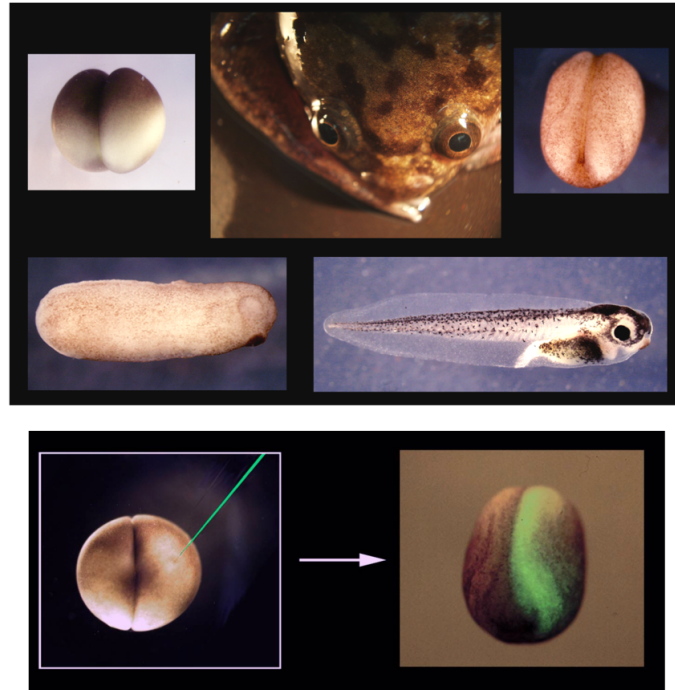


Figure 1.17 *Xenopus* as a model organism

Xenopus laevis is a powerful model organism to perform large scale biochemical and genomic studies as it is easy to obtain hundreds of synchronously developing embryos and development is external. Since the left right axis is established at the two-cell stage, one half of the embryo can be manipulated while the other half functions as an internal control. These embryos undergo rapid development, reaching tadpole stages within a few days post-fertilization.

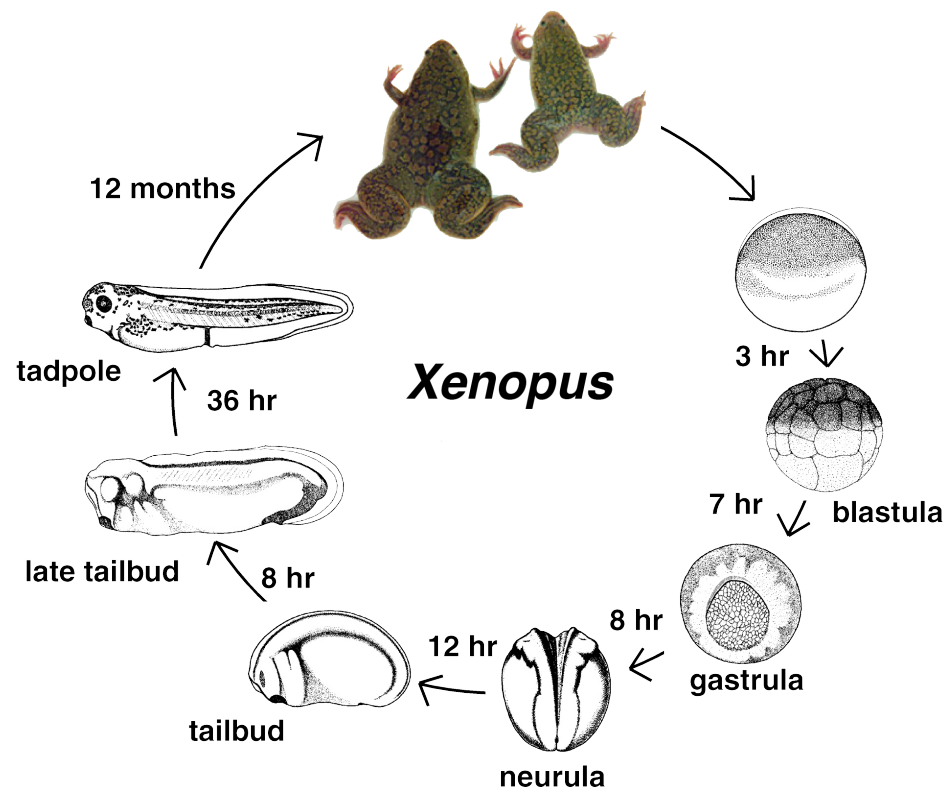


Figure 1.18 *Xenopus laevis* developmental life cycle

Schematic representation of the stages of *Xenopus* development. *Xenopus* embryos develop fairly rapidly making this model easy for perturbation experiments. (Adapted from Xenbase)

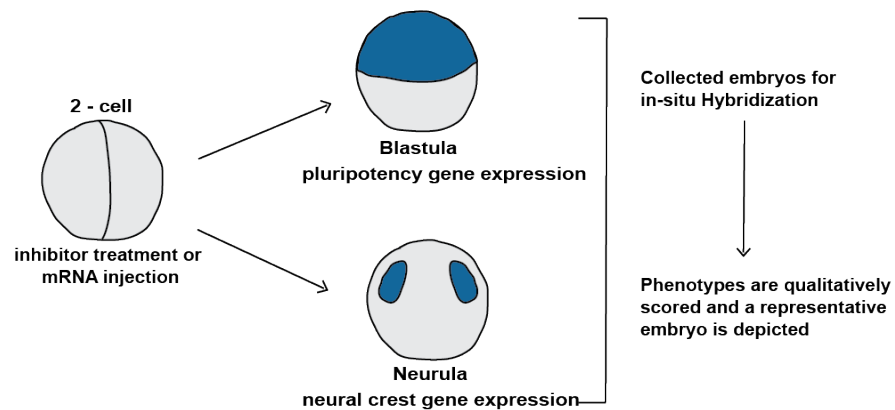


Figure 1.19 Embryological workflow

Xenopus embryos can be injected/treated with inhibitor at the 2-cell stage and cultured until blastula (Stage 9) or neurula (Stage 13-17) stages, at which point they are fixed and processed for *in situ* hybridization. Digoxigenin-labelled antisense RNA probes are used to detect the mRNA of interest. Phenotypes are scored qualitatively, and a representative embryo is chosen to depict results.

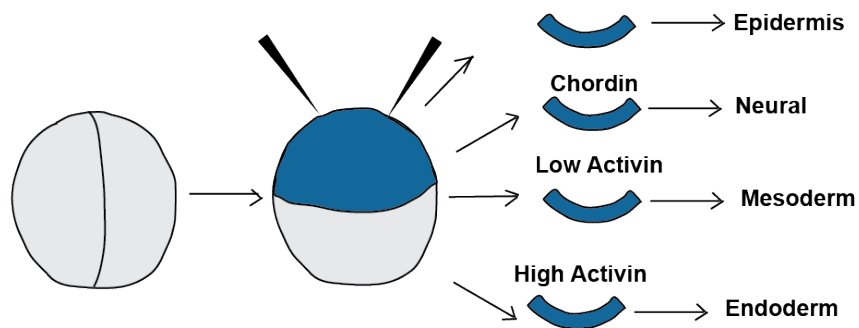


Figure 1.20 Schematic of animal cap assay

Pluripotent cells of blastula animal cap of *Xenopus* embryos can be dissected out and cultured with exogenous factors to induce different cellular fates. When cultured in isolation, these cells default to an epidermal state. By treating with varying levels of activin, or by perturbing BMP signaling with Chordin, the explants can be reprogrammed to form mesodermal, endodermal and neural cells.

cell in the early embryo allows for targeted gene knock-out, knockdown and overexpression studies (Figure 1.18).

In particular, *Xenopus* is a great system to study developmental and stem cell biology. Frogs produce thousands of embryos that can be easily manipulated for experimentation, in a highly reproducible manner (Figure 1.19). The pluripotent cells of the blastula animal pole can be easily dissected from the embryo and cultured in isolation for studying stem cell potential and lineage restriction. Further, since these explants are naïve, exogenous factors can be used to induce various cell fates (Figure 1.20). Another useful feature of this system is that the left right halves of the embryo are established with the first cleavage. It is hence possible to manipulate one half of the embryo while retaining the other half as a perfect internal control. Finally, due to the large amount of material available, this system can be used for large scale and high throughput experiments such as chemical inhibitor screens and genomic studies.

Historically, genomics and genetics in *Xenopus* has been challenging because these animals are pseudotetraploid. However, the successful sequencing and characterization of the *Xenopus* genome is a recent and major breakthrough for the community that has made it possible to perform large scale genomic studies in this system. This has also allowed for genome editing techniques such as CRISPR to be used in *Xenopus*, furthering the experimental tractability of this system. Thus, *Xenopus* is an ideal model organism and invaluable tool in which to perform studies for understanding molecular and biochemical processes contributing to stem cell regulation during early embryonic development.

Specified questions to be addressed in this thesis

Epigenetic state of the cell is critical for its stem cell potential as well as lineage restriction. Recent work from the lab has shown that neural crest cells have the ability to retain their stem attributes while the rest of the embryo is undergoing lineage restriction. However, the mechanisms utilized by these cells to stay in a suspended state of pluripotency is still unknown. While we have some idea regarding the signaling cues involved and the transcriptional regulation of this process, very little is known about the role of epigenetic regulation of neural crest stem cell maintenance. Thus, the over-arching goal of this thesis is to identify epigenetic factors that are necessary for neural crest stem cell maintenance and characterize the mechanisms utilized for maintaining the pluripotency of these cells and further our understanding of the neural crest stem cell state.

In Chapter 2 of this thesis, I report a novel role for histone deacetylase (HDAC) activity in neural crest stem cell maintenance. I identify that HDAC activity is necessary for the neural crest formation as well as for the pluripotency of blastula cells. I show that loss of HDAC activity results in premature expression of markers of lineage restriction pushing the cells out of a pluripotent state. Further, I provide evidence that histone acetylation is low in neural crest cells and blastula cells suggestive that HDACs are functioning in a similar manner to maintain histone acetylation low in these cell types. Finally, I demonstrate that HDAC activity has the ability to enhance cells to be reprogrammed to a neural crest state.

In Chapter 3, I further explore the mechanisms of epigenetic regulation of stem cell maintenance. Using a genome wide approach, I show that loss of HDAC activity results in global changes in gene expression with equal number of genes upregulated and downregulated

suggestive that HDACs play an important regulatory role in these cells. I elucidate that HDACs control histone acetylation at loci of pluripotency genes differently from lineage markers. Further, I explore the crosstalk between the mechanisms of HDAC activity and epigenetic reader Brd4 activity in neural crest formation and stem cell maintenance. Finally, I explore the dynamic changes in abundance of histone modifications during lineage restriction and characterize epigenetic modifications closely associated with stem cell maintenance.

In Chapter 4, I utilize large scale transcriptomics approach to characterize the gene expression profile of neural crest cells and identify novel factors that may have interesting roles in the maintenance of pluripotency of these cells. Using differential analysis, WGCNA and PCA/sparsePCA methodologies, I characterize an early and late neural crest signature and identify novel factors previously unknown to be involved in neural crest formation. Lastly, in Chapter 6, I present evidence detailing the role of transcription factor, Snail, in the maintenance of pluripotency during embryonic development and demonstrate that neural crest factors Snail1/2 physically interact with HDAC1/2 in *Xenopus*.

Chapter 2

**Histone Deacetylases are essential for
the formation and maintenance of the
vertebrate neural crest**

The Neural Crest, a progenitor population that drove vertebrate evolution, retains the broad developmental potential of the blastula cells it is derived from, even as neighboring cells undergo lineage-restriction. The mechanisms that enable these cells to preserve their developmental potential remain poorly understood. Here we explore the role that Histone Deacetylase (HDAC) activity plays in this process. We show that HDAC activity is essential for formation of neural crest, as well as for proper patterning of the early ectoderm. The requirement for HDAC activity initiates in naïve blastula cells; HDAC inhibition causes loss of pluripotency gene expression and blocks the ability of blastula stem cells to contribute to lineages of the three embryonic germ layers. We find that pluripotent naïve blastula cells and neural crest cells are both characterized by low levels of histone acetylation and show that increasing HDAC1 levels enhances the ability of blastula cells to be reprogrammed to a neural crest state. Together, these findings elucidate a previously uncharacterized role for HDAC activity in establishing the neural crest stem cell state.

Introduction

First described by Wilhelm His 150 years ago, the neural crest is a vertebrate progenitor population distinguished by its ability to contribute cell types associated with multiple germ layers to the vertebrate body plan (Hall, 2000; Sauka-Spengler & Bronner-Fraser, 2008). Acquisition of these cells, referred to as the “Zwischenstrang” by His, played a central role in the evolution of vertebrates, by layering a myriad of novel structures onto the simple chordate body plan (His, 1868; Le Douarin & Dupin, 2012). Understanding the mechanisms that control the genesis of these important cells is essential to understanding how vertebrates evolved and is linked to understanding the establishment and maintenance of their developmental potential.

We recently demonstrated that much of the transcriptional circuitry that controls the potency of neural crest cells is shared with pluripotent blastula cells (the *Xenopus* equivalent of mammalian inner cell mass cells), and proposed that neural crest cells arise through retention of the characteristics of those earlier cells (Buitrago-Delgado et al., 2015). This revised model for neural crest formation raises fundamental questions regarding how these cells escape lineage restriction, and retain their potency, until after the basic chordate body plan has been laid down. These mechanisms are certain to involve signaling pathways and transcription factors previously linked to neural crest formation (Prasad et al., 2012; Taylor & LaBonne, 2007). For example, the transcription factors Snail and Sox5, which are both essential to the formation of the neural crest, are first required for the pluripotency of blastula cells, as are BMP and FGF signaling (Buitrago-Delgado et al., 2015; Geary & LaBonne, 2018; LaBonne & Bronner-Fraser, 2000; Nordin & LaBonne, 2014). Less understood is the role that chromatin remodelers, and regulation of the epigenetic state, may play in the retention of pluripotency leading to establishment of the neural crest (N. Hu, Strobl-Mazzulla, & Bronner, 2014).

The epigenetic landscape of a pluripotent cell determines its cellular competency, and impacts its lineage choices (Atlasi & Stunnenberg, 2017; M. Li, Liu, & Belmonte, 2012). The maternally defined chromatin state has been shown to be critical for controlling gene expression during embryonic development (Hontelez et al., 2015); (Liang & Zhang, 2012). In cultured embryonic stem (ES) cells, histone acetylation plays an essential role in ensuring appropriate gene expression; acetylation of histone H3 at specific lysine residues (H3K9/14Ac and H3K27Ac) marks important developmentally regulated genes in murine and human ES cells as well as in early embryonic development (Bogdanovic, Fernandez-Miñán, Tena, la Calle-Mustienes, et al., 2012a; Creyghton et al., 2010; Gupta et al., 2014; Hezroni et al., 2011;

Karmodiya et al., 2012; Rada-Iglesias et al., 2011). Histone deacetylases (HDACs), enzymes tasked with removing acetyl marks from histones, have also been shown to play a critical role in maintaining pluripotency in cultured ES cells (Dovey et al., 2010; Jamaladdin et al., 2014). HDAC1 occupies the promoter regions of pluripotency genes in these cells, suggesting a positive regulatory role in maintaining the stem cell state and HDAC containing complexes such as Sin3A-HDAC and NuRD have been linked to promotion of pluripotency in ES cells and during somatic cell reprogramming (Baltus et al., 2009; Saunders et al., 2017).

While studies in cultured ES cells suggest that histone acetylation plays a critical if complex role in controlling pluripotency and lineage restriction, surprisingly little is known about the roles and regulation of HDAC activity/histone acetylation in the control of pluripotency *in vivo*, during embryonic development. In contrast to ES cell cultures, pluripotency is transient in early embryos, and lineage restriction events are dynamic and spatially controlled. We hypothesized that understanding how regulation of histone acetylation impacts the pluripotency of naïve blastula cells might shed light on the epigenetic mechanisms that contribute to a subset of cells escaping lineage restriction to form the neural crest. *Xenopus* embryos provide an ideal model for such studies, as explanted blastula cells retain the transient pluripotency characteristic of cells *in vivo* and require no exogenous factors for survival or retention of potential. Remarkably, although decades worth of experiments utilizing these cells have shaped our understanding of the signals that direct pluripotent cells to adopt specific lineage states, much less is known about how pluripotency is controlled and maintained in naïve blastula cells.

In this study, we examine the role of HDACs and histone acetylation in the establishment of the neural crest and the maintenance of pluripotency in *Xenopus* embryos. We show that

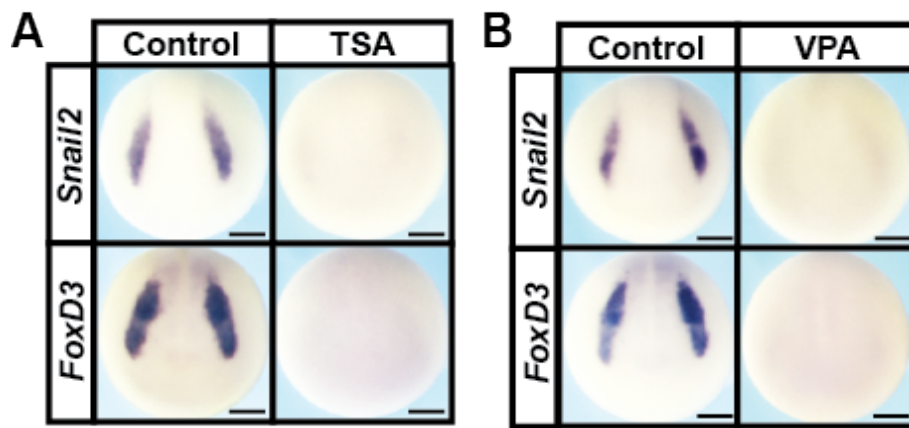


Figure 2.1 HDAC activity is necessary for neural crest formation

In situ hybridization examining *Snail2* and *FoxD3* expression in neurula stage (stage 15) embryos treated with vehicle control (DMSO or water) or Trichostatin A(200nM) (A) or Valproic Acid(10mM) (B). Loss of HDAC activity leads to loss of expression of neural crest markers. Embryos were treated at mid-gastrula stages (Stage 11) and grown until mid-neurula stages (Stage 15). Scale bars: 250μM.

HDAC activity is critical for formation of the neural crest stem cell population, and for the pluripotency of blastula stem cells. Inhibiting HDAC activity causes naïve blastula cells to lose pluripotency, and this is accompanied by precocious and aberrant expression of genes that direct multiple lineage states. We find that pluripotent blastula cells and neural crest cells are both characterized by low levels of histone acetylation, further emphasizing the similarities between these cell types. Finally, we show that increasing HDAC1 activity in blastula cells enhances their reprogramming to a neural crest state. Together these findings provide novel insights into the epigenetic mechanisms that control the maintenance of pluripotency in the developing embryo, and show that the regulation of HDAC activity is essential to establishing the neural crest state. These results shed important new mechanistic light on the genesis of a cell type central to the evolution of vertebrates.

Results

HDAC activity is essential for neural crest and placode formation

HDAC activity can be temporally controlled in developing embryos by employing potent small molecule inhibitors such as Trichostatin A (TSA) and Valproic Acid (VPA) (Göttlicher et al., 2001; Yoshida et al., 1990). TSA inhibits Class I and IIB families of HDACs while VPA specifically targets the Class I family. Treatment of *Xenopus* embryos with these inhibitors at early gastrula stages (Nieuwkoop and Faber stage 10.5-11) results in the complete absence of neural crest cells, as evidenced by loss of expression of *Snail2* (TSA-100%, n=72; VPA-100%, n=74) and *FoxD3* (TSA-100%, n=65; VPA -100%, n=70) in comparison to vehicle treated control embryos (*Snail2* (DMSO-0%, n=61; water-0%, n=65); *FoxD3* (DMSO-0%, n=66; water-0%, n=68))(Figure 2.1 A,B). This loss of neural crest cells is accompanied by developmental

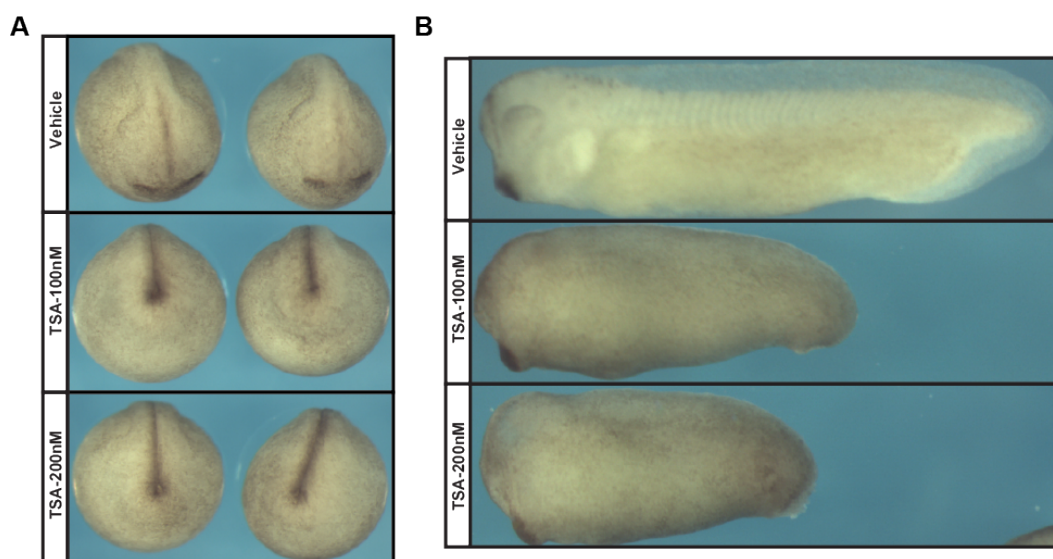


Figure 2.2 TSA treatment prevents normal embryological development.

Brightfield images of embryos treated with vehicle-control (DMSO) or TSA (100nM, 500nM) at Stage 10.5. Embryos were imaged at control Stage 19 and control Stage 30. TSA treatment results in abnormal embryological development and stunted growth. Embryos were treated at mid-gastrula stages (Stage 11) and grown until late-neurula stages (Stage 19)(A) and tadpole stages (Stage 30).

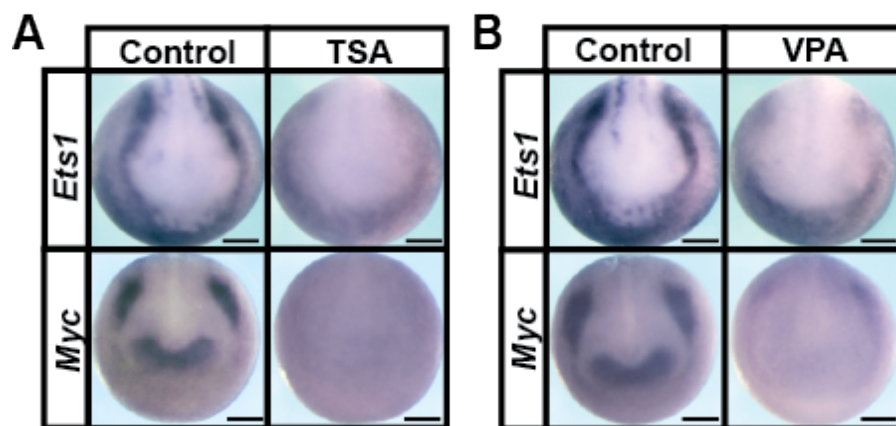


Figure 2.3 HDAC activity is required for expression of neural crest markers.

In situ hybridization examining *Ets1* and *Myc* expression in neurula stage (stage 15) embryos treated with vehicle control (DMSO or water) or Trichostatin A (200nM) (A) or Valproic Acid (10mM) (B). Loss of HDAC activity leads to loss of expression of neural crest markers. Embryos were treated at mid-gastrula stages (Stage 11) and grown until mid-neurula stages (Stage 15). Scale bars: 250 μ M.

malformations in cranial regions and axial truncations (Figure 2.2 A,B) (Almouzni et al., 1994; Gurvich et al., 2005).

Based on these findings we asked if HDAC activity is essential for establishing the neural plate border (NPB) region, where neural crest cells and cranial placodes, another vertebrate novelty, reside (Groves & LaBonne, 2014). The NPB is characterized by expression of a number of key transcription factors including *Ets1*, *Myc*, *TFAP2*, *Msx1/2*, and *Zic1* (Groves & LaBonne, 2014; Milet & Monsoro-Burq, 2012; Prasad et al., 2012). Treatment with TSA or VPA resulted in embryos with significantly reduced expression of several NPB factors including *Myc* (TSA: 100%, n= 78; VPA: 100%, n=71), *Ets1* (TSA: 100%, n=61; VPA: 100%, n=64) when compared to vehicle treated control embryos (*Myc* (DMSO: 0%, n= 64; water: 0%, n=76), *Ets1* (DMSO: 0%, n=61; water: 0%, n=57)) (Figure 2.3 A,B), as well as *Ap2* and *Msx2* (Figure 2.4 A,B). Notable among these is *Myc*, one of the four original factors that Yamanaka and colleagues showed could “reprogram” somatic cells to form iPSCs, and prior to that had been shown to control stem cell potential in the neural crest (Bellmeyer et al., 2003; Takahashi et al., 2007). These results demonstrate that HDAC activity is essential for establishing the neural plate border region, and the vertebrate specific progenitor cell populations that reside there. Consistent with this, the cranial placode region fail to form in TSA/VPA treated embryos, as evidenced by loss of expression of *Six1* (TSA: 100%, n =77 (control: 0%, n=62); VPA: 100%, n=74 (control: 0%, n=76)) (Figure 2.5 A,B).

We noted that the expression of *Ets1* and *Myc* outside of the NPB, in the prospective epidermis, was less affected by HDAC inhibition (Figure 2.3 A,B). We therefore examined the effects of TSA/VPA treatment on the establishment of other ectoderm-derived cell types. At neurula stages, we observed an expansion of the neural plate, as marked by *Sox2* expression, into

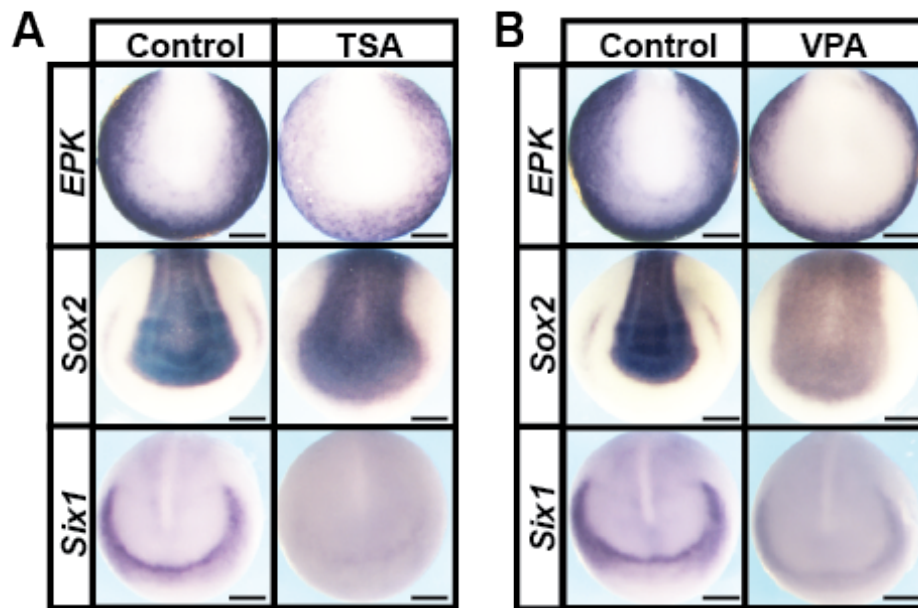


Figure 2.4 HDAC activity is required for proper ectodermal patterning.

In situ hybridization examining *EPK* (epidermal), *Sox2* (neural plate) and *Six1* (placode) expression in neurula stage (stage 15) embryos treated with vehicle control (DMSO or water) or Trichostatin A (200nM) (A) or Valproic Acid (20mM) (B). Loss of HDAC activity leads to loss of expression of epidermal marker, *EPK* and expansion of neural plate marker, *Sox2*. HDAC activity is necessary for expression of placodal marker, *Six1*. Embryos were treated at mid-gastrula stages (Stage 11) and grown until mid-neurula stages (Stage 15). Scale bars: 250μM.

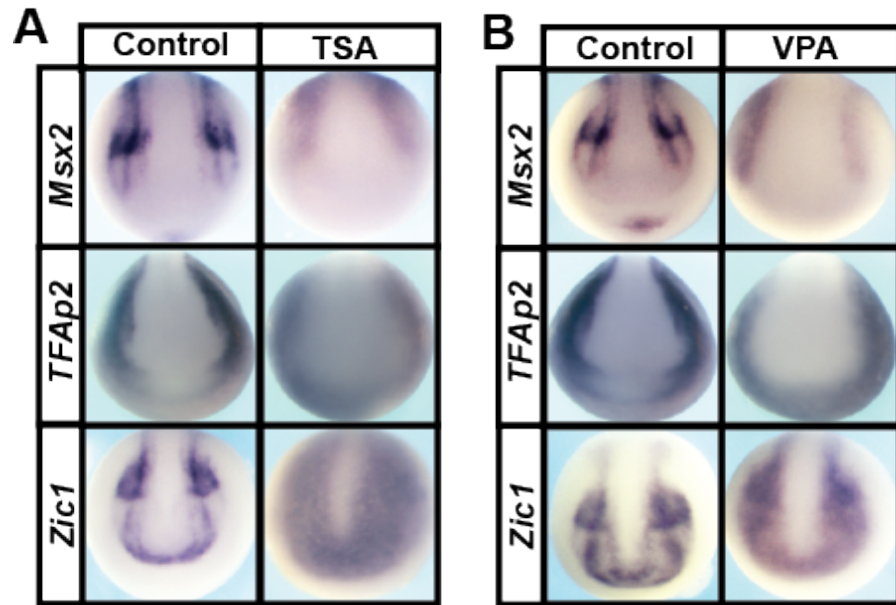


Figure 2.5 Loss of HDAC activity disrupts neural plate border formation.

In situ hybridization examining neural plate border marker expression: *Msx2*, *TFAP2* and *Zic1* expression in neurula stage (stage 15) embryos treated with vehicle control (DMSO or water) or Trichostatin A(200nM) (A) or Valproic Acid(20mM) (B). Loss of HDAC activity leads to loss of expression of neural plate border markers *Msx2* and *TFAP2*. *Zic1*, which marks both the neural plate and neural crest is expanded after TSA/VPA treatment consistent with the expansion of the neural plate. Embryos were treated at mid-gastrula stages (Stage 11) and grown until mid-neurula stages(Stage 15).

the regions that would normally form neural crest, placodes, and epidermis (TSA: 100%, n = 62(control: 0%, n=65); VPA: 100%, n= 58(control: 0%, n=52)) (Figure 2.4 A,B). We also observed that expression of *Zic1*, a gene that marks both neural and NPB cells, is expanded after HDAC inhibition (TSA: 100%, n=70 (control: 0%, n= 68); VPA: 100%, n=66 (control: 0%, n= 63)), consistent with an expansion of the neural plate (Figure 2.5 A,B). *Tfap2*, which is expressed in the epidermis and the NPB lost only its NPB expression upon HDAC inhibition (TSA: 100%, n = 66 (control: 0%, n= 66); VPA: 100%, n= 64 (control: 0%, n= 61)) (Figure 2.5 A,B). TSA and VPA treatment resulted in lowered expression of *EPK*, including 'salt and pepper' like expression at the neural plate border, but the overall domain of expression for this epidermal marker was largely unchanged (TSA: 100%, n = 68 (control: 0%, n= 60); VPA: 100%, n = 56 (control: 0%, n= 54)) (Figure 2.4 A,B). Together, these results indicate that HDAC activity plays an essential role in patterning of the early ectoderm, including establishing sharp boundaries for the prospective epidermis and CNS. The most prominent effect of blocking HDAC activity, however, is a failure to establish the neural crest and placode populations at the neural plate border.

HDAC activity is required for reprogramming to a neural crest state

To further investigate the striking loss of neural crest stem cells observed in embryos treated with HDAC inhibitors, we asked if TSA/VPA treatment could also prevent explants of pluripotent blastula cells from being reprogrammed to a neural crest state. This reprogramming can be achieved following introduction of Pax3/Zic1 or Wnt/Chd, and leads to robust expression of neural crest regulatory factors including *Snail2* and *FoxD3* (Hong & Saint-Jeannet, 2007; LaBonne & Bronner-Fraser, 1998; Plouhinec et al., 2014). We found that treatment with either

TSA or VPA prevented reprogramming, as evidenced by a failure to induce expression of markers of the neural crest cell state in response to either Pax3/Zic1 (TSA: Snail2-100%, n=26 (control: 18%, n=28), FoxD3-100%, n=23 (control: 18%, n=28); VPA: Snail2-100%, n=28 (control: 18%, n=28), FoxD3-100%, n=25 (control: 13%, n=31)) or Wnt/Chd (TSA: Snail2-100%, n=26, FoxD3- 100%, n=25; VPA: Snail2-100%, n=27, FoxD3- 100%, n=28) expression (Figure 2.6 A,B and Figure 2.7 A,B). These results indicate that the requirement for HDAC activity to establish the neural crest stem cell population is direct, and further suggests that it may be linked to the retention of pluripotency in blastula stem cells.

HDAC1 is necessary for neural crest formation

As TSA and VPA are broad spectrum HDAC inhibitors, we wished to identify which HDACs were specifically required for neural crest formation. Romidepsin (RMD) is an inhibitor that specifically targets HDAC1 and HDAC2 (Furumai et al., 2002). RMD treatment completely inhibited expression of neural crest markers in Wnt/Chd reprogrammed animal cap explants (Snail2-100%, n=27 (control: 13%, n=30), FoxD3-100%, n=30 (control: 15%, n=27) indicating that HDAC1 and/or HDAC2 activity was required for neural crest formation ((Figure 2.8).

HDAC1 is much more prominently expressed than *HDAC2* in both early embryos and naïve blastula cells. (Xenbase (<http://www.xenbase.org>), Figure 2.9). We therefore further investigated the requirement for HDAC1 by designing a translation blocking morpholino that targets both alleles. We found that morpholino-mediated depletion of HDAC1 phenocopied the effects of pharmacological HDAC inhibition on neural crest formation, as evidenced by loss of *Snail2* expression (loss -75%, n=52) (Figure 2.10). Interestingly, we found that these effects could be rescued by either HDAC1 (rescue - 72%, n=54) or HDAC2 (rescue - 60%, n=53) (Fig.11).

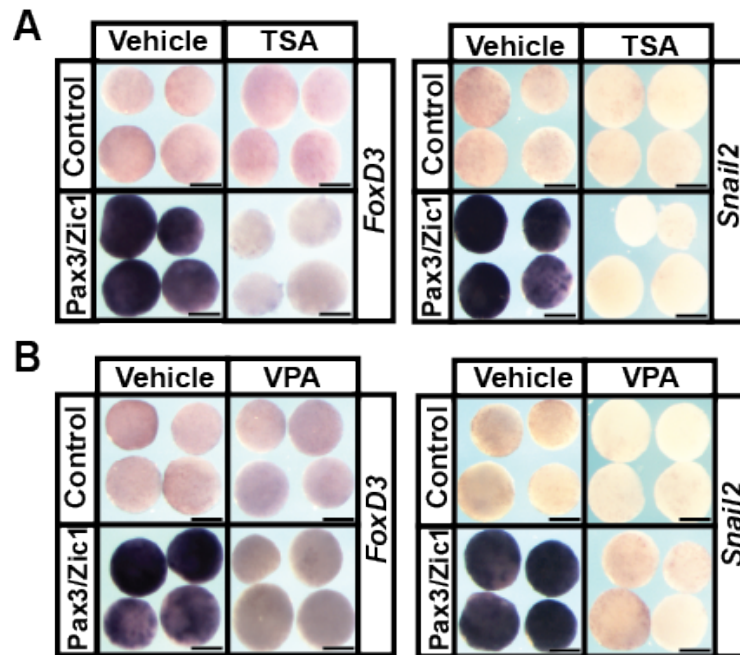


Figure 2.6 HDAC activity is necessary for blastula explants to be reprogrammed to neural crest state using Pax3/Zic1.

In situ hybridization examining *Foxd3* and *Snail2* expression in animal cap explants reprogrammed to neural crest state using Pax3/Zic1 and treated with vehicle control (DMSO or water) or Trichostatin A(500nM) (A) or Valproic Acid(10mM) (B). Loss of HDAC activity leads to loss of expression of *FoxD3* and *Snail2* representative of a failure to form the neural crest. Explants were cultured alongside sibling embryos grown until late-neurula stages (Stage 18). Scale bars: 250 μ M.

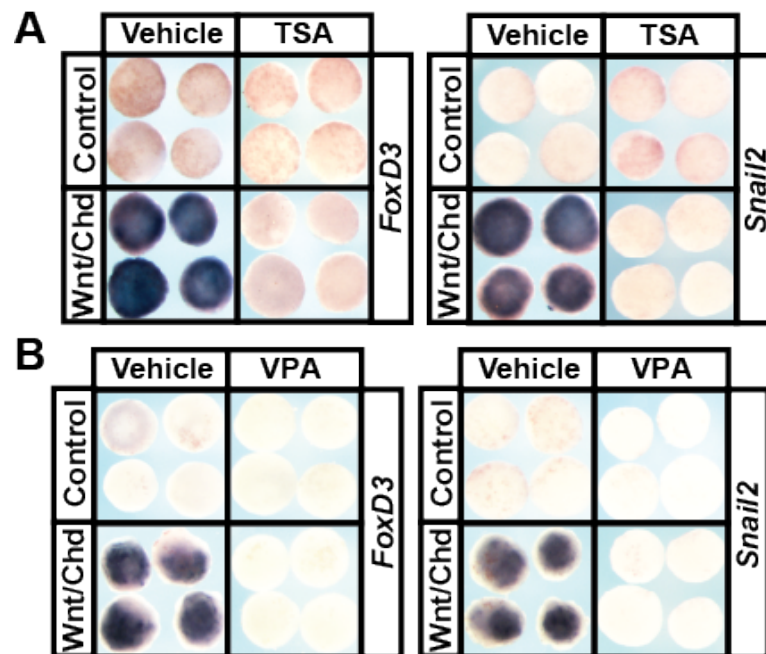


Figure 2.7 HDAC activity is necessary for blastula explants to be reprogrammed to neural crest state using Wnt/Chd.

In situ hybridization examining *Foxd3* and *Snail2* expression in animal cap explants reprogrammed to neural crest state using Wnt/Chd and treated with vehicle control (DMSO or water) or Trichostatin A(500nM) (A) or Valproic Acid(10mM) (B). Loss of HDAC activity leads to loss of expression of *FoxD3* and *Snail2* representative of a failure to form the neural crest. Explants were cultured alongside sibling embryos grown until late-neurula stages (Stage 18).

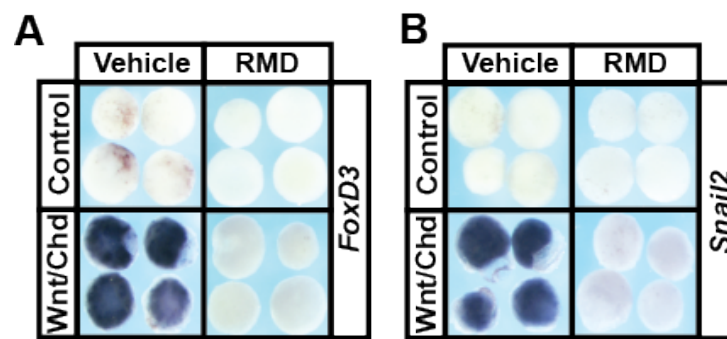


Figure 2.8 HDAC1/2 activity is necessary for blastula explants to be reprogrammed to neural crest state.

In situ hybridization examining *Foxd3* and *Snail2* expression in animal cap explants reprogrammed to neural crest state using Wnt/Chd and treated with vehicle control (DMSO) or Romidepsin(2.5 μ M) (A,B) Loss of HDAC1/2 activity leads to loss of expression of *FoxD3* and *Snail2* representative of failure to form the neural crest and recapitulates effects seen after TSA/VPA treatment. Explants were cultured alongside sibling embryos grown until late-neurula stages (Stage 18).

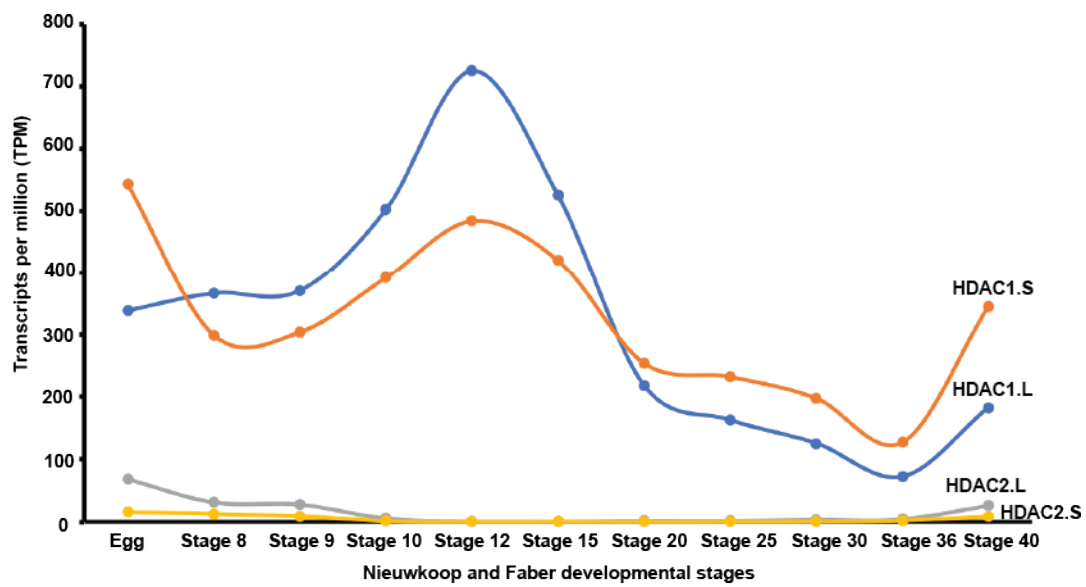


Figure 2.9 *HDAC1* and *HDAC2* expression during *Xenopus* development

Graph depicting gene expression in transcripts per million (TPM) for HDAC1 and HDAC2 at various stages of *Xenopus* development. HDAC1 is more strongly expressed during *Xenopus* development than HDAC2. Data adapted from Xenbase.

This suggests that these two HDACs do not have distinct activities in this context. Further, we found that introducing both HDAC1 and HDAC2 (rescue - 62%, n=45) did not enhance the rescue (Figure 2.10). Consistent with a role in regulating neural crest genesis, *Hdac1* is broadly expressed at early neurula stages, and becomes heightened in the neural crest by late neurula (Stage 18/19) stages (Figure 2.11) ((Z. Zhang et al., 2017). *Hdac1* is maternally provided, expressed from early cleavage stages, and highly enriched in pluripotent cells at blastula stages (Figure 2.11), an expression profile consistent with a role in maintaining pluripotency and progenitor states (Carneiro et al., 2011).

HDAC activity is essential for proper gene expression in pluripotent blastula cells

Since HDAC activity is essential for pluripotent blastula cells to transit to a neural crest state, we wondered if this might reflect an essential role in maintaining the pluripotency of those blastula stem cells. Accordingly, we examined the effects of TSA/VPA treatment on the expression of pluripotency genes in blastula embryos. Strikingly, inhibition of HDAC activity led to dramatically decreased expression of genes linked to pluripotency including *Vent2* (TSA: 100%, n=71 (control: 0%, n=70); VPA: 100%, n=68 (control: 0%, n=63)), *Oct25* (TSA: 98%, n=64 (control: 0%, n=63); VPA: 92%, n=71 (control: 0%, n=62)), *Sox3* (TSA: 100%, n=74 (control: 0%, n=79); VPA: 100%, n=70 (control: 0%, n=69)) as well as *TFAP2* (TSA: 100%, n=67 (control: 0%, n=63); VPA: 100%, n=71 (control: 0%, n=68)) and *Id3* (TSA: 100%, n=54 (control: 0%, n=68); VPA: 100%, n=61 (control: 0%, n=73)) in stage 9 embryos (Figure 2.12 & Figure 2.13). Similar decreases were observed when gene expression changes were examined quantitatively in explants of pluripotent cells using qPCR (Figure 2.14). We observed reduction in the expression of pluripotency genes previously examined by *in situ* hybridization, as well as

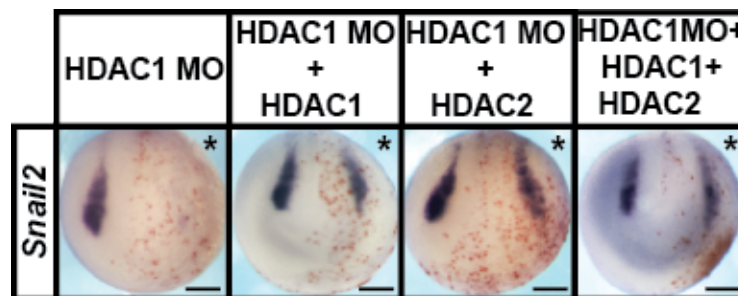


Figure 2.10 HDAC1 activity is essential for neural crest formation.

In situ hybridization examining *Snail2* expression in whole embryos after morpholino mediated knockdown of *Hdac1* and rescued with co-injection of HDAC1, HDAC2, or HDAC1+HDAC2 mRNA. Loss of HDAC activity leads to loss of expression of *Snail2* representative of failure to form the neural crest, and recapitulates the effect seen after TSA/VPA treatment. This phenotype can be rescued by overexpression of HDAC1, HDAC2 or HDAC1 and 2 in combination. Embryos were injected at 8-cell stage and collected at mid-neurula stages (Stage 15). Scale bars: 250 μ M.

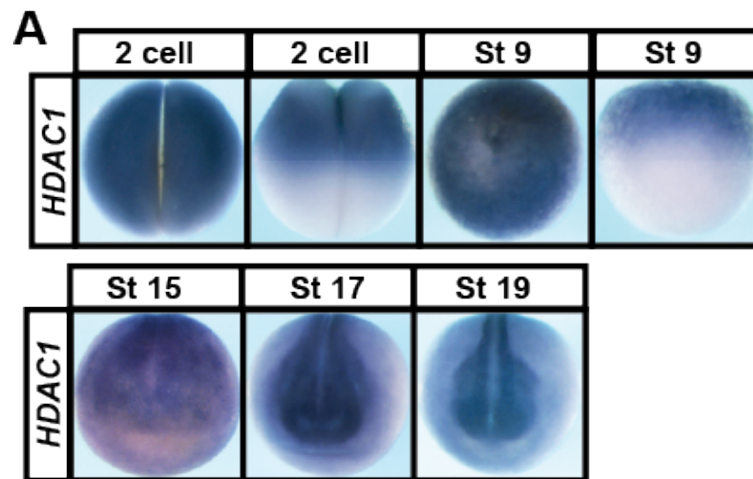


Figure 2.11 *HDAC1* expression during development.

In situ hybridization examining *HDAC1* expression in whole embryos over various stages of development. *HDAC1* is maternally deposited and expressed in the pluripotent blastula cells. By late neurula stages (St17-19), there is heightened expression of *HDAC1* seen in the neural crest forming regions. Wild type *Xenopus* embryos were collected at first cleavage (2 cell), blastula (Stage 9) and neurula stages (Stage 15, 17 and 19).

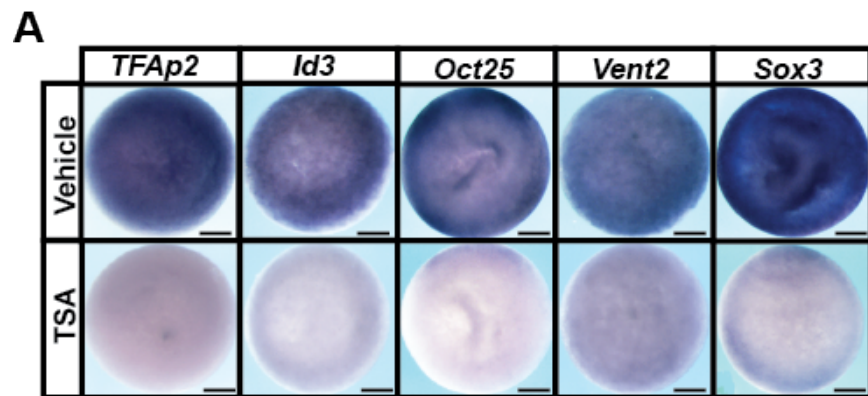


Figure 2.12 HDAC activity is essential for pluripotency gene expression in blastula cells.

In situ hybridization examining *TFAp2*, *Id3*, *Oct25*, *Vent2* and *Sox3* expression in whole embryos treated with vehicle (DMSO) or Trichostatin A (500nM). Loss of HDAC activity leads to dramatic reduction of expression of pluripotency genes in blastula cells. Embryos were treated at first cleavage (2 cell stage) and grown until blastula stage (Stage 9). Scale bars: 250μM.

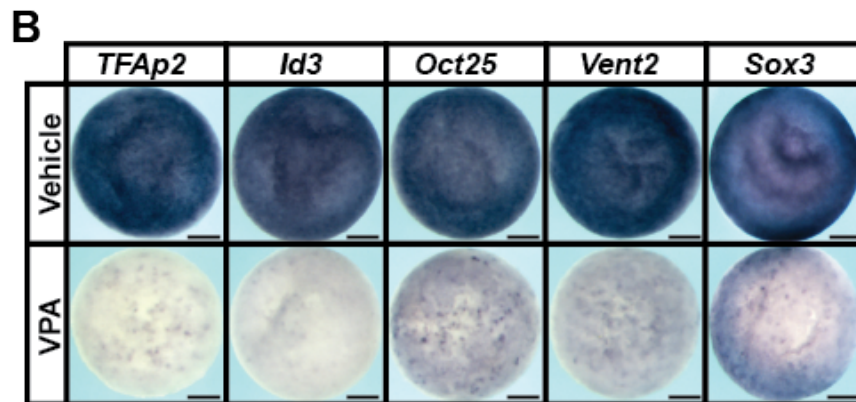


Figure 2.13 Blocking HDAC activity with VPA lead to loss of pluripotency gene expression in blastula cells.

In situ hybridization examining *TFAp2*, *Id3*, *Oct25*, *Vent2* and *Sox3* expression in whole embryos treated with vehicle (water) or VPA (20mM). Loss of HDAC activity leads to dramatic reduction of expression of pluripotency genes in blastula cells. Embryos were treated at first cleavage (2 cell stage) and grown until blastula stage (Stage 9). Scale bars: 250 μ M.

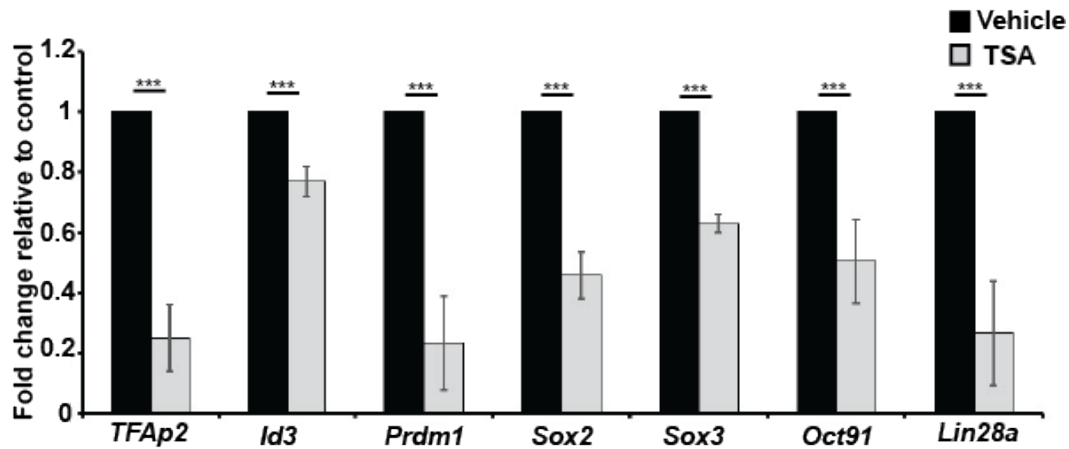


Figure 2.14 HDAC activity is required pluripotency gene expression in pluripotent animal cap explants.

Quantitative RT-PCR examining *TFAP2*, *Id3*, *Prdm1*, *Sox2*, *Sox3*, *Oct91* and *Lin28a* expression in animal cap explants treated with vehicle (DMSO) or Trichostatin A (500nM). Disruption of HDAC activity leads to significant loss of expression of pluripotency genes in blastula cells. P value : ***<0.005

in *Oct91*, *Lin28a* and *Prdm1*, factors that also have been shown to be involved in the maintenance of pluripotency.

The pluripotency of explanted naïve blastula cells is transient; these cells will adopt an epidermal state in the absence of other inductive signals, expressing the epidermal marker *EPK*, while down-regulating expression of pluripotency markers, such as *Sox3*. Interestingly, we found that following TSA treatment, explants failed to transit to an epidermal state and express *EPK* (100%, n=28), unlike control explants (control: 0%, n=29) (Figure 2.15A). Conversely, we observed sustained expression of *Sox3* in TSA treated explants (100%, n=29 (control: 0%, n=25)) (Figure 2.15B). *Sox3* expression could be indicative of either a pluripotent state or a neural progenitor state, as *Sox3* marks both these cell populations. To distinguish between these possibilities, we examined expression of another pluripotency marker, *Oct60*, and found that unlike *Sox3* its expression was not maintained in TSA treated explants (Figure 2.16A). Thus, HDAC inhibition is neither blocking exit from pluripotency, nor causing cells to prematurely transit to an epidermal state. TUNEL assays also demonstrated that HDAC inhibition does not lead to a significant increase in cell death in these explants (Figure 2.16B).

HDAC activity is essential for the pluripotency of blastula cells

Given the striking changes in gene expression noted in pluripotent blastula cells following HDAC inhibition, we next asked if these cells had lost their pluripotency by challenging them to adopt specific lineage fates. Isolated *Xenopus* animal poles cells (“animal caps”) can be induced to form any cell type in the embryo given appropriate developmental cues (Ariizumi & Asashima, 2001). For example, these cells will transit to a neural state if treated with BMP inhibitors such as noggin or chordin, and can be induced to form mesoderm or

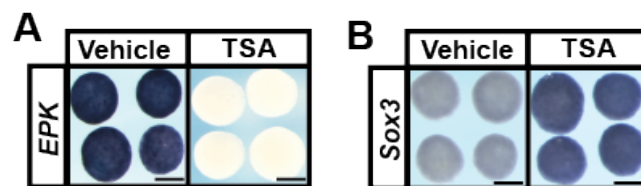


Figure 2.15 TSA treated explants do not default to an epidermal state

In situ hybridization examining *EPK* and *Sox3* expression in animal cap explants treated with vehicle control (DMSO) or Trichostatin A (500nM) (A,B). Loss of HDAC1/2 activity leads to loss of expression of *EPK*, suggesting that TSA treated explants were not defaulting to an epidermal state. Interestingly, TSA treatment results in a sustained expression of *Sox3*. Explants were cultured alongside sibling embryos grown until late-gastrula stages (Stage 13). Scale bars: 250 μ M.

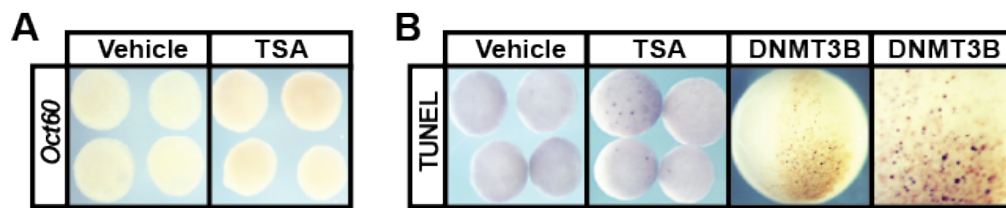


Figure 2.16 TSA treated explants do not default to an epidermal state

(A) *In situ* hybridization examining *Oct60* explants treated with vehicle control (DMSO) or Trichostatin A (500nM). Loss of HDAC1/2 activity does not lead to an increase in pluripotency, as evidenced by no change in *Oct60* expression. Interestingly, TSA treatment results in a sustained expression of *Sox3*. (B) TUNEL assay examining cell death in the vehicle and TSA treated animal cap explants alongside positive control DNMT3B injected embryos. No increase in cell death is observed in the TSA treated explants. Explants were cultured alongside sibling embryos until neural plate stages (stage 13).

endoderm with low or high doses of TGF-beta factors such as activin (Asashima, Nakano, Shimada, et al., 1990a; Asashima, Nakano, Uchiyama, et al., 1990b; Lamb et al., 1993; Y. Sasai, Lu, Steinbeisser, & De Robertis, 1995) (Figure 2.17). To determine if pluripotent blastula cells can form neural tissue when HDAC activity is inhibited, animal pole regions explanted from embryos previously injected with mRNA encoding Chordin were treated with vehicle or inhibitors TSA/VPA. Explants treated with HDAC inhibitors were unable to adopt a neural fate, as evidenced by loss of expression of *Nrp1* (TSA: 100%, n=30 (control: 4%, n=27); VPA: 100%, n=26 (control: 4%, n=27)) and *Sox11* (TSA: 100%, n=23 (control: 0%, n=24); VPA: 96%, n=23 (control: 3%, n=30)) (Figure 2.18 A,B). When explants were challenged to form mesoderm in response to activin treatment, control explants displayed robust expression of mesodermal genes such as *Xbra* (DMSO: 100%, n=29; water: 100%, n=30) and *MyoD* (DMSO: 100%, n=26; water: 100%, n=29) (Figure 2.19 A,B). By contrast, explants treated with HDAC inhibitors failed to form mesoderm as evidenced by failure to express these markers (TSA: *Xbra*-100%, n=28, *MyoD* -100%, n=28; VPA: *Xbra*- 100%, n=26, *MyoD* - 100%, n= 29) (Figure 2.19 A,B). Similarly, inhibitor treated explants fail to form endoderm in response to high doses of activin, as seen by failure to express genes such as *Endodermin* (TSA: 96%, n=28 (control: 0%, n=25); VPA:96%, n=27 (control: 0%, n=27)) and *Sox17* (TSA: 96%, n=25 (control: 0%, n=26); VPA:96%, n=28 (control: 3%, n=30)) (Figure 2.20 A,B). Taken together, the inability of naïve blastula cells to adopt neural, mesodermal or endodermal states following HDAC inhibition, combined with the observed loss of pluripotency gene expression in these cells, strongly suggests that HDAC activity is essential for the maintenance of pluripotency in blastula animal pole cells.

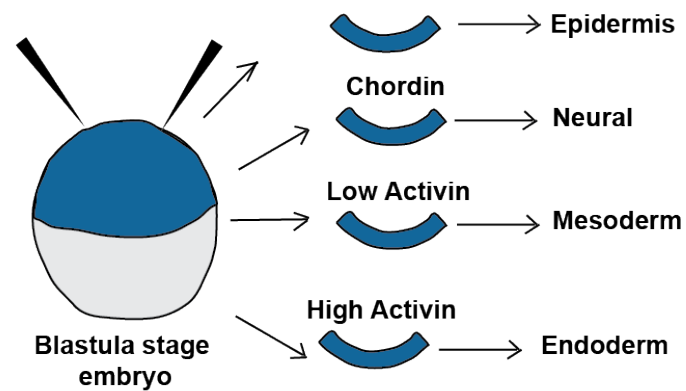


Figure 2.17 Animal cap explants can be induced to give rise to various fates
Schematic representing the animal cap explant assay. Explants can be induced with various signals to give rise to different germ layers : ectoderm(epidermis and neural), mesoderm and endoderm.

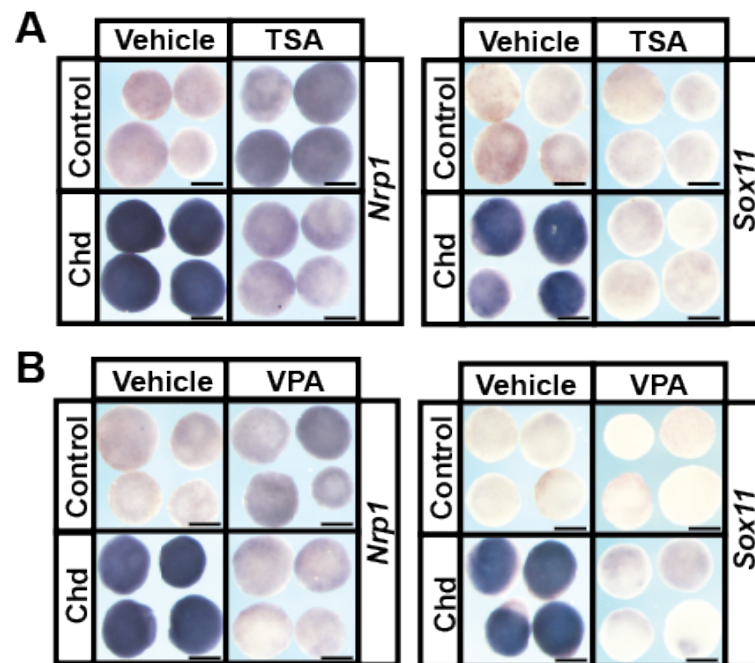


Figure 2.18 Loss of HDAC activity results in an inability to form neural fate

In situ hybridization examining *Nrp1* and *Sox11* expression in Chordin (Chd) induced animal cap explants treated with vehicle control (DMSO/water) or Trichostatin A (500nM) or Valproic Acid (10mM) (A,B). Loss of HDAC1/2 activity leads to inability to form neural fate as evidenced by loss of *Nrp1* and *Sox11* expression. Explants were cultured alongside sibling embryos grown until late-neurula stages (Stage 18). Scale bars: 250μM.

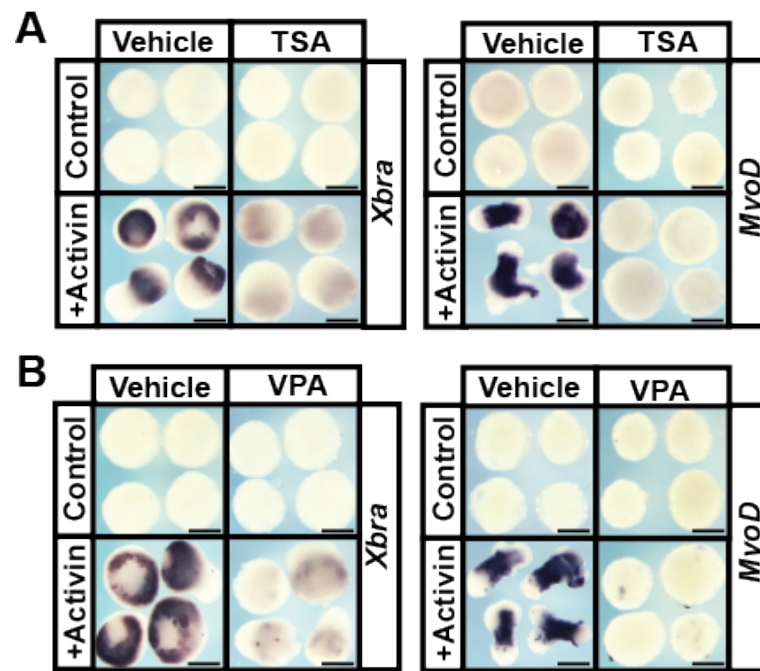


Figure 2.19 Loss of HDAC activity results in an inability to respond to activin and form mesoderm

In situ hybridization examining *Xbra* and *MyoD* expression in activin treated animal cap explants (low-medium dose) treated with vehicle control (DMSO/water) or Trichostatin A (500nM) or Valproic Acid (10mM) (A,B). Loss of HDAC1/2 activity causes explants to lose competency to respond to activin signals to form mesoderm as evidenced by loss of *Xbra* and *MyoD* expression. Explants were cultured alongside sibling embryos grown until mid-gastrula stages (Stage 11.5) for *Xbra* and late-neurula stages (Stage 18) for *MyoD*. Scale bars: 250 μ M.

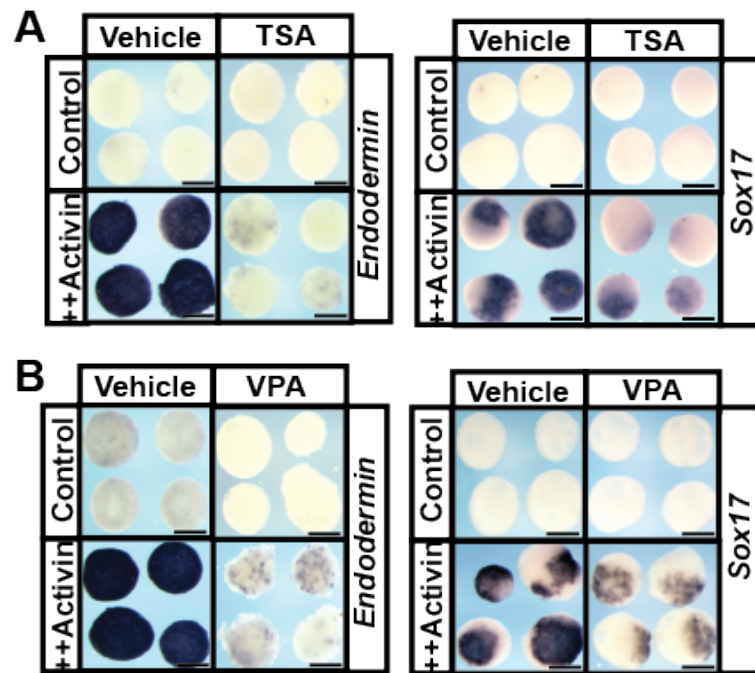


Figure 2.20 Loss of HDAC activity results in an inability form endoderm lineage

In situ hybridization examining *Endodermin* and *Sox17* expression in activin treated animal cap explants (low-medium dose) treated with vehicle control (DMSO/water) or Trichostatin A (500nM) or Valproic Acid (10mM) (A,B). TSA/VPA treatment causes explants to lose competency to respond to activin signals to form endoderm as evidenced by loss of *Endodermin* and *Sox17* expression. Explants were cultured alongside sibling embryos grown until mid-gastrula stages (Stage 11.5). Scale bars: 250 μ M.

Inhibition of HDAC activity results in precocious expression of multi-lineage markers

Given that blastula explants were no longer pluripotent following HDAC inhibition, yet were also not prematurely lineage restricted to an epidermal state, we wished to determine what the state of these cells was. We were intrigued by the expression of *Sox3* in TSA treated explants at levels consistent with a weakly neuralized state (Figure 2.15), despite the apparent inability of explanted cells to adopt definitive neural fates (Fig.2.18A,B). We wondered if markers of other lineages might be similarly expressed following HDAC inhibition. To test this, we examined the expression of genes pertaining to the adoption of mesodermal (*MyoD*), endodermal (*Sox17*) and neural (*Sox11*) states in animal pole explants over developmental time via *in situ* hybridization. Control explants should not express any of these markers at stage 13, by which time cells should have lost pluripotency and become restricted to an epidermal state. However, in TSA treated explants we observed significant expression of *MyoD* (100%, n=26 (control: 0%, n=24)), *Sox17* (100%, n=28 (control: 0%, n=20)) and *Sox11* (96%, n=24 (control: 5%, n=20)) (Figure 2.21 A,B,C). As with *Sox3*, these levels were lower than would be induced following activin or chordin-mediated induction of specific lineage states. We used qPCR to quantify changes in the expression of genes linked to various lineage states at both stage 9, when control cells are pluripotent, and stage 13, when those cells should have become lineage restricted. Striking increases in lineage-linked gene expression were noted even in blastula stage explants treated with TSA, including *Olig2* and *Sox11* (neural markers), *Sox17* and *VegT* (endodermal markers), and *MyoD* and *Xbra* (mesodermal markers), and this aberrant expression was maintained through neurula stages (Figure 2.22). We compared the levels of *Sox17* and *Xbra* expression in TSA treated explants to the levels seen following activin treatment and found that they were significantly lower (Figure 2.23A, B).

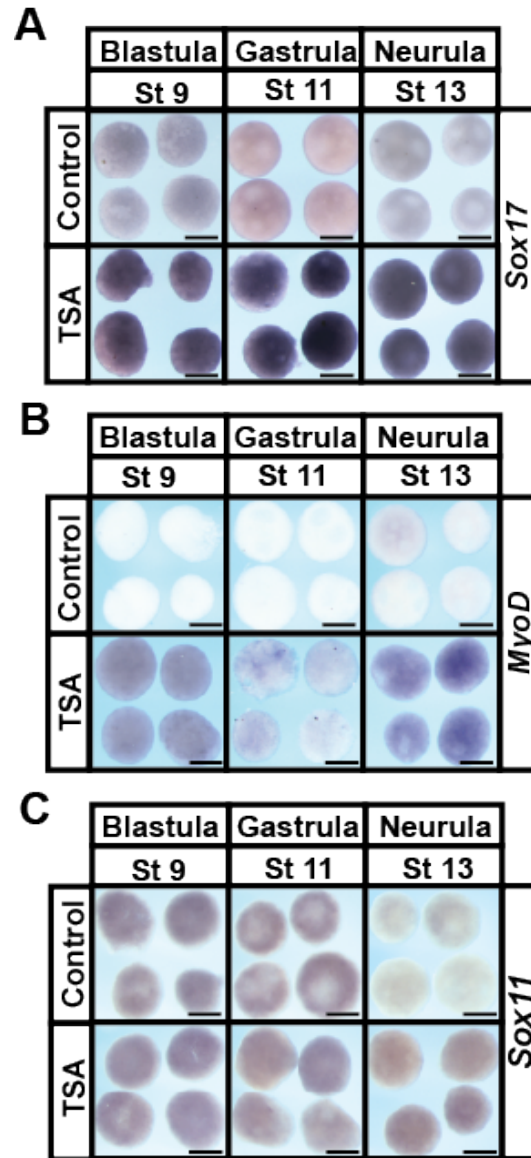


Figure 2.21 Loss of HDAC activity results precocious expression of markers of various lineages

In situ hybridization examining *MyoD*, *Sox17* and *Sox11* expression in aging animal cap explants treated with vehicle control (DMSO) or Trichostatin A (500nM) (A,B,C). TSA treatment results in precocious expression of markers of various lineages as evidenced by increased expression of *MyoD* (mesoderm), *Sox17* (endoderm) and *Sox11* (neural). Explants were cultured alongside sibling embryos grown from blastula (Stage 9) to mid-gastrula (Stage 11) until late-gastrula stages (Stage 13). Scale bars: 250 μ M.

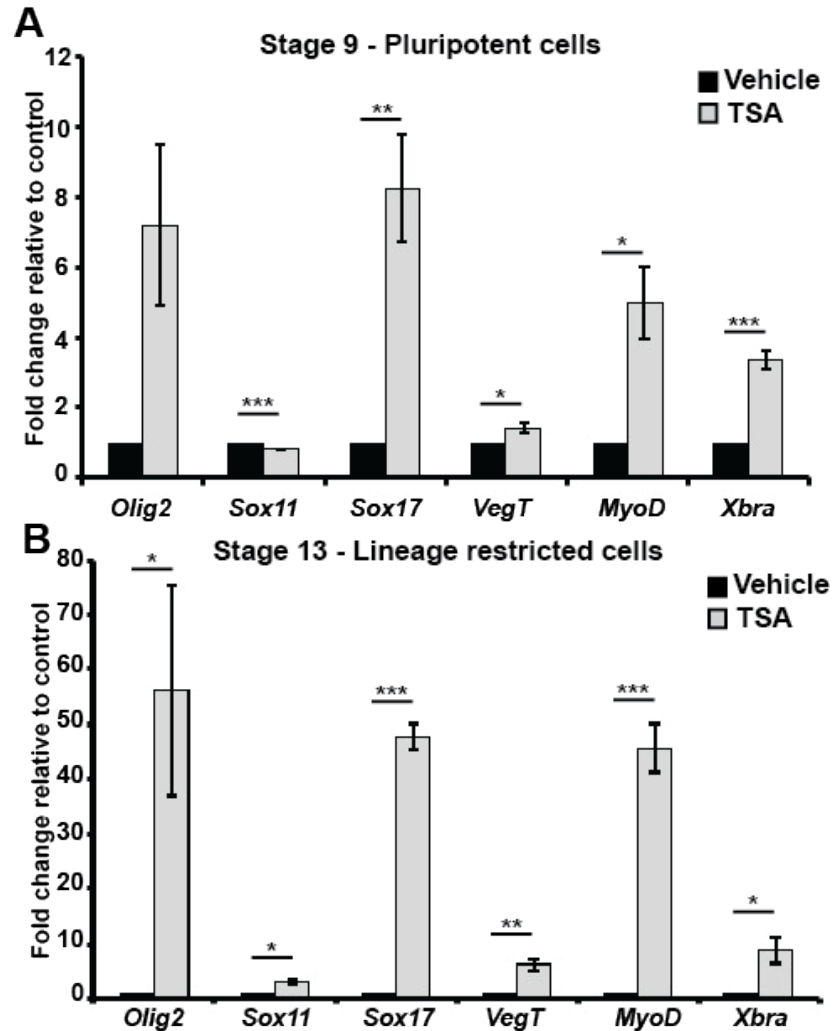


Figure 2.22 Loss of HDAC activity results precocious expression of markers of various lineages

Quantitative RT-PCR examining lineage marker gene expression in aging animal cap explants treated with vehicle (DMSO) or inhibitor (TrichostatinA - 500nM) (A,B,C). TSA treatment results in precocious expression of markers of various lineages as evidenced by increased expression of *Xbra* and *MyoD* (mesoderm), *VegT* and *Sox17* (endoderm) and *Olig2* and *Sox11* (neural). Explants were cultured alongside sibling embryos grown from blastula (Stage 9) until late-gastrula stages (Stage 13). (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$)

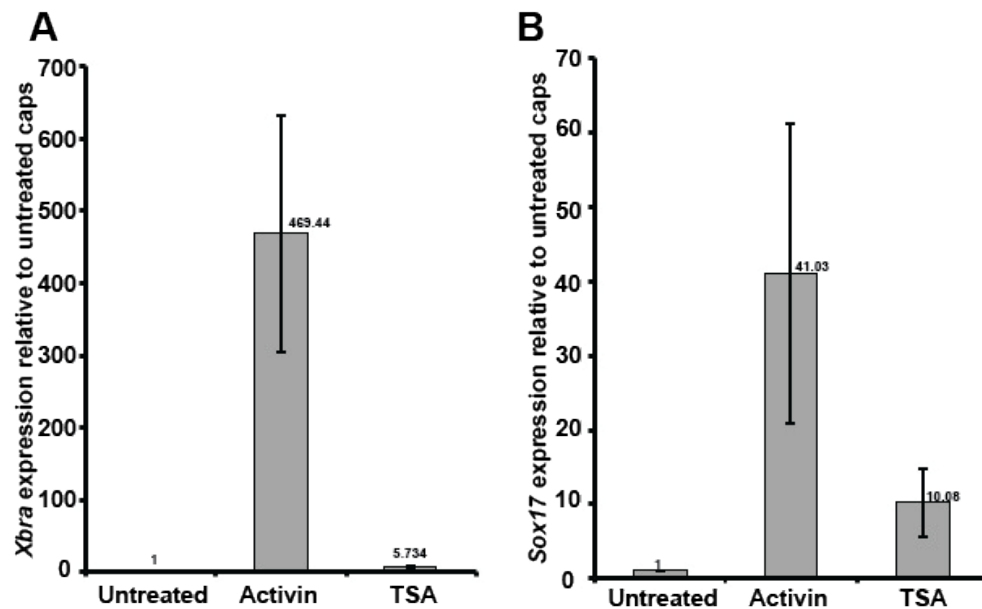


Figure 2.23 TSA treatment causes low level of expression of mesoderm/endoderm markers in comparison to Activin induction

Quantitative RT-PCR examining lineage marker gene expression in aging animal cap explants treated with vehicle (DMSO) or inhibitor (TrichostatinA -500nM) or Activin (low-mesoderm induction/high-endoderm induction) (A,B). TSA treatment results in precocious expression of *Xbra* and *Sox17* but these levels are not as high as expression after Activin induction. Explants were cultured alongside sibling embryos grown until mid-gastrula stages (Stage 11).

Given the aberrant expression of multiple lineage markers observed in TSA-treated explants, we wished to examine makers of lineage determination in later stage embryos. To determine effects on CNS development, we utilized a transgenic line in which GFP expression is driven by the N- β -tubulin promoter *Xla.Tg(tubb2b:mapt-GFP*; (J. K. Huang, Dorey, Ishibashi, & Amaya, 2007). To validate the line in our hands, Tubb2-GFP embryos were co-stained with an E7-N-Tubulin antibody (Figure 2.24). TSA treated Tubb2-GFP embryos showed a pronounced increase in expression of the Tubb2-GFP reporter, particularly in the anterior CNS and in the eye (Figure 2.25), consistent with the expanded expression of Sox3 noted in response to TSA treatment (Fig.1E). In order to determine when this enhanced neural commitment commenced, we followed the expression of the Tubb2-GFP reporter in both TSA and vehicle treated embryos as they developed. Strikingly, strong expression of Tubb2-GFP was noted in anterior regions of TSA treated embryos as early as stage 16, a time when no GFP is detected in control embryos, indicating that CNS development is not delayed as a consequence of HDAC inhibition (Figure 2.26A). The precocious and expanded expression of this neural reporter in response to TSA treatment was not similarly accompanied by an increase in staining for Muscle Actin (Figure 2.26B). Indeed, somatic muscle was shifted caudally in TSA treated embryos as compared to vehicle treated controls, possibly as a consequence of enhanced neural development in anterior regions.

Histone acetylation increases as cell transit from pluripotency to lineage restriction

HDACs function by removing acetyl groups from lysine residues in the N-terminal tail of histones. Since specific histone marks, including H3K9Ac and H3K27Ac, have previously been

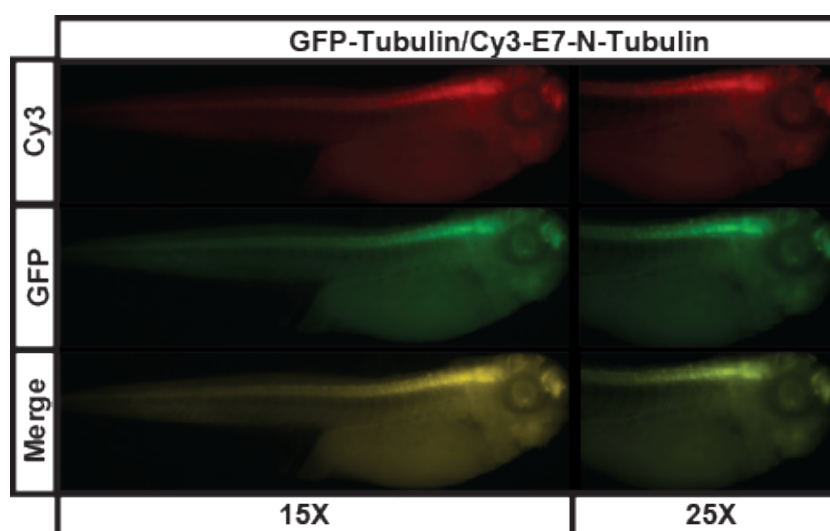


Figure 2.24 Validation of GFP-Tubulin transgenics

Immunofluorescence staining on Tubb2-GFP transgenic embryos co-staining GFP with N-Tubulin E7 antibody. Strong overlap was seen between GFP and N-Tubulin antibody staining at early tadpole stages (Stage 25).

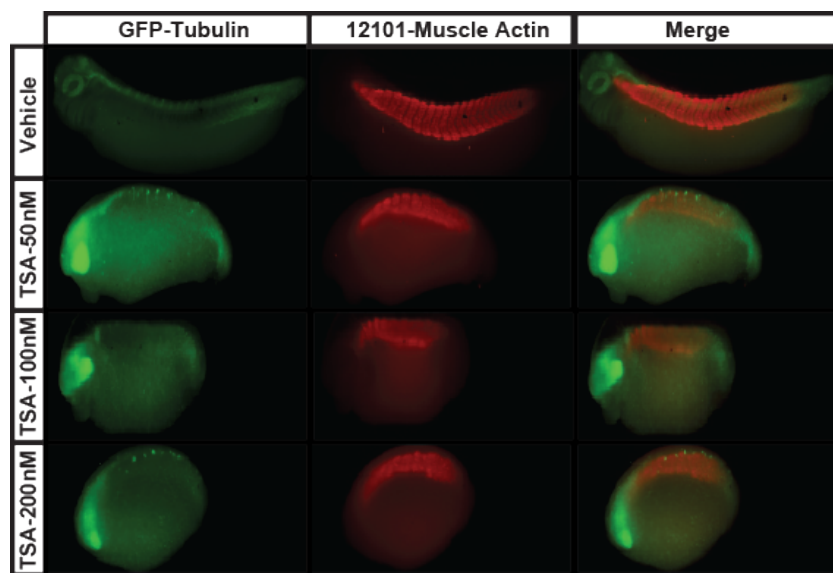


Figure 2.25 TSA treatment enhances expression of Tubb2-GFP reporter

Immunofluorescence staining on Tubb2-GFP transgenic embryos co-staining GFP with muscle actin 12101 antibody after treatment with vehicle or increasing amounts of TSA (50nM-200nM). TSA treatment resulted in pronounced increase in expression of the Tubb2-GFP reporter, especially in the anterior CNS and in the eye. Embryos were treated at mid-gastrula stages (stage 11).

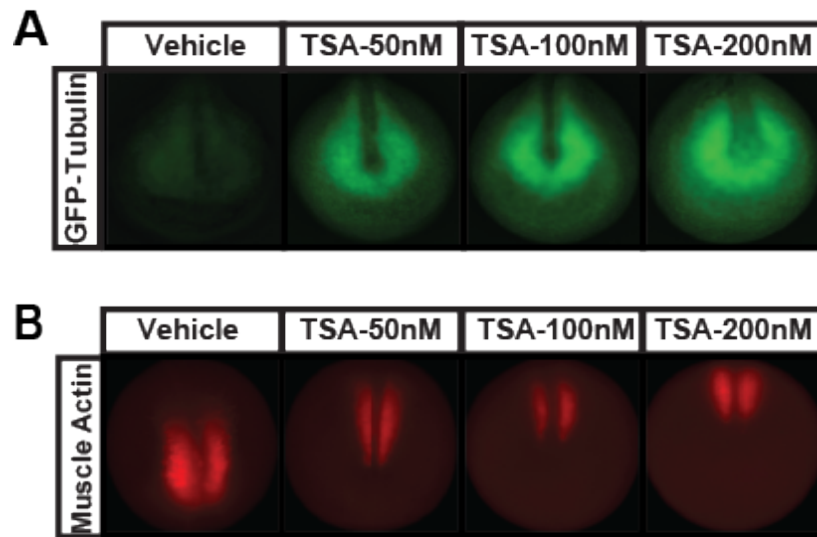


Figure 2.26 TSA treatment causes precocious expression of Tubb2-GFP reporter

(A) Tubb2-GFP transgenic embryos following treatment with vehicle or TSA (50nM – 200nM). Embryos were treated at mid-gastrula stages (stage 11). TSA treatment results in a enhanced and precocious expression of Tubb2-GFP reporter. (B) Immunofluorescence examining 12101- muscle actin staining on embryos after treatment with vehicle or TSA (50nM – 200nM). Embryos were treated at mid-gastrula stages (stage 11). Somatic muscle actin expression is shifted caudally in TSA treated embryos in comparison to vehicle control.

linked to determining the developmental state of a cell, we were interested in understanding when these marks accumulate as pluripotent cells progress through lineage restriction (Creyghton et al., 2010; Karmodiya et al., 2012). Western blot analysis of *Xenopus* animal pole explants showed that global levels of H3K9Ac and H3K27Ac are very low at blastula stages, when cells are pluripotent, and increase as cells progress from gastrula to neurula stages and become lineage restricted (Figure 2.27A). While total H3 levels also increase in these explants, the observed changes in H3K9Ac and H3K27Ac are significant even when normalized to total H3 levels (Figure 2.27B). Interestingly, TSA treatment dramatically increases the levels of H3K9Ac and H3K27Ac histone acetylation even in pluripotent cells (Stage 9) and more so in lineage restricted explants (Stage 13) (Figure 2.28, 2.29A,B). This suggests that HDACs actively maintain low levels of H3K9Ac and H3K27Ac in pluripotent cells, and that increases in these marks might underlie the loss of pluripotency observed following TSA/VPA treatment.

HDAC activity promotes retention of pluripotency

Given that low levels of H3K9Ac and H3K27Ac correlate with pluripotency, we wondered if increasing HDAC activity might promote or sustain pluripotency gene expression animal pole cells. mRNA encoding HDAC1 was injected into 2-cell *Xenopus* embryos targeting the animal pole, and explants were isolated at blastula stages. These explants display dramatically reduced levels of H3K9Ac and H3K27Ac relative to control explants (Figure 2.30). When we examined the changes in gene expression characteristic of progression from the pluripotent to lineage restricted state in these explants, we found HDAC1 activity promoted sustained expression of Sox3 (96%, n=28 (control: 0%, n=29)) and Oct91, Oct60 (Figure 2.31) and a failure to initiate EPK expression (90%, n=29 (control: 3%, n=30)) (Figure 2.32),

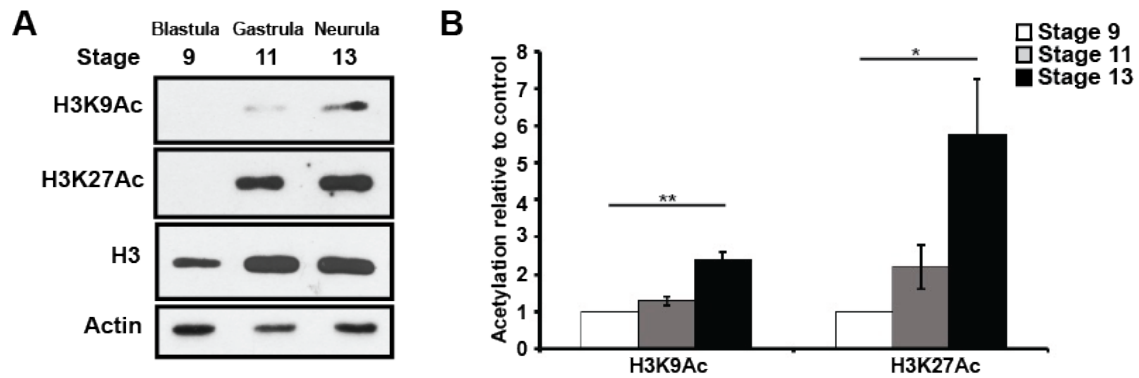


Figure 2.27 Histone acetylation is increase with lineage restriction

Western blot analysis of aging animal caps examining H3K9Ac and H3K27Ac alongside total H3 levels via chemiluminescence (A) and quantified using Odyssey (B). H3K9Ac/H3K27Ac is low in pluripotent cells and increases with lineage restriction. Explants were cultured alongside sibling embryos until blastula (Stage 9), mid-gastrula (Stage 11) and neural plate stages (Stage 13). * $P < 0.05$, ** $P < 0.01$.

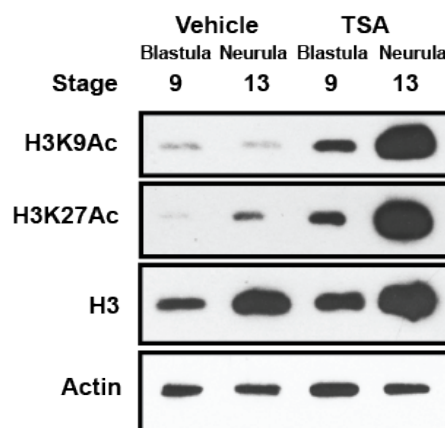


Figure 2.28 TSA treatment dramatically increases histone acetylation in animal cap explants

Western blot analysis of aging animal caps treated with vehicle (DMSO) or inhibitor (TSA-500nM) examining H3K9Ac and H3K27Ac alongside total H3 levels via chemiluminescence. H3K9Ac/H3K27Ac increases dramatically after TSA treatment. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

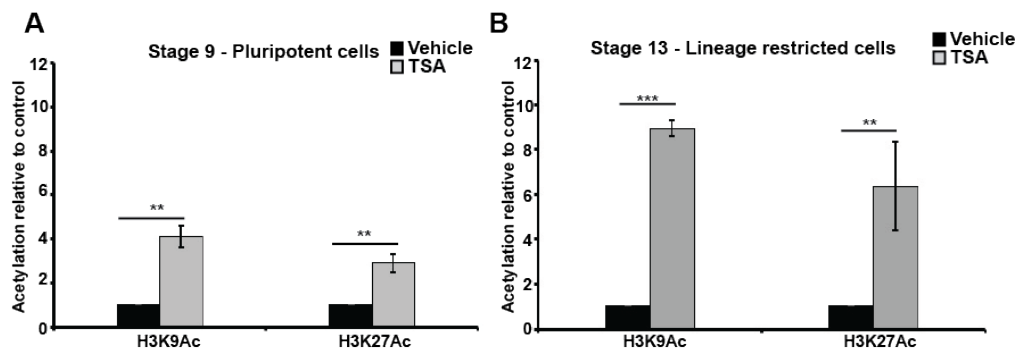


Figure 2.29 TSA treatment causes significant increase in histone acetylation in aging animal cap explants

Western blot analysis of aging animal caps treated with vehicle (DMSO) or inhibitor (TSA-500nM) examining H3K9Ac and H3K27Ac alongside total H3 levels and quantified using Odyssey. H3K9Ac/H3K27Ac increases dramatically after TSA treatment. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

P< 0.01, *P<0.005.

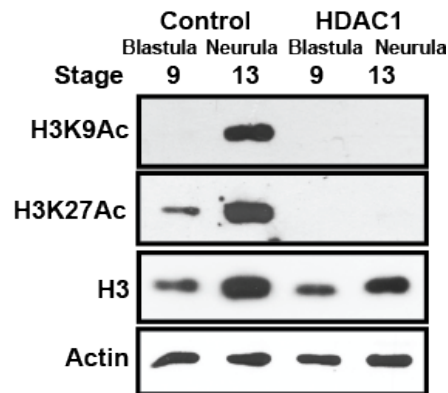


Figure 2.30 TSA treatment causes significant increase in histone acetylation in aging animal cap explants

Western blot analysis of aging animal caps from control embryos or embryos injected with HDAC1 mRNA examining H3K9Ac and H3K27Ac alongside total H3 levels. H3K9Ac/H3K27Ac decreases dramatically after HDAC1 overexpression. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

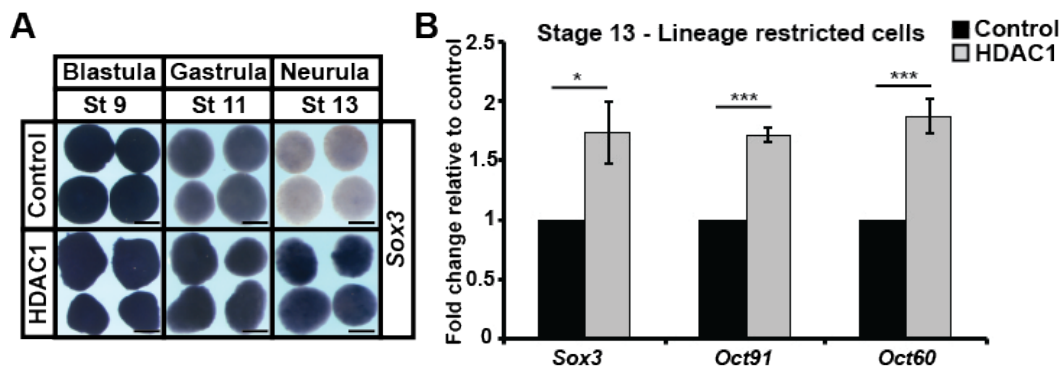


Figure 2.31 HDAC1 activity promotes pluripotency gene expression in animal cap explants

In situ hybridization examining *Sox3* (A) and quantitative RT-PCR examining *Sox3*, *Oct91* and *Oct60* (B) expression in aging animal caps from control embryos or embryos injected with HDAC1 mRNA. HDAC1 over-expression increases pluripotency gene expression in the aging explants. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9), mid-gastrula (Stage 9) and neural plate stages (Stage 13).

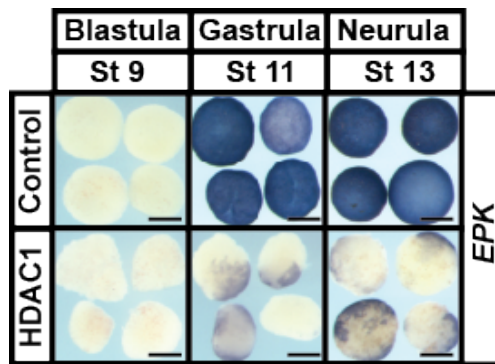


Figure 2.32 Increased HDAC1 activity prevents normal lineage restriction in animal cap explants

In situ hybridization examining *EPK* expression in aging animal caps from control embryos or embryos injected with HDAC1 mRNA. HDAC1 over-expression prevents normal lineage restriction in these explants as evidenced by loss of *EPK* expression. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9), mid-gastrula (Stage 11) and neural plate stages (Stage 13).

consistent with a retention of pluripotency. Similar increases in Sox3 expression were observed when HDAC1 (87%, n=31) or HDAC2(81.5%, n=38) or HDAC1 and 2 in combination (86%, n=36) (control: 5%, n=40) were expressed at high levels suggesting that HDAC1/2 might be functionally redundant in this context (Figure 2.33). No changes were observed in the expression of *Nrp1* or *Sox11* (neural markers) upon HDAC1 overexpression, confirming that these explants were not defaulting to a neural state ((*Nrp1*: Chd : 100%, n = 28, HDAC: 0%, n=29),(*Sox11*: Chd : 100%, n = 28, HDAC1: 0%, n=28)) (Figure 2.34).

HDAC activity promotes the neural crest state

We have proposed that neural crest cells arise as a consequence of retaining characteristics of earlier, pluripotent blastula cells. We would therefore expect neural crest cells, like pluripotent blastula cells, to be characterized by low levels of histone acetylation, and we hypothesized that increased HDAC activity might help promote reprogramming to a neural crest state. To test this, Pax3 and Zic1 were expressed in animal pole cells at levels that only weakly promote expression of neural crest markers. Strikingly, we found that co-expression of HDAC1 significantly enhanced expression of *FoxD3* (89%, n=36) and *Snail2* (81%, n=31) compared to Pax3/Zic1 alone (*FoxD3*: 20%, n=44) (*Snail2*: 10%, n=30)) (Figure 2.35). The ability of HDAC1 to enhance neural crest formation is consistent with the loss of neural crest observed when HDAC activity is inhibited and suggests that proper levels of HDAC activity are critical for establishment of neural crest cells.

A prediction of our revised model for neural crest origins is that Pax3/Zic1-mediated reprogramming to a neural crest state should at least partially preserve or restore the levels of H3K9Ac and H3K27Ac characteristic of pluripotent blastula cells. To test this, we examined

H3K9Ac and H3K27Ac levels in the Pax3/Zic1 reprogrammed explants. We found that by stage 13 the levels of these marks are significantly reduced in Pax3/Zic1 explants compared to age matched control explants, and approximate the levels found at stage 9, when control explants are pluripotent (Figure 2.36). These data indicate that neural crest cells and pluripotent blastula cells share key aspects of their epigenetic state. Interestingly, while we observed an initial increase in H3K9Ac levels at stage 9, which likely accompany transcriptional changes occurring in response to Pax3/Zic1 activity, no similar increase in H3K27Ac is observed (Figure 2.37, 2.38). Taken together, our findings demonstrate an essential role for HDAC activity in promoting both the pluripotent and neural crest states and suggest that retaining the histone marks characteristic of pluripotent blastula cells is a key aspect of establishing the neural crest stem cell population.

Discussion

2018 marks the 150th anniversary of the discovery of the Neural Crest, the primary synapomorphy of vertebrates, by Wilhelm His in 1868. The neural crest is distinguished by its retention of stem cell attributes long past the time when neighboring cells in the early embryo have undergone lineage restriction. Understanding the mechanisms underlying this maintenance of pluripotency is key to understanding the evolution of vertebrates, and is important for leveraging the power of these cells for regenerative medicine. In this study, we report a novel role for HDAC activity and histone acetylation in the maintenance of pluripotency, and the genesis of neural crest cells, in *Xenopus*. We show that HDAC activity is required for the formation of the neural crest, and for the pluripotency of the blastula stem cells they are derived from. Inhibition of HDAC activity using the chemical inhibitors TSA or VPA results in precocious expression of markers of multiple lineages, and an accompanying inability to commit

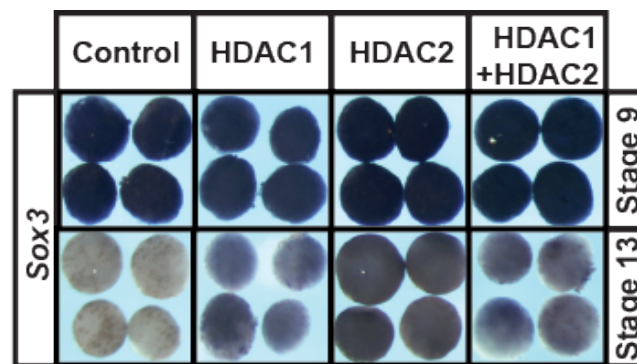


Figure 2.33 HDAC1/2 perform functionally redundant roles for promoting pluripotency gene expression

In situ hybridization examining *Sox3* expression in aging animal caps from control embryos or embryos injected with HDAC1, HDAC2 or HDAC1+HDAC2 mRNA. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

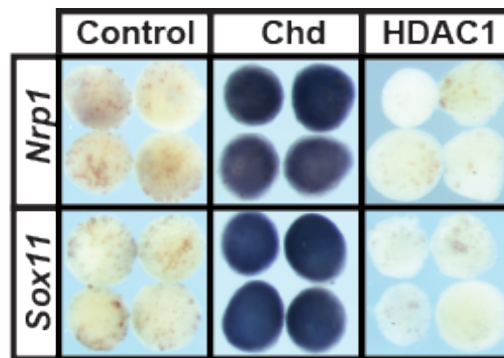


Figure 2.34 HDAC1 activity does not cause animal cap explants to default to neural state

In situ hybridization examining *Sox11* and *Nrp1* expression in aging animal caps from control embryos or embryos injected with HDAC1 or embryos injected with Chordin mRNA. Explants from Chd injected embryos strongly expressed neural markers Sox11 and Nrp1, while HDAC1 over expression did not result in the induction of neural markers. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until late neurula stages (Stage 18).

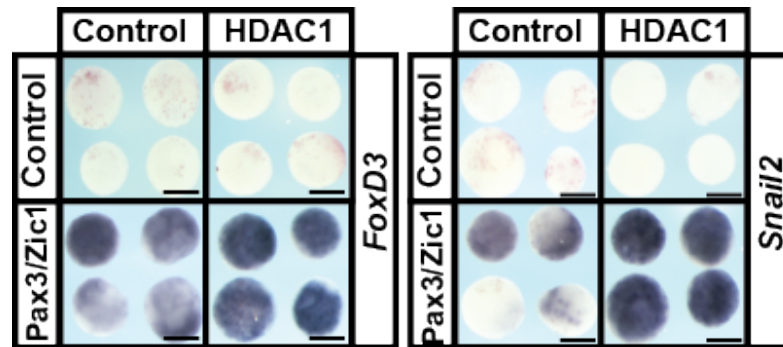


Figure 3.35 HDAC1 activity enhances the ability of animal explants to be reprogrammed to a neural crest state.

In situ hybridization examining *FoxD3* and *Snail2* expression in animal cap explants induced with Pax3/Zic1 at levels titrated for weak neural crest establishment, with/without co-expression of HDAC1. Co-expression of HDAC1 enhanced the ability of explants to be reprogrammed to a neural crest state as evidenced by increased *FoxD3* and *Snail2* expression. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until late neurula stages (Stage 18).

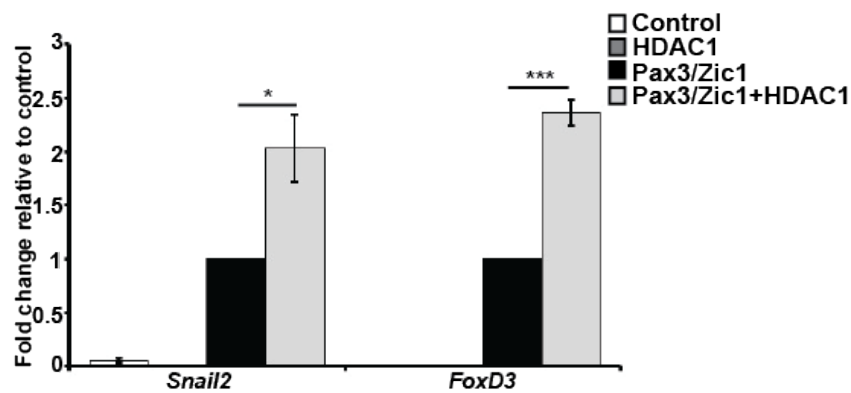


Figure 2.36 HDAC activity promotes neural crest state

Quantitative RT-PCR examining *FoxD3* and *Snail2* expression in animal caps explants induced with Pax3/Zic1 at levels titrated for weak neural crest establishment, with/without co-expression of HDAC1. Co-expression of HDAC1 enhanced the ability of explants to be reprogrammed to a neural crest state as evidenced by increased *FoxD3* and *Snail2* expression. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until late neurula stages (Stage 18).

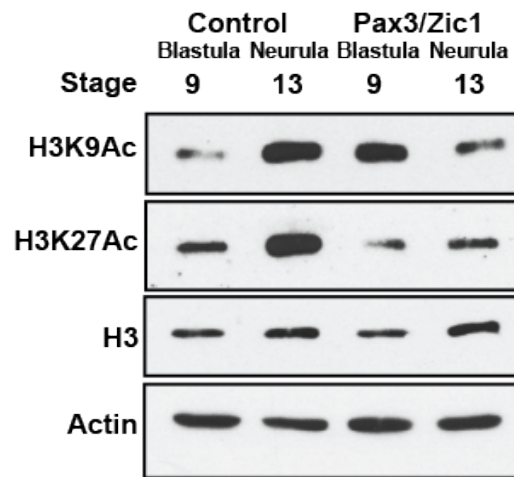


Figure 2.37 Neural crest cells retain low levels of histone acetylation similar to blastula cells

Western blot analysis of aging animal caps from control embryos or embryos injected with Pax3/Zic1 mRNA examining H3K9Ac and H3K27Ac alongside total H3 levels via chemiluminescence. H3K9Ac/H3K27Ac levels are greatly reduced by Stage 13 in neural crest explants and resembles levels close to blastula cells (Stage 9). Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

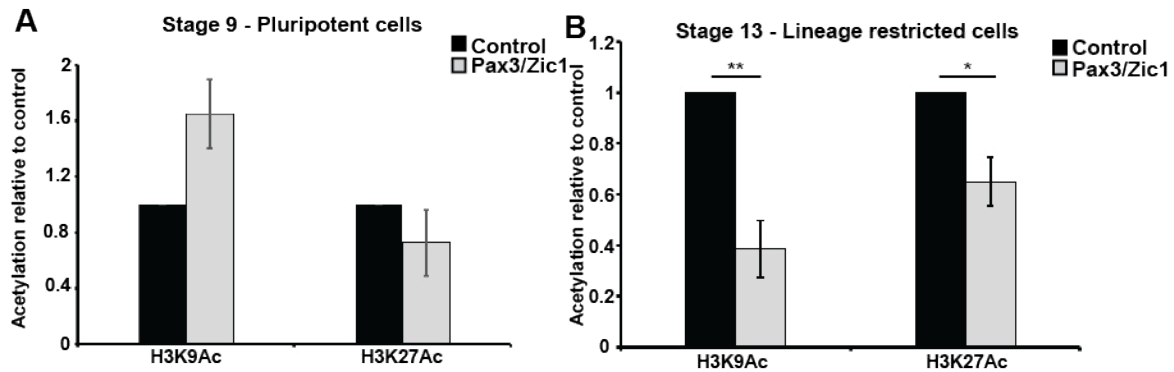


Figure 2.38 Histone acetylation levels in neural crest explants is similar to blastula pluripotent explants

Western blot analysis of aging animal caps from control embryos or embryos injected with Pax3/Zic1 mRNA examining H3K9Ac and H3K27Ac alongside total H3 levels and quantified via Odyssey. H3K9Ac/H3K27Ac levels are significantly reduced by Stage 13 in neural crest explants when compared to control explants. Embryos were injected bilaterally at the 2-cell stage and explants were dissected at Stage 8-9. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

to a specific lineage fate. Finally, we show that pluripotent blastula cells and neural crest cells are both characterized by low levels of H3K9Ac and H3K27Ac acetylation, and that increased HDAC1 activity promotes reprogramming to a neural crest state. Together these findings provide novel insights into the epigenetic mechanisms that control the maintenance of pluripotency in the early embryo and show that regulation of HDAC activity is essential to establishing neural crest stem cells.

Our study is the first to functionally examine epigenetic control of pluripotency in the naïve blastula cells of early *Xenopus* embryos, and its contribution to the retention of potential that underlies the genesis of the neural crest. By contrast, a number of studies have examined a role for epigenetic regulation in the later maintenance, migration and differentiation of neural crest cells in other systems. Epigenomic profiling of human neural crest cells derived from hESCs revealed several neural crest specific enhancers that were marked by H3K27Ac and H3K4me1 and correlated with previously identified chick neural crest enhancer regions (Rada-Iglesias et al., 2012). This suggests that the neural crest state is under a tight regulation by chromatin remodelers. The *de novo* methyltransferase DNMT3A has been shown to be expressed predominantly in the neural crest in avian embryos, and its loss results in downregulation of neural crest specifier genes (N. Hu et al., 2012). Moreover, Jmjd2A, a histone demethylase, has been shown to be recruited to the promoter regions of Snail2 and Sox10, and thus necessary for neural crest specification (Strobl-Mazzulla et al., 2010). Likewise, it has been shown that chromatin remodeler CHD7 associated with PBAF is essential for the activation of the neural crest transcriptional circuitry (Bajpai et al., 2010). Interestingly, during formation of the craniofacial skeleton in mouse, pre-migratory neural crest cells display pre-patterned, poised (H3K27me3/H3K4me2) chromatin confirmation states that are maintained in post-migratory

cells, and may contribute to the plasticity that allows position specific cues to direct neural crest differentiation post migration (Minoux et al., 2017).

HDACs have also been linked to the regulation of neural crest transcription, migration and differentiation. For example, HDAC1 has been shown to physically interact with Ets1 to control the expression of *Id3*, a BMP target gene (C. Wang et al., 2015). Similarly, Snail2 has been shown to recruit the HDAC-Sin3A complex to repress *Cad6B* expression, suggesting that HDAC activity is important for EMT and migration (Strobl-Mazzulla & Bronner, 2012). Interestingly, injection of HDAC inhibitors into the closing neural tube of chick embryos resulted not only in defects in neural tube closure, but also defects in dorsal neural tube patterning and increased expression of a subset of neural crest markers including Pax3 and Sox10 (Murko et al., 2013). This suggests that HDAC activity may have distinct consequences for neural crest cells at different developmental time points. Indeed, at later stages HDAC activity has been shown to regulate the formation of a number of neural crest derivatives including cartilage, melanocytes, cardiomyocytes and components of the PNS and has also been shown to be required for tail regeneration in *Xenopus* (Cunliffe & Casaccia-Bonnet, 2006; Ignatius et al., 2008; 2013; Jacob et al., 2014; Montgomery et al., 2007; Pillai, Coverdale, Dubey, & Martin, 2004; Tseng et al., 2011).

Not surprisingly, HDAC inhibition has developmental consequences beyond the neural crest. In mouse, germ-line deletion of HDACs results in embryonic lethality at e10.5, indicating a requirement of HDAC function during key early developmental decisions in this system (Lagger et al., 2002; Montgomery et al., 2007). Similarly, treatment of gastrula or later stage *Xenopus* embryos with VPA or other HDAC inhibitors results in severe developmental abnormalities such as axial malformations and neural tube defects (Gurvich et al., 2005).

Nevertheless, our finding that in early embryos neural crest cells and neural plate border cells are more sensitive to HDAC inhibition than are other cell types, including CNS and epidermal cells, suggests a critical link to the regulation of developmental potential. Consistent with this, a number of studies have shown that HDACs are fundamental for the maintenance of pluripotency and/or appropriate differentiation in cultured embryonic stem cells (Dovey et al., 2010; Jamaladdin et al., 2014; Kidder & Palmer, 2012). In concordance with our findings here, HDAC inhibition in ES cell culture has been shown to result in both negative and positive changes in gene expression (Jamaladdin et al., 2014; Karantzali et al., 2008; Zupkovitz et al., 2006). Interestingly, genome wide mapping of HDAC occupancy has found their enrichment on some active genes that may be regulated by a dynamic cycle of histone acetylation and deacetylation (Z. Wang et al., 2009).

While there is a wealth of literature on the role of HDACs in the maintenance of pluripotency in cultured embryonic stem cells, a single clear model has not emerged from this work. HDAC containing complexes have been shown to assemble on, and promote expression of, pluripotency genes in ESCs and pre-iPSCs, suggestive of a positive regulatory role for HDACs in pluripotency (Baltus et al., 2009; Saunders et al., 2017). Moreover, the reported effects of HDAC inhibition in cultured ES cells appear to vary significantly dependent on the inhibitor utilized, the concentration and duration of treatment and the source of the stem cells (murine vs human). Studies in cultured mouse and human ES cells have revealed that levels of histone acetylation (H3K9/14Ac) increase and decrease dynamically when the cells are induced to differentiate (Hezroni et al., 2011; P. Liu et al., 2015; Markowetz, Mulder, Airoidi, Lemischka, & Troyanskaya, 2010; Melcer et al., 2012; Moussaieff et al., 2015; Qiao et al., 2015). Moreover, HDAC inhibition during differentiation of mEpiSCs prevents neural

differentiation, whereas in hESCs the effects are time dependent; early inhibition leads to maintenance of pluripotency while later inhibition promotes neural fate commitment (P. Liu et al., 2015; Qiao et al., 2015; J. Yang et al., 2014). It has been suggested that HDAC inhibition may promote progression of naïve mESCs towards a primed mEpiSCs state, and hESCs to an earlier state, but in both cases it promotes self-renewal of these cells (Ware et al., 2009).

Our findings in early *Xenopus* embryos demonstrate a clear role for HDAC activity in the maintenance of pluripotency in naïve blastula stem cells. Importantly, we further show that maintenance of this activity is linked to the events that preserve the potency in a subset of these cells leading to formation of the neural crest. Interestingly, inhibition of HDAC activity led to enhanced expression of genes linked to multiple lineage states, in addition to a loss of pluripotency markers, yet these cells were unable to give rise to any lineage tested. We find that low levels of H3K9Ac and H3K27Ac acetylation are a shared feature of both pluripotent blastula cells and neural crest cells, building on our recent work showing that a requirement for high Map kinase activity and low Akt activity is similarly shared by these cell types (Geary & LaBonne, 2018). Importantly, we also find that increased HDAC1 activity can enhance reprogramming to a neural crest state, which may have implications for regenerative medicine. While future work will need to address the genome occupancy of HDACs at genes up and down-regulated following HDAC inhibition, these studies shed important new light on the epigenetic mechanisms that control maintenance of pluripotency, the establishment of neural crest stem cells, and the evolution of vertebrates.

Materials and Methods

Embryological methods

Wildtype *Xenopus laevis* embryos were obtained using standard methods and staged according to Nieuwkoop and Faber (1994). For animal cap assays, ectodermal explants were manually dissected at early blastula (St. 8-9) from embryos microinjected with the indicated mRNA and/or treated with the specified inhibitor at the 2-cell stage and cultured in 1x MMR until sibling embryos reached the denoted stage. mRNA for microinjection was *in vitro* transcribed from a linearized DNA template using the SP6 mMessage mMachine Kit (Ambion). Pax3-GR and Zic1-GR expression explants were dissected from embryos treated at St8 with 10uM Dexamethasone (Sigma-Aldrich) as previously described (Buitrago-Delgado et al., 2015). For Morpholino experiments, a translation blocking HDAC1 morpholino (Gene Tools, Sequence: 5' GAGTCAGCGCCATTTTCCTTC 3') was injected at the 8 cell stage in isolation or co-injected with HDAC1/2 mRNA. For activin experiments, animal cap explants from control or inhibitor treated embryos were dissected at blastula stages and were cultured with recombinant activin protein (R&D Systems) at a final concentration of 20-40ng/mL for mesoderm, and 100ng/mL for endoderm induction in 1x MMR supplemented with 0.1% BSA as a carrier. Manipulated embryos/explants were processed for *in situ* hybridization by fixing in 1x MEMFA and dehydrating in 100% methanol. *In situ* hybridization was performed using digoxigenin labelled RNA probes and developed using BM Purple substrate (Roche) (LaBonne & Bronner-Fraser, 1998). For TUNEL staining, DMSO/TSA treated explants were processed alongside a DNMT3B mRNA injected embryos (positive control) and processed as previously described (Bellmeyer et al., 2003). Results shown are representative of at least three independent experiments.

Western blot analysis

Animal cap explants (20 – 40 explants) were collected at the indicated stages and lysed in TNE lysis buffer (50mM Tris-HCl [pH 7.4], 150mM NaCl, 0.5mM EDTA, and 0.5% Triton X-100) supplemented with protease inhibitors (Aprotinin, Leupeptin and PMSF) and complete Mini tablet (Roche). SDS-PAGE and western blot analysis was used to detect proteins and modifications using the following antibodies: H3K9Ac (#9649, Cell Signaling, 1:2000), H3K27Ac (ab4729, Abcam, 1:2000), H3 (#3638 and #4499, Cell Signaling, 1:1000) and Actin (A2066, Sigma-Aldrich, 1:4000). For enhanced chemiluminescence based detection, HRP-conjugated secondary antibodies were used. Results shown are representative of three independent experiments. For detection and quantification using the LiCOR-Odyssey platform, blots were incubated simultaneously with primary antibodies for both H3K9Ac/H3K27Ac and H3. Histone acetylation was detected using the IRDye® secondary antibodies and proteins amounts were quantified using the Image Studio™ Lite software. Relative histone acetylation (H3K29Ac or H3K27Ac) was calculated against total H3 levels. Represented is the mean of three independent biological replicates with error bars depicting the standard error of mean (SEM). An unpaired, two-tailed t-test was utilized to determine significance.

RNA isolation, cDNA synthesis and qRT-PCR

RNA was isolated from control or manipulated animal cap explants (10-30 explants) using Trizol (Life Technologies) followed by LiCl precipitation. 1µg of purified RNA was used as a template for synthesizing cDNA using High Capacity Reverse Transcription Kit (Life Technologies). qRT-PCR was performed using SYBR® Premix ExTaq II (Takara Bio) and

detected using BioRad CFX96 Connect system. The primer sequences used are available in the supplementary figures. Expression was normalized to ornithine decarboxylase (ODC) and fold change was calculated relative to control samples of the same stage. Represented is the mean of at least three independent biological replicates with error bars depicting the standard error of mean (SEM). An unpaired, two-tailed t-test was utilized to determine significance.

Immunofluorescence Analysis

Wildtype and Tubb2-GFP transgenic (*Xla.Tg(tubb2b:mapt-GFP)*) *Xenopus Laevis* embryos were blocked in Whole mount Block Solution(WMBS) (155mM NaCl, 10mM Tris-Cl, pH7.5, 10% FBS, 5%DMSO) and incubated overnight with primary antibodies : 12/101-Actin (DSHB, 1:15), GFP(Rabbit GFP, A-11122, Life Technologies, 1:250), E7-Tubulin(DSHB, 1:100). After washing with Tris Buffer Saline with 0.1% Triton, the embryos were re-blocked in WMBS and incubated overnight with secondary antibodies: Goat anti-Rabbit - Alexa Fluor 488, A-11008, Life Technologies (1:250), Goat Anti-Mouse Cy3, 115-165-146, Jackson Labs (1:500).

DNA Constructs and Inhibitors

Full length *Xenopus Laevis* HDAC1, HDAC2 was obtained from the *Xenopus* ORFeome and sub-cloned into a pCS2 vector for microinjection and into a pGEM-T vector for RNA probe synthesis. Pax3-GR and Zic1-GR constructs were a kind gift from Jean-Pierre Saint-Jeannet (New York University). For HDAC inhibition, embryos/explants were treated with Trichostatin A (Sigma-Aldrich) at a final concentration of 200-500nM or Valproic Acid (Sigma-Aldrich) at a final concentration of 10-20mM at the noted stage or Romidepsin (Sigma-Aldrich) at a final concentration of 5-15μM.

Animals

All animal procedures were approved by the Institutional Animal Care and Use Committee, Northwestern University, and are in accordance with the National Institutes of Health's "Guide for the Care and Use of Laboratory Animals". Tubb2-GFP (*Xla.Tg(tubb2b:mapt-GFP)*) transgenic line was obtained from the National Xenopus Resource (www.mbl.edu/xenopus/) and generation of this line has been described (Huang et al., 2007).

Chapter 3

Mechanisms of epigenetic regulation of stem cell maintenance

The embryonic neural crest is a unique vertebrate stem cell population that has the ability to retain its stem cell attributes while the rest of the embryo is being lineage restricted. While much is known about the signaling factors and transcription factors that are tasked with maintaining the pluripotency of these cells, we are still unaware of the epigenetic mechanisms that regulate neural crest stem cell maintenance. We recently uncovered a novel role for histone deacetylase (HDAC) activity for maintaining the stem cell attributes of neural crest cells and pluripotent blastula cells, and we found that low levels of histone acetylation are a shared feature of these two cell types. In this study, we extend our previous work and explore the mechanisms through which HDACs and histone acetylation control pluripotency in blastula stem cells. Using genome wide techniques like RNASeq and ChIPSeq, we analyze the global changes in gene expression in response to loss of HDAC activity and investigate the regulation of histone acetylation at genomic loci of pluripotency genes and lineage specific genes during lineage restriction. Further, we compare the mechanisms utilized by HDACs and epigenetic readers, BET proteins in the maintenance of pluripotency. Finally, using mass spectrometry, we explore global changes in abundance of histone modifications that are important for stem cell maintenance and lineage restriction during embryonic development.

Introduction

The embryonic neural crest is a vertebrate stem cell population that distinguishes itself by its unique capability to contribute to multiple germ layers (Hall, 2000; Sauka-Spengler & Bronner-Fraser, 2008). This novel characteristic has puzzled biologists for decades as these cells appear to have a greater developmental potential than the cells from which they arose, defying the embryonic paradigm of progressive lineage restriction (Hoppler & Wheeler, 2015). While it

was previously believed that these cells gained these stem cell attributes through inductive events, recent work from our lab has provided evidence that neural crest cells share a similar transcriptional program and developmental potential to the pluripotent blastula cells in the early embryo and may arise due to a retention of stem cell attributes. Indeed, it was found that several transcription factors such as *Myc*, *FoxD3*, *Tfap2*, *Id3*, *Snail* are shared by these cell states suggesting that the molecular circuitry of pluripotent cells is at-least partially preserved in neural crest cells (Buitrago-Delgado et al., 2015). This revised model for neural crest formation raises interesting questions regarding the mechanisms adopted by these cells to preserve the pluripotent regulatory circuitry of their blastula precursors. Given the transient and complex nature of the pluripotent state, this process would necessitate the calibrated interplay of signaling molecules, transcription factors, and epigenetic modifiers (Niwa, 2007) (Habibi & Stunnenberg, 2017). Interestingly, transcription factors like Snail and Sox5 that are integral to the formation of the neural crest, have been shown to be expressed in and essential for the maintenance of pluripotency of blastula cells (Buitrago-Delgado et al., 2015; Nordin & LaBonne, 2014). Further, we have identified that BMP signaling as well as the FGF-MAPK cascade is critical for maintaining the neural crest stem cell population (Geary & LaBonne, 2018; Nordin & LaBonne, 2014). However, the role of chromatin remodelers that regulate the epigenetic state of neural crest stem cell state is not well understood. Recently, we identified a novel role for histone deacetylase (HDAC) activity in maintaining the pluripotency of neural crest cells and blastula stem cells (Chapter 2). We found that loss of HDAC activity results in the upregulation of markers of different lineages resulting in a loss of pluripotency. We identified that the pluripotent state is associated with low global levels of H3K9Ac and H3K27Ac, and HDACs regulate the neural crest stem cell state by maintaining these cells in a low level of histone

acetylation similar to blastula cells. However, the mechanisms by which HDACs and histone acetylation contribute to the regulation of pluripotency of these cells remains largely unclear.

Histone acetylation, as a major regulator of transcription, plays a vital role in ensuring appropriate gene expression and maintaining the pluripotent stem cell state. The level of histone acetylation in embryonic stem cells is critical for proper cellular fate commitment and differentiation (P. Liu et al., 2015; Melcer et al., 2012; Qiao et al., 2015). Interestingly, specific histone marks such as H3K9/14Ac and H3K27Ac have been shown to mark important developmentally regulated genes in ES cells and are necessary for regulating gene expression in these cells (Creyghton et al., 2010; Karmodiya et al., 2012). Maintaining appropriate levels of histone acetylation seems to be critical for stem cell maintenance as well as for the onset of lineage restriction. Indeed, our work and others have found that an increase in the levels of histone acetylation through HDACi results in a premature loss of pluripotency due to precocious expression of markers of other lineages (Karantzali et al., 2008) (Chapter 2). Thus, regulating the levels of histone acetylation through HDAC and HAT activity seem to be necessary for maintaining the pluripotent stem cell state.

HDACs, in particular Class I HDACs, are very ubiquitous and play important roles in a variety of cellular contexts including embryonic development, stem cell maintenance and differentiation (Haberland, Montgomery, & Olson, 2009b). HDACs lack intrinsic DNA binding ability and recruited to target genes by their direct association with transcriptional activators and repressors, and through their role in large multiprotein complexes like NuRD and Sin3A complexes (Y. Li & Seto, 2016; Seto & Yoshida, 2014). Thus, their activity varies with cellular context and the availability of binding partners in the cell. As low histone acetylation is thought to be a closed and repressed chromatin state, this had promoted the idea of HDACs been

traditionally believed to be repressors of gene expression. However, more recently it has been found that HDACs might have a positive regulatory role in controlling gene expression. It was found that both p300 and HDACs are enriched at the genomic loci of active genes suggesting a dynamic regulation of histone acetylation and deacetylation at active genes and more integral role of HDACs in regulating transcription than was previously known (Z. Wang et al., 2009). Interestingly, it was recently shown that HDACs have a positive regulatory role in regulating gene expression in pluripotent cells. HDAC1/2 double KO results in the reduced expression of pluripotency genes and loss of cell viability (Jamaladdin et al., 2014). Strikingly, the Sin3A-HDAC complex been implicated in promoting pluripotency in ES cells and during somatic cell reprogramming (Baltus et al., 2009) (Saunders et al., 2017). Despite these studies, no unified model has as yet emerged to describe the mechanisms through which HDAC activity regulates pluripotency, especially *in vivo* during embryonic development.

Histone acetylation in the cell is recognized by a class of epigenetic factors with protein domains called bromodomains which have the capability to selectively target acetylated lysine and recruit transcriptional regulatory elements (Marmorstein & Zhou, 2014). Interestingly, these readers of histone acetylation have also been shown to be critical for the maintenance of pluripotency further providing evidence that the levels of histone acetylation are under tight control during lineage restriction. The Bromodomain reader, Brd4, has been shown to interact with acetylated H4 to regulate pluripotency in ES cells as well interact with Oct4 to regulate the pluripotency network (Gonzales-Cope et al., 2016; T. Wu et al., 2015). BRD4 knockout is embryonically lethal, and it has been shown that BRD4 activity is required to maintain pluripotent state in embryonic stem cells (Di Micco et al., 2014; Fernandez Alonso et al., 2017; Houzelstein et al., 2002; W. Liu et al., 2014). This would suggest that the level of histone

acetylation is of critical importance for maintaining pluripotency and ensuring appropriate lineage restriction. Strikingly, studies have as yet not uncovered how histone acetylation and epigenetic readers regulate pluripotency during embryogenesis.

Apart from just histone acetylation, histones are susceptible to other posttranslational modifications (PTM) such as methylation, phosphorylation and ubiquitination. Histone methylation, both on lysine and arginine residues have been shown to be critical for gene activation and repression and are under tight control during stem cell maintenance and lineage restriction (Y. Zhang & Reinberg, 2001). It is hence not surprising that dynamic changes in the abundance of histone modifications occur during embryonic development. Studies have looked at the levels of different histone modifications during early and late embryonic development. Interestingly, it was found that there appears to a hierarchy in the deposition of histone modifications during development to control spatial regulation of gene expression (Akkers et al., 2009). Another study used mass spectrometry to characterize the abundance of histone modifications during different stages of *Xenopus* development – blastula (Stage 9), gastrula (Stage 12), neurula (Stage 18) and tadpole stages (Stage 37) and found interesting changes take place in the global abundance of histone modifications during embryonic development (Schneider et al., 2011). However, it was yet unknown which histone modifications are necessary for maintaining pluripotency and which ones might be required for controlling lineage restriction.

In this study, we explore the mechanisms through which HDACs regulate pluripotency of blastula and neural crest cells. Using genome wide RNA sequencing, we identify that loss of HDAC activity results in global changes in gene expression with half of the genes upregulated and half the genes being downregulated. We observe that several pluripotency factors are

downregulated, while lineage specific genes are upregulated. We utilize ChIP-sequencing and ChIP-qPCR to identify that histone acetylation is regulated differently at the genomic loci of pluripotent genes vs lineage specific genes, suggesting that HDACs function to maintain histone acetylation low at the loci of lineage genes in blastula cells. Further, we compare the mechanisms through which HDACs and BET proteins regulate pluripotency in these cells. Finally, we explore the dynamic changes in histone modifications that occur during lineage restriction and identify ones that are enriched in the pluripotent state and increase with lineage restriction. Taken together, these studies enhance our knowledge of the epigenetic regulation of stem cell maintenance during embryonic development.

Results

Transcriptomic analysis in pluripotent cells depleted for HDAC activity

Our earlier work identified that HDAC activity is necessary for the pluripotency of blastula cells and neural crest cells; and that low levels of histone acetylation are a shared feature of these two cell types (Chapter 2). Since HDAC activity is so vital for the maintenance of the pluripotent state, we sought to determine the mechanism through which HDACs function to regulate pluripotency in these cells. To this end, we carried out a global transcriptomic analysis to explore changes in gene expression in pluripotent and lineage restricted cells with and without HDAC activity. Animal caps were treated with vehicle or Trichostatin A, a HDAC inhibitor and collected at pluripotent stages (Stage 9) as well as lineage restricted stages (Stage 13) (Figure 3.1). Hierarchical clustering revealed that there were no significant batch effects seen between the 2 biological replicates (Figure 3.2). We utilized a standard NGS pipelines (STAR-HTSeq-

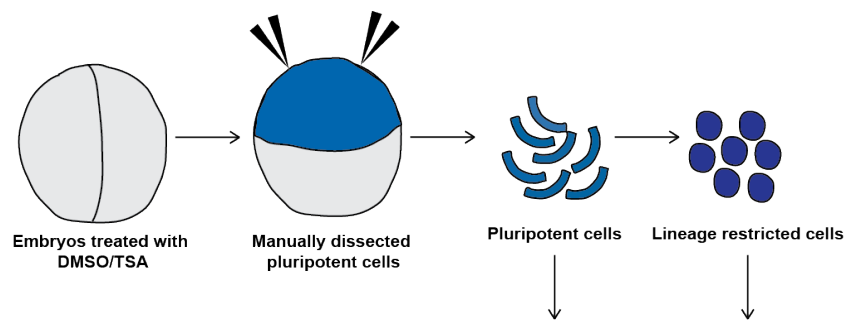


Figure 3.1 Animal cap explants RNA-sequencing workflow

Schematic representation of RNA-sequencing experiment of animal cap explants with/without TSA treatment. Animal cap explants were dissected from embryos were treated at the two-cell stage with vehicle (DMSO) or Inhibitor (TSA). Explants were cultured in inhibitor containing media until sibling embryos reached blastula stages (Stage 9) or neural plate stages (Stage 13) and collected for RNA isolation, library preparation and sequencing.

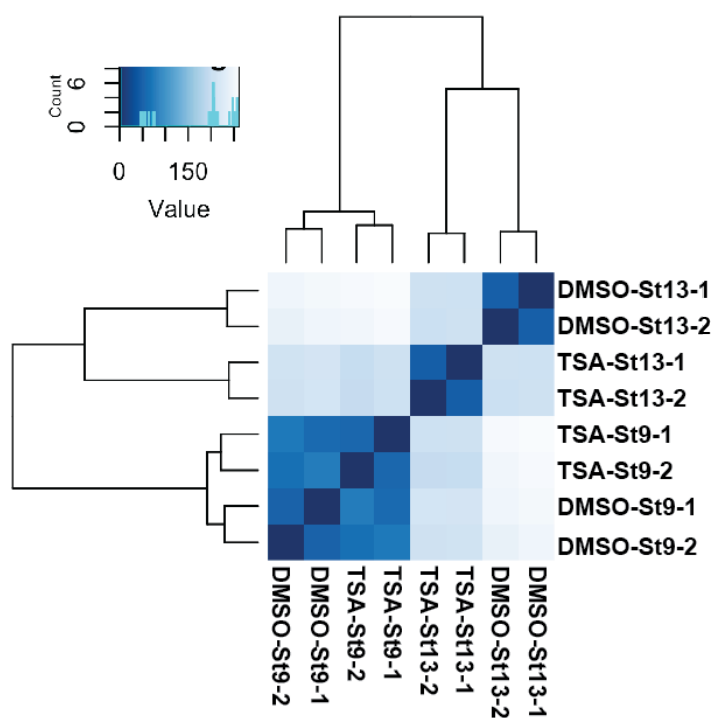


Figure 3.2 Hierarchical clustering of RNAseq biological replicates

Heatmap depicting the hierarchical clustering of RNAseq biological replicates based on the Euclidean distance between samples. No significant batch effects are seen as similar sample cluster together irrespective of biological replicate.

MA-plot (A) and volcano plot (B) depicting gene expression changes in pluripotent cells after TSA treatment. Loss of HDAC activity results in ~50% of the genes to be upregulated, and 50% of the genes to be downregulated. Only 2% of the genes are significantly changed after TSA treatment (920 genes).

MA-plot (A) and volcano plot (B) depicting gene expression changes in pluripotent cells after TSA treatment. Loss of HDAC activity results in ~50% of the genes to be upregulated, and 50% of the genes to be downregulated. Only 2% of the genes are significantly changed after TSA treatment (920 genes).

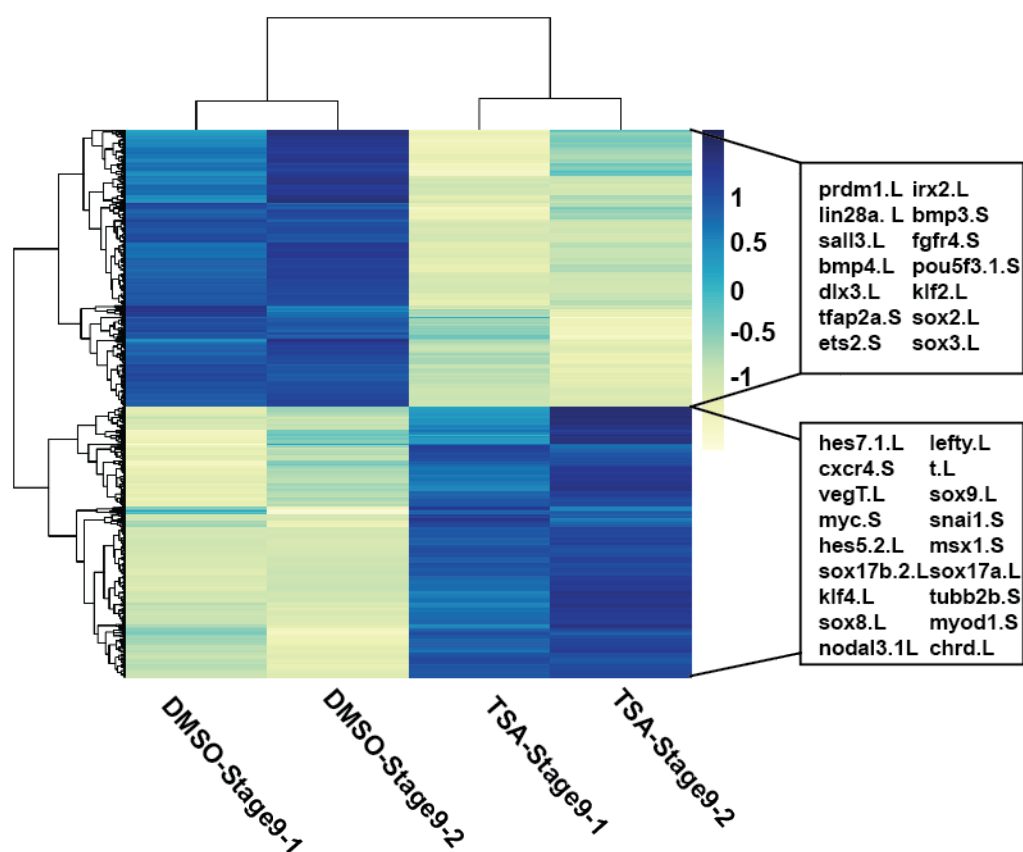


Figure 3.4 TSA treatment results in loss of pluripotency gene expression and upregulation of lineage genes

Heatmap depicting genes that are significantly differentially expressed after loss of HDAC activity in pluripotent cells. Genes that are downregulated are enriched for genes with known roles in the maintenance of pluripotency, while several lineage specific genes are upregulated after TSA treatment.

DESeq2) to identify differentially expressed genes after TSA treatment and genes were considered significant at a statistical threshold of adjusted Pvalue < 0.05.

HDAC activity controls appropriate gene expression in pluripotent cells

We expected to see global changes in gene expression in response to loss of HDAC activity. However, in the pluripotent cells, only 2% of the genes (920 genes) were significantly changed after TSA treatment. This suggested that HDAC activity is not performing a global function but is instead involved in a very specific and regulated function in these cells. Strikingly, in spite of the fact that HDACs are traditionally thought to function as repressors of genes expression, TSA treatment resulted in equal number of genes upregulated and equal number of genes downregulated (Figure 3.3A,B). These data suggest that HDACs play a vital and possibly positive role in regulating appropriate gene expression in pluripotent cells. Indeed, this is consistent with our results and other studies that have suggested that HDAC activity is necessary for pluripotency gene expression (Chapter 2) (Baltus et al., 2009; Saunders et al., 2017). When we examined genes that were downregulated after TSA treatment, there are a number of genes that play known roles in the maintenance of pluripotency, including *Lin28a*, *TFAP2a*, *Sox2*, *Sox3*. Interestingly, we find that genes that were upregulated included several genes that are known to be required for lineage restriction such as *MyoD1*, *VegT*, *Lefty*, *Msx1*, *Tubb2b*, *Xbra* etc (Figure 3.4). Strikingly, loss of HDAC activity resulted in the upregulation of lineage specific genes of different cell fates simultaneously as seen the by the upregulation of mesodermal, endodermal and neural markers in these pluripotent cells. This is consistent with the phenotypes we had seen previously via quantitative RT-PCR as well as reported by others previously in studies in ESCs (Dovey et al., 2010; Karantzali et al., 2008) (Chapter 2).

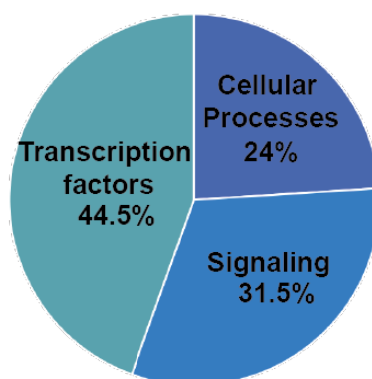


Figure 3.5 Functional categories of genes of differentially expressed genes after loss of HDAC activity

Pie chart representing the functional categories of top 200 significantly genes differentially expressed after TSA treatment. These are predominantly transcription factors and signaling molecules, with several genes involved in cellular processes such as cell cycle, apoptosis, cell death etc.

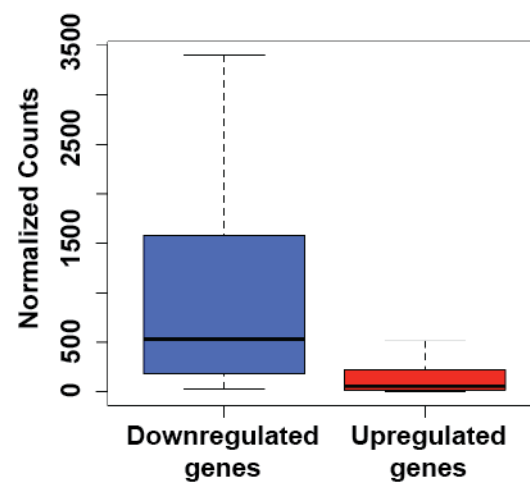


Figure 3.6 Comparative analysis of upregulated and downregulated genes after TSA treatment

Box plot depicting mean expression (normalized counts) of genes upregulated vs downregulated after TSA treatment. Genes that are downregulated are highly expressed in pluripotent cells while upregulated genes predominantly have low expression in these cells.

Interestingly, several neural crest genes such as *Sox9*, *Snail*, *Myc* are also upregulated following TSA treatment suggesting that precise control of HDAC activity is important for neural crest formation (Figure 3.4). A comparative analysis of the functions of top 200 significantly changed genes after TSA treatment shows that these genes are predominantly transcription factors and signaling molecules as well as several genes involved in other cellular processes such as cell cycle, apoptosis and cell division (Figure 3.5). This suggests that HDAC activity is necessary for regulating the expression of critical developmental genes in pluripotent cells. Indeed, we find that the mean expression of genes that are upregulated after TSA treatment is significantly lower than the expression of genes that are downregulated after TSA treatment (Figure 3.6). These data indicate that during normal development, HDACs are necessary for ensuring appropriate gene expression and that loss of HDAC activity results in aberrant gene expression pushing the cells out of a pluripotent state.

HDACs are recruited to DNA through interaction with context specific transcription factors and binding partners (Seto & Yoshida, 2014). We hypothesized that HDACs might be recruited to the genomic loci of upregulated vs downregulated genes by different transcription factors in a context specific manner. In order to identify potential interacting partners of HDACs in this context, we performed motif analysis on the promoter regions (-1.5kb from the TSS) of the upregulated vs downregulated genes and identified the enrichment for motifs of different transcription factors that have been previously described to interact with HDACs. Interestingly, when we compare the motif enrichment at the promoters of differentially changed genes after TSA treatment, we find enrichment for several transcription factors including Klf5, GAGA, Sp1. However, this enrichment is seen both in upregulated and downregulated genes to varying

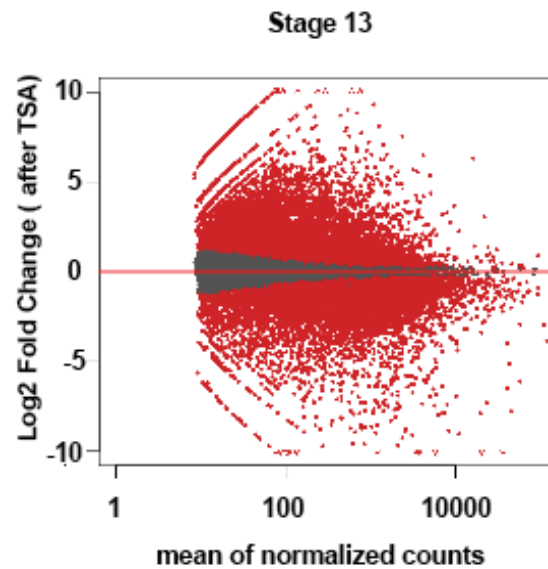


Figure 3.7 TSA treatment causes dramatic changes in gene expression

MA-plot depicting gene expression changes in lineage restricted cells after TSA treatment. Loss of HDAC activity results in ~50% of the genes to be upregulated, and 50% of the genes to be downregulated. 15198 genes are differentially expressed after TSA treatment in Stage 13 animal cap explants.

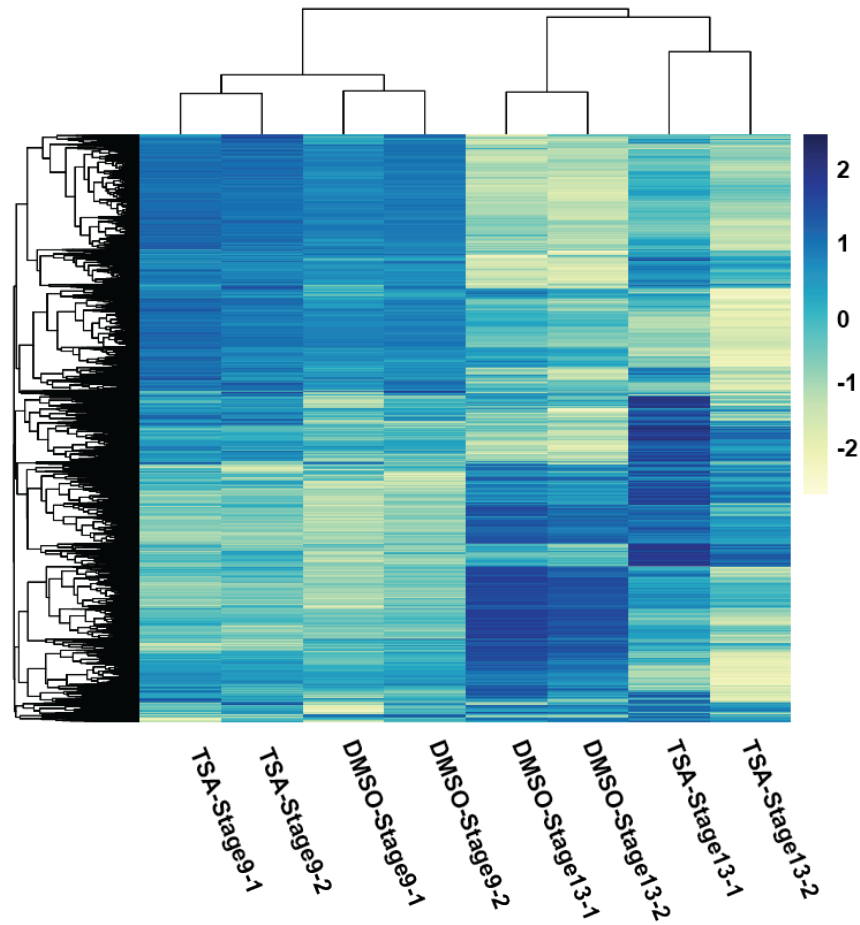


Figure 3.8 Global changes in gene expression after TSA treatment

Heatmap depicting gene expression changes following loss of HDAC activity as cells progress from a pluripotent to a lineage restricted state. TSA treatment results in loss of pluripotency and disrupts normal lineage restriction to epidermal state.

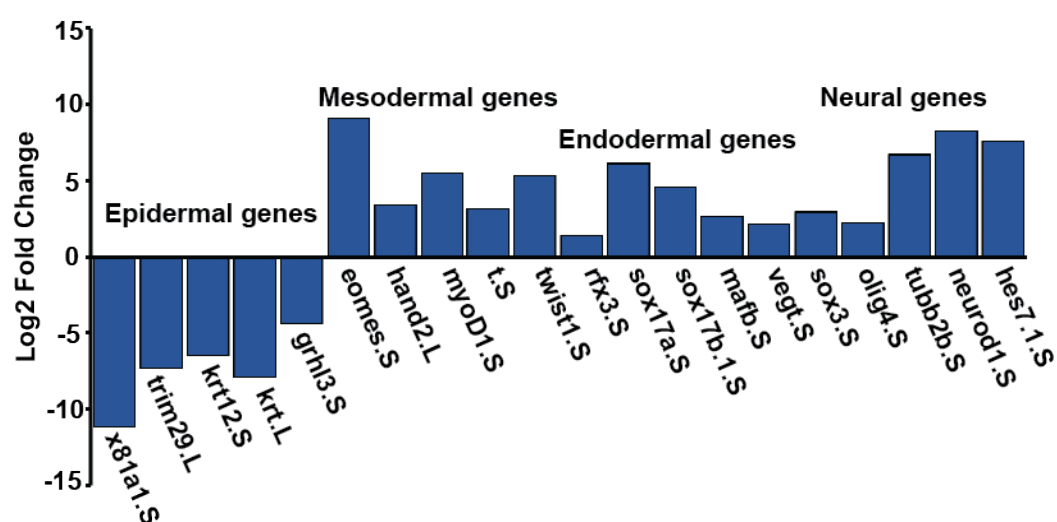


Figure 3.9 Loss of HDAC activity disrupts normal lineage restriction

Graph representing log fold change quantified through differential analysis of RNAsequencing data from Stage 13 animal caps +/-TSA. TSA treatment results in disruption of normal lineage restriction as seen by loss of epidermal expression and simultaneous upregulation of mesodermal, endodermal and neural lineage specific genes.

degrees suggestive that the mechanism through which HDACs are differentially regulating these genes cannot be explained purely through differential binding with partners.

Loss of HDAC activity leads to failure in normal lineage restriction

Given that we had identified that loss of HDAC activity causes genome wide transcriptomic changes in pluripotent cells, we further sought to characterize how loss of HDAC activity affected normal lineage restriction. Dissected animal caps cultured in isolation will naturally lineage restrict and default to an epidermal state (Ariizumi & Asashima, 2001). We had previously identified that TSA treatment resulted in loss of pluripotency of blastula cells and inability of the cells to give rise to any lineage (Chapter 2). Interestingly, global changes in gene expression take place during the process of lineage restriction in animal cap cells after loss of TSA treatment (Figure 3.7, 3.8). 15198 genes are significantly differentially expressed in Stage 13 animal cap explants after loss of HDAC activity. We find that while there are a set of genes that are strongly turned on in the control animal caps defaulting to an epidermal state, TSA treatment results in downregulation of these genes and upregulation of others. While the control animal caps defaulted to an epidermal state as seen by an upregulation of known epidermal markers, TSA treatment resulted in a sustained expression of markers of different lineages and disrupted normal lineage restriction in these cells (Figure 3.9). Taken together, this data suggests that HDACs are required to control expression of lineage genes, and they regulate lineage genes differently than pluripotency genes.

H3K27Ac enrichment is lower at lineage genes when compared to pluripotency genes

Our data suggested that HDACs were differentially regulating expression of lineage genes when compared to pluripotent genes. Histone acetylation is often considered as a proxy for

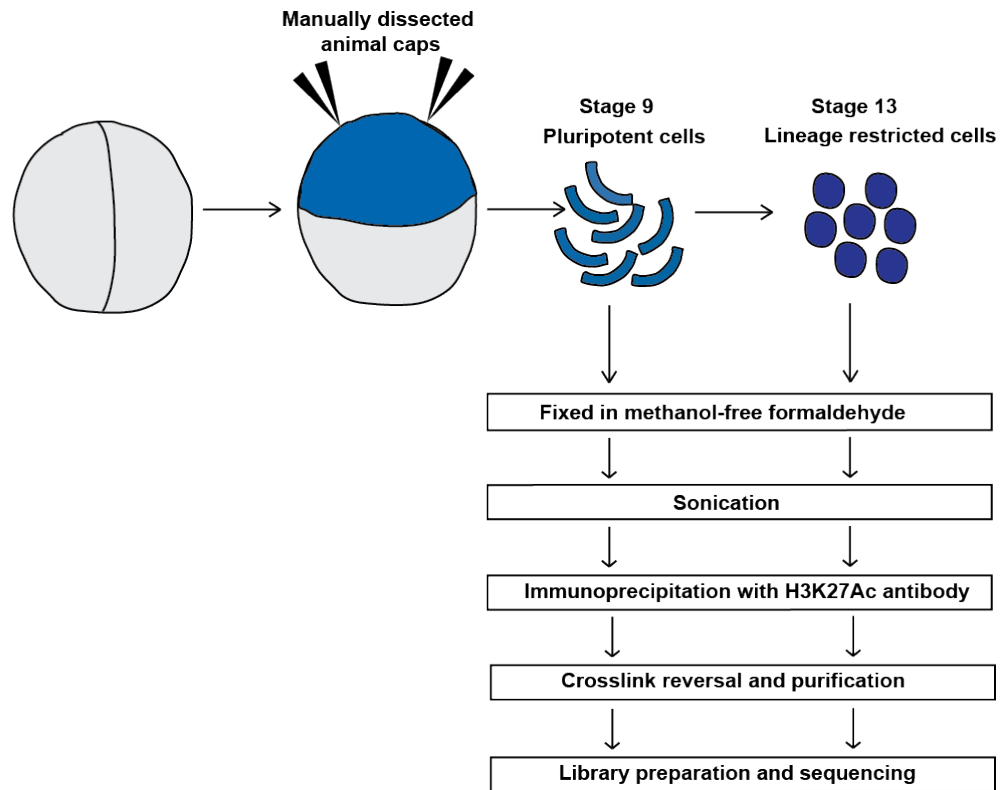


Figure 3.10 ChIP-seq with animal cap explants workflow

Schematic representation of H3K27Ac ChIP-seq of animal cap explants. Animal cap explants were dissected from WT embryos and collected and fixed when sibling embryos reached blastula pluripotent stages (Stage 9) and neurula lineage restricted stages (Stage 13). Purified ChIP-DNA was obtained by immunoprecipitation with H3K27Ac antibody and crosslink reversal and sent for library preparation and sequencing.

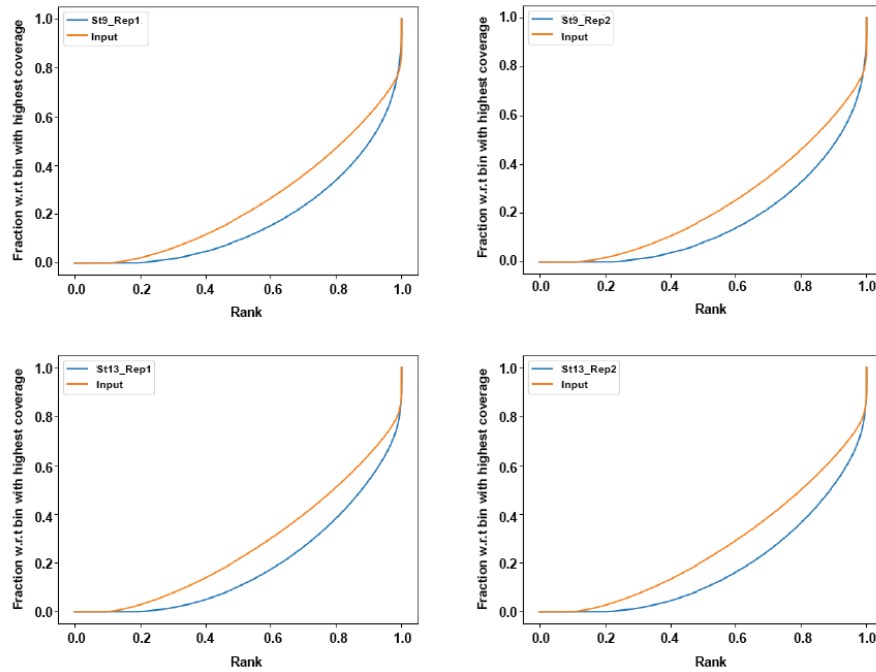


Figure 3.11 Quality control for H3K27Ac ChIP-seq

Fingerprint plots comparing ChIP-seq coverage with input coverage. In all samples, we observe a deviation of ChIP sample from input sample suggesting the pulldown was performed successfully.

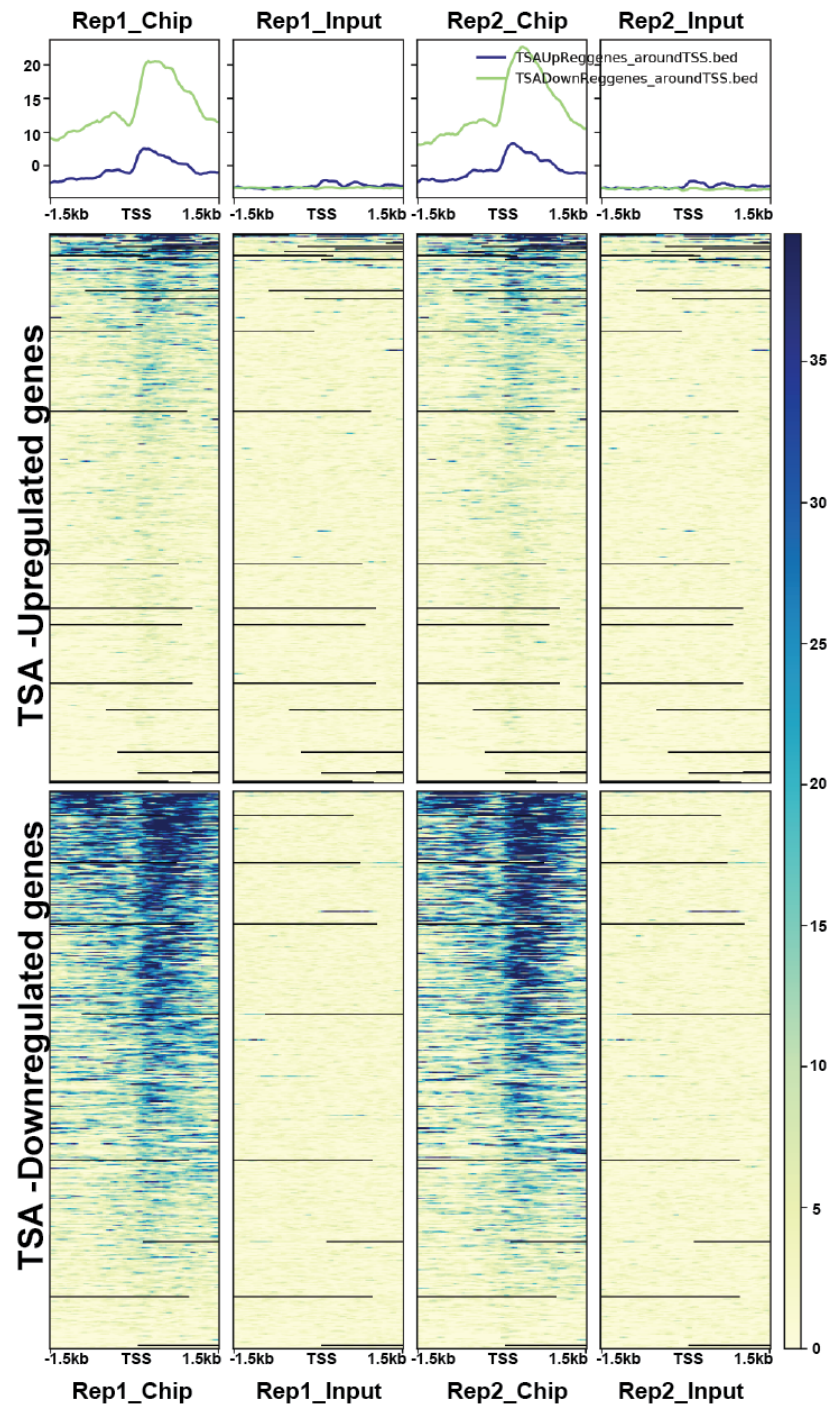


Figure 3.12 H3K27Ac enrichment is higher in genes downregulated after TSA treatment than upregulated by TSA

Heatmap representing H3K27Ac enrichment at genomic loci of genes downregulated after TSA treatment vs genes upregulated after TSA treatment in pluripotent blastula cells (Stage 9). H3K27Ac enrichment is very high at the loci of downregulated genes and low enrichment is observed at the loci of upregulated genes.

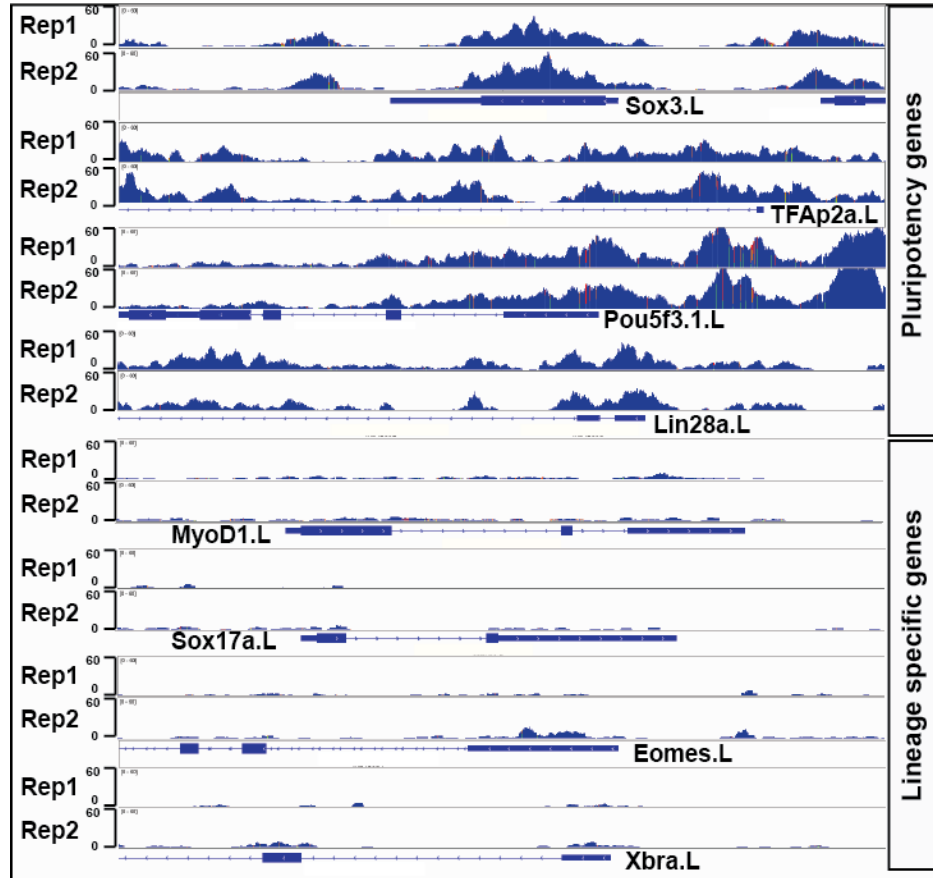


Figure 3.13 H3K27Ac enrichment is higher at pluripotency genes vs lineage genes

ChIP-Seq tracks comparing H3K27Ac enrichment at genomic loci of pluripotency genes vs lineage specific genes in pluripotent blastula cells (Stage 9). H3K27Ac enrichment is very high at the loci of pluripotency genes and low enrichment is observed at the loci of lineage specific markers.

HDAC binding and activity as HDACs do not directly bind to DNA. In order to pinpoint the mechanism through which HDACs are regulating pluripotency in blastula cells, we performed ChIP-Seq to compare H3K27Ac enrichment at genomic loci of pluripotent and lineage genes, as well as at the loci of genes that are upregulated vs downregulated by TSA (Figure 3.10). Quality control was performed on the aligned reads and a plot of the cumulative read coverage for each sample was plotted which showed that ChIPseq signal differed sharply from the background (input) signal suggestive that we had good enrichment of H3K27Ac peaks in these samples (Figure 3.11). A global view of H3K27Ac in all genes shows strong H3K27Ac enrichment in ChIP samples in comparison to input in pluripotent cells, and these peaks are enriched around the TSS of the genes. H3K27Ac is typically thought to mark the enhancer regions, however lack of enhancer architecture and position data prevented us from exploring H3K27Ac enrichment at the enhancers.

Given that TSA treatment strongly upregulated some genes while downregulated others, we were interested in comparing the levels of histone acetylation at the loci of these two gene categories. Interestingly, when we compare the H3K27Ac enrichment at the genomic loci of these genes we find that H3K27Ac is low at the genes that upregulated after loss of HDAC activity, while high enrichment is observed at the loci of genes that are downregulated (Figure 3.12). This suggested to us that HDACs might be functioning differently to regulate expression of these two gene categories and that histone acetylation is low at the loci of lineage specific genes (predominantly upregulated after TSA treatment). Indeed, we find the H3K27Ac enrichment is much higher at the genomic loci of pluripotent genes when compared to lineage specific genes in pluripotent cells (Figure 3.13). We observed similar high enrichment of

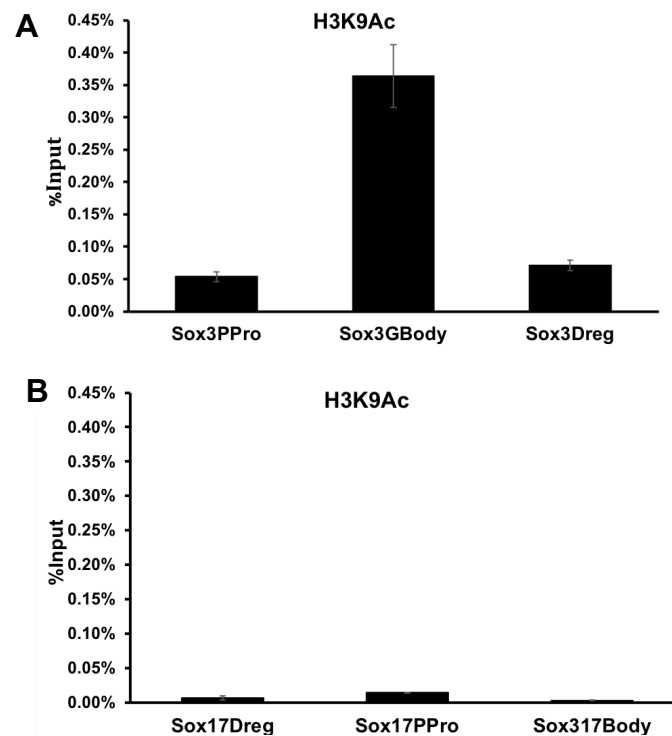


Figure 3.14 H3K9Ac enrichment is higher at pluripotency genes vs lineage genes by ChIP-qPCR

ChIP-qPCR comparing H3K9Ac enrichment at genomic loci of pluripotency gene (Sox3) vs lineage specific gene (Sox17) in pluripotent blastula cells (Stage 9). H3K9Ac enrichment is very high at the loci of Sox3 and low enrichment is observed at the loci of Sox17.

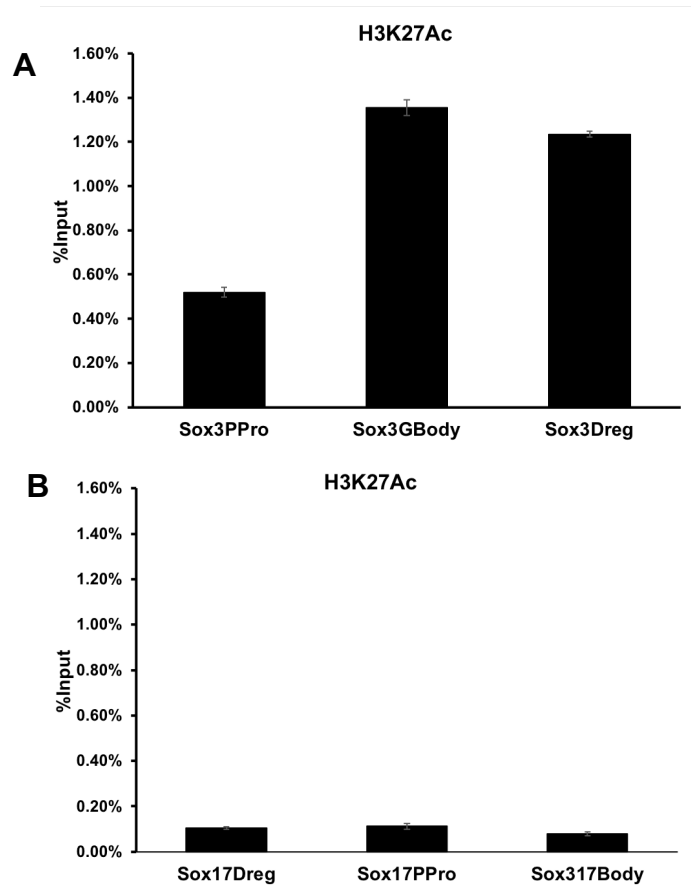


Figure 3.15 H3K27Ac enrichment is higher at pluripotency genes vs lineage genes by ChIP-qPCR

ChIP-qPCR comparing H3K27Ac enrichment at genomic loci of pluripotency gene (Sox3) vs lineage specific gene (Sox17) in pluripotent blastula cells (Stage 9). H3K27Ac enrichment is very high at the loci of Sox3 and low enrichment is observed at the loci of Sox17.

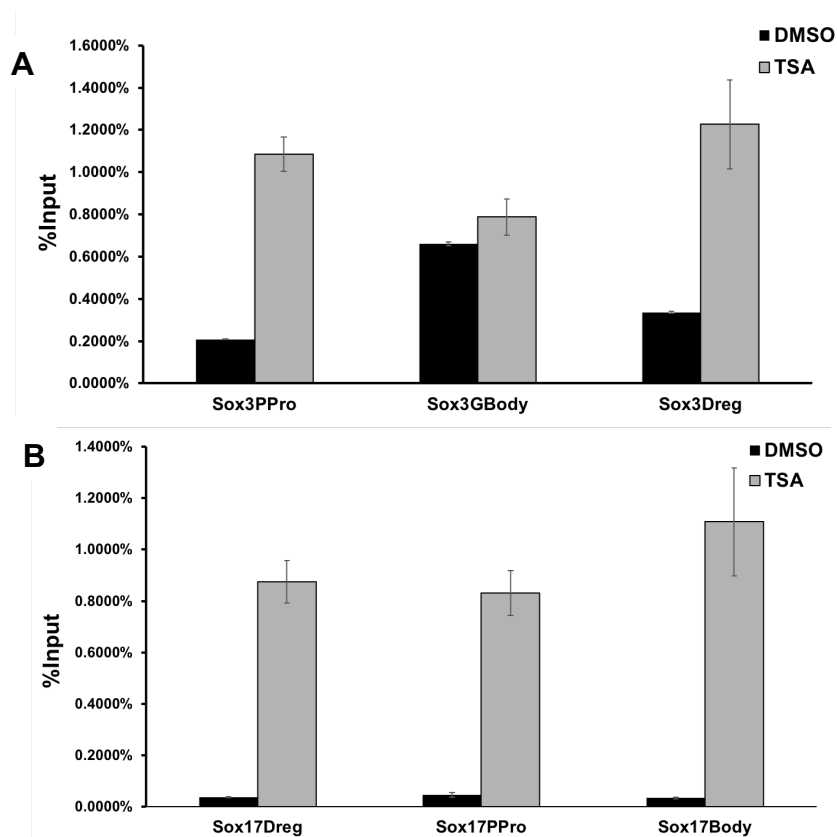


Figure 3.16 H3K27Ac enrichment at pluripotency and lineage genes after TSA treatment

ChIP-qPCR comparing H3K27Ac enrichment at genomic loci of pluripotency gene (Sox3) vs lineage specific gene (Sox17) in pluripotent blastula cells (Stage 9) with/without TSA treatment. H3K27Ac enrichment is strongly increased at the loci of Sox17, while less effects are observed at the loci of Sox3.

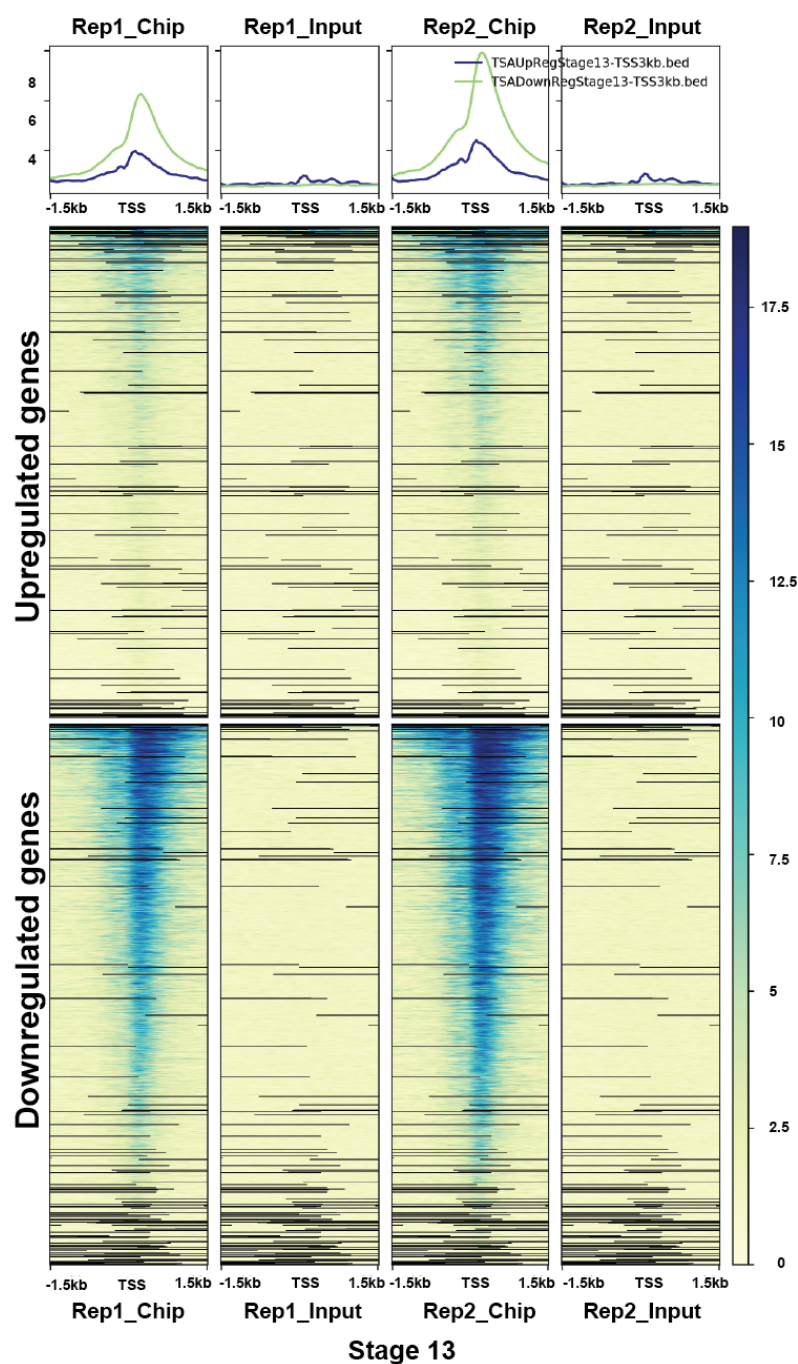


Figure 3.17 Comparing H3K27Ac enrichment at genes up/downregulated after TSA treatment in Stage 13 explants

Heatmap representing H3K27Ac enrichment at genomic loci of genes downregulated after TSA treatment vs genes upregulated after TSA treatment in early neurula cells (Stage 13). H3K27Ac enrichment is very high at the loci of downregulated genes and low enrichment is observed at the loci of upregulated genes.

H3K27Ac and H3K9Ac at the genomic loci of Sox3 (pluripotency gene) via ChIP-qPCR, while low enrichment was measured at the loci of Sox17 (lineage gene) (Figure 3.14, 3.15). This suggests that HDACs are preferentially acting at the genomic loci of lineage genes by maintaining histone acetylation low at these genomic loci. To test our hypothesis, we performed ChIP-qPCR with or without TSA treatment and looked to see how H3K27Ac enrichment changed at the genomic loci of pluripotency vs lineage genes. Interestingly, we find that TSA treatment results in a dramatic increase in H3K27Ac enrichment at the loci of Sox17 (lineage gene) when compared to the change seen in Sox3 (Figure 3.16). This suggests that HDACs function to maintain the pluripotency of these cells by preventing/controlling the expression of lineage genes till instructive cues/signals are received to de-repress the expression of the factors. This data also gave us insights that low levels of histone acetylation are present in these pluripotent cells.

BET protein readers regulate pluripotency through different mechanisms than HDACs

Our data suggested that low levels of histone acetylation are present in these pluripotent cells and we sought to identify the importance of maintaining the appropriate level of histone acetylation. To this end, we decided to employ a chemical inhibitor, IBET, that blocks the activity of the BET family of proteins which are bona-fide readers of histone acetylation. Bromodomain proteins specifically recognize histone acetylation on lysine residues and regulate the onset of transcription (Marmorstein & Zhou, 2014). Interestingly, IBET-mediated inhibition of BET protein activity resulted in loss of expression of pluripotency genes in blastula embryos (Figure 3.18A) (Credit: Paul Huber). Further, loss of BET protein activity results in a loss of neural crest formation (Figure 3.18B) (Credit: Paul Huber). This was a very intriguing

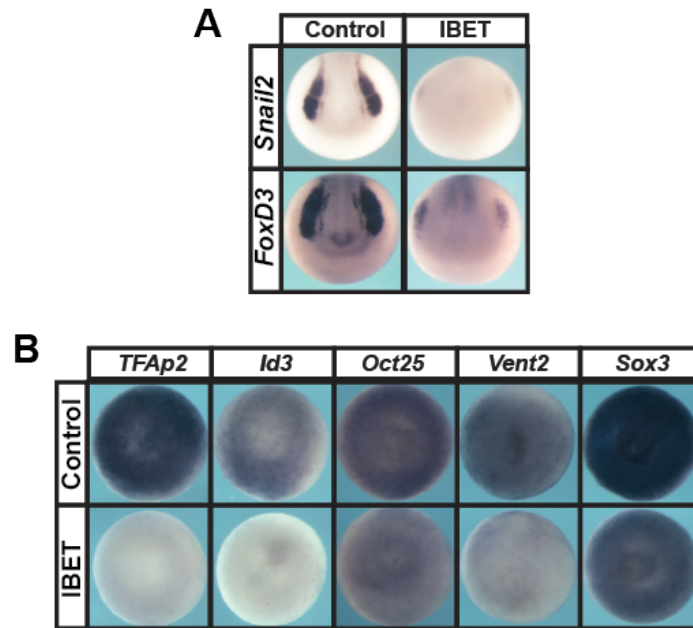


Figure 3.18 Loss of BET activity leads to inhibition of pluripotency gene expression and loss of neural crest formation

(A) *In situ* hybridization examining *Snail2* and *FoxD3* expression in neurula stage (stage 15) embryos treated with vehicle control (DMSO or water) or IBET (250 μ M). Loss of BET activity leads to loss of expression of *Snail2* and *FoxD3*. (B) *In situ* hybridization examining *TFAp2*, *Id3*, *Oct25*, *Vent2* and *Sox3* expression in blastula stage (stage 9) embryos treated with vehicle control (DMSO or water) or IBET (250 μ M). Loss of BET activity leads to loss of expression of *Snail2* and *FoxD3*. Loss of BET activity leads to loss of pluripotency gene expression. Embryos were treated at 2-cell stage and grown until blastula (Stage 9) or mid-neurula stages (Stage 15).

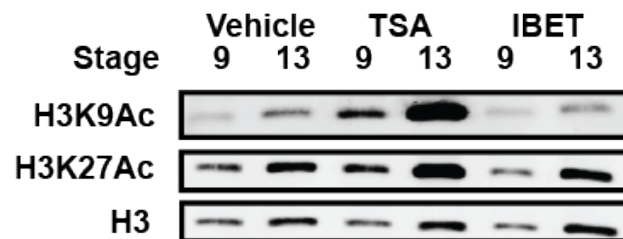


Figure 3.19 Loss of BET protein activity does not affect histone acetylation

Western blot analysis of aging animal caps treated with vehicle (DMSO), or HDAC inhibitor (TSA-500nM) or BET inhibitor (IBET-250 μ M) examining H3K9Ac and H3K27Ac alongside total H3 levels via chemiluminescence. H3K9Ac/H3K27Ac is increased dramatically by TSA while IBET treatment does not affect histone acetylation. Explants were cultured alongside sibling embryos until blastula (Stage 9) and neural plate stages (Stage 13).

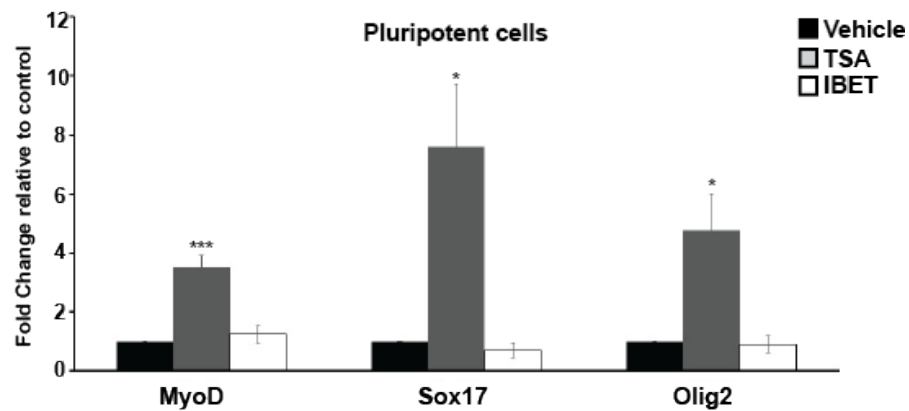


Figure 3.20 Loss of BET protein activity does not cause upregulation of lineage genes like loss of HDAC activity

Quantitative RT-PCR examining lineage marker gene expression in aging animal cap explants treated with vehicle (DMSO) or HDAC inhibitor (TSA -500nM) or BET inhibitor (IBET-250μM). TSA treatment results in precocious expression of markers of various lineages such as *MyoD*, *Sox17* and *Olig2* but are unaffected by IBET treatment. Explants were cultured alongside sibling embryos grown from blastula (Stage 9) until late-gastrula stages (Stage 13). (*P<0.05, ***P<0.005)

and unexpected finding as the increase of histone acetylation as seen through TSA treatment resulted in the same phenotype as the loss in the ability of a cell to read existing histone acetylation. We were hence interested in identifying the mechanism through which BRD proteins and low levels of histone acetylation are important for the pluripotency of these cells. We first wanted to test to see if there was any feedback loop that existed within the system, and if IBET treatment was causing a spurious increase in histone acetylation. We find that while TSA treatment resulted in a dramatic increase in both H3K9Ac and H3K27Ac in pluripotent cells, IBET treatment did not cause any change in the levels of histone acetylation (Figure 3.19). Our previous studies had identified that TSA treatment results in spurious expression of markers of all lineage simultaneously and that HDACs control pluripotency by preventing expression of these lineage markers, we next asked if BET proteins regulate pluripotency through a similar mechanism to HDACs. Interestingly, using quantitative RT-PCR we found that while TSA treatment caused dramatic upregulation of *Olig2*, *Sox17* and *MyoD* expression in pluripotent cells, IBET had no effect on the expression of these lineage specific genes (Figure 3.20). This would suggest that HDACs and BET proteins regulate pluripotency through different mechanisms.

BRD proteins use distinct mechanisms to regulate pluripotency

In order to characterize the mechanisms through which BET proteins regulate pluripotency, we took a transcriptomic approach and compared the changes in gene expression in response of TSA and IBET treatment in pluripotent cells. Both TSA and IBET treatment results in global changes in gene expression with 920 genes significantly differentially regulated after TSA treatment and 957 genes differentially regulated after IBET treatment. Interestingly, we find

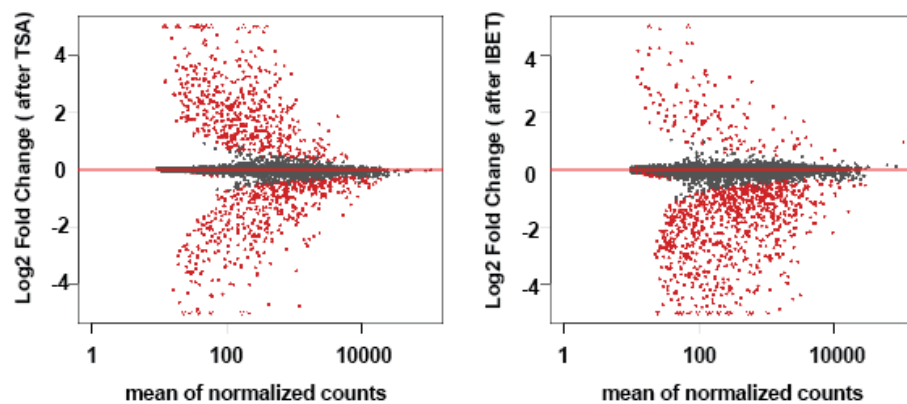


Figure 3.21 HDAC and BET proteins regulate pluripotency through distinct mechanisms

MA-plot depicting gene expression changes in pluripotent cells after TSA treatment (A) and IBET treatment (B). Loss of HDAC activity results equal upregulation vs downregulation of gene expression while loss of BET activity predominantly results in down regulation of gene expression. 957 genes (~2%) are significantly differentially expressed after IBET treatment.

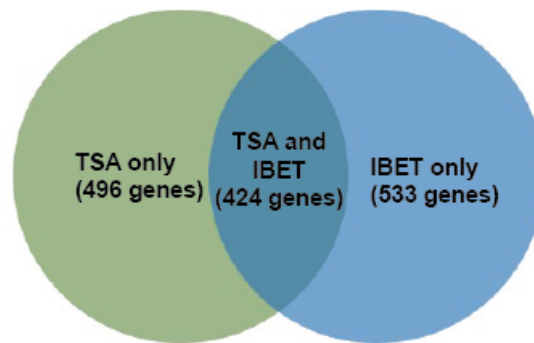


Figure 3.22 Comparing gene expression changes after loss of HDAC and BET activity

Venn diagram depicting genes that are affected by both TSA and IBET treatment or only by one or the other. ~50% of the genes are co-regulated by both HDAC and BET activity.

Category A	TSA and IBET (DOWN)
Category B	TSA and IBET (UP)
Category C	TSA(UP),IBET (DOWN)

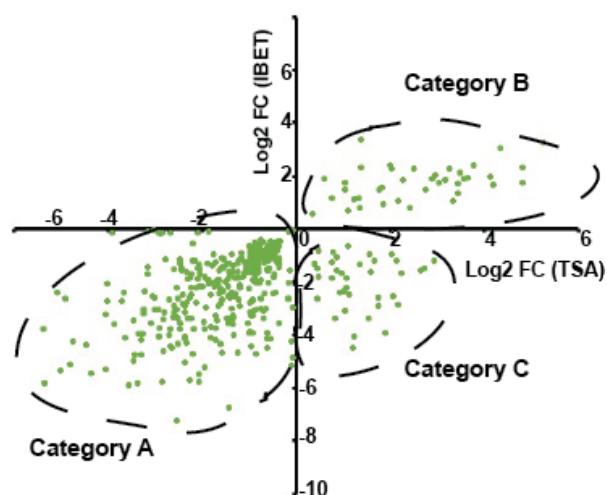


Figure 3.23 Examining genes that are co-regulated by HDAC and BET activity

Graph depicting categories of genes that are affected by both TSA and IBET treatment. Category A represents genes are downregulated by both TSA and IBET, while Category B consists of genes that upregulated by both TSA and IBET. Category C includes genes that upregulated by TSA and down-regulated by IBET.

Category A (TSA,IBET - negative FC)				Category B (TSA,IBET - postive FC)		Category C (TSA -positive FC/IBET-negative FC)	
cyp26a1.S	pou5f3.1.S	msl2.S	LOC108715183	mafb.S		kif4.L	
LOC108715370	smad4.1.L	LOC108700186	ppp1r15b-like.S	myod1.L		plk2.L	
prdm1.S	LOC108709393	fosl2.L	LOC398207.S	ccndx.S		not.S	
lhx5.S	pkn1.L	sf3b5.S	LOC108712563	snai1.S		sox17b.2.L	
lhx5.L	foxj1.L	mat2a.S	ggt1.L	LOC108719460		LOC108701190	
sfrp2.S	jag1.L	mansc1.S	LOC108696067	cldn5.S		LOC108715571	
tfap2a.L	lin28a.S	foxi4.1.L	Xetrov90029963m.L	cldn5.L		lmo4.2.S	
gadm.L	acvr1b.L	fbxl14.L		LOC108709229		lefty.L	
LOC100490918.L	e2f7.L	midn.S		wnt9b.S		flrt3.L	
arg1.L	aldh111.L	ccnf.L		cebpa.S		LOC108714950	
cxcl12.L	tead4.S	LOC108712829		h2afj.L		aplnr.L	
tfap2a.S	LOC108698541	grhl3.L	LOC108716428	myc.L		kif17.L	
dlc1.L	dupp6.S	krt12.S	LOC108717283	mafb.L		rasd1.S	
ptgs2.S	hes3.1.S	cdc25b.S	sp5l.L	LOC108712526		not.L	
prdm14.S	trim29.L	sox3.S	parp2.L	sox9.L		rasl11b.L	
sds.L	LOC108720009	lpar6.L	spg21.S	cebpa.L		LOC108719558	
akap12.L	plekhg3.L	fsbp-like.L	eif5.L	txnip.S		aplnr.S	
dlx5.S	hdgf.L	Xelaev18039004m	dlx5.L	sox17b.1.S		mdm2.L	
fhdc1.L	socs3.L	pnhd.S	Xelaev18001619m	h2afj.S		c3ar1.L	
LOC108704959	tfap2c.S	LOC108718087	kctd15.L	egr1.L		cxcr4.S	
has2.L	tgif2.S	rftg.L	LOC108719533	jun.L		LOC108718345	
ckap4.S	LOC108711252	e2f3.L	LOC108716155	txnip.L		Xelaev18005230m	
sox21.L	LOC108719547	tmem45b.L	rbm47.S	ptf1a		tiparp.S	
htra1.S	trim29.S	tsc22d1.L	LOC108702843	LOC108703754		foxi1.L	
plekhn1.L	irg1.L	ephb3.L	Xelaev18036963m	LOC108703185		hes5.2.L	
nedd9.L	lpar4.L	LOC108703190	sbno1.L	jund.S		ccng1.L	
prickle1.L	LOC108708269	top2b.L	tdgf1.3.L	myc.S		LOC108699852	
fgfr4.S	LOC108700085	MGC81970	LOC108700239	LOC100192369		nudt22.L	
xk81a1.L	rbm5.L	krt.L	LOC100158389	LOC108711588		fam212a.L	
LOC108708488	fbxl14.S	sox11.L	LOC108708501	LOC108695465		endod1.S	
bmp4.L	LOC108699970	LOC108700915	irs4.L	cebpd.L		LOC108696734	
LOC108701135	sema3f.L	sox11.S	rbm5.S	flil.L		LOC108695555	
sall4.L	stx19.L	cxcl11.L	LOC108720027	egr1.S		metnl.L	
dlx6.S	hes3.1.L	fzd4.L	zfp36.S	rhub.S		pvr11-like.2.L	
irx2.L	smad7.L	foxd4l1.2.L	LOC108705666	Xetrov90018873m.S		hs3st1.L	
fam174b.L	LOC108698345	hal.1.L	LOC108704510	LOC108716539		relt.L	
LOC414675	gpr4.S	elf1.L	LOC108698063	akap6-like.S		tuba4a.S	
LOC108714480	foxd4l1.1.S	gprc5c.L	dusp1.S	fam83d.L		LOC108695565	
dhrr3.S	LOC108711587	meis3.L	sp5l.S				
tcf7l1.L	LOC108699303	cdt1.S	LOC108696199				
ptgs2.L	tcf7l1.S	ccnt2.L	cdk9.L				
irx1.S	gas1.S	srsf10.L	fbxo34.S				
xarp.L	dennd2c.L	LOC108713138	rhebl1.S				
bmp4.S	s1pr5.L	emp2.S	srsf5.L				
nedd9.S	LOC108698659	spry2.L	LOC108704931				
LOC108711414	irx1.L	dlx3.L	spry2.S				
hes6.1.L	meis3.S	slc35a2.S	xbp1.S				
LOC108696405	LOC108699606	foxi4.2.S	sbk1.L				
rhou.L	chst2.L	LOC108719479	lrwd1.L				
LOC108695360	rbmx.L	cdh1.S	dido1.L				
LOC108696018	mst1.L	eif5.S	Xelaev18017809m				
g2e3.L	LOC108715729	skil.S	Xelaev18035319m				
znf238.2.L	foxj1.S	LOC108720018	LOC108702226				
grhl3.S	Xelaev18029313m	Xelaev18022784m	slc30a2.L				
pcdh7.L	epha4.L	mx3.S	cdc14b.S				
chst7.S	hic2.L	ccnj.S	tbcel.L				
LOC495419	rsad2.L	sema3f.S	ftth1a				
ccne2.L	LOC108710862	fut1.S	foxh1.2.L				
akap2.S	gpr65.S	mex3b.L	LOC108713085				
apln.L	LOC108696887	arl13a.L	gpm2.S				
ptafr.L	apold1.L	zc3h10.S	rfa4.L				
prdm1.L	LOC100158340	Xelaev18013159m	znf750.L				
LOC108703074	LOC108717145	LOC108702469	rarg.S				
sox21.S	Xetrov90004873m.L	LOC108699980	crx.L				
rgcc.L	LOC108701005	fbxo5.L	pou5f3.1.L				
lpar6.S	LOC108718188	esrp1.S	alkbh5.L				
socs3	Xelaev18018260m	cbarp.L	coq10b.L				
tuba1a.S	kif2.L	LOC108701004	tp53inp1.L				
entpd2.S	LOC108695976	fgfr4.L	upk1b.L				
zic1.L	btg1.L	LOC108713056	robo3.S				
arl4c.S	prss36.L	ackr3.S	c6orf62.S				
zmcm3.L	chst7.L	hnmpa1.S	LOC108711338				
LOC108699863	admp2.L	LOC108698547	MGC81393				
LOC108710268	gnb3.L	mid1ip1.L	ppp1r15b.L				
sertad3.L	pnrc2.S	LOC108704926	LOC108711731				
LOC108710290	zfp36l2.L	sat1.L	gas211.L				
a2m.S	kiaa1324l.L	LOC108701003	wdr43.L				
sox3.L	admp.S	sall4b	eomes.L				
sall1.L	Xelaev18000483m	LOC108698352	xbp1.L				
LOC108708416	amot2.S	LOC108702925	rhebl1.L				
akap2.L	LOC108713024	rmf222.S					
LOC108704161	ventx1.2.S	c15orf39.L					
lin28a.L	ccnt2.S	rnd1.S					
usp43.L	sox2.L	prph.L					
LOC108714525	LOC108698564	LOC108696061					
sh3bp4.L	Xelaev18004347m	LOC108705686					
pnrc2.L	eef1a1.L	srsf7.L					
tgif2.L	rbmx.S	isg2012.L					
tnfrsf101.L	LOC108715416	kif2.S					

Table 3.1 Genes that are significantly changed by TSA and IBET

that while TSA treatment leads to equal upregulation vs downregulation of genes, IBET predominantly downregulates genes expression in these pluripotent cells (Figure 3.21). We next wondered how many genes are coregulated by both TSA and IBET (Figure 3.22). We find that 424 genes are significantly altered by both TSA and IBET. These genes fall into 3 categories: Ones that are both downregulated by both TSA and IBET (Category A) – pluripotency genes like *Oct91*, *Sox2* and *TFAp2a* fall in this category. Category B has genes that are upregulated by both TSA and IBET, and Category C has genes that upregulated by TSA but downregulated by IBET (Figure 3.23). This suggests that while there are some genes that are altered by TSA and IBET in a similar manner, this might be happening through different mechanisms. Indeed, we find that over half of the total genes mis-regulated by IBET treatment (533 genes of a total 957 genes) are unchanged after TSA treatment (Figure 3.22). Interestingly, a large percentage of those genes are downregulated, and some genes are upregulated as depicted in the heatmap, as well shown via quantitative RT-PCR (Figure 3.24, 3.25). This suggested that BET proteins regulate the pluripotency of blastula cells through distinct mechanisms than HDACs. We hypothesized that the BET proteins might be regulating gene expression during zygotic genome activation occurring at the mid-blastula transition (MBT). To test this hypothesis, we took advantage of data available on Xenbase to curate a list of 3086 genes that were zygotically transcribed at MBT (Session et al., 2016). Of these 3086 genes, 249 genes (8%) were significantly altered by IBET treatment (Figure 3.26A). This would suggest that BET proteins do not globally control transcriptional activation during MBT, but instead are performing a very specific role and regulating only a subset of these genes. We looked more closely at the function of the genes that BET proteins regulate at MBT and find that these genes are predominantly transcription factors and signaling molecules, factors that are critical during these stages of development (Figure

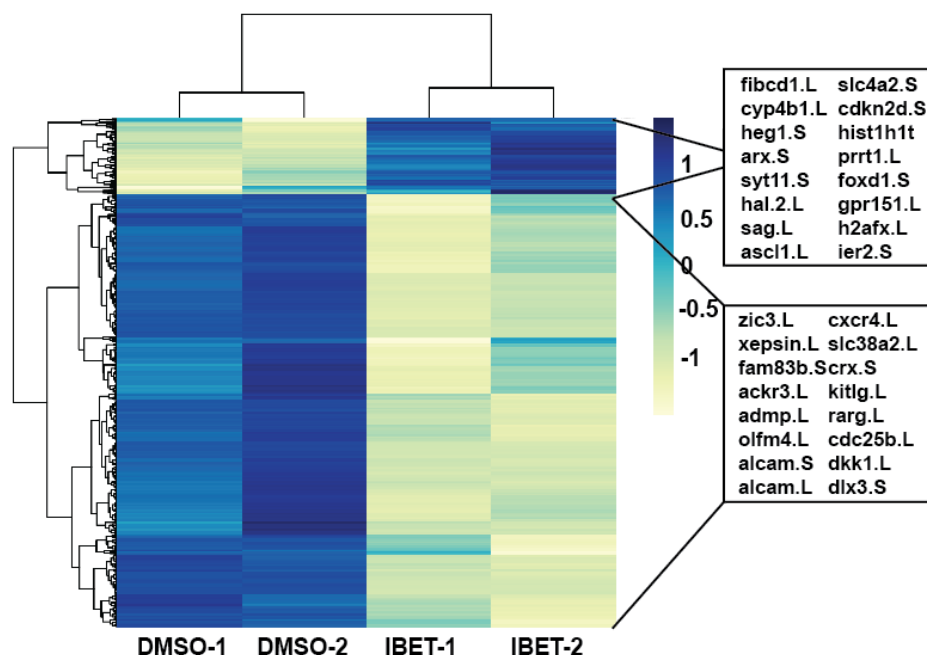


Figure 3.24 Genes affected specifically after loss of BET protein activity

Heatmap depicting genes that are significantly differentially expressed after loss of BET protein activity in pluripotent cells. The predominant effect seen after IBET treatment is downregulation of gene expression.

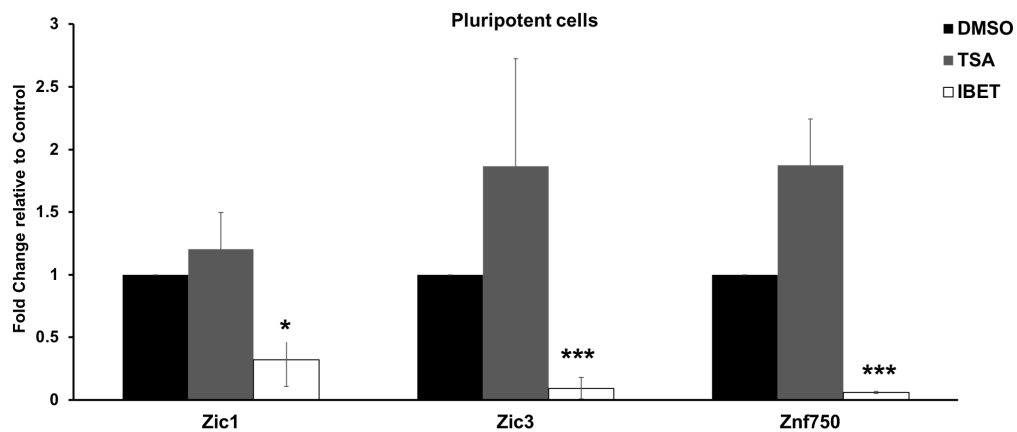


Figure 3.25 BET proteins regulate several genes not affected by TSA treatment

Quantitative RT-PCR examining gene expression in aging animal cap explants treated with vehicle (DMSO) or HDAC inhibitor (TSA -500nM) or BET inhibitor (IBET-250 μ M). IBET treatment results in downregulation of *Zic1*, *Zic3* and *Znf750* but are unaffected by TSA treatment. Explants were cultured alongside sibling embryos grown until blastula stages (Stage 9) (* $P < 0.05$, *** $P < 0.005$)

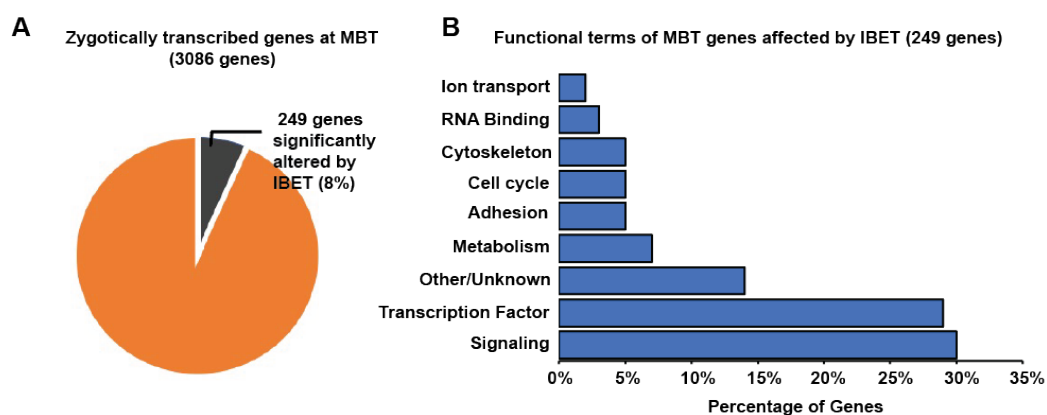


Figure 3.26 Mechanism of BET activity regulating zygotic gene transcription

(A) Pie chart depicting that IBET affects only 8% of the genes that are zygotically transcribed at MBT. (B) Functional terms for zygotically transcribed MBT genes affected by IBET show enrichment for signaling molecules and transcription factors.





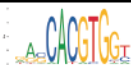
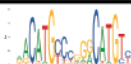
Motif	Transcription factor	Percentage of genes
	FoxD3	91% (222/244)
	TFAP2A	89% (216/244)
	CEBP-A	82% (201/244)
	YY1	81% (198/244)
	Max/Myc	77% (188/244)
	p53	67% (164/244)

Figure 3.27 Motif analysis of promoters of genes altered by IBET at MBT

Table depicting motif analysis of promoters of genes altered by IBET at MBT. Enrichment is observed for transcription factor binding of known factors that have been previously characterized to be involved with BET proteins and have known functions in stem cell maintenance and neural crest formation.

3.26B). We next sought to characterize how BET proteins could be regulating a specific subset of genes at this stage. One hypothesis is that BET proteins are recruited by different transcription factors to the genomic loci of these genes. We looked at the enrichment of transcription factor binding motifs in the promoter regions of these genes, and we find motifs for several known factors involved in pluripotency such TFAP2A, FoxD3, Myc/Max which have previously been shown to be interacting partners of BRD4 (Figure 3.27). This would suggest that BET proteins might be functioning to regulate a specific subset of genes by using context specific complexes to regulate function in pluripotent cells during embryonic development.

Global changes in histone modifications during lineage restriction

Given that we had identified that histone acetylation is critical for maintaining the pluripotent state, we sought to identify other histone post translational modifications (PTM) that might play an important role in stem cell maintenance and lineage restriction. To this end, we carried out a high throughput mass spectrometry experiment utilizing a technique known as epiproteomics to quantify the abundance of different histone modifications during embryonic development (Figure 2.28) (Work done in collaboration with Jeannie Carmillo, Northwestern Proteomics Core). Utilizing this technique, we were able to obtain the abundance of 85 different histone modifications during 6 developmental timepoints, from blastula to early neurula stages (Figure 2.28). 85 histone PTMs were successfully detected and quantified in all samples including marks such as H3K4me1/2/3, H3K9Ac, H3.1/3K27me1/2/3, H3.1/3K27Ac.

The most significant observation in this study is that we find that dynamic changes take place in the abundance of histone modifications during embryonic development. The quantitative

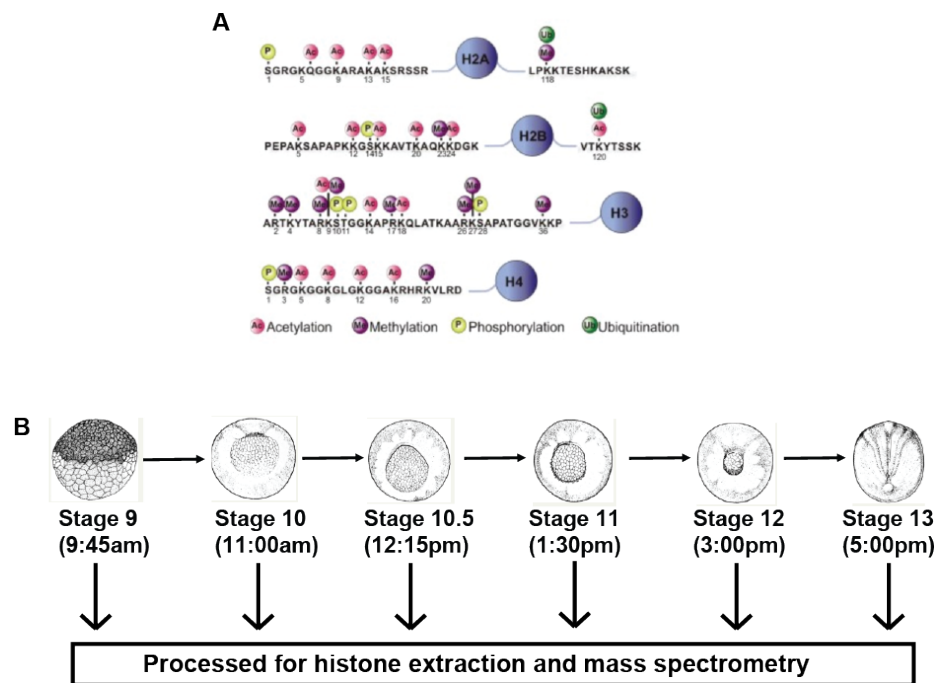


Figure 3.28 Epiproteomic analysis of early *Xenopus* development
 (A) Diagrammatic representation of histone modification assayed via epiproteomics mass spectrometry methodology (B) Schematic representation of epiproteomic experiment in early *Xenopus* embryos. Embryos were collected according to a pre-determined time and temperature regimen.

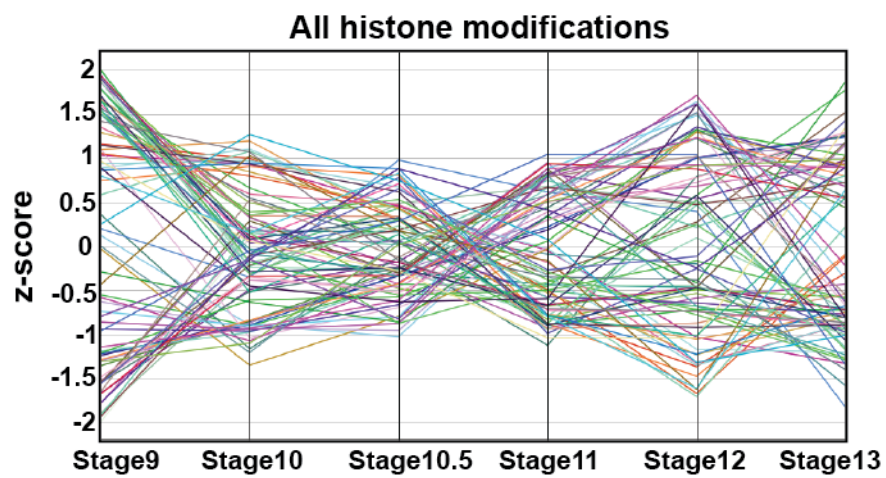


Figure 3.29 Dynamic changes in histone modifications over developmental time
Parallel coordinates plot depicting changes in all modifications over time. Dynamic changes are observed in the abundance of the histone modifications from blastula to neurula stages.

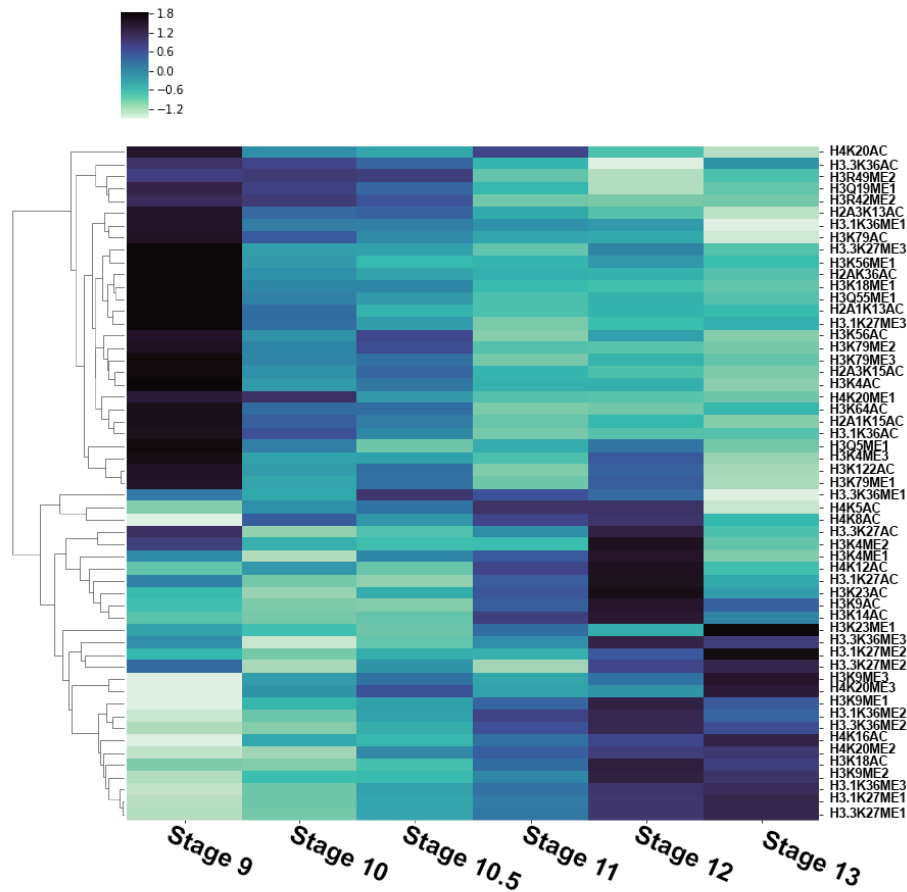


Figure 3.30 Epigenomic signature of early embryonic development

Heatmap depicting hierarchical clustering of abundance of histone modification during various stages of early development. Several modifications are high in the pluripotent blastula cells while others increase during lineage restriction and are high at neurula stages.

dataset allowed us to perform head to head comparisons of the abundance of histone PTMs during these early stages of *Xenopus* development. The data was obtained as a percentage abundance of all detected histone modifications. From this, we calculated average abundance across three biological replicates, and plotted parallel coordinate plots and heatmap from the data using Python (Refer to Appendix 2 for Python scripts for analysis). Indeed, when we trace the changes in all detectable modifications over time, we find that the abundance of histone PTMs change both monotonically and non-monotonically as the embryo gradually gets lineage restricted (Figure 3.29). This would suggest that there are histone modifications that are highest in the early embryo, while others that increase as embryo begins lineage restriction. Interestingly, when we perform hierarchical clustering on the data and plot it as heatmap, we find several modifications that highest in the pluripotent cells and drop off dramatically as the embryo begins gastrulation e.g. H4K20Ac, H4K20Me1, H3.3K36Ac, H3.3K27ME3, H3K79ME2/3, H3K79Ac (Figure 3.30, 3.32). H3.3K27ME3, in particular, is very interesting because it is a repressive mark that had previously been shown and thought to increase with lineage restriction (Bogdanovic, van Heeringen, & Veenstra, 2012b; Schneider et al., 2011). However, our data would suggest that earlier in development, the H3.3K27ME3 might actually be high when lineage specific genes are kept turned off and be removed as the embryo undergoes gastrulation and starts to express lineage specific genes. However, later in development, it probably increases as early developmental genes are turned off. We also find that there are a set of histone modifications that are low in the early embryo and seem to increase as lineage restriction proceeds such as H3K18AC, H3.1K27ME1, H3.1K27ME3, H4K16AC, H3K9ME1 (Figure 3.30, 3.31). There are also marks that changes dynamically over time such as H3K4ME1 and H3K79ME1 (Figure 3.33). Interestingly, we find that similar to what we had previously seen,

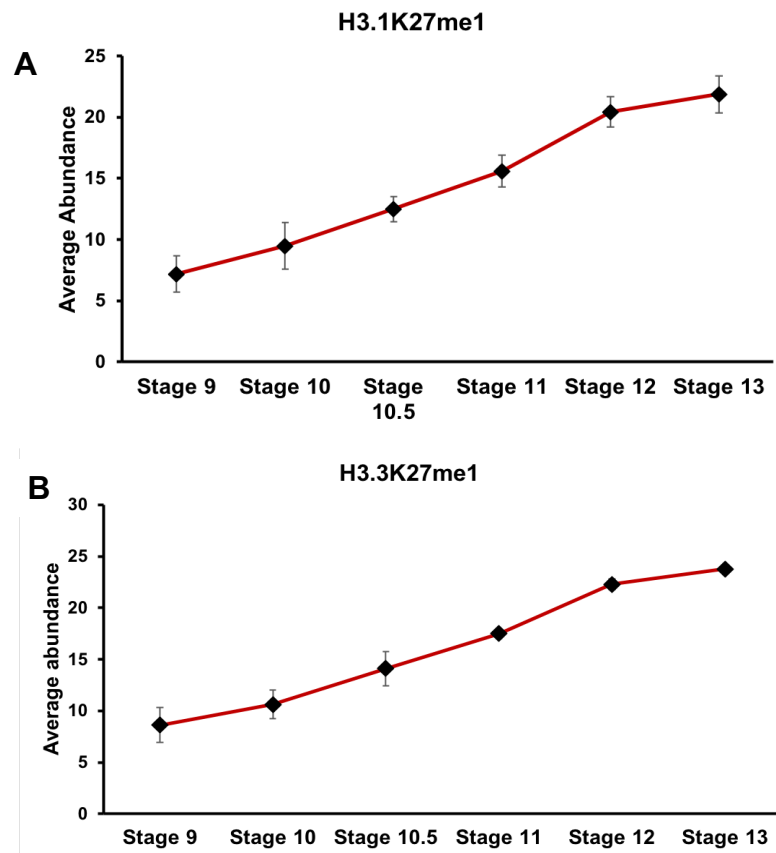


Figure 3.31 Histone modifications that increase during lineage restriction
 (A), (B) Graph depicting changes in average abundance of H3.1K27Me1 and H3.3K27Me1 over developmental time. These marks have low abundance in blastula embryos and increases over developmental time.

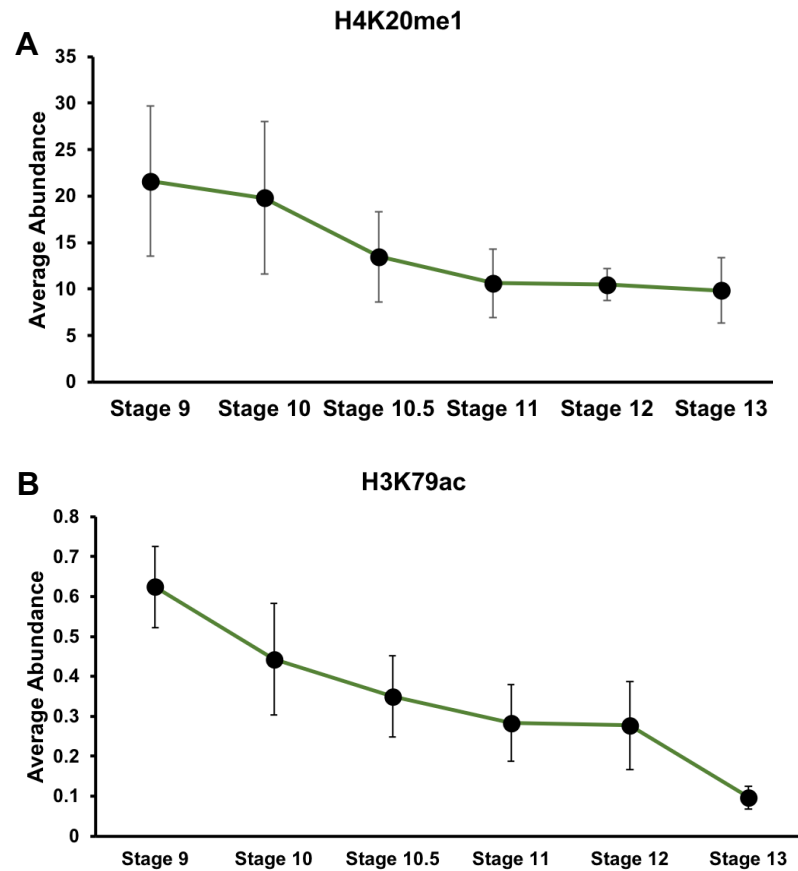


Figure 3.32 Histone modifications that decrease during lineage restriction

(A), (B) Graph depicting changes in average abundance of H4K20Me1 and H3K79Ac over developmental time. These marks have low abundance in blastula embryos and increases over developmental time.

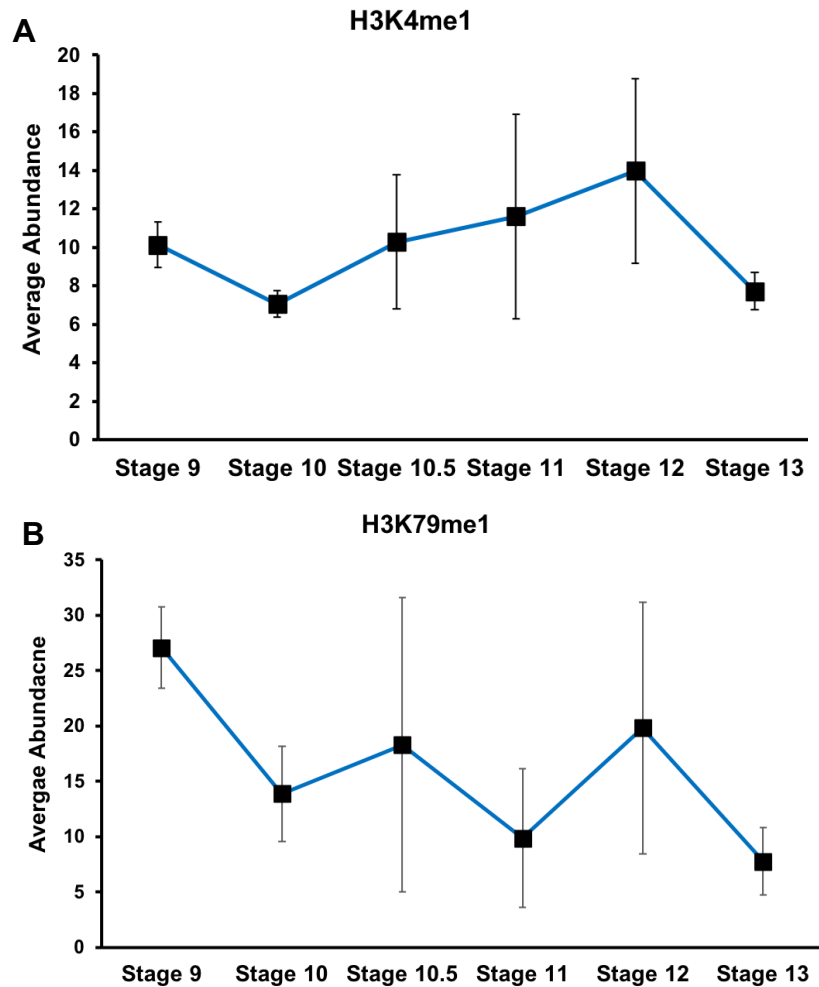


Figure 3.33 Histone modifications that change variably during lineage restriction

(A), (B) Graph depicting changes in average abundance of H3K4Me1 and H3K79me1 over developmental time. These marks change their abundance non-monotonically over developmental time.

H3.1/3K27AC and H3K9AC are low at Stage 9 and increase during lineage restriction. However, we find that in this whole embryo dataset that the peak expression of these marks is seen at Stage 12 while we had previously characterized this to be the case at Stage 13 in animal cap explants. This suggests that there might be cell type specific differences in histone PTM abundance. Indeed, we find that neural crest cells retain low levels of both H3K9AC and H3K27Ac. It would be interesting to further characterize the abundance of histone modifications in different cell fates. Thus, a concerted histone signature seems to exist that defines the early embryo, and another which describes the lineage restricted state. These studies describe histone PTMs that we had previously not known were involved in stem cell maintenance and would be interesting to follow up on.

Discussion

Epigenetic regulation is a critical component of the regulatory circuit that controls stem cell maintenance. Histone post translational modifications (PTMs) and chromatin remodelers tasked with writing, erasing and reading these modifications have been found to be critical for stem cell maintenance. These epigenetic factors have also been implicated to have vital roles during embryonic development. Indeed, loss of critical chromatin remodelers like members of the PRC2 complex, HDACs and others have been shown to be embryonically lethal (Dovey et al., 2010) (Pasini, Bracken, Jensen, Lazzerini Denchi, & Helin, 2004) (O'Carroll et al., 2001). This suggests that chromatin remodelers play important role in stem cell maintenance during embryonic development. Interestingly, we recently uncovered a novel mechanism for epigenetic regulation of pluripotency of neural crest cells, a unique embryonic stem cell population that has the capability to retain its stem cell attributes. We found that HDAC activity was critical for

neural crest formation and pluripotent blastula cells and HDACs function to keep histone acetylation low in both these cells types. Here, we extend our previous work to identify the mechanisms through which HDACs function to maintain pluripotency in blastula cells. Using RNASeq and ChIPSeq, we find that HDACs are regulating the expression of lineage specific genes in these cells preventing their expression until the developmental cues are received to allow transcriptional activation. Further, we characterize that a low level of histone acetylation is critical for maintaining pluripotency, and BRD proteins and HDACs function through distinct mechanisms to regulate pluripotency. This suggests that the regulation of the precise levels of histone acetylation is critical for maintaining the stem cell state. Finally, we utilized mass spectrometry to identify histone PTMs that might play important roles in the maintenance of pluripotency and lineage restriction. This work is of high significance as it furthers our understanding of the mechanisms that regulate pluripotency during embryonic development and gives us novel insights into the epigenetic control of stem cell maintenance.

Our findings that HDAC inhibition causes equivalent up and down regulation of gene expression is very interesting as it calls to question the notion that HDACs are predominantly transcriptional repressors of gene expression. Indeed, the dynamic changes in gene expression after TSA treatment suggest that HDACs play a more context specific and complex role in regulation gene expression. Indeed, recent work has identified that HDACs are bound to the promoters of active genes and have a more integral role to play in transcription (Z. Wang et al., 2009). Interestingly, our results as well have other work have suggested that HDACs may play a positive regulatory role in stem cell maintenance (Baltus et al., 2009; Saunders et al., 2017) (Chapter 2). Indeed, HDACs have been shown to function to regulate pluripotency gene expression in ESCs and iPSCs (Baltus et al., 2009; Saunders et al., 2017).

Although HDAC activity is critical for maintenance of pluripotency, our data also suggests that a low level of histone acetylation is critical for maintenance of pluripotency. Our work identified that the level of histone acetylation is under tight control in pluripotent cells; not only does disruption in the levels of histone acetylation lead to a loss in pluripotency and defects in neural crest formation, the loss in the ability to read histone acetylation also results in the same phenotype. This is consistent with previous reports that have shown that BRD proteins, such as BRD4, is critical for the maintenance of the stem cell state (Gonzales-Cope et al., 2016). Strikingly, BRD4 has been shown to regulate pluripotency gene expression in these cells (Horne et al., 2015; W. Liu et al., 2014; T. Wu et al., 2015). Our data also suggests that the effects seen due to loss of BRD protein activity is at least, in part, due to loss of transcriptional activation at the mid-blastula transition. Future work should be directed at identifying the mechanisms through which BRD proteins are able to regulate a specific subset of genes specifically during zygotic genomic activation. Our data suggests that it might be thought the interaction with separate binding partners, however more work needs to be done identify if these might be direct interactions and if transcription factors like FoxD3 or TFAP2 are preferentially binding to the same sites as BRD4 in these cells.

Our data provided tremendous insights into the role of histone acetylation in the maintenance of pluripotency, but we do not yet have a complete picture of the epigenomic landscape during stem cell maintenance during embryonic development. Several histone post translational modifications have been shown to be important for pluripotency in ES cells such as H3K4Me3 and H3K27me3. However, there are many other histone modifications the role of which has not be characterized in stem cell maintenance. In particular, not much is known regarding the abundance of histone modifications during embryonic development. Our study is

the first to systematically investigate the dynamics of abundance of histone modifications at the early stages of *Xenopus* development using mass spectrometry. The advantage of the technique is that it allows us to assay in an unbiased manner and compare the abundance of different modifications between the different developmental times and identify modifications that play important roles in pluripotency that can be followed up on in the future. Our data identified several histone modifications that are present in high quantities in the blastula embryos, suggestive that these are required for the maintenance of pluripotency. Interestingly, we found also that several histone PTMs are accumulated in embryo during development and peaked after gastrulation suggestive that these maybe important for lineage restriction. This also suggests that the levels of histone PTMs, and their deposition must be under tight control during embryonic development. Our study provides novel epigenetic insights into the control of pluripotency during embryonic development and provides the basis for further studies to gain mechanistic knowledge into the epigenetic control of the stem cell state.

Materials and Methods

Embryological methods

Wildtype *Xenopus laevis* embryos were obtained using standard methods and staged according to Nieuwkoop and Faber (1994). For animal cap assays, ectodermal explants were manually dissected at early blastula (St. 8-9) from embryos treated with the specified inhibitor at the 2-cell stage and cultured in 1x MMR until sibling embryos reached the denoted stage. Manipulated embryos/explants were processed for *in situ* hybridization by fixing in 1x MEMFA and dehydrating in 100% methanol. *In situ* hybridization was performed using digoxigenin

labelled RNA probes and developed using BM Purple substrate (Roche) (LaBonne & Bronner-Fraser, 1998). Results shown are representative of at least three independent experiments.

Western blot analysis

Animal cap explants (20 – 40 explants) were collected at the indicated stages and lysed in TNE lysis buffer (50mM Tris-HCl [pH 7.4], 150mM NaCl, 0.5mM EDTA, and 0.5% Triton X-100) supplemented with protease inhibitors (Aprotinin, Leupeptin and PMSF) and complete Mini tablet (Roche). SDS-PAGE and western blot analysis was used to detect proteins and modifications using the following antibodies: H3K9Ac (#9649, Cell Signaling, 1:2000), H3K27Ac (ab4729, Abcam, 1:2000), H3 (#3638 and #4499, Cell Signaling, 1:1000) and Actin (A2066, Sigma-Aldrich, 1:4000). For detection and quantification using the LiCOR-Odyssey platform, blots were incubated simultaneously with primary antibodies for both H3K9Ac/H3K27Ac and H3. Histone acetylation was detected using the IRDye® secondary antibodies. Results shown are representative of three independent experiments.

RNA isolation, cDNA synthesis and qRT-PCR

RNA isolation and cDNA synthesis for qRT-PCR were performed as previously described (Buitrago-Delgado et al., 2015). For each condition, 10 animal pole caps were collected in eppendorf tubes (in 1xMMR), frozen on dry ice, and placed at -80°C until beginning of RNA isolation. Samples were lysed in 250 µL homogenization buffer [50 mM NaCl, 50 mM Tris-Cl (pH7.5), 5 mM EDTA (pH 8.0), 0.5% SDS, 200 µg/ml proteinase K (added fresh)] and placed at 37°C for 1 hour. RNA was then purified by phenol-chloroform extraction followed by ethanol precipitation and resuspended in 42 µL molecular grade water. Samples were treated

with RQ1 DNase (Promega) at 37°C for 1 hour. RNA was again purified by phenol-chloroform extraction followed by ethanol precipitation and then resuspended in 21 µL molecular grade water. cDNA was synthesized using High-Capacity cDNA Reverse Transcription Kits (Fisher) and used for qRT-PCR with SYBR Premix Ex Taq II (Tli RNaseH Plus), Bulk (Takara) on a CFX Connect™ Real-Time PCR Detection System (Bio-Rad). The primer sequences used are available in Chapter 2. Expression was normalized to ornithine decarboxylase (ODC) and fold change was calculated relative to control samples of the same stage. Represented is the mean of at least three independent biological replicates with error bars depicting the standard error of mean (SEM). An unpaired, two-tailed t-test was utilized to determine significance.

RNA-sequencing sample preparation

RNA isolation for RNA-seq was performed using TRIzol™-LiCl method. Animal cap explants were collected at stages 9 and 13 (twenty-five caps per condition) in 500µL TRIzol and flash frozen in liquid nitrogen. 500µL of TRIzol was then added to each sample and caps were homogenized by pipetting and vortexed for 15 seconds. RNA was then isolated by chloroform extraction followed by LiCl and ethanol precipitation. Purified RNA was resuspended in 30µL of molecular grade water and RNA from two days was combined together to form one biological replicate for sequencing. Purified RNA was submitted to Northwestern's Sequencing Core for library prep (TruSeq mRNA Library Prep kit) and sequencing (75bp single end on Illumina NextSeq).

Chromatin Immunoprecipitation (ChIP)

Chromatin Immunoprecipitation assays were carried out with 50 animal cap explants for qPCR and 150 animal cap explants per sample for sequencing. Explants were manually dissected from wild-type or vehicle/inhibitor treated blastula stage embryos and collected at blastula stages (Stage 9) or neurula stages (Stage 13) and fixed for 30mins in 1% Formaldehyde (methanol-free) (Thermo-Fischer) on agarose lined plates. Explants were lysed in ChIP lysis buffer (50mM Tris-HCl (pH7.4), 1%NP-40, 0.25%sodium deoxycholate, 150mM NaCl, 1mM EDTA, 0.1% SDS, 0.5mM DTT). The lysis buffer was supplemented with a protease inhibitor cocktail (Sigma), phenylmethylsulfonyl fluoride, N- ethylmaleimide (Sigma) and iodoacetamide (Sigma), Phosphatase inhibitor II and III (Sigma). Immunoprecipitation for myc-tagged proteins was performed using α -H3K27Ac antibody (ab4729, Abcam) and α -H3K9Ac antibody (#9649, Cell Signaling) on Protein G magnetic beads (Dynabeads, Invitrogen #100-04D) for 2 hours at 4°C. Beads were washed in four different wash buffers, 5 minutes each at 4°C. Wash buffer 1: (0.1% SDS, 1% TritonX-100, 2mM EDTA, 20mM Tris-HCl pH 8.0, 150mM NaCl) Wash buffer 2: (0.1% SDS, 1% TritonX-100, 2mM EDTA, 20mM Tris-HCl pH 8.0, 500mM NaCl) Wash buffer 3: (0.25M LiCl, 1% NP-40, 1% Sodium deoxycholate, 1mM EDTA, 10mM Tris-Cl pH 8.0, 150mM NaCl) Wash buffer 4: (10mM Tris-Cl pH 8.0, 1mM EDTA). Protein-DNA complexes were eluted off magnetic beads with ChIP elution buffer (50mM Tris-Cl pH 8.0, 10mM EDTA, 1% SDS, 150mM NaCl) at 65°C overnight. Protein was digested with Proteinase K (Ambion) for 4 hours at 55°C. Samples were subsequently phenol/chloroform extracted and ethanol precipitated. For sequencing, DNA samples were submitted directly to the Northwestern Sequencing Core for library prep (TruSeq ChIPSeq kit) and 75bp single-end Illumina

sequencing. For qPCR, the precipitated DNA was column purified using the NucleoSpin PCR Clean-up kit (Clontech). qPCR was performed using SYBR Premix (Clontech #RR820W) and primers for 3 genomic loci of Sox3 and Sox17 as well as proximal promoter of eEF1 α . Primer sequences are listed below.

Sample collection for mass spectrometry

Whole embryos (5) were collected at 6 different developmental time points according to a fixed time and temperature regimen (Credit: Experimental framework adapted from Kristin Johnson). Briefly, *in vitro* fertilization was done at 2pm and the embryos were aged overnight till blastula stages at 14°C. At precisely 8:30am, the embryos were moved to 18°C, and the first sample (Stage 9) was collected at 9:45am. The embryos were then moved to 20°C and aged till Stage 13. Embryos were collected consecutively at 11 am (Stage 10), 12:15pm (Stage 10.5), 1:30pm (Stage 11), 3pm (Stage 12) and 5pm (Stage 13). The embryos were collected into Eppendorf tubes and all the extra liquid was removed and the samples were flash frozen. The samples were submitted to the Northwestern Proteomics Core for preparation for mass spectrometry and “Histone Panel B” analysis was performed.

DNA Constructs and Inhibitors

For HDAC inhibition, embryos/explants were treated with Trichostatin A (Sigma-Aldrich) at a final concentration of 500nM. For BRD protein inhibition, embryos/explants were treated with IBET762 (Cayman) at a final concentration of 250 μ M.

Animals

All animal procedures were approved by the Institutional Animal Care and Use Committee, Northwestern University, and are in accordance with the National Institutes of Health's "Guide for the Care and Use of Laboratory Animals".

Bioinformatics Analysis:

RNA-sequencing Analysis

The obtained reads were checked for quality using FAST-QC (Babraham Bioinformatics) and aligned to the *Xenopus Laevis* genome 9.2 (Xenbase) using STAR aligner (Dobin A et al, 2012). The aligned reads were counted using HTSeq Counts (Anders S et al, 2014). Differential analysis was done in R using the DESeq2 package applying standard log-fold shrinkage procedures and TSA dataset and IBET datasets were processed separately with their own controls (Love M et al, 2014). Genes were considered significantly changed for $\text{padj} < 0.05$. For the MBT analysis, Egg-Stage, Stage 8 and Stage 9 data was obtained from Xenbase (Session et al, 2017). The IBET and MBT datasets were matched to find the overlapping genes, and only those were considered for the analysis. The MBT dataset was filtered to remove genes which did not have $\text{TPM} > 0.5$ in any of the three stages. Genes that are zygotically transcribed were identified based on the criteria of $\text{Stage8TPM/EggStage}$ or $\text{Stage9TPM/EggStage}$ greater than 5. Bash and R scripts used in the analysis are found in the appendix (Chapter 6).

ChIP-sequencing Analysis

The obtained reads were checked for quality using FAST-QC (Babraham Bioinformatics) and aligned to the *Xenopus Laevis* genome 9.2 (Xenbase) using bowtie2. The obtained samfiles

were converted to bam and sorted and indexed using samtools. The indexed bam files were visualized using Integrated Genome V (IGV). For generating the heatmaps, bigwig files were generated from the indexed bamfiles using deeptools bamCoverage, followed by computeMatrix and plotHeatmaps. The bed files containing the genomic coordinates of the +/-1.5kb from TSS was generated using a custom R script. All bash and R scripts used in the analysis are found in the appendix (Chapter 6).

Chapter 4

Exploring the circuitry controlling pluripotency of neural crest stem cells

Neural crest cells have a unique ability to retain their stem cell attributes while rest of the embryo undergoes lineage restriction. These cells employ a novel mechanism to preserve their stem cell state allowing them to give rise to cell types from multiple germ layers. While we have some insights into the regulatory circuit that controls the pluripotency of these cells, we still have much to learn regarding the mechanisms employed by neural crest cells to retain their stem cell attributes. Here, we utilize a genome wide transcriptomic approach and computational analysis to further characterize the gene regulatory circuitry that is involved in the control of pluripotency of neural crest cells. Using RNAsequencing, we compare global gene expression changes between neural crest cells and epidermal cells to identify factors that are involved in neural crest formation. Further, we use WGCNA and PCA/sparse-PCA methodologies to investigate the differences between the early and late neural crest signature and identify genes that play important roles in neural crest stem cell maintenance. This study furthers our understanding of the transcriptional landscape and epigenetic control of this unique cell type that has contributed to the evolution of vertebrates.

Introduction

Embryonic development is characterized by the gradual restriction of lineage potential as embryo progresses from a totipotent zygote to terminally differentiated cells. Neural crest cells are an exception to this paradigm of progressive lineage restriction and have the unique capability to retain their stem cell attributes. These cells stay in a suspended state of pluripotency far longer than their cellular neighbors and give rise to a diverse array of cell types that are considered ectodermal such as melanocytes and peripheral ganglia, as well as mesodermal in nature e.g. chondrocytes and smooth muscle cells (LaBonne & Bronner-Fraser, 1999; Prasad et

al., 2012). This multi-germ layer developmental potential has led to neural crest cells being described classically as the fourth germ layer and its acquisition has led to evolution of vertebrates (Bronner & LeDouarin, 2012; Hall, 2000).

Neural crest cell development is a highly controlled process that occurs in a step wise manner. Through a highly coordinated process, these cells begin as a population of stem cells and then begin to lineage restrict to give rise to derivatives, albeit later than the rest of the embryo. This suggests that exist a rigid temporal control that allows for the maintenance of temporary state of pluripotency but also allows for lineage restriction to begin at the appropriate developmental time. The gene regulatory network that controls neural crest development must consist of factors that are responsible for the maintenance of pluripotency, but also genes that push the cells out of pluripotency and begin the process of differentiation.

We have a fairly thorough understanding of the process through which neural crest cells are specified and begin to lineage restrict. The interplay of FGF, WNT and BMP signals set up a zone of competence at the border of the neural ectoderm and non-neural ectoderm that drives the expression of neural plate border specifiers (Sauka-Spengler & Bronner-Fraser, 2008). These factors such *Pax3*, *Zic1* and *TFAp2*, in turn, act in concert to drive the expression of neural crest specifiers that demarcate the cells that will form the future neural crest. Interestingly, neural crest specifiers such as *Snail2* and *Twist* function reiteratively later in neural crest development and are necessary for these cells to undergo epithelial to mesenchymal transition (Simões-Costa & Bronner, 2015). While decades of research have revealed the precise components of the complex gene regulatory network that control neural crest formation, much is left to be elucidated regarding the early stages of neural crest formation and the genes involved in the process of maintenance of pluripotency of neural crest cells.

Neural crest cells are considered to be a multipotent stem cell population and have the capability to give rise to mesodermal and ectodermal derivatives. Recent work from our lab has shown that these cells might have the ability to give rise to endodermal derivatives at least *in vitro*, suggestive that neural crest cells might be pluripotent (Buitrago-Delgado et al., 2015). Interestingly, single cell lineage tracing experiments have shown that neural crest cells give rise to several derivatives *in vivo* (Bronner-Fraser & Fraser, 1988). Further, *in vitro* clonal analysis have shown that neural crest cells have the ability to self-renew (Baroffio et al., 1988; Calloni et al., 2009; Trentin et al., 2004). Indeed, studies have found that neural crest cells grown at clonal density can continue to undergo self-renewal for 10 days in culture (Stemple & Anderson, 1992). These data suggest that neural crest is a stem cell population and must express factors that are tasked with the maintenance for pluripotency of these cells.

Neural crest cells have been found to express several transcription factors that have been described to have important role in maintenance of pluripotency. Spatial genomic analysis in chick identified that there is signature of pluripotency gene expression in early undifferentiated neural crest cells which has characteristic high Oct4, Nanog and Klf4 expression that sets these cells apart from neural stem cells (Lignell et al., 2017). Strikingly, recent work from our lab identified that there are several genes that are shared between blastula cells and neural crest cells suggestive that a common transcriptional circuitry regulates the pluripotency of these two cell types (Buitrago-Delgado et al., 2015). Key among these is the proto-oncogene Myc, which has been identified as a factor that is essential for the maintenance of pluripotency and reprogramming of somatic cells (Takahashi et al., 2007). Interestingly, Myc, and its downstream effector Id3, have also been shown to be essential for neural crest formation and stem cell maintenance (Bellmeyer et al., 2003) (Light et al., 2005). In a similar manner, several core

pluripotency factors have been shown to be expressed in neural crest cells such as *Vent2(Nanog)*, *Oct25 (Oct4)*, *Lin28a* (Buitrago-Delgado et al., 2015) (Bhattacharya et al., 2018) (Lignell et al., 2017). Interestingly, SoxB factors, Sox2/3 are present in blastula cells but are replaced by the related SoxE factors in neural crest cells (Buitrago-Delgado et al., 2018). This suggests that there are interesting similarities and differences in the transcriptional landscape between these cell types that contribute the regulation of pluripotency of these cells.

We also find that there are several neural crest factors that are expressed in blastula cells. Indeed, we identified that several neural crest factors *Snail*, *Sox5*, *Pax3*, *TFap2*, *Id3* are expressed earlier than previously thought and are present in blastula cells. Further, we find that not only are these factors expressed in blastula cells but also seem to have functional roles maintain pluripotency in these cells. Strikingly, we found *Snail* and *Sox5* are essential for the pluripotency of blastula cells (Buitrago-Delgado et al., 2015) (Chapter 6, Appendix). Further, the shared features of these two cell types extends beyond just the transcriptional landscape. In recent work from our lab, we have identified that FGF/MAPK signaling is important for pluripotency and stem cell maintenance of both blastula and neural crest cells (Geary & LaBonne, 2018). We also recently identified that HDACs and histone acetylation are necessary for the maintenance of pluripotency of blastula and neural crest cells (Chapter 2,3). This suggests that strong links exist in the control of pluripotency of blastula and neural crest cells.

While we have made tremendous strides in enhancing our knowledge regarding the processes that are involved in stem cell maintenance of neural crest cells, we still have much to learn in order to gain a complete picture of regulatory circuitry that controls the pluripotency of these cells. The ascent of genomic technologies has given us the capability to assay like never before the global transcriptional and epigenetic landscape (Simões-Costa & Bronner, 2013).

Instead of using a candidate approach, these technologies provide us with the capability to explore in an unbiased manner the factors that contribute to neural crest stem cell maintenance. Further, we can use these methodologies to identify how these different factors relate to each other and how they integrate into the network that contributes to the pluripotency of these cells. This methodology also provides us with a framework which we can use to identify candidate genes that we can then go back to the laboratory and test using traditional approaches to identify their function in the process.

Here, we use a genome-wide transcriptomic approach to explore and characterize the regulatory circuitry that controls the pluripotency of blastula and neural crest cells. We find global differences in the transcriptional landscape between neural crest cells and epidermal cells. We further identify that there exists an early and late neural crest signature during neural crest formation. We leveraged WGCNA and PCA/sparse-PCA methodologies to identify novel factors that are involved in neural crest formation and postulate that factors contributing to the early neural crest signature are likely involved in stem cell maintenance, while late neural crest genes are involved in the differentiation of these cells. Together, these data give a framework within which we can further our understanding regarding the control of stem cell potential of this unique embryonic cell population.

Results

Transcriptomic analysis of neural crest formation

Our earlier work identified that neural crest cells have a unique ability to retain their stem cell attributes, yet much is left to be learned about the mechanisms that control this process and the factors that contribute significantly to the maintenance of pluripotency of these cells. Hence,

we sought to explore further the gene regulatory circuitry that is responsible for maintaining the neural crest stem cell state. To this end, we performed a genome-wide RNA-sequencing experiment comparing animal cap explants reprogrammed to a neural crest state and control explants at different stages (Figure 4.1). We took advantage of the fact that animal caps explants can be reprogrammed to a neural crest state using exogenously injected Wnt/Chd and we were able to obtain neural crest cells and age matched control explants (default to epidermal state) (LaBonne & Bronner-Fraser, 1998). Animal caps were collected at blastula (Stage 9), early neurula (Stage 13) and late neurula (Stage 17). This experimental set up provided us with a unique advantage to simultaneously assess the dynamic changes in gene expression that occur during lineage restriction, both as a function of developmental time and cell state. *In situ* hybridization was performed on caps as a control to ensure that there was uniform reprogramming to the neural crest state as seen by expression of *Snail2* and *FoxD3* for each of the biological replicates prior to sequencing (Data not shown). We performed hierarchical clustering on the samples and no significant batch effects were observed in the data suggestive that the biological replicates are true replicates and the changes we observe are not due to random sample differences (Figure 4.2). Dynamic changes in gene expression occurs as animal cap explants proceed from a pluripotent blastula state to an epidermal or neural crest state (Figure 4.3).

Differential analysis identifies differences between neural crest explants and control explants

Animal cap explants gradually and naturally default to an epidermal state, while the neural crest explants reach an early neural crest state (Stage 13) and a more specified neural crest

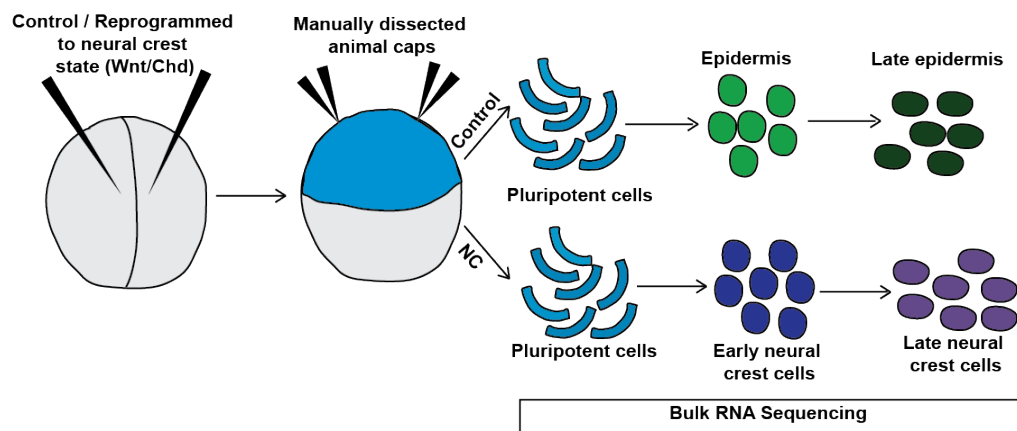


Figure 4.1 Animal cap explants RNA-sequencing workflow

Schematic representation of RNA-sequencing experiment of animal cap explants from control embryos or embryos injected with Wnt/Chd. Animal cap explants were dissected from embryos at blastula stages and explants were cultured until sibling embryos reached blastula stages (Stage 9), early neurula stages (Stage 13) and late neurula stages (Stage 17) and collected for RNA isolation, library preparation and sequencing.

Heatmap depicting the hierarchical clustering of RNAseq biological replicates based on the Euclidean distance between samples. No significant batch effects are seen as similar samples cluster together irrespective of biological replicate.

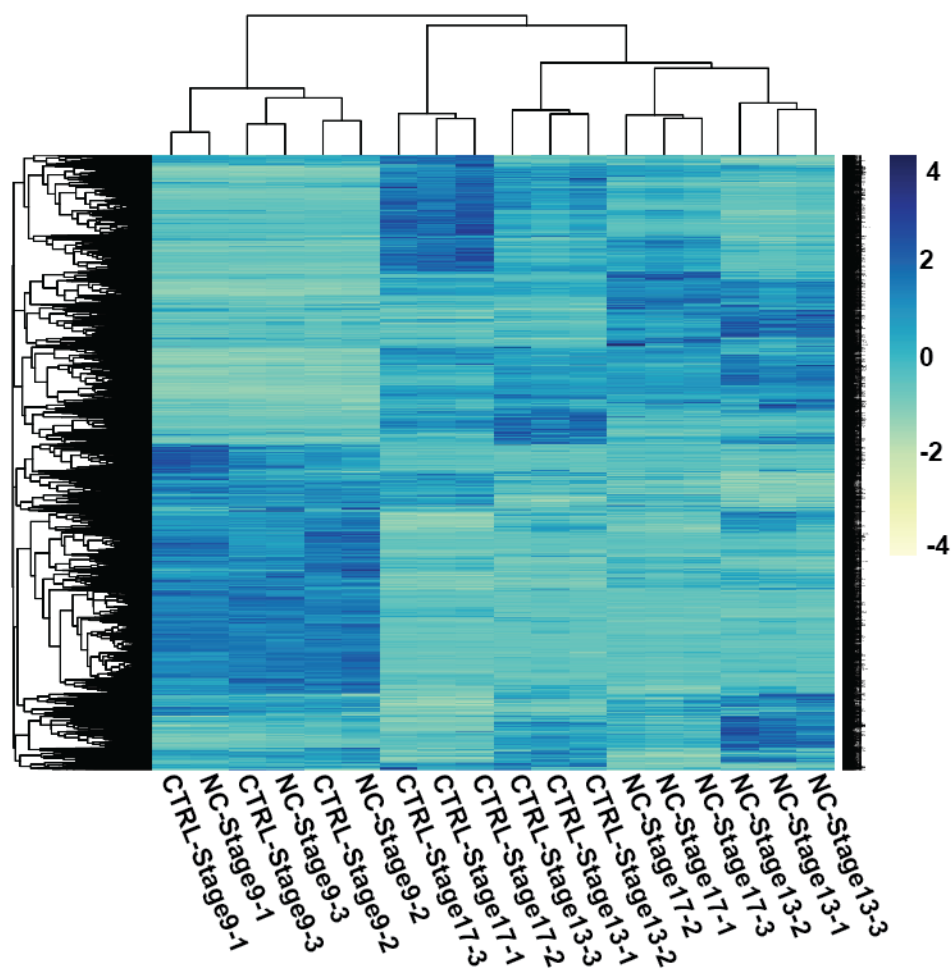


Figure 4.3 Global changes in gene expression during lineage choices

Heatmap depicting changes in gene expression as animal cap explants progress from blastula stage to epidermal lineage or reprogrammed to a neural crest state.

(Stage 17). We hoped to analyze the differences in gene expression between the neural crest cells and the aged matched control explants to explore further the transcriptional landscape of neural crest cells at various stages in development. DESeq2 was utilized to perform stage specific pairwise differential analysis and estimate log fold changes, and a threshold for adjusted pvalue was set at $pvalue < 0.05$ (Love, Huber, & Anders, 2014). As expected, we see differentially expressed genes between neural crest explants (Wnt/Chd) and control explants at all stages (Figure 4.4). Interestingly, we find minimal changes in gene expression (no. of differentially expressed genes = 101) seen at blastula stages between control and neural crest (Wnt/Chd) explants which suggests that at blastula stages these two cell types are almost equivalent (Figure 4.5). We observed very high expression of Wnt8 and Chordin detected at these stages, but it is likely an artifact from the detection of injected mRNA at these stages. Strikingly, we find that *Nodal* is strongly upregulated in these cells, suggestive that TGF-Beta signaling might be playing an important role in early neural crest development.

By early neurula stages (Stage 13), we find that there are dramatic changes in gene expression between the control and neural crest explants. Strikingly, 6619 genes are differentially expressed ($p < 0.05$, $LFC > |1|$) between the early neural crest and epidermal cells (Figure 4.6). Notable among the genes that are significantly upregulated in neural crest cells are known neural crest genes such as *FoxD3*, *Sox9*, *Snail*, *Snai2* and neural plate border markers such as *Pax3*, *Zic1*, *Msx1* etc. Indeed, the gene expression profile matches closely our expectation based on our prior knowledge of the NC-GRN and gave us confidence that our experimental set up was sound and we have successfully reprogrammed these cells to a neural crest state. We also saw a downregulation of several epidermal markers such as *xk81a1.L* (EPK), *krt17.L*, *trim29.L* etc in

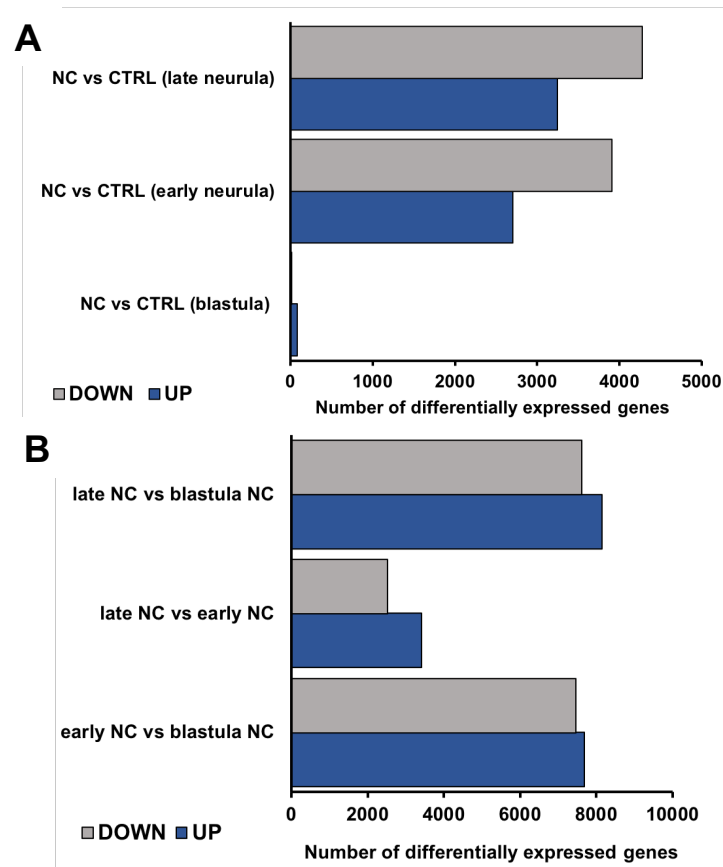


Figure 4.4 Differential expression between control and neural crest explants
 Bar charts depicting differentially expressed genes between control and neural crest explants at all stages (A) and different stages of neural crest explants (B).

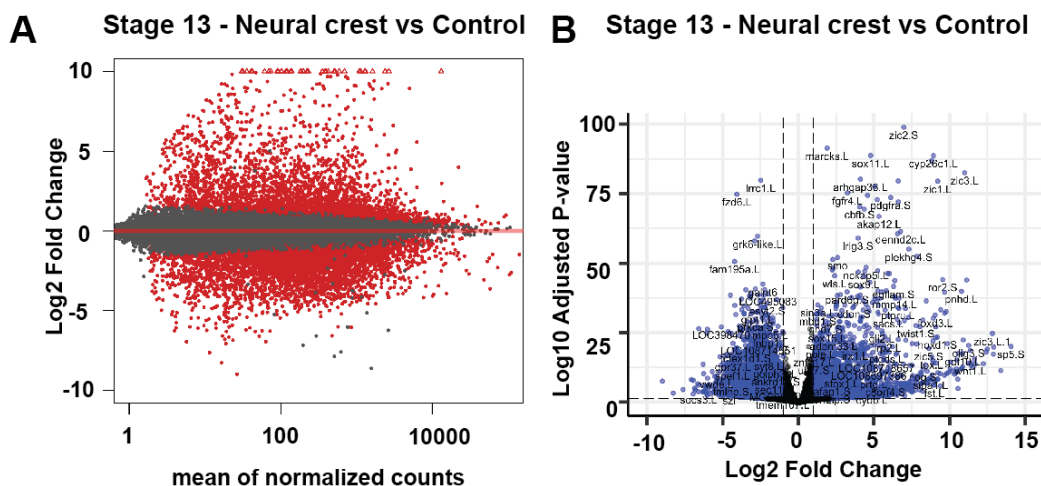


Figure 4.6 Changes in gene expression between control and neural crest explants at early neurula stage

MA-plot (A) and volcano plot (B) depicting gene expression changes in animal cap explants from control embryos or embryos injected with Wnt/Chd. 6619 genes are differentially expressed between control explants and neural crest explants at Stage 13.

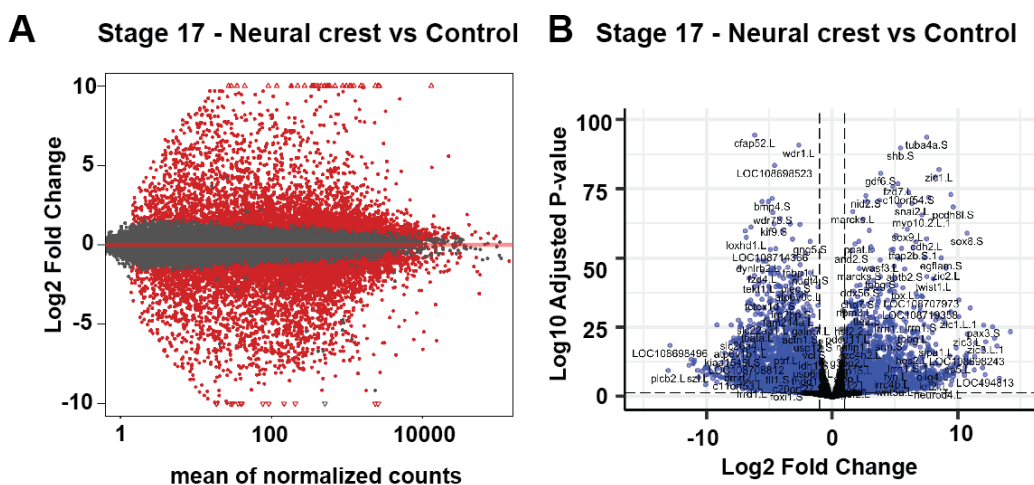


Figure 4.7 Changes in gene expression between control and neural crest explants at late neurula stage

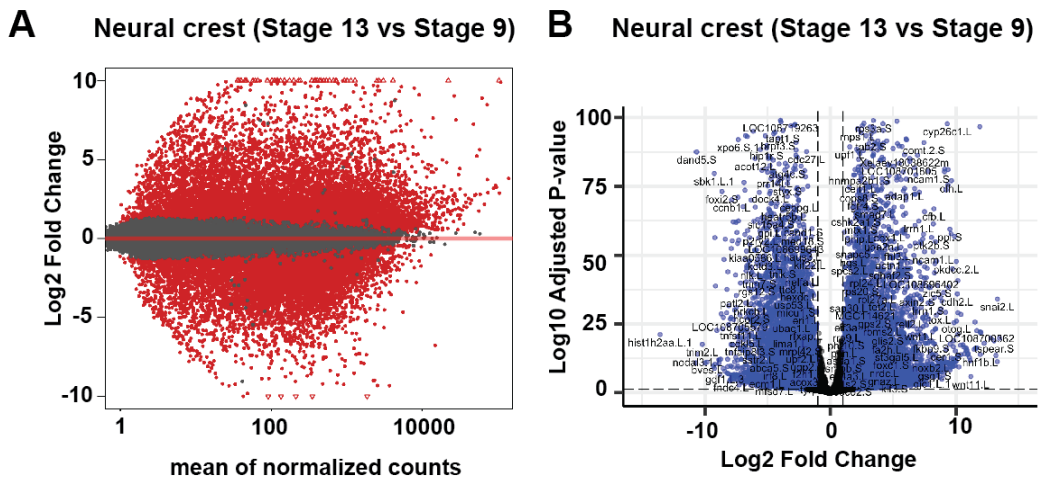
MA-plot (A) and volcano plot (B) depicting gene expression changes in animal cap explants from control embryos or embryos injected with Wnt/Chd. 7526 genes are differentially expressed between control explants and neural crest explants at Stage 17.

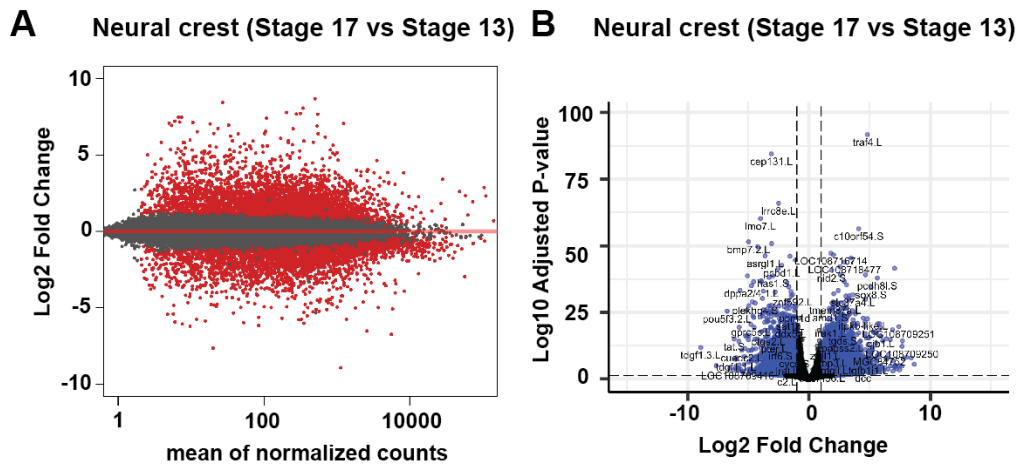
the neural crest in comparison to the control explants, confirming that these cells were forming the neural crest at the expense of defaulting to an epidermal state.

In order to gain a complete understanding of the temporal control of neural crest formation, we also sought to characterize the premigratory neural crest state. By Stage 17 in *Xenopus*, the neural crest cells are specified, and it is just prior to the onset of migration. Interestingly, when we compare neural crest explants and control explants at this late neurula stage (Stage 17), we find that several late neural crest genes such as *Sox10*, *Snai2*, *FoxD3*, *Zeb2*, *Twist* etc have strong expression in the neural crest explants. These genes are known to be markers of late neural crest formation and have roles in promoting epithelial to mesenchymal transition. About 7526 genes are significantly differentially expressed between the neural crest explants and control explants at these stages ($p < 0.05$, $LFC > |1|$) (Figure 4.7). Similar to stage 13, we find that several epidermal genes are strongly downregulated in the neural crest explants. These data suggest that we had successfully obtained transcriptomic data of early premigratory neural crest development, and we could now use this data to explore further the circuitry that controls the pluripotency of these cells.

Analysis of early neural crest development reveals an early and late neural crest signature

Given that we had successfully generated a neural crest transcriptomic dataset, we next wanted to characterize the temporal program of gene expression during the process of neural crest formation. Our initial analysis revealed that global changes in gene expression were taking place when the cells were forming neural crest instead of epidermal cells. We further wanted to characterize the changes in expression as cells progressed from an early (neural plate border) to a late neural crest state (neural crest progenitor). Interestingly, we find that large scale changes in





gene expression occurs during the course of neural crest development. 15143 genes are differentially expressed between neural crest explants at blastula and early neurula stages (Figure 4.8). This is particularly interesting because it suggests that while pluripotent blastula cells share a common transcriptional circuitry, the levels of these factors are markedly different between these two cell types. Strikingly, 5940 genes are differentially expressed between neural crest explants at early neural stages and late neurula stages (Figure 4.9). This suggests that there are changes in the transcriptional landscape during the early stages of neural crest specification. Based on these results, we postulated that there are a set of genes expressed early in neural crest development (characteristic of the neural plate border) that are required for the maintenance of pluripotency of these cells while a set of genes that is turned out on later in neural crest development (characteristic of a neural crest progenitor) that would be required for the later specification and migration of these cells. Indeed, when we look closer at the expression of genes that we know are involved in neural crest formation, we find genes that are considered bona-fide neural crest markers such as *FoxD3* and *Sox9* that increase monotonically over neural crest development (Figure 4.10). There are also genes that are late neural crest genes that are low early in development and are turned on dramatically at the late neural crest stage such *Sox10* and *Twist* (Figure 4.11). Further, there are set of genes that are high in the early neural crest stage such as *Pax3* and *Zic1* which start to go down in expression as the neural crest begins to become more specialized (Figure 4.12). This data suggests that there exists an early and late neural crest gene signature; and we hypothesized that early signature represent genes that play important roles in the maintenance of pluripotency and the late neural crest signature contributed to neural crest specification.

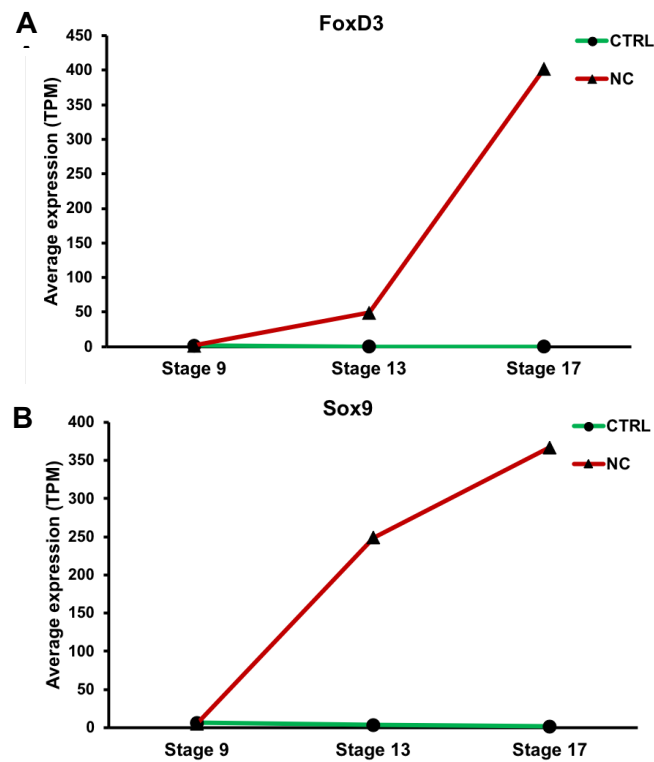


Figure 4.10 Expression of neural crest genes

Average expression (in TPM) of bonafide neural crest markers that increase monotonically during early neural crest development.

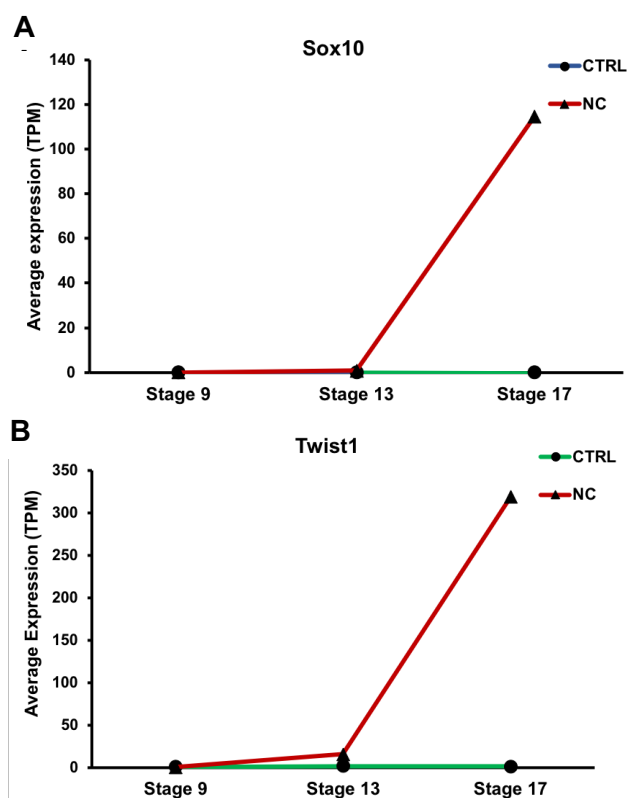


Figure 4.11 Expression of late neural crest genes (specification)

Average expression (in TPM) of neural crest genes that have low expression at the early neural crest state and increase in expression at the later neural crest state.

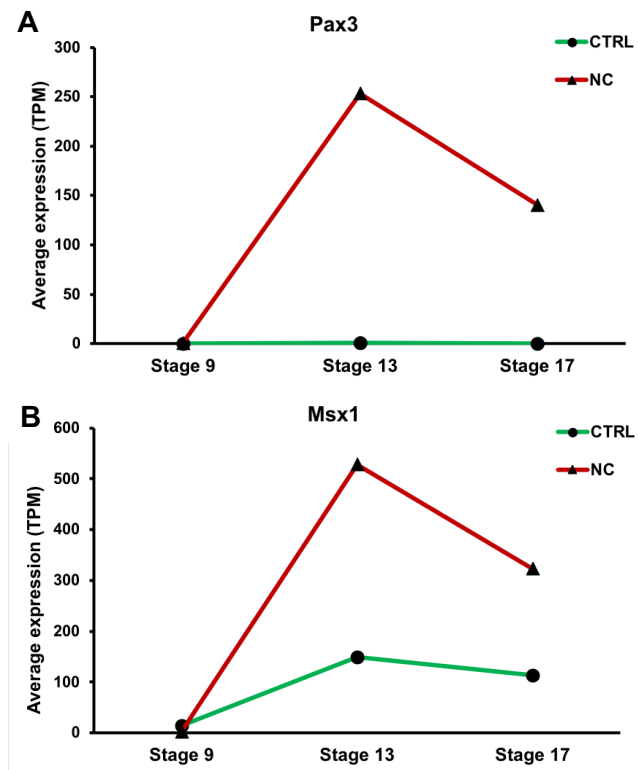


Figure 4.12 Expression of early neural crest genes (border)

Average expression (in TPM) of neural crest genes that have high expression at the early neural crest state and go down in expression at the late neural crest state.

WGCNA identifies genes that highly correlated with an early and late neural crest signature

Our differential analysis revealed that dynamic changes in gene expression occurred during the process of neural crest formation, we sought to explore further the temporal program of gene expression during this process. We hypothesized that there existed an early (neural plate border) and a late neural crest signature (neural crest progenitor), and we hoped to identify in an unbiased manner the genes that contribute to these different cell states. To this end, we took advantage of an established network-based approach called WGCNA to identify genes that are highly correlated with the early or the late neural crest state. Weighted Gene Co-expression network analysis maximizes a network's scale free topology and identifies clusters of genes that are highly correlated and relates them to an external sample trait (Langfelder & Horvath, 2008) (Figure 4.13). This method performs pair wise comparisons on the genes and clusters them based on their gene expression. We performed WGCNA on our dataset (filtered and normalized DESeq2 gene counts) for the top 15000 most differentially expressed genes across 18 samples. Sample clustering based on the Euclidean distance between samples visualized as a dendrogram revealed that there were no significant outlier samples in our dataset (Figure 4.14). In order to build a network from the dataset, we first calculated the appropriate soft thresholding power (β) to ensure a scale-free topology (Figure 4.15). As shown, we changed the power value step by step to determine the optimal value that the network connectivity is smooth. The $\beta = 8$ was determined as the lowest power for which the scale free topology fit index curve flattens out (Figure 4.15). We next constructed an unsigned network from the data that takes into consideration both positively and negatively correlated genes. Briefly, the calculated adjacency

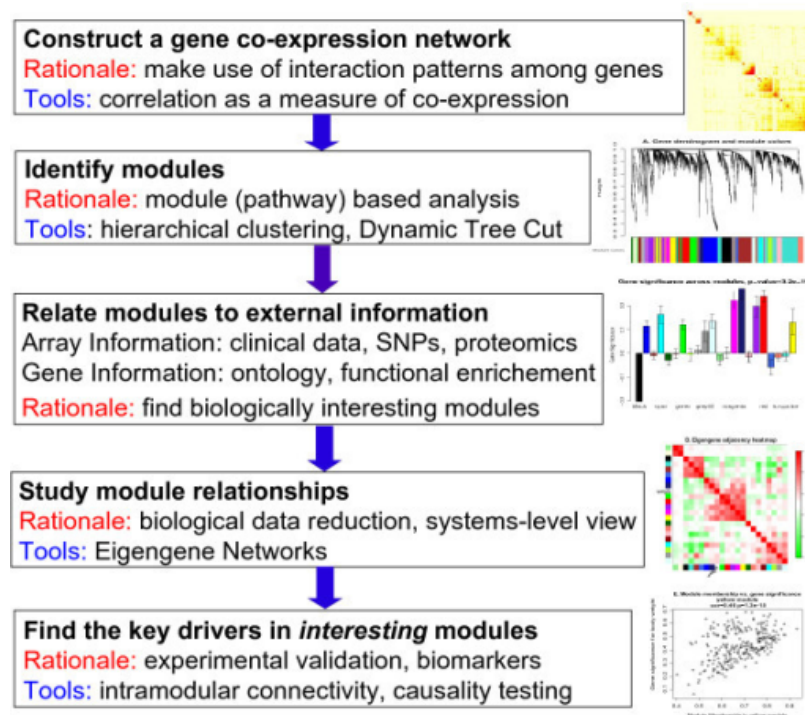


Figure 4.13 Weighted Co-expression Network Analysis

Framework for building co-expression networks from gene expression (RNASeq) data. (Adapted from Langfelder and Horvath, 2011)

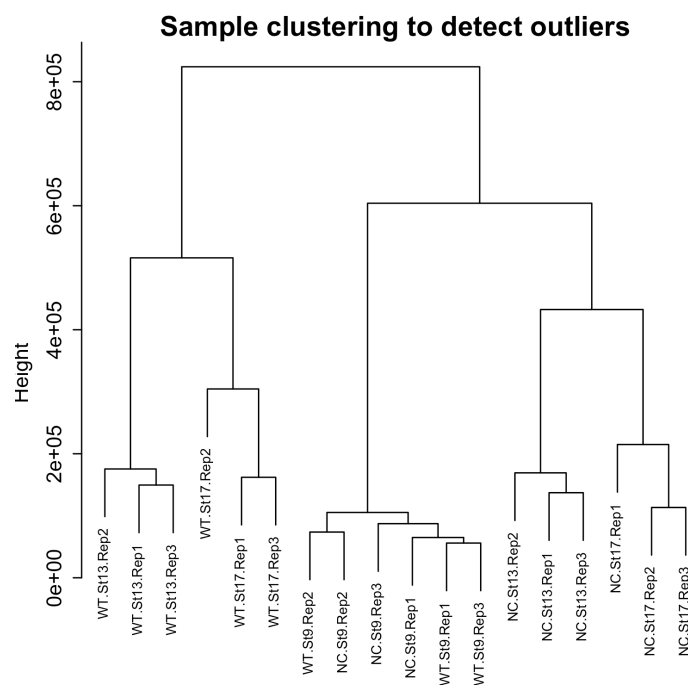


Figure 4.14 Sample clustering to detect outliers

Prior to performing WGCNA, samples were clustered together based on euclidean distance. No outliers were detected between the 18 samples.

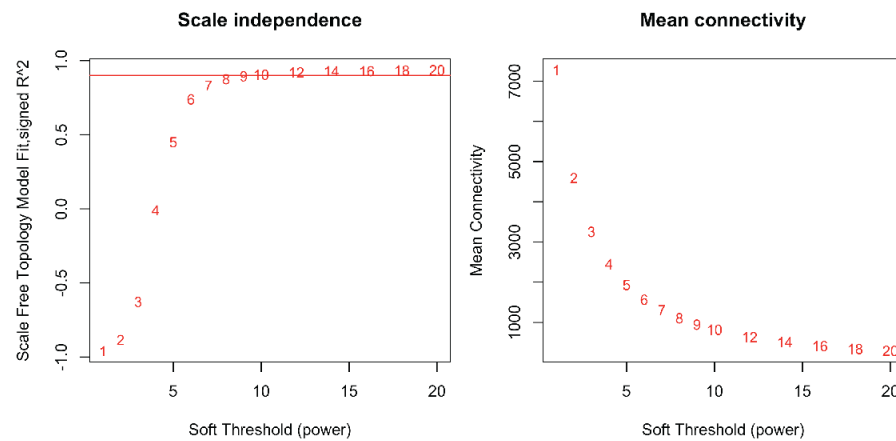


Figure 4.15 Determination of power beta value in WGCNA

Analysis performed to determine the power beta value from the adjacency matrix using WGCNA. The weighted parameter power value was determined using the scale free criterion. To ensure average connectivity was smooth, $\beta = 8$ was chosen for topology fitting results (A) and mean connectivity (B).

matrix at $\beta = 8$ was further transformed to topological overlap matrix (TOM) using WGCNA package. The hierarchical clustering on all the genes were performed to generate a dendrogram. By using dynamic tree cutting, the functional clusters (modules) were obtained from the constructed gene dendrogram. The topology over matrix (TOM) among all the genes in the dataset depicts the constructed network and module gene membership (Figure 4.16).

By clustering correlated genes together, 9 independent co-expression gene modules were identified from the data (Figure 4.17). Finally, we related the module colors to the external sample trait data, in this case, the developmental time and cell state. Strikingly, we find that there is at least one module that is highly correlated with control and neural crest explants at Stage 13 and Stage 17. Module green is highly correlated with control explants at St13, while Stage 17 control explants have high correlation with two modules, black and brown. Interestingly, we find that there is a module (blue) of 3578 genes that is highly correlated with the early neural crest state (Stage13) ($R^2=0.83$) while there is another module (yellow) of 666 genes that is strongly correlated with the late neural crest state (Stage 17) ($R^2 = 0.91$). Visualizing the genes of each module using a heatmap shows significant enrichment of module blue genes in early neural crest cells (Stage 13) and module yellow genes in late neural crest explants (Figure 4.18, 4.19). This suggests that there exists an early and late neural crest gene signature and the control of the transcriptional landscape at each stage is necessary to ensure maintenance of pluripotency or promotion of lineage restriction. Given the developmental state, we expect that the early neural crest signature likely involves genes that are required for stem cell maintenance while the late neural crest signature involves genes important for lineage specification and migration of these cells.

Network Heatmap - All genes

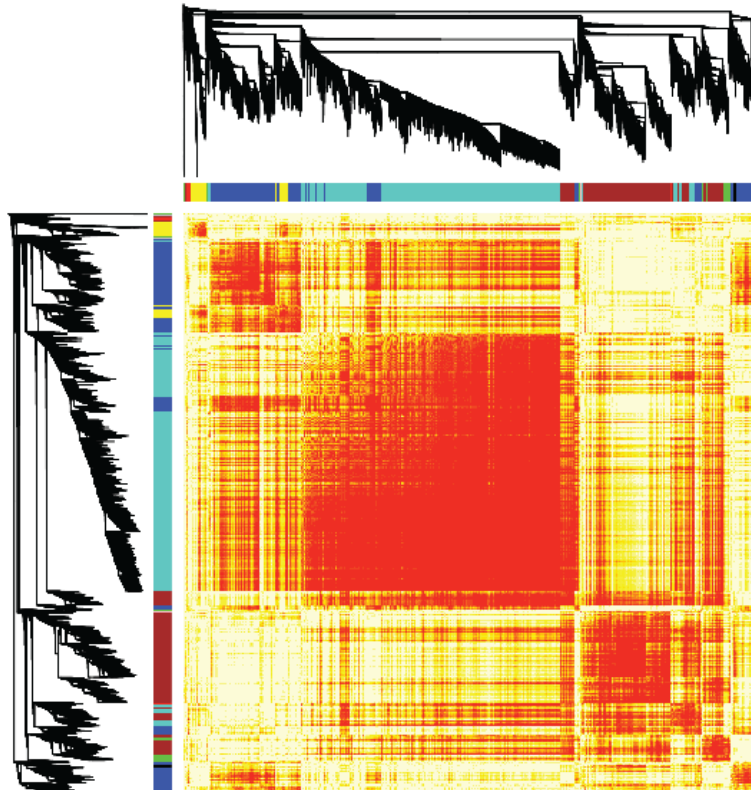


Figure 4.16 WGCNA on top 15000 most varying genes

Heatmap depicting the Topology Overlap matrix (TOM) generated from WGCNA performed on the top 15000 most varying genes in the dataset. WGCNA identified 9 distinct functional modules which are depicted with different colors.

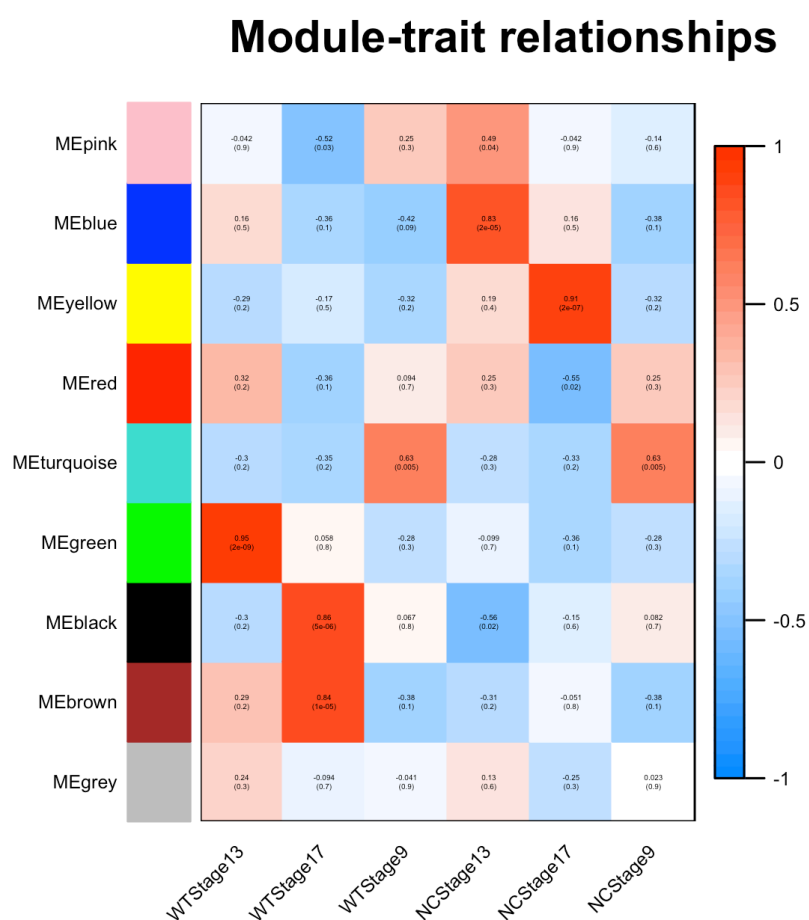


Figure 4.17 Relating modules to sample type – developmental time/cell state

Heatmap representing the correlation between modules and sample type from WGCNA performed on the top 15000 most varying genes in the dataset. Module blue is highly correlated with the early neural crest state (Stage 13) and module yellow is highly correlated with the late neural crest state (Stage 17).

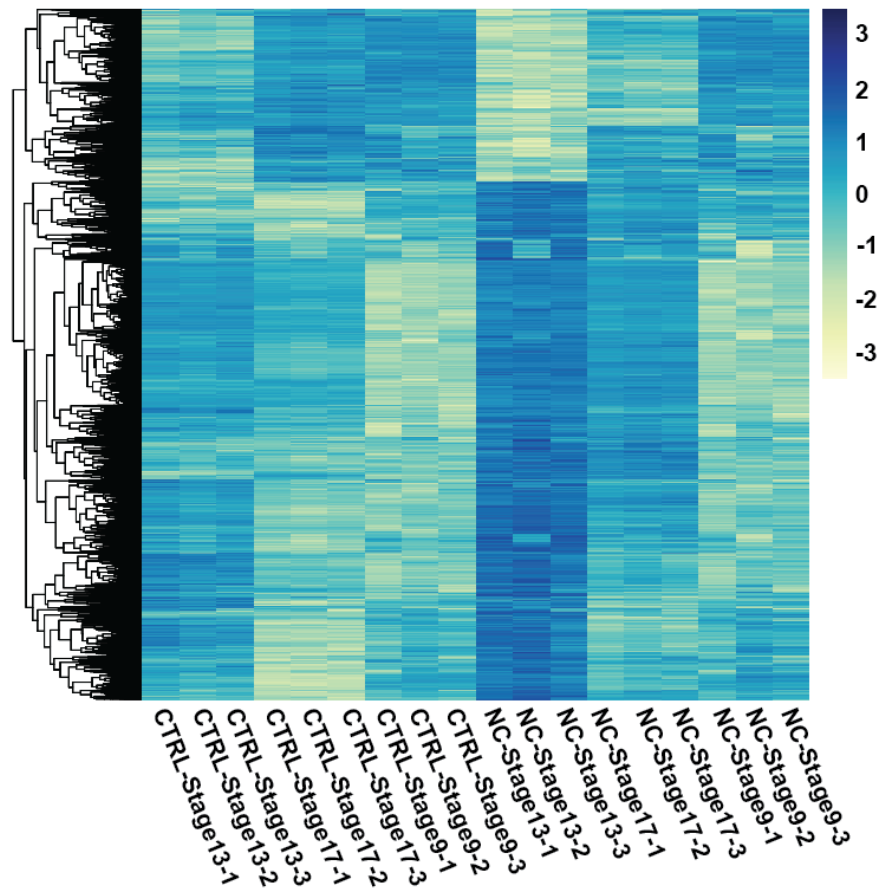


Figure 4.18 Module blue genes are highly correlated with early neural crest state

Heatmap depicting normalized count data across all samples for module blue genes. Module blue genes are highly correlated with the early neural crest state. About 70% of the genes are strongly expressed in Stage 13 neural crest explants while the rest are strongly downregulated in these cells.

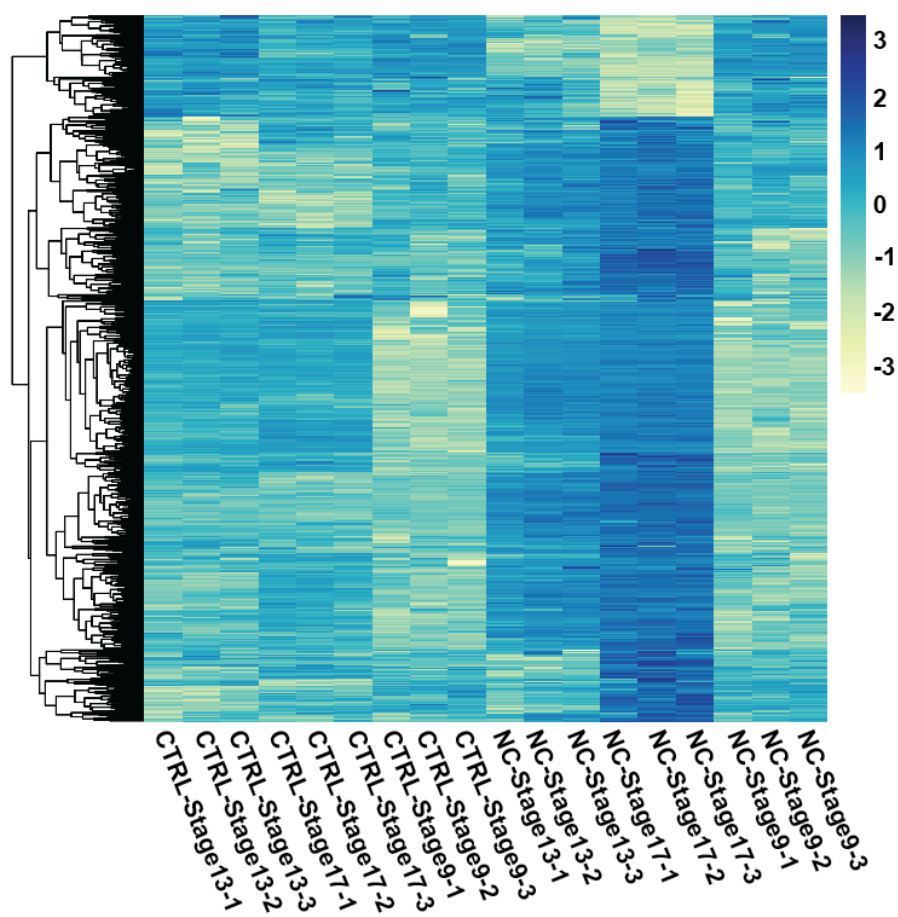


Figure 4.19 Module yellow genes are highly correlated with late neural crest state

Heatmap depicting normalized count data across all samples for module yellow genes. Module yellow genes are highly correlated with the early neural crest state. About 80% of the genes are strongly expressed in Stage 17 neural crest explants while the rest are strongly downregulated in these cells.

The late neural crest module consists of genes that we have previously known to play important roles in neural crest formation. Strikingly, several genes such as *Twist1*, *Snai2*, *Sox10* that are known to be involved in late neural crest development and EMT/migration are found in this module. Interestingly, we find that the genes in module blue which is highly correlated with early neural crest state involve ones with known roles in the maintenance of pluripotency of these cells. Key among these is *HDAC1* which we previously identified to be critical for neural crest stem cell maintenance (Chapter 2). Further, several other genes such as *Fgfr4*, *Snai1*, *Myc*, *Pax3*, *TFAP2a*, *lin28a* that have been shown to be necessary for the control of pluripotency in these cells are also found in module. While typically WGCNA assigns a gene to a single module, the pseudo-tetraploidy *Xenopus* provided us with a unique advantage as the two gene alloalleles could be independently assigned to modules. Strikingly, we find that for several of the key neural crest factors, one alloallele is found in the blue module while the other alloallele is found in the yellow module. This suggests that there is a core set of genes that are necessary for neural crest development at all stages such as *Sox9* (L form in blue, S for in yellow), *Ets1* (L form in blue, S for in yellow) and *Myc* (L form in blue, S for in yellow). Interestingly, we find that while the *Sox2* and *Sox9* seems to be expressed in the early neural crest state, this is replaced by *Sox9* and *Sox10* in late neural crest state.

Our data also provided us with a framework wherein to identify new genes with hitherto uncharacterized roles in neural crest formation. In both the blue and yellow module, we find that genes that had not previously been known to be involved in neural crest formation. Interestingly, in the blue module, we find genes such as *Sox15*, *hmgb3*, *ddx5*, *znf346* and others which have been shown to be involved in stem cell maintenance that would be interesting genes to follow up on to identify their function in these cells. Of particular interest are several chromatin remodelers

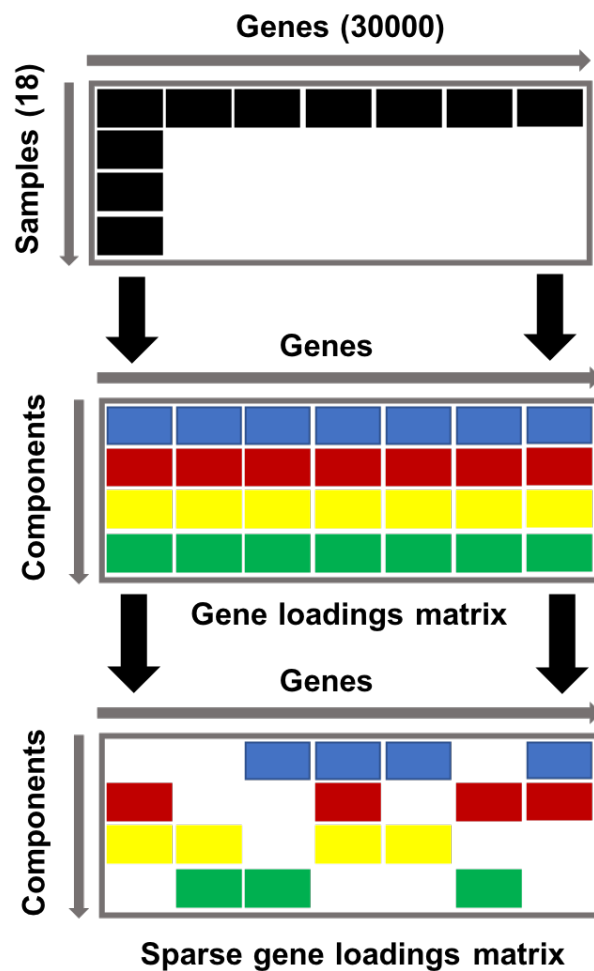


Figure 4.20 Schematic representation of principal component analysis and sparse PCA

Diagrammatic representation of PCA and sparse PCA analysis of gene expression data set. PCA is performed on gene matrix of 30K genes and 18 samples, generating principal components with weights/loadings for all genes. Sparse PCA enforce penalty parameter driving some of those loadings to zero and retains the important non-zero components (genes) that contribute the most to the variance in the data.

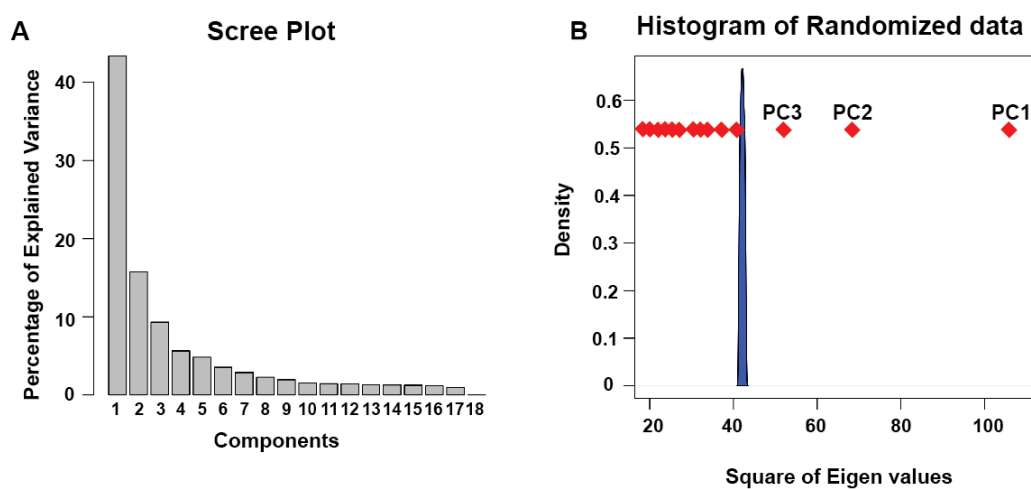


Figure 4.21 PCA identifies three significant components

Scree plot depicting the components generated by PCA. The first three components explain ~65% of the variance in the data

Histogram representing eigen values generated from performing PCA on randomized data matrices. The first three principal components have eigen values greater than the distribution and can be considered significant.

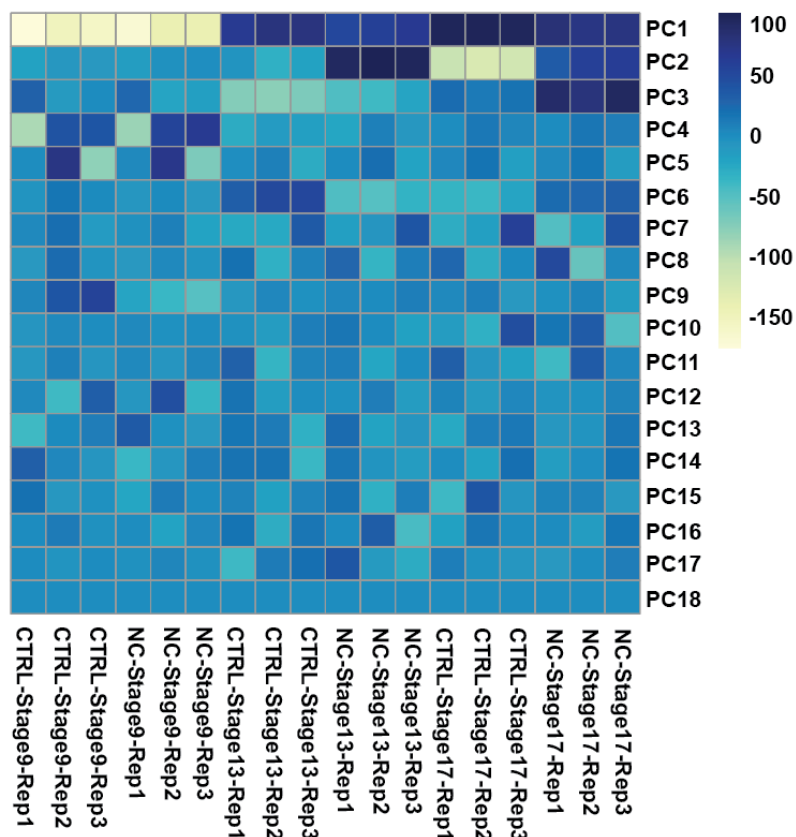


Figure 4.22 Gene expression data projected onto principal component space

Heatmap for visualization of gene expression data projected onto principal component space (dot product of the gene matrix and the eigen vector matrix). Large scale changes are seen along the first three principal components across the samples. The first principal component varies with time irrespective of the cell state while the second and third principal components vary due to cell state.

such *Mta1*, *Mbd1*, *Chd7*, *Prdm1*, *Prmt1* which might play an important role in epigenetic regulation of the neural crest stem cell state. Similarly, in the late neural crest gene module, we find genes such as *Gcn1*, *Fam212a*, *Ptk7* and others that would be very interesting to characterize the function for in these cells.

Principal Component Analysis identifies 3 significant components that explain dynamics of neural crest cell fate decisions

Our data showed that there existed set of genes that are highly correlated with the early neural crest state and a set of genes that are highly correlated with a late neural crest signature. We desired to use another independent method to identify genes that belong to an early vs late neural crest signature, as well as further explore the dynamics of gene expression changes that occur during lineage restriction of these cells. In order to do so, we decided to use principal component analysis to explore the dynamics of gene expression changes that were occurring in these cells as they made the adopted the neural crest state over developmental time (Credit: Analysis framework adapted from Dr. Simon Freedman and Dr. Madhav Mani, Northwestern University). Principal component analysis is a statistical technique for determining key variables in a multidimensional dataset and can be used to simplify the analysis and visualization of multidimensional datasets. Here, we performed PCA on the gene matrix of 30128 genes (variables) and 18 samples (observations) (Figure 4.20). PCA identified 17 principal components (eigen values) that accounted for varying degrees of the variance in the data (Figure 4.21). Parallel analysis, wherein the distribution of the eigen values generated from 1000 scrambled matrices from the data was plotted, revealed that the first three principal components can be considered significant while the rest of the principal components can be considered to be formed

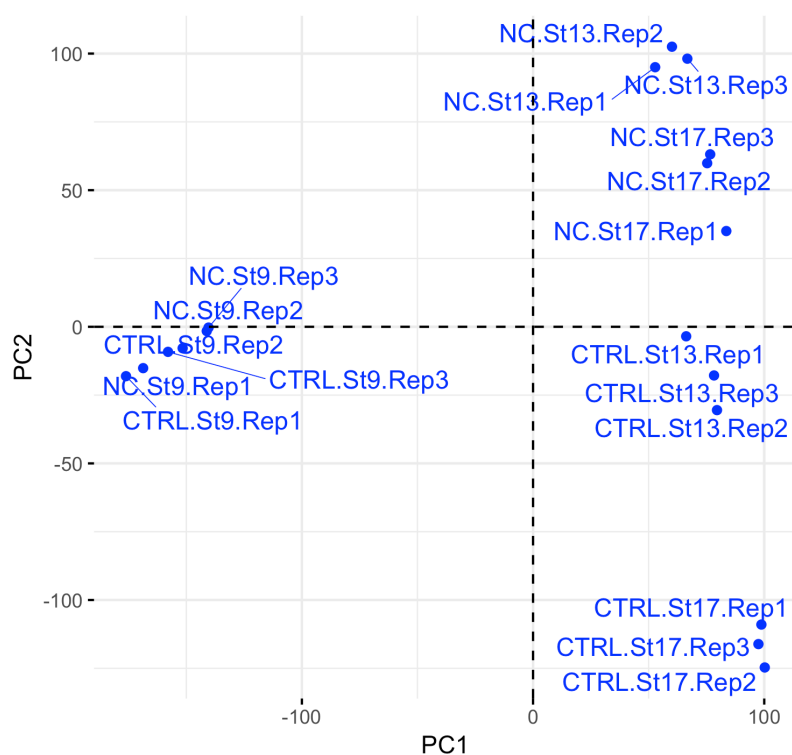


Figure 4.23 Biplot of PC1 vs PC2

Biplot for visualization of gene expression data comparing PC1 vs PC2. Large scale differences are seen along the first two principal components across the samples. The first principal component varies with time irrespective of the cell state while the second vary due to cell state.

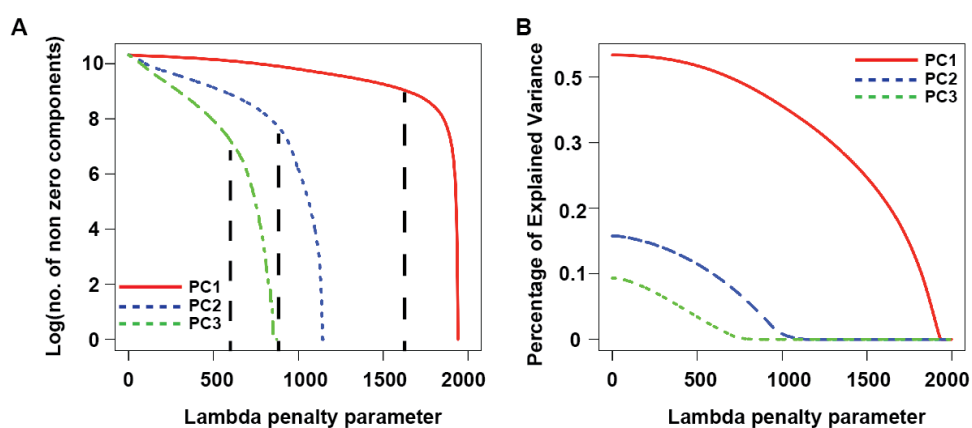


Figure 4.24 Determination of Lambda penalty parameter for sparsePCA

(A) Graph of number of non-zero components for PC1, PC2 and PC3 with varying lambda penalty parameter. Number of non-zero components decreases with increasing lambda value. At Lambda = 1600,830,680, the number of non-zero components rapidly drops to zero.

(B) Graph of percentage of explained variance for PC1, PC2 and PC3 with varying lambda penalty parameter. Percentage of explained variance decreases with increasing lambda value.

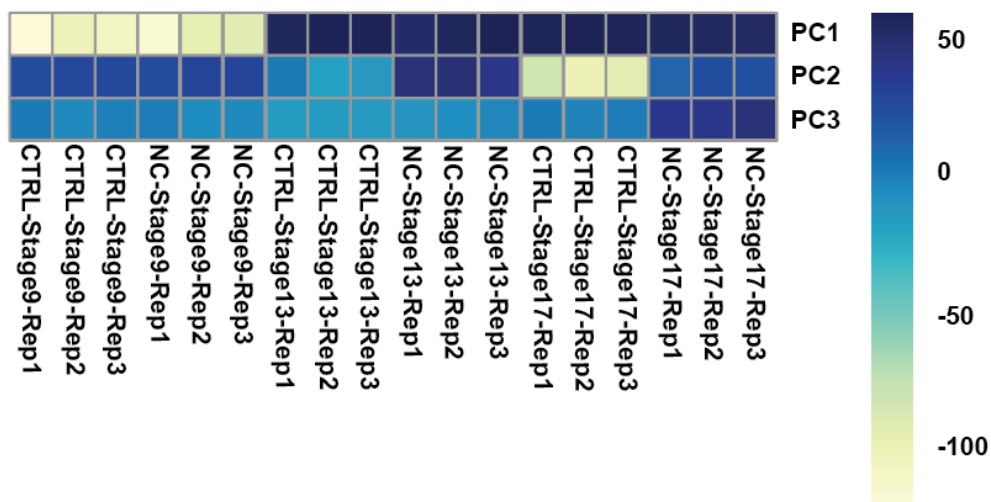


Figure 4.25 Sparse principal components from gene expression data.

Heatmap for visualization of gene expression data projected onto principal component space after sparse PCA. Principal components generated after sparsePCA appear similar to principal components from the PCA, suggesting that enforcing sparsity did not result in loss of variance in the data.

	Penalty parameter	No. of non-zero components
PC1	1600	8666
PC2	830	3888
PC3	630	669

Figure 4.26 Sparse principal components from gene expression data.

Number of non-zero components generated for each principal component after enforcing penalty parameter.

purely due to noise (Figure 4.21). These first three components explain about 65% of the variance in the data (Figure 4.21). Visualization of the data in PCA space revealed very interesting features of the first 3 principal components (Figure 4.22, 4.23). Principal component 1 (PC1) seems vary due to developmental time irrespective of whether the samples are epidermis or neural crest (Figure 4.22, 4.23). This would suggest that there are a set of genes that vary during the process of lineage restriction wherein the cell state or lineage choice does not matter. We observed that principal component 2 (PC2) described the variance in the data in direction of the early neural crest state (positive) as well a negative variance due to the formation of early epidermis (Figure 4.22, 4.23). Interestingly, the third principal component (PC3) appears to vary greatly in the direction of the late neural crest state. These data suggest that we have identified 3 principal components that are sufficient to explain a large part of variance in the dataset and we can utilize these components to explore further the key variables that contribute the most to the variance in the data.

Each principal component is a linear combination of all the variables (genes) and has the weight/loadings which give us information on how much each gene contributes to each component. While PCA helped us to identify that the first three principal components were sufficient to explain most of the variance in the data, there were still 30128 gene loadings that contributed to each principal component. In order to identify which genes contributed the most to each principal component, we performed sparse principal component analysis (sparsePCA) on the data (Credit: Analysis framework adapted from Dr. Simon Freedman and Dr. Madhav Mani, Northwestern University). SparsePCA is a natural extension of PCA which is well suited to high-dimensional data such as gene expression data where $p \gg n$. The elastic-net approach in PCA formulates PCA as a regression-type optimization problem and obtains sparse loadings by

integrating a lasso penalty into the regression criterion. This method enforces a sparsity penalty to the data that reduces some of the loadings to zero while retaining others as the ones that contribute the most to the variance in the principal components. In order to apply sparse PCA on our dataset, we first identified the lambda (λ) penalty parameter value that is optimized for explained variance and sparsity in this context. As shown, we systematically varied the λ value (0-2000) and calculated the corresponding number of non-zero components generated. The $\lambda=$ (1600,830,680) was identified to be the value just before the number of non-zero components dropped sharply to zero (Figure 2.24).

Based on the identified λ criterion, sparsePCA was performed on the dataset. The principal components obtained following sparsePCA closely resemble the components from the original PCA and gave us confidence that enforcing sparsity still retained most of the variance in the data (Figure 4.25). Interestingly, we find that after sparsePCA, the first principal component has retained about 8848 genes (Figure 4.26). This is of particular interest as it suggests that these genes are ones that change over developmental time and don't change with lineage choice. Further, sparsePCA retained 3888 genes in PC2 and 669 genes in PC3 (Figure 4.26). Fascinatingly, PC2 and PC3 several genes have important roles in neural crest formation as we had predicted from the PCA. Our results indicate that PC2 consists of a list of genes that contribute to the early neural crest signature, while PC3 consists of genes that involved in a late neural crest signature.

Comparing early and late neural crest state from sparsePCA and WGCNA

Given that, we had identified genes through independent methods that contributed to an early and late neural crest signature, we decided to compare the gene lists that we obtained from

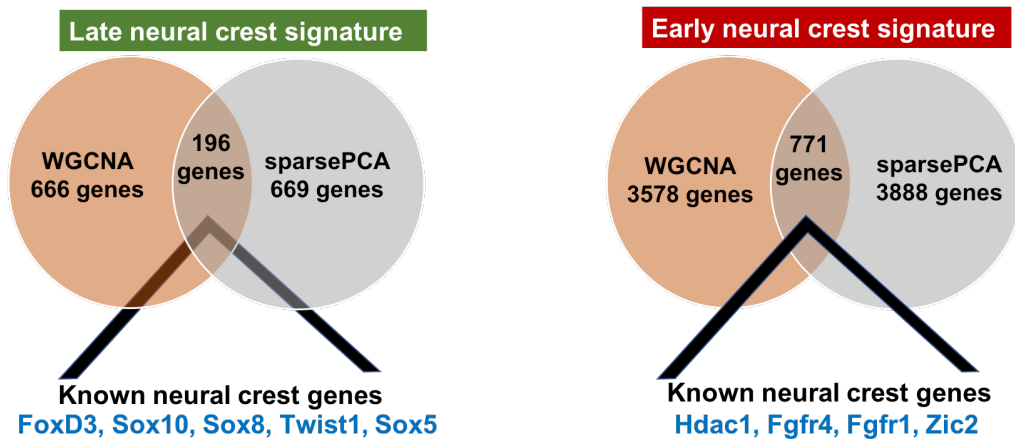


Figure 4.27 Comparing WGCNA and sparsePCA

Venn diagrams comparing early and late neural crest genes identified by WGCNA and sparsePCA. Late neural crest signature has 196 genes that overlap from the two analysis, while early neural crest signature has 771 genes common between the two datasets. These contain genes with known roles in neural crest development and several that have been previously uncharacterized to have a function in neural crest formation.

Early neural crest (neural plate border) genes from WGCNA and PCA analysis									
prmt1.L	chd7.S	supt16h.S	srrm2.S	eif2s3.S	rp2.S	Xelaev18026940m	abt1.L	ube2v2.L	aggf1.S
ilf3.S	snrpb.L	eftud2.L	iws1.S	mgc75753.S	sart3.L	rrs1.L	rbm17.L	LOC108712258	kdf1.S
hnmpab.L	MGC80893	fgfr4.S	ubtf.L	ppan.S	nup153.L	ddx52.S	dmap1.L	LOC108718145	ism7.L
prp8.S	smarce1.L	saftb.S	kif17.L	acsl3.L	bzw1.L	tpbg.S	cdca5.S	c11orf95.L	bcl7a.S
ddx5.L	ran.L	nat10.L	g3bp2.L	zfr.S	elav1.S	sp3.L	wis.L	sgol1.L	rab10.S
hnmpu.S	mat2a.S	plrg1.S	eif3.S	bcas2.L	csnk2a2.S	cdk2.L	phf8.L	LOC108706873	supg1.S
LOC108701569	alyref.L	tnpo2.L	thoc1.L	rrp1b.L	znf217.L	cp5f3.L	klhd3.S	nme4.L	kans12.L
anp32c.L	hnmpul2.S	zic2.S	ppp2r2a.S	sub1.L	chd7.L	dnajc21.L	ddx49.S	LOC398134	edrf1.S
krt18.S	sltm.S	smarcc1.S	khdrbs1.L	asx1.L	LOC108706979	mtf2.L	farp2.S	LOC108711717	cox5b.1.L
hnmpa0.L	xarp.L	dhx15.S	LOC108711518	umps.L	eif3h.L	LOC108712771	cnn3.S	cxorf56.L	arhgap29.L
nolc1.L	eftud2.S	ranbp1.S	cnbp.S	hnmpu1.S	rbbp7.L	nol11.L	zc3hc1.L	LOC108714503	kbtbd7.S
fubp1.L	ddx3x.L	rbm14.S	LOC108709624	LOC108709624	LOC108698074	hspa9.S	nup35.L	LOC108711717	LOC108712891
ddx39a.S	sf3b3.L	eif5b-like.S	LOC108711249	LOC108711249	fgfr1.S	thoc3.L	LOC108717044	LOC108714503	ddx47.S
g3bp1.S	ddx46.L	marcks.L	LOC108711485	LOC108711485	LOC398269	cacybp.S	nfrkb.L	LOC108704086	fbxo28.L
hmgb3.S	ewsr1.S	actf6a.S	ddx24.L	mcm4.L	ehmt2.L	dnajc2.S	disp1.L	rqcd1.L	prkag2.S
sr5f7.S	sf3b1.S	ncoas5.S	utp6.L	cct5b	Xelaev18006002m	cdca5.L	cbfb.L	brix1.L	arid4a.L
znf326.S	hdac1.S	snrpd1.L	srsf11.S	LOC108719141	pnisr.L	LOC100191026	msl1.S	rangap1.L	kiaa1191.S
eif4g2.L	tbl.S	ssbp3.S	LOC108698691	ccnj.L	LOC108695374	bysl.S	prc1.2.S	hdgfrp2.L	zswim6.S
hnmpa0.S	LOC108698297	srsf10.S	snu13.S	ubtf.S	rad23b.L	naa40.L	tial1.S	hirip3.L	LOC108701773
dknp1.L	nop58.L	mfpap1.L	LOC108716231	pum3.S	usp4.L	ckw15.S	LOC108712345	best2.S	wnt11b.L
smarcc1.L	top1.1.L	cdh3.L	prrc2c.S	LOC108695527	bcl7a.L	cebpz.S	c16orf72.L	ube2s.S	metap2.S
snmp70.L	LOC108696656	snrpd3.L	eif3.L	nucd.L	mdh1.L	banf1.L	stox1.S	LOC108710994	zntf19.S
sr5f3.S	rrp12.S	supt5h.S	tbl3.L	rbm39.L	LOC733425	sbno1.S	Xelaev18008693m	zbtb2.L	spop.L
sr5f1.S	srrm2.L	mcm5.L	abce1.L	dbf1.S	trim28.L	ceif1.S	fam50a.S	tpbg.L	mis12.L
sr5f6.L	rbm25.S	smc4.L	gsp1.L	e2f3.L	syf2.S	pa2g4.S	znf852.S	znf852.S	znf451.S
sox15.L	dnntt2p.L	eif3b.L	eif3.L	cadm1.S	LOC108708118	mcm2.S	gja7.S	actr8.S	c11orf58.L
hist2h3d.L	rbmx.S	niplb.L	ncapg.L	LOC108711598	letm1.L	csnk2b.L	xpo1.L	zbtb2.S	LOC108717209
fgfr4.L	snu13.L	rbm28.L	gtf2f1.L	tcf3.L	polr2e.L	tcf12.L	nup85.L	ppie.L	ttgfi1.L
LOC108708917	pdcd11.L	ppig.L	sox11.L	ddx42.L	cdon.S	prps1.L	phf5a.L	ccnj.S	Xelaev18025740m
sr5f6.S	supg2.L	znf346.L	Xelaev18014773m	LOC108713808	dek.L	cnot3.S	usp4.S	xdx28.L	cdk2ap1.L
anp32e.L	marcks.S	nap111.L	pum1.S	taf13.S	epb41.L	usp14.S	LOC108714684	LOC108698855	tkf2.S
eif4at1.S	znf346.L	LOC397764	snrnp40.L	paip2b.L	mtf2.S	snrpb2.L	polr2e.S	LOC108709577	stard3.S
tuba1c.L	api5.S	acn1.S	taf15.L	nol10.L	lyar.S	Xelaev18026939m	rrp15.S	cdca7.S	pmsb2.S
ddx23.S	LOC108719497	sic25a5.S	LOC108709846	npm3.S	cyrr1.L	xcrc1.L	snrpe.L	med29.L	pak11p1.S
xarp-like.S	tmop.L	akirin1.S	LOC108709846	nm3.S	rsd1.L	rsd1.L	zfx.S	nov2.L	trim33.L
snrpb.S	ddx3x.S	prpf19.L	tia1.L	mmd3.S	LOC108707047	gar1.S	Xelaev18002771m	paxip1.L	cks2.L
ctnnb1.L	smc1a.L	ptges3.L	ucl2.S	irx1.L	cdkn2aipn1.S	ubr7.S	ppan.L	ttc1.S	hes6.L
g3bp1.L	ppp13a.L	cdc51.L	ncoas5.L	dnajc8.S	LOC100101277	dhx33.L	phf10.S	kbtbd7.L	ism6.L
mybbp1a.L	api5.L	ddu3.L	brd7.L	rbpb4.S	smarcb1.S	csnk2b.S	plekhg3.L	vezf1.L	ism8.S
snmp70.S	akap12.L	tcf3.S	cbfb.S	baz1b.L	krr1.S	psmc4.L	exoc9.L	trim36.L	meaf6.S
ran.S	eif4a3.S	ncapd2.S	rarg.S	szrd1.L	LOC108697536	rexo4.L	med15.L	ubi5.S	pomp.L
sr5f4.L	prpf6.L	LOC108702873	LOC108709681	rxo4.L	LOC108704127	rrs1.S	psma2.S	c19orf52.S	pomp.S
LOC108716102	pnn.S	mta2.L	LOC108717002	snmp27.S	LOC108704127	ccnt2.L	zc3h15.L	vps72.S	ackr3.L
nucks1.S	hnmpu.S	snrpd3.S	sumo2.S	u2surp.S	snmp27.S	depor.L	mi1b1.S	bard1.L	rbm15.L
hdacl.L	sltm.L	anp32e.S	ttf1-like.S	hnp2.S	snmp27.S	morf411.S	ss18.L	snx22.L	mrpgb.L
znf638.S	mcm2.L	cnns3.L	zmyzm4.S	pest1.S	snmp27.S	tcp1.S	LOC108714958	cd82.L	wdr46.S
etf1.S	cnns3.L	asx1.L	LOC733385	cbx5.S	poir2f.S	thg11.L	chd1.L	thoc7.S	paxip1.S
cwc25.L	ppid.L	kpnad.L	prpf4.L	psmd8.L	kn1.L	pbx1.S	LOC108717071	stx2.L	mkks.L
fbxo11.L	ppp3r1.L	eif2s3.L	prpf4.L	eif3g.L	LOC108702709	LOC108702709	shq1.L	psma6.L	chpt1.L
ism3.L	zzz3.L	apex1.L	saal1.L	mgf.L	LOC108715607	LOC108715607	ctnnb1.L	snrpg.L	cdca4.L
sox11.S	trap1.L	lypia2.L	lypia2.L	ddx18.L	LOC108715607	LOC108708901	phf5a.S	rbm12b.L	gppb111.L
fam64a.L	gtf2f2.S	pbx1.L	gabpa.L	cdc25.L	MGC78849	vars.S	med16.L	tbl1xr1.L	kbtbd2.L
eif4e1b.S	mfaf23.S	gabpa.L	gabpa.L	abcf3.S	LOC108703077	msant4d.S	abl1.S	znhit6.S	bt3f14.L
mcm4.S	rmmt.L	pel11.L	gpm.S	apola.L	LOC108703077	nup62.L	mars.L	gpr180.S	lias.S
LOC108715148	slbp.S	senp5.L	tcf7.S	slc7a3.L	LOC108703077	ppp4r2.S	arglu1.L	spp13.S	rmf2.L
wrm.L	hmgb1.L	dhx16.L	znrf82.L	rbpb4.L	LOC108703077	LOC108703077	gpr180.S	psmd7.S	MGC145518.L
cdk13.L	diexf.L	cdca4.S	snrpd1.S	eif3k.S	Xelaev18002950m	ddx51.L	gpr180.S	nkap.S	ssca1.S
nol11.S	bxn14a.S	srfbp1.L	snrpd1.S	c21orf91.L	cadm1.L	ddx42.S	arglu1.L	LOC108707719	rrp7a.S
rrp9.L	msl1.L	snrpd1.L	snrpd1.S	snn.L	LOC108707719	actk2.L	gpr180.S	noc4.L	c12orf43.L
rgma.L	cp5f3.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	aurka.S	gpr180.S	hsf2.L	hes2.L
psma7.L	skp2.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	ngdn.L	gpr180.S	eif6.S	
LOC108708189	scat8.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	gsp11.S	gpr180.S	sp5b4.L	
LOC108714318	sic39a6.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	cepi1.L	gpr180.S	phf5a.L	
chchd4.S	eif6.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	serinc5.S	gpr180.S	ck2.S	
nkx2.L	tra2a.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	glt2h4.S	gpr180.S	emg1.S	
ddx10.L	xpo5.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	rnf113a.S	gpr180.S	prkag2.L	
c14orf79.S	hnmpd.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	bcor.S	gpr180.S	exosc6.L	
vipas39.L	serh2.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	MGC83090	gpr180.S	zbtb33.L	
fam58a.S	imp3.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	ar3.L	gpr180.S	sap301.S	
gan.L	LOC108712797	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108696403	gpr180.S	smim3.S	
chchd4.L	Xelaev18032192m	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	fam206a.S	gpr180.S	mrp137.L	
mett10.S	tc39c.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108706892	gpr180.S	int5.L	
c14orf166.L	sic10a7.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108705719	gpr180.S	nemp1.S	
LOC108696673	rab25.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108714527	gpr180.S	fbxo5.L	
ube2v2.S	mrp140.S	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108695505	gpr180.S	LOC108719636	
taf12.L	LOC108706849	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108715571	gpr180.S	LOC108698854	
wdr89.L	tfb2m.L	snrpd1.L	snrpd1.S	LOC108712074	LOC108707719	LOC108715571	gpr180.S	desi2.L	
								nuf2.L	

Table 4.1 Early neural crest (border) genes from WGCNA and sparsePCA

Late neural crest (progenitor) genes from WGCNA and PCA analysis				
sox8.L	fhl3.L	slc4a7.L	rnf125.L	map2k4.L
foxd3.L	arhgap28.S	nr2f1.L	rtn2.L	fras1.L
sox10.S	nos1.L	b4galt4.S	LOC108719031	LOC108706949
LOC108696179	rock2.L	fam83g.S	LOC108702063	pdcd2.S
pcdh8l.S	megf8.S	epb41l1.L	Xetrov90003447m.S	LOC108707496
sox8.S	ptcd3.L	uri1.S	wdr86.L	patz1.L
LOC100191024	ltv1.L	wnt10a.L	fndc3a.L	elp2.S
fhl3.S	nrp2.S	prelid3b.L	sept5.L	dvl1.L
cyp26c1.S	setd6.L	padi4.S	LOC108715811	mmp28.L
fam212a.L	urb2.L	asb9.L	LOC108711787	csgalnact2.L
pdgfra.S	LOC108710100	ercc1.L	trnt1.L	pbd1.S
sox10.L	LOC108698799	polr3d.L	parl.L	kif2c.S
twist1.S	MGC84224	hdac7.S	mef2c.L	fip1l1.L
tnfrsf19.L	txndc5.L	sncaip.L	s1pr1.L	LOC108716708
LOC108699362	psmd12.L	tmem222.L	aimp2.S	pfkm.S
twist1.L	dchs1.L	nckap5l.S	bccip.L	aamp.S
tuba4a.S	sema3d.L	LOC108716914	LOC108704156	psmd2.L
pdgfra.L	ppp1r7.L	ptprd.S	mettl17.L	LOC108709525
rapgef2.L	mpdz.L	LOC108712586	LOC108713445	kif4a.L
prtg.L	cldnd1.L	zfyve16.L	brat1.L	prrc2a.L
ptk7.L	mfbp2.L	LOC100492888-like.L	hip1.L	LOC108714364
myo10.2.S	stoml2.L	tubb2b.S	elf2b4.L	sdhb.S
mafB.S	cltcl1.S	adam17.L	ssbp4.S	serpinh1.S
LOC108713443	ednra.S	elf2b2.L	iffo1.L	egr2.L
rapgef2.S	nid2.S	st3gal5.S	LOC108713835	sox5
tnfrsf19.S	elmo1.S	gabap1-b	LOC108695783	LOC108696525
mafB.L	LOC108718707	mitf.S	h2afy2.S	LOC108709882
lrpprc.L	LOC108716200	stx1b.L	anos1.S	hapln3.L
traf4.L	nxn.L	tshz3.L	LOC108709293	slit2.S
shb.S	zfp36l2.L	tmem119.L	LOC108718657	lars.L
urb1.L	wnk1.S	afap1l1.S	iqsec1.L	slc39a5.L
LOC108698051	c10orf54.S	LOC108701483	prdx1.L	acss3.L
LOC108710030	robo1.L	aifm3.L	LOC108717348	
flrt3.L	mef2c.S	surf2.L	pdcd2l.L	
ror1.L	wsb1.S	aimp1.S	mrpl46.L	
nrnc.L	LOC108717147	mettl13.S	LOC108717763	
prtg.S	gpn1.L	fancm.L	LOC108707168	
LOC108701464	cib2.S	tenm4.L	samd11.S	
abtb2.S	nfasc.S	sqrll.L	psip1.S	
cept1.S	mycn.S	tmem116.S	Xelaev18016214m	
agpat3.S	LOC108717965	cadps.L	ldlr3d3.L	

Table 4.2 Late neural crest (progenitor) genes from WGCNA and sparsePCA

these two methods. Interestingly, for the early neural crest signature (neural plate border), we find 771 genes that are identified commonly between these two methodologies while 196 genes in the late neural crest signature (neural crest progenitor) (Figure 4.27) (Table 4.2). These include genes such as *HDAC1*, *Fgfr1/4*, *Zic2* and others that have known roles in neural crest stem cell maintenance (Figure 4.27) (Table 4.1). Fascinatingly, we can identify several new genes that have previously uncharacterized roles in neural crest stem cell maintenance, such as *Prmt1*, *Sox15*, *Znf326*, *Ddx5*, *Chd7*, *ILF3*, *Hmgb3* and others (Table 4.1). This suggests that this methodology can be used to identify genes that might be contributing to the maintenance of the neural crest stem cell state as well as differentiate genes that are important for later neural crest specification.

Discussion

Understanding the gene regulatory network that controls the pluripotency of neural crest cells will give us novel insights into the mechanisms of maintenance of stem cell attributes which has broad implications for stem cell therapy and regenerative medicine. Neural crest cells have the unique ability to retain their stem cell attributes early in development yet undergo lineage restriction later in order to give rise to a diverse array of derivatives. This interesting facet of neural crest stem cells suggest that these cells must possess a gene regulatory circuitry that allows not only for maintenance of pluripotency of these cells for the precise developmental time, but also controls the exit from the pluripotent state and integration into the gene regulatory network that controls neural crest specification. While decades of research have characterized the NC-GRN that is necessary for the specification and later differentiation of neural crest cells, we still have limited knowledge regarding the regulatory landscape of control of pluripotency of

these cells. Here, we use a genome-wide transcriptomic approach to characterize the circuitry that contributes to the maintenance of pluripotency of the neural crest. Our data provides a unique paradigm for us to investigate this process in greater detail and explore further the temporal control of early neural crest development. We identified that large changes in gene expression occur over developmental time and while lineage choices are being made in *Xenopus* animal cap explants. Interestingly, we found that the transcriptional landscape also changes dramatically during the process of neural crest development. We identified that there exists an early and late neural crest signature, with early NC genes involved in stem cell maintenance while late NC genes play important roles in EMT and lineage restriction. Strikingly, we found that there are a set of genes that change over developmental time (during restriction) which do not conform to the formation of a particular cell fate. This study has provided us with a unique framework within which to identify candidate factors with novel roles in the neural crest stem cell maintenance which we can follow up mechanistically for their function in the process of neural crest formation.

Our data provides two critical snapshots during neural crest development, an early neural crest state (Stage 13) where the cells have taken the neural crest identity but are yet partially pluripotent and the late neural crest state (Stage 17), when the neural crest is more specified and getting ready to undergo EMT. Our initial analysis revealed a complex relationship between the early and late neural crest state, with several shared genes (*Sox9*, *Myc*) that are necessary to maintain the neural crest identity. Indeed, another study recently identified that these genes are significantly higher in the premigratory neural crest cells when compared to migratory neural crest cells suggestive of their role in stem cell maintenance (Lignell et al., 2017). Interestingly,

this study suggests that there is a core cluster of genes that are necessary for neural crest formation in chick (Lignell et al., 2017).

Our data also shows that there exists some degree of synergy between pluripotent blastula cells and neural crest cells. Interestingly, as we had previously seen via *in situ* hybridization, there are several shared genes between neural crest cell and pluripotent blastula cells (Buitrago-Delgado et al., 2015). Through our transcriptomic analysis, we find that the core pluripotency genes, *Oct25 (Oct4)*, *Vent2 (Nanog)* and *Sox2*, are highly enriched in the neural crest cells suggestive of a shared transcriptional landscape with the pluripotent blastula cells. Several neural crest factors such *FoxD3*, *Myc*, *Snail*, *Pax3* and others are expressed in blastula cells. However, the levels of expression of these factors is markedly different between these two cell types. This would suggest that while the same transcription factors are functioning in these two cell types, their contributions to the process might be different and this would be an interesting avenue for future investigation.

Our data also suggests that fundamental differences exist between the pluripotent blastula cells and early neural crest cells. From our initial analysis, we find that there is a large number of genes that are differentially expressed between the neural crest explants at early neurula and blastula stages. This would suggest that by early neurula stages (Stage 13), the neural crest explants have begun to deviate from blastula explants. This could be attributed to the timing i.e. a slightly earlier stage of neural crest explants might be closer to the pluripotent state and it would hence be interesting to perform a time course to assess the time point at which the transcriptional landscape of neural crest cells begin to diverge from blastula cells. Our data also suggests that there a set of genes that change over developmental time, and differences in gene expression between early neural crest and blastula explants can be attributed to this as well.

The advent of genomic technologies has provided us with the ability to assay in an unbiased manner the transcriptional landscape of the cell at various stages of cellular differentiation. This, in turn, provides us with the ability to generate hypothesis, and identify novel factors involved in the process as well as characterize the big picture, of how these different factors relate to one another to form a network of regulatory control. Combining differential analysis methods with more rigorous computational analysis such as WGCNA and PCA/sparsePCA provides us with an unbiased framework to investigate the dynamics of the processes. We have used this study to gain extensive information regarding the genes involved in neural crest development and can be used to identify factors that are required for the control of pluripotency. We hypothesize that genes that contribute to the early neural crest signature would likely be involved in the stem cell maintenance. The identification of new and novel factors involved in neural crest formation raises additional questions regarding the mechanism through which these factors function in neural crest formation, and how they integrate into the gene regulatory network that allows for neural crest specification. These studies will further our knowledge regarding the mechanisms through which pluripotency is spatially and temporally controlled in these cells and give us tremendous insights into the processes that led to the formation of the neural crest and the evolution of vertebrates.

Materials and Methods

Embryological methods

Wildtype *Xenopus laevis* embryos were obtained using standard methods and staged according to Nieuwkoop and Faber (1994). For animal cap assays, ectodermal explants were manually dissected at early blastula (St. 8-9) from control or embryos injected with Wnt/Chd

mRNA bilaterally at the 2-cell stage and cultured in 1x MMR until sibling embryos reached the denoted stage. Embryos/explants were aged using a predetermined timing and temperature regimen (Credit: Experimental paradigm adapted from Kristin Johnson). Briefly, *in vitro* fertilization was performed at 2pm and fertilized embryos were grown at 14°C till blastula stages (embryos were briefly maintained at RT (18°C) during injection). At precisely 8:30am, embryos were brought to RT (18°C) and animal caps explants were dissected. At 9:45am, the Stage 9 (blastula) sample was collected. Animal cap explants and control embryos were moved to 20°C to age to Stage 13 (collected at 5pm) and to 14°C to age to Stage 17 (collected the next morning). mRNA for microinjection was *in vitro* transcribed from a linearized DNA template using the SP6 mMessage mMachine Kit (Ambion). Manipulated embryos/explants were processed for *in situ* hybridization by fixing in 1x MEMFA and dehydrating in 100% methanol. *In situ* hybridization was performed using digoxigenin labelled RNA probes and developed using BM Purple substrate (Roche) (LaBonne & Bronner-Fraser, 1998). Results shown are representative of at least three independent experiments.

RNA-sequencing sample preparation

RNA isolation for RNA-seq was performed using TRIzolTM-LiCL method. Animal cap explants from control and Wnt/Chd injected embryos were collected at stages 9, 13 and 17 (twenty-five caps per condition) in 500µL TRIzol and flash frozen in liquid nitrogen. 500µL of TRIzol was then added to each sample and caps were homogenized by pipetting and vortexed for 15 seconds. RNA was then isolated by chloroform extraction followed by LiCl and ethanol precipitation. Purified RNA was resuspended in 30µL of molecular grade water (one biological

replicate for sequencing). Purified RNA was submitted to Northwestern's Sequencing Core for library prep (TruSeq mRNA Library Prep kit) and sequencing (75bp single end on Illumina NextSeq).

Animals

All animal procedures were approved by the Institutional Animal Care and Use Committee, Northwestern University, and are in accordance with the National Institutes of Health's "Guide for the Care and Use of Laboratory Animals".

Bioinformatics Analysis: RNA-sequencing Analysis

The obtained reads were checked for quality using FAST-QC (Babraham Bioinformatics). Alignment and counting of reads was performed using Bowtie2 and RSEM. Briefly, a *Xenopus* reference transcriptome was generated using rsem-prepare-reference from the *Xenopus Laevis* 9.2 (from Xenbase) with the GTF file for gene annotation. The reads were aligned to transcriptome using Bowtie2 within RSEM (Xenbase) and counted using rsem-calculate-reference. TPM/count data was obtained from RSEM. Differential analysis was done in R using the DESeq2 package applying standard log-fold shrinkage procedures (Love et al., 2014). Genes were considered significantly changed for $\text{padj} < 0.05$, and log fold change threshold was set as 1.

For WGCNA, the normalized count data was obtained from DESeq2 and the top 15000 genes with highest variance across all samples were used for the analysis. WGCNA was performed in R using the WGCNA package according to the pipeline detailed in the online tutorial. For Principal Component analysis, the TPM values for all genes/samples was obtained from RSEM, and the data was filtered to only include genes with $\text{TPM} > 0.5$ in at least one sample.

The filtered geneset of 30128 genes was used to perform PCA using `prcomp` function in R. For sparsePCA, the `arrayspc` function from the `elasticnet` package was used. All bash and R scripts used in the analysis are found in the appendix (Chapter 6).

Chapter 5

General Discussion

The neural crest is unique embryonic stem cell population that contributes to a multitude of structures within the vertebrate embryo. The acquisition of these structures has added a layer of complexity onto the simple chordate body plan allowing for a larger cranial and jaw structures, more advanced brain, as well skin pigmentation that has led to the evolution of vertebrates (Le Douarin & Dupin, 2012; Prasad et al., 2012). Since its discovery 150 years ago, neural crest cells have fascinated biologists because they seemingly defy the embryonic paradigm of progressive lineage restriction and give rise to derivatives from multiple germ layers. Neural crest cells form a diverse array of derivatives *in vivo* and *in vitro* that have ectodermal origins such as neurons, glia and melanocytes, and others that are considered mesodermal in nature such as chondrocytes, adipocytes and smooth muscle cells (Bronner-Fraser & Fraser, 1988; Sauka-Spengler & Bronner-Fraser, 2008). Recent work from our lab has also identified that neural crest cells have the ability to respond to endoderm inducing cues *in vitro* (Buitrago-Delgado et al., 2015). This remarkable potency of neural crest cells has raised fundamental questions regarding the mechanisms through which neural crest cells maintain their stem cell attributes. Understanding the control of neural crest stem cell state will give us novel insights regarding the mechanisms through which stem cells are maintained during embryonic development and is of utmost significance to advance stem cell therapies and regenerative medicine.

Traditionally, it was thought that neural crest cells gained developmental potential through an inductive event, i.e. a subset of cells within the ectoderm regained their stem cell attributes which led to formation of neural crest. However, recent work from our lab has provided evidence to support an alternate model, that suggests that neural crest cells arise through a retention of stem cell attributes, i.e. a subset of blastula cells have the ability to stay

pluripotent while the rest of embryo undergoes lineage restriction, and these cells give rise to the neural crest (Buitrago-Delgado et al., 2015). This would suggest that neural crest do not travel back up Waddington's landscape and gain developmental potential, but instead have the unique capability to delay the onset of lineage restriction (Figure 5.1) (Hoppler & Wheeler, 2015). This parsimonious model fits well with our classical view of embryonic development and progressive lineage restriction, but also raises many questions regarding the mechanisms utilized by the cells to evade cues that promote lineage restriction within the embryo. What are the factors that are involved in the maintenance of the pluripotent stem cell state and what are similarities and differences exist between neural crest cells and pluripotent blastula cells from which they are derived? My dissertation aims to address these questions and add to our current knowledge regarding the control of pluripotency of the neural crest stem cell state. The findings presented here provide novel insights into the epigenetic control of neural crest stem cell maintenance and identifies the mechanisms through which these epigenetic factors contribute to pluripotency. Specifically, I demonstrate that HDAC activity and histone acetylation are critical for the formation of the vertebrate neural crest, and I explore further the gene regulatory circuitry that controls the neural crest stem cell state.

HDAC activity is essential for blastula and neural crest cells

Most of our knowledge regarding the epigenetic control of the stem cell state has been gained through experiments conducted in human and mouse ESCs and iPSCs. While studies in ESCs have given some cellular insights regarding the pluripotent epigenetic state, these *in vitro* studies provide us with a very poor understanding of the functional role of HDAC activity and histone acetylation in maintaining pluripotency *in vivo* during embryonic development. Several

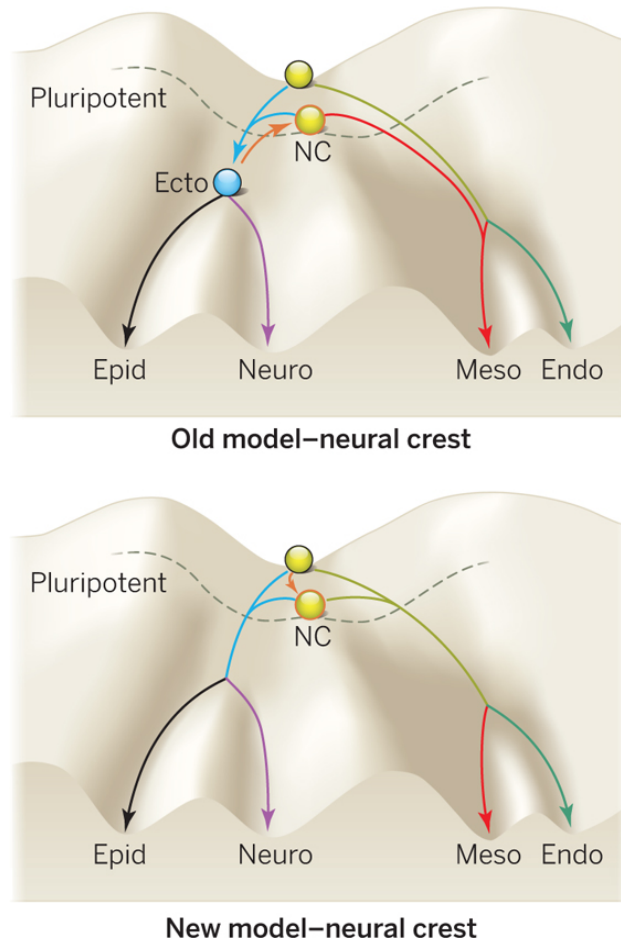


Figure 5.1 Old and new model of neural crest formation

Old and new model of neural crest formation projected on Waddington's landscape. Traditionally, neural crest cells were thought to gain developmental potential through an inductive event. New model indicates that neural crest cells arise through retention of stem cell attributes.

discrepancies have been found when comparing mESCs and hESCs highlighting the importance of timing (naïve vs primed pluripotent state) for understanding dynamic cellular context (Weinberger, Ayyash, Novershtern, & Hanna, 2016). However, as pluripotent cells in the embryo exist only transiently, knowledge regarding pluripotency and lineage restriction *in vivo* remains elusive. In this regard, neural crest cells provide us with unique opportunity to explore the nature of pluripotency *in vivo* in a cell type where it is sustained and can provide insights into the broader mechanism controlling pluripotency in a developing embryo.

In Chapter 2, I demonstrate that HDAC activity is essential for neural crest stem cell maintenance. Interestingly, inhibition of HDAC activity results in a loss of expression of neural crest markers and a failure to form the neural crest. Further, I identified that increased HDAC activity enhances neural crest formation. This is of particular significance as it is the first study to identify the requirement of HDAC activity in early neural crest formation. Given that neural crest cells are derived through the retention of pluripotency of blastula cells, I tested to see if HDAC activity is also required for the maintenance of pluripotency of these cells. Strikingly, I found that HDAC activity is necessary for the expression of pluripotency markers *Oct25 (Oct4)*, *Sox3* and *Vent2 (Nanog)* in blastula cells. This is consistent with previous reports that have shown that HDACs function to positively regulate gene expression in pluripotent cells. In these studies, it was found that HDACs functioning through the larger HDAC-Sin3A complex is necessary to drive pluripotency gene expression in ESCs and iPSCs (Baltus et al., 2009; Saunders et al., 2017). However, not much is known about the role of Sin3A and other complex members in the maintenance of pluripotency within the embryo. It would be very interesting to identify if HDACs are functioning with the Sin3A complex in this context. Strikingly, Sin3A has been found to be expressed in the neural crest forming regions in tadpole embryos,

suggestive that it is expressed in the right place and time to be involved in neural crest formation (Damjanovski, Sachs, & Shi, 2000). It would be fascinating to characterize the expression of Sin3A and other members of the complex during early *Xenopus* development and identify potential mechanistic links between HDAC and the Sin3A complex activity in blastula and neural crest cells.

In addition to driving pluripotency gene expression, I also demonstrate that HDAC activity is essential for the competency of blastula cells to respond to inductive cues to form different cell fates (including neural crest). I found that HDACi results in aberrant expression of markers of various lineages that resulted in a loss of pluripotency and an inability to commit to any lineage. This would suggest that HDAC activity is necessary to regulate appropriate gene expression in pluripotent blastula cells and neural crest cells. Interestingly, in Chapter 3, through genome wide RNA-sequencing of animal cap explants with/without HDAC activity, I identified that HDAC inhibition results in equal upregulation and downregulation of gene expression. Strikingly, this represents only 2% of total genes, suggestive that HDACs are performing a very specific and targeted role in these cells. This result raises interesting questions regarding the mechanisms through which HDACs are regulating a specific subset of genes within the embryo in a context dependent manner despite being ubiquitously expressed within the embryo and is an exciting avenue of future investigation. In particular, it would be interesting to identify how HDACs are positively regulating some genes while negatively regulating others. Through chromatin immunoprecipitation experiments, I elucidate that levels of histone acetylation are low at the genes that are upregulated by loss of HDAC activity which would suggest that HDACs are functioning to maintain histone acetylation low at these genomic loci, thereby preventing their transcriptional activation. However, this does not discount that HDACs could be functioning

through an alternate mechanism to positively regulate some genes as previously reported and must be investigated further.

One mechanism through which HDACs could be regulating genes differently is through its interaction/recruitment by context specific transcription factors. HDACs have been shown to interact/ form complexes with various transcription factors to mediate their function. For instance, HDACs have been shown to be recruited to the E-cadherin promoter by Snail2 during EMT (Cano et al., 2000). Interestingly, in Chapter 6 (appendix), I demonstrate that HDAC1 and HDAC2 interact with Snail1 and Snail2 in a temporally dependent manner during early *Xenopus* development. This suggests that during development HDACs potentially interact with different transcription factors in a context dependent manner in order to carry out its many functions and is likely the mechanism through which HDACs are controlling a subset of genes differently than other genes. It would be fascinating to identify all the binding partners of HDACs in these cells, and further characterize if these TFs are recruited to the genomic loci of genes that are up/downregulated after HDACi. I performed motif analysis comparing the transcription start site (TSS) of these genes, and it was not immediately apparent through a single transcription factor motif if there was differential binding of TFs and this warrants further investigation. These studies would provide us with mechanistic insights into how HDACs might be regulating a specific subset of genes in stem cells.

While the primary substrate for HDACs is histones, we must not overlook that HDACs also acetylate other proteins within the cell. HDAC1/2, in particular, are found primarily in the nucleus and most likely function on histones (Seto & Yoshida, 2014). My work identified that HDAC1/2 are most likely to be the family member involved in this context, however, we cannot disregard that other HDACs might also play important roles in blastula cells and might be

necessary to acetylate other proteins. Indeed, post translational modification of proteins is an excellent method to regulate transcription factor function. It has previously been shown that Sox9 is sumoylated and this regulates its function during neural crest formation (P.-C. Lee et al., 2012). In this regard, it would be very interesting to identify which proteins undergo post translational acetylation in pluripotent blastula cells for which HDACs may be important. It would be very intriguing to perform mass spectrometry to detect all acetylated proteins in pluripotent blastula cells and neural crest cells and compare them to acetylome seen after HDAC inhibition to identify substrates of HDACs in these cells. One hypothesis could be that potentially HDAC function directly through transcription factor interaction to regulate expression of lineage markers while indirectly activate transcription of other genes by acetylation of transcription factors. This is a very thought-provoking hypothesis and these answers will further our understanding on how HDACs, a ubiquitous chromatin remodeler with a myriad of cellular roles can function in a context specific manner.

Histone acetylation and stem cell maintenance

Histone acetylation plays a complex role during stem cell maintenance and lineage restriction. Studies in ESCs (mouse and human ESCs) have identified that there are several histone acetylation marks that are critical for stem cell maintenance. While histone acetylation marks has been shown to mark important developmentally regulated genes in ESCs in cell culture as well during early embryonic development, a clear picture has not yet emerged regarding the role of maintenance of pluripotency during development (Creyghton et al., 2010; Karmodiya et al., 2012; P. Liu et al., 2015). In Chapter 2, I demonstrate that the pluripotent state is associated with low levels of histone acetylation and the levels of histone acetylation increase

as cells proceed from a pluripotent to lineage restricted state. While this is surprising since high levels of acetylation have been typically considered to be associated with more open chromatin and a pluripotent state, this is consistent with our previous findings that HDACs are required for stem cell maintenance suggesting that they are functioning to keep histone acetylation to maintain the pluripotent state (Chapter 2). My data suggests histone acetylation is maintained low at the genomic loci of lineage specific genes in pluripotent cells (by HDACs). This would suggest that the traditional model that depicts the open chromatin landscape in pluripotent cells that allows for these cells to give rise to all cell types might be too simple. In reality, it appears that the epigenetic state of stem cell state is fairly dynamic, and there is tight epigenetic control of gene expression: pluripotency genes are activated and lineage genes are suppressed.

Given that neural crest cells are derived from the pluripotent blastula cells, I wondered if HDACs functioned to maintain histone acetylation low in these cells as well. Fascinatingly, I found that low levels of histone acetylation are a shared feature of blastula and neural crest cells, and that HDACs are performing a similar function in these two related cell types. This exciting finding that HDAC activity is an essential and shared requirement of blastula and neural crest cells begs the question regarding the mechanisms through which it is controlling pluripotency in both these cell types and its temporal function as cells proceed from pluripotent to a neural crest state. It would very interesting to profile the histone acetylation in neural crest cells through ChIPSeq and compare the genome wide enrichment of H3K9Ac and H3K27Ac in neural crest and blastula cells. Based on our previous results we would hypothesize that histone acetylation would be high at the loci of pluripotency genes and some neural crest genes, but predominantly histone acetylation would be maintained low genome wide due to HDAC activity. Such studies

would give us an in-depth understanding regarding the epigenomic landscape of neural crest cells compared to blastula cells and would be a fascinating question to pursue.

While it is true that HDAC activity is critical for stem cell maintenance, we also find that low levels of histone acetylation are necessary to maintain the pluripotent stem cell state. In Chapter 3, I elucidate the importance of low levels of histone acetylation as blocking BRD protein activity (reader of histone acetylation) results in a loss of pluripotency and neural crest formation. This is particularly fascinating as it is unexpected to see that an increase in histone acetylation as seen due to TSA treatment, has the same effect as the loss in the ability of a cell to read the histone acetylation. Curiously, we find that while TSA and IBET treatment has the same broad effect on the embryos, HDACs and BRD proteins regulate pluripotency through distinct mechanisms. While HDACs are controlling the expression of lineage markers till it is an appropriate time for them to be expressed, it is likely that BRD proteins are functioning at least in part by regulating transcription during the mid-blastula transition. It would be interesting to further explore how BRD proteins are regulating a specific subset of genes in these cells. Not unexpectedly, we find that the genes regulated by IBET have high H3K27Ac enrichment suggesting that BRD proteins are being recruited to these sites. However, we still do not have a clear picture of how these proteins are regulating a specific subset of genes during MBT. It is likely that they might be functioning through an interaction with context specific transcription factors. We found enrichment for number of TF motifs in the promoter regions of genes regulated by IBET, and it would be useful to further identify if these TFs interact with BRD proteins and if they show differential binding at the promoter regions of the genes affected by IBET. It would be also interesting to further characterize the targets of BRD protein activity in neural crest cells and compare them to those in the pluripotent blastula cells.

Our data that low levels of histone acetylation are important for the maintenance of pluripotency during development suggest that there is balance of HDAC activity and HAT activity that is necessary to maintain the appropriate levels of histone acetylation and at the right loci. Interestingly, while our work has identified that HDAC1/2 are the players that are required for neural crest stem cell maintenance, we still do not have much information regarding the HATs that are involved in the process. HATs, such as p300/CBP and Gcn5, have been previously implicated to be necessary for stem cell maintenance (Fang et al., 2014; L. Wang et al., 2018). It would be fascinating to characterize the HATs that are involved in the process of maintenance of pluripotency of blastula and neural crest cells. Given the tight control of histone acetylation during the process of neural crest formation, it is likely that HDACs and HATs might be working dynamically, and in a highly coordinated fashion in order to regulate the exact levels of histone acetylation and at specific loci in these cells and is an intriguing hypothesis to explore further.

The epigenomic landscape of embryonic development

While the primary focus of the work in this thesis has been histone acetylation, the importance of other histone marks in the maintenance of pluripotency must not be forgotten. Histone methylation, like histone acetylation is under tight control during stem cell maintenance and lineage restriction, and histone methyl transferases and histone demethylases play an important role regulating the stem cell state. Histones are methylated frequently on both lysine and arginine residues, and several of these marks have previously been shown to be important for pluripotency. Top among these are H3K4me3 and H3K27me3 that have been implicated separately and together to be involved in stem cell maintenance in ESCs and iPSCs. The

epigenomic landscape of ESCs has been fairly well studied, and several unique features such as poised promoters and bivalent promoters or active enhancer signatures have been identified (B. E. Bernstein, Mikkelsen, Xie, Kamal, Huebert, Cuff, Fry, Meissner, Wernig, Plath, Jaenisch, Wagschal, Feil, Schreiber, & Lander, 2006b) (Azuara et al., 2006) (Spivakov & Fisher, 2007). However, much less is known regarding the epigenomic landscape during early embryonic development. In Chapter 3, I use an unbiased mass spectrometry approach to characterize the dynamics of abundance of various histone modifications during embryonic development. I identify histone modifications that are high in the pluripotent embryo, and that decrease in abundance as the embryo begins to gastrulate and others that are high in the lineage restricted embryos. This suggests that there exists an epigenomic signature that corresponds to a pluripotent state with the different modifications present at varying levels. Interestingly, in our mass spectrometry dataset no histone phosphorylation, ubiquitination or biotinylation was detected suggestive that these marks are not present in high concentration in *Xenopus* embryos. The caveat to this study is of course there are several different populations of cells within the embryo and we do not get a clear picture regarding the specific abundance of different histone modifications in a particular cell type. An interesting question raised by this work is that does abundance of different histone modifications change with each cell fate? Are individual modifications important or is it a combinatorial presence/absence of certain modifications that is required to direct cell fate? Indeed, the identification of bivalent promoters (H3K4Me3, H3K27me3) suggest that it might be a combinatorial effect of multiple modifications and this warrants further investigation. Further, it would be very interesting to understand the dynamic changes that take place during lineage restriction by performing a similarly timed experiment in aging animal cap explants. This would give us a platform to better analyze the changes in

abundance of histone modifications over time, but also give us the ability to compare the changes in pure population of pluripotent cells to the whole embryo. Further, by inducing different cell fates, we can identify if there are particular modifications that enriched in a certain cell fate vs another. In particular, it would be very interesting to study the epigenomic landscape of neural crest cells and compare them to the state of the pluripotent blastula cells. This would provide us with a better understanding of the epigenomic changes that occur during cell fate decisions and help us identify histone modifications that important for the maintenance of pluripotency as well as lineage restriction, and chromatin remodelers that are potentially involved in the process. In order to gain the complete picture of the changes in the genomic landscape that take place over lineage restriction, alongside histone modifications it would be very important to characterize the changes in the chromatin landscape and DNA methylation. In this regard, it would be fascinating to employ ATAC-Seq and bi-sulphite sequencing in aging animal cap explants to explore further the regions of open and closed chromatin during lineage restriction. Such studies would provide us with tremendous knowledge regarding the changes in the epigenomic landscape during lineage restriction and would be an exciting future avenue to study.

Gene regulatory circuitry controlling neural crest stem cell state

Characterizing the gene regulatory circuitry that is controlling neural crest stem cell maintenance is necessary to identify the mechanisms that are involved in the control of pluripotency of these cells. Using a candidate approach, I demonstrated a previously uncharacterized role of HDAC1/2 in neural crest stem cell maintenance (Chapter 2). Further, I identified a novel role for Snail proteins to maintain the pluripotency of these cells which had previously been thought to function only in the later specification and migration of neural crest

cells (Chapter 6). However, there are still large gaps in our knowledge regarding the network that allows for these cells to be maintained in a suspended state of pluripotency. In Chapter 4, I address this by performing a genome scale transcriptomic study on control and neural crest animal cap explants to expand on our previous knowledge regarding the network that controls the pluripotency of these cells and identify factors that might have novel roles in the maintenance of stem cell attributes of neural crest cells. Using a network-based approach such as WGCNA as well PCA/sparse-PCA based methodologies, I explore further the gene regulatory circuitry that controls the neural crest stem cell state. I demonstrate that there is an early and late neural crest signature with the early signature comprising of genes that are involved in the maintaining the stem cell state, while the late signature corresponds to genes involved in lineage restriction of these cells. Our results provide a framework wherein we have identified genes that are potentially involved in this process and can now test systematically and mechanistically their individual contributions to the pluripotency network and characterize further how they fit into the gene circuitry controlling the pluripotency of these cells.

Although large similarities exist between pluripotent cells and neural crest cell, there many differences between these two cell types as well. Indeed, there is a differential requirement of SoxB1 and SoxE factors in blastula cells and neural crest cells respectively(Buitrago-Delgado et al., 2018). Strikingly, on global transcriptomic level, we find that while many of same factors are expressed in both these cell types, there are dramatic differences in the levels that we observe in neural crest explants at blastula and neurula stages. While at blastula stages (Stage 9) these cells are very similar, dramatic changes in gene expression occur over developmental time. This would suggest that that during the process of development, neural crest cells become different from the pluripotent blastula cells. It would be very interesting to identify at which point neural

crest explants begin to deviate from blastula cells in terms of gene expression, and the dynamics of the changes that take place over the process of neural crest formation. To this end, an ideal experiment would be to perform a time series of gene expression at all stages between Stage 9 (blastula) and Stage 13 (neurula) and follow the dynamic changes of gene expression that occur over that time period. Further, using ATAC-Seq, it would be interesting to compare the changes in gene expression to changes in open regions of chromatin that are present to further characterize the transcriptional and epigenetic landscape of neural crest formation, and explore further the similarities and differences between these two cell types.

While we are aware that global transcriptional changes occur during neural crest formation, it is hard to characterize how much of these changes are being translated to protein levels in the cell. It would hence be very interesting to compare the protein levels in neural crest cells and pluripotent blastula cells, as well as identify the changes in these levels as the cells proceed from pluripotent to a defined neural crest state. Another interesting question that arises from this work is, how homogenous is the population of blastula cells that are destined to become the neural crest on a transcriptional and epigenetic level? The advent of single cell technologies would help us explore this further as it would allow us to assay neural crest formation at a single cell resolution as they progress from a blastula state to a neural crest state and is an exciting area for future investigation.

Concluding remarks

Through my thesis work, I have uncovered a novel role for HDAC activity and histone acetylation in maintaining the neural crest stem cell state. I found that HDAC activity is necessary for the pluripotency of blastula cells and neural crest cells. Further, I demonstrate that

low levels of histone acetylation are a shared feature of pluripotent blastula cells and neural crest cells. Interestingly, I have identified that HDACs function by maintaining histone acetylation low at the genomic loci of lineage markers, and HDACi results in aberrant expression of markers of various lineages pushing the cells out of pluripotency. The ability of the cell to read the low levels of histone acetylation is also very important, as loss of BRD protein activity also results in a loss of pluripotency and neural crest formation. Strikingly, I have found that HDACs and BRD proteins regulate pluripotency through distinct mechanisms. Further, I explored the changes in the epigenomic landscape as an embryo proceeds through early development and identified histone modifications that are associated with the pluripotent state. Finally, I have characterized the gene expression changes that occur during the process of neural crest development using transcriptomic analysis and identified an early and late neural crest signature that presumably contributes to stem cell maintenance and lineage restriction of these cells. Taken together, these results advance our knowledge regarding the mechanism of control of pluripotency during embryonic development and provide vital insights into the formation and maintenance of the neural crest.

My work clearly demonstrates that the transcriptional and epigenetic control of pluripotency of the neural crest as well as stem cell maintenance in general is tightly regulated during embryonic development. While we have resolved several unanswered questions in the field, many remain to be addressed. Is there a specific level of histone acetylation that needs to be maintained in order to predispose the cells to become neural crest? What are the specific roles of each histone modifications in the maintenance of pluripotency of the neural crest and is there an epigenomic signature of neural crest state? What are different chromatin remodelers that are involved in the process and how do the different components of regulatory circuitry integrate in

order to maintain the neural crest state? How heterogeneous is the population of early neural crest progenitors – do they all have same degrees of potency? We have now begun to understand a little bit more about the maintenance of the neural crest stem cell state, and these insights would be useful as we continue explore further the similarities and differences between the regulatory networks mediating the potency of both stem cell populations. The answers to these questions will provide us with a framework to more clearly understand that origins and formation of this unique stem cell population, the neural crest, and understand further the evolution of vertebrates.

Significance of the thesis work

Understanding the mechanisms utilized by neural crest cells to retain their stem cell attributes will provide us with tremendous knowledge to advance stem cell therapies and regenerative medicine. My thesis work provides novel insights into the epigenetic mechanisms utilized by neural crest cells to retain their pluripotency while the rest of the embryo undergoes lineage restriction, as well adds to our fundamental knowledge regarding how the epigenetic state regulates stem cell maintenance within the developing embryo. It elucidates the molecular players involved in stem cell maintenance during embryonic development and expands our knowledge regarding the gene regulatory circuitry controlling pluripotency of neural crest cells. This work also helps us reconcile some of the discrepancies observed regarding the role of HDAC activity and histone acetylation in stem cell maintenance in ES cells. Further, given that HDACs have been broadly implicated in cancers, understanding the mechanisms through which HDACs might contribute to cancer stem cell maintenance is highly significant and provides us with more clues regarding possible therapeutic avenues that can be exploited in cancers by targeting HDACs. Ultimately, the work presented in this thesis enhances our knowledge

regarding maintenance of stem cell potency during embryonic development and broadens our understanding of the genesis of this novel cell type, the neural crest.

References

- Adamo, A., Sesé, B., Boue, S., Castaño, J., Paramonov, I., Barrero, M. J., & Izpisua Belmonte, J. C. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nature Cell Biology*, 13(6), 652–659. <http://doi.org/10.1038/ncb2246>
- Adams, G. E., Chandru, A., & Cowley, S. M. (2018). Co-repressor, co-activator and general transcription factor: the many faces of the Sin3 histone deacetylase (HDAC) complex. *The Biochemical Journal*, 475(24), 3921–3932. <http://doi.org/10.1042/BCJ20170314>
- Akkers, R. C., van Heeringen, S. J., Jacobi, U. G., Janssen-Megens, E. M., François, K.-J., Stunnenberg, H. G., & Veenstra, G. J. C. (2009). A hierarchy of H3K4me3 and H3K27me3 acquisition in spatial gene regulation in *Xenopus* embryos. *Developmental Cell*, 17(3), 425–434. <http://doi.org/10.1016/j.devcel.2009.08.005>
- Alberga, A., Boulay, J. L., Kempe, E., Dennefeld, C., & Haenlin, M. (1991). The snail gene required for mesoderm formation in *Drosophila* is expressed dynamically in derivatives of all three germ layers. *111*(4), 983–992.
- Almouzni, G., Khochbin, S., Dimitrov, S., & Wolffe, A. P. (1994). Histone Acetylation Influences both Gene Expression and Development of *Xenopus laevis*. *Developmental Biology*, 165(2), 654–669. <http://doi.org/10.1006/dbio.1994.1283>
- Alsdorf, R., & Wyszynski, D. F. (2005). Teratogenicity of sodium valproate. *Expert Opinion on Drug Safety*, 4(2), 345–353.
- Ariizumi, T., & Asashima, M. (2001). In vitro induction systems for analyses of amphibian organogenesis and body patterning. *45*(1), 273–279.
- Asashima, M., Nakano, H., Shimada, K., Kinoshita, K., Ishii, K., Shibai, H., & Ueno, N. (1990a). Mesodermal induction in early amphibian embryos by activin A (erythroid differentiation factor). *Roux's Archives of Developmental Biology : the Official Organ of the EDBO*, 198(6), 330–335. <http://doi.org/10.1007/BF00383771>
- Asashima, M., Nakano, H., Uchiyama, H., Davids, M., Plessow, S., Loppnow-Blinde, B., et al. (1990b). The vegetalizing factor belongs to a family of mesoderm-inducing proteins related to erythroid differentiation factor. *Die Naturwissenschaften*, 77(8), 389–391.
- Atlasi, Y., & Stunnenberg, H. G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nature Reviews. Genetics*. <http://doi.org/10.1038/nrg.2017.57>
- Aybar, M. J., Nieto, M. A., & Mayor, R. (2003). Snail precedes slug in the genetic cascade required for the specification and migration of the *Xenopus* neural crest. *130*(3), 483–494.

- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., et al. (2006). Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*, 8(5), 532–538. <http://doi.org/10.1038/ncb1403>
- Baggiolini, A., Varum, S., Mateos, J. M., Bettosini, D., John, N., Bonalli, M., et al. (2015). Premigratory and migratory neural crest cells are multipotent in vivo. *Cell Stem Cell*, 16(3), 314–322. <http://doi.org/10.1016/j.stem.2015.02.017>
- Bajpai, R., Chen, D. A., Rada-Iglesias, A., Zhang, J., Xiong, Y., Helms, J., et al. (2010). CHD7 cooperates with PBAF to control multipotent neural crest formation. *Nature*, 463(7283), 958–962. <http://doi.org/10.1038/nature08733>
- Baltus, G. A., Kowalski, M. P., Tutter, A. V., & Kadam, S. (2009). A positive regulatory role for the mSin3A-HDAC complex in pluripotency through Nanog and Sox2. *The Journal of Biological Chemistry*, 284(11), 6998–7006. <http://doi.org/10.1074/jbc.M807670200>
- Baroffio, A., Dupin, E., & Le Douarin, N. M. (1988). Clone-forming ability and differentiation potential of migratory neural crest cells., 85(14), 5325–5329.
- Battle, R., Alba-Castellón, L., Loubat-Casanovas, J., Armenteros, E., Francí, C., Stanisavljevic, J., et al. (2013). Snail1 controls TGF- β responsiveness and differentiation of mesenchymal stem cells., 32(28), 3381–3389. <http://doi.org/10.1038/onc.2012.342>
- Bellmeyer, A., Krase, J., Lindgren, J., & LaBonne, C. (2003). The protooncogene c-myc is an essential regulator of neural crest formation in xenopus. *Developmental Cell*, 4(6), 827–839.
- Berdasco, M., & Esteller, M. (2011). DNA methylation in stem cell renewal and multipotency. *Stem Cell Research & Therapy*, 2(5), 42. <http://doi.org/10.1186/scrt83>
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006a). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), 315–326. <http://doi.org/10.1016/j.cell.2006.02.041>
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006b). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), 315–326. <http://doi.org/10.1016/j.cell.2006.02.041>
- Bhattacharya, D., Rothstein, M., Azambuja, A. P., & Simões-Costa, M. (2018). Control of neural crest multipotency by Wnt signaling and the Lin28/let-7 axis. *eLife*, 7, 1241006. <http://doi.org/10.7554/eLife.40556>

- Blythe, S. A., Cha, S.-W., Tadjuidje, E., Heasman, J., & Klein, P. S. (2010). beta-Catenin primes organizer gene expression by recruiting a histone H3 arginine 8 methyltransferase, Prmt2. *Developmental Cell*, 19(2), 220–231. <http://doi.org/10.1016/j.devcel.2010.07.007>
- Bogdanovic, O., Fernandez-Miñán, A., Tena, J. J., la Calle-Mustienes, de, E., Hidalgo, C., van Kruysbergen, I., et al. (2012a). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*, 22(10), 2043–2053. <http://doi.org/10.1101/gr.134833.111>
- Bogdanovic, O., van Heeringen, S. J., & Veenstra, G. J. C. (2012b). The epigenome in early vertebrate development. *Genesis (New York, N.Y. : 2000)*, 50(3), 192–206. <http://doi.org/10.1002/dvg.20831>
- Bronner, M. E., & LeDouarin, N. M. (2012). Development and evolution of the neural crest: an overview., 366(1), 2–9. <http://doi.org/10.1016/j.ydbio.2011.12.042>
- Bronner, M. E., & Simões-Costa, M. (2016). The Neural Crest Migrating into the Twenty-First Century. *Current Topics in Developmental Biology*, 116, 115–134. <http://doi.org/10.1016/bs.ctdb.2015.12.003>
- Bronner-Fraser, M., & Fraser, S. E. (1988). Cell lineage analysis reveals multipotency of some avian neural crest cells., 335(6186), 161–164. <http://doi.org/10.1038/335161a0>
- Buitrago-Delgado, E., Nordin, K., Rao, A., Geary, L., & LaBonne, C. (2015). NEURODEVELOPMENT. Shared regulatory programs suggest retention of blastula-stage potential in neural crest cells. *Science (New York, N.Y.)*, 348(6241), 1332–1335. <http://doi.org/10.1126/science.aaa3655>
- Buitrago-Delgado, E., Schock, E. N., Nordin, K., & LaBonne, C. (2018). A transition from SoxB1 to SoxE transcription factors is essential for progression from pluripotent blastula cells to neural crest cells. *Developmental Biology*, 444(2), 50–61. <http://doi.org/10.1016/j.ydbio.2018.08.008>
- Calloni, G. W., Glavieux-Pardanaud, C., Le Douarin, N. M., & Dupin, E. (2007). Sonic Hedgehog promotes the development of multipotent neural crest progenitors endowed with both mesenchymal and neural potentials., 104(50), 19879–19884. <http://doi.org/10.1073/pnas.0708806104>
- Calloni, G. W., Le Douarin, N. M., & Dupin, E. (2009). High frequency of cephalic neural crest cells shows coexistence of neurogenic, melanogenic, and osteogenic differentiation capacities. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), 8947–8952. <http://doi.org/10.1073/pnas.0903780106>

- Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., et al. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression., *2*(2), 76–83. <http://doi.org/10.1038/35000025>
- Carneiro, K., Donnet, C., Rejtar, T., Karger, B. L., Barisone, G. A., Díaz, E., et al. (2011). Histone deacetylase activity is necessary for left-right patterning during vertebrate development. *BMC Developmental Biology*, *11*(1), 29. <http://doi.org/10.1186/1471-213X-11-29>
- Chambers, I. (2004). The molecular basis of pluripotency in mouse embryonic stem cells., *6*(4), 386–391. <http://doi.org/10.1089/clo.2004.6.386>
- Chambers, I., & Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency., *136*(14), 2311–2322. <http://doi.org/10.1242/dev.024398>
- Chen, X., Vega, V. B., & Ng, H.-H. (2008). Transcriptional regulatory networks in embryonic stem cells., *73*(0), 203–209. <http://doi.org/10.1101/sqb.2008.73.026>
- Cornell, R. A., & Eisen, J. S. (2002). Delta/Notch signaling promotes formation of zebrafish neural crest by repressing Neurogenin 1 function. *Development (Cambridge, England)*, *129*(11), 2639–2648.
- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. <http://doi.org/10.1073/pnas.1016071107>
- Cunliffe, V. T., & Casaccia-Bonnet, P. (2006). Histone deacetylase 1 is essential for oligodendrocyte specification in the zebrafish CNS. *Mechanisms of Development*, *123*(1), 24–30. <http://doi.org/10.1016/j.mod.2005.10.005>
- Damjanovski, S., Sachs, L. M., & Shi, Y. B. (2000). Multiple stage-dependent roles for histone deacetylases during amphibian embryogenesis: implications for the involvement of extracellular matrix remodeling. *The International Journal of Developmental Biology*, *44*(7), 769–776.
- Dang, H., Ding, W., Emerson, D., & Rountree, C. B. (2011). Snail1 induces epithelial-to-mesenchymal transition and tumor initiating stem cell characteristics., *11*(1), 396. <http://doi.org/10.1186/1471-2407-11-396>
- de Crozé, N., Maczkowiak, F., & Monsoro-Burq, A. H. (2011). Reiterative AP2a activity controls sequential steps in the neural crest gene regulatory network. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(1), 155–160. <http://doi.org/10.1073/pnas.1010740107>

- DeLaurier, A., Nakamura, Y., Braasch, I., Khanna, V., Kato, H., Wakitani, S., et al. (2012). Histone deacetylase-4 is required during early cranial neural crest development for generation of the zebrafish palatal skeleton. *BMC Developmental Biology*, 12(1), 16. <http://doi.org/10.1186/1471-213X-12-16>
- Di Micco, R., Fontanals-Cirera, B., Low, V., Ntziachristos, P., Yuen, S. K., Lovell, C. D., et al. (2014). Control of Embryonic Stem Cell Identity by BRD4-Dependent Transcriptional Elongation of Super-Enhancer-Associated Pluripotency Genes. *Cell Reports*, 9(1), 234–247. <http://doi.org/10.1016/j.celrep.2014.08.055>
- Dovey, O. M., Foster, C. T., & Cowley, S. M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18), 8242–8247. <http://doi.org/10.1073/pnas.1000478107>
- Dupin, E., & Sommer, L. (2012). Neural crest progenitors and stem cells: from early development to adulthood. *Developmental Biology*, 366(1), 83–95. <http://doi.org/10.1016/j.ydbio.2012.02.035>
- Eastham, A. M., Spencer, H., Soncin, F., Ritson, S., Merry, C. L. R., Stern, P. L., & Ward, C. M. (2007). Epithelial-mesenchymal transition events during human embryonic stem cell differentiation., 67(23), 11254–11262. <http://doi.org/10.1158/0008-5472.CAN-07-2253>
- Eckschlager, T., Plch, J., Stiborova, M., & Hrabeta, J. (2017). Histone Deacetylase Inhibitors as Anticancer Drugs. *International Journal of Molecular Sciences*, 18(7), 1414. <http://doi.org/10.3390/ijms18071414>
- Falkenberg, K. J., & Johnstone, R. W. (2014). Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nature Reviews. Drug Discovery*, 13(9), 673–691. <http://doi.org/10.1038/nrd4360>
- Fang, F., Xu, Y., Chew, K.-K., Chen, X., Ng, H.-H., & Matsudaira, P. (2014). Coactivators p300 and CBP maintain the identity of mouse embryonic stem cells by mediating long-range chromatin structure. *Stem Cells (Dayton, Ohio)*, 32(7), 1805–1816. <http://doi.org/10.1002/stem.1705>
- Farthing, C. R., Ficz, G., Ng, R. K., Chan, C.-F., Andrews, S., Dean, W., et al. (2008). Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLOS Genetics*, 4(6), e1000116. <http://doi.org/10.1371/journal.pgen.1000116>
- Fernandez Alonso, R., Davidson, L., Hukelmann, J., Zengerle, M., Prescott, A. R., Lamond, A., et al. (2017). Brd4-Brd2 isoform switching coordinates pluripotent exit and Smad2-dependent lineage specification. *EMBO Reports*, 18(7), 1108–1122. <http://doi.org/10.15252/embr.201643534>

- Furumai, R., Matsuyama, A., Kobashi, N., Lee, K.-H., Nishiyama, M., Nakajima, H., et al. (2002). FK228 (depsipeptide) as a natural prodrug that inhibits class I histone deacetylases. *Cancer Research*, 62(17), 4916–4921.
- Gans, C., & Northcutt, R. G. (1983). Neural crest and the origin of vertebrates: a new head. *Science (New York, N.Y.)*, 220(4594), 268–273. <http://doi.org/10.1126/science.220.4594.268>
- Geary, L., & LaBonne, C. (2018). FGF mediated MAPK and PI3K/Akt Signals make distinct contributions to pluripotency and the establishment of Neural Crest. *eLife*, 7, 139. <http://doi.org/10.7554/eLife.33845>
- Glavic, A., Silva, F., Aybar, M. J., Bastidas, F., & Mayor, R. (2004). Interplay between Notch signaling and the homeoprotein Xiro1 is required for neural crest induction in *Xenopus* embryos. *Development (Cambridge, England)*, 131(2), 347–359. <http://doi.org/10.1242/dev.00945>
- Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: a landscape takes shape., 128(4), 635–638. <http://doi.org/10.1016/j.cell.2007.02.006>
- Gonzales-Cope, M., Sidoli, S., Bhanu, N. V., Won, K.-J., & Garcia, B. A. (2016). Histone H4 acetylation and the epigenetic reader Brd4 are critical regulators of pluripotency in embryonic stem cells. *BMC Genomics*, 17(1), 95. <http://doi.org/10.1186/s12864-016-2414-y>
- Gou, Y., Li, J., Wu, J., Gupta, R., Cho, I., Ho, T.-V., et al. (2018). Prmt1 regulates craniofacial bone formation upstream of Msx1. *Mechanisms of Development*, 152, 13–20. <http://doi.org/10.1016/j.mod.2018.05.001>
- Göttlicher, M., Minucci, S., Zhu, P., Krämer, O. H., Schimpf, A., Giavara, S., et al. (2001). Valproic acid defines a novel class of HDAC inhibitors inducing differentiation of transformed cells. *The EMBO Journal*, 20(24), 6969–6978. <http://doi.org/10.1093/emboj/20.24.6969>
- Groves, A. K., & LaBonne, C. (2014). Setting appropriate boundaries: Fate, patterning and competence at the neural plate border., 389(1), 2–12. <http://doi.org/10.1016/j.ydbio.2013.11.027>
- Gupta, R., Wills, A., Ucar, D., & Baker, J. (2014). Developmental enhancers are marked independently of zygotic Nodal signals in *Xenopus*. *Developmental Biology*, 395(1), 38–49. <http://doi.org/10.1016/j.ydbio.2014.08.034>
- Gurvich, N., Berman, M. G., Wittner, B. S., Gentleman, R. C., Klein, P. S., & Green, J. B. A. (2005). Association of valproate-induced teratogenesis with histone deacetylase inhibition in vivo. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, 19(9), 1166–1168. <http://doi.org/10.1096/fj.04-3425fje>

- Haberland, M., Mokalled, M. H., Montgomery, R. L., & Olson, E. N. (2009a). Epigenetic control of skull morphogenesis by histone deacetylase 8. *Genes & Development*, 23(14), 1625–1630. <http://doi.org/10.1101/gad.1809209>
- Haberland, M., Montgomery, R. L., & Olson, E. N. (2009b). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews. Genetics*, 10(1), 32–42. <http://doi.org/10.1038/nrg2485>
- Habibi, E., & Stunnenberg, H. G. (2017). Transcriptional and epigenetic control in mouse pluripotency: lessons from in vivo and in vitro studies. *Current Opinion in Genetics & Development*, 46, 114–122. <http://doi.org/10.1016/j.gde.2017.07.005>
- Hall, B. K. (2000). The neural crest as a fourth germ layer and vertebrates as quadroblastic not triploblastic., 2(1), 3–5.
- Hanna, L. A., Foreman, R. K., Tarasenko, I. A., Kessler, D. S., & Labosky, P. A. (2002). Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo., 16(20), 2650–2661. <http://doi.org/10.1101/gad.1020502>
- Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., et al. (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*, 6(5), 479–491. <http://doi.org/10.1016/j.stem.2010.03.018>
- Hemavathy, K., Ashraf, S. I., & Ip, Y. T. (2000). Snail/slug family of repressors: slowly going into the fast lane of development and cancer., 257(1), 1–12.
- Hezroni, H., Sailaja, B. S., & Meshorer, E. (2011). Pluripotency-related, valproic acid (VPA)-induced genome-wide histone H3 lysine 9 (H3K9) acetylation patterns in embryonic stem cells. *The Journal of Biological Chemistry*, 286(41), 35977–35988. <http://doi.org/10.1074/jbc.M111.266254>
- His, W. (1868). Untersuchungen über die erste Anlage des Wirbeltierleibes: die erste Entwicklung des Hühnchens im Ei, Vogel FCW, Leipzig.
- Hong, C.-S., & Saint-Jeannet, J.-P. (2007). The activity of Pax3 and Zic1 regulates three distinct cell fates at the neural plate border., 18(6), 2192–2202. <http://doi.org/10.1091/mbc.E06-11-1047>
- Hong, C.-S., Park, B.-Y., & Saint-Jeannet, J.-P. (2008). Fgf8a induces neural crest indirectly through the activation of Wnt8 in the paraxial mesoderm. *Development (Cambridge, England)*, 135(23), 3903–3910. <http://doi.org/10.1242/dev.026229>
- Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S. J., Bogdanovic, O., Lister, R., & Veenstra, G. J. C. (2015). Embryonic transcription is controlled by maternally defined chromatin state. *Nature Communications*, 6, 10148. <http://doi.org/10.1038/ncomms10148>

- Hoppler, S., & Wheeler, G. N. (2015). DEVELOPMENTAL BIOLOGY. It's about time for neural crest. *Science (New York, N.Y.)*, 348(6241), 1316–1317. <http://doi.org/10.1126/science.aab2719>
- Horne, G. A., Stewart, H. J. S., Dickson, J., Knapp, S., Ramsahoye, B., & Chevassut, T. (2015). Nanog requires BRD4 to maintain murine embryonic stem cell pluripotency and is suppressed by bromodomain inhibitor JQ1 together with Lefty1. *Stem Cells and Development*, 24(7), 879–891. <http://doi.org/10.1089/scd.2014.0302>
- Hou, Z., Peng, H., Ayyanathan, K., Yan, K.-P., Langer, E. M., Longmore, G. D., & Rauscher, F. J. (2008). The LIM protein AJUBA recruits protein arginine methyltransferase 5 to mediate SNAIL-dependent transcriptional repression., 28(10), 3198–3207. <http://doi.org/10.1128/MCB.01435-07>
- Houzelstein, D., Bullock, S. L., Lynch, D. E., Grigorieva, E. F., Wilson, V. A., & Beddington, R. S. P. (2002). Growth and early postimplantation defects in mice deficient for the bromodomain-containing protein Brd4. *Molecular and Cellular Biology*, 22(11), 3794–3802.
- Hu, N., Strobl-Mazzulla, P. H., & Bronner, M. E. (2014). Epigenetic regulation in neural crest development. *Developmental Biology*, 396(2), 159–168. <http://doi.org/10.1016/j.ydbio.2014.09.034>
- Hu, N., Strobl-Mazzulla, P., Sauka-Spengler, T., & Bronner, M. E. (2012). DNA methyltransferase3A as a molecular switch mediating the neural tube-to-neural crest fate transition. *Genes & Development*, 26(21), 2380–2385. <http://doi.org/10.1101/gad.198747.112>
- Huang, J. K., Dorey, K., Ishibashi, S., & Amaya, E. (2007). BDNF promotes target innervation of *Xenopus* mandibular trigeminal axons in vivo. *BMC Developmental Biology*, 7(1), 59. <http://doi.org/10.1186/1471-213X-7-59>
- Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A. E., & Melton, D. A. (2008). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnology*, 26(7), 795–797. <http://doi.org/10.1038/nbt1418>
- Ignatius, M. S., Moose, H. E., El-Hodiri, H. M., & Henion, P. D. (2008). colgate/hdac1 Repression of foxd3 expression is required to permit mitfa-dependent melanogenesis. *Developmental Biology*, 313(2), 568–583. <http://doi.org/10.1016/j.ydbio.2007.10.045>
- Ignatius, M. S., Unal Eroglu, A., Malireddy, S., Gallagher, G., Nambiar, R. M., & Henion, P. D. (2013). Distinct functional and temporal requirements for zebrafish Hdac1 during neural crest-derived craniofacial and peripheral neuron development. *PloS One*, 8(5), e63218. <http://doi.org/10.1371/journal.pone.0063218>

- Jacob, C., Lötscher, P., Engler, S., Baggiolini, A., Varum Tavares, S., Brügger, V., et al. (2014). HDAC1 and HDAC2 control the specification of neural crest cells into peripheral glia. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, 34(17), 6112–6122. <http://doi.org/10.1523/JNEUROSCI.5212-13.2014>
- Jacques-Fricke, B. T., & Gammill, L. S. (2014). Neural crest specification and migration independently require NSD3-related lysine methyltransferase activity. *Molecular Biology of the Cell*, 25(25), 4174–4186. <http://doi.org/10.1091/mbc.E13-12-0744>
- Jamaladdin, S., Kelly, R. D. W., O'Regan, L., Dovey, O. M., Hodson, G. E., Millard, C. J., et al. (2014). Histone deacetylase (HDAC) 1 and 2 are essential for accurate cell division and the pluripotency of embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 111(27), 9840–9845. <http://doi.org/10.1073/pnas.1321330111>
- Kajita, M., McClinic, K. N., & Wade, P. A. (2004). Aberrant expression of the transcription factors snail and slug alters the response to genotoxic stress., 24(17), 7559–7566. <http://doi.org/10.1128/MCB.24.17.7559-7566.2004>
- Kaltschmidt, B., Kaltschmidt, C., & Widera, D. (2012). Adult craniofacial stem cells: sources and relation to the neural crest. *Stem Cell Reviews*, 8(3), 658–671. <http://doi.org/10.1007/s12015-011-9340-9>
- Karantzali, E., Schulz, H., Hummel, O., Hubner, N., Hatzopoulos, A., & Kretsovali, A. (2008). Histone deacetylase inhibition accelerates the early events of stem cell differentiation: transcriptomic and epigenetic analysis. *Genome Biology*, 9(4), R65. <http://doi.org/10.1186/gb-2008-9-4-r65>
- Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H., & Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics*, 13(1), 424. <http://doi.org/10.1186/1471-2164-13-424>
- Kelly, R. D. W., Chandru, A., Watson, P. J., Song, Y., Blades, M., Robertson, N. S., et al. (2018). Histone deacetylase (HDAC) 1 and 2 complexes regulate both histone acetylation and crotonylation in vivo. *Scientific Reports*, 8(1), 14690. <http://doi.org/10.1038/s41598-018-32927-9>
- Kidder, B. L., & Palmer, S. (2012). HDAC1 regulates pluripotency and lineage specific transcriptional networks in embryonic and trophoblast stem cells. *Nucleic Acids Research*, 40(7), 2925–2939. <http://doi.org/10.1093/nar/gkr1151>
- Kim, H., Kang, K., Ekram, M. B., Roh, T.-Y., & Kim, J. (2011). Aebp2 as an epigenetic regulator for neural crest cells. *PloS One*, 6(9), e25174. <http://doi.org/10.1371/journal.pone.0025174>

- King, M. W., Roberts, J. M., & Eisenman, R. N. (1986). Expression of the c-myc proto-oncogene during development of *Xenopus laevis*, 6(12), 4499–4508.
- Kong, Y., Grimaldi, M., Curtin, E., Dougherty, M., Kaufman, C., White, R. M., et al. (2014). Neural crest development and craniofacial morphogenesis is coordinated by nitric oxide and histone acetylation. *Chemistry & Biology*, 21(4), 488–501.
<http://doi.org/10.1016/j.chembiol.2014.02.013>
- LaBonne, C., & Bronner-Fraser, M. (1998). Neural crest induction in *Xenopus*: evidence for a two-signal model. *Development (Cambridge, England)*, 125(13), 2403–2414.
- LaBonne, C., & Bronner-Fraser, M. (1999). Molecular mechanisms of neural crest formation. *Annual Review of Cell and Developmental Biology*, 15(1), 81–112.
<http://doi.org/10.1146/annurev.cellbio.15.1.81>
- LaBonne, C., & Bronner-Fraser, M. (2000). Snail-related transcriptional repressors are required in *Xenopus* for both the induction of the neural crest and its subsequent migration., 221(1), 195–205. <http://doi.org/10.1006/dbio.2000.9609>
- Lagger, G., O'Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., et al. (2002). Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *The EMBO Journal*, 21(11), 2672–2681.
<http://doi.org/10.1093/emboj/21.11.2672>
- Lamb, T. M., Knecht, A. K., Smith, W. C., Stachel, S. E., Economides, A. N., Stahl, N., et al. (1993). Neural induction by the secreted polypeptide noggin. *Science (New York, N.Y.)*, 262(5134), 713–718.
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. <http://doi.org/10.1186/1471-2105-9-559>
- Le Douarin, N. M., & Dupin, E. (2012). The neural crest in vertebrate evolution. *Current Opinion in Genetics & Development*, 22(4), 381–389.
<http://doi.org/10.1016/j.gde.2012.06.001>
- Lee, P.-C., Taylor-Jaffe, K. M., Nordin, K. M., Prasad, M. S., Lander, R. M., & LaBonne, C. (2012). SUMOylated SoxE factors recruit Grg4 and function as transcriptional repressors in the neural crest. *The Journal of Cell Biology*, 198(5), 799–813.
<http://doi.org/10.1083/jcb.201204161>
- Leptin, M. (1991). twist and snail as positive and negative regulators during *Drosophila* mesoderm development., 5(9), 1568–1576.

- Li, M., Liu, G.-H., & Belmonte, J. C. I. (2012). Navigating the epigenetic landscape of pluripotent stem cells. *Nature Reviews Molecular Cell Biology*, 13(8), 524–535. <http://doi.org/10.1038/nrm3393>
- Li, X., Yang, H., Huang, S., & Qiu, Y. (2014). Histone deacetylase 1 and p300 can directly associate with chromatin and compete for binding in a mutually exclusive manner. *PloS One*, 9(4), e94523. <http://doi.org/10.1371/journal.pone.0094523>
- Li, Y., & Seto, E. (2016). HDACs and HDAC Inhibitors in Cancer Development and Therapy. *Cold Spring Harbor Perspectives in Medicine*, 6(10), a026831. <http://doi.org/10.1101/cshperspect.a026831>
- Liang, G., & Zhang, Y. (2012). Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. *Cell Research*, 23(1), 49–69. <http://doi.org/10.1038/cr.2012.175>
- Light, W., Vernon, A. E., Lasorella, A., Iavarone, A., & LaBonne, C. (2005). Xenopus Id3 is required downstream of Myc for the formation of multipotent neural crest progenitor cells., 132(8), 1831–1841. <http://doi.org/10.1242/dev.01734>
- Lignell, A., Kerosuo, L., Streichan, S. J., Cai, L., & Bronner, M. E. (2017). Identification of a neural crest stem cell niche by Spatial Genomic Analysis. *Nature Communications*, 8(1), 1830. <http://doi.org/10.1038/s41467-017-01561-w>
- Lin, T., Ponn, A., Hu, X., Law, B. K., & Lu, J. (2010). Requirement of the histone demethylase LSD1 in Snail-mediated transcriptional repression during epithelial-mesenchymal transition. *Oncogene*, 29(35), 4896–4904. <http://doi.org/10.1038/onc.2010.234>
- Lin, Y., Li, X.-Y., Willis, A. L., Liu, C., Chen, G., & Weiss, S. J. (2014). Snail1-dependent control of embryonic stem cell pluripotency and lineage commitment. *Nature Communications*, 5, 3070. <http://doi.org/10.1038/ncomms4070>
- Liu, P., Dou, X., Liu, C., Wang, L., Xing, C., Peng, G., et al. (2015). Histone deacetylation promotes mouse neural induction by restricting Nodal-dependent mesendoderm fate. *Nature Communications*, 6, 6830. <http://doi.org/10.1038/ncomms7830>
- Liu, W., Stein, P., Cheng, X., Yang, W., Shao, N.-Y., Morrissey, E. E., et al. (2014). BRD4 regulates Nanog expression in mouse embryonic stem cells and preimplantation embryos. *Cell Death and Differentiation*, 21(12), 1950–1960. <http://doi.org/10.1038/cdd.2014.124>
- Loh, Y.-H., Zhang, W., Chen, X., George, J., & Ng, H.-H. (2007). Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes & Development*, 21(20), 2545–2557. <http://doi.org/10.1101/gad.1588207>

- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <http://doi.org/10.1186/s13059-014-0550-8>
- Lukoseviciute, M., Gavriouchkina, D., Williams, R. M., Hochgreb-Hagele, T., Senanayake, U., Chong-Morrison, V., et al. (2018). From Pioneer to Repressor: Bimodal foxd3 Activity Dynamically Remodels Neural Crest Regulatory Landscape In Vivo. *Developmental Cell*, 47(5), 608–628.e6. <http://doi.org/10.1016/j.devcel.2018.11.009>
- Luo, T., Lee, Y.-H., Saint-Jeannet, J.-P., & Sargent, T. D. (2003). Induction of neural crest in *Xenopus* by transcription factor AP2alpha. *Proceedings of the National Academy of Sciences of the United States of America*, 100(2), 532–537. <http://doi.org/10.1073/pnas.0237226100>
- Ma, P., & Schultz, R. M. (2016). HDAC1 and HDAC2 in mouse oocytes and preimplantation embryos: Specificity versus compensation. *Cell Death and Differentiation*, 23(7), 1119–1127. <http://doi.org/10.1038/cdd.2016.31>
- Maguire, L. H., Thomas, A. R., & Goldstein, A. M. (2015). Tumors of the neural crest: Common themes in development and cancer. *Developmental Dynamics : an Official Publication of the American Association of Anatomists*, 244(3), 311–322. <http://doi.org/10.1002/dvdy.24226>
- Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells., 133(4), 704–715. <http://doi.org/10.1016/j.cell.2008.03.027>
- Markowetz, F., Mulder, K. W., Airoidi, E. M., Lemischka, I. R., & Troyanskaya, O. G. (2010). Mapping dynamic histone acetylation patterns to gene expression in nanog-depleted murine embryonic stem cells. *PLoS Computational Biology*, 6(12), e1001034. <http://doi.org/10.1371/journal.pcbi.1001034>
- Marmorstein, R., & Zhou, M.-M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harbor Perspectives in Biology*, 6(7), a018762–a018762. <http://doi.org/10.1101/cshperspect.a018762>
- Matsukawa, S., Miwata, K., Asashima, M., & Michiue, T. (2015). The requirement of histone modification by PRDM12 and Kdm4a for the development of pre-placodal ectoderm and neural crest in *Xenopus*. *Developmental Biology*, 399(1), 164–176. <http://doi.org/10.1016/j.ydbio.2014.12.028>
- Mauhin, V., Lutz, Y., Dennefeld, C., & Alberga, A. (1993). Definition of the DNA-binding site repertoire for the *Drosophila* transcription factor SNAIL., 21(17), 3951–3957.
- Medeiros, D. M., & Crump, J. G. (2012). New perspectives on pharyngeal dorsoventral patterning in development and evolution of the vertebrate jaw. *Developmental Biology*, 371(2), 121–135. <http://doi.org/10.1016/j.ydbio.2012.08.026>

- Meissner, Alexander, Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770. <http://doi.org/10.1038/nature07107>
- Melcer, S., Hezroni, H., Rand, E., Nissim-Rafinia, M., Skoultchi, A., Stewart, C. L., et al. (2012). Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation. *Nature Communications*, 3, 910. <http://doi.org/10.1038/ncomms1915>
- Milet, C., & Monsoro-Burq, A. H. (2012). Neural crest induction at the neural plate border in vertebrates., 366(1), 22–33. <http://doi.org/10.1016/j.ydbio.2012.01.013>
- Milet, C., Maczkowiak, F., Roche, D. D., & Monsoro-Burq, A.-H. (2013). Pax3 and Zic1 drive induction and differentiation of multipotent, migratory, and functional neural crest in *Xenopus* embryos., 110(14), 5528–5533. <http://doi.org/10.1073/pnas.1219124110>
- Milstone, Z. J., Lawson, G., & Trivedi, C. M. (2017). Histone deacetylase 1 and 2 are essential for murine neural crest proliferation, pharyngeal arch development, and craniofacial morphogenesis. *Developmental Dynamics : an Official Publication of the American Association of Anatomists*, 246(12), 1015–1026. <http://doi.org/10.1002/dvdy.24563>
- Minoux, M., Holwerda, S., Vitobello, A., Kitazawa, T., Kohler, H., Stadler, M. B., & Rijli, F. M. (2017). Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science (New York, N.Y.)*, 355(6332), eaal2913. <http://doi.org/10.1126/science.aal2913>
- Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., et al. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular Cell*, 30(6), 755–766. <http://doi.org/10.1016/j.molcel.2008.05.007>
- Montgomery, R. L., Davis, C. A., Potthoff, M. J., Haberland, M., Fielitz, J., Qi, X., et al. (2007). Histone deacetylases 1 and 2 redundantly regulate cardiac morphogenesis, growth, and contractility. *Genes & Development*, 21(14), 1790–1802. <http://doi.org/10.1101/gad.1563807>
- Morrison, G. M., & Brickman, J. M. (2006). Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development., 133(10), 2011–2022. <http://doi.org/10.1242/dev.02362>
- Morrison, S. J., White, P. M., Zock, C., & Anderson, D. J. (1999). Prospective identification, isolation by flow cytometry, and in vivo self-renewal of multipotent mammalian neural crest stem cells., 96(5), 737–749.
- Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., et al. (2015). Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early

- differentiation of embryonic stem cells. *Cell Metabolism*, 21(3), 392–402.
<http://doi.org/10.1016/j.cmet.2015.02.002>
- Mundell, N. A., & Labosky, P. A. (2011). Neural crest stem cell multipotency requires Foxd3 to maintain neural potential and repress mesenchymal fates., 138(4), 641–652.
<http://doi.org/10.1242/dev.054718>
- Muñoz, W. A., & Trainor, P. A. (2015). Neural crest cell evolution: how and when did a neural crest cell become a neural crest cell. *Current Topics in Developmental Biology*, 111, 3–26.
<http://doi.org/10.1016/bs.ctdb.2014.11.001>
- Murko, C., Lager, S., Steiner, M., Seiser, C., Schoefer, C., & Pusch, O. (2013). Histone deacetylase inhibitor Trichostatin A induces neural tube defects and promotes neural crest specification in the chicken neural tube. *Differentiation; Research in Biological Diversity*, 85(1-2), 55–66. <http://doi.org/10.1016/j.diff.2012.12.001>
- Nieto, M. A. (2002). The snail superfamily of zinc-finger transcription factors., 3(3), 155–166.
<http://doi.org/10.1038/nrm757>
- Niwa, H. (2007). How is pluripotency determined and maintained? *Development (Cambridge, England)*, 134(4), 635–646. <http://doi.org/10.1242/dev.02787>
- Nordin, K., & LaBonne, C. (2014). Sox5 Is a DNA-Binding Cofactor for BMP R-Smads that Directs Target Specificity during Patterning of the Early Ectoderm. *Developmental Cell*, 31(3), 374–382. <http://doi.org/10.1016/j.devcel.2014.10.003>
- O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., & Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. *Molecular and Cellular Biology*, 21(13), 4330–4336. <http://doi.org/10.1128/MCB.21.13.4330-4336.2001>
- Ochoa, S. D., Salvador, S., & LaBonne, C. (2012). The LIM adaptor protein LMO4 is an essential regulator of neural crest development., 361(2), 313–325.
<http://doi.org/10.1016/j.ydbio.2011.10.034>
- Pasini, D., Bracken, A. P., Jensen, M. R., Lazzerini Denchi, E., & Helin, K. (2004). Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *The EMBO Journal*, 23(20), 4061–4071. <http://doi.org/10.1038/sj.emboj.7600402>
- Peinado, H., Ballestar, E., Esteller, M., & Cano, A. (2004). Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex., 24(1), 306–319.
- Peinado, H., Olmeda, D., & Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype?, 7(6), 415–428.
<http://doi.org/10.1038/nrc2131>

- Penzel, R., Oschwald, R., Chen, Y., Tacke, L., & Grunz, H. (1997). Characterization and early embryonic expression of a neural specific transcription factor xSOX3 in *Xenopus laevis*, 41(5), 667–677.
- Phillips, B. T., Kwon, H.-J., Melton, C., Houghtaling, P., Fritz, A., & Riley, B. B. (2006). Zebrafish *msxB*, *msxC* and *msxE* function together to refine the neural-nonneural border and regulate cranial placodes and neural crest development. *Developmental Biology*, 294(2), 376–390. <http://doi.org/10.1016/j.ydbio.2006.03.001>
- Pillai, R., Coverdale, L. E., Dubey, G., & Martin, C. C. (2004). Histone deacetylase 1 (HDAC-1) required for the normal formation of craniofacial cartilage and pectoral fins of the zebrafish. *Developmental Dynamics : an Official Publication of the American Association of Anatomists*, 231(3), 647–654. <http://doi.org/10.1002/dvdy.20168>
- Plouhinec, J.-L., Roche, D. D., Pegoraro, C., Figueiredo, A. L., Maczkowiak, F., Brunet, L. J., et al. (2014). Pax3 and Zic1 trigger the early neural crest gene regulatory network by the direct activation of multiple key neural crest specifiers. *Developmental Biology*, 386(2), 461–472. <http://doi.org/10.1016/j.ydbio.2013.12.010>
- Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits., 9(4), 265–273. <http://doi.org/10.1038/nrc2620>
- Prasad, M. S., Sauka-Spengler, T., & LaBonne, C. (2012). Induction of the neural crest state: control of stem cell attributes by gene regulatory, post-transcriptional and epigenetic interactions., 366(1), 10–21. <http://doi.org/10.1016/j.ydbio.2012.03.014>
- Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., et al. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell*, 163(1), 68–83. <http://doi.org/10.1016/j.cell.2015.08.036>
- Pünzeler, S., Link, S., Wagner, G., Keilhauer, E. C., Kronbeck, N., Spitzer, R. M., et al. (2017). Multivalent binding of PWWP2A to H2A.Z regulates mitosis and neural crest differentiation. *The EMBO Journal*, 36(15), 2263–2279. <http://doi.org/10.15252/emj.201695757>
- Qi, H. H., Sarkissian, M., Hu, G.-Q., Wang, Z., Bhattacharjee, A., Gordon, D. B., et al. (2010). Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature*, 466(7305), 503–507. <http://doi.org/10.1038/nature09261>
- Qiao, Y., Wang, R., Yang, X., Tang, K., & Jing, N. (2015). Dual roles of histone H3 lysine 9 acetylation in human embryonic stem cell pluripotency and neural differentiation. *The Journal of Biological Chemistry*, 290(4), 2508–2520. <http://doi.org/10.1074/jbc.M114.603761>

- Rada-Iglesias, A., Bajpai, R., Prescott, S., Brugmann, S. A., Swigut, T., & Wysocka, J. (2012). Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell*, 11(5), 633–648. <http://doi.org/10.1016/j.stem.2012.07.006>
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470(7333), 279–283. <http://doi.org/10.1038/nature09692>
- Reynolds, N., Latos, P., Hynes-Allen, A., Loos, R., Leaford, D., O'Shaughnessy, A., et al. (2012). NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell*, 150(5), 583–594. <http://doi.org/10.1016/j.cell.2012.02.020>
- Sasai, N., Mizuseki, K., & Sasai, Y. (2001). Requirement of FoxD3-class signaling for neural crest determination in *Xenopus*. *Development*, 128(13), 2525–2536.
- Sasai, Y., Lu, B., Steinbeisser, H., & De Robertis, E. M. (1995). Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in *Xenopus*. *Development*, 121(6), 333–336. <http://doi.org/10.1038/376333a0>
- Satijn, D. P., Hamer, K. M., Blaauwen, den, J., & Otte, A. P. (2001). The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Molecular and Cellular Biology*, 21(4), 1360–1369. <http://doi.org/10.1128/MCB.21.4.1360-1369.2001>
- Sauka-Spengler, T., & Bronner-Fraser, M. (2008). A gene regulatory network orchestrates neural crest formation. *Nature Reviews Molecular Cell Biology*, 9(7), 557–568. <http://doi.org/10.1038/nrm2428>
- Sauka-Spengler, T., Meulemans, D., Jones, M., & Bronner-Fraser, M. (2007). Ancient evolutionary origin of the neural crest gene regulatory network. *Developmental Cell*, 13(3), 405–420. <http://doi.org/10.1016/j.devcel.2007.08.005>
- Saunders, A., Huang, X., Fidalgo, M., Reimer, M. H., Faiola, F., Ding, J., et al. (2017). The SIN3A/HDAC Corepressor Complex Functionally Cooperates with NANOG to Promote Pluripotency. *Cell Reports*, 18(7), 1713–1726. <http://doi.org/10.1016/j.celrep.2017.01.055>
- Scerbo, P., Girardot, F., Vivien, C., Markov, G. V., Luxardi, G., Demeneix, B., et al. (2012). Ventx factors function as Nanog-like guardians of developmental potential in *Xenopus*. *PLoS ONE*, 7(5), e36855. <http://doi.org/10.1371/journal.pone.0036855>
- Schneider, T. D., Arteaga-Salas, J. M., Mentele, E., David, R., Nicetto, D., Imhof, A., & Rupp, R. A. W. (2011). Stage-specific histone modification profiles reveal global transitions in the *Xenopus* embryonic epigenome. *PLoS ONE*, 6(7), e22548. <http://doi.org/10.1371/journal.pone.0022548>

- Schwarz, D., Varum, S., Zemke, M., Schöler, A., Baggiolini, A., Draganova, K., et al. (2014). Ezh2 is required for neural crest-derived cartilage and bone formation. *Development (Cambridge, England)*, 141(4), 867–877. <http://doi.org/10.1242/dev.094342>
- Sen, R., Pezoa, S. A., Carpio Shull, L., Hernandez-Lagunas, L., Niswander, L. A., & Artinger, K. B. (2018). Kat2a and Kat2b Acetyltransferase Activity Regulates Craniofacial Cartilage and Bone Differentiation in Zebrafish and Mice. *Journal of Developmental Biology*, 6(4), 27. <http://doi.org/10.3390/jdb6040027>
- Session, A. M., Uno, Y., Kwon, T., Chapman, J. A., Toyoda, A., Takahashi, S., et al. (2016). Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*, 538(7625), 336–343. <http://doi.org/10.1038/nature19840>
- Seto, E., & Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*, 6(4), a018713–a018713. <http://doi.org/10.1101/cshperspect.a018713>
- Simões-Costa, M., & Bronner, M. E. (2013). Insights into neural crest development and evolution from genomic analysis. *Genome Research*, 23(7), 1069–1080. <http://doi.org/10.1101/gr.157586.113>
- Simões-Costa, M., & Bronner, M. E. (2015). Establishing neural crest identity: a gene regulatory recipe. *Development (Cambridge, England)*, 142(2), 242–257. <http://doi.org/10.1242/dev.105445>
- Singh, N., Trivedi, C. M., Lu, M., Mullican, S. E., Lazar, M. A., & Epstein, J. A. (2011). Histone deacetylase 3 regulates smooth muscle differentiation in neural crest cells and development of the cardiac outflow tract. *Circulation Research*, 109(11), 1240–1249. <http://doi.org/10.1161/CIRCRESAHA.111.255067>
- Snape, A., Wylie, C. C., Smith, J. C., & Heasman, J. (1987). Changes in states of commitment of single animal pole blastomeres of *Xenopus laevis*. *119*(2), 503–510.
- Solter, D. (2006). From teratocarcinomas to embryonic stem cells and beyond: a history of embryonic stem cell research. *7*(4), 319–327. <http://doi.org/10.1038/nrg1827>
- Spivakov, M., & Fisher, A. G. (2007). Epigenetic signatures of stem-cell identity. *Nature Reviews. Genetics*, 8(4), 263–271. <http://doi.org/10.1038/nrg2046>
- Staszkiwicz, J., Power, R. A., Harkins, L. L., Barnes, C. W., Strickler, K. L., Rim, J. S., et al. (2013). Silencing histone deacetylase-specific isoforms enhances expression of pluripotency genes in bovine fibroblasts. *Cellular Reprogramming*, 15(5), 397–404. <http://doi.org/10.1089/cell.2013.0026>

- Stemple, D. L., & Anderson, D. J. (1992). Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell*, 71(6), 973–985.
- Strobl-Mazzulla, P. H., & Bronner, M. E. (2012). A PHD12-Snail2 repressive complex epigenetically mediates neural crest epithelial-to-mesenchymal transition. *The Journal of Cell Biology*, 198(6), 999–1010. <http://doi.org/10.1083/jcb.201203098>
- Strobl-Mazzulla, P. H., Sauka-Spengler, T., & Bronner-Fraser, M. (2010). Histone demethylase Jmjd2A regulates neural crest specification. *Developmental Cell*, 19(3), 460–468. <http://doi.org/10.1016/j.devcel.2010.08.009>
- Szabó, A., & Mayor, R. (2018). Mechanisms of Neural Crest Migration. *Annual Review of Genetics*, 52(1), 43–63. <http://doi.org/10.1146/annurev-genet-120417-031559>
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors., 131(5), 861–872. <http://doi.org/10.1016/j.cell.2007.11.019>
- Tan, Y., Xue, Y., Song, C., & Grunstein, M. (2013). Acetylated histone H3K56 interacts with Oct4 to promote mouse embryonic stem cell pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), 11493–11498. <http://doi.org/10.1073/pnas.1309914110>
- Taylor, K. M., & LaBonne, C. (2005). SoxE factors function equivalently during neural crest and inner ear development and their activity is regulated by SUMOylation., 9(5), 593–603. <http://doi.org/10.1016/j.devcel.2005.09.016>
- Taylor, K. M., & LaBonne, C. (2007). Modulating the activity of neural crest regulatory factors., 17(4), 326–331. <http://doi.org/10.1016/j.gde.2007.05.012>
- Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A., & Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers., 145(6), 875–889. <http://doi.org/10.1016/j.cell.2011.05.017>
- Tien, C.-L., Jones, A., Wang, H., Gerigk, M., Nozell, S., & Chang, C. (2015). Snail2/Slug cooperates with Polycomb repressive complex 2 (PRC2) to regulate neural crest development. *Development (Cambridge, England)*, 142(4), 722–731. <http://doi.org/10.1242/dev.111997>
- Trentin, A., Glavieux-Pardanaud, C., Le Douarin, N. M., & Dupin, E. (2004). Self-renewal capacity is a widespread property of various types of neural crest precursor cells., 101(13), 4495–4500. <http://doi.org/10.1073/pnas.0400629101>

- Tseng, A.-S., Carneiro, K., Lemire, J. M., & Levin, M. (2011). HDAC activity is required during *Xenopus* tail regeneration. *PloS One*, 6(10), e26382. <http://doi.org/10.1371/journal.pone.0026382>
- van Heeringen, S. J., Akkers, R. C., van Kruijsbergen, I., Arif, M. A., Hanssen, L. L. P., Sharifi, N., & Veenstra, G. J. C. (2014). Principles of nucleation of H3K27 methylation during embryonic development. *Genome Research*, 24(3), 401–410. <http://doi.org/10.1101/gr.159608.113>
- Vastenhouw, N. L., Zhang, Y., Woods, I. G., Imam, F., Regev, A., Liu, X. S., et al. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. *Nature*, 464(7290), 922–926. <http://doi.org/10.1038/nature08866>
- Vega-Lopez, G. A., Cerrizuela, S., Tribulo, C., & Aybar, M. J. (2018). Neurocristopathies: New insights 150 years after the neural crest discovery. *Developmental Biology*. <http://doi.org/10.1016/j.ydbio.2018.05.013>
- Villarejo, A., Cortés-Cabrera, A., Molina-Ortiz, P., Portillo, F., & Cano, A. (2014). Differential role of Snail1 and Snail2 zinc fingers in E-cadherin repression and epithelial to mesenchymal transition., 289(2), 930–941. <http://doi.org/10.1074/jbc.M113.528026>
- Wakamatsu, Y., Maynard, T. M., & Weston, J. A. (2000). Fate determination of neural crest cells by NOTCH-mediated lateral inhibition and asymmetrical cell division during gangliogenesis. *Development (Cambridge, England)*, 127(13), 2811–2821.
- Wang, C., Kam, R. K. T., Shi, W., Xia, Y., Chen, X., Cao, Y., et al. (2015). The Proto-oncogene Transcription Factor Ets1 Regulates Neural Crest Development through Histone Deacetylase 1 to Mediate Output of Bone Morphogenetic Protein Signaling. *The Journal of Biological Chemistry*, 290(36), 21925–21938. <http://doi.org/10.1074/jbc.M115.644864>
- Wang, J., Kumar, R. M., Biggs, V. J., Lee, H., Chen, Y., Kagey, M. H., et al. (2011). The Msx1 Homeoprotein Recruits Polycomb to the Nuclear Periphery during Development. *Developmental Cell*, 21(3), 575–588. <http://doi.org/10.1016/j.devcel.2011.07.003>
- Wang, L., Koutelou, E., Hirsch, C., McCarthy, R., Schibler, A., Lin, K., et al. (2018). GCN5 Regulates FGF Signaling and Activates Selective MYC Target Genes during Early Embryoid Body Differentiation. *Stem Cell Reports*, 10(1), 287–299. <http://doi.org/10.1016/j.stemcr.2017.11.009>
- Wang, Z., Zang, C., Cui, K., Schones, D. E., Barski, A., Peng, W., & Zhao, K. (2009). Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell*, 138(5), 1019–1031. <http://doi.org/10.1016/j.cell.2009.06.049>
- Ware, C. B., Wang, L., Mecham, B. H., Shen, L., Nelson, A. M., Bar, M., et al. (2009). Histone Deacetylase Inhibition Elicits an Evolutionarily Conserved Self-Renewal Program in

- Embryonic Stem Cells. *Cell Stem Cell*, 4(4), 359–369.
<http://doi.org/10.1016/j.stem.2009.03.001>
- Watanabe, A., Yamada, Y., & Yamanaka, S. (2013). Epigenetic regulation in pluripotent stem cells: a key to breaking the epigenetic barrier. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1609), 20120292–20120292.
<http://doi.org/10.1098/rstb.2012.0292>
- Weinberger, L., Ayyash, M., Novershtern, N., & Hanna, J. H. (2016). Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nature Reviews Molecular Cell Biology*, 17(3), 155–169. <http://doi.org/10.1038/nrm.2015.28>
- Williams, S. R., Aldred, M. A., Kaloustian, Der, V. M., Halal, F., Gowans, G., McLeod, D. R., et al. (2010). Haploinsufficiency of HDAC4 causes brachydactyly mental retardation syndrome, with brachydactyly type E, developmental delays, and behavioral problems. *American Journal of Human Genetics*, 87(2), 219–228.
<http://doi.org/10.1016/j.ajhg.2010.07.011>
- Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., et al. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature*, 557(7704), 256–260.
<http://doi.org/10.1038/s41586-018-0080-8>
- Wu, T., Pinto, H. B., Kamikawa, Y. F., & Donohoe, M. E. (2015). The BET family member BRD4 interacts with OCT4 and regulates pluripotency gene expression. *Stem Cell Reports*, 4(3), 390–403. <http://doi.org/10.1016/j.stemcr.2015.01.012>
- Wysocka, J., Swigut, T., Milne, T. A., Dou, Y., Zhang, X., Burlingame, A. L., et al. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*, 121(6), 859–872. <http://doi.org/10.1016/j.cell.2005.03.036>
- Wyszynski, D. F., Nambisan, M., Surve, T., Alsdorf, R. M., Smith, C. R., Holmes, L. B., Antiepileptic Drug Pregnancy Registry. (2005). Increased rate of major malformations in offspring exposed to valproate during pregnancy. *Neurology*, 64(6), 961–965.
<http://doi.org/10.1212/01.WNL.0000154516.43630.C5>
- Xie, W., Song, C., Young, N. L., Sperling, A. S., Xu, F., Sridharan, R., et al. (2009). Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Molecular Cell*, 33(4), 417–427. <http://doi.org/10.1016/j.molcel.2009.02.004>
- Yang, J., Tang, Y., Liu, H., Guo, F., Ni, J., & Le, W. (2014). Suppression of histone deacetylation promotes the differentiation of human pluripotent stem cells towards neural progenitor cells. *BMC Biology*, 12(1), 95. <http://doi.org/10.1186/s12915-014-0095-z>

- Yoshida, M., Kijima, M., Akita, M., & Beppu, T. (1990). Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *The Journal of Biological Chemistry*, 265(28), 17174–17179.
- Young, R. A. (2011). Control of the embryonic stem cell state. *Cell*, 144(6), 940–954. <http://doi.org/10.1016/j.cell.2011.01.032>
- Zhang, Y., & Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development*, 15(18), 2343–2360. <http://doi.org/10.1101/gad.927301>
- Zhang, Z., Lei, A., Xu, L., Chen, L., Chen, Y., Zhang, X., et al. (2017). Similarity in gene-regulatory networks suggests that cancer cells share characteristics of embryonic neural cells. *The Journal of Biological Chemistry*, 292(31), 12842–12859. <http://doi.org/10.1074/jbc.M117.785865>
- Zhu, L.-F., Hu, Y., Yang, C.-C., Xu, X.-H., Ning, T.-Y., Wang, Z.-L., et al. (2012). Snail overexpression induces an epithelial to mesenchymal transition and cancer stem cell-like properties in SCC9 cells., 92(5), 744–752. <http://doi.org/10.1038/labinvest.2012.8>
- Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., et al. (2006). Negative and Positive Regulation of Gene Expression by Mouse Histone Deacetylase 1. *Molecular and Cellular Biology*, 26(21), 7913–7928. <http://doi.org/10.1128/MCB.01220-06>

Chapter 6

Appendix 1:

**Snail proteins are necessary for stem
cell maintenance**

The Snail family of zinc finger transcription factors are considered to be key regulators of EMT during development and cancer. More recently, they have been implicated in playing functional roles in the formation of cancer stem cells and promoting stem cell attributes. However, the mechanisms by which Snail factors mediate these new roles are still unknown. The neural crest (NC) is an excellent model to ask questions about the function of Snail factors, as they are necessary for both the specification of the NC and its subsequent migration. Since the neural crest is a multipotent stem cell population, it provides an ideal context to further study the role of Snail factors in the maintenance of stem cell attributes. We sought to determine if Snail factors play analogous roles in the maintenance of potency of blastula cells and NC cells in *Xenopus* development.

Snail family transcription factors play important roles during embryonic development and also in tumorigenesis and metastasis. Known for their central role in promoting epithelial to mesenchymal transition (EMT), delamination and migration, these factors have been implicated in the acquisition of invasive migratory behavior, promotion of stem-cell like characteristics and resistance to apoptosis (Nieto, 2002). They were first identified for their role in mesodermal differentiation in *Drosophila* and represent a group of evolutionarily conserved proteins with a characteristic zinc finger domain of the C2H2 type (Alberga, Boulay, Kempe, Dennefeld, & Haenlin, 1991; Hemavathy, Ashraf, & Ip, 2000). Snail proteins have been subsequently identified in several species including humans and other vertebrates. All identified members of this family have a highly conserved C terminal region containing 4-6 zinc fingers and a varying N terminal domain. The Zn fingers function as sequence specific DNA binding domains that recognize an E2-box type element C/A (CAGGTG) (Mauhin, Lutz, Dennefeld, & Alberga, 1993). In vertebrates, there are two members of the Snail family, *Snail1* and *Snail2* (*Slug*) that

have some overlap in their expression patterns and potential redundancy in functionality (Hemavathy et al., 2000; Nieto, 2002). Studies in chick and mouse have shown that there is an inverted expression pattern of *Snail1* and *Snail2* expression in the neural crest and mesoderm suggestive that they can mediate similar functions. While *Snail1* is expressed during mesoderm and neural crest development in mice, *Snail2* is predominantly expressed in those regions in chick. In *Xenopus*, it has been reported that both *XSnail1* and *XSnail2*(*Slug*) are first found to be expressed in the dorsal marginal zone from Stage 10 onwards, above the dorsal blastopore lip. By midgastrula stages, *XSnail1* is expressed in the ectoderm in an arc that surrounds the prospective neural plate (Aybar, Nieto, & Mayor, 2003). However, recently, an expression profile of *XSnail1* and *XSnail2* done in the lab identified that *XSnail1* but not *XSnail2* is expressed maternally from early cleavage stages (Buitrago-Delgado et al., 2015). The expression of *XSnail1* is strong in the blastula at Stage 9 and dorsally in the organizer at Stage 10 (Figure 6.1). After gastrula stages, *XSnail1* expression concentrates to the neural plate border and becomes specific to the NC by neurula stages (Figure 6.1). On the other hand, *XSnail2* is strongly expressed first in the neural plate border regions during late gastrulation and becomes restricted to the NC by Stage 16-17 (Figure 6.2).

Structurally, *Snail1* and *Snail2* are very similar in the C-terminal zinc finger regions and their N-terminal SNAG domain. However, they have differences in their central proline-serine rich region. *Snail2* has the 5 zinc fingers and a characteristic regulatory ‘Slug’ domain that has been found in all vertebrate *Snail2* genes, whereas *Snail1* has 4 Zinc fingers and a regulatory domain with a destruction box and Nuclear Export Signal box (Figure 6.3) (Peinado, Olmeda, & Cano, 2007).

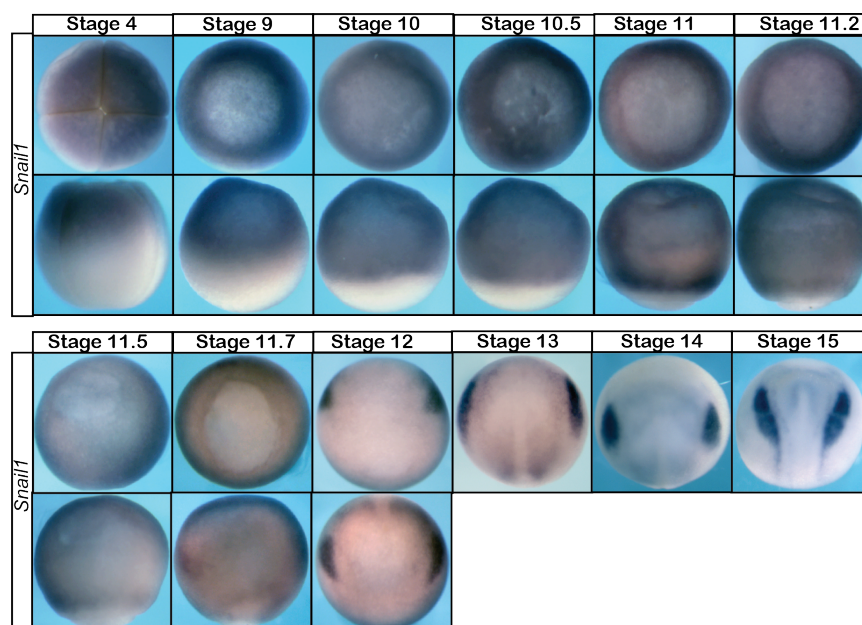


Figure 6.1 *Snail1* expression profile during early *Xenopus* development
In situ hybridization examining *Snail1* expression from early cleavage stages to neurula stages during *Xenopus* development. *Snail1* is maternally provided and expressed in the pluripotent blastula cells and resolves at later stages to the neural crest.

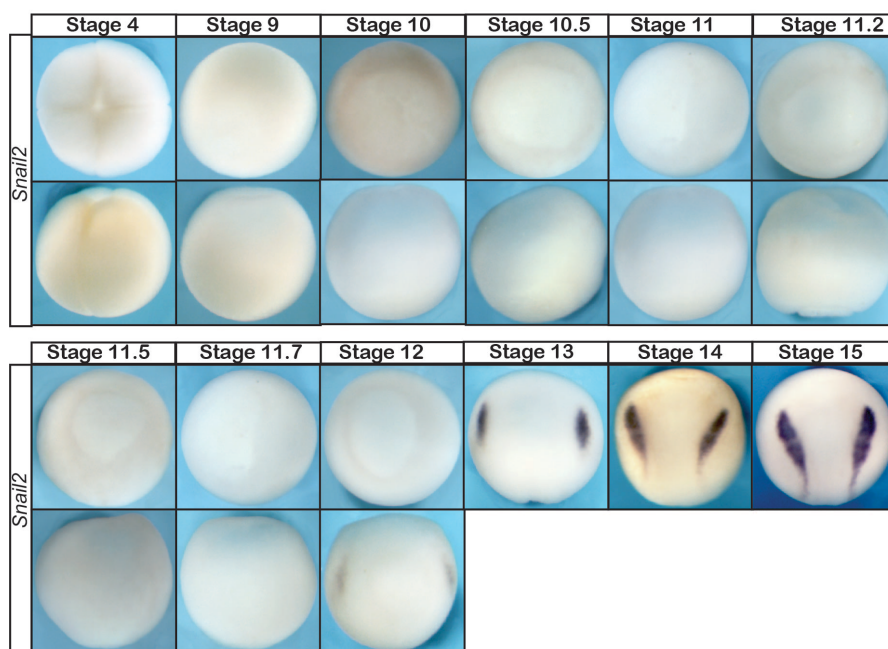


Figure 6.2 *Snail2* expression profile during early *Xenopus* development

In situ hybridization examining *Snail2* expression from early cleavage stages to neurula stages during *Xenopus* development. *Snail2* is not maternally provided or expressed in the pluripotent blastula cells but is turned on towards the end of gastrulation in the prospective neural crest.

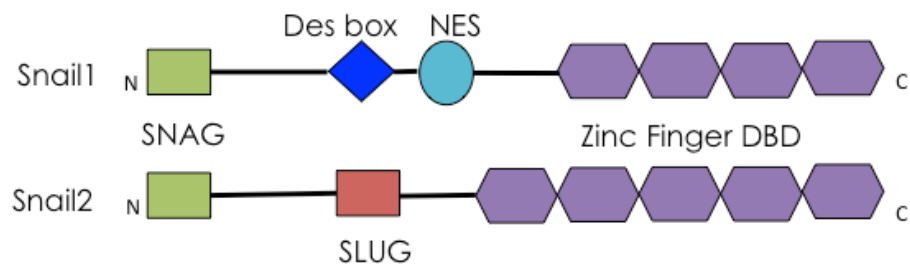


Figure 6.3 Cartoon depicting domains of Snail1 and Snail2

Snail1 and Snail2 contain Zinc finger DNA binding domains with 4 and 5 Zinc fingers respectively and a conserved N-terminal SNAG domain.

Although the zinc fingers of these proteins are highly conserved and structurally very similar, differential requirement of individual zinc fingers has been identified (Villarejo, Cortés-Cabrera, Molina-Ortíz, Portillo, & Cano, 2014). While the first two zinc fingers are essential for Snail function, zinc fingers 3 and 4 are necessary for Snail2 function. Owing to their highly conserved nature, there seems to be potential redundancy in their functionality (Hemavathy et al., 2000).

XSnail1 is described as the first gene to be expressed in the prospective neural crest, and to precede *XSnail2* (*Slug*) in the genetic cascade that leads to the specification of the neural crest (Aybar et al., 2003). *Snail2* has been shown to play a dual role in the neural crest in that it is essential for the specification of the neural crest as well as its subsequent migration (LaBonne & Bronner-Fraser, 2000). The loss of Snail2 function using a dominant negative form of the protein leads to a subsequent loss of the expression of other neural crest regulatory factors and *Snail2* itself. Additionally, overexpression of *XSnail2* leads to an ectopic expansion of the neural crest. Taken together, these results show that *Snail2* is essential for the formation of the neural crest.

Recent work has identified new roles for Snail proteins in the maintenance of stem cell attributes and lineage commitment in ESCs, iPSCs and cancer stem cells (Dang, Ding, Emerson, & Rountree, 2011; Eastham et al., 2007; Zhu et al., 2012). While the function of Snail in conferring stem cell like characteristics to migratory cancer cells by promoting EMT and inhibiting apoptosis etc has been under speculation for several years, a more direct role of Snail in this process has started to emerge recently (Polyak & Weinberg, 2009). There are several studies that have tried to dissect out the role that Snail proteins might play in the pluripotency process, reporting opposing conclusions. While some studies say that *Snail* plays an important role in endowing stem cell characteristics, others report that Snail proteins promote lineage

commitment, differentiation and exit from the stem cell state. One group reported that *Snail1* overexpression is sufficient for inducing an EMT mediated conferring of mesenchymal stem cell like characteristics to mammary gland epithelial cells in culture (Mani et al., 2008). *Snail1* has also been shown to be required for maintaining mesenchymal stem cell potency in a TGF-beta responsive manner (Batlle et al., 2013). On the other hand, a recent report shows *Snail1* to be required for nudging mESCs towards mesodermal fates, while repressing the formation of neuroectodermal fates (Y. Lin et al., 2014). Interestingly, this group also reported that Snail is not necessary for the maintaining the potency of the ESCs, rather, it is required for mediating a WNT and EMT driven exit from the stem cell state. Furthermore, Snail factors have been reported to assist cells in evading apoptosis and repressing cell cycle progression, thereby moving them closer to a cancer stem cell state (Kajita, McClinic, & Wade, 2004). However, all these studies were done *in vitro* and there has yet not been any experiments performed *in vivo* identifying a direct role for Snail factors in the maintenance of stem cell attributes.

While the function of Snail proteins in *Xenopus* is well defined in neural crest formation and migration, as well as mesoderm development, not much is known about the function of Snail factors in the blastula animal pole cells. Recent evidence in the lab has suggested that the neural crest arises due to the retention of stem cell attributes. Snail factors are expressed in the blastula cells and get restricted to the border regions as development proceeds. According to our new hypothesis, the expression pattern of *Snail1* is spatially and temporally consistent with the formation of the neural crest. This would suggest that Snail factors could play key roles in the maintenance of potency of the blastula cells during development as they proceed to become neural crest cells. This new role for Snail proteins is at its infancy stages of characterization, and definitely not clearly understood. Investigating their function in the blastula cells provides an

ideal *in vivo* context in which to study the role of Snail proteins in the maintenance of potency as well as test their ability to confer stem cell characteristics. Given the contradictory reports of Snail protein function in these processes, detailed investigation is required to dissect out the true role of Snail factors in the maintenance of stem cell attributes and lineage commitment.

Snail proteins are required for pluripotency of blastula cells

In order to dissect out the role of Snail factors in this process, we wanted to first determine their necessity for pluripotency gene expression. Overexpression of the Zinc finger DNA binding domain of Snail2 (Δ Snail) serves as a dominant negative to functionally knockdown both Snail1 and Snail2 function and causes loss of neural crest formation (Figure 6.4) (LaBonne & Bronner-Fraser, 2000). Interestingly, loss of Snail function in blastula embryos results in loss of expression of pluripotency markers *Sox3*, *TFAp2*, *Id3*, *Vent2* and *Oct25* (Figure 6.5). This suggests that Snail protein function is necessary for the regulating pluripotency gene expression in blastula cells.

As NC factors such as Snail are required for maintaining expression of factors linked to pluripotency in animal pole cells, including *Oct25*, *Vent2* and *Sox2/3*, we hypothesized that cells depleted for Snail1 would no longer be competent to respond to inducing signals. To test this hypothesis, blastula animal explants from control embryos or embryos injected with Δ Snail, were treated with moderate doses of activin. The control explants strongly responded to Activin induction to form mesoderm as seen by *Xbra* and *MyoD* expression but this responsiveness is lost in explants depleted for Snail function has been blocked (Figure 6.6). As Snail factors have roles in mesoderm formation endogenously, a more rigorous test of their contributions to the stem cell potential of animal pole cells is to ask if these cells retain the capacity to form endoderm when

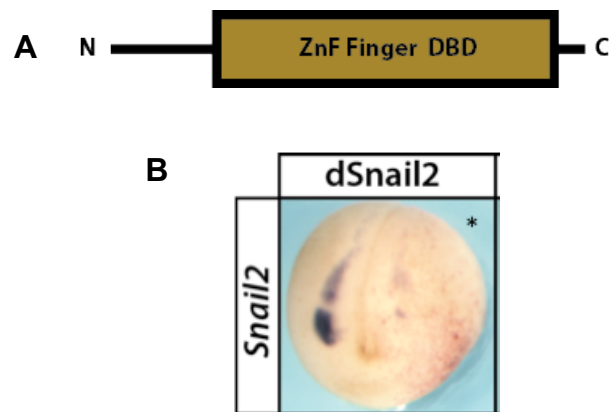


Figure 6.4 Snail function is necessary for neural crest formation

Diagrammatic representation of dominant negative Snail construct
In situ hybridization examining Snail2 expressed in response to unilateral injection of dominant negative Snail protein. Loss of Snail function results in loss of expression of Snail2 when compared to control suggestive of failure to form the neural crest.

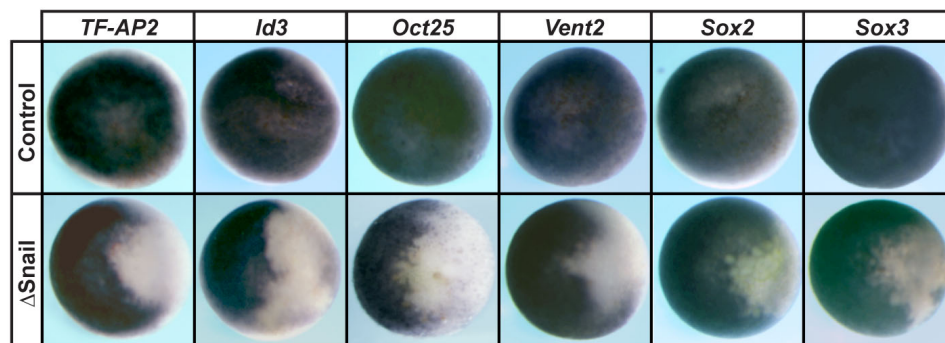


Figure 6.5 Snail function is necessary for pluripotency gene expression
In situ hybridization examining pluripotency gene expression in response to unilateral injection of dominant negative Snail. Loss of Snail function results in loss of expression of *TFAP2*, *Id3*, *Oct25*, *Vent2* and *Sox3* when compared to control.

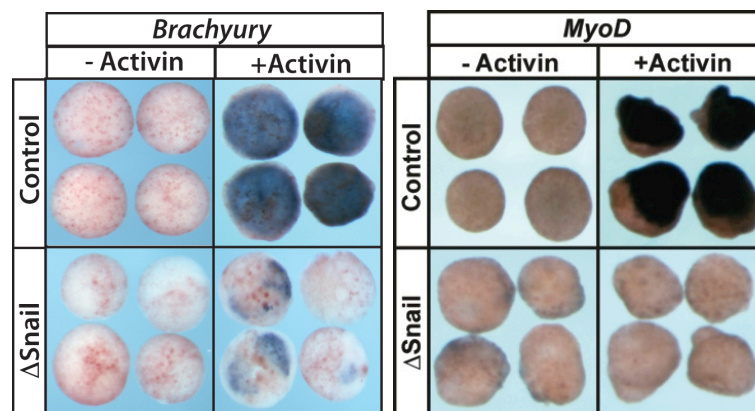


Figure 6.6 Snail function is necessary for blastula cells to form mesoderm

In situ hybridization examining *Brachyury* and *MyoD* after mesoderm induction with activin with/without unilateral injection of dominant negative Snail. Loss of Snail function results in loss of expression of mesodermal markers suggesting the explants have lost the competency to respond to Activin treatment.

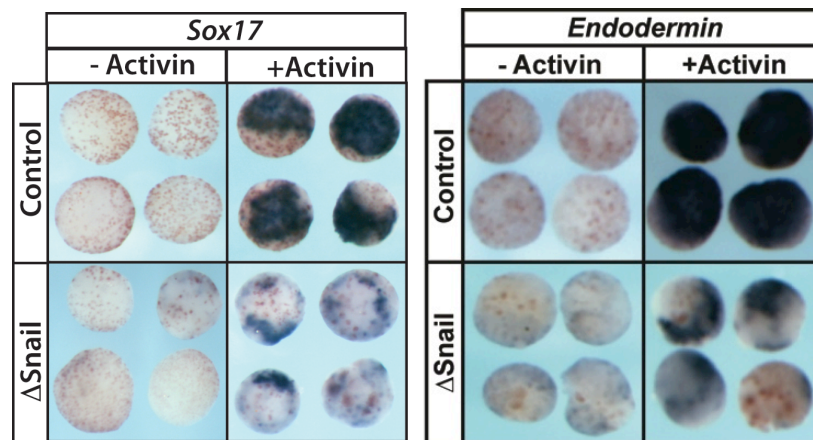


Figure 6.7 Snail function is necessary for blastula cells to form endoderm

In situ hybridization examining *Sox17* and *Endodermin* after endoderm induction with activin with/without unilateral injection of dominant negative Snail. Loss of Snail function results in loss of expression of endodermal markers suggesting the explants have lost the competency to respond to Activin treatment.

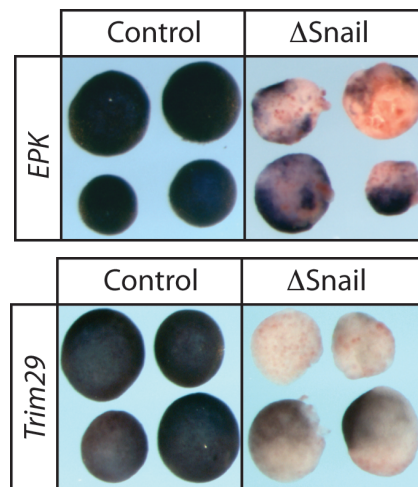


Figure 6.8 Snail function is necessary for blastula cells to form epidermis

In situ hybridization examining *EPK* and *Trim29* in aging animal cap explants with/without unilateral injection of dominant negative Snail. Loss of Snail function results in loss of expression of epidermal markers suggesting the explants do not default to an epidermal state.

Snail function is blocked. Blastula animal pole cells will adopt endodermal fates in response to high doses of activin, and express endoderm specific markers such as *Endodermin* and *Sox17*. Strikingly, we found that blastula explants in which Snail function had been blocked could no longer give rise to endoderm (Figure 6.7). This is a highly significant finding as Snail proteins are not expressed in, and do not function in, endoderm formation during normal development. Loss of activin-mediated endoderm induction must therefore reflect a general lack of competence on the part of Snail depleted animal pole cells to respond to lineage restricting signals. This is strong evidence that Snail plays a required role in the transcriptional network that controls the potency of blastula animal pole cells. Given that Snail protein function seems to be necessary for the potency of blastula cells, we wondered if Snail proteins were required for normal lineage restriction of these cells. The animal cap explants will default to an epidermal state if cultured in isolation. Interestingly, we found that explants depleted for Snail function have an impaired ability to form epidermis as seen by loss of *EPK* and *Trim29* expression (Figure 6.8). These data suggest that Snail protein activity is integral to regulatory network controlling the potency and lineage restriction of blastula cells.

SNAG domain is essential to mediate Snail protein function in blastula cells

Given that we identified that Snail protein function is essential for neural crest formation and maintaining the stem cell state, we next wondered regarding the mechanisms through which Snail proteins might be regulating pluripotency. The SNAG domain of Snail proteins is 9 amino acid conserved domain that has previously been described to be essential for Snail protein function. Indeed, the SNAG domain has been shown by our lab and others to be essential for Snail proteins to interact with binding partners in the cell. We hoped to characterize if the SNAG

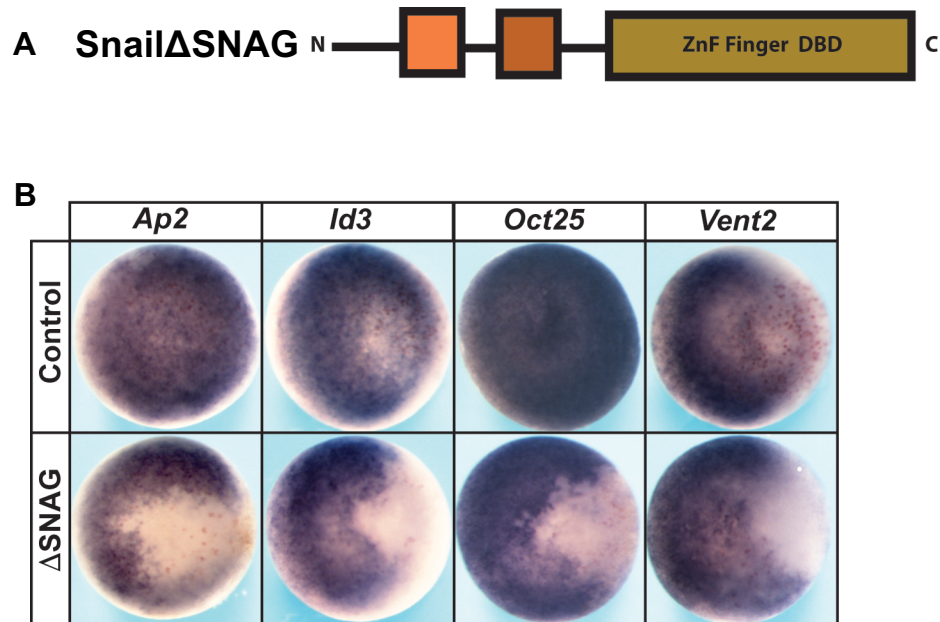


Figure 6.9 Snail function is necessary for neural crest formation

(A) Diagrammatic representation of Snail1 Δ SNAG construct
 (B) *In situ* hybridization examining pluripotency gene expression in response to unilateral injection of Snail1 Δ SNAG. Snail1 Δ SNAG overexpression results in loss of expression of *TFAp2*, *Id3*, *Oct25*, *Vent2* and *Sox3* when compared to control.

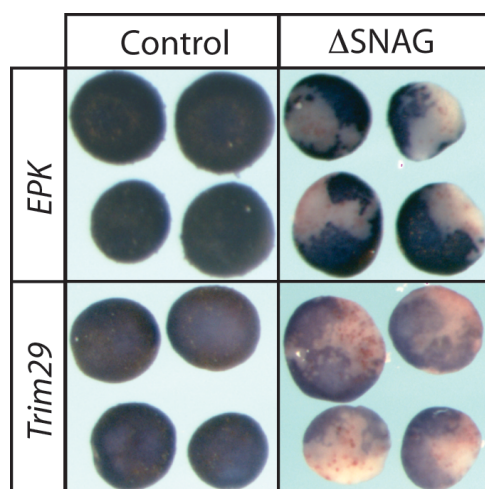


Figure 6.10 Snail1 Δ SNAG results in loss in epidermal formation

In situ hybridization examining *EPK* and *Trim29* in aging animal cap explants with/without unilateral injection of Snail1 Δ SNAG. Snail1 Δ SNAG results in loss of expression of epidermal markers suggesting the explants do not default to an epidermal state.

domain is also required for the maintenance of pluripotency of these cells. To this end, I generated a *Snail1* Δ SNAG construct, which contained the entire *Snail* construct with the SNAG domain deleted. I injected the *Snail1* Δ SNAG construct unilaterally into embryos at the 2 cell stage and collected them at blastula stages to assay if there were any changes in pluripotency gene expression. Strikingly, I found that loss of the SNAG domain completely phenocopied loss of *Snail* protein function through the dominant negative. I found that the *Snail1* Δ SNAG construct results in loss of *Oct25*, *Vent2*, *Tfap2* and *Id3* expression (Figure 6.9). Further, animal cap explants dissected from embryos injected with *Snail1* Δ SNAG construct fail to default to an epidermal state similar to what was observed after loss of *Snail* protein function (Figure 6.10). This suggests that the SNAG domain is necessary to mediate the function of *Snail* proteins in blastula cells.

We were further interested in identifying if the SNAG domain was sufficient to restore *Snail1* function in the blastula. An artificial construct was generated with the SNAG domain fused to dominant negative, Δ *Snail*. Presence of the SNAG domain fused to Δ *Snail* rescues the effects on gene expression caused by Δ *Snail* (Figure 6.11). This suggests that SNAG domain is necessary to mediate *Snail* function in blastula, and *Snail* proteins are functioning through the SNAG domain likely to recruit binding partners to regulate pluripotency.

***Snail1* interacts with the core pluripotency network**

Given that we discovered that *Snail* were necessary for pluripotency of blastula cells and were functioning through their SNAG domain to regulate this function, we hypothesized that *Snail* might be mediating pluripotency through a physical interaction with the core pluripotency network. The proteins of the pluripotency network often function together to mediate the

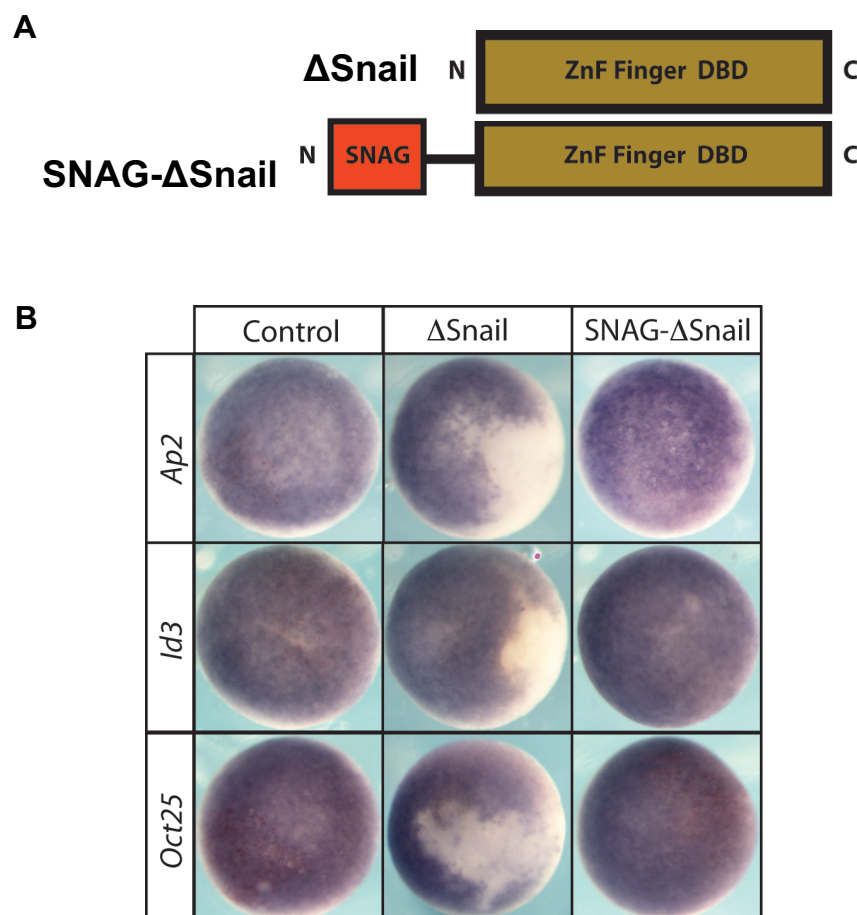


Figure 6.11 Snail function is necessary for neural crest formation

(A) Diagrammatic representation of Snail1 Δ SNAG construct

(B) *In situ* hybridization comparing pluripotency gene expression in response to unilateral injection of Δ Snail and SNAG- Δ Snail. SNAG Δ Snail construct rescues loss of expression of *TFAp2*, *Id3*, *Oct25*, *Vent2* and *Sox3* observed by overexpression of Δ Snail.

function in ESCs. Indeed, Oct25 and Sox2 and Nanog physically interact. Interestingly, Snail1 but not Snail2 was recently found to interact with Nanog and required for reprogramming of iPSCs. Strikingly, previous work from our lab also showed that Snail2 interacted with Vent2 in *Xenopus* via co-immunoprecipitation experiments (Ann Vernon, unpublished data). This suggested to us that it was likely that Snail proteins were functioning through physical interaction with the core pluripotency network and we decided to investigate this further. To this end, we performed co-immunoprecipitation experiments with Snail1 and Snail2 with members of the core pluripotency network such Oct25, Oct60, Vent2 and cMyc. As previously reported, we saw an interaction with Vent2 but did not see any evidence for interaction with Oct25 and Oct60 (unpublished, data not shown). Strikingly, we saw a strong interaction of Snail1 with Myc protein. This was very interesting because this is the first report of Snail protein interaction with cMyc. Curiously, we find that Snail1 but not Snail2 has strong interaction with cMyc (Figure 6.12). This is of particular interest as both Snail and cMyc have been implicated to be required for the maintenance of pluripotency of the neural crest. This would suggest that Snail and Myc might be functioning together in this context and warrants further investigation. Given the novelty of this finding, we were curious whether this interaction was conserved among different species. To test this, I performed co-immunoprecipitation experiments with Snail proteins from the following species: *Drosophila* (D), *Amphioxus* (A), *Ciona* (C), Zebrafish (Z), Mouse (M) alongside *Xenopus* (X) Snail. Strikingly, we find that only *Drosophila* and *Xenopus* Snail strongly interact with Myc (Figure 6.13). This suggests that Snail protein interaction with Myc might be a potentially conserved interaction, and it would very interesting to further characterize the domains required for this interaction and compare the sequences of *Xenopus* and *Drosophila*

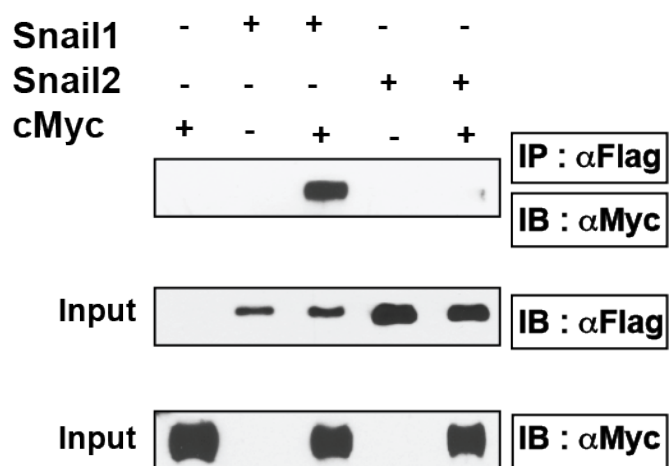


Figure 6.12 Snail1 but not Snail2 interacts with cMyc

Co-immunoprecipitation examining the interaction between Snail1 and Snail2 with cMyc. Snail1 interacts with cMyc but Snail2 does not.

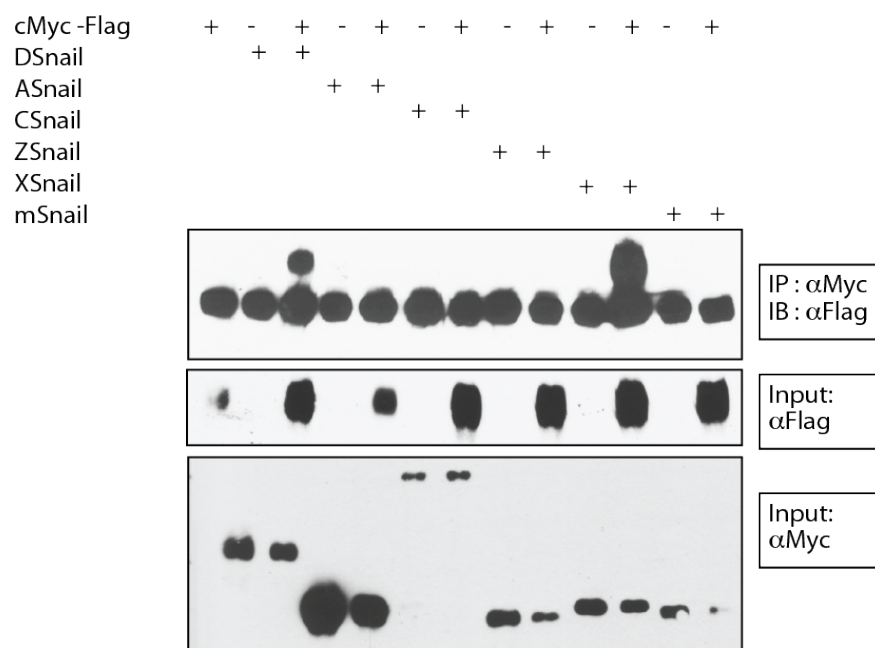


Figure 6.13 *Xenopus* and *Drosophila* Snail1 interacts with cMyc

Co-immunoprecipitation examining the interaction between Snail1 from different species with cMyc. *Xenopus* and *Drosophila* Snail1 interact with cMyc but not other species.

Snail to understand further why these proteins interact while Snail proteins from other species do not. These data suggest that Snail proteins play an important role in pluripotency maintenance and might be working through a physical interaction with core pluripotency protein network.

Snail1 and Snail2 interact with HDACs in *Xenopus*

Our findings regarding the role of Snail proteins and HDACs in stem cell maintenance and neural crest formation led us to wonder further whether these two proteins might be functioning together in this context. Interestingly, it has previously been identified that Snail proteins physically interact with HDAC1 in cell culture (Peinado et al., 2004). This interaction is mediated through the SNAG domain. Given that the SNAG domain is critical for Snail protein function in stem cell maintenance, we hypothesized that Snail proteins might be functioning through a interaction with HDAC1 to mediate their role in blastula cells. Indeed, our lab previously showed that Snail1 and Snail2 in *Xenopus* interact with mouse HDAC1 (Ochoa et al., 2012). We were interested in characterizing the interaction of Snail1 and Snail2 with HDAC1 during early *Xenopus* development. To this end, we performed co-immunoprecipitation experiments with *Xenopus* Snail1 and Snail2 and HDAC1 by microinjecting tagged versions of the proteins at the 2-cell stage. Interestingly, we find that both Snail1 and Snail2 interact with HDAC1 at Stage 10.5 in these embryos. We were further curious to see if there was any temporal control of the interaction of Snail1 and Snail2 with HDAC1. Strikingly, we find that there is temporal control of Snail1 and Snail2 interacting with HDAC1. While Snail2 interacts with HDAC1 at Stage 8 and Stage 10, Snail1 only interacts with HDAC1 at the later stages. This suggests that HDAC1 and Snail2 might be functioning together in the regulation of pluripotency

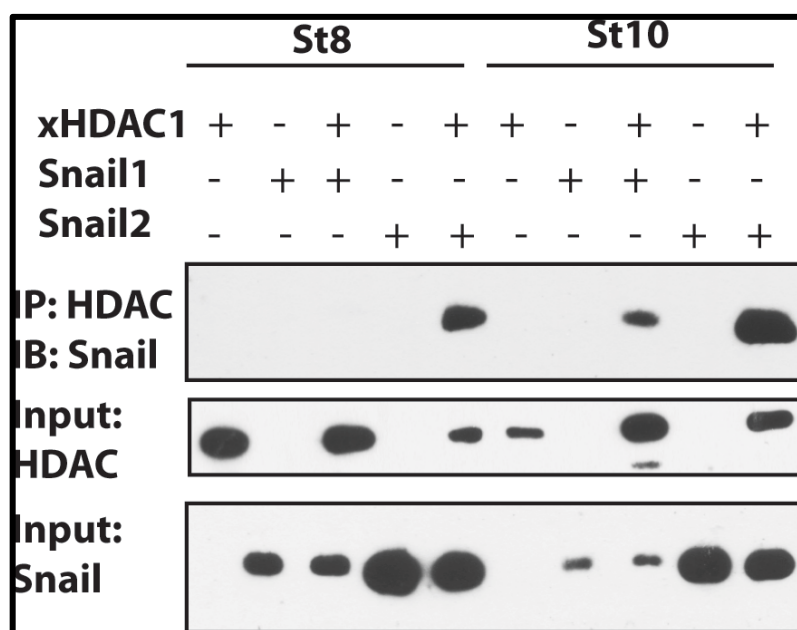


Figure 6.14 Temporal control of Snail1/2 interaction with HDAC1

Co-immunoprecipitation examining the interaction between Snail1 and Snail2 with HDAC1. Snail1 interacts with HDAC1 only at Stage 10, while Snail2 interacts with HDAC1 at Stage 8 and Stage 10

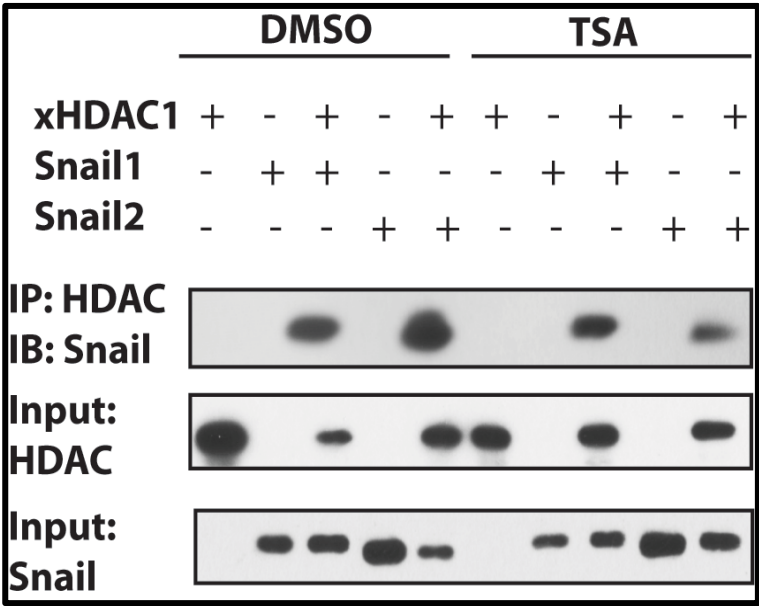
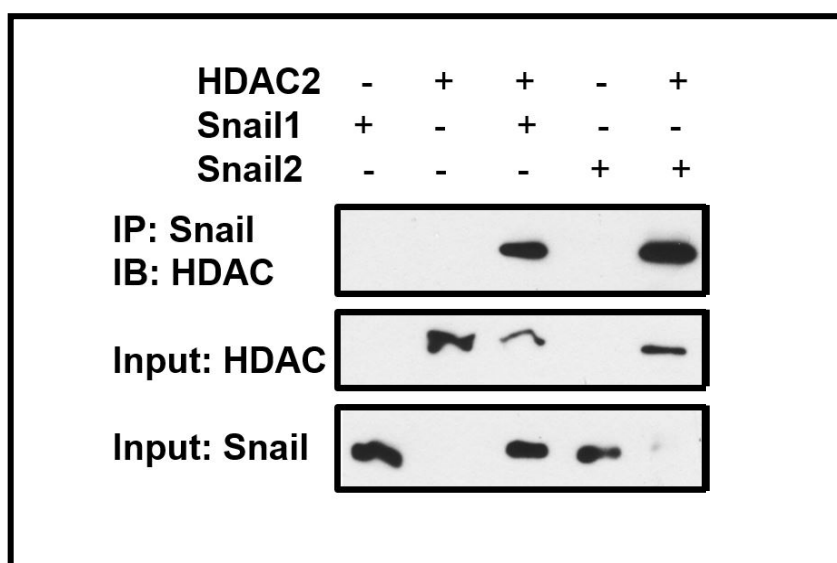


Figure 6.15 Snail1/2 interaction with HDAC1 is partially dependent on acetylation activity

Co-immunoprecipitation examining the interaction between Snail1 and Snail2 with HDAC1 with or without TSA treatment. TSA treatment partially disrupts interaction between Snail1/Snail2 with HDAC1.



in the blastula cells. However, this early role might be switched for a late role of Snail1/2 with HDAC1 in EMT at later stages of development. We further wanted to test if this interaction was dependent on the deacetylase activity of HDAC. To this end, we tested if blocking HDAC activity using TSA would result in a disruption of the interaction between Snail1/2 with HDAC1. While TSA treatment does lead to a reduction of the interaction between Snail1/2 with HDAC1, the interaction is largely maintained. This suggests that this interaction is not dependent on HDAC1 activity. Overall, these data suggest that Snail proteins function potentially alongside HDACs in blastula and neural crest cells to regulate pluripotency.

Chapter 6

Appendix 2:

Bioinformatics Analysis scripts

RNA-Sequencing Analysis

FastQC Analysis – Shell script

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=04:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N fastqc
#MSUB -j oe

module load fastqc/0.11.5

## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR

fastqc ../fastafiles/filename.fastq --outdir= ../fastqcreports
```

STAR Alignment – Shell script

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=04:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N star
#MSUB -j oe

module load STAR/2.6.0
module load samtools
module load boost
module load gcc/6.4.0
module load java
module load fastqc/0.11.5

## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR

STAR --runThreadN 6 \
```

```
--genomeDir ../xenopus_index \
--readFilesIn ../fastafiles/filename.fastq \
--outFileNamePrefix results/outputfilename_ \
--outSAMtype BAM SortedByCoordinate \
--outSAMunmapped Within \
--outSAMattributes Standard
```

HTSEQ Count – Shell script

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=04:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N htseq
#MSUB -j oe

module load STAR/2.6.0
module load samtools
module load boost
module load gcc/6.4.0
module load java
module load python
module load fastqc/0.11.5

## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR

python -m HTSeq.scripts.count -f bam -r pos -s reverse -i gene_name \
results/Filename_Aligned.sortedByCoord.out.bam \
../../genomefiles/XENLA_9.2_Xenbase.GTF > counts_new/outputfilename.txt
```

RSEM – Shell script

RSEM – Prepare reference transcriptome

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=02:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
```



```
#MSUB -N rsem
#MSUB -j oe
```

```
module load bowtie2/2.2.6
module load tophat/2.1.0
module load samtools
module load boost
module load gcc/4.8.3
module load java
module load fastqc/0.11.5
module load homer
```

These are shell commands. Note that all MSUB commands come first.

```
cd $PBS_O_WORKDIR
```

```
../../../../../../software/rsem/1.2.28/bin/rsem-prepare-reference --gtf
../../../../genomefiles/XENLA_9.2_Xenbase.GTF \
--bowtie2 --bowtie2-path
../../../../../../software/bowtie2/2.2.6/bin/ \
../../../../genomefiles/XL9_2.fa xeno92genome
```

RSEM – Calculate expression

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=07:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N rsem
#MSUB -j oe
```

```
module load bowtie2/2.2.6
module load tophat/2.1.0
module load samtools
module load boost
module load gcc/4.8.3
module load java
module load fastqc/0.11.5
module load homer
```

These are shell commands. Note that all MSUB commands come first.

```
cd $PBS_O_WORKDIR
```

```

../../../../../software/rsem/1.2.28/bin/rsem-calculate-expression -p 8 \
--bowtie2 --bowtie2-path ../../../../../software/bowtie2/2.2.6/bin/ \
--estimate-rspd \
--append-names \
--output-genome-bam \
--forward-prob 0 \
../fastafiles/filename.fastq \
xeno92genome ../bamfiles/outputfilename

```

DESeq2 – R script

```

library(BiocInstaller)
library(tximport)
library(geneplotter)
library(RColorBrewer)
library(gplots)
library(readr)
library(DESeq2)
library(EnhancedVolcano)
library(pheatmap)

## importing the RSEM unnormalized count data for use with DESeq - done using tximport
directory <- ("/Users/arao/Documents/LAB/Genomics/RNASEQ-DATA/NC-
RNASEQ/Bowtie2-RSEM pipeline/RSEMoutput/") ## path to folder where the files are stored
filenames = list.files(directory)
names = c("CTRL_St13","CTRL_St13","CTRL_St13",
          "CTRL_St17","CTRL_St17","CTRL_St17",
          "CTRL_St9","CTRL_St9","CTRL_St9",
          "NC_St13","NC_St13","NC_St13",
          "NC_St17","NC_St17","NC_St17",
          "NC_St9","NC_St9","NC_St9")
group = c("CTRL_St13","CTRL_St13","CTRL_St13",
          "CTRL_St17","CTRL_St17","CTRL_St17",
          "CTRL_St9","CTRL_St9","CTRL_St9",
          "NC_St13","NC_St13","NC_St13",
          "NC_St17","NC_St17","NC_St17",
          "NC_St9","NC_St9","NC_St9") ## condition levels required for DESeq2 analysis
samples = data.frame(filenames,names, group) ## created a df with filenames and sample names
files <- file.path(directory, filenames) ## constructs a path to the count files
all(file.exists(files)) ## ensuring that 'files' points to the count data
names(files) = samples$names ## renaming the file names to desired sample names
rna.txi <- tximport(files, type = "rsem", txIn = FALSE, txOut = FALSE) ## actual tximport
rna.txi$length[rna.txi$length == 0] <- 1
head(rna.txi$counts)

```

```

## creating DESeq2 dataset
## tximported countdata + sample info + design conditions

ddsTxi = DESeqDataSetFromTximport(rna.txi,colData = samples, design = ~group)

## Generate count data without filtering
ddsTxi = estimateSizeFactors(ddsTxi)
counts = counts(ddsTxi, normalized = TRUE)
write.csv(counts, "NCRnaseqcounts_nofilter.csv")

## filter out rows with low sums
ddsTxi <- ddsTxi[ rowSums(counts(ddsTxi)) >= 10, ]

##Generate count data after filtering
ddsTxi = estimateSizeFactors(ddsTxi)
counts_afterfilter = counts(ddsTxi, normalized = TRUE)
write.csv(counts_afterfilter, "NCRnaseqcounts_afterfilter.csv")

##heatmap for sample to sample distances
rld <- rlogTransformation(ddsTxi, blind=TRUE)
distsRL <- dist(t(assay(rld)))
mat <- as.matrix(distsRL)
rownames(mat) <- colData(ddsHTSeq)$condition
colnames(mat) <- colData(ddsHTSeq)$condition
hmcol <- colorRampPalette(brewer.pal(9, "Blues"))(255)

tiff("heatmap_NCRnaseq.tiff", height = 5, width = 5, units = 'in', res = 300)
heatmap.2(mat, trace="none", col = rev(hmcol), margin=c(13, 13), labCol = colnames(mat),
cexRow = 0.6, cexCol = 0.6)
dev.off()

## PCA for all samples

tiff("NCRnaseq-PCA.tiff", height = 5, width = 5, units = 'in', res = 300)
plotPCA(rld, intgroup = c("group"))
dev.off()

## Starting differential analysis

## comparing CTRLtoNC-St9
ddsTxi$group = relevel(ddsTxi$group, ref = "CTRL_St9")
##runs the differential analysis
ddsSt9 <- DESeq(ddsTxi)
## gets results after applying the logfoldshrinkage using apeglm method
resultsSt9 = lfcShrink(ddsSt9, coef = "group_NC_St9_vs_CTRL_St9",

```

```

        type = c("apeglm"))
resultsSt9 <- resultsSt9[order(resultsSt9$padj),]
head(resultsSt9)
summary(resultsSt9)

write.csv(as.matrix(resultsSt9), "NCSt9vsCTRLSt9.csv")

tiff("NCSt9vsCTRLSt9_MApot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsSt9, xlim = c(1, 1e5), ylim = c(-10,10), ylab = "Log2 Fold Change")
dev.off()

tiff("NCSt9vsCTRLSt9_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsSt9,
  lab = rownames(resultsSt9),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-15, 15),
  ylim = c(0, -log10(10e-75)),
  transcriptPointSize = 1,
  transcriptLabSize = 2.5,
  xlab = bquote(~Log[2]~ "fold change"),
  ylab = bquote(~Log[10]~adjusted~italic(P)),
  col=c("black", "black", "black", "blue")
)
dev.off()

## comparing NCSt17-NCSt9
ddsTx1$group = releval(ddsTx1$group, ref = "NC_St9")
##runs the differential analysis
ddsSt9 <- DESeq(ddsTx1)
## gets results after applying the logfoldshrinkage using apeglm method
resultsA = lfcShrink(ddsSt9, coef = "group_NC_St17_vs_NC_St9",
  type = c("apeglm"))
resultsA <- resultsA[order(resultsA$padj),]
head(resultsA)
summary(resultsA)
write.csv(as.matrix(resultsA), "NCSt17vsNCSt9.csv")

##comparing NCSt13 to CTRL St9
ddsNCSt13vsCTRLSt9 <- DESeq(ddsTx1)
##runs the differential analysis
## gets results after applying the logfoldshrinkage using apeglm method

```

```

resultsNCSt13vsCTRLSt9 = lfcShrink(ddsNCSt13vsCTRLSt9, coef =
"group_NC_St13_vs_CTRL_St9",
                                type = c("apeglm"))
resultsNCSt13vsCTRLSt9 <-
resultsNCSt13vsCTRLSt9[order(resultsNCSt13vsCTRLSt9$padj),]

write.csv(as.matrix(resultsNCSt13vsCTRLSt9), "NCSt13vsCTRLSt9.csv")

tiff("NCSt13vsCTRLSt9_MApplot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsNCSt13vsCTRLSt9, xlim = c(1, 1e5), ylim= c(-10,10), ylab= "Log2 Fold
Change")
dev.off()

tiff("NCSt13vsCTRLSt9_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsNCSt13vsCTRLSt9,
  lab = rownames(resultsNCSt13vsCTRLSt9),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-15, 15),
  ylim = c(0, -log10(10e-100)),
  transcriptPointSize = 1,
  transcriptLabSize = 2.5,
  xlab = bquote(~Log[2]~ "fold change"),
  ylab = bquote(~-Log[10]~adjusted~italic(P)),
  col=c("black", "black", "black", "blue")
)
dev.off()

## comparing CTRLtoNC-St13
ddsTxi$group = releval(ddsTxi$group, ref = "CTRL_St13")
##runs the differential analysis
ddsSt13 <- DESeq(ddsTxi)
## gets results after applying the logfoldshrinkage using apeglm method
resultsSt13 = lfcShrink(ddsSt13, coef = "group_NC_St13_vs_CTRL_St13",
                        type = c("apeglm"))
resultsSt13 <- resultsSt13[order(resultsSt13$padj),]
head(resultsSt13)
summary(resultsSt13)

write.csv(as.matrix(resultsSt13), "NCSt13vsCTRLSt13.csv")

tiff("NCSt13vsCTRLSt13_MApplot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsSt13, xlim = c(1, 1e5), ylim= c(-10,10), ylab= "Log2 Fold Change")

```

```

dev.off()

tiff("NCSt13vsCTRLSt13_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsSt13,
  lab = rownames(resultsSt13),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-10, 15),
  ylim = c(0, -log10(10e-100)),
  transcriptPointSize = 1,
  transcriptLabSize = 2.5,
  xlab = bquote(~Log[2]~ "fold change"),
  ylab = bquote(~-Log[10]~adjusted~italic(P)),
  col=c("black", "black", "black", "blue")
)
dev.off()

## comparing CTRLtoNC-St17
ddsTxigroup = releve(ddsTxigroup, ref = "CTRL_St17")
##runs the differential analysis
ddsSt17 <- DESeq(ddsTxigroup)
## gets results after applying the logfoldshrinkage using apeglm method
resultsSt17 = lfcShrink(ddsSt17, coef = "group_NC_St17_vs_CTRL_St17",
  type = c("apeglm"))
resultsSt17 <- resultsSt17[order(resultsSt17$padj),]
head(resultsSt17)
summary(resultsSt17)

write.csv(as.matrix(resultsSt17), "NCSt17vsCTRLSt17.csv")

tiff("NCSt17vsCTRLSt17_MApplot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsSt17, xlim = c(1, 1e5), ylim= c(-10,10), ylab= "Log2 Fold Change")
dev.off()

tiff("NCSt17vsCTRLSt17_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsSt17,
  lab = rownames(resultsSt17),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-15, 15),

```

```

        ylim = c(0, -log10(10e-100)),
        transcriptPointSize = 1,
        transcriptLabSize = 2.5,
        xlab = bquote(~Log[2]~ "fold change"),
        ylab = bquote(~Log[10]~adjusted~italic(P)),
        col=c("black", "black", "black", "blue")
    )
dev.off()

## comparing NCSt9toNC-St13
ddsTxigroup = releval(ddsTxigroup, ref = "NC_St9")
##runs the differential analysis
ddsNCSt9vsST13 <- DESeq(ddsTxigroup)
## gets results after applying the logfoldshrinkage using apeglm method
resultsNCSt9vsST13 = lfcShrink(ddsNCSt9vsST13, coef = "group_NC_St13_vs_NC_St9",
                              type = c("apeglm"))
resultsNCSt9vsST13 <- resultsNCSt9vsST13[order(resultsNCSt9vsST13$padj),]

write.csv(as.matrix(resultsNCSt9vsST13), "NCSt13vsNCSt9.csv")

tiff("NCSt13vsNCSt9_MApplot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsNCSt9vsST13, xlim = c(1, 1e5), ylim= c(-10,10), ylab= "Log2 Fold Change")
dev.off()

tiff("NCSt13vsNCSt9_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsNCSt9vsST13,
                lab = rownames(resultsNCSt9vsST13),
                x = "log2FoldChange",
                y = "padj",
                FCcutoff = 1,
                pCutoff = 0.05,
                xlim = c(-15, 15),
                ylim = c(0, -log10(10e-100)),
                transcriptPointSize = 1,
                transcriptLabSize = 2.5,
                xlab = bquote(~Log[2]~ "fold change"),
                ylab = bquote(~Log[10]~adjusted~italic(P)),
                col=c("black", "black", "black", "blue")
    )
dev.off()

## comparing NCSt13toNCSt17
ddsTxigroup = releval(ddsTxigroup, ref = "NC_St13")
##runs the differential analysis

```

```

ddsNCSt17vsSt13 <- DESeq(ddsTxi)
## gets results after applying the logfoldshrinkage using apeglm method
resultsNCSt17vsSt13 = lfcShrink(ddsNCSt17vsSt13, coef = "group_NC_St17_vs_NC_St13",
                                type = c("apeglm"))
resultsNCSt17vsSt13 <- resultsNCSt17vsSt13[order(resultsNCSt17vsSt13$padj),]

write.csv(as.matrix(resultsNCSt17vsSt13), "NCSt17vsSt13.csv")

tiff("NCSt17vsSt13_MAp1ot.tiff", height = 5, width = 5, units = 'in', res = 300)
plotMA(resultsNCSt17vsSt13, xlim = c(1, 1e5), ylim= c(-10,10), ylab= "Log2 Fold Change")
dev.off()

tiff("NCSt17vsSt13_Volcanoplot_2.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(resultsNCSt17vsSt13,
  lab = rownames(resultsNCSt17vsSt13),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-15, 15),
  ylim = c(0, -log10(10e-100)),
  transcriptPointSize = 1,
  transcriptLabSize = 2.5,
  xlab = bquote(~Log[2]~ "fold change"),
  ylab = bquote(~-Log[10]~adjusted~italic(P)),
  col=c("black", "black", "black", "blue")
)
dev.off()

St17vs13 = read.csv("St17vs13-forVolcano.csv", header = TRUE, row.names = 1)
tiff("17vs13_Volcanoplot.tiff", height = 5, width = 5, units = 'in', res = 300)
EnhancedVolcano(St17vs13,
  lab = rownames(St17vs13),
  x = "log2FoldChange",
  y = "padj",
  FCcutoff = 1,
  pCutoff = 0.05,
  xlim = c(-10, 10),
  ylim = c(0, -log10(10e-75)),
  transcriptPointSize = 1,
  transcriptLabSize = 2.5,
  xlab = bquote(~Log[2]~ "fold change"),
  ylab = bquote(~-Log[10]~adjusted~italic(P)),
  col=c("black", "black", "black", "blue")
)

```



```
dev.off()
```

```
select <- order(rowMeans(counts(ddsTx,normalized=TRUE)),
  decreasing=TRUE)[1:1000]
df <- as.data.frame(colData(dds)[,c("condition","type")])
pheatmap(assay(ntd)[select,], cluster_rows=FALSE, show_rownames=FALSE,
  cluster_cols=FALSE, annotation_col=df)
```

WGCNA – R script

```
library(WGCNA)
library(tidyverse)
library(pheatmap)
library(DESeq2)
library(RColorBrewer)
##essential command
options(stringsAsFactors = FALSE)
allowWGCNAThreads()

##load the data and give the rownames and the columnnames
geneData = read.csv("Top15K-Var-countsforWGCNA.csv", row.names = 1)
dim(geneData)
names(geneData)

## transpose the data
datExpr0 = as.data.frame(t(geneData))

##check if all the genes are good for the wgcna analysis
gsg = goodSamplesGenes(datExpr0, verbose = 3)
gsg$allOK

## check if there are any sample outliers by clustering
sampleTree = hclust(dist(datExpr0), method = "average")
tiff("sampleClustering_15KtopVar.tiff", height = 5, width = 5, units = 'in', res = 300)
par(cex = 0.6);
par(mar = c(0,4,2,0))
plot(sampleTree, main = "Sample clustering to detect outliers", sub="", xlab="", cex.lab = 1.5,
  cex.axis = 1.5, cex.main = 2)
dev.off()

## importing condition information
traitData = read.csv("traits.csv", row.names = 1)
traitDatainv = as.matrix(t(traitData))
```

```

dim(traitData)
names(traitData)
row.names(traitData)

Samples = rownames(datExpr0)
traitRows = row.names(traitData)

# Re-cluster samples
sampleTree2 = hclust(dist(datExpr0), method = "average")
# Convert traits to a color representation: white means low, red means high, grey means missing
entry
traitColors = numbers2colors(traitRows, signed = FALSE)
# Plot the sample dendrogram and the colors underneath.
tiff("sampleClusteringwithcategories_top15KVar.tiff", height = 5, width = 5, units = 'in', res =
300)
plotDendroAndColors(sampleTree2, traitColors,
                    groupLabels = names(traitData),
                    main = "Sample dendrogram and trait heatmap")
dev.off()

## saving all the files that have been created
save(datExpr0, traitData, file = "WGCNA-01-dataInput.RData")

# Choose a set of soft-thresholding powers
powers = c(c(1:10), seq(from = 12, to=20, by=2))
# Call the network topology analysis function
sft = pickSoftThreshold(datExpr0, powerVector = powers, verbose = 5)
# Plot the results:
sizeGrWindow(9, 5)
par(mfrow = c(1,2))
cex1 = 0.9;
# Scale-free topology fit index as a function of the soft-thresholding power
tiff("softThresholding1.tiff", height = 5, width = 5, units = 'in', res = 300)
plot(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
     xlab="Soft Threshold (power)",ylab="Scale Free Topology Model Fit,signed R^2",type="n",
     main = paste("Scale independence"))
text(sft$fitIndices[,1], -sign(sft$fitIndices[,3])*sft$fitIndices[,2],
     labels=powers,cex=cex1,col="red")
abline(h=0.90,col="red")
dev.off()
# this line corresponds to using an R^2 cut-off of h

# Mean connectivity as a function of the soft-thresholding power
tiff("softThresholding2.tiff", height = 5, width = 5, units = 'in', res = 300)

```

```

plot(sft$fitIndices[,1], sft$fitIndices[,5],
     xlab="Soft Threshold (power)", ylab="Mean Connectivity", type="n",
     main = paste("Mean connectivity"))
text(sft$fitIndices[,1], sft$fitIndices[,5], labels=powers, cex=cex1, col="red")
dev.off()

bwnet = blockwiseModules(datExpr0, maxBlockSize = 15000,
                        power = 8, TOMType = "signed", minModuleSize = 30,
                        reassignThreshold = 0, mergeCutHeight = 0.25,
                        numericLabels = TRUE,
                        saveTOMs = TRUE,
                        saveTOMFileBase = "NCRnaseq-Top15KVar-signed",
                        verbose = 3)
table(bwnet$colors)

tiff("geneDendrogram-top15KVar.tiff", height = 5, width = 5, units = 'in', res = 300)
# Convert labels to colors for plotting
mergedColors = labels2colors(bwnet$colors)
# Plot the dendrogram and the module colors underneath
plotDendroAndColors(bwnet$dendrograms[[1]], mergedColors[bwnet$blockGenes[[1]]],
                    "Module colors",
                    dendroLabels = FALSE, hang = 0.03,
                    addGuide = TRUE, guideHang = 0.05)
dev.off()

moduleLabels = bwnet$colors
moduleColors = labels2colors(bwnet$colors)
MEs = bwnet$MEs
geneTree = bwnet$dendrograms[[1]]
save(MEs, moduleLabels, moduleColors, geneTree,
     file = "Rnaseq-02-networkConstruction-auto.RData")

# Define numbers of genes and samples
nGenes = ncol(datExpr0)
nSamples = nrow(datExpr0)
# Recalculate MEs with color labels
MEs0 = moduleEigengenes(datExpr0, moduleColors)$eigengenes
MEs = orderMEs(MEs0)
moduleTraitCor = cor(MEs, traitData, use = "p")
moduleTraitPvalue = corPvalueStudent(moduleTraitCor, nSamples)

# Will display correlations and their p-values
tiff("moduletraitrelationship-Top15KVar.tiff", height = 5, width = 5, units = 'in', res = 300)
textMatrix = paste(signif(moduleTraitCor, 2), "\n(",

```

```

        signif(moduleTraitPvalue, 1), "")", sep = "");
dim(textMatrix) = dim(moduleTraitCor)
par(mar = c(6, 8.5, 3, 3));
# Display the correlation values within a heatmap plot
labeledHeatmap(Matrix = moduleTraitCor,
  xLabels = names(traitData),
  yLabels = names(MEs),
  ySymbols = names(MEs),
  colorLabels = FALSE,
  colors = blueWhiteRed(50),
  textMatrix = textMatrix,
  setStdMargins = FALSE,
  cex.text = 0.2,
  cex.lab = 0.5,
  zlim = c(-1,1),
  main = paste("Module-trait relationships"))
dev.off()

geneset1 = names(datExpr0)[moduleColors == "blue"]
write.csv(geneset1, "moduleblue-NCSt13.csv")

geneset2 = names(datExpr0)[moduleColors == "yellow"]
write.csv(geneset2, "moduleyellow-NCSt17.csv")

geneset3 = names(datExpr0)[moduleColors == "black"]
write.csv(geneset3, "moduleblack-WTSt17.csv")

geneset4 = names(datExpr0)[moduleColors == "brown"]
write.csv(geneset4, "modulebrown-WTSt17.csv")

geneset5 = names(datExpr0)[moduleColors == "green"]
write.csv(geneset5, "modulegreen-WTSt13.csv")

library(ggplot2)
library(pheatmap)
library(RColorBrewer)

set1 = geneData
set1a = set1 + 0.001
set1b = log10(set1a)
head(set1b)
set1c = na.omit(set1b)
library(gplots)
library(RColorBrewer)

```

```

hmc col= colorRampPalette(brewer.pal(9,"YlGnBu"))(255)

tiff("NCRnaseq_heatmap_New.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(as.matrix(set1c), col = hmc col, trace = "none", fontsize_row = 1,
          fontsize_col = 10, scale = "row")
dev.off()

tiff("NCRnaseq_heatmap3.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(as.matrix(geneData), col = hmc col, trace = "none", fontsize_row = 1, fontsize_col = 10,
          scale = "row",
          clustering_distance_rows = "euclidean", clustering_distance_cols = "euclidean")
dev.off()

# Calculate topological overlap anew: this could be done more efficiently by saving the TOM
# calculated during module detection, but let us do it again here.
dissTOM = 1-TOMsimilarityFromExpr(datExpr0, power = 8)
# Transform dissTOM with a power to make moderately strong connections more visible in the
heatmap
plotTOM = dissTOM^7
# Set diagonal to NA for a nicer plot
diag(plotTOM) = NA
# Call the plot function
tiff("TOM_plot.tiff", height = 5, width = 5, units = 'in', res = 300)
TOMplot(plotTOM, geneTree, moduleColors, main = "Network heatmap plot, all genes")
dev.off()

# Recalculate module eigengenes
MEs = moduleEigengenes(datExpr0, moduleColors)$eigengenes
# Isolate weight from the clinical traits
condition = as.data.frame(traitData$);
names(weight) = "weight"
# Add the weight to existing module eigengenes
MET = orderMEs(cbind(MEs, weight))
# Plot the relationships among the eigengenes and the trait
tiff("relationships.tiff", height = 5, width = 5, units = 'in', res = 300)
par(cex = 0.9)
plotEigengeneNetworks(MEs, "", marDendro = c(0,4,1,2), marHeatmap = c(3,4,1,2), cex.lab =
0.8, xLabelsAngle
= 90)
dev.off()
# Plot the dendrogram
tiff("dendrogram.tiff", height = 5, width = 5, units = 'in', res = 300)
par(cex = 1.0)
plotEigengeneNetworks(MEs, "Eigengene dendrogram", marDendro = c(0,4,2,0),

```

```

        plotHeatmaps = FALSE)
dev.off()
# Plot the heatmap matrix (note: this plot will overwrite the dendrogram plot)
tiff("adjacencymatrix.tiff", height = 5, width = 5, units = 'in', res = 300)
par(cex = 1.0)
plotEigengeneNetworks(MEs, "Eigengene adjacency heatmap", marHeatmap = c(3,4,2,2),
        plotDendrograms = FALSE, xLabelsAngle = 90)
dev.off()

chooseTopHubInEachModule(
    datExpr0,
    moduleColors,
    omitColors = "grey",
    power = 8,
    type = "signed")

## heatmaps for module blue and module yellow

set1 = geneData[c(as.matrix(geneset1)),]
head(set1)
set1a = set1 + 0.001
set1b = log10(set1a)
head(set1b)
set1c = na.omit(set1b)
library(gplots)
library(RColorBrewer)
hmc1=colorRampPalette(brewer.pal(9,"YlGnBu"))(255)

tiff("NC13-modulebluegene_heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(as.matrix(set1c),col = hmc1, trace = "none", fontsize_row = 1,
        fontsize_col = 10, scale = "row", show_rownames = FALSE, cluster_cols = FALSE)
dev.off()

set2 = geneData[c(as.matrix(geneset2)),]
head(set2)
set2a = set2 + 0.001
set2b = log10(set2a)
head(set2b)
set2c = na.omit(set2b)
hmc2=colorRampPalette(brewer.pal(9,"YlGnBu"))(255)
tiff("NC17-moduleyellowgene_heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(as.matrix(set2c),col = hmc2, trace = "none", fontsize_row = 1,
        fontsize_col = 10, scale = "row", show_rownames = FALSE, cluster_cols = FALSE)
dev.off()

```

PCA – R script

```

library(ggplot2)
library(RColorBrewer)
library(reshape2)
library(pracma)
library(elasticnet)
library(pheatmap)
library(factoextra)

## inputing the TPM matrix
data = read.csv("/Users/arao/Documents/LAB/Genomics/RNASEQ-DATA/NC-
RNASEQ/PCA/FilteredTPM3.csv",
               header = TRUE, row.names = 1)
##getting the transpose of the matrix
tdata = as.data.frame(t(data))

#performin the PCA - with scaling
pc = prcomp(tdata, scale. = TRUE)

##looking at the results of the PCA
names(pc)
print(pc)
pc$x[,1]
eigvecs = pc$rotation
Scaledeigvecs = Scaledpc$rotation
Scaledeigvecs[,1]
eigvecs[,1]

## rotation is the eigenvectors
## sdev is the square root of the eigvalues
## x is the dot product of the PCA eigvectors and the original data
## x are the principal components
## x = scale(as.matrix(tdata)) %*% eigvecs [calculated manually]

##plotting the first pc
plot(pc$x[,1], type = "l", main = "")

##plotting PC1 vs PC2 or PC2 vs PC3- two different plotting method
plot(pc$x[, 2], pc$x[, 3], type = "p", main = "PCA", xlab = "PC1", ylab = "PC2",
     col = "red")
ggplot(as.data.frame(pc$x) , aes( x = pc$x[, 2], y = pc$x[, 3])) + geom_point() +
geom_text(label = row.names (pc$x), = TRUE)

```

```

##plotting the variance explained - also called the scree plot - using inbuilt plot function
p.variance.explained <- pc$sdev^2 / sum(pc$sdev^2)
tiff("scree-plot.tiff", height=5, width = 5, units = 'in', res = 300)
barplot(100*p.variance.explained, las=2, xlab="", ylab="% Variance Explained")
dev.off()

## using factoextra package that makes better PCA based plots
tiff("Screeplot-New.tiff", height=5, width = 5, units = 'in', res = 300)
fviz_eig(pc)
dev.off()

## using factoextra package that makes better PCA based plots - PC1 vs PC2
tiff("PC1vsPC2.tiff", height=5, width = 5, units = 'in', res = 300)
fviz_pca_ind(pc,
             col.ind = "blue",
             repel = TRUE)+ labs(title="PCA", x = "PC1", y = "PC2")
dev.off()

## creating a null distribution from the data
i = 1
tdata = as.matrix(tdata)
randomizeeigvalssq = matrix(,18)
for(i in 1:1000)
{
  randomtdata = apply(tdata,2,sample)
  rpc = prcomp(randomtdata, scale. = TRUE)
  eigvals = rpc$sdev
  randomizeeigvalssq = rbind(randomizeeigvalssq,c(eigvals))
}

SqREigvals = as.data.frame(randomizeeigvalssq[2:1001, ])
SqREigvals = SqREigvals[,1:17]
SqREigvals1 = SqREigvals^2
SqREigvals1 = SqREigvals1/30186
d = density(as.matrix(SqREigvals1))
hist(as.matrix(SqREigvals1), xlab = "Eigen Values/N")
tiff("randomizeddataEigvals.tiff", height=5, width = 5, units = 'in', res = 300)
plot(d, xlab= "Eigen values", col="blue", main = "Histogram of randomized data", xlim =
c(20,110))
polygon(d, col = "blue")
dev.off()

SqREigvals1= scale(SqREigvals1)
d1 = density(as.matrix(SqREigvals1))
plot(d1)

```



```

hmc=colorRampPalette(brewer.pal(9,"YlGnBu"))(255)
dim(pc$x)
pheatmap(t(pc$x))
pc1loading = pc$rotation[,1]
pc1loading = as.matrix(pc1loading)
tiff("PCA-space-heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(t(pc$x),col = hmc,
         trace = "none", fontsize_row = 10, fontsize_col = 10,
         cluster_rows = FALSE)
dev.off()
eigvec1 = pc$x[,1]
plot(eigvec1)
PCAprj = as.matrix(pc$x)

write.csv(PCAprj, "PCAprj.csv")

```

```

hmc=colorRampPalette(brewer.pal(9,"YlGnBu"))(255)
dim(pc$x)
pheatmap(t(pc$x))
pc1loading = pc$rotation[,1]
pc1loading = as.matrix(pc1loading)
tiff("PCAnoscale-space-heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(t(pc$x),col = hmc,
         trace = "none", fontsize_row = 10, fontsize_col = 10,
         cluster_rows = FALSE)
dev.off()

loadings = pc$x
loadings = loadings[c("CTRL.St9.Rep1","CTRL.St9.Rep2","CTRL.St9.Rep3",
"NC.St9.Rep1","NC.St9.Rep2","NC.St9.Rep3",

"CTRL.St13.Rep1","CTRL.St13.Rep2","CTRL.St13.Rep3","NC.St13.Rep1","NC.St13.Rep2","
NC.St13.Rep3",
"CTRL.St17.Rep1","CTRL.St17.Rep2","CTRL.St17.Rep3", "NC.St17.Rep1",
"NC.St17.Rep2", "NC.St17.Rep3"),]

tiff("PCA-nocluster-heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(t(loadings),col = hmc,
         trace = "none", fontsize_row = 10, fontsize_col = 10,
         cluster_rows = FALSE, cluster_cols = FALSE)
dev.off()
plot(pc)

```

```

cov = cov(tdata)

## sparse PCA
##scaling tdata because sparsePCA does not have automatic scaling functionality
sdata = scale(tdata)
mean(sdata[,1])
std(sdata[,1])

## identifying the penalty parameters from the data
## manually identified the upper limit to be 2000 - by which the first eig vec all components are
zero
colnames(comp) = ("eigvec1", "eigvec2", "eigvec3")
comp = matrix(.,3)
pev = matrix(.,3)
i = 0
for (i in 0:2000)
{
  tspc = arrayspc(x = sdata, K =3,
    para = c(i,i,i),
    trace = TRUE)
  tspcloadings = tspc$loadings
  num1 = sum(tspcloadings[,1] != 0)
  num2 = sum(tspcloadings[,2] != 0)
  num3 = sum(tspcloadings[,3] != 0)
  comp = rbind(comp, c(num1,num2,num3))
}

comp = matrix(.,3)
pev = matrix(.,3)
i = 0
for (i in 0:2000)
{
  Trialspc = arrayspc(x = sdata, K =3,
    para = c(i,i,i),
    trace = TRUE)
  pev = rbind(pev, Trialspc$pev)
  Trialspcloadings = Trialspc$loadings
  num1 = sum(Trialspcloadings[,1] != 0)
  num2 = sum(Trialspcloadings[,2] != 0)
  num3 = sum(Trialspcloadings[,3] != 0)
  comp = rbind(comp, c(num1,num2,num3))
}
comp = comp[2:2000,]
comp = cbind(comp, c(0:1998))
colnames(comp) = c("PC1", "PC2", "PC3", "lambda")

```

```

comp = as.data.frame(comp)
pev = pev[2:2000,]
pev = cbind(pev, c(0:1998))
pev = as.data.frame(pev)
colnames(pev) = c("PC1", "PC2", "PC3", "lambda")

tspc$pev
print(tspc)

tiff("Penaltyparameters.tiff", height = 5, width = 5, units = 'in', res = 300)
plot(comp$lambda, log(comp$PC1), type = "l", lwd = 3, xlab = "Lambda - Penalty parameter",
      ylab = "Log(non zero components)", col = "red", lty = 1)
lines(comp$lambda, log(comp$PC2), lwd = 3, type = "l", lty = 3, col = "blue")
lines(comp$lambda, log(comp$PC3), type = "l", lwd = 3, lty = 4, col = "green")
legend(0.2, legend = c("PC1", "PC2", "PC3"), col = c("red", "blue", "green"), lty = c(1, 2, 3), cex =
0.8)
dev.off()

tiff("PEV.tiff", height = 5, width = 5, units = 'in', res = 300)
plot(pev$lambda, pev$PC1, type = "l", xlab = "Lambda - penalty parameter",
      ylab = "Percentage of Explained Variance", col = "red", lty = 1, lwd = 3)
lines(pev$lambda, pev$PC2, type = "l", lty = 2, col = "blue", lwd = 3)
lines(pev$lambda, pev$PC3, type = "l", lty = 3, col = "green", lwd = 3)
legend(1600, 0.4, legend = c("PC1", "PC2", "PC3"), col = c("red", "blue", "green"), lty = c(1, 2, 3),
cex = 0.8)
dev.off()

tiff("PEVvsComponents.tiff", height = 5, width = 5, units = 'in', res = 300)
plot(comp$PC1, pev$PC1, type = "l", xlab = "No. of non-zero components",
      ylab = "Percentage of Explained Variance", col = "red", lty = 1, lwd = 3)
lines(comp$PC2, pev$PC2, type = "l", lty = 2, col = "blue", lwd = 3)
lines(comp$PC3, pev$PC3, type = "l", lty = 3, col = "green", lwd = 3)
legend(1600, 0.4, legend = c("PC1", "PC2", "PC3"), col = c("red", "blue", "green"), lty = c(1, 2, 3),
cex = 0.8)
dev.off()

ggplot(data = comp, aes(x=lambda, y = PC1 + geom_line() + geom_point()))
dotmatrix = as.data.frame(scale(as.matrix(tdata)) %*% tspcloadings)
dotmatrix = dotmatrix[c("CTRL.St9.Rep1", "CTRL.St9.Rep2", "CTRL.St9.Rep3",
"NC.St9.Rep1", "NC.St9.Rep2", "NC.St9.Rep3",
"CTRL.St13.Rep1", "CTRL.St13.Rep2", "CTRL.St13.Rep3", "NC.St13.Rep1", "NC.St13.Rep2", "
NC.St13.Rep3",
"CTRL.St17.Rep1", "CTRL.St17.Rep2", "CTRL.St17.Rep3", "NC.St17.Rep1",
"NC.St17.Rep2", "NC.St17.Rep3"),]

```

```

tspc = arrayspc(x = sdata, K = 3,
               para = c(1600,830,680),
               trace = TRUE)
pev = rbind(pev, tspc$pev)
print(tspc)
tspcloadings = tspc$loadings
num1 = sum(tspcloadings[,1] != 0)
num2 = sum(tspcloadings[,2] != 0)
num3 = sum(tspcloadings[,3] != 0)

tiff("Nsparsecomponents-nocluster-heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(t(dotmatrix),col = hmcol,
         trace = "none", fontsize_row = 10, fontsize_col = 10, cellwidth = 10, cellheight = 10,
         cluster_rows = FALSE, cluster_cols = FALSE)
dev.off()
row.names(tspcloadings)= row.names(data)

write.csv(tspcloadings, "spclloadingsNew.csv")

```

Heatmaps – R script

```

library(ggplot2)
library(pheatmap)
library(RColorBrewer)

data = read.csv("IBET_countdata_filtered.csv", header = TRUE, row.names = 1)
genes = read.table("IBETonlygenes.txt")

set1 = data[c(as.matrix(genes)),]
head(set1)
set1a = set1 + 0.001
set1b = log10(set1a)
head(set1b)
set1c = na.omit(set1b)
hmcol=colorRampPalette(brewer.pal(9,"Blues"))(255)

tiff("IBETonlygenes_heatmap.tiff", height = 5, width = 5, units = 'in', res = 300)
pheatmap(as.matrix(set1c),col = hmcol, trace = "none", fontsize_row = 1, fontsize_col = 10,
scale = "row")
dev.off()

```

ChIP-Seq Analysis

Alignment using Bowtie2 – Shell script

Building reference genome

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=02:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N bowtie2
#MSUB -j oe

module load bowtie2/2.2.6
module load tophat/2.1.0
module load samtools
module load boost
module load gcc/4.8.3
module load java
module load fastqc/0.11.5
module load homer
## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR

bowtie2-build ../../genomefiles/XL9_2.fa ../bowtieindex/xeno92genome
```

Alignment

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=02:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N bowtie2
#MSUB -j oe

module load bowtie2/2.2.6
module load tophat/2.1.0
module load samtools
```

```

module load boost
module load gcc/4.8.3
module load java
module load fastqc/0.11.5
module load homer
## These are shell commands. Note that all MSUB commands come first.

```

```
cd $PBS_O_WORKDIR
```

```

bowtie2 -t -p 6 -q -x ../bowtieindex/xeno92genome -U fastqfiles/WTSt13_1_Chip.fastq -S
samfiles/WTSt13_1_Chip.sam
bowtie2 -t -p 6 -q -x ../bowtieindex/xeno92genome -U fastqfiles/WTSt13_1_input.fastq -S
samfiles/WTSt13_1_input.sam

```

Filter Bam files – Shell script

```

#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=02:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=1:ppn=6
#MSUB -N bowtie2
#MSUB -j oe

module load bowtie2/2.2.6
module load tophat/2.1.0
module load samtools
module load boost
module load gcc/4.8.3
module load java
module load fastqc/0.11.5
module load homer
## These are shell commands. Note that all MSUB commands come first.

```

```
cd $PBS_O_WORKDIR
```

```

samtools view -F 0x04 -b samfiles/WTSt13_1_Chip_sorted.bam >
finalbamfiles/WTSt13_1_Chip_FnS.bam
samtools view -F 0x04 -b samfiles/WTSt13_2_Chip_sorted.bam >
finalbamfiles/WTSt13_2_Chip_FnS.bam
samtools view -F 0x04 -b samfiles/WTSt13_1_input_sorted.bam >
finalbamfiles/WTSt13_1_input_FnS.bam
samtools view -F 0x04 -b samfiles/WTSt13_2_input_sorted.bam >
finalbamfiles/WTSt13_2_input_FnS.bam

```

Deeptools – Shell script

Convert Bam to BigWig

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=10:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=2:ppn=10
#MSUB -N deeptools
#MSUB -j oe

module load samtools
module load python
module load deeptools
## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR

bamCoverage -b finalbamfiles/WTSt13_1_Chip_FnS.bam \
-o bigWig/WTSt13_1.bw \
-of bigwig \
--normalizeUsing RPGC \
--effectiveGenomeSize 2300000000 \
-p max
```

computeMatrix

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=10:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=2:ppn=10
#MSUB -N deeptools
#MSUB -j oe

module load samtools
module load python
module load deeptools
## These are shell commands. Note that all MSUB commands come first.

cd $PBS_O_WORKDIR
```

```
computeMatrix scale-regions \
-R IBETMBT1500bparoundTSS.bed MBTnoIBET_1500bparoundTSS.bed \
-S bigWig/WTSt9_1.bw bigWig/WTSt9_1_input.bw bigWig/WTSt9_2.bw
bigWig/WTSt9_2_input.bw \
-o vizfigures/matrixIBETMBT_TSS3kb.gz \
--outFileSortedRegions vizfigures/matrixIBETMBT_TSS3kb.bed
```

plotHeatmaps

```
#!/bin/bash
#MSUB -A b1042
#MSUB -q genomics
#MSUB -l walltime=10:00:00
#MSUB -M anjalirao2017@u.northwestern.edu
#MSUB -l nodes=2:ppn=10
#MSUB -N deeptools
#MSUB -j oe

module load samtools
module load python
module load deeptools
## These are shell commands. Note that all MSUB commands come first.
```

```
cd $PBS_O_WORKDIR
```

```
plotHeatmap -m vizfigures/matrixIBETMBT_TSS3kb.gz \
--outFileName vizfigures/matrixIBETMBT_TSS3kb.png \
--dpi 300 \
--colorMap YlGnBu \
--samplesLabel WTSt9_Rep1 WTSt9_Rep1_Input WTSt9_Rep2 WTSt9_Rep2_Input \
--startLabel 1.5kb \
--endLabel 1.5kb \
--xAxisLabel DistancefromTSS
```

Generating Upstream sequences for motif analysis – Rscript

```
library(BiocInstaller)
library(GenomicFeatures)
library(Rsamtools)

gfffile = ("/Users/arao/Documents/LAB/Genomics/Genome/XENLA_9.2_Xenbase2.gff3")
txdb_gff = makeTxDbFromGFF(file=gfffile, format = c("gff3"), dataSource = "Xenbase",
organism = "Xenopus laevis", circ_seqs = DEFAULT_CIRC_SEQS,
dbxrefTag = "gene")
```



```

siggenes = read.table("siggenes.txt")

# # get a list of transcripts associated with each gene
tx_by_gene <- transcriptsBy(txdb_gff, 'gene')
# # reduce transcripts on each locus
genic.tx <- reduce(tx_by_gene)
# # convert from GRangesList to Granges
genes.gr <- unlist(genic.tx)
## Change to whatever distance you want up/down
TSS <- promoters(genes.gr, upstream=1500, downstream = 1500)
##Than write to a bed file.
TSS.df = as.matrix(TSS)
export.bed(TSS, "1500bparoundTSS-allgenes.bed")

```

Proteomics Analysis – Python Script

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import plotly
import plotly.offline as py
import plotly.graph_objs as go
from plotly.offline import download_plotlyjs, init_notebook_mode, plot, iplot
import cufflinks as cf
import seaborn as sns

```

```
WholeEmbryodata = pd.read_csv("WholeEmbryodataset.csv", index_col='Mark' )
```

```
subWEdata = WholeEmbryodata[WholeEmbryodata['Modification'] != "UN"]
```

```
subWEdata.head()
```

```
WEdataforplot = subWEdata[['Stage 9','Stage 10','Stage 10.5','Stage 11','Stage 12','Stage 13']]
```

```
WEdataforplot.head()
```

```

f, ax = plt.subplots(figsize=(10, 6))
corr = WEdataforplot.corr()
hm = sns.heatmap(round(corr,2), annot=True, ax=ax, cmap="coolwarm",fmt='.2f',
                  linewidths=.05)
f.subplots_adjust(top=0.93)
t= f.suptitle('Stage Wise Correlation Heatmap', fontsize=14)

```

```

cols = ['Stage 9', 'Stage 10', 'Stage 10.5', 'Stage 11', 'Stage 12', 'Stage 13']
pp = sns.pairplot(WEdataforplot[cols], size=1.8, aspect=1.8,
                  plot_kws=dict(edgecolor="k", linewidth=0.5),
                  diag_kind="kde", diag_kws=dict(shade=True))

fig = pp.fig
fig.subplots_adjust(top=0.93, wspace=0.3)
t = fig.suptitle('Stage Pairwise Plots', fontsize=14)

WEdataforplot.hist()

transpose = pd.DataFrame.transpose(WEdataforplot)

transpose.head()

f, ax = plt.subplots(figsize=(30, 16))
corr = transpose.corr()
hm = sns.heatmap(round(corr,2), annot=True, ax=ax, cmap="coolwarm",fmt='.2f',
                  linewidths=.05)
f.subplots_adjust(top=0.93)
t= f.suptitle('Correlation Heatmap', fontsize=14)

Stage9 = WEdataforplot['Stage 9']

Stage9 = Stage9.sort_values(ascending = False)

Stage9.head()

max = WEdataforplot.max(axis=1)

min = WEdataforplot.min(axis=1)

mean = WEdataforplot.mean(axis=1)

stdev = WEdataforplot.std(axis =1)

change = (max - min)

change = change.sort_values(ascending=False)

change10 = change[0:10]

change10 = pd.DataFrame(change[0:10])

change10

```

```

short = change10[['Mark']]

zscore = pd.read_csv("zscore.csv")

zscore = zscore[['Mark','Stage 9','Stage 10','Stage 10.5','Stage 11','Stage 12','Stage 13']]

x = ['H3: K23AC', 'H3: K23AC', 'H3: K14AC', 'H3.3: K36ME2', 'H3: K79ME1','H3.1:
K36ME2',
'H3.3: K27ME1',
'H3.1: K27ME1',
'H4: K20ME2',
'H3: K9ME2',
'H4: K20ME1' ]

zscore_sub = zscore.loc[zscore['Mark'].isin(x)]

zscore_sub

zscore_sub = zscore_sub[['Mark','Stage 9','Stage 10','Stage 10.5','Stage 11','Stage 12','Stage 13']]

f, ax = plt.subplots(figsize=(15, 8))
plt.title('Top 10 most changed histone modifications')
pd.plotting.parallel_coordinates(
    zscore_sub,'Mark', color = ('black',
    'brown','blue','green','yellow','darkmagenta','aqua','deeppink','teal','orange'))
plt.savefig('parallel_coordinates.png')

f, ax = plt.subplots(figsize=(15, 8))
plt.title('All histone modifications')
pd.plotting.parallel_coordinates(
    zscore,'Mark')
ax.legend().remove()
plt.savefig('parallel_coordinates_All.png')

color = sns.color_palette("RdBu_r", 7)
s = sns.clustermap(WEdatatorplot, col_cluster = False, z_score = 0, cmap = 'mako_r',robust =
'True', figsize = (15,15))
plt.savefig('heatmap.png')

```