January 2023

Understanding NANOG's Role During Cell Differentiation

Oluwatobiloba Samuel Aminu

UNDERSTANDING NANOG'S ROLE DURING CELL DIFFERENTIATION

by

Oluwatobiloba Samuel Aminu

Bachelor of Science, University of Lagos, Nigeria (2018)

A Thesis

Submitted to the Graduate Faculty

Of the

University of North Dakota

In partial fulfillment of the requirements

For the degree of

Master of Science, Biomedical Sciences

Grand Forks, North Dakota

August, 2023

Name: Oluwatobiloba Aminu

Degree: Master of Science

This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Junguk Hur
1C49E6598068480...
Junguk Hur

DocuSigned by:
Motoki Takaku
EEA02046E2E4415...
Motoki Takaku

DocuSigned by:
Susan Eliazer
9E628E5D1DE0489...
Susan Eliazer

DocuSigned by:
Manu Manu
E0D49834CCAC470...
Manu Manu

This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C733403...
Chris Nelson
Dean of the School of Graduate Studies

7/6/2023
Date

PERMISSION

Title: UNDERSTANDING NANOG's ROLE DURING CELL DIFFERENTIATION

Department: Biomedical Sciences

Degree: Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Oluwatobiloba Samuel Aminu

August, 2023

TABLE OF CONTENTS

LIST OF FIGURES vi

ACKNOWLEDGEMENTS vii

DEDICATION ix

ABSTRACT x

INTRODUCTION AND LITERATURE REVIEW 1

LITERATURE REVIEW 3

NAIVE TO PRIMED PLURIPOTENCY 16

METHODS 20

RESULTS 23

DISCUSSION 44

NEUROECTODERMAL DIFFERENTIATION 47

METHODS 48

RESULTS 50

DISCUSSION 55

FUTURE DIRECTIONS 57

REFERENCES 59

LIST OF FIGURES

Figure 1. STAGES OF EARLY DEVELOPING EMBRYO....4

Figure 2. DEVELOPMENTAL POTENTIAL OF MOUSE ESC....6

Figure 3. PLURIPOTENCY GENE NETWORK....8

Figure 4. NANOG'S CHANGING EXPRESSION PATTERN....11

Figure 5. CURRENT MODEL OF PLURIPOTENCY....13

Figure 6. NAIVE TO PRIMED PLURIPOTENCY MODEL....18

Figure 7. METHODOLOGY NAIVE TO PRIMED TRANSITION....21

Figure 8. DISTANCE MATRIX....24

Figure 9. PRINCIPAL COMPONENT ANALYSIS PLOT....25

Figure 10. REPRESENTATIVE GENES....27

Figure 11. NANOG DIFFERENTIALLY EXPRESSED GENES....28

Figure 12. NANOG CISTROME....30

Figure 13. DIFFERENTIAL NANOG BINDING AND CHROMATIN ACCESSIBILITY HEATMAP....32

Figure 14. NAIVE ESC MOTIFS....34

Figure 15. DAY 1 MOTIFS....35

Figure 16. DAY 2 MOTIFS....36

Figure 17. NANOG REGULATED GENES....38

Figure 18. STAGE SPECIFIC EMBRYONIC DEVELOPMENT GENES (DEVELOPING EMBRYO)....

40

Figure 19. STAGE-SPECIFIC EMBRYONIC DEVELOPMENT GENES (NANOG).....41

Figure 20. BIOLOGICAL PROCESS FROM NAÏVE TO PRIMED STATE 42

Figure 21. METHODOLOGY NEUROECTODERMAL DIFFERENTIATION 49

Figure 22. NANOG CISTROME IN NEUROECTODERMAL DIFFERENTIATION.....51

Figure 23. ENRICHED MOTIFS AT UNINDUCED STAGE 53

Figure 24. ENRICHED MOTIFS AT DIFFERENTIATED STAGE 54

Figure 25. ENRICHMENT ANALYSIS OF GENES CLOSEST TO LHX1 MOTIFS 54

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to several individuals who have played a significant role in the completion of my thesis. First and foremost, I am deeply grateful to my supervisor, Dr. Hur, for his invaluable guidance and support throughout the entire research process. His mentorship has been instrumental in shaping the direction of my work.

I would also like to extend my appreciation to the members of my committee, particularly Dr. Takaku, whose insightful feedback and constructive criticism have been crucial in advancing my research. I am thankful to Dr. Eliazer for her constant support and to Dr. Manu, whose innovative thinking has been an inspiration to me.

I would like to express my sincere gratitude to the members of my laboratory, Christianah, Brett, Hasin, and Kai, for fostering a nurturing and inclusive environment where I could freely pursue scientific inquiry and learn from my mistakes.

I am indebted to GB for engaging in countless hours of stimulating scientific discussions, which have greatly enriched my understanding and perspective. I also thank Dr. Nechaev for his impromptu scientific conversations.

Lastly, I would like to acknowledge and thank all my friends and classmates for their constant support and motivation. Together, we have pushed each other to strive for excellence and become better versions of ourselves. To all these individuals, I say thank you for making this journey a fulfilling and enriching experience.

DEDICATION

To God, for granting me the strength, health, and perseverance to complete this thesis.

To my parents, for their unwavering support throughout my academic pursuits.

To my siblings, for their love and encouragement.

To myself, for showing up consistently.

ABSTRACT

Pluripotency is characterized by the capacity to self-renew and differentiate into all cell types. This state relies on both extrinsic factors, such as leukemia inhibitory factor and bone morphogenetic protein, and intrinsic factors, including pluripotent transcription factors.

While the role of pluripotent transcription factors in stem cell maintenance is well understood, their functions beyond pluripotency remain unclear. In this study, we focused on NANOG, a core member of the pluripotency network, and investigated its roles in two different models of differentiation: the transition from the naive to the primed pluripotent state and neuroectodermal differentiation.

Our findings demonstrate that NANOG plays a dual role. It functions as a pluripotency maintenance factor during the undifferentiated state and can direct mesodermal specification during early differentiation. We propose that this mechanism is achieved through the indirect recruitment of co-transcription factor LHX1. These results reinforce the fundamental concept of pluripotent transcription factors such as NANOG in governing cell fate decisions and subsequent developmental programs.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Molecular biology has long been plagued by a question that has eluded scientists for decades: how does a single cell transform into a fully functioning organism? Developmental biology, a field dedicated to uncovering the mysteries of this process, is partly focused on understanding the mechanisms that dictate cell fate determination. In essence, cell fate decision refers to the complex processes that determine the final type of cell a given cell will become. The process of pluripotency and differentiation, two distinct but related processes, are critical for understanding how a cell's fate is determined. By shedding light on this intricate process, we can gain a greater understanding of how a cell transitions to a specific fate.

The commitment of a cell to a specific fate is a complex process that involves two stages: specification and determination.[1] When a cell is specified to a particular fate, it is still reversible and can transition to another cell type. However, once a cell is determined, its fate is irreversible and will differentiate into a specific cell type.

Transcription factors play a critical role in regulating gene expression during cell fate determination. They can work in complex combinations, each binding to specific sites in DNA to activate or repress target genes. Different combinations of transcription factors are required for different cell types, and the precise timing and level of expression are essential for accurate cell fate decisions. As a cell progresses from a less restrictive to a more restrictive state during fate determination, transcription factors bind to regulatory regions in DNA and shape the gene expression landscape, ultimately

leading to the expression of genes that define a cell's fate. The interplay between these factors, along with external signaling cues and environmental factors, guides a cell towards its final fate. However, before a cell fate is specified, it must undergo a series of developmental transitions, starting from its pluripotent condition and through intermediate states into cell lineages that activate subsequent developmental events.

This thesis aims to investigate the role of NANOG in cell fate determination, focusing on two models of differentiation; naive to primed pluripotency and neuroectodermal differentiation. To achieve this, we will employ a comprehensive approach integrating multiple omics techniques, including Assay for Transposable Accessible Chromatin (ATAC), Chromatin Immunoprecipitation (ChIP), and RNA sequencing. Through this approach, we seek to enhance our understanding of NANOG's involvement in the differentiation process. Ultimately, our work will contribute to the growing body of evidence of pluripotent factors acting as lineage specifiers.

LITERATURE REVIEW

EMBRYOGENESIS

Embryogenesis begins with the formation of a zygote resulting from the fusion of male sperm cells and female oocytes. The zygote undergoes multiple rounds of cell division until it reaches the 8 - 16 cell stage, after which it forms the compacted morula (Figure 1), usually around four days post-conception.[2] During the early stages, cells have the highest developmental potential to differentiate into both embryonic and extraembryonic tissues (also referred to as Totipotency).

The cellular potency of cells is important in determining its ability to differentiate into various cell types. Cells possess different levels of potency, ranging from totipotency, where cells can give rise to all cell types including the placenta, to pluripotency, where cells can differentiate into all cell types except the placenta, to multipotency, where cells can differentiate into a limited number of cells within a particular lineage, and unipotency, where cells can only differentiate into one type of cell (Figure 2). The level of potency that a cell possesses places a constraint on the cell type it can differentiate into and thus influences the cell fate decision.

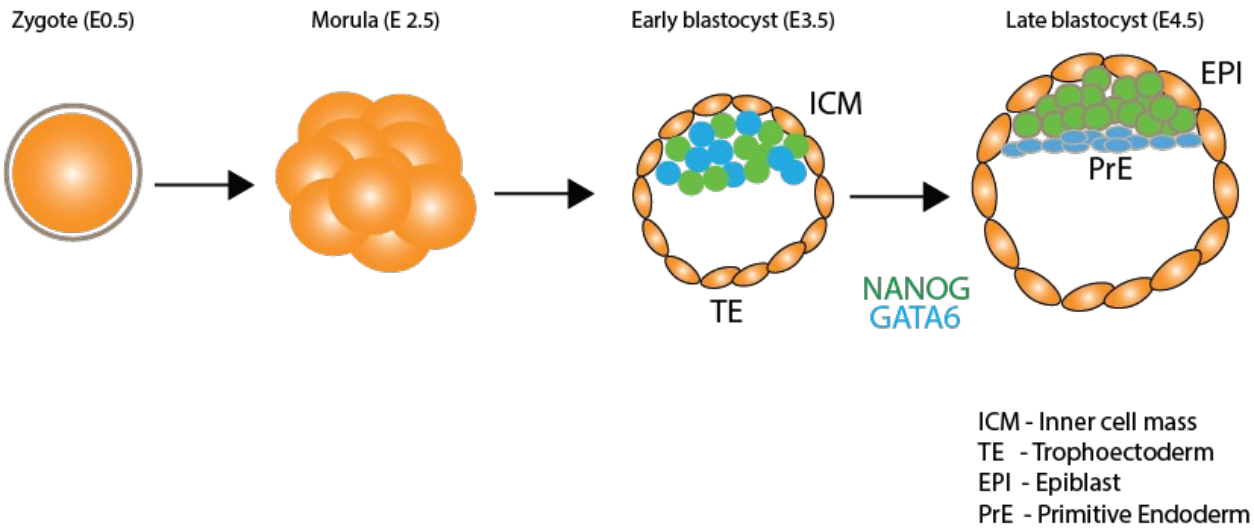

Figure 1. STAGES OF EARLY DEVELOPING EMBRYO

After the fusion of the male sperm cell and the oocyte, the zygote is formed. The first cell lineage decision takes place when the inner cell mass (ICM) differentiates from the trophectoderm.

The morula is an early-stage embryo consisting of a compact mass of 16 cells.[3] Through a process called blastulation, it undergoes transformation, developing into the blastocyst. Blastulation involves a series of cellular rearrangements and differentiation events that eventually lead to the compaction of cells.[3] The compaction process is facilitated by the formation of desmosomes and gap junctions between cells, which allows them to come together and form distinct layers of cells. The outer layers of cells, known as the trophoblast, are responsible for implantation and the formation of the placenta, while the inner cell mass (ICM) gives rise to the embryo proper. This marks the first lineage specification event. Subsequently, the ICM divides into the hypoblast, which contributes to the formation of the extraembryonic tissues and the epiblast, which gives rise to the three germ layers (ectoderm, mesoderm, and endoderm) in a process known as gastrulation.[4]

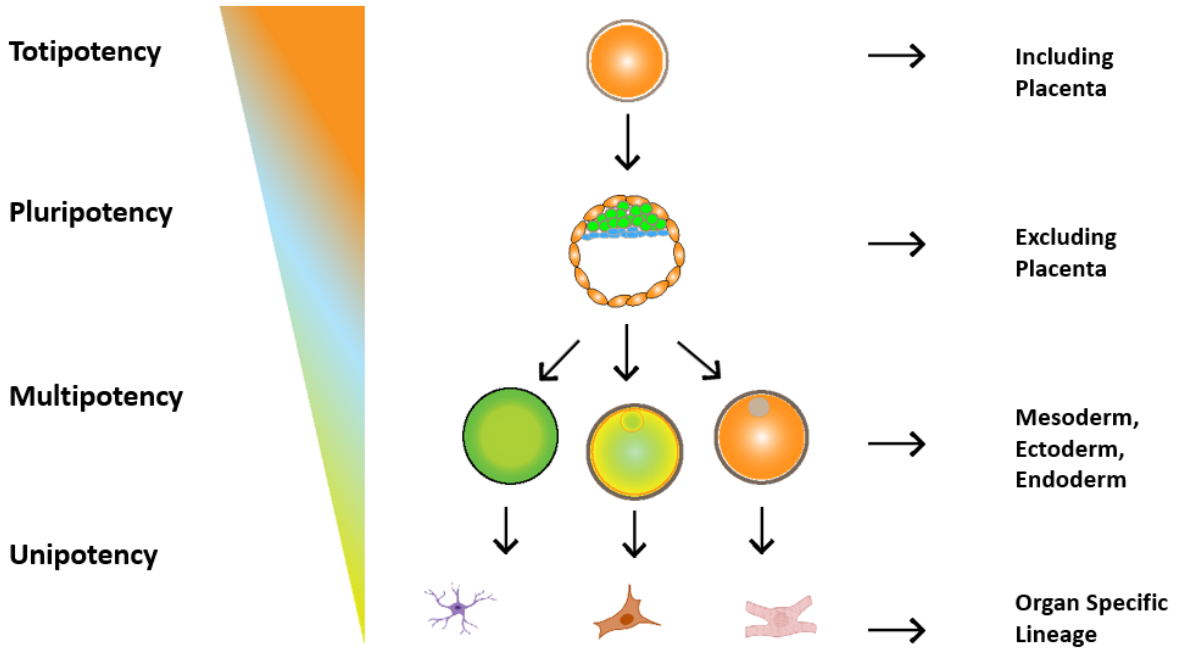

Figure 2. DEVELOPMENTAL POTENTIAL OF MOUSE ESC

As an embryo develops, it moves from a less restrictive state to a more restrictive state, which signals lineage priming, prior to specification.

ESC AS A MODEL SYSTEM TO STUDY DEVELOPMENT

Embryonic stem cells (ESCs) were first derived by explanting blastocysts or ICMs on a layer of “feeder” cells in a medium containing fetal calf serum.[5] They are defined by two properties: first, their ability to self-renew allows for long-term culture while maintaining their differentiating ability; second, they are pluripotent, meaning they can differentiate into all types of germ cell layers of the body. ESCs serve as a valuable model for studying the signaling environment of pluripotency and the molecular mechanism governing early development. Researchers have exploited the potential of these cells in multiple areas, including modeling disease, drug discovery, regenerative medicine, and cellular reprogramming.[6–8]

PLURIPOTENCY NETWORK

Both *in vitro* and *in vivo* studies have identified that pluripotency is maintained by a combination of specific extracellular signals and hierarchical gene regulatory networks.[9] The statement that transcription factors rule pluripotency has been established multiple times[9–11], but the most important pluripotency transcription factors (core) have been identified, which includes OCT4, SOX2, and NANOG. These core pluripotent transcription factors are known to participate in cooperative interactions to regulate similar target genes.[12] In addition to the primary pluripotency network, there is a secondary network (Figure 3) that includes *Esrrb*, *Klf2, cMyc, Fgf4, Tbx3, Dax1, Stat3,* etc.[13]

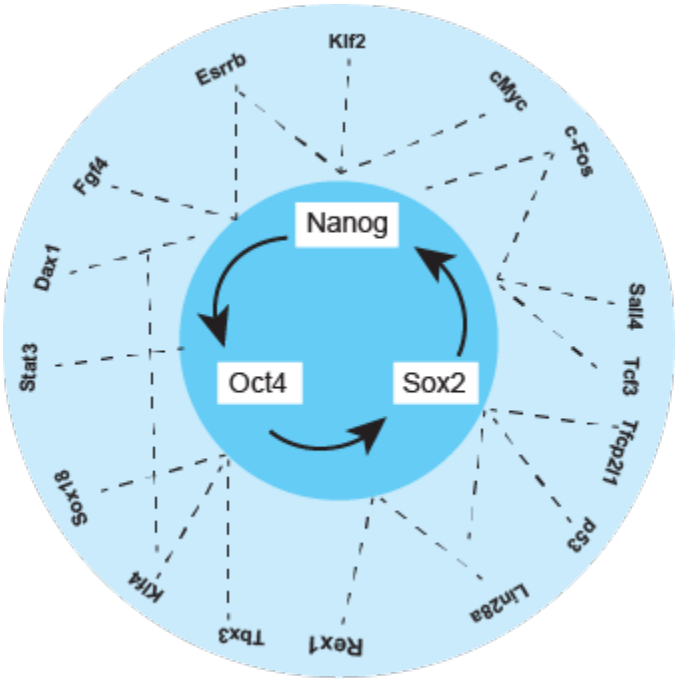

Figure 3. PLURIPOTENCY GENE NETWORK

Multiple genes regulate the activation and maintenance of pluripotency in vivo and in vitro, at the core, *Nanog*, *Sox2,* and *Oct4* form the primary pluripotency network, which is further maintained by the secondary gene network, including *Esrrb*, *Klf*, *Tbx3,* etc. (modified from Parfitt and Shen)[13]

CORE PLURIPOTENT TRANSCRIPTION FACTORS

The core of the pluripotent network consists of three transcription factors, namely octamer-binding OCT4, the SRY family transcription factor SOX2 and the homeobox transcription factor NANOG.[14] In the mouse blastocyst, OCT4 mRNA and protein are present in the inner cell mass (ICM) but not in the trophectoderm.[15] OCT4 plays a crucial role in establishing and maintaining pluripotency, as its deficiency leads to the differentiation of ES cells into the trophectoderm, resulting in cell death.[16] Additionally, *Pou5f1*-null embryos fail to form a pluripotent ICM, further supporting the critical role of OCT4 in pluripotency maintenance.[17] The naïve and primed pluripotent stem cell states are two distinct states that have been identified in vitro to have distinct development potential, expression of genes, chromatin landscapes. Although *Oct4* is expressed both in the naive and primed pluripotent stem cells, its mechanism of expression is controlled by distal enhancers and proximal enhancers, respectively.[18] When generating induced pluripotent stem cells (iPSCS), OCT4 is the only reprogramming factor that is constant in most differentiating protocols,[19] underscoring its significance in the differentiation process.

SOX2 belongs to a group of proteins called HMG (High-mobility group) and is important for self-renewal and pluripotency maintenance.[23] During embryogenesis, SOX2 is expressed within the ICM and extraembryonic tissues[20], and after gastrulation, it is predominantly expressed in the central nervous system, explaining its role in governing ESCs into the neuroectoderm. Mouse ESCs deficient in SOX2 will differentiate primarily into the trophectoderm.[21] SOX2 and OCT4 form a heterodimer and cooperatively regulate downstream targets (e.g., expression of genes involved in

early development *Nanog, Fgf4, Utf1*)[15,22,23], they also regulate each other's expression. It has been shown that SOX2 mutations affect the formation of the OCT4-SOX2 heterodimer conformation on DNA, which is crucial for establishing the pluripotent transcriptional network.[24]

NANOG is a DNA-binding homeobox transcription factor, another core member of the pluripotent transcription factor.[25] The name was inspired by the story of the Irish Legend of Oisín and Niamh and the Land of Eternal Youth, beauty, health, and joy - Tír Na nÓg. Until its discovery in 2003, the pluripotency of mouse embryonic stem cells was majorly attributed to four transcription factors OCT4, SOX2, FOXD3, and STAT3.[25] The importance of Nanog as a core member was established when it bypassed the requirement for leukemia inhibitory factor (LIF) in mESC self-renewal by activating LIF-responsive genes, in particular *Esrrb*.[25, 26] *Nanog* is a homeobox gene that, like other homeobox genes, plays a role in body segmentation and patterning. [25] It displays a critical role in acquiring pluripotency, but it becomes dispensable once pluripotency is achieved.[28] ESCs deficient in *Nanog* lose pluripotency and differentiate into the extraembryonic endoderm lineage.[29] The expression pattern of *Nanog* undergoes significant changes during the transition of cells from pre- to post-implantation stages. During this process, Nanog plays a crucial role in maintaining pluripotency of the inner cell mass (ICM) and is also involved in the transition from embryonic stage E3.5 to E4.5 (Figure 4). Specifically, *Nanog* is expressed in a specific subset of cells in the ICM that give rise to the pluripotent epiblast, while the remaining cells deficient in Nanog will differentiate into the primitive endoderm driven by endoderm-specific genes such as *Gata*4/6.[30]

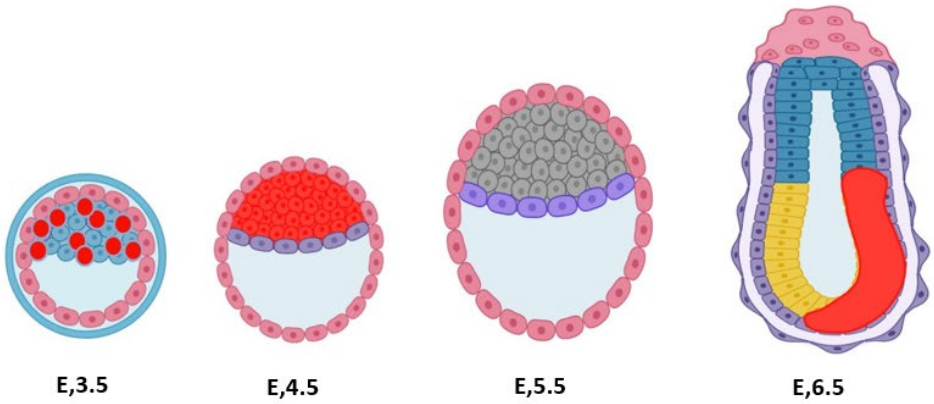

Figure 4. NANOG'S CHANGING EXPRESSION PATTERN

Nanog displays a dynamic expression pattern from the pre-implantation blastocyst stage to the post-implantation epiblast. Cells colored red is indicative of NANOG expression.

CURRENT MODEL OF PLURIPOTENCY

Given the current evidence of these pluripotent factors, to maintain pluripotency and promote differentiation, it is proposed that the core ESC transcription factors act to maintain pluripotency by either activating the expression of other pluripotency-associated factors while repressing lineage-specific genes[31] or by activating their own gene expression and that of each other (Figure 5).[14]

In an ideal scenario, if pluripotent factors were solely involved in repressing lineage-specific genes, overexpressing these core transcription factors would be expected to enhance pluripotency maintenance. However, contrary to this expectation, overexpression often leads to the induction of differentiation. Overexpression of *Oct4* promotes differentiation into the primitive endoderm and mesoderm lineages.[16] *Sox2* overexpression triggers the expression of cell markers for a variety of differentiated cells, including neuroectoderm and mesoderm, but not endoderm.[32] Overexpression of *Esrrb*, *Tbx3* induces endodermal differentiation.[33,34] In addition, Nanog overexpression directs definitive endoderm differentiation in hESCs.[35]

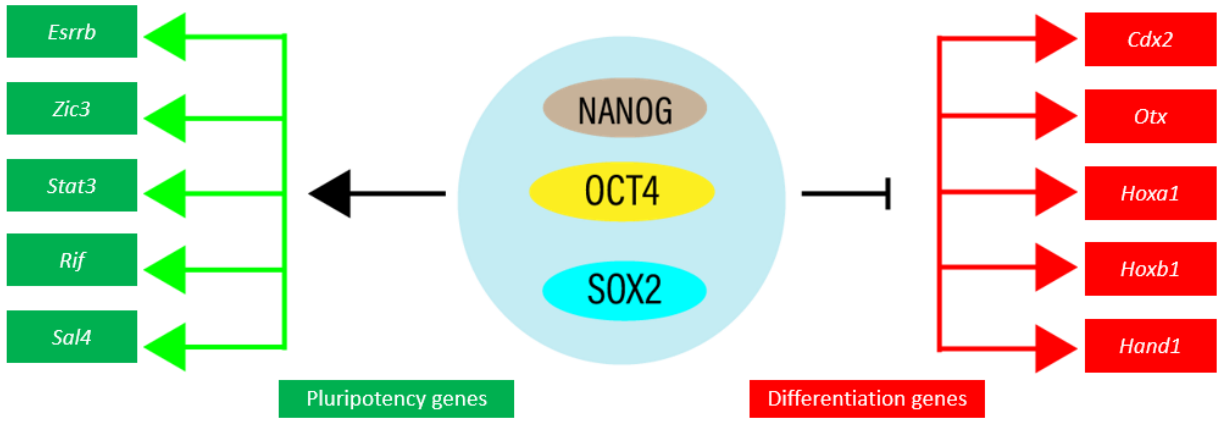

Figure 5. CURRENT MODEL OF PLURIPOTENCY

Pluripotent factors are known to act via inhibition of differentiation genes and activation of pluripotency genes.

PLURIPOTENCY FACTORS AS LINEAGE SPECIFIERS

The observations established in the previous section strongly support the notion that pluripotency factors not only contribute to maintaining pluripotency but also possess lineage-specifying functions. It is becoming increasingly clear that these factors play a crucial role in the process of lineage specification. In other words, pluripotency factors act as lineage specifiers or have functionally significant roles in determining specific cell lineages during the process of differentiation.

Studies investigating core cell state transition revealed that the core pluripotency factors OCT4 and SOX2 play opposing roles in promoting mesodermal and neuroectodermal specification, respectively.[36] OCT4 has been shown to promote mesodermal specification, while suppressing genes associated with neuroectodermal specification.[36] Conversely, SOX2 has been shown to promote neuroectodermal specification while suppressing genes associated with mesodermal specification, further highlighting their critical role in directing early cell fate decisions.[36] In hESC, these core pluripotent factors also control the expression of Eomesodermin (*EOMES*), one of the lineage markers for endodermal specification. *EOMES* can, in turn, interact with SMAD2/3 to control endoderm formation.[35]

Despite these links with cell fate decisions, it might be hard to reconcile what specific role a pluripotent factor plays. For example, *Nanog* has been classified as a mesendodermal-class gene, like *Oct4* and *Tbx3*.[60] It is also an epiblast marker and a primitive endoderm marker.[36] *Nanog* shows multiple facets in pluripotency maintenance and lineage specification in early development.

PROBLEM STATEMENT

We have examined the potential role of NANOG in differentiation, given its distinctive expression pattern as previously described. Our overall goal was to gain a deeper understanding of the specific role of Nanog during differentiation.

To achieve this goal, we established two specific objectives. Firstly, we aimed to investigate the function of NANOG during the transition from naive to primed pluripotency. Secondly, we aimed to explore the role of NANOG in neuroectodermal differentiation.

CHAPTER 2

NAIVE TO PRIMED PLURIPOTENCY

Distinct in vitro states have been identified for mESCs, starting from the naive or ground state, progressing through the epiblast-like cells (EpiLCs), and finally, the epiblast stem cells (EpiSCs).[37] Both the EpiLCs and the EpiSCs are representatives of the primed state and show distinct morphologies, cytokine dependence, gene expression, and epigenetic profiles compared to the ESCs (naive state).[38]

In the naive state, mESCs derived from the pre-implantation blastocyst display an unbiased developmental potential, representing a stage prior to lineage decision. They express general naive pluripotency factors, including *Oct4* and *Sox2.*[38] However, during the transition to the primed state after implantation, the naive marker genes are suppressed, and lineage specification genes such as *Fgf5 (a post-implantation marker)* and other lineage-specific factors like *T and Foxa2* are upregulated. [38] This primed state displays lineage priming, characterized by early transcription of genes associated with lineage specification and a repressive chromatin structure.

mESC can be maintained *in vitro* in the naive state using a defined media, which consists of two kinase inhibitors known as “2i” to block mitogen-activated kinase (MEK)/ERK and glycogen synthase kinase-3(GSK3) signaling pathways and leukemia inhibitory factor (LIF).[39] Removal of 2i leads to a dismantling of the naive pluripotency network leading to a transition into the primed state (EpiSCs). EpiSCs can also be generated and maintained by culturing naive pluripotent mESC in Fibroblast growth factor (FGF) or activin A (Figure 6).[40] Extensive research has been focused on the role

of pluripotent factors in maintaining the naive state.[14] However, the transition from the naive state to the primed state, specifically during the Naive to EpiLCs stage, has received increasing attention. Regulators such as OCT4, OTX2, and ZIC3 have been identified as critical players in this transition, which involves significant changes in the chromatin landscape and gene expression patterns.[41, 42] The functional significance of NANOG, a core pluripotent factor, in this specific transition remains largely unexplored.

To contribute to our understanding of the regulatory networks underlying the transition from ESCs to EpiLCs, the current study focused on investigating the cistrome (binding sites on DNA) of NANOG during the early stages of this transition in mouse ESCs. We specifically analyzed regions that showed differential Nanog binding and differential chromatin accessibility. Additionally, we integrated gene expression data analysis to gain further insights into the associated molecular events. Through this investigation, we aimed to shed light on the specific role of NANOG in the context of the Naive to EpiLCs transition and provide a better understanding of the molecular mechanisms involved.

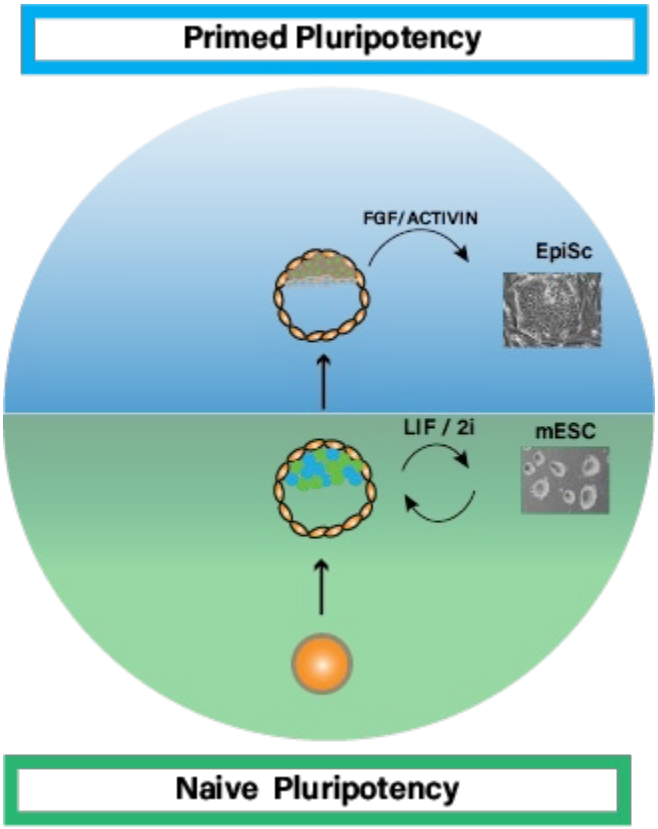

Figure 6. NAIVE TO PRIMED PLURIPOTENCY MODEL

Embryonic stem cells can transition from naive state to primed state via defined culture conditions. Naive state is maintained via 2i inhibitor or LIF while primed condition via activin/FGF.

HYPOTHESIS STATEMENT

Based on the evidence of pluripotent factors also acting as lineage specifiers, we hypothesize that NANOG has a preferential role for mesodermal specification during the transition from naive to primed pluripotency. Furthermore, we propose that NANOG has a dual role during this transition, acting as a pluripotent factor during the naive state and a lineage specifier in the primed state. Our aim is to provide further evidence supporting the link between pluripotency maintenance and lineage specification and to establish NANOG as a key regulator of this process.

STUDY GOALS

The primary objectives of this study are to identify the specific targets of NANOG during the transition from naive to primed pluripotency and to integrate ChIP-sequencing and gene expression data to interrogate the NANOG target gene network during this transition. The results of this study could enhance our comprehension of the molecular mechanism of NANOG underlying pluripotency maintenance & lineage decision and how its targets change during the transition from naive to primed pluripotent state. In addition, the results could lead to the development of new approaches for directing cell fate decisions in regenerative medicine.

METHODS

Dataset and Study Design

RNA Sequencing

The study design for all the datasets used is described in Figure 7. Raw sequencing reads were downloaded from the NCBI Gene Expression Omnibus (GEO) database (accession ID: GSE138818).[43] Mouse embryonic stem cells were maintained in 2i+LIF conditions at time 0h, serum was added to the medium at 12hr, and 2i+LIF was subsequently removed, allowing the cells to differentiate. Three time points were obtained in triplicate: 0h (naive), 12h (transition), and 24h (primed).

ChIP Sequencing

Raw sequencing reads were downloaded from the NCBI Gene Expression Omnibus database (GEO) (accession ID: GSE71933).[44] Embryonic stem cells that were initially in the naive pluripotent state were cultured in a medium containing basic fibroblast growth factor (bFGF) and activin A. As a result of this culture, they developed into cells with an epiblast-like phenotype known as EpiLCs, which are considered to be in a primed state at day two of differentiation.

ATAC Sequencing

Raw sequencing reads were downloaded from the EBI European Bioinformatics Institute ArrayExpress database (ID: E-MTAB-7207). The accessible chromatin regions were profiled at ESC as they transitioned to EpiLCs over two days using transposase-

accessible chromatin sequencing (ATAC-seq).[41] Day 2 of differentiation is defined as the primed state.

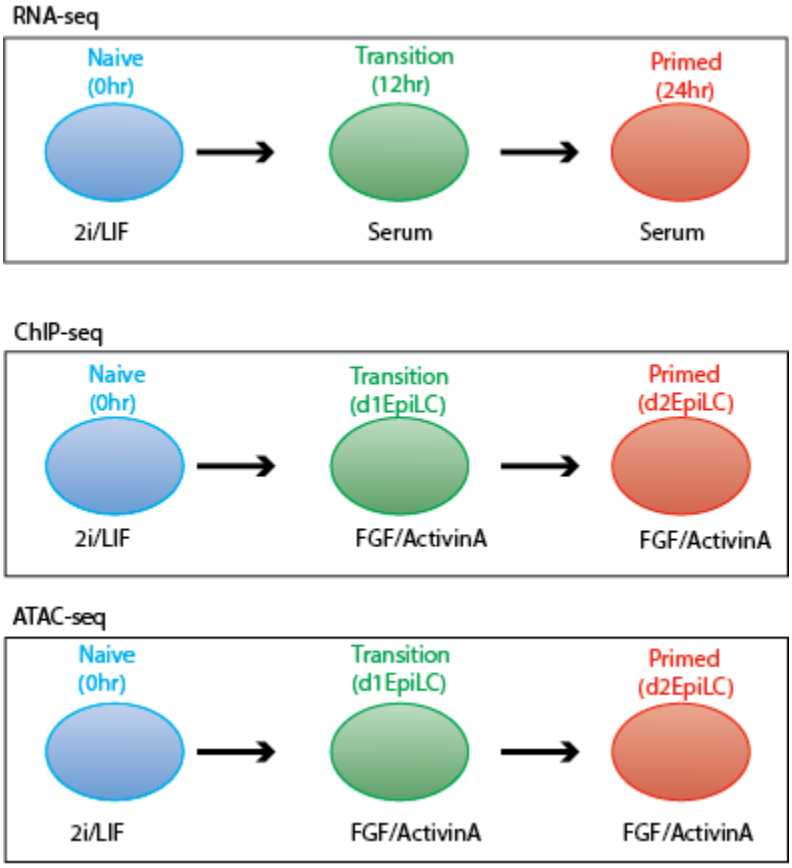

Figure 7. METHODOLOGY NAIVE TO PRIMED TRANSITION

Diagram showing the naive to primed transition for multi-omics analysis.

BIOINFORMATICS ANALYSIS

Preprocessing

The software utilized default settings for all experiments except when otherwise mentioned. Prior to the analysis, the raw sequencing reads underwent trimming and filtering procedures to eliminate adapter sequences using Trimmomatic[61]. The resulting filtered reads were then aligned to the National Center for Biotechnology Information build mm10 using Bowtie2 (ChIP-Seq reads)[45] and HISAT2[46] (RNA-seq reads). Only uniquely mapped reads were retained for further analysis. Finally, the generation of heatmaps was accomplished using ComplexHeatmap[62] and deepTools.[47]

RNA-Seq Data Analysis.

Initially, the data from each replicate were compared to assess their similarity through a Principal Component Analysis (PCA). After confirming data quality, raw read counts were extracted using featureCounts.[63] The mean of the raw counts was calculated, and the resulting matrix was scaled using z-score adjustment. To identify differentially expressed genes, DESeq2[48] was employed, and genes with an adjusted p-value below 0.05 were designated as significantly different. To conduct an enrichment analysis, clusterProfiler[49] was utilized.

ChIP and ATAC-Seq Data Analysis

The identification of peaks was carried out on the individual replicates using MACS2.[50] To ensure the selection of highly reliable peak sets, the IDR (Irreproducible Discovery Rate)[51] method was employed, considering only peaks with a p-value less

than 0.05, indicating reproducibility across replicates. All subsequent analyses were conducted based on these selected peak sets.

Next, using the mergePeak function from HOMER[52] (d = 100), we identified peaks that exhibited differential NANOG binding in at least two conditions. These differentially bound peaks were then merged with the accessible peaks, resulting in a union set of 6,156.

RESULTS

Determination of *Nanog* cistrome

First, we tested the hypothesis that inhibition of *Nanog* expression in the pluripotent naive cells will lead to a decrease in the expression of lineage-specific markers, indicating that *Nanog* plays a role in early cell fate decisions. To investigate this, we conducted an analysis of the mESC RNA-sequencing dataset, comparing gene expression in *Nanog* knockout cells to wild-type cells as control. To assess the overall similarities between the samples, the Euclidean distance was calculated (Figure 8), after which an apparent clustering of the replicates was observed. Furthermore, principal component analysis (PCA) revealed a distinct separation of the samples based on their genotype PC1 (62%) variance and the timing of differentiation PC2 (29%) variance (Figure 9).

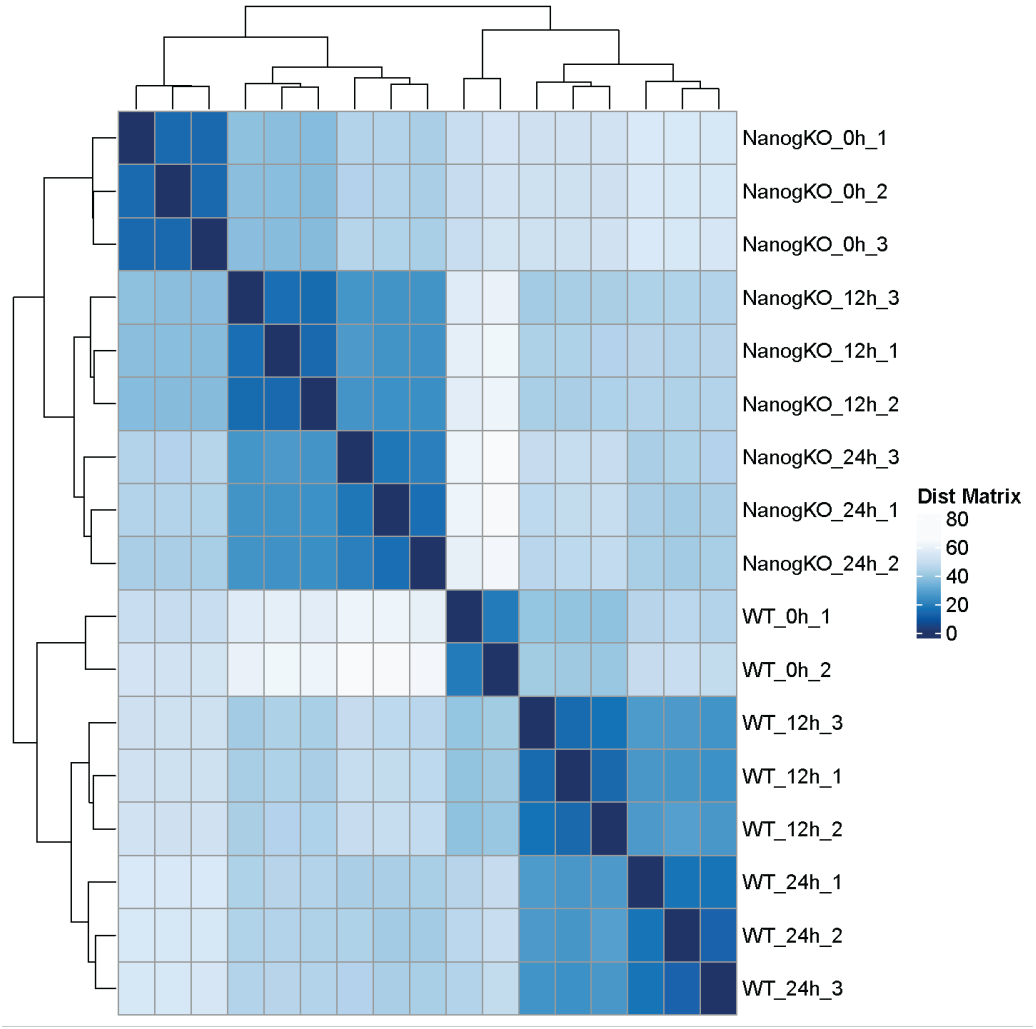

Figure 8. DISTANCE MATRIX

The distance matrix is used to provide a numerical representation of the similarities between the knockout and wildtype samples. A positive correlation is observed across the replicates.

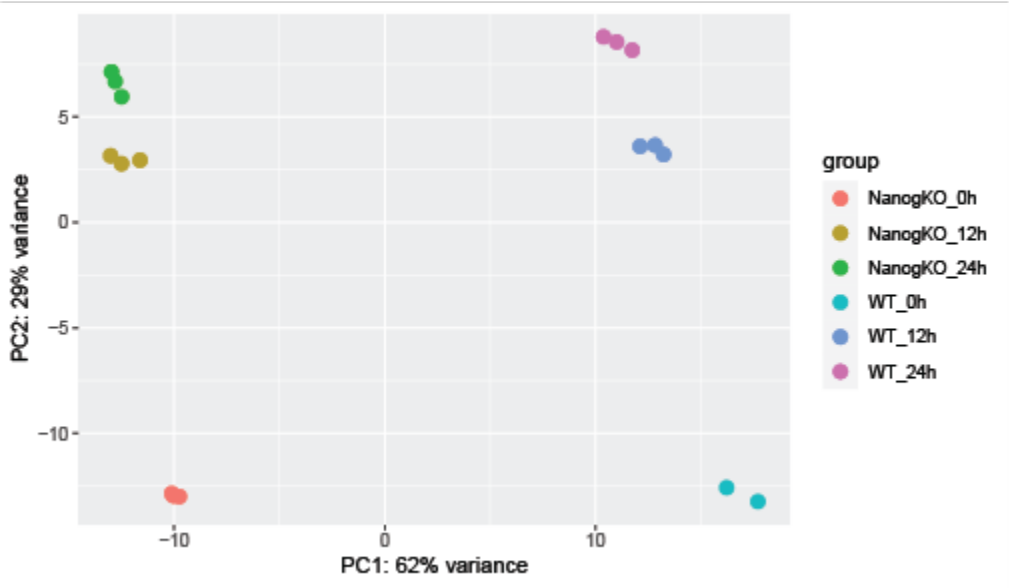

Figure 9. PRINCIPAL COMPONENT ANALYSIS PLOT

PCA shows a clear separation of the samples across the genotype and time of differentiation.

Next, we focused on selecting representative genes from various lineages, including mesodermal, neuroectodermal, and endodermal lineage, that exhibited changes in their expression dynamics from the naive to primed condition. Specifically, we identified genes that were repressed by *Nanog* as representatives of the neuroectodermal and endodermal lineages, as these genes showed increased expression in Nanog mutant cells. Conversely, the knockdown of *Nanog* led to the decreased expression of mesodermal genes suggesting that *Nanog* may be required for their activation during the transition from naive to primed pluripotency (Figure 10).

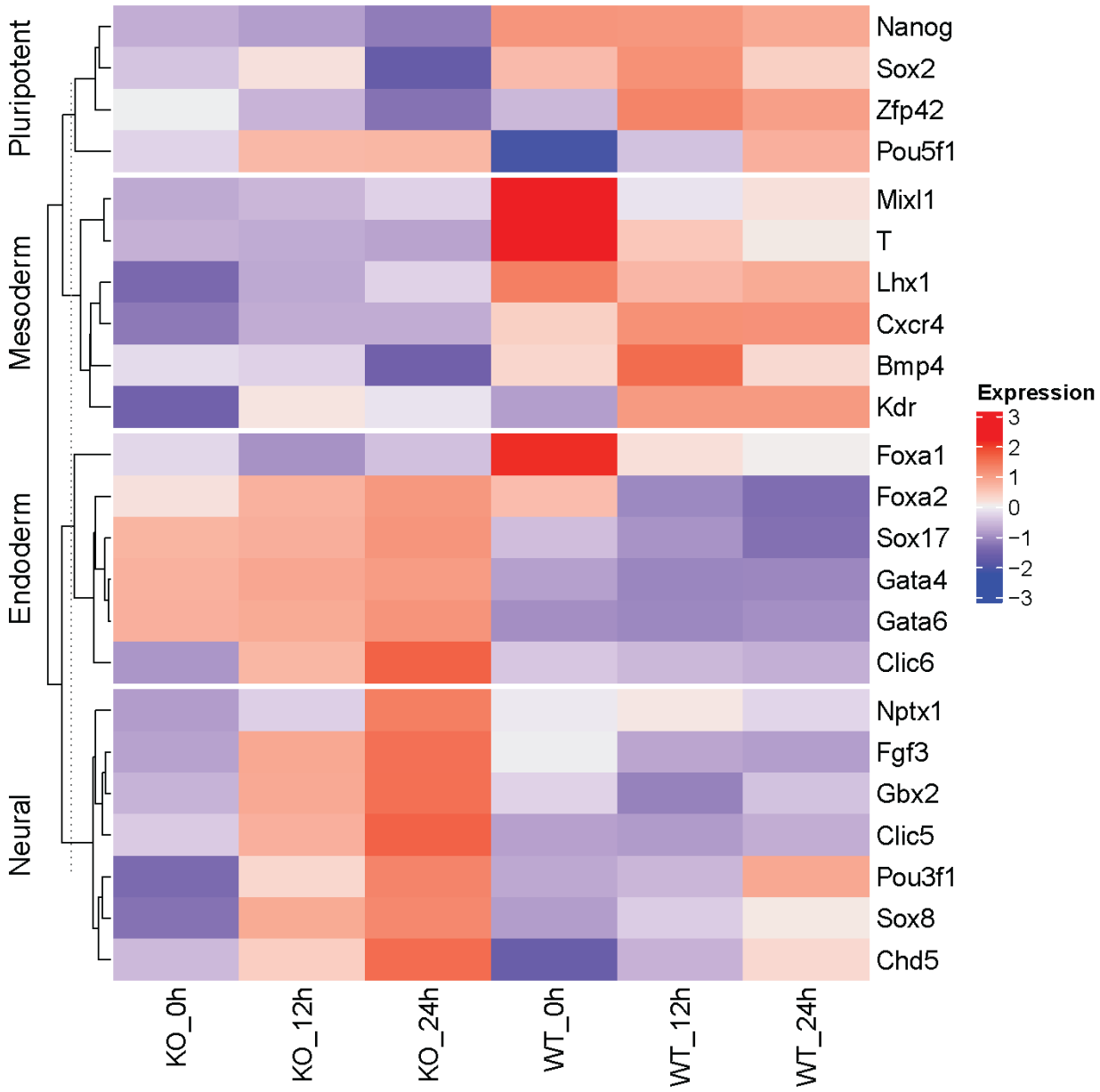

Figure 10. REPRESENTATIVE GENES

Heat Map showing the dynamic expression of lineage marker genes during naive to primed pluripotent state. Representative genes associated with mesodermal, endodermal, and neuroectodermal specifications were selected based on previous literature.[53, 54]

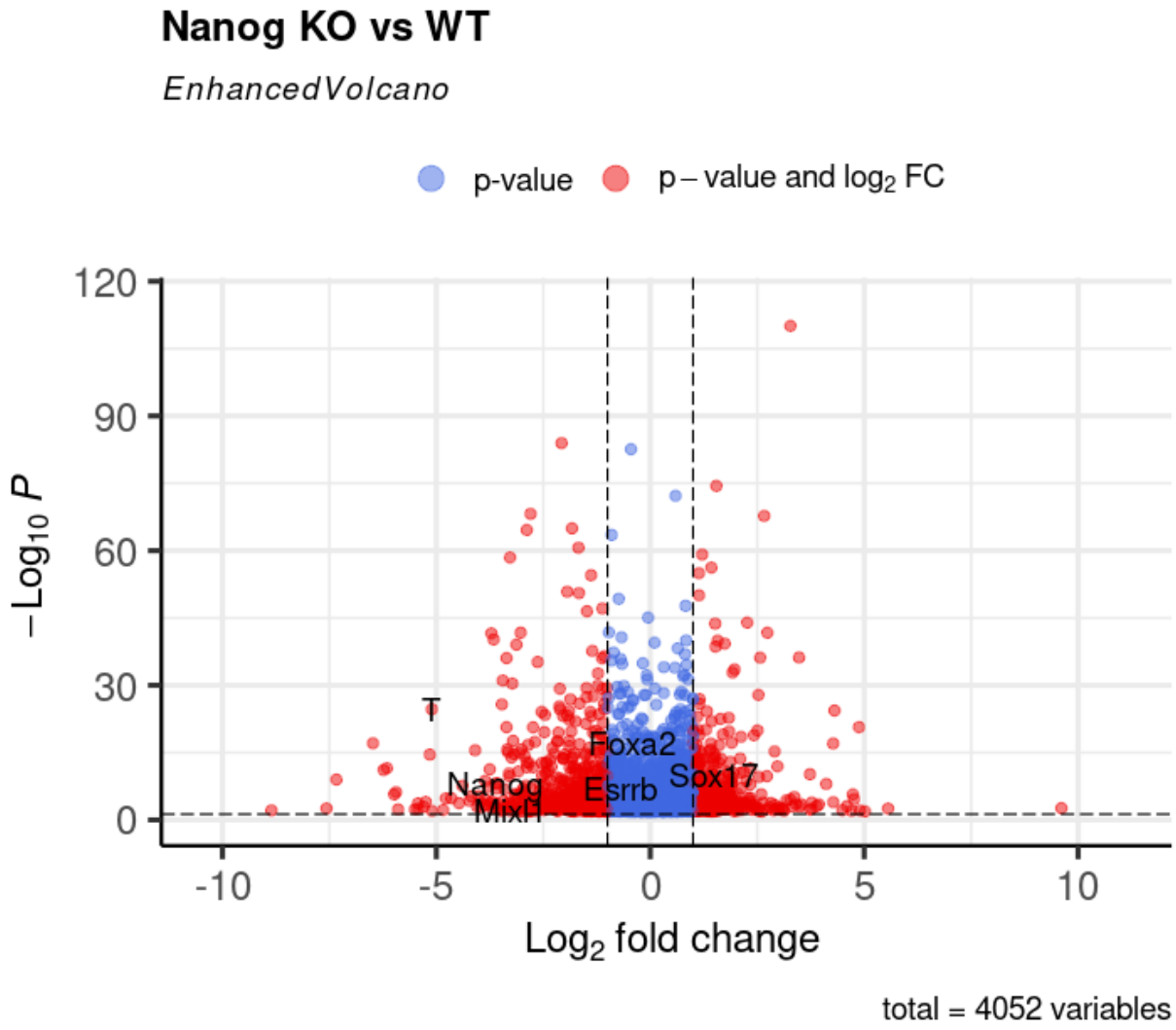

Figure 11. NANOG DIFFERENTIALLY EXPRESSED GENES

Volcano plot showing the differentially expressed Nanog genes during the naive to primed transition (Adj p-value < 0.05).

By employing the likelihood ratio test, we conducted an analysis to identify differentially expressed genes across time. Our analysis revealed a total of 4,052 genes that showed significant differential expression (Adj p-value < 0.05). These genes included the known developmental genes *Nanog*, *T*, *Sox*, *Esrrb,* etc.

To further explore the role of NANOG, we asked whether these differentially expressed genes are direct targets of NANOG. Considering that the depletion of *Nanog* resulted in a large number of changes, including both direct and indirect consequences, we integrated published chromatin immunoprecipitation sequencing (ChIP-seq) data to identify direct target genes of NANOG.

Although the time points of the gene expression data and the ChIP-seq data do not perfectly align, we reasoned that a temporal alignment could be performed. We took into account that the naive ESCs correspond to the 0-hour time point, and the day 2 EpiLCs correspond to the 24-hour time point. By aligning these time points, we aimed to identify direct target genes of NANOG relevant to the transition from the naive to the primed state.

To investigate the regulatory potential of NANOG, we analyzed the publicly available genome-wide binding data. Our initial focus was on the bound sites in EpiLCs at day 1, representing the transition state. We identified a total of 44,810 reproducible bound sites, (Figure 12) demonstrating NANOG's occupancy (IDR p-value < 0.05). Interestingly, the majority of these sites were found in intergenic regions, consistent with Nanog roles at enhancer regions. The NANOG-bound regions were further associated with 11,861 target genes using the nearest-neighbor model.

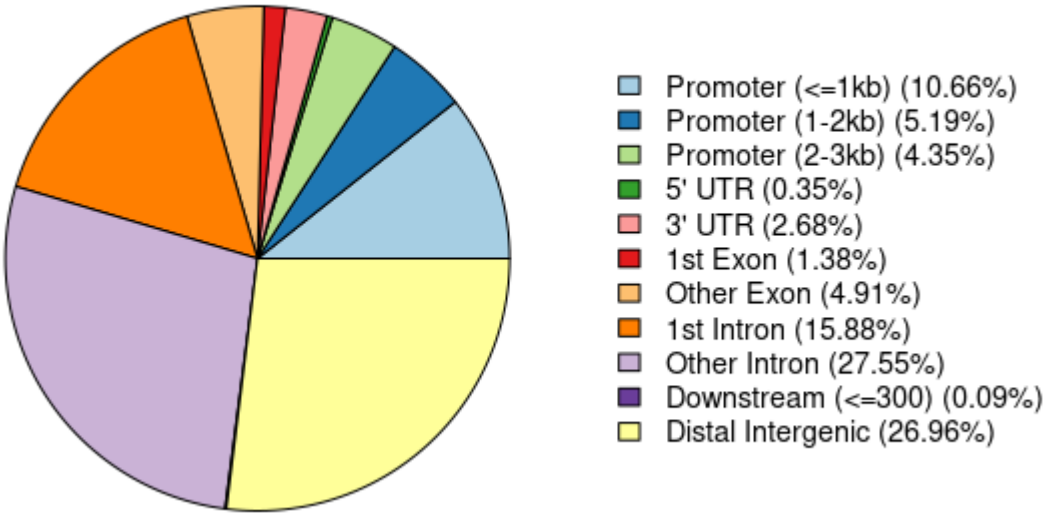

Figure 12. NANOG CISTROME

Pie chart showing the peak annotation of NANOG-bound regions at day 1 (transition). Promoter is defined as -3 to +3 kb.

To further understand the dynamics of NANOG binding and its association with changing chromatin accessibility profiles during the transition from naive to primed pluripotency, we integrated publicly available ATAC-seq data. Our analysis focused on identifying sites with differential binding patterns of NANOG and chromatin accessibility between the two conditions (Figure 13). By using the individual peaks from this analysis, we visualized the NANOG bound sites at the three time points Naive esc, Day 1, and Day 2. These regions were further divided into four distinct clusters to represent genomic loci where NANOG binding, and chromatin accessibility may have a significant impact on the regulation of gene expression and cell fate decisions during the transition process.

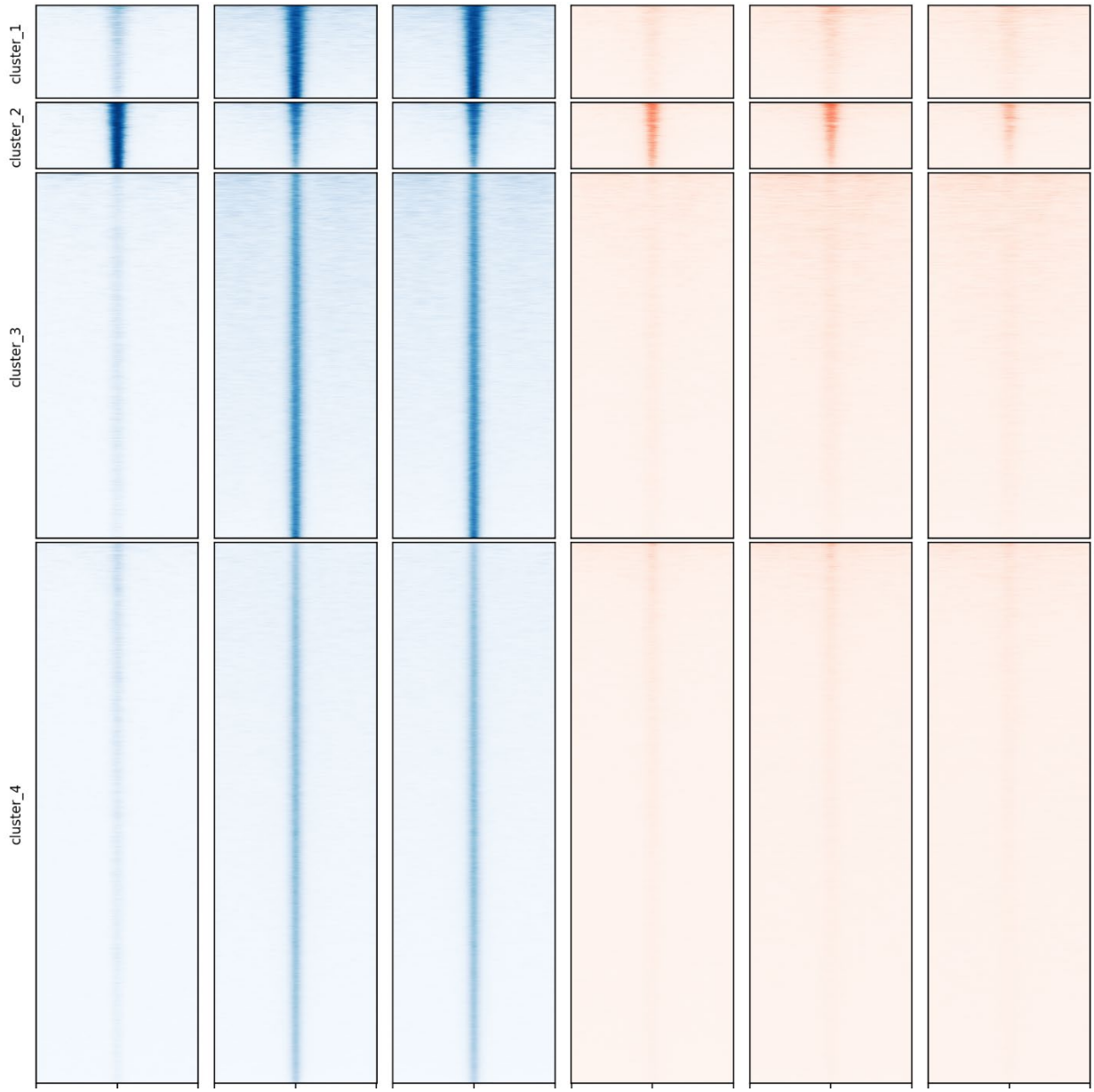

Figure 13. DIFFERENTIAL NANOG BINDING AND CHROMATIN ACCESSIBILITY HEATMAP

Heat map showing differential Nanog binding and accessible chromatin. NANOG binding is summarized in the blue heat map (left), while Chromatin accessibility is visualized in the red one (right).

Utilizing HOMER to identify the enriched motifs within the peaks of interest, cluster 2 and cluster 4 were particularly interesting as they included motifs for several transcription factors known to be involved in the transition from ESC to EpiLCs, including Pou5f1/Oct4, Nanog and Sox2 (Figures 14, 15, and 16). The significant enrichment presence of these motifs suggests that these transcription factors may interact with Nanog and potentially collaborate in regulating cell fate decisions during the transition. Interestingly, at the transition day 1 (Figure 16), there was enrichment of LIM homeobox 1 (LHX1) motif, which is known to play a role in lineage differentiation, and during gastrulation, it is known to contribute to axial mesendoderm.[55] In addition, 35.4% and 55.1% of the peaks were contained for the target sequence of NANOG and LHX1, respectively.

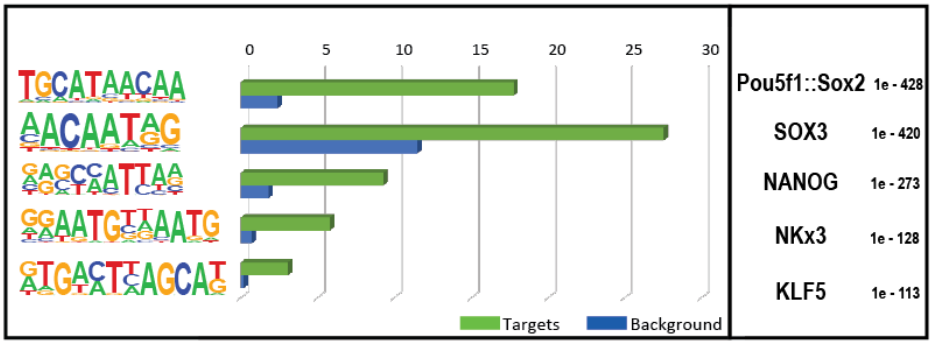

Figure 14. NAIVE ESC MOTIFS

Bar graph showing the top five enriched motifs from naive esc time point, including Pouf1::Sox2, SOX3, NANOG, Nkx3 and KLF5.

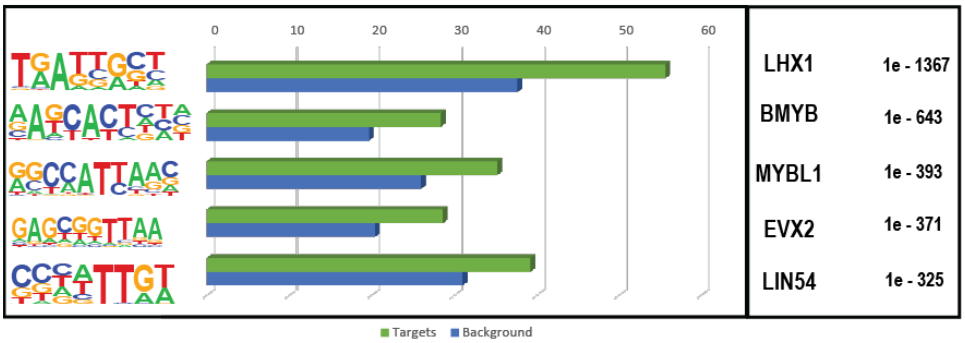

Figure 15. DAY 1 MOTIFS

Bar graph showing the top five enriched motifs from day 1 of bound NANOG and accessible peaks.

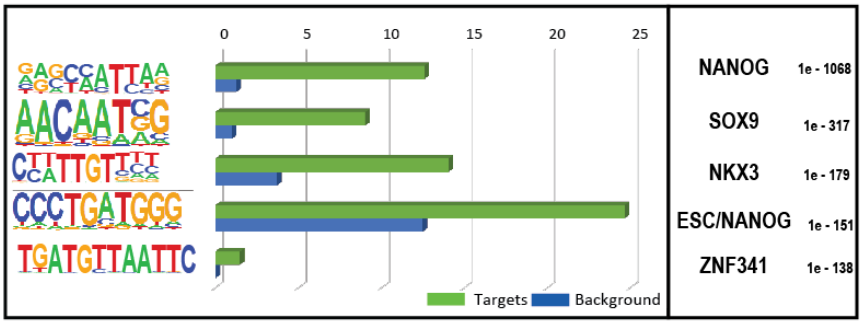

Figure 16. DAY 2 MOTIFS

Bar graph showing the top five most significantly enriched motifs at day 2; Bars indicate the percentages of the input peaks.

NANOG Dependent Gene Regulatory Event.

Having established the NANOG cistrome and its genome-wide changes, we next asked whether NANOG influences gene expression during these transitions. Our results integrating ChIP and RNA sequencing data revealed NANOG exerts both activating and repressive effects on gene expression during this transition. Specifically, we found that a similar proportion of genes were downregulated during the transition (Figure 17), with 49.7% (n = 4,047) of genes showing decreased expression at day 1 and 53% (n = 3,703) at day 2. Conversely, 50.3% of genes showed increased expression on day 1 and 47% on day 2. This suggests that Nanog can act as both an activator and a repressor during the transition from naive to primed pluripotency.

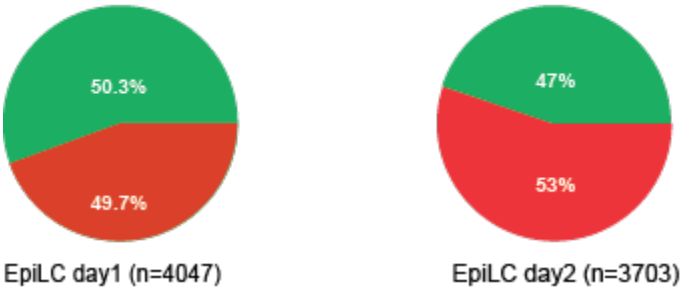

Figure 17. NANOG REGULATED GENES

Pie chart showing the percentages of the differentially regulated Nanog genes (log2foldchange > 0).

In addition, we examined the correlation between the expression of Nanog-regulated genes and genes relevant to early mouse embryonic development.[53] We observed a correlation of expression, with representative genes from ICM being most similar to the naive ESC, post-implantation epiblast similar to the primed ESC, while the pre-implantation epiblast genes resemble the transition state (Figure 18). We next asked whether Nanog is relevant in the regulation of marker genes of early mouse embryonic development. By comparing our Nanog dataset with the developmental stage-specific datasets (Figure 19), we found a similarity and a reflection of gene expression transitions at different developmental stages. This correlation suggests that Nanog plays a role in regulating the expression of genes representative of the pre- and post-implantation epiblast. Among these genes, we identified known regulators of the transition from naive to primed pluripotency, such as *Otx2* and *Pou3f1*.[42] This provides further evidence for the involvement of Nanog in the regulation of genes that are critical for the transition from the naive to primed pluripotent state.

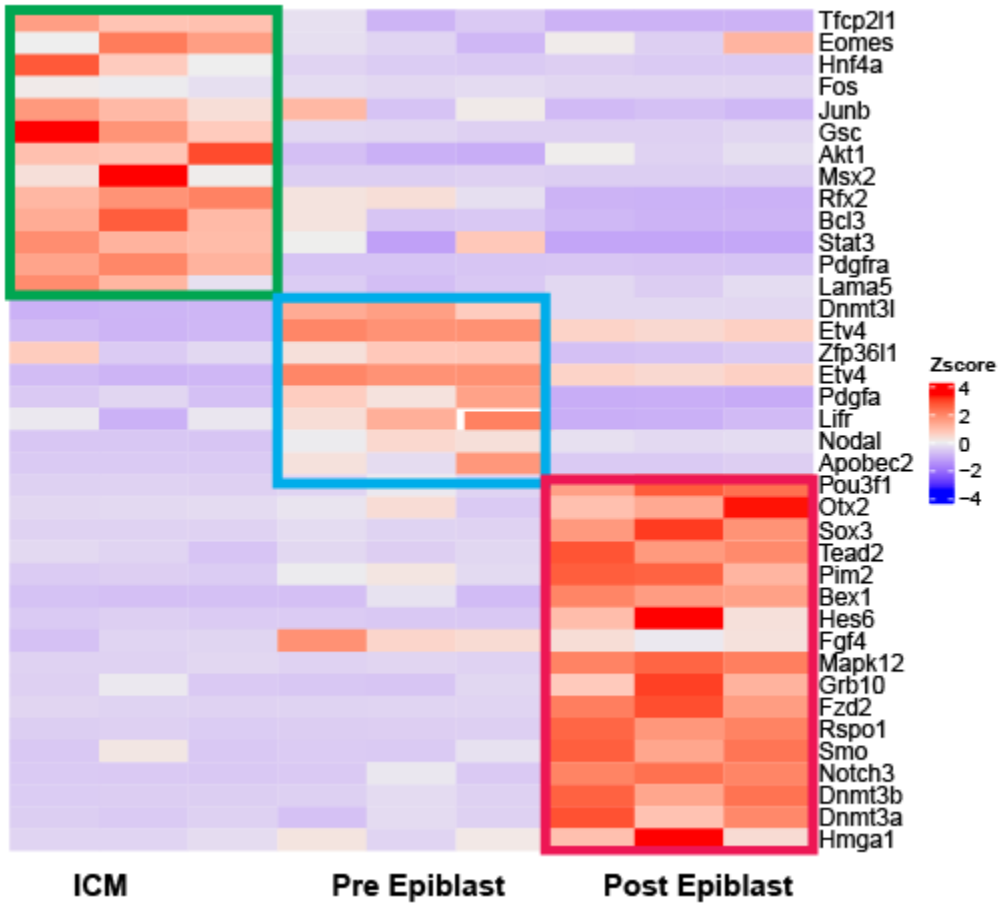

Figure 18. STAGE SPECIFIC EMBRYONIC DEVELOPMENT GENES (DEVELOPING EMBRYO)

Heat map showing the expression of representative genes at developmental stages, ICM, preimplantation, and post-implantation epiblast in a developing embryo.

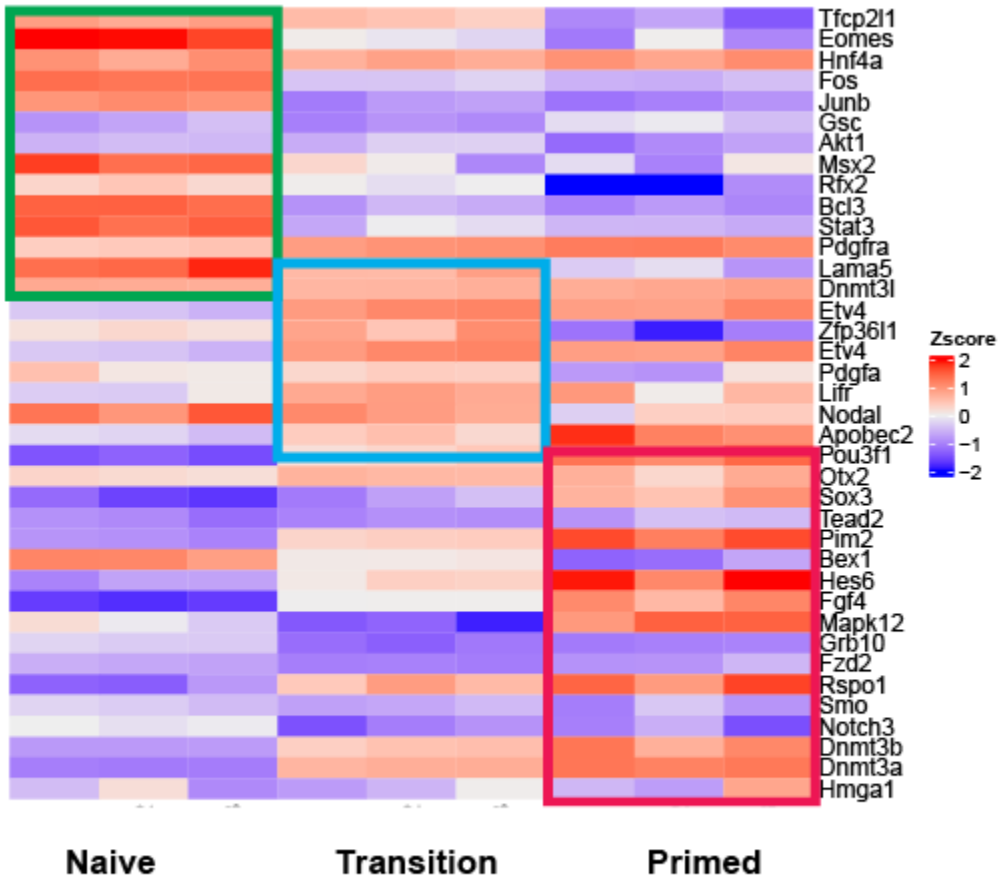

Figure 19. STAGE-SPECIFIC EMBRYONIC DEVELOPMENT GENES (NANOG)

Heat map showing the expression of representative genes at developmental stages, ICM, pre-, and post-implantation epiblast for mouse ESC.[53]

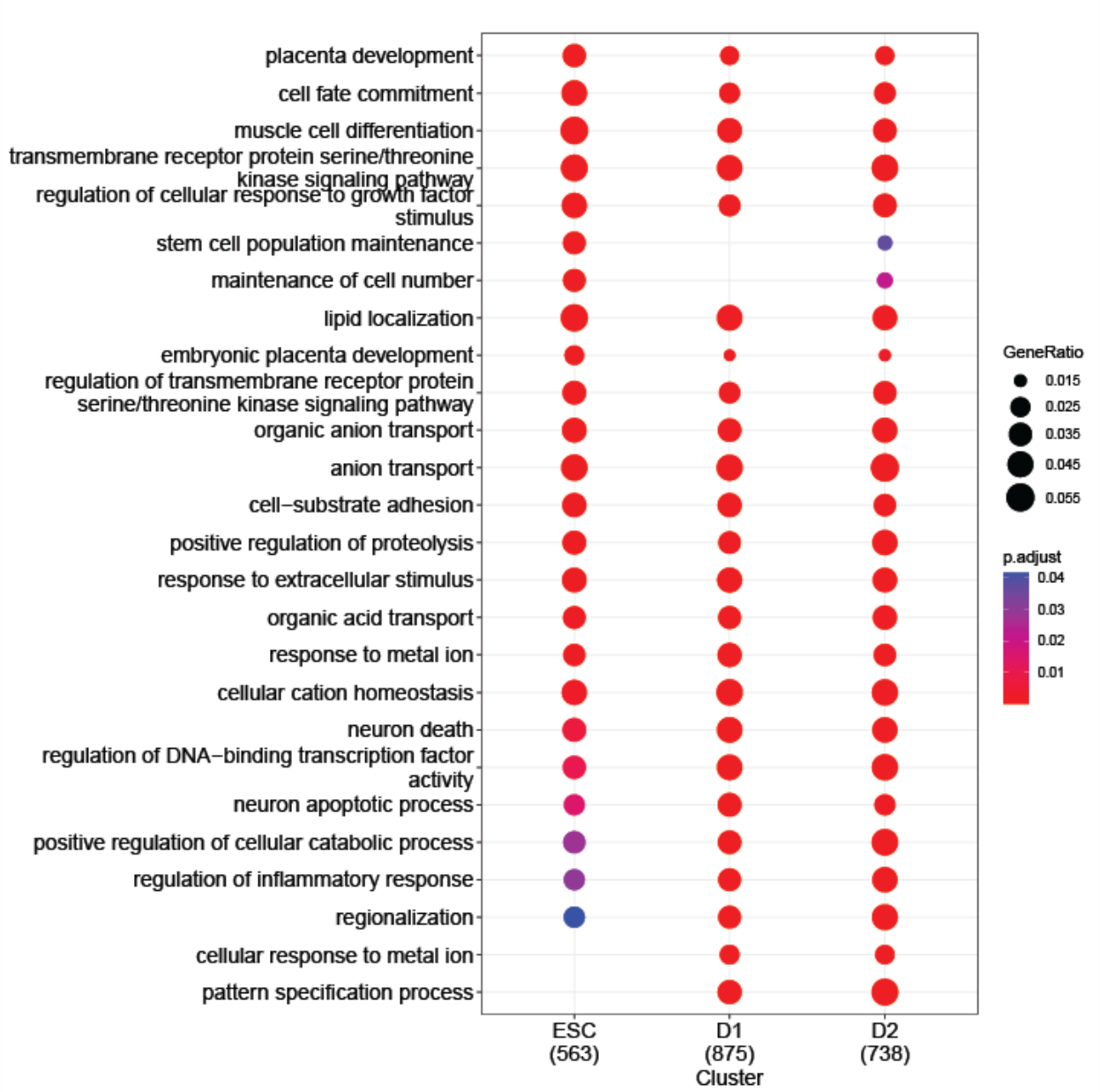

Figure 20. BIOLOGICAL PROCESS FROM NAÏVE TO PRIMED STATE

Dot plot showing both distinct and conserved biological processes from naive to primed state (p-adjusted < 0.05).

Finally, we were interested in the specific target genes that are activated or repressed by NANOG, which would reveal the molecular mechanism underlying pluripotency maintenance and lineage specification. In examining the biological process of the genes regulated by Nanog across the different stages of differentiation, we observed both conserved and distinct functions.

Across all stages, we found several developmental processes conserved, including cell fate commitment, placenta development, muscle cell differentiation, (Figure 20) etc. The conservation of these processes suggests NANOG's involvement in essential cellular functions during both pluripotency maintenance and lineage specification. However, we also observed distinctions in biological processes among the developmental stages. In the naive (ESC) state, we found an upregulation of processes related to stem cell maintenance and the maintenance of cell number. This suggests that during the Naive state, there is an emphasis on self-renewal and the preservation of the pluripotent cell population. However, these processes were downregulated at the day 1 transition stage and less significant at the day 2 primed state, indicating a shift in cellular priorities as differentiation progresses.

Interestingly, we observed an upregulation of the pattern specification process and regionalization at the day 1 transition and day 2 primed state, which was absent in the naive state. This suggests that during the transition from naive to primed pluripotency, there is a heightened focus on the establishment of specific cellular patterns and identities, potentially reflecting the initiation of lineage specification.

DISCUSSION

Several regulators of transition from naive (ESCs) to the primed (EpiLCs) have been previously studied, despite NANOG's importance it is unclear how its role changes during the transition. Through integrated analysis of multiple omics data, we aimed to understand the molecular mechanism underlying NANOG's function.

In our RNA sequencing analysis, we observed that the knockout of *Nanog* resulted in the downregulation of mesodermal representative genes. This finding is consistent with previous evidence classifying *Nanog* as a mesendodermal gene, along with Oct4 and Tbx3.[36] The downregulation of mesodermal genes upon *Nanog* knockout suggests its involvement in the regulation of mesodermal lineage specification.
While our data suggests downregulation of mesendodermal genes, it is possible that NANOG's role may vary depending on the specific cellular context and culture condition. However, it is important to note that our ESCs were cultured in a naive mouse condition, which involves the presence of 2i and LIF. This raises an interesting point of comparison with another study conducted on human ESCs (hESCs).[35] In that study, NANOG was proposed to promote definitive endoderm specification, which is one of the three primary germ layers. This finding contrasts with our observations in mESCs; hence, it would be necessary to conduct comparative studies between mouse and human ESCs in the context of NANOG function and its impact on lineage specification. Furthermore, some studies have also demonstrated that primed mouse epiblast and hESCs share similar culture conditions, gene expression patterns, and signaling responses.[18]

Our investigation into the cistrome of NANOG has provided further insights into the regulatory landscape governing the transition from naive to primed pluripotency. Through clustering analysis, we identified enriched motifs of known transcription regulators such as Zic, GRHL2, and Pou5f1::Sox2, indicating NANOG's interaction with a complex regulatory network involved in orchestrating this transition. Of particular interest is the presence of the Pou5f1::Sox2 motif, which is commonly observed in NANOG binding.[12] This motif suggests a combinatorial binding of NANOG with OCT4 and SOX2, two key pluripotency factors, to direct gene regulation. This finding aligns with previous studies highlighting the cooperative interactions among these transcription factors during development.[23] Furthermore, the enrichment of the LHX1 homeobox motif suggests that NANOG may be involved in the indirect recruitment of the LHX1 transcription factor. This transcription factor is known for its lineage specification during gastrulation and has been implicated in mesendoderm development.[55] However, whether this enrichment is a direct consequence of the changing role of NANOG remains a subject for further investigation.

Previous research has emphasized the role of the transcription factor OTX2 as a major regulator of the transition from naive to primed pluripotency.[42] It has been suggested that overexpression of *Otx2* is necessary for the global reorganization of OCT4 binding. Given the close interplay between NANOG and OCT4, it would be interesting to explore how the overexpression of *Otx2* affects global NANOG binding and its impact on the regulatory dynamics during the transition.

Lastly, the change in gene expression patterns also provided insights into its biological role. Evidence exists for maintaining pluripotency at the naive state and

regionalization, pattern formation during the transition, and primed state from the enrichment analysis (Figure 20). It is important to note that our enrichment analysis result provides only a preliminary understanding of the potential biological functions of Nanog during the transition. Further research is still needed to understand the transcriptional regulatory network between these genes and their specific roles in the context of NANOG regulation.

CHAPTER 3
NEUROECTODERMAL DIFFERENTIATION

During development, embryos usually undergo a process of gastrulation, which involves seperation into the three germ layers mesoderm, ectoderm, and endoderm.[4] The ectoderm layer differentiates to form the neural tubes, neural crest cells, and epidermis. This differentiation process is collectively regulated by morphogens known as organizers, which coordinate embryonic patterning and involve major signaling pathways such as Wnt signaling, bone morphogenetic proteins, activin, Nodal signaling, and retinoic acid.[56] As such, one method to induce neural differentiation is via retinoic acid.[58]

Retinoic acid, a metabolite of vitamin A, plays a crucial role in cell development and differentiation. When retinoic acid enters the nucleus, it binds to specific receptors called retinoic acid receptors (RARα, β, or γ) to form a heterodimer on the DNA. This interaction can activate the expression of retinoic acid-responsive genes by binding to retinoic acid-responsive elements (RAREs), which can ultimately trigger the expression of genes involved in cell differentiation and development.[57]

One of the immediate early genes to be transcribed during RA differentiation is the Homeobox gene *Hoxa1,* which possesses enhancer elements containing RARE. The mechanism of RA-induced neuroectodermal differentiation involves the promotion of neural gene expression and repression of mesodermal gene expression.[58] In addition, the RA-induced differentiation process involves the downregulation or dismantling of the transcriptional network that maintains the pluripotent state of ESCs. However, despite the dismantling of this network, there is evidence to suggest that

certain aspects of the ESC pluripotent transcriptional network may still have functional roles to play even after this process.

Previous studies have identified and characterized the regions bound by Hoxa1 on a genome-wide basis in differentiating mouse ESCs. It was established that a cross-regulatory interaction occurs between Nanog and Hoxa1 through a mechanism of mutual repression on similar target genes. HOXA1 binds to the regulatory regions of *Nanog*, while NANOG binds to the 3' enhancer of *Hoxa1*.[59] Therefore we characterized the genome-wide binding properties of NANOG and its regulatory interactions during early neuroectodermal differentiation.

METHODS

Dataset and Study Design

All raw sequencing data were obtained from NCBI Sequence Read Archive (SRA) database (Accession ID: SRP079975). KH2 ES cells were grown using N2B27+2i media to establish pluripotent conditions and then changed to a differentiation media [DMEM + 10% (vol/vol) Serum + NEAA + 0.03 μM RA]. Quality checks of the raw reads were processed similarly to the previous section. The reads were aligned to the University of California, Santa Cruz (UCSC) mm9 genome with bowtie2.[45] The Peaks were called with macs2, parameters "-p 0.25 -m 5 50". The top 100,000 peaks by p value, for each replicate, were compared with irreproducible discovery rate (IDR).[51] The Design is described in Figure 21.

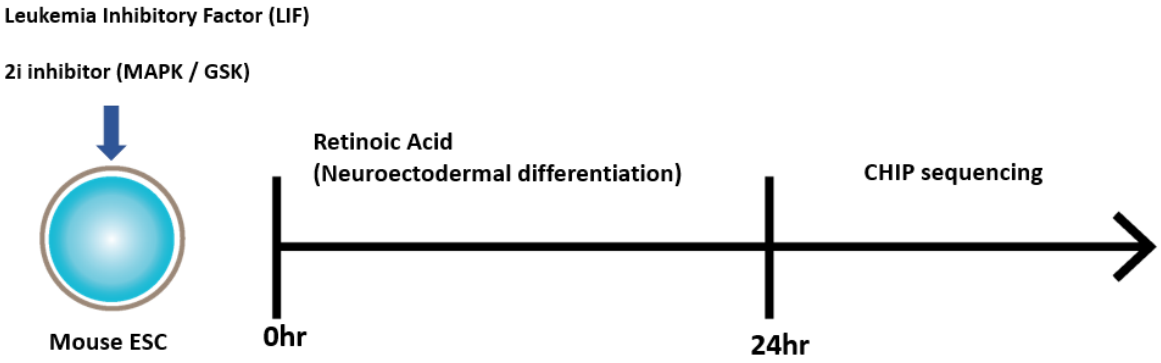

Figure 21. METHODOLOGY NEUROECTODERMAL DIFFERENTIATION

Diagram showing the differentiation of mouse ESC in retinoic acid for 24 hours following differentiation.

BIOINFORMATICS ANALYSIS

We focused on the genomic loci bound by NANOG during differentiation and how motif enrichment changes. The task of identifying sequence preference of a transcription factor is called Motif discovery. Using the HOMER module findMotifsGenome.pl mm10 -size -50, 50, we conducted the enrichment analysis. HOMER scores motifs by looking for motifs with differential enrichment between two sets of sequences.

RESULTS

NANOG BOUND REGIONS

The analysis from the chromatin immunoprecipitation sequencing identified 42,752 and 12,388 bound Nanog peaks in the uninduced and 24 hours of retinoic acid differentiation (Figure 22), respectively (idr $p < 0.01$). Of these sites, 15,547 loci were shared between both time points. In addition, Nanog binding was enriched at Transcription start sites (promoter) +-1kb at 24 hours of differentiation compared to the uninduced stage. Still, most of the binding occurred at intergenic and intragenic sites.

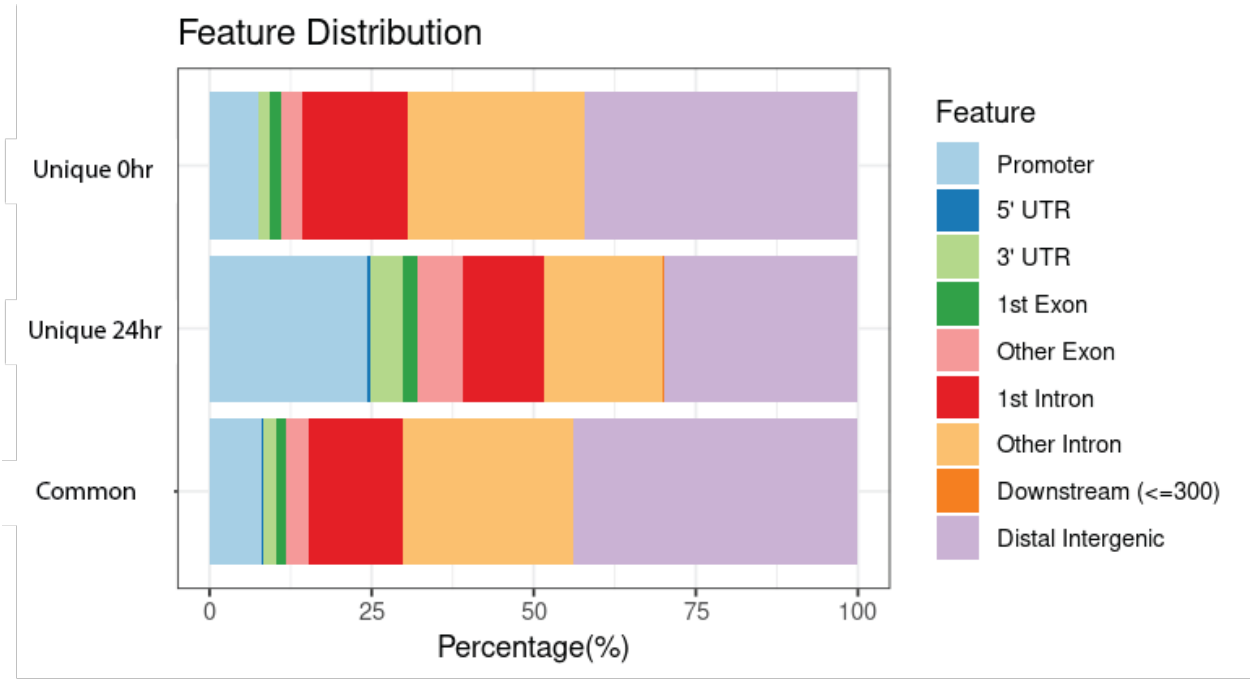

Figure 22. NANOG CISTROME IN NEUROECTODERMAL DIFFERENTIATION

Bar plot showing peak profiling, following NANOG binding. Most of the sites are bound to intergenic and intragenic regions.

MOTIF ANALYSIS

Motif analysis provides valuable insights into the mechanisms underlying transcription factor binding and their regulatory roles. In our study, we conducted a motif search on the NANOG uniquely bound regions in both (0hr) ESCs and (24hr) differentiated cells to gain a deeper understanding of the distinct mechanisms and potential regulatory interactions involved.

In the undifferentiated ESC cells, we observed a significant enrichment of motifs corresponding to pluripotent transcription factors such as NANOG, SOX3, POU5F1-SOX2, and KLF, among others (Figure 23). These findings align with our understanding of the role of these transcription factors in maintaining pluripotency and self-renewal in ESCs. The enrichment of these motifs suggests that NANOG collaborates with other pluripotent factors to regulate gene expression and maintain the undifferentiated state of ESCs.

Interestingly, at 24 hours of RA treatment, we observed a significant enrichment of the LHX1 motif (Figure 24). LHX1 is not traditionally considered a pluripotent factor but is classified as a homeobox gene. Homeobox genes are known to play critical roles in embryonic development and cell fate determination. The presence of the LHX1 motif within the NANOG binding regions suggests a potential functional interaction between NANOG and LHX1 during the early stages of differentiation. We also performed enrichment analysis of the genes closest to LHX1 motifs (Figure 25) and found biological processes which included pattern specification process, gland development, regionalization etc.

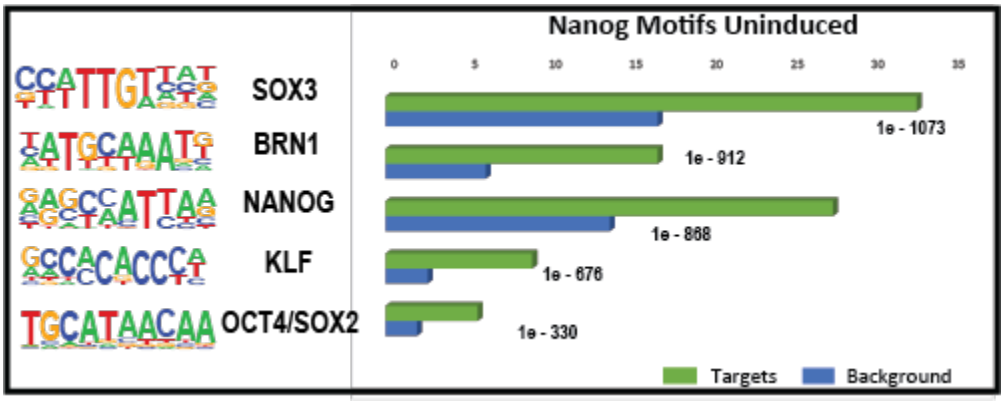

Figure 23. ENRICHED MOTIFS AT UNINDUCED STAGE

Bar plot showing the enriched motifs at the uninduced stage. The graph is in percentage.

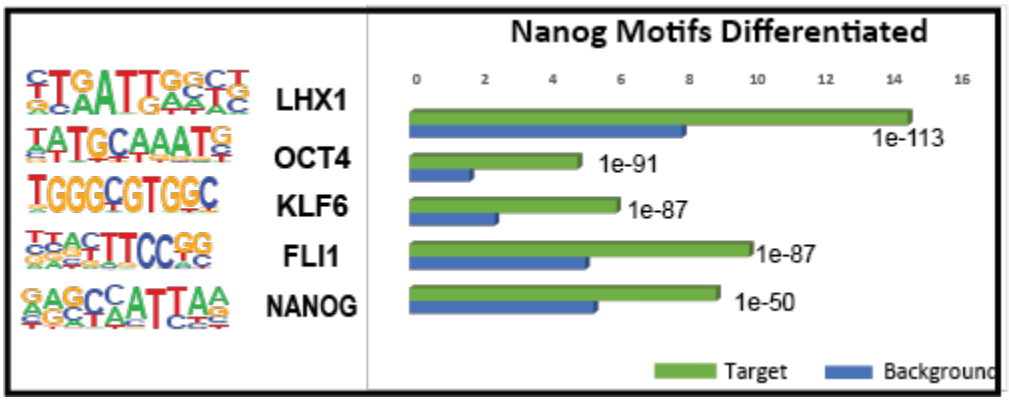

Figure 24. ENRICHED MOTIFS AT DIFFERENTIATED STAGE

Bar plot showing the enriched motifs at 24 hours of differentiation, the motifs are in percentages.

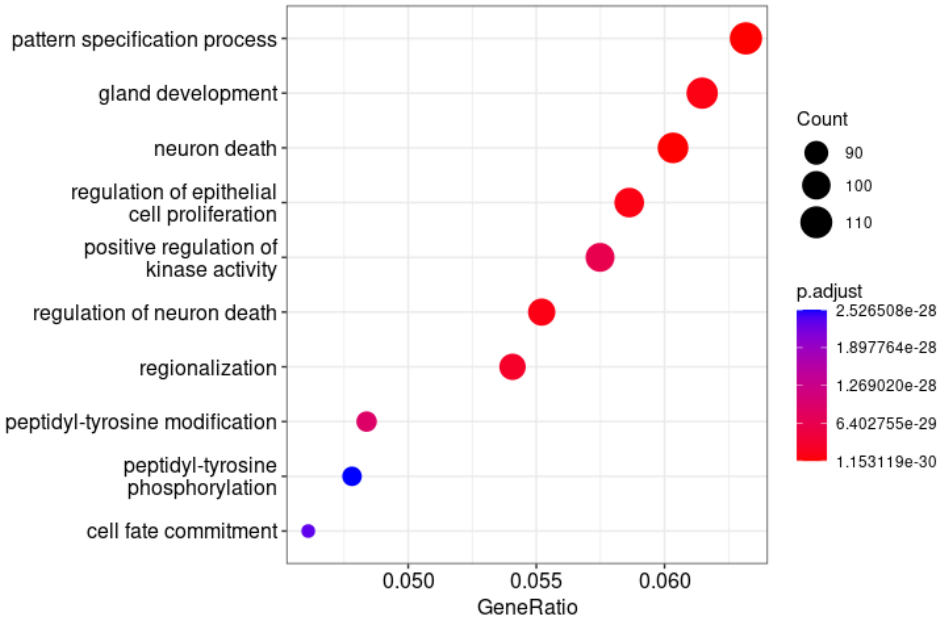

Figure 25. ENRICHMENT ANALYSIS OF GENES CLOSEST TO LHX1 MOTIFS

DISCUSSION

In this study, we conducted a genome-wide analysis of NANOG binding events during retinoic acid (RA) - induced neuroectodermal differentiation. Our hypothesis was that the enrichment of specific motifs indicates the binding of transcription factors and that changes in the motif enrichment during differentiation reflect changes in co-transcriptional activity.

Interestingly, we observed the enrichment of pluripotency factors in the uninduced state, consistent with the known role of NANOG in maintaining the pluripotent state. However, at 24 hours of RA differentiation, we observe a shift in motif enrichment, including the enrichment of a new class of motif LHX1. This finding is particularly intriguing given that LHX1 has been implicated in lineage specification. As LHX1 is a homeobox gene like *Nanog*, it may be premature to draw conclusions about its functional role based solely on motif enrichment. However, this finding is still compelling due to its occurrence in the previous naive to primed differentiation model. Conducting further ChIP sequencing studies or functional assays specific to LHX1 could provide a better understanding of any potential connection between these two transcription factors during development. These studies would therefore provide more direct evidence regarding the binding of the transcription factors to the identified motifs and help elucidate their functional implications in gene regulation and lineage specification.

In conclusion, our study revealed significant motif enrichments of pluripotent factors in undifferentiated ESC cells and lineage specification factors in mouse ESCs undergoing RA-induced differentiation. These findings suggest potential collaborations

and regulatory interactions between NANOG and lineage-specific transcription factors during early differentiation, shedding light on the regulatory mechanisms governing cell fate determination.

CHAPTER 4
FUTURE DIRECTIONS

Our study focused on two models of differentiation, the naive to primed transition and neuroectodermal differentiation, to investigate the role of NANOG in these processes. We observed that NANOG exhibits a dual role that is distinct between the naive and primed states. In the naive state, NANOG appears to be involved in the maintenance of pluripotency. In contrast, during the primed state, NANOG's role shifts towards lineage specification, which is supported by the enrichment of LHX1 transcription factor that has been implicated in mesendoderm development. The similar enrichment of LHX1 during neuroectodermal development further suggests NANOG's role is context dependent.

Our findings highlight the potential role of Nanog as a pluripotency factor that also directs mesodermal specification during the transition from naive to primed pluripotency. While our findings provide evidence supporting NANOG's involvement in mesodermal development, further research is necessary to validate and expand upon these observations. One limitation of our study is the reliance on marker genes to distinguish between lineages, as some markers can be expressed in multiple lineages. Therefore, it is crucial to validate lineage identities through functional assays or immunostaining.

Additionally, we acknowledge that NANOG's multifaceted role may stem from its transcriptional heterogeneity in cell fate decisions. While our study provided valuable insights, the bulk RNA sequencing analysis approach we used limits our ability to fully profile the transcriptional landscape at the individual cell level. Future investigations

should employ single-cell analysis techniques to unravel the heterogeneity within *Nanog*-expressing cells and determine its influence on cell fate determination. Another limitation is the lack of profiling of loss-of-function studies for *Nanog* in cells undergoing mesodermal differentiation in ESCs. Examining Nanog in this context would further elucidate its importance during mesodermal specification.

Furthermore, understanding the interplay between Nanog and other transcription factors is crucial for elucidating its role in regulating target genes. Since many loci bound by Nanog involve combinations of multiple transcription factors, it is essential to explore these cross-combinations and their effects on target gene regulation.

Although we initially studied different motif combinations, and assumed that their presence suggests transcription factor binding, we realized that this assumption is not sufficient for studying combinatorial transcriptional regulation. Conducting a ChIP-seq experiment of the various implicated motifs is necessary. Therefore, we can only suggest the involvement of the LHX1 in neuroectodermal differentiation based on its enrichment, without reaching a definitive conclusion.

Finally, we propose a model where combinatorial transcription factor binding can influence the role of NANOG. In conclusion, our study contributes to understanding NANOG in the context of pluripotency maintenance and lineage specification.

REFERENCES

1. Ross Granville Harrison, 1870-1959. *Biogr. Mem. Fellows R. Soc.* **7**, 110–126 (1961).
2. Nichols, J. & Smith, A. The origin and identity of embryonic stem cells. *Dev. Camb. Engl.* **138**, 3–8 (2011).
3. Mercader, A., Valbuena, D. & Simón, C. Human Embryo Culture. in *Methods in Enzymology* vol. 420 3–18 (Academic Press, 2006).
4. Vang, J. J. Developmental Biology, Tenth Edition by Scott F. Gilbert (Hardcover. *Dev. Biol. Tenth Ed. Scott F Gilbert Hardcover*.
5. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
6. Nir, T., Melton, D. A. & Dor, Y. Recovery from diabetes in mice by β cell regeneration. *J. Clin. Invest.* **117**, 2553–2561 (2007).
7. Shi, Y. Induced pluripotent stem cells, new tools for drug discovery and new hope for stem cell therapies. *Curr. Mol. Pharmacol.* **2**, 15–18 (2009).
8. Singh, V. K., Kalsan, M., Kumar, N., Saini, A. & Chandra, R. Induced pluripotent stem cells: applications in regenerative medicine, disease modeling, and drug discovery. *Front. Cell Dev. Biol.* **3**, (2015).
9. Li, M. & Belmonte, J. C. I. Ground rules of the pluripotency gene regulatory network. *Nat. Rev. Genet.* **18**, 180–191 (2017).
10. Apostolou, E. *et al.* Genome-wide Chromatin Interactions of the Nanog Locus in Pluripotency, Differentiation, and Reprogramming. *Cell Stem Cell* **12**, 699–712 (2013).
11. Kim, J. *et al.* An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell* (2008) doi:10.1016/j.cell.2008.06.004.

12. Göke, J. *et al.* Combinatorial Binding in Human and Mouse Embryonic Stem Cells Identifies Conserved Enhancers Active in Early Embryonic Development. *PLoS Comput. Biol.* **7**, e1002304 (2011).
13. Parfitt, D.-E. & Shen, M. M. From blastocyst to gastrula: gene regulatory networks of embryonic stem cells and early mouse embryogenesis. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130542 (2014).
14. Young, R. A. Control of the Embryonic Stem Cell State. *Cell* (2011) doi:10.1016/j.cell.2011.01.032.
15. Palmieri, S. L., Peter, W., Hess, H. & Schöler, H. R. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev. Biol.* **166**, 259–267 (1994).
16. Niwa, H., Miyazaki, J. & Smith, A. Quantitative Expression of Oct-3/4 Defines Differentiation, Dedifferentiation or Self-Renewal of ES Cells. *Nat. Genet.* (2000) doi:10.1038/74199.
17. Nichols, J. *et al.* Formation of Pluripotent Stem Cells in the Mammalian Embryo Depends on the POU Transcription Factor Oct4. *Cell* **95**, 379–391 (1998).
18. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
19. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183–193 (2016).
20. Multipotent cell lineages in early mouse development depend on SOX2 function. https://genesdev.cshlp.org/content/17/1/126.long.
21. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).
22. Chew, J.-L. *et al.* Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Mol. Cell. Biol.* **25**, 6031–6046 (2005).

23. Rodda, D. J. *et al.* Transcriptional Regulation of Nanog by OCT4 and SOX2. *J. Biol. Chem.* **280**, 24731–24737 (2005).
24. Reményi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059 (2003).
25. Allouba, M. H., ElGuindy, A. M., Krishnamoorthy, N., Yacoub, M. H. & Aguib, Y. E. NaNog: A pluripotency homeobox (master) molecule. *Glob. Cardiol. Sci. Pract.* **2015**, 36 (2015).
26. Festuccia, N. *et al.* Esrrb Is a Direct Nanog Target Gene That Can Substitute for Nanog Function in Pluripotent Cells. *Cell Stem Cell* (2012) doi:10.1016/j.stem.2012.08.002.
27. Chambers, I. *et al.* Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell* **113**, 643–655 (2003).
28. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
29. Mitsui, K. *et al.* The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell* **113**, 631–642 (2003).
30. Silva, J. *et al.* Nanog Is the Gateway to the Pluripotent Ground State. *Cell* **138**, 722–737 (2009).
31. Yeo, J.-C. & Ng, H.-H. The transcriptional regulation of pluripotency. *Cell Res.* **23**, 20–32 (2013).
32. Kopp, J. L., Ormsbee, B. D., Desler, M. & Rizzino, A. Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *Int. J. Cell Cloning* (2008) doi:10.1634/stemcells.2007-0951.
33. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).

34. Lu, R., Yang, A. & Jin, Y. Dual Functions of T-Box 3 (Tbx3) in the Control of Self-renewal and Extraembryonic Endoderm Differentiation in Mouse Embryonic Stem Cells*. *J. Biol. Chem.* **286**, 8425–8436 (2011).
35. Teo, A. K. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **25**, 238–250 (2011).
36. Thomson, M. *et al.* Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers. *Cell* **145**, 875–889 (2011).
37. Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. *Cell* **146**, 519–532 (2011).
38. Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130540 (2014).
39. Betschinger, J. *et al.* Exit from Pluripotency Is Gated by Intracellular Redistribution of the bHLH Transcription Factor Tfe3. *Cell* **153**, 335–347 (2013).
40. Guo, G. *et al.* Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**, 1063–1069 (2009).
41. Yang, S.-H. *et al.* ZIC3 Controls the Transition from Naive to Primed Pluripotency. *Cell Rep.* **27**, 3215-3227.e6 (2019).
42. Buecker, C. *et al.* Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. *Cell Stem Cell* **14**, 838–853 (2014).
43. Barral, A. *et al.* Nanog regulates Pou3f1 expression at the exit from pluripotency during gastrulation. *Biol. Open* **8**, bio046367 (2019).
44. Murakami, K. *et al.* NANOG alone induces germ cells in primed epiblast in vitro by activation of enhancers. *Nature* **529**, 403–407 (2016).

45. Langmead, B. & Salzberg, S. L. Fast Gapped-Read Alignment With Bowtie 2. *Nat. Chem. Biol.* (2012) doi:10.1038/nmeth.1923.

46. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

47. Deeptools2 bamCoverage. *NGS Analysis* https://learn.gencore.bio.nyu.edu/chipseq-analysis/exercise-part1-deeptools2-bamcoverage/ (2018).

48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

49. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

50. Zhang, Y. *et al.* Model-Based Analysis of ChIP-Seq (MACS). *Genome Biol.* (2008) doi:10.1186/gb-2008-9-9-r137.

51. Qunhua Li, James B. Brown, Haiyan Huang, & Peter J. Bickel. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

52. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S1097276510003667?token=40B61D9A232140F61BF6D5881BE87E0C0EA28DF87D13C852CFB83FE9B20050404A81FDF0641FF0F624E13BFDD04A0543&originRegion=us-east-1&originCreation=20230207150352 doi:10.1016/j.molcel.2010.05.004.

53. Boroviak, T. *et al.* Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis. *Dev. Cell* **35**, 366–382 (2015).

54. Evseenko, D. *et al.* Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc. Natl. Acad. Sci.* **107**, 13742–13747 (2010).

55. McMahon, R., Sibbritt, T., Salehin, N., Osteil, P. & Tam, P. P. L. Mechanistic insights from the LHX1-driven molecular network in building the embryonic head. *Dev. Growth Differ.* **61**, 327–336 (2019).

56. Kumar, V., Park, S., Lee, U. & Kim, J. The Organizer and Its Signaling in Embryonic Development. *J. Dev. Biol.* **9**, 47 (2021).

57. Gudas, L. J. & Wagner, J. A. Retinoids regulate stem cell differentiation. *J. Cell. Physiol.* **226**, 322–330 (2011).

58. Bain, G., Ray, W. J., Yao, M. & Gottlieb, D. I. Retinoic Acid Promotes Neural and Represses Mesodermal Gene Expression in Mouse Embryonic Stem Cells in Culture. *Biochem. Biophys. Res. Commun.* **223**, 691–694 (1996).

59. De Kumar, B. *et al.* Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells. *Proc. Natl. Acad. Sci.* **114**, 5838–5845 (2017).

60. Weidgang, C. E., Seufferlein, T., Kleger, A. & Mueller, M. Pluripotency Factors on Their Lineage Move. *Stem Cells International* **2016**, e6838253 (2015).

61. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.

62. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016 Sep 15;32(18):2847-9. doi: 10.1093/bioinformatics/btw313. Epub 2016 May 20. PMID: 27207943.

63. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014 Apr 1;30(7):923-30. doi: 10.1093/bioinformatics/btt656. Epub 2013 Nov 13. PMID: 24227677.

DATASET

DATASET	CONDITION	SOURCE	ID	CHAPTERS
RNA-Seq	NANOG KO	Gene Expression Omnibus (GEO)	GSE138818	2
ChIP-Seq	TRANSITION FROM ESCs to EpiLC's	Gene Expression Omnibus (GEO)	GSE71932	2
ATAC-Seq	TRANSITION FROM ESCs to EpiLC's	ArrayExpress	E-MTAB-7207	2
ChIP-Seq	RETINOIC ACID	NCBI Sequence Read Archive	SRP079975	3

SOFTWARE AND ALGORITHMS

TOOL	VERSION	ID
Bowtie2	2.5.1	http://bowtie-bio.sourceforge.net/bowtie2/ index.shtml
HOMER	4.11	http://homer.ucsd.edu/homer/
DESeq2	1.8.3	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ComplexHeatmap	2.1.4.0	https://jokergoo.github.io/ComplexHeatmap-reference/book/
deepTools	2	https://deeptools.readthedocs.io/en/develop/index.html
clusterProfiler	4.6.2	https://yulab-smu.top/biomedical-knowledge-mining-book/index.html
ChIPseeker	1.35.0	http://bioconductor.org/packages/devel/bioc/vignettes/ChIPseeker/inst/doc/ChIPseeker.html

ABBREVIATIONS

mESCS - Mouse Embryonic stem cells
EpiLCs - Epiblast like cells
EpiSCs - Epiblast stem cells
2i - two kinase inhibitor
LIF - Leukemia Inhibitory factor
MAPK - Mitogen activated protein kinase
ERK - Extracellular signal-regulated kinases
GSK - Glycogen synthase kinase
FGF - Fibroblast growth factor
PCA - Principal component analysis
ATAC - Assay for Transposable accessible chromatin
ChIP - Chromatin Immunoprecipitation
RNA - Ribonucleic Acid
Seq - Sequencing
ESC - Embryonic stem cell