UNIVERSITY OF CALIFORNIA,

IRVINE

Comparative Functional Genomics of Mammalian Endodermal and Neuronal Development

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Developmental and Cellular Biology

By

Christina Wilcox Thai

Dissertation Committee:

Professor Ali Mortazavi, Chair

Professor Tallie Z. Baram

Professor Tom Schilling

Professor Peter Donovan

2021

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGEMENTS

I would like to thank many people for their help and support throughout my Ph.D. journey.

First, I would like to thank and express my gratitude to my advisor Professor Ali Mortazavi. Dr. Mortazavi introduced me to computational biology, genomics, and developmental biology. With his support I was able to take part in many opportunities I would otherwise not have access too. This includes traveling to classes, seminars, and conferences as well as access to the science and projects that I was involved in during my years at UC Irvine. He was heavily involved in the projects included in this dissertation. I appreciate everything that he has done for me and the opportunities he has presented to me over the years.

I would also like to thank the other members of my thesis committee, Dr. Tallie Z. Baram, Dr. Tom Schilling, and Dr. Peter Donovan for their continued support and guidance throughout this thesis.

I would like to thank the Conte Center at UC Irvine. I would especially like to thank Dr. Tallie Z. Baram for funding support as well as her integral role in many of the projects within this dissertation.

I would like to also thank my family, especially my parents for their constant support through thick and thin. Their guidance and support was invaluable to my dissertation work. I would like to thank my husband for his continued support and encouragement throughout my graduate studies.

I would like to thank Dr. Shan (Mandy) Jiang for her guidance and friendship throughout my graduate school experience. I would also like to thank Dr. Annabel Short for her guidance and input in multiple projects within this dissertation. She has been a great friend and mentor throughout the last few years.

I would like to thank my lab mates, past and present, for their wonderful friendship and overall support throughout all aspects of graduate student life, including guidance in computational biology and wet lab experiments. I especially would like to thank Kate and Bea for being the greatest of friends and giving continued emotional support as well as sharing their experiences in science and in life. I would also like to thank Kate for organizing all the fun social outings over the years.

Finally, I would like to thank the GHTF at UCI, especially Melanie Oakes who was instrumental in organizing the sequencing of samples in a thorough and accelerated manner.

CURRICULUM VITAE

Christina Wilcox Thai

www.linkedin.com/in/christinawilcoxthai

EDUCATION

University of California, Irvine
PhD in Developmental and Cell Biology — Jun 2021

University of California, Davis
Masters of Science in Nutritional Biology — Jun 2015

California State University, Sacramento
Bachelor of Science in Biomedical Sciences — Dec 2012

EXPERIENCE

Graduate Student Researcher, Mortazavi Lab, Irvine, CA — Apr 2016-present

- Collaborated with a team of scientists to investigate the effects of early life stress on the development of the brain in mice and human subjects
- Managed a small team of undergraduate researchers to understand how changes in epigenetic regulatory networks contribute to the differentiation of embryonic stem cells to neuronal cell types in 3 different mammalian species
- Taught multiple discussion and lab sections including a freshmen biology, anatomy lab, and cell biology where I was responsible for developing the syllabuses, creating teaching material, writing and grading quizzes, exams, and written assignments

Graduate Student Researcher, Hawkes Lab, Davis, CA — Jan 2014-May 2015

- Taught a systemic physiology lab course where I developed the syllabus, taught discussion sections, graded lab reports, quizzes, and oral examinations
- Lead a small team of undergraduate researchers in investigating the effects of silencing specific dietary proteins on liver cancer prognosis
- Managed all aspects of study design, data collection, and analysis

Graduate Student Researcher, Keen Lab, Davis, CA — Jun 2013-Dec 2013

- Explored the effects of Flavonoid rich foods on the cardiovascular system and overall health of 14-18-year-old male human subjects
- Aided with the recruitment of human subjects and their legal guardians into the study
- Assisted participants in understanding the study, their responsibility as a participant, and obtaining informed consent

Undergraduate Student Researcher, Stem Cell Partners, LLC, — Dec 2012-Dec 2013

Collaborated with a team of fellow undergraduate researchers in culturing mesenchymal stem cells extracted from primary adipose tissue obtained from liposuction

SKILLS

Laboratory: Next Generation Sequencing (NGS), Illumina sequencing, Data analysis, RNA-sequencing, single cell sequencing techniques, Assay for Transposase-Accessible Chromatin (ATAC) sequencing, Cell culture, Gene delivery, CRISPR techniques, Cloning, Immunochemistry, Fluorescent microscopy, Fluorescent Activated Cell Sorting (FACS), Electrophoresis, Immunoblots, PCR, RT-qPCR, animal dissections, mouse colony maintenance, mouse embryo isolation, clinical patient recruitment, informed consent

Computational Analysis: Python, R Statistical Program, Microsoft Office, Adobe Photoshop

TEACHING EXPERIENCE

Teaching Assistant for Anatomy 2018-2019

University of California, Irvine

Responsibilities: taught lab sections; held office hours

Dr. Justin Shaffer, professor.

Teaching Assistant for Cell Biology 2017

University of California, Irvine

Responsibilities: taught discussion sessions; held office hours

Dr. Christine Sutterlin and Dr. Grant MacGregor

Teaching Assistant for Biology Organisms to Ecosystems 2016

University of California, Irvine

Responsibilities: Developed Syllabus; taught discussion sections; held office hours; graded quizzes, and student oral presentations; and administered all final grades. Dr. Justin Shaffer, professor.

Teaching Assistant for Systemic Physiology Lab 2014-2015

University of California, Davis

Responsibilities: Developed Syllabus; taught lab and discussion sections; held office hours; graded lab reports, quizzes, and student oral presentations; and administered all final grades. Dr. Erwin Bautista, professor.

PAPERS

Thai, CW., Jiang, S., Roxas, YD., Zeng, C., McGill, CJ., Ma, XY., and Mortazavi, A., Comparative Chromatin Dynamics of Stem Cell Differentiation in Human and Rat. (submitted)

Thai, CW., Short, AK., Mortazavi, A., Baram, TZ. Single-Cell Transcriptomics of hypothalamic CRH-expressing neurons informs mechanisms by which Early-Life Adversity sculpts stress-responses enduringly. (in preparation)

ABSTRACT OF THE DISSERTATION

Comparative Functional Genomics of Mammalian Endodermal and Neuronal Development

By

Christina Wilcox Thai

Doctor of Philosophy in Developmental and Cell Biology

University of California, Irvine, 2021

Professor Ali Mortazavi, Chair

Development and differentiation are complex cellular processes that are regulated by a cohort of transcription factors (TFs) and signaling molecules that can be investigated at the levels of cells, tissues, and whole organisms. Regulation of these processes is, in part, controlled by changes in DNA accessibility surrounding protein-DNA binding sites around promoter and enhancer regions of downstream target genes. Differentiation of cell types that are homologous between species are controlled by conserved networks of regulatory elements driving gene expression. In order to identify conserved and species-specific aspects of gene expression and chromatin accessibility during cell differentiation, we conducted a time-course study of differentiation in two different species between two cell lineages. We are interested in investigating the extent of conservation of developmental gene expression between species and what we can learn from these similarities including their epigenetic basis. We hypothesize that the conservation of developmental expression is encoded at the epigenetic level in changes to open chromatin and DNA methylation. The environment of an organism also influences development, both on a whole organism level and during brain development. Epigenetic

changes such as differences in DNA methylation, and accessibility of TFs can have lasting effects on the phenotype of the organism. Although some epigenetic changes, such as changes in the cellular transcriptome, are localized to the affected cell or tissue, other epigenetic changes such as changes in DNA methylation can occur on a systemic scale and affect more than one organ or developmental process. During sensitive periods of development, environmental factors may have serious consequences for an organism's health. Immune disorders and psychiatric diseases have both been linked to disturbances in epigenetic regulation.

In this thesis, I describe my studies of chromatin dynamics and TF expression as well as environmental impacts on the epigenetics of brain development in multiple mammalian species. In Chapter 2 I strive to determine conserved or species-specific candidate developmental regulatory elements in distantly-related species. I find that definitive endoderm (DE) differentiation from embryonic stem cells is more conserved than neural progenitor cell (NPC) differentiation, with 22% and 8% conservation of candidate cis-regulatory elements, between human and rat, respectively.

In Chapter 3 I focus on investigating the developmental plasticity of epigenetic regulation in corticotropin-releasing hormone (CRH)-expressing hypothalamic paraventricular nucleus (PVN)-neurons during early life adversity. The differentially expressed genes between early life adversity (ELA) and control (CTL) suggest that ELA PVN neurons undergo more cellular stress and neuronal activity than control neurons. Glutamatergic PVN neurons of the ELA condition may have altered energy production or are not able to properly maintain excitability or synaptic transmission.

In Chapter 4 I evaluate DNA methylation changes in a number of individuals between birth and one year of life to determine the correlation to a range of life experiences. Individuals

from the high unpredictability group had striking increases in methylation, suggesting reduced expression, of genes involved in cell junction organization, actin filament-based processes and differentiation. They also had decreased methylation, suggesting overexpression, of genes associated with cell proliferation, growth factor signaling, and cell motility. From such data we may be able to predict genes involved in early life experiences and their possible long-term consequences.

CHAPTER 1

Introduction

Neurodevelopment in Model Organisms and Humans

Chapter 1

Introduction: Neurodevelopment in Model Organisms and Humans

1.1 Introduction

How cells differentiate and respond to their environment is a vast question. In this thesis we explore two very different areas of research addressing this question. In the first part, I investigate the GRNs that control cell fate using time-courses of RNA-seq and ATAC-seq. More specifically, I analyze how conserved cell-type specific time courses are between human and rat. At the other extreme I studied the effect of the environment on epigenetics using early life adversity in a rodent model as well as in human infants. Given that these are quite distinct fields of research I will cover the background for each of these projects separately.

My three projects include the conservation of the transcriptome and chromatin dynamics of transcription factors involved in neural progenitor cells (Fig. 1.1) and definitive endoderm (Fig. 1.2) differentiation in human and rat, early life adversity and it's effects on the transcriptome of the Crh+ PVN neurons in the hypothalamus, and the impact of early life adversity in humans on the methylation profiles of buccal swab samples between infants and one year old children.

Understanding developmental biology, or how an organism develops from a single cell at fertilization, to a complex, multi-cellular organism, gives great insights into the mechanisms of nominal processes and manifestations of diseases [1]. Mammalian development is a very complex and well-regulated process[1,2]. After fertilization, the zygote undergoes cleavage, or rapid cell division, eventually forming the blastocyst, which is made up of two layers, an inside layer termed the inner cell mass (ICM), and an outer layer termed the trophoblast [2]. The ICM will go on to form the entire embryo, and the trophoblast will form the extraembryonic structures like the placenta [2]. If extracted, the ICM can be cultured *in vitro*, as embryonic stem cells

(ESCs). These cells, like their *in vivo* counterparts, can differentiate into any cell in the organism and are a powerful tool to understand development and differentiation *in vitro* [1–5].

To understand the processes that occur during development, it is important to utilize both *in vivo* and *in vitro* models of development. For instance, rodents and other mammalian model organisms have been used to study *in vivo* development of the human condition. Due to their widespread use as model organisms, it is important to understand the similarities and differences in developmental processes between rodents and humans. Since there are moral restrictions on studying *in vivo* development in humans, it is important to have a good *in vitro* model system, such as human ESCs, for comparison [1,6].

Brain development and more specifically neurogenesis, which is the formation and development of neurons in the brain, remains poorly understood [7]. One reason for this is its vast complexity. For example, the human brain contains approximately 100 billion neurons at birth, each with at least one or more synaptic connections to other neurons [8]. On average about 250,000 new neurons are formed per minute in the developing brain during prenatal development [8]. Given the vast complexity of the brain, understanding how each neuron or group of neurons is formed and affected by its connections is an enormous task. There is also the added complexity of the environmental impact on the developing brain; stress, malnutrition, as well as other environmental impacts have all been shown to influence brain development [9]. Maternal prenatal stress has been shown to have an effect on parasympathetic control of the fetal heart, placenta function, and maternal stress hormone levels that increase fetal activity and decrease fetal weight [10]. Malnutrition prior to and after conception has been associated with learning disorders, and impacts the learning control centers of the brain, and the ability to adapt to one's environment [11].

The complexity of neurogenesis during brain development and the possible environmental influences make the study of neurogenesis a challenging process. The lack of access to human brain samples for research is another reason why there is still a lot to learn about neurogenesis and the developing brain. Developmental neuroscience relies heavily on model organisms. Model organisms have played a large role in the study of neural development due to the moral implications of studying the developing human brain. Model organisms and humans share many similarities, especially key milestones in neuronal development, that make developmental studies in model organisms comparable to human development [12].

1.2 Mammalian Early Development and the Formation of the Neural Tube

Some of the first key events in early development are axis formation and gastrulation. In mice, cells need to determine positional information along the body axes, which may be initiated before gastrulation [13]. The formation of a structure called the primitive streak establishes embryonic symmetry and determines the site of gastrulation [14]. The three-body axes are the anterior-posterior (AP), the dorsal-ventral (DV), and the left-right axis. Later, the cells will be assigned to a fate based on their position along these axes [13].

The formation of the three germ layers (ectoderm, mesoderm, and endoderm) in gastrulation is one of the most important developmental processes in the life of an organism as these cells lose the indefinite self-renewal capacity and become more differentiated [3,6,13]. During gastrulation, environmental signals determine which differentiation path each cell will take, either becoming progenitors of mesendoderm (ME) or neural ectoderm (NE) [15,16]. ME cells are bipotent progenitors, either differentiating into endoderm or mesoderm cells.

The formation of ME progenitors is rigorously controlled by a group of signaling molecules and transcription factors (TF) of the Nodal, Wnt, BMP, FGF, GATA, Sox, and Fox

families among others [15,17]. The formation of NE progenitors is controlled by Sox2, Foxp1, Rbpj, Dnmt3a, and Zfp532. Nodal signaling inhibits the formation of NE [17] whereas expression of the transcription factor Sox2 drives the formation of NE and represses the ME fate [15], suggesting that differentiation into these two progenitor types is rigorously controlled via gradients of signaling molecule and transcription factor expression in different parts of the embryo [15]. ME progenitors undergo an epithelial to mesenchymal transition during their differentiation that initiates their migration through the primitive streak to form the endoderm and mesoderm layers [14,16]. The cells near the anterior portion of the primitive streak start expressing Foxa2 and become DE. DE forms into the foregut, midgut, and hindgut via regional AP inductive signals from adjacent mesoderm. Foregut cells will express Hhex, Sox2, and Foxa2 and hindgut cells will express Cdx1, Cdx2, and Cdx4 [18]. Ectodermal precursors, which are anterior to the primitive streak [1419,20], form the neural plate which, by induction from a mesodermal structure called the notochord, fold into the neural tube and become neuroectoderm [21].

While the neural plate folds into the neural tube, the resulting neuroectodermal cells pattern themselves to create distinct categories of neural progenitor cells (NPCs) that will further differentiate to form either the forebrain, midbrain, hindbrain, or spinal cord [22,23]. Since the notochord induces the formation of the neural tube and thus the neuroectoderm, proximity to the notochord determines the fate of the cells in the neural tube [24–26]. The cells in the ventral midline of the neural tube create the floorplate, which further directs differentiation into the different classes of NPCs [25,27,28]. The ventral portion of the neural tube will eventually give rise to motor neurons. Cells that are farther away from the ventral midline form sensory neurons of the CNS [24,29]. A third population of cells is formed where the folded edges of the neural plate meet, called the neural crest. Cells in this population migrate away from the neural tube and are exposed to different kinds of external stimuli and as a result, will form many different types of neuronal and non-neuronal cells [24,29]. Neural crest cells give rise to neurons and glia of the

sensory and visceral motor ganglia, the sensory cells of the adrenal gland, and the intestinal nervous system. They will also form non-neuronal structures like pigment cells, cartilage, and bone [24,29]. It is important to understand the underlying signaling molecules and transcription factors that drive these different processes. Although we can study the mechanisms of some of these processes *in vivo* using animal models, we also need to understand how these processes relate to humans using *in vitro* methods. There are no published functional genomics studies of NPC or DE differentiation using rat ESCs. We first adapted mouse protocols for NPC and DE differentiation in rat and assayed open chromatin and gene expression changes during the differentiation time-courses. We compare Rat and Human NPCs and DE differentiation to investigate transcription factor expression and accessibility surrounding possible regulatory elements of these cell types.

1.3 *in vitro* Methods for Studying Neurodevelopment

The differentiation of ESCs into NPCs and mature neurons is a powerful *in vitro* method for studying the development and differentiation of neurons [23,30]. *In vitro* models of neurodevelopment recapitulate morphology changes and signaling pathways that occur *in vivo* [30,31]. For instance, the neuro-rosette formation seen *in vitro* ESC differentiation to NPCs resembles the *in vivo* neuroectoderm and the cells express many of the same markers found *in vivo* [26,31]. Some of these markers include general neural markers like βIII-Tubulin, and Nestin, whereas Sox2 and Pax6 are expressed in neural stem cells, Tbr2 and Ctip2 are forebrain progenitor markers, and Lmx1a, Otx2, and Foxa2 are midbrain progenitor markers, and Zic1 is a hindbrain marker [23,26,31, 85]. Taken together, ESC differentiation into NPCs shows changes in morphology and gene expression over a similar timeframe, differentiation state and markers as *in vivo* patterning and gene expression [85]. We can use *in vitro* methodologies to study neuronal development in an accurate manner.

Another complication of studying early neural development *in vitro* is that numerous neural progenitor cell subtypes exist, many of which can be recapitulated *in vitro* [23,32,33]. Comparative studies need to take into account the difference between NPC subtypes and how they are different. These subtypes have been shown to be different in their location in the CNS *in vivo,* morphology, as well as their gene expression profiles. Subtypes of NPCs that have been recapitulated *in vitro* include forebrain, midbrain, hindbrain and spinal cord [23]. There are also larger differences between embryonic and adult NPCs. Embryonic NPCs have the ability to self-renew as well as asymmetrically divide to a neural progenitor cell and a neuron daughter cell [33].

1.4 Rodents as a genetically accessible mammalian model of Brain Development

Since *in vitro* methods have been established for the study of neural development we can start to compare model organisms to humans to determine the conservation and accuracy of using model organisms to study development. Rodents are the most utilized model organisms for biomedical research, especially neurodevelopment and neuroscience related research [34,35]. Up until recently, mouse models were used in the majority of neuroscience related genetic studies because of technique availability in mice that was not possible, or was very difficult in rats. However, as the genetic toolkit for rats catches up with mice, rats are becoming increasingly popular in neuroscience related research [35–37]. Rats are genetically and physically different from mice, diverging between 15-20 million years ago, and using rats as a model organism can have clear advantages over mice [34]. Rats are much easier to handle and are less stressed when handled by researchers. This is especially important in neuroscience research due to the physiological changes stress can cause on the brain and other systems. Rats also have larger brains, which makes neuroscience related studies much easier to conduct [34]. There is also increasing evidence of dramatic differences between the rat brain and that of

the mouse. The prolonged development of the rat brain, increased adult neurogenesis, as well as the more complex, and flexible behavioral range of the rat compared with mice suggest that the rat brain may be the better model for understanding the human [38]. An area where available research is lacking for rats is *in vitro* studies. Now that rats are becoming popular for genetic research, it is important to catch up to mice in *in vitro* signalling pathway, developmental regulation, and evolutionary comparison studies, among others. Since we are starting to understand the differences between rats and mice, the differences between rat and human need to be investigated to further validate the usefulness of rat as a model organism.

1.5 Rat as a Model Organism for Human Brain Development

One of the most prominent developmental differences between rat and human is the gestational time. Rodents have a gestational period of 18-24 days depending on the species, mice are at the beginning of this range (18-19 days) and rats are in the middle of this range with a gestational period of 22 days [39]. Humans have a much longer gestation period of 280 days, or 40 weeks [40]. The short gestational time period in rats and other rodents gives an advantage to researchers for increased productivity and decreased cost. However, this difference can have drawbacks as there are large time differences between when cells and tissues form. For example, neural tube development occurs around gestational day 10.5 in rats, about halfway through their gestational time period. In humans, the beginning of this event occurs earlier in development, between gestational day 24 and 28 out of a gestation period of 280 days [12,41]. Longer gestation periods have been associated with more complex and developed neural networks [42]. In humans, the earlier formation of the neural tube in relation to gestational time can be explained by the increased complexity of the human neurological system or longer gestational period compared to rat [39]. While the timescale is different between these two species, the sequence of major events in brain development remain remarkably similar [12,41].

Both rodents and humans are considered to be born in an underdeveloped state, and are considered to be altricial species [41]. Another difference which needs to be considered between rodent and human development is that rats also have a large litter size and can have around 9 fetuses *in utero* at a time. Whereas 98% of human pregnancies have one fetus [40]. A smaller number of fetuses in human versus rodents correlates with larger fetal size. Even though both species have offspring that are very dependent on maternal care. Rodents and humans are born at different neurodevelopmental stages. The human brain at birth is at a similar developmental stage as rodents at PND 7 [41]. These differences need to be considered when using rats as a model to study brain development.

1.6 Comparison of Human and Rat Embryonic Stem Cells

For developmental studies, a commonly used *in vitro* system is ESC differentiation. Before starting a developmental *in vitro* comparison study, it is important to understand the differences between rat and human ESCs as well as differing culturing methods between these two species. Two ESC states have been described, the first described is the primed state, which is a later post-implantation pluripotent state. The second state is the naive state which is considered an earlier, pre-implantation ground state ESC [23]. Both states are pluripotent and can make all cells of the embryo but are distinguished by gene expression signatures, chromatin states, and reliance on extracellular signals to remain pluripotent [43]. Recently, a transition state between naive and primed ESCs but distinct from either state has been described [43]. The ESC state depends heavily on the method of extraction and the gestational age of the harvested ICM. For studies that use established cell lines, the cell states have been well investigated. Human H1 ESCs are in a primed state [44]. Rat DAc8 ESCs are thought to be in-between the naive state and primed state. Rat ESCs are cultured in a medium containing two inhibitors (2i), which in mouse ESCs maintain a naive-like cell state. However rat ESCs express Cdx2, which

is a trophoblast marker not expressed in naive mouse ESCs [45].. It is important to understand the states of the cells from each species for *in vitro* differentiation to know what factors are active at the start of experimentation.

1.7 Transcription Factor and Chromatin Dynamics in ESC Differentiation into Neuronal Lineages

At its core, neuronal lineage specification and differentiation are driven by chromatin remodeling causing accessibility changes around TF binding sites near promoter and enhancer regions called cis-regulatory modules (CRMs) along with downstream signaling molecules, which collectively form Gene Regulatory Networks (GRNs) [46,47]. The identification of CRMs that are required for the regulation of gene expression during neuronal development are important for broadening our understanding of the conservation or divergence of development between species [48].

One way to study CRMs is through a method termed chromatin footprinting, which requires deep sequencing of DNA in chromatin to investigate where TFs might be bound to the DNA. One method commonly used in footprinting analysis is Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) where open chromatin (or accessible) regions are sequenced deeply. In a deeply sequenced open chromatin region, a reproducible decrease in reads inside an otherwise open promoter or enhancer suggests a TF binding site [49]. Even though only about 5% of individual DNA bases are conserved between rat and human, TF binding sites are about 20% conserved, meaning that these sites are present in the same homologous sequence in both species [50]. On the GRN level there is about a 95% conservation between mouse and human, suggesting that the evolutionary conservation of regulatory

networks is under selection, but the CRMs and individual bases are less so, allowing for plasticity between species [50].

A key area of research that has been understudied in rats is the regulation of lineage specification in neural development via transcription factors and signaling molecules. One way to study this is through the use of ESC differentiation. Comparative studies of ESC differentiation into neuronal lineages found there are a number of neuronal markers and transcription factors that are conserved between human and mouse [51]. Similar studies need to be conducted in rats to determine the conservation of this model organism with humans.

1.8 Brain and Circuit Development

A circuit refers to a set of interconnected components that work together and perform a particular function. The definition of a brain circuit is slightly more complicated. A brain circuit can consist of a group of neurons that receives electrochemical inputs that the circuit reacts to and transmits to other brain circuits until a particular outcome is achieved. Or a brain circuit can refer to a network of brain regions that integrate information and perform more complex functions [52]. The formation of neural circuits depends on a very complex set of genetic programs, neurodevelopmental events, and environmental stimuli. Because of this, neural circuits are dynamic and flexible and can be vulnerable during specific times in their development [52].

In the prenatal brain, as developing neurons mature into post-mitotic neurons, they migrate via a diverse set of signaling gradients, neighboring neuronal cells, and structural glial cells. As these developing neurons reach their destination, they form axons and dendrites to connect with the appropriate synaptic partners. These synaptic connections in early development are often transient and aid in the process of the development of the more stable

mature circuit connections [52]. One of the more important specialized groups of neural cells forms the hypothalamus.

1.9 Formation and Function of the Hypothalamus, PVN, and Stress Pathway in the Brain

The hypothalamus is located in the ventral forebrain and regulates energy homeostasis, fluid balance, stress, growth, reproductive behavior, emotion, and circadian rhythms [53]. The telencephalon and diencephalon forms from the anterior-ventral part of the neural tube by graded and specifically timed WNT signaling. The hypothalamus is generated from the diencephalon. The developing hypothalamus can be divided into the alar plate, basal plate, floor plate, and the intrahypothalamic diagonal. Expression of Sim1 and Pax6 are markers of the alar plate. Nkx2.1 is a marker of the basal plate. The alar plate gives rise to the paraventricular nucleus of the hypothalamus (PVN) and supraoptic nucleus (SON), whereas the basal plate forms the arcuate nucleus (ARC), dorsomedial nucleus of the hypothalamus (DMH) and ventromedial nucleus of the hypothalamus (VMH). Many TFs play a role in the development and maturation of the PVN, including: Sim1, Sim2, Arnt2, Brn2, Otp, and COUP-TFII [53].

The PVN contains heterogeneous parvocellular neurons, magnocellular neurons, and long-projecting neurons. Parvocellular neurons project to the median eminence and in response to environmental stress secrete corticotropin-releasing hormone (CRH) to initiate the hypothalamic-pituitary-adrenal (HPA) axis. CRH positive neurons in the PVN receive the synaptic input from catecholaminergic neurons in the locus coeruleus, nucleus of the solitary tract (NTS) of the brainstem, and ventrolateral medulla to activate the HPA axis in response to stress [53]. In the HPA axis, CRH prompts the secretion of adrenocorticotropic hormone (ACTH) in the anterior pituitary (AP), causing the secretion of glucocorticoids in the adrenal gland [53]. At

baseline, adrenocortical glucocorticoids are crucial for maintaining homeostasis by regulating glucose, hormone, and neurotransmitter levels in the brain and peripheral tissues. Environmental stress causes rapid but transient increases in glucocorticoid secretion, the magnitude of which depends on the duration and intensity of the stimulus. Glucocorticoids normally function by altering gene expression via glucocorticoid receptors (GRs) on cells, which cause a signaling cascade that will eventually enter the nucleus and either increase or inhibit expression of target genes [54]. CRH positive neurons can also release AVP and OXT, which cause secretion of ACTH in the AP and secretion of glucocorticoids [53].

The HPA axis is mediated by various hormonal signals and distinct neuronal inputs. Glucocorticoids, as well as being a target of HPA axis activation, provide negative feedback on the HPA axis. This negative feedback limits HPA axis activation by inhibiting expression and secretion of CRH and vasopressin in the hypothalamus as well as ACTH secretion and expression of its precursor protein, proopiomelanocortin (POMC). In contrast to their roles in the AP, another negative feedback mechanism is the region-specific dendritic release of AVP and OXT within the PVN and SON, which can inhibit the secretion of HPA axis hormones [55].

1.10 Cellular and Molecular Diversity of CRH Positive Neurons of the PVN

There are three different subdivisions of the PVN, each marked with a unique pattern of gene expression. Gad2 and Ntng1 expression is highly enriched in anterior PVN. Vglut2, Crh, and Avp expression has been found in the middle PVN, and Vglut2, Gad2, Npy1r. Crh, Reln, Ntng1, Pdyn, Oxt, and Avp are highly expressed in posterior PVN. Co-expression with CRH differs based on the anterior-posterior position of the neurons. For example, the majority of CRH neurons in the posterior PVN co-express Npy1r, where only a small number of CRH neurons co-express Npy1r in the anterior or middle PVN [56].

The PVN has the highest concentration of CRH positive neurons in the brain, and this population can be further divided into three major subclasses. The first is parvocellular, termed for their small size, CRH neurons found in the anterior and medial–dorsal region of the PVN with axons projecting to the external zone of the median eminence. These CRH neurons co-express other peptides including vasopressin (VP), enkephalins, cholecystokinin and angiotensin II. In response to stress this subclass of CRH neurons expresses a small amount of VP and secretes VP into pituitary portal circulation. There are other dopaminergic neurons that intermingle with the CRH neurons in this area [57]. The second subclass of CRH positive neurons in the PVN are dorso-laterally located magnocellular neurons sending oxytocin or vasopressin projections to the posterior pituitary through the internal zone of the median eminence. These neurons release peptides to the peripheral circulation [57]. The third subclass of CRH positive neurons in the PVN is found in the dorsal, medial–ventral and lateral parvocellular subdivisions. These autonomic neurons have projections to the brainstem and spinal cord. These neurons express CRH as well as other neuropeptides and are involved in regulating the physiological connection between the sympathetic nervous system and the adrenal medulla, or sympathoadrenal system [57]. Single-cell sequencing techniques can further help classify individual neurons and distinct neuronal populations in the PVN as well as other areas of the brain. By investigating the transcriptome of individual neurons, we can start to understand the overwhelming complexity of individual neurons that are the fundamental building blocks of neural networks.

1.11 Early Life Stress and its Effect on Brain Development

Postnatal brain and brain circuit development as well as maturation are regulated by genetics and environmental inputs via epigenetic mechanisms [58,59]. DNA methylation, histone acetylation, and noncoding RNAs such as microRNAs are all examples of epigenetic mechanisms that may be affected by environmental stimuli. A wide range of chromatin

modifications have been shown to be dynamically regulated in the developing and adult brain. During critical periods of neural circuit development, many are unstable and open to influences from the environment [59]. In humans and other mammals, early life adversity (ELA), which can be caused by maternal inputs such as poverty, parental loss, mistreatment, and neglect are associated with consequences to neural circuit development that can have lasting consequences including impairments in cognition and emotional regulation. In rodents and primates that experienced some form of ELA, hypersecretion of CRH has been detected years after the stress exposure period [59,60]. Genetic abnormalities or polymorphisms of proteins in the stress pathway have also been found to play a role in increased risk of adverse effects after ELA. For example, CRH type 1 receptor (CRHR1) has been shown to have polymorphisms that affect the adult phenotype after ELA [59]. Important genes within the stress pathway have been shown to be epigenetically regulated, suggesting that ELA can have lasting impacts on expression of these genes. For example, altered promoter methylation of GRs has been detected in subjects with a history of childhood abuse [59]. ELA does not only affect the HPA axis and CRH secretion, but also results in decreases in expression of hormones that are involved in social biology and bonding such as oxytocin as well as increases in inflammatory markers in patients [59].

1.12 Common Model Organisms and Techniques used to Study Early Life Adversity

Methods for studying ELA using animal models have been developed to mimic the effect of adverse environments on human brain development [61,62]. There are two common mammalian model organisms used to study the consequences of ELA on adult outcomes; non-human primates and rodents. The use of non-human primates to study ELA has distinct advantages. For example, non-human primates are similar socially and behaviorally to humans. One of the most notable differences between non-human primate models and rodent models is

the HPA axis similarity between humans and other primates [86]. In rodents, GR expression is low at birth, but starts to increase during development, whereas in the primate brain GR expression levels do not seem to change from birth to adulthood. The differences between rat, non-human primates, and human HPA axis is still not well understood. Similar to humans, non-human primates form strong and lasting mother-infant bonds. The primate mother provides nourishment, protection, social lessons, and regulates the infant's fear and stress responses [63,86]. In rodents, levels of maternal care, like feeding, licking and grooming regulate the pup's fear and stress responses. Although similar to primate interactions, there are differences in maternal interactions between rodents and primates.

Some examples of ELA models in non-human primates include peer-rearing versus maternal rearing, where infants are either raised in a nursery until postnatal day (PND) 30, then placed with similar aged peers for 6 months or raised in their mothers' care until 6 months when they are placed in similar peer-inhabited environments. The peer-reared infants show higher incidence of risk-taking and anxiety later in life [63]. Other examples of non-human primate models of ELA include maternal social subordination, infant maltreatment (similar to child abuse), and maternal separation [63]. Although there are clear advantages to studying ELA in non-human primates, the studies are time consuming and expensive due to their need to be performed in specialized facilities with experienced researchers due to the increased risk of physical harm to the researchers and pathogen exposure. Due to these drawbacks, rodent models of ELA are by far the most common model system used to study ELA. Rodents also have a very early sensitive period of brain development and relatively short gestational period, time to weaning, and well-established ELA models and stress tests used to study ELA and its effects on the adult stress pathway [64].

A well-characterized preclinical rodent model of ELA is the limited bedding and nesting (LBN) model [65]. The LBN model mimics an impoverished environment, inducing unpredictable maternal care, during a critical period of brain development [61,66,67]. It entails rodent mothers and their litters kept on a wire mesh with very limited bedding and about half of the bedding materials normally given. The LBN model does not affect the duration or number of maternal interactions with the pups. However, the quality of maternal care is disrupted, which is measured by its consistency. This unpredictable maternal care is due to the increase in the frequency that the mother leaves the nest area to search for bedding materials [65]. Maternal separation is another common technique used to study ELA (in both rodents and non-human primates) and has been shown to have similar effects on brain development as the LBN model. However maternal separation can only be used for short periods of time due to the risk of pup/infant malnutrition and hypothermia [63,64].

The ELA model environment is typically given between PND 2 and PND 9-10 for rodents because this has been shown to be the sensitive period [61]. The timing and duration of when the stress is applied to have the greatest impact on brain development has been extensively studied in rodents [64]. For instance, in rodents a 24-hour maternal separation between PND 3 to 4 leads to increased responsiveness to experiential stressors later in life, whereas a similar maternal separation performed just a few days later, between PND 7 to 8 or 11 to 12, results in decreased responsiveness to later stressors. Similar results have been found for learning and memory deficits due to rodent maternal separation. Separation on PND 3 causes active avoidance and a reaction to external stimuli, such as conditioned freezing task, impairment, whereas separation on PND 9 improves performance [64]. Taken together the sensitive period for brain development in rodents seems to be within PND 2-9.

Models of ELA induce enduring long-term consequences [61] including reduced reward seeking [68–71], impairments in hippocampal memory [65,72–75], and altered development of fear inhibition pathways [76]. Combined, these data suggest that there are complex and overlapping neural developmental processes early in life and adverse experiences during these sensitive periods of development may drastically influence brain development.

1.13 Methods and Techniques used for Studying ELA in Humans

Studies surrounding ELA in humans are typically observational, with no treatment and control groups due to moral implications. However, because of this the ELA is assessed on a spectrum of adversity, and not specifically in one group or another. Some typical studies conducted in humans include children who are exposed to poor or a low amount of care for long hours in early development, children exposed to severe deprivation (i.e. orphanages), maternal depression studies, poverty conditions, and studies of neglect or abuse [61,77]. These studies suggest that ELA during sensitive periods of development has lasting impacts on brain development, which leads to increases in incidences of neurological diseases, such as depression, anhedonia, and drug dependence, later in life.

Environmental signals that generate stress in the adult such as restraint are not stressful to a newborn. This is due to the immaturity of both the developing stress circuit and the brain. The developing processes in the brain during this period include axonal and dendritic growth, synaptic stabilization, and synaptic pruning [59,78]. This age-specific stress sensitivity has led to the idea of an increased sensitivity period in brain development [61,64]. In humans, by the end of the first year of life, events that used to elicit distress, wariness, and inhibition of approach in younger infants no longer produce elevations in cortisol, suggesting that the first 12 months of life account for the hypersensitive period of brain development in humans. This suggests a

disconnection between activation of the HPA axis on the adrenal level from behavioral distress and negative emotion [79]. Although moral implications of human studies shy away from perturbation experiments, it is important to find other non-invasive ways to study the effects of ELA on brain development in humans. Analysis of peripheral DNA methylation using blood cells is a non-invasive way to examine the epigenetic effects of ELA on brain development

1.14 Investigating Epigenetic Changes via DNA Methylation

Recent studies have started to look at molecular changes in CRH neurons such as alterations in CRH expression caused by stress [80,81]. However, depending on the timing and duration of the stress, CRH levels can either increase or decrease [77,78]. These changes in expression are most likely due to the maintenance of the long-lasting CRH gene repression or activation involving epigenetic mechanisms. For instance, increased DNA methylation at the CRH promoter region led to transcriptional repression in some studies [78], while others found that decreased DNA methylation at the CRH promoter caused no change in expression [82].

On the same note, gene expression responses to stress later in life can be predicted by maternal interactions with offspring during sensitive periods of brain development. Augmented maternal care, i.e. high rates of maternal licking and grooming in rodents, has been associated with reduced anxiety-like behaviors and increased negative feedback to the HPA axis via enhanced GR reactivity [83]. DNA methylation studies have corroborated these findings with methylation and expression changes in the GR-encoding gene Nr3c1. Similar findings in the human NR3C1 gene expression and promoter methylation have been shown following ELA [83]. The epigenetic mechanisms that govern changes in gene expression of stress-related genes during and after ELA remain to be fully understood.

Although it is already known that DNA methylation changes correlate with gene expression changes that occur as a result of ELA, it would be interesting to determine whether these DNA methylation changes could provide a useful epigenetic signature of ELA. A recent study determined that intra-individual DNA methylation pattern changes distinguish rats raised in ELA conditions from control rats [84]. These pattern changes between ELA and control may be able to predict ELA signatures in individuals and possibly be correlated to human experiences [84]. It would be interesting to investigate if a similar signature can be found in humans over a spectrum of ELA.

1.15 Thesis Overview

The central theme of my thesis is to understand developmental conservation of gene regulation during mammalian evolution as well as the developmental plasticity of epigenetic regulation. Using a combination of both experimental and computational methods, I examined epigenetic changes primarily at a single cell level.

In Chapter 2, I investigate integrative analysis of mRNA expression and open chromatin dynamics during lineage specification. We asked the question: how conserved are the transcriptome and chromatin dynamics of transcription factors involved in NPC and DE differentiation in human and rat? We hypothesize that a select group of conserved regulatory elements are deployed dynamically to regulate gene expression during the process of definitive endoderm (DE) and/or neural progenitor cell (NPC) differentiation. I study shared and cell-type specific transient differentiation markers from NPC (Fig. 1.1) and DE (Fig. 1.2) lineages in both human and rat. Specifically, I investigate the expression and open chromatin dynamics of key TFs that may regulate differentiation into each cell type and their candidate cis-regulatory elements (cCREs). I show that DE differentiation is more conserved in both cCREs and

chromatin accessibility than NPC differentiation. The similarity of early development of DE between species and the complexity of later formation of NPCs and species-specific developmental time differences may account for the differences in conservation seen between NPC and DE differentiations. The goal of this study was to quantify conserved and species-specific chromatin dynamics during embryonic stem cell (ESC) differentiation to DE and to NPCs. To achieve this we collected a daily time-course of gene expression and chromatin accessibility in rat and human for both cell types. Differentiation of homologous cell types in different species is controlled by conserved networks of regulatory elements driving gene expression. We provide the first global comparison of transcriptional complexity and chromatin dynamics between human and rat for DE and NPC differentiation. The information obtained from this study shows that rats are a good model organism for both neural and early developmental studies, but it is important to understand the differences on a genetic and epigenetic level between humans and the model organisms used to study basic and translational science.

In Chapter 3, we study single-cell transcriptomics of hypothalamic CRH-expressing neurons. We asked the question: how does early life adversity effect the transcriptome of the Crh positive PVN neurons in the hypothalamus? I use single-cell RNA sequencing (scRNA-seq) to capture the heterogeneity of cellular responses and define differential expression associated with ELA in the rodent hypothalamus. I discover interesting changes in gene expression profiles of stress-sensitive CRH-neurons in the PVN following ELA. This study utilizes an animal model, the mouse limited bedding and nesting (LBN) model, of an impoverished environment with unpredictable maternal care which causes negative cognitive and emotional outcomes. We try to understand single-cell transcriptomic changes that occur as a result of ELA and how these changes influence the development of stress-sensitive CRH-neurons in the PVN. We used single cell methods in this study to highlight heterogeneity within the CRH positive PVN neuron

population between the ELA and control condition. It is known that both genetic and environmental factors contribute to the growth and maturation of neurons and synapses. However, the mechanisms by which ELA alters epigenetic signatures within an individual CRH-neuron, and the long-term functional consequences of these changes, remain unknown. Such changes in the epigenetic programs of individual stress-sensitive neurons suffice to induce long-lasting changes within these cells, influencing gene expression. By conducting differential expression analysis between the ELA and control conditions we were able to determine that stress-sensitive neurons from the ELA condition may be maturing faster and increasing their energy production due to increased synaptic input. These changes seem to only affect the excitatory glutamatergic neurons and not neurons that express GABAergic neurotransmitters. Disfunction of the stress circuit following ELA can be at least partially explained by CRH positive PVN neurons from the ELA condition maturing faster with altered energy production compared to cells from the control condition. Since the population of CRH positive neurons of the PVN are made up of mostly glutamatergic neurons, it is not surprising that we did not find significant differences between the ELA and control conditions in GABAergic neurons.

In Chapter 4, we compare Intra-individual peripheral methylation signatures of diverse early life experiences. We hypothesize that intra-individual methylome signatures show transcription-driven alterations of cellular growth and function in a spectrum of human early life adversity. I show that a DNA signature of ELA can be used to distinguish individual children with a spectrum of early life experiences. I use an intra-individual approach, using a similar computational method as previously done in rats, to test “delta” methylation, which is the change in methylation that occurs between birth and the first year of life, in each of the study groups of individuals over a spectrum of ELA experiences. Using this method, I found striking differences that correlate with age and experience. During the first year of life, in humans, the brain is more sensitive to influences by environmental inputs. ELA is one of the largest

environmental risk factors for the development of abnormal brain circuits and psychiatric diseases. These later pathologies are at least partially encoded by underlying neural epigenetic pathways established during ELA. How these epigenetic pathways are encoded and if there are epigenetic signatures that can be used to predict potential adverse effects of ELA remains unknown. In this study, we employ a non-invasive buccal swab collection method to investigate systemic DNA methylation changes that occur even outside the brain to study ELA in humans over a diverse range of experiences. In the future, we will apply machine learning and other computational techniques to determine an epigenetic signature that can distinguish the risk of developing psychiatric disorders after adverse ELA experiences.

1.16 Figures

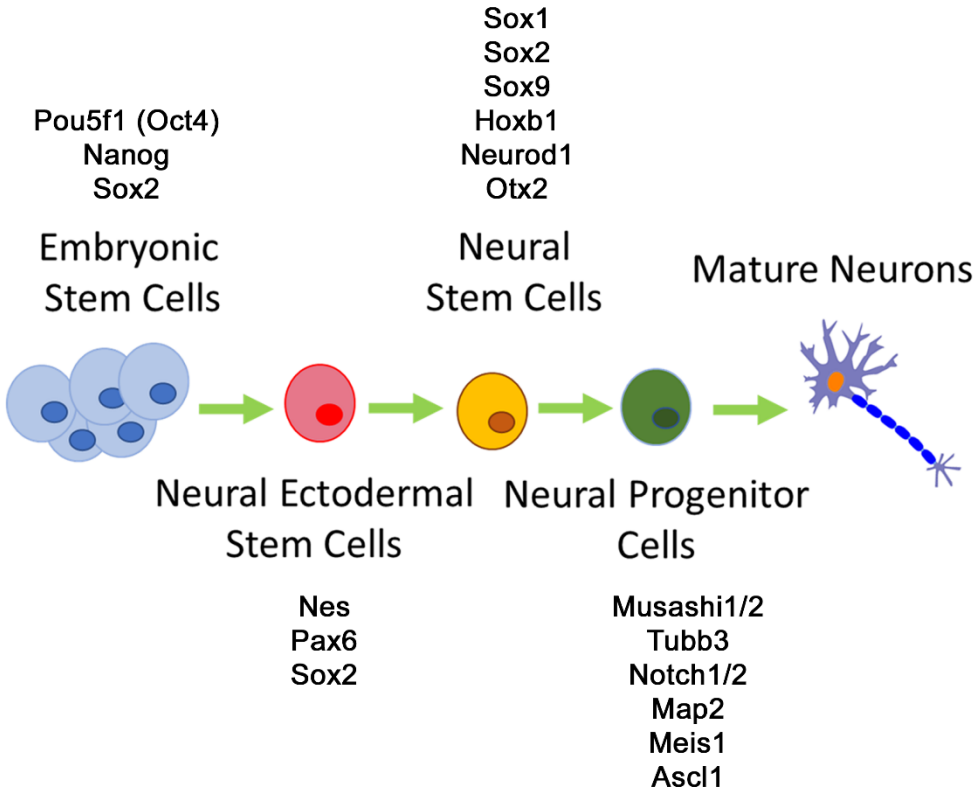

Figure 1.1 Neural Progenitor Cell Differentiation *in vitro*

Different stages of neural progenitor cell differentiation in both rat and human and the markers defining each stage.

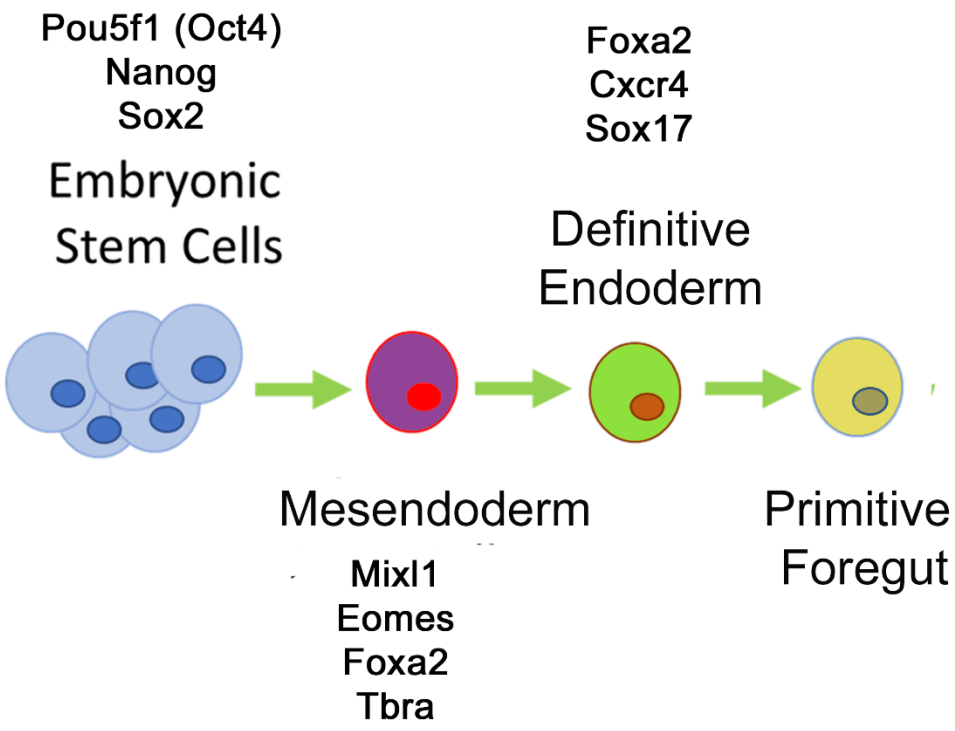

Figure 1.2 Definitive Endoderm Differentiation *in vitro*

Different stages of definitive endoderm differentiation in both rat and human and the markers defining each stage.

1.20 References

1. Fu, J., Warmflash, A. & Lutolf, M. P. Stem-cell-based embryo models for fundamental research and translation. *Nat. Mater.* **20**, 132–144 (2021).

2. Molnar, C. & Gair, J. *Concepts of Biology: First Canadian Edition*. (2013).

3. Benitah, S. A. & Frye, M. Stem cells in ectodermal development. *J. Mol. Med.* **90**, 783–790 (2012).

4. Mehta, R. H. Sourcing human embryos for embryonic stem cell lines: problems & perspectives. *Indian J. Med. Res.* **140 Suppl**, S106–11 (2014).

5. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).

6. Nishikawa, S.-I., Jakt, L. M. & Era, T. Embryonic stem-cell culture as a tool for developmental cell biology. *Nat. Rev. Mol. Cell Biol.* **8**, 502–507 (2007).

7. Budday, S., Steinmann, P. & Kuhl, E. Physical biology of human brain development. *Front. Cell. Neurosci.* **9**, 257 (2015).

8. Ackerman, S. *Discovering the Brain*. (National Academies Press, 1992).

9. Shonkoff, J. P. & Phillips, D. A. *From Neurons to Neighborhoods: The Science of Early Childhood Development*. (National Academies Press (US), 2014).

10. Kinsella, M. T. & Monk, C. Impact of maternal stress, depression and anxiety on fetal neurobehavioral development. *Clin. Obstet. Gynecol.* **52**, 425–440 (2009).

11. Morgane, P. J. *et al.* Prenatal malnutrition and development of the brain. *Neurosci. Biobehav. Rev.* **17**, 91–128 (1993).

12. Semple, B. D., Blomgren, K., Gimlin, K., Ferriero, D. M. & Noble-Haeusslein, L. J. Brain development in rodents and humans: Identifying benchmarks of maturation and vulnerability to injury across species. *Prog. Neurobiol.* **106-107**, 1–16 (2013).

13. Perea-Gomez, A. & Meilhac, S. M. Formation of the Anterior-Posterior Axis in Mammals. *Principles of Developmental Genetics* 171–188 (2015) doi:10.1016/b978-0-12-405945-0.00010-7.

14. Grubb, B. J. Developmental Biology, Eighth Edition. Scott F. Gilbert, editor. *Integrative and Comparative Biology* vol. 46 652–653 (2006).

15. Thomson, M. *et al.* Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889 (2011).

16. Yamamoto, Y. *et al.* Random migration of induced pluripotent stem cell-derived human gastrulation-stage mesendoderm. *PLoS One* **13**, e0201960 (2018).

17. Wei, S. & Wang, Q. Molecular regulation of Nodal signaling during mesendoderm formation. *Acta Biochim. Biophys. Sin.* **50**, 74–81 (2018).

18. Muhr, J. & Ackerman, K. M. Embryology, Gastrulation. in *StatPearls* (StatPearls Publishing, 2020).

19. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266 (1993).

20. Rivera-Pérez, J. A., Jones, V. & Tam, P. P. L. Culture of Whole Mouse Embryos at Early Postimplantation to Organogenesis Stages. *Methods in Enzymology* 185–203 (2010) doi:10.1016/s0076-6879(10)76011-0.

21. Ansari, A. & Pillarisetty, L. S. Embryology, Ectoderm. in *StatPearls* (StatPearls Publishing, 2020).

22. Pankratz, M. T. *et al.* Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. *Stem Cells* **25**, 1511–1520 (2007).

23. Alsanie, W. F. *et al.* Specification of murine ground state pluripotent stem cells to regional neuronal populations. *Sci. Rep.* **7**, 16001 (2017).

24. Purves, D. *et al. Neuroscience*. (2018).

25. Patten, I. & Placzek*, M. The role of Sonic hedgehog in neural tube patterning. *Cellular and Molecular Life Sciences* vol. 57 1695–1708 (2000).

26. Sathananthan, A. H. Neural stem cells in neurospheres, embryoid bodies, and central nervous system of human embryos. *Microsc. Microanal.* **17**, 520–527 (2011).

27. Ulloa, F. & Martí, E. Wnt won the war: antagonistic role of Wnt over Shh controls dorso-ventral patterning of the vertebrate neural tube. *Dev. Dyn.* **239**, 69–76 (2010).

28. Hatakeyama, J. *et al.* Hes genes regulate size, shape and histogenesis of the nervous system by control of the timing of neural stem cell differentiation. *Development* **131**, 5539–5550 (2004).

29. Bronner, M. E. Formation and migration of neural crest cells in the vertebrate embryo. *Histochemistry and Cell Biology* vol. 138 179–186 (2012).

30. Vincent, P. H., Benedikz, E., Uhlén, P., Hovatta, O. & Sundström, E. Expression of Pluripotency Markers in Nonpluripotent Human Neural Stem and Progenitor Cells. *Stem Cells Dev.* **26**, 876–887 (2017).

31. Topol, A., Tran, N. N. & Brennand, K. J. A Guide to Generating and Using hiPSC Derived NPCs for the Study of Neurological Diseases. *Journal of Visualized Experiments* (2015) doi:10.3791/52495.

32. Guillaume, D. J. & Zhang, S.-C. Human embryonic stem cells: a potential source of transplantable neural progenitor cells. *Neurosurg. Focus* **24**, E3 (2008).

33. Martínez-Cerdeño, V. & Noctor, S. C. Neural Progenitor Cell Terminology. *Front. Neuroanat.* **12**, 104 (2018).

34. Ellenbroek, B. & Youn, J. Rodent models in neuroscience research: is it a rat race? *Dis. Model. Mech.* **9**, 1079–1087 (2016).

35. Suckow, M. A., Claire Hankenson, F., Wilson, R. P. & Foley, P. L. *The Laboratory Rat*. (Academic Press, 2019).

36. Homberg, J. R., Wöhr, M. & Alenina, N. Comeback of the Rat in Biomedical Research. *ACS Chem. Neurosci.* **8**, 900–903 (2017).

37. Brenowitz, E. A. & Zakon, H. H. Emerging from the bottleneck: benefits of the comparative approach to modern neuroscience. *Trends Neurosci.* **38**, 273–278 (2015).

38. Snyder, J. S. *et al.* Adult-born hippocampal neurons are more numerous, faster maturing, and more involved in behavior in rats than in mice. *J. Neurosci.* **29**, 14484–14495 (2009).

39. Clancy, B., Finlay, B. L., Darlington, R. B. & Anand, K. J. S. Extrapolating brain development from experimental species to humans. *Neurotoxicology* **28**, 931–937 (2007).

40. Andersen, M. D. *et al.* Animal Models of Fetal Medicine and Obstetrics. *Experimental Animal Models of Human Diseases - An Effective Therapeutic Strategy* (2018) doi:10.5772/intechopen.74038.

41. Chini, M. & Hanganu-Opatz, I. L. Prefrontal Cortex Development in Health and Disease: Lessons from Rodents and Humans. *Trends Neurosci.* **44**, 227–240 (2021).

42. Kim, D.-J. *et al.* Longer gestation is associated with more efficient brain networks in preadolescent children. *Neuroimage* **100**, 619–627 (2014).

43. Dundes, C. E. & Loh, K. M. Bridging naïve and primed pluripotency. *Nature cell biology* vol. 22 513–515 (2020).

44. Takahashi, S., Kobayashi, S. & Hiratani, I. Epigenetic differences between naïve and primed pluripotent stem cells. *Cell. Mol. Life Sci.* **75**, 1191–1203 (2018).

45. Hong, J. *et al.* A focused microarray for screening rat embryonic stem cell lines. *Stem Cells Dev.* **22**, 431–443 (2013).

46. Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).

47. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4936–4942 (2005).

48. Hardison, R. C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* **13**, 469–483 (2012).

49. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).

50. Stergachis, A. B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365–370 (2014).

51. Fujiwara, N. & Cave, J. W. Partial Conservation between Mice and Humans in Olfactory Bulb Interneuron Transcription Factor Codes. *Front. Neurosci.* **10**, 337 (2016).

52. Tau, G. Z. & Peterson, B. S. Normal development of brain circuits. *Neuropsychopharmacology* **35**, 147–168 (2010).

53. Qin, C., Li, J. & Tang, K. The Paraventricular Nucleus of the Hypothalamus: Development, Function, and Human Diseases. *Endocrinology* **159**, 3458–3472 (2018).

54. Deng, Q. *et al.* Rapid Glucocorticoid Feedback Inhibition of ACTH Secretion Involves Ligand-Dependent Membrane Association of Glucocorticoid Receptors. *Endocrinology* **156**, 3215–3227 (2015).

55. Landgraf, R. & Neumann, I. Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. *Front. Neuroendocrinol.* **25**, 150–176 (2004).

56. Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370**, (2020).

57. Aguilera, G. & Liu, Y. The molecular physiology of CRH neurons. *Front. Neuroendocrinol.* **33**, 67–84 (2012).

58. Yasui, D., Peedicayil, J. & Grayson, D. R. *Neuropsychiatric Disorders and Epigenetics*. (Academic Press, 2016).

59. Bale, T. L. *et al.* Early life programming and neurodevelopmental disorders. *Biol. Psychiatry* **68**, 314–319 (2010).

60. Heim, C. & Nemeroff, C. B. The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol. Psychiatry* **49**, 1023–1039 (2001).

61. Walker, C.-D. *et al.* Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* **20**, 421–448 (2017).

62. Nishi, M., Horii-Hayashi, N. & Sasagawa, T. Effects of early life adverse experiences on the brain: implications from maternal separation models in rodents. *Front. Neurosci.* **8**, 166 (2014).

63. Wakeford, A. G. P., Morin, E. L., Bramlett, S. N., Howell, L. L. & Sanchez, M. M. A review of nonhuman primate models of early life stress and adolescent drug abuse. *Neurobiol Stress* **9**, 188–198 (2018).

64. Chen, Y. & Baram, T. Z. Toward Understanding How Early-Life Stress Reprograms Cognitive and Emotional Brain Networks. *Neuropsychopharmacology* **41**, 197–206 (2016).

65. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* **149**, 4892–4900 (2008).

66. Molet, J., Maras, P. M., Avishai-Eliner, S. & Baram, T. Z. Naturalistic rodent models of chronic early-life stress. *Dev. Psychobiol.* **56**, 1675–1688 (2014).

67. Ivy, A. S., Brunson, K. L., Sandman, C. & Baram, T. Z. Dysfunctional nurturing behavior in rat dams with limited access to nesting material: a clinically relevant model for early-life stress. *Neuroscience* **154**, 1132–1142 (2008).

68. Bolton, J. L. *et al.* Anhedonia Following Early-Life Adversity Involves Aberrant Interaction of Reward and Anxiety Circuits and Is Reversed by Partial Silencing of Amygdala Corticotropin-Releasing Hormone Gene. *Biol. Psychiatry* **83**, 137–147 (2018).

69. Levis, S. C. *et al.* On the early life origins of vulnerability to opioid addiction. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0628-5.

70. Bolton, J. L. *et al.* Early-life adversity facilitates acquisition of cocaine self-administration and induces persistent anhedonia. *Neurobiol Stress* **8**, 57–67 (2018).

71. Molet, J. *et al.* Fragmentation and high entropy of neonatal experience predict adolescent emotional outcome. *Transl. Psychiatry* **6**, e702 (2016).

72. Naninck, E. F. G. *et al.* Chronic early life stress alters developmental and adult neurogenesis and impairs cognitive function in mice. *Hippocampus* **25**, 309–328 (2015).

73. Bath, K. G. *et al.* Early life stress leads to developmental and sex selective effects on performance in a novel object placement task. *Neurobiol Stress* **7**, 57–67 (2017).

74. Brunson, K. L. *et al.* Mechanisms of late-onset cognitive decline after early-life stress. *J. Neurosci.* **25**, 9328–9338 (2005).

75. Molet, J. *et al.* MRI uncovers disrupted hippocampal microstructure that underlies memory impairments after early-life adversity. *Hippocampus* **26**, 1618–1632 (2016).

76. Bath, K., Manzano-Nieves, G. & Goodwill, H. Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Horm. Behav.* **82**, 64–71 (2016).

77. Lupien, S. J., McEwen, B. S., Gunnar, M. R. & Heim, C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* **10**, 434–445 (2009).

78. Regev, L. & Baram, T. Z. Corticotropin releasing factor in neuroplasticity. *Front. Neuroendocrinol.* **35**, 171–179 (2014).

79. Gunnar, M. R. & Donzella, B. Social regulation of the cortisol levels in early human development. *Psychoneuroendocrinology* **27**, 199–220 (2002).

80. Baram, T. Z. & Hatalski, C. G. Neuropeptide-mediated excitability: a key triggering mechanism for seizure generation in the developing brain. *Trends Neurosci.* **21**, 471–476 (1998).

81. Liu, Y. *et al.* Salt-inducible kinase is involved in the regulation of corticotropin-releasing hormone transcription in hypothalamic neurons in rats. *Endocrinology* **153**, 223–233 (2012).

82. McClelland, S., Korosi, A., Cope, J., Ivy, A. & Baram, T. Z. Emerging roles of epigenetic mechanisms in the enduring effects of early-life stress and experience on learning and memory. *Neurobiol. Learn. Mem.* **96**, 79–88 (2011).

83. Bale, T. L. Epigenetic and transgenerational reprogramming of brain development. *Nat. Rev. Neurosci.* **16**, 332–344 (2015).

84. Jiang, S. *et al.* Intra-individual methylomics detects the impact of early-life adversity. *Life Sci Alliance* **2**, (2019).

85. Zhang J, Jiao J. Molecular Biomarkers for Embryonic and Adult Neural Stem Cell and Neurogenesis. *Biomed Res Int*. **2015**,727542. (2015) doi: 10.1155/2015/727542. Epub 2015 Sep 1. PMID: 26421301; PMCID: PMC4569757.

86. Sanchez MM. The impact of early adverse care on HPA axis development: nonhuman primate models. *Horm Behav*. Nov;50(**4**):623-31 (2006). doi: 10.1016/j.yhbeh.2006.06.012. Epub 2006 Aug 15. PMID: 16914153.

Chapter 2

Comparative Chromatin Dynamics of Stem Cell Differentiation in Human and Rat

Notes:

(1) Dr. Shan Jiang and I equally contributed to the material in this chapter. She designed and performed Definitive Endoderm experiments, as well as providing important suggestions on interpreting the results. I designed and performed Neural Progenitor Cell experiments. I also analyzed the data.

(2) Yuka Roxas performed cell maintenance support for embryonic stem cells and some differentiation into neuronal lineages

(3) Cassandra McGill performed quality checks and sequenced experimental samples.

(4) Savanna Ma performed quality checks and sequenced experimental samples.

(5) Dr.Ali Mortazavi conceived the idea and provided continued support and guidance throughout the project.

Chapter 2

Comparative Chromatin Dynamics of Stem Cell Differentiation in Human and Rat

2.1 Abstract

Differentiation of cell types homologous between species are controlled by conserved networks of regulatory elements driving gene expression. In order to identify conservation of gene expression and chromatin accessibility during cell differentiation in mammals we collected a daily time-course of gene expression and chromatin accessibility in rat and human during embryonic stem cell differentiation to definitive endoderm (DE) as well as to neural progenitor cells (NPCs). We identify shared and cell-type specific transient differentiation markers in each species, including key transcription factors that may regulate differentiation into each cell-type and their candidate cis-regulatory elements (cCREs) including genes such as *GBX2, EGR1,* and *PBX1* for NPC differnetiation and *ETV1, GBX2,* and *TGIF1* for DE differentiation. We found 8% of candidate cis-regulatory elements with significant transcriptional overlaps are shown to be conserved in NPC differentiation whereas 22% of significantly overlapping candidate cis-regulatory are shown to be conserved in DE differentiation between rat and human (Figure 2.7).. We provide the first global comparison of transcriptional complexity and chromatin dynamics between human and rat for DE and NPC differentiation.

2.2 Introduction

A fundamental question in biology is how lineage specification is regulated by a large cohort of transcription factors (TFs) and signaling molecules. Lineage specification and differentiation are driven, at least in part, by chromatin remodeling causing changes in the accessibility of TF binding sites near promoter and enhancer cis-regulatory modules (CRMs) of other TFs, which collectively form Gene Regulatory Networks (GRNs) [1,2]. Changes in

expression of key TFs are crucial in the regulation of differentiation. The identification of CRMs that are required for the regulation of gene expression are important for broadening our understanding of development and the conservation or divergence of gene regulation between species and cell lineages [3].

Previous functional genomics studies have shown that binding of stage-specific TFs to *cis*-regulatory elements (cREs) controls precise spatiotemporal gene expression. Previous research on *cis*-regulatory element gain or loss (i.e. turnover) has historically been on specific TFs or histone modifications, rather than genome-wide surveys of species-specific or conserved TF-DNA interactions, preventing the construction of comprehensive GRNs [4]. Furthermore, TF-DNA interactions can be investigated using cRE assays such as Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) for TF footprints. TF occupancy near a gene combined with the gene's expression reveals a regulatory connection which forms the basis of GRNs [5]. While individual TF to gene interactions may be conserved weakly, GRNs are highly conserved across species, which makes the comparison of GRNs important for studying evolutionary differences between species [6].

Mammalian embryonic stem cells (ESCs) are defined by their unique ability to self-renew and to generate all lineages of the organism. ESCs divide asymmetrically, producing a daughter cell with more limited differentiation properties [7]. The formation of the three germ layers (ectoderm, mesoderm, and endoderm) during gastrulation is one of the most important developmental processes in the life of an organism.. This differentiation process can be mimicked *in vitro* with ESCs [7]. The specification of the germ layers is accomplished through the activation of lineage-specific GRNs [7,8]. In this study we compare the differentiation of rat and human ESCs into ectodermal NPCs and DE using a time-course of RNA-seq and ATAC-seq in

order to determine shared and species-specific evolutionary and regulatory mechanisms that drive development of these two lineages between rodents and humans.

In the past decade, several groups have established protocols from human ESCs to efficiently differentiate into pancreatic and hepatic cells *in vitro* and to characterize changes in gene expression during the differentiation process [9–11]. However, these studies have primarily focused on later stages of differentiation and the generation of functional cells, thus leaving the initiation of endoderm formation understudied. Commonly used rodent models such as mouse and rat have proven surprisingly challenging for robust endoderm layer generation *in vitro* [12–14]. There have been several temporal *in vitro* studies of human ESC to NPC differentiation assessing gene regulatory and cRE dynamics. Some of these stages include differentiation to postmitotic neurons [16] and early neural differentiation [17]. Techniques used to study these processes include in-vitro methods such as cerebral organoids [15], primary cultures of human NPCs and the developing brain [18], and stage-specific transcriptional networks in iPSC neurodevelopment [19,20]. Evolutionary studies have also compared gene expression programs between species such as chimpanzee versus human [17], or mouse versus chimpanzee [18], both in NPCs and in DE differentiation[18]. However, the temporal dynamics of NPC and DE differentiation lineages have not been compared in humans, and the dynamics of rat and human have not been compared for either lineage. Methods for *in vitro* monolayer differentiation of human ESCs into both DE [9,10,12–14,21,22], and NPCs [23–27] are available. While multiple protocols for mouse differentiation of ESCs to DE [9,10,12–14,21,22] and NPCs [23–27] have been published, the differentiation of rat ESCs into either lineage has not been previously studied using functional genomics.

Our hypothesis is that a select group of conserved regulatory elements are deployed dynamically to regulate gene expression during the process of DE and/or NPC differentiation. In

order to quantify the extent of conservation of GRNs between rats and humans during lineage specification, we differentiated ESCs into NPC and DE. In order to accomplish this, we first adapted mouse differentiation protocols for rat. We then collected daily time-courses of differentiation for each cell type and species using RNA-seq and ATAC-seq to measure gene expression and chromatin accessibility changes, respectively. We then compared the conservation level of both gene expression and chromatin accessibility in the two species. We compared the RNA-seq and ATAC-seq time-courses between the different lineages in rat, then in human to quantify the dynamics of the transient expressed genes, transcription factor motif enrichment, and to determine their shared function using gene ontology. We then compared human and rat gene modules in each cell type to identify conserved as well as species-specific TF-to-gene connections. Finally, we compared these modules to each other to determine the extent of evolutionarily conserved regulation between the species.

2.3 Results

2.3.1 Generation of Definitive Endoderm and Neural Progenitor Cells in Two Species

We induced the differentiation of rat and human embryonic stem cells (ESCs) into ectodermal NPCs and DE to quantify the transcriptional and candidate *cis*-regulatory element (cCRE) dynamics that drive cellular commitment. Previously developed methods for *in vitro* monolayer differentiation of human ESCs into both DE and NPCs were used (Figure 2.1A, Figure S2.1A) [28,29]. We adapted previously published ESC mouse differentiation protocols for NPC to rat (Figure 2.1A, Figure S2.1A), which have not been previously described [12–14,23]. For DE differentiation, the protocols for both rat and human rely on the activation of Nodal signaling to induce the expression of TFs and target genes in DE formation. Protocols add exogenous Activin A (Nodal analog) to mimic high concentrations of Nodal in order to produce DE cells with

expression of key marker genes like *MIXL1, CER1, SOX17, FOXA2*, and *CXCR4* [21,30–33]. For NPC differentiation, the protocols for rat and human both rely on Smad2/3 inhibition or the inhibition of the TGF-β pathway[23,28]. However, in human NPC differentiation dual Smad inhibition was used, whereas only the ALK2 and ALK3 inhibitor LDN193189 was used in rat (see methods). Rat NPC differentiation also relies on Fgf2 to further differentiate the cells into a neuronal cell fate [23,34]. Terminal NPC differentiation was reached by day 8 in both rat and human (Figure 2.1A and Figure S2.1A), and terminal DE differentiation was reached by day 7 in rat and day 5 in human (Figure 2.1B, Figure S2.1B). During each differentiation time-course, morphology changed from colony-like to the formation of neuro-rosettes for NPC differentiation (Figure S2.1B and 2.1C) [35] and pebble/cobblestone morphology for DE differentiation (Figure S2.1D and 2.1E) [36]. We observed morphology changes as early as day 1 of differentiation for both cell-types (Figure S2.1B-E). Thus, the rat cells changed from their ESC morphology to the expected cell-type specific morphology over the course of NPC and DE differentiation.

Differentiation of both DE and NPC in both species requires the downregulation of key pluripotent genes, *OCT4* and *NANOG* (Figure 2.1B). However, lingering expression of pluripotency markers has been shown in humans, but does not occur in rodents [37,38]. In order to examine differentiation on a global level for these lineages we checked RNA-seq timepoints daily during differentiation and found that key endodermal markers (*CXCR4, MIXL1, FOXA2, EOMES, GSC, GATA4, SOX17,* and *GATA6*) are activated in both species during DE differentiation and key neuronal markers (*NGFR* [39], *NEFM* [40], *NCAM1* [41], *NES, PAX6, MAP2, ASCL1* [42], and *MEIS1* [43] are activated during NPC differentiation (Figure 2.1B). Interestingly, we found that *FOXA1* and *SOX7* were differentially expressed in rat but not human DE differentiation. Overall, the increased expression of the lineage specific markers in each species and the decrease in expression of ESC markers suggest successful differentiation into each of the lineages in both species.

2.3.2 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC and DE in Rat

We next focused on the comparison of expression and cCRE dynamics between NPC and DE in rat to identify candidate regulators for the differentiation of each cell type. We collected a daily time-course of differentiation in duplicate for RNA-seq and ATAC-seq for a total of 32 RNA-seq and 32 ATAC-seq datasets in rat. We detected 13,112 genes expressed over 1 transcript per million (TPM) in two or more samples across both NPC and DE differentiation (Table 2.1). We used maSigPro [44] to identify 10,846 differentially expressed genes that form 18 clusters of distinct expression profiles across time (Figure 2.2A, Table 2.2). Five clusters, comprising 2,859 genes (26%), decrease over time in either one or both differentiation time-courses. Eight clusters comprising 4,323 genes (40%) increase during NPC differentiation (labeled in blue). Five clusters comprising 3,664 genes (34%) increase during DE differentiation (labeled in green). We plotted the correlation of the expression profiles over each differentiation time-course (Figure S2.2A) as well as plotted the samples on UMAP (Figure S2.2B). UMAP1 separates samples by differentiation time and UMAP2 represents the difference between the two differentiation time-courses. There are larger changes in the differentially expressed genes later in each time-course (Figure S2.2A and S2.2B), suggesting a transition between early and late differentiation about mid-way through each time-course (Figure S2.2B and S2.2C). We then focused on the TFs expressed in each of the maSigPro clusters, as they may play a role in the regulation of differentiation into each cell type. As expected, TFs that showed increased expression in NPC clusters include those important in neuronal development and overall expression of early neurons such as *Mef2a* (RR8), *Mef2c* (RR11), and *Tead2* (RR1) (Figure 2.2A)[45]; [46,47]; [48]. Similarly, the TFs in clusters that increase during DE differentiation include important markers of DE development and differentiation such as *Eomes* (RR9), *Foxa1* (RR9), and *Lhx1* (RR9). As expected, the five tan clusters with decreased expression over the time-

course include important ESC TFs such as *Oct4* (RR7), *Nanog* (RR12), and *Sox2* (RR18). The TFs found in cell-type specific clusters in rat suggest that they are expressed in the appropriate pattern to regulate differentiation into NPC and DE.

We used gene ontology (GO) to analyze clusters higher in NPCs (blue) and DE (green) (Figure 2.2B) to investigate functional roles of TFs which increased in expression in a cell-type specific manner. Cluster RR11, which sharply increases in expression after day 6 of the NPC time course, is enriched for GO terms such as forebrain development ($p < 6x10^{-12}$) and includes important TFs such as *Pax6* and *Mef2c* (Figure 2.2B). Cluster RR9, which increases in expression after day 3 of the DE time-course, is enriched for GO terms such as digestive tract morphogenesis ($p < 2x10^{-9}$) and includes TFs like *Pitx2, Sox17,* and *Gata6* (Figure 2.2B). We also performed a GO analysis of the five tan clusters of ESC high genes (Figure S2.3A). GO analysis of differentially expressed genes clustered by expression profile reveals that genes with shared functions are co-expressed during NPC and DE differentiation along with key TFs.

Next, we investigated changes in chromatin accessibility in the differentiation time-courses using ATAC-seq by clustering differential cCREs as previously described for gene expression (Figure 2.2C, Table 2.3). For genes with more than one associated cCRE, each cCRE is analyzed independently, therefore a gene can have cCREs that are in different clusters. We detected 51,417 accessible chromatin regions in two or more replicates across both NPC and DE differentiation (Table 2.1). Of these, 25,665 differential cCREs grouped into 24 clusters of distinct patterns across time. We observe a decrease in accessibility during differentiation in 9,375 (37%) of cCRE comprising nine clusters (labeled in tan, Figure 2.2C). Clusters that increased in accessibility in the NPC time-course (labeled in blue) accounted for 5,124 (20%) of differential cCRE. Clusters that increased in accessibility in the DE time-course (labeled in green) accounted for 8,383 (33%) of cCRE. Clusters with shared changes in both

time-courses (black) accounted for 2,762 of genes (10%) of cRE. Similar to the gene expression analysis, we calculated the correlation of the cCRE profiles over each differentiation time-course (Figure S2.2C) as well as plotted the samples on UMAP (Figure S2.2E). UMAP1 represents time and UMAP2 represents the difference between NPC and DE differentiation. Both the correlation plot and UMAP suggest there are substantial changes in accessibility during differentiation (Figure S2.2B).

We then focused on the accessible regions surrounding TF genes that are also differentially expressed (Figure 2.2C). TFs found to be differential in both expression and accessibility profiles may be key in the regulation of differentiation for their respective cell types. Concordant cCREs have an accessibility profile that matches their gene expression profile in the specific cell-type. Discordant cCREs have an accessibility profile that does not match the gene expression profile. Examples of TFs found with concordant differential expression and chromatin accessibility include *Mef2a*, which increases in expression in NPC (RR8) and has one concordant cCRE (RA18) (Fig 2C). *Mef2a* has been shown to regulate neuron differentiation [49]. Another TF that shows an increase in accessibility in NPCs is *bhlhe40* (RR2) (Figure 2.2C). There are three differential cCRE surrounding *Bhele40*, which is needed for proper neuronal function [50]. Each *Bhlhe40* cCRE has a different accessibility profile, one of which is concordant (RA7) (Figure 2.2C). A TF that shows an increase in accessibility in DE is *Eomes* (RR9), which is essential for DE specification (Figure 2.2C) [31]. There are five differential cCREs surrounding *Eomes*, four of which are concordant (Figure 2.2C). Of the differentially expressed TFs, 63% have at least one cCRE that is concordant to its expression profile. This suggests that cell-type specific TF expression in rat is accompanied by differential accessibility of cCREs in a concordant direction, which we analyze in greater detail below.

To explore functional roles of cell-type specific TF cCREs we used GO for clusters increasing in NPCs (blue) and the DE (green, Figure 2.2D). Cluster RA7, which increases in accessibility after day 1 in NPC differentiation, is enriched for GO terms like brain development ($p < 2x10^{-9}$) and includes regions around important TFs like *Gbx2, Notch1,* and *Rara* (Figure 2.2D). Cluster RA17, which increases in accessibility after day 2 in NPC differentiation, is enriched for GO terms including brain development ($p < 4x10^{-14}$). Cluster RA19, which is highest in accessibility at day 4 of DE differentiation, is enriched for GO terms including tube formation ($p < 1x10^{-11}$) and includes regions near *Gata3*. We also analyzed GO terms from each ESC-like cCRE cluster (Figure S2.3B).

We investigated which TF motifs are found in differentially accessible regions in each cluster using Homer (Figure S2.3C, S2.3D) (see methods). Motifs found in ESC-like clusters include Sox2, Pou5f1 (Oct4), and Nanog. DE cluster motif enrichments include Tead2, Foxa1, and Eomes (Figure S2.3C). Many motifs were enriched in both cell-types. The Eomes motif is enriched in eight clusters (all $p < 5x10^{-19}$), three of which are in the DE group. NPC cluster motif enrichments include Tead4, Sox2, and Maz. As expected, we find the Sox2 motif in nine clusters (all $p < 5x10^{-26}$), including the ESC and NPC groups, suggesting Sox2 is required for regulation of both ESC maintenance and NPC differentiation [51]. The Tcf12 motif was enriched in 3 clusters (all $p < 2x10^{-18}$) and has been shown to be important in NPC differentiation [52,53]. In conclusion, GO term analysis paired with *de novo* motif enrichment highlights TFs with binding sites in cell-type and time-course specific cCREs.

To mine clusters for differential gene expression and cCRE interactions we analyzed the relationship between differential RNA and ATAC clusters in rat. We used the overlaps to build a contingency table that contains the association of each of the differentially accessible regions with the mRNA clusters (see methods, Table 2.10). We detected 55 sets of significantly

enriched interactions between ATAC and RNA clusters. Of these, 22 have concordant significantly enriched interactions with both the expression and accessibility higher in the same cell-type (NPC (blue), DE (green), or ESC (tan)). 29 are discordant significantly enriched interactions (i.e. expression is high in NPC and accessibility is high in DE), and an additional four significant interactions in which the accessibility is not associated with any cell-type (black). We focused on cCREs surrounding TFs in each significantly overlapping cluster and found 231 instances of cCREs for TFs (totaling 109 TFs). Some of the TFs that have a shared affiliation with NPC clusters include *Mef2a* and *Bhlhe40*. Some of the TFs with shared DE affiliations include *T* and *Foxb1*. TFs in ESC affiliated clusters are *Tcf7l1*, *Nanog*, and *Foxp1*. All of the TFs listed above have increased expression and promoter accessibility in their respective cell types and probably play an important role in the regulation of differentiation in rat. We investigated the TFs that have significant overlapping clusters from different groups (i.e. increased expression in NPC and increased accessibility in DE) between DE and NPC clusters such as *Tgif1*, *Notch1*, and *Jun*. Some of these genes have important roles in the regulation of differentiation into one or both cell-types in human [54–57], but have not been previously shown in rat. These TFs either have increased expression, or increased promoter accessibility, in one of the time-courses suggesting that these TFs are candidate regulatory genes for at least one of the differentiation time-courses in rat.

We then mined the accessible regions with significant overlap (from Figure 2.2E) for TF footprints (see methods, Figure 2.2F). The most enriched footprints are Klf, Sp1, and Zic, which are found in multiple ESC, DE, and NPC sets of enriched interactions and are probably needed for the regulation of differentiation of both lineages as well as early differentiation and ESC maintenance. Footprint enrichments for Atf3 are mostly exclusive to significant NPC overlaps, whereas Sox17 and Eomes footprints are mainly exclusive to significant DE cluster overlaps. As expected, Nanog and Sox2 have significant footprint enrichment in ESC cluster overlaps. Jund

and Junb footprints are found in clusters with significant overlapping clusters with NPC and DE affiliations. Although we expected a higher fraction of concordant significant overlapping clusters with matching gene expression and chromatin accessibility for their respective cell types, we still have many candidate regulatory genes for each cell type.

2.3.3 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC and DE in Human

We repeated the same analysis for human ES cell differentiation using the human H1 ESC line to identify candidate regulators for the differentiation of each cell type. We collected 27 RNA-seq human datasets (see methods), resulting in 17,661 genes expressed over 1 TPM in two or more replicates across both NPC and DE time-courses (Table 2.1). We used maSigPro to identify 5,755 differentially expressed genes grouped into 18 distinct clusters (Figure 2.3A, Table 2.4). Four clusters containing 1,338 genes (23%) decrease over time. Seven clusters comprising 1,638 genes (29%) increase during NPC differentiation. Another seven clusters containing 2,779 genes (48%) increase during DE differentiation. We plotted the correlation of the expression profiles over each differentiation time-course (Figure S2.4A) as well as plotting the samples on a UMAP (Figure S2.4B). UMAP1 represents the difference between the two differentiation time-courses, and UMAP2 represents differentiation time with larger changes in the differentially expressed genes later in each time-course suggesting a clear difference between early and late differentiation as seen in rat (Figure S2.4A-C). Clusters that increased during NPC differentiation included TFs important in neurodevelopment and expression in early neurons such as *SOX2*, *FOXP1*, and *MEF2A*[57–59]. Clusters that increased during DE differentiation included TFs that are important in endoderm development and early differentiation including *FOXA2*, *GSC*, and *EOMES* [30,31,60,61]. The ESC-like clusters, which decrease in expression in both of the time-courses, include interesting TFs involved in ESC

maintenance including *NANOG, OCT4*, and *TCF7L1*. Many of these TFs are shared between rat and human; some examples are, *PAX6, MEF2A,* and *MEIS2* for NPC differentiation and *GATA4, LHX1,* and *SOX17* for DE differentiation. The TFs found in cell-type specific clusters in human suggest they are expressed in the appropriate pattern to regulate differentiation into either NPC or DE.

We performed GO analysis of clusters increasing in NPC (blue) or DE (green) (Figure 2.3B). Cluster HR14, which increases in expression after day 3 in the NPC time-course, is enriched for GO terms such as brain development ($p < 4x10^{-9}$) and forebrain development ($p < 1x10^{-4}$) and includes *MEIS2*. Cluster HR17, which has high expression on days 2 and 3 of the DE time-course, is enriched for GO terms such as endoderm formation ($p < 7x10^{-5}$) and includes *EOMES* and *LHX1*. Next, we considered clusters in the ESC group (tan, Figure S2.5A). For example, Cluster HR13, which decreases after day 1 in NPC differentiation and is highest in day 2 in DE differentiation, is enriched for GO terms such as embryonic morphogenesis ($p < 4x10^{-6}$).

Next, we investigated changes in chromatin accessibility in the human differentiation time-courses using ATAC-seq (Figure 2.3C). We collected 27 ATAC-seq human datasets (see methods), resulting in 75,271 accessible chromatin regions in two or more replicates (Table S2.1). We clustered temporally differential cCRE as performed in rat. We found 19,630 differentially accessible cCRE that grouped into 24 clusters (Figure 2.2C, Table 2.5). We observed a loss in accessibility in 4,636 (24%) of cCRE that form eight clusters. We observed six clusters that increased in chromatin accessibility in the NPC lineage (labeled in blue) accounting for 4,828 (25%) of differential cCRE. We found six clusters that increased in accessibility in the DE time-course, totaling 8,480 (42%) of the differential cCRE. Four clusters shared changes in accessibility in the two time-courses (black) totaling 1,686 (9%) of the

differential cCRE. We plotted the correlation of cCRE profiles for each cell type (Figure S2.4D) and plotted the differential cCRE from each sample on a UMAP (Figure S2.4E). UMAP1 represents differentiation time and UMAP2 represents the difference between cell types. Similar to what we found in rat, human differential cCRE show substantial differences between early and late stages (Figure S2.4B). We then focused on cCREs near TFs that are also differentially expressed to find which cCREs show increasing or decreasing accessibility and expression in the same direction (concordant) or different directions (discordant) changes in both techniques (Figure 2.3C). TCF4, a TF that supports differentiation into neurons, expression increases during human NPC differentiation [62]. TCF4 has 3 out of 8 associated differential cCREs and 3 of these cCREs with concordant increases in accessibility to the RNA expression profile. *GBX2*, a neuroectodermal marker, increases in expression over the NPC time-course and has 4 out of 9 concordant differential cCRE [63]. *RXRG*, which is involved in endoderm development [64], increases in expression over the DE time-course and has two differential cCREs, one of which concordant. Of the differential TF cCREs, 66% have at least one cCRE that is concordantly changing with its expression profile. Similar to what we found in rat, this suggests that TFs found in cell-type specific expression clusters show concordant differential accessibility of at least one associated cCRE.

Next, to explore functional roles of cell-type specific TF cCREs we analyzed each cCRE cluster using GO for clusters increasing in NPC (blue) and DE (green, Figure 2.3D). Cluster HA18, with the greatest accessibility on day 5 of the NPC time-course, is enriched for GO terms such as forebrain development ($p < 2x10^{-11}$) which includes *GBX2, LEF1,* and *FOXB1*. Cluster HA3, which increases in accessibility after day 1 of DE differentiation, is enriched for GO terms like gland development ($p < 1x10^{-7}$) and includes *PITX2*, *RARA* and *RARG*.

We single out the TFs found in both Human and Rat NPC and DE differentiation. Many of the TFs upregulated in both cell-types are upregulated in both rat and human, such as *LHX1, bHLHE40,* and *RARA* for NPC differentiation and *PITX2, RARA,* and *CER1* for DE differentiation (Figure 2.2D and 2.3D) and are important for regulation of differentiation into their specific cell-types [50,61,65–68]. We also used GO to investigate ESC specific clusters (Figure S2.5B Cluster HA12, which decreases in expression after day 3 in NPC differentiation and day 1 in DE differentiation, is enriched for GO terms like embryonic morphogenesis ($p < 8x10^{-9}$). GO analysis of cCRE clusters shows that sets of TFs with shared functions are increasing in accessibility in similar temporal patterns in human and many of these regions show similar patterns in rat.

We then investigated which *de novo* motifs are found in each cluster using Homer (Figure S2.5C-D). We found POU5F1, SOX2, and NANOG motif enrichment in ESC clusters, whereas NPC clusters were enriched for MEIS2, MEF2A, and SP1 motifs. The MEF2A motif is enriched in two clusters (both $p < 1x10^{-38}$), both are NPC clusters. Motifs enriched in DE clusters include EOMES, OTX2, and GSC (Figure S2.5C). In conclusion, *de novo* motif enrichment highlights TFs with binding sites found in differential cCRE with time-course specific functionality in human.

Next, we analyzed the relationship between differential cCREs and expression clusters in human, similar to what we did in rat (see methods, Figure 2.3E). We detected 20 sets of significantly enriched interactions (Table 2.11). Of these, 13 had concordant significantly enriched clusters, with expression and accessibility clusters falling into the same category, NPC (blue), DE (green), or ESC (tan). Seven discordant significantly enriched clusters, and one significant interaction in which the accessibility cluster is unaffiliated with a specific cell-type (black). We focused on cCREs surrounding TFs in each significantly overlapping clusters

and found 93 instances of cCREs for TFs (totaling 50 unique TFs). Examples of TFs with shared affiliations with NPC clusters include *FOXB1* and *PAX8*. Interestingly, these two TFs are in the same RNA cluster and have cCREs that are in the same ATAC clusters, which suggests coordinated regulation in NPC differentiation. Some of the TFs with shared DE affiliations include *PITX2*, *FOXA2*, and *OTX2*. Of note, many of the cCREs with significant overlaps in the DE group had multiple cCRE regions with the same cluster overlap identity, *SOX17, PITX2,* and *OTX2* all had multiple regions in significant overlapping clusters with the same cCRE cluster affiliations. TFs in ESC affiliated clusters include *POU5F1*. All of the TFs listed above have increased expression and promoter accessibility in their respective cell types, making them clear candidates for regulatory activity in their respective differentiations. We investigated the TFs that have significant overlapping clusters from different groups (i.e. increased expression in NPC and increased accessibility in DE) between DE and NPC clusters such as *RXRG*. This TF has increased expression in the DE time-course and increased accessibility in the NPC time-course in human, suggesting it is a strong regulatory factor candidate for at least one of the differentiation time-courses.

We then mined the ATAC regions with significant overlaps from Figure 2.3E for footprints (see methods, Figure 2.3F). The most enriched footprints include SOX2, SOX17, and SP5. These motifs are found in multiple ESC, DE, and NPC enriched interactions and are probably needed for the regulation of differentiation of both lineages and ESC maintenance. Footprint enrichments for E2F1 and bHLHE40 are mainly exclusive to significant NPC cluster overlaps. OTX2, GATA3, and GSC that are mainly exclusive to significant DE cluster overlaps. NANOG has motifs in ESC cluster overlaps. Whereas SMAD3 and FOXP1 motifs are found in clusters with significant overlapping clusters with NPC and DE affiliations. Since the majority of the significantly overlapping clusters are concordant, we have many cCREs for TFs for each cell

type. There are a few discordant cCREs, which are also interesting to investigate because they have differing promoter accessibility and expression in their respective cell types. The discordant cCREs may still be involved in the regulation of differentiation possibly via repression.

2.3.4 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC Between Rat and Human

We then focused on one-to-one orthologs that are expressed in at least one species and detected 9,898 genes expressed above 5 TPM in at least two replicates across NPC differentiation in either species. We computed the differential expression between rat and human during differentiation into NPC using maSigPro (Figure 2.4A, Table 2.6). We found 3,562 differentially expressed orthologs grouped into 18 clusters. We separated the differentially expressed clusters by early, late, and species-specific expression profiles. For the early group, expression profiles were highest in both human and rat during the first half of the time-courses (roughly from days 0 to 5). We found four clusters accounting for 1,244 genes (35%) that had high expression in the early group (light blue). For the late group, both human and rat expression profiles peaked in expression later in the NPC time-course (from days 4 to 8). We found six clusters accounting for 1,117 genes (31%) that had increased in expression in the late group (dark blue). For the species-specific group, expression profiles had an opposite trend in each species (i.e. early group in rat, and late group in human). We determined seven clusters totaling 1,201 genes (34%) have species-specific expression profiles (black). Two of these clusters have high expression in the late group in rat and the early group in human. The other five clusters are in the early group in rat and the late group in human. We focused on TFs found in each cluster as they may play a role in the regulation of NPC differentiation in both species. TFs found in the early group include important ESC maintenance genes like *NANOG*, and

POU5F1, and important early neuronal differentiation genes, *ZIC3*, *TGIF1*, and *GBX2* [57,63]. TFs found in the later expression clusters including, *RARA*, *TEAD2*, and *PBX1* are more likely to be involved in the final stages of NPC differentiation or NPC maintenance [69]. The TFs in the species-specific clusters are interesting because they might have differing regulation in each species. These TFs include; *LEF1*, *MEIS2*, and *KLF2* [70]. While 66% of expressed orthologs show conserved changes in expression between the two species, another 34% show species-specific behavior. The TFs found in each of the three groups are expressed in the appropriate temporal pattern to regulate different temporal stages of NPC differentiation in one or both species.

Next, we investigated each cluster using GO and separated clusters by their group affiliation, starting with the early group (Figure 2.4B). Early cluster JNR1 is enriched for GO terms such as negative regulation of neuron apoptotic process ($p < 2x10^{-4}$) and GABA synthesis, release, reuptake, and degradation ($p < 1x10^{-4}$). We next investigated GO terms for clusters from the late group (Figure 2.4C). Late cluster JNR3 is enriched for GO terms such as neural tube development ($p < 4x10^{-5}$) and includes *RARG, SALL2,* and *TEAD2*. Next we investigated GO terms found for the clusters found in the species-specific group (Figure 2.4D). Species-specific cluster JNR9, which is very highly expressed in day 0 in rat, with a sharp decrease on day 1 and increases after day 4 in human, is enriched for GO terms such as regulation of neuron death ($p < 5x10^{-5}$). GO analysis of gene expression indicates species-specific clusters with differential expression between the two species during NPC differentiation are enriched for a distinct subset of neuronal GO terms.

Next, we compared the changes in chromatin accessibility in NPC differentiation between rat and human (Figure 2.4E) by clustering cCREs as previously described. We detected 25,599 uniquely alignable accessible chromatin regions in at least two replicates

across NPC differentiation in each species. We found 7,400 differential cCRE that grouped into 18 clusters (Table 2.7). We classified the cCRE clusters into three different groups: early, late, and species-specific. We found six clusters accounting for 2,444 cCRE (33%) that were the most accessible in the early group. There are seven clusters accounting for 2,708 cCRE (37%) in the late group. There are five clusters consisting of 2,248 cCRE (30%) in the species-specific group. Clusters in the early group included cCREs surrounding TFs that are a mix of ESC markers, including *POU5F1, TCF7L1*, and NRSF/*REST*, and early neuronal differentiation markers, including; *MEIS2*, *FOXP1*, and *ZIC3*. cCREs in the late group include TFs that are important in the final stages of NPC differentiation in both species; including, *EGR1*, *GBX2*, and *TEAD4*. Arguably, the most interesting are cCREs in the species-specific group, as these cCREs may be important in species-specific aspects of differentiation in each species. The cCREs in species-specific clusters include *PBX1*, *ATF3*, and *KLF3*. While 70% of alignable cCRE show conserved changes in chromatin accessibility between the two species, another 30% show species-specific behavior.

Next, we used GO to analyze each cCRE cluster, separating the clusters into early (light blue), late (dark blue) and species-specific (pink), starting with the early group (Figure 2.4F). Early cluster JNA9 is enriched for GO terms such as brain development ($p < 9x10^{-11}$) and embryonic morphology ($p < 5x10^{-7}$) which include cCREs for *GBX2, TGIF2,* and *FOXP1*. Clusters in the late group were analyzed for GO terms (Figure 2.4G). Late cluster JNA 16 is enriched for GO terms such as regulation of neuron differentiation ($p < 2x10^{-12}$) and includes cCREs for *SOX11, NOTCH1,* and *MEF2A*. These cCREs are accessible in a pattern conducive to conserved regulation in NPC differentiation in rat and human. We then focused on species-specific clusters (Figure 2.4H). Species-specific cluster JNA3, which increases in accessibility in rat after day 1 and decreases in human after day 6, is enriched for GO terms such as brain development ($p < 2x10^{-8}$) and tissue morphogenesis ($p < 2x10^{-14}$) and includes *LHX1, NOTCH1,*

and *PBX1*. Since these cCREs differ in their temporal accessibility in each species, they may play species-specific roles during NPC differentiation, while cCREs in early or late clusters may play a more conserved role in the same process.

2.3.5 Distinct Transcriptional and Chromatin Accessibility Trajectories for DE Between Rat and Human

We analyzed the expression one-to-one orthologs during DE differentiation in both species using maSigPro and detected 7,320 differentially expressed genes (Figure 2.5A). These genes were further classified into 12 clusters with distinct stem, mesendodermal, or definitive endodermal profiles (Figure 2.5A, Table 2.8). We separated the clusters by early, late, and species-specific expression profiles. We found 2,691 genes (37%) in five clusters that had increased expression in human and rat DE time-courses early in DE differentiation (light green). We found 2,666 (36%) genes in four clusters with increased later expression (dark green). Clusters with genes expressed at the late stage were highly enriched with endodermal markers that showed gradual activation towards the end of DE differentiation. TFs found in the late group included *GATA4*, *SOX17*, and *GSC*. Brachyury/T (JDR6), which is a mesendodermal marker [71,72], had the highest expression level at day 2 differentiation for human and day 4 differentiation for rat and then maintained the expression level towards the end of time-course. Similar to what we observed in the NPC differentiation, we also identified species-specific expression group profiles between two species. The remaining 1963 (27%) genes in three clusters have more pronounced species-specific expression profiles (black). JDR5 and JDR12 exhibit early expression in rat and late expression in human while cluster JDR4 has late activation in rat and early activation in human. These clusters include a set of TFs associated with cell cycle regulation and biosynthetic processes were enriched in species-specific clusters such as *ETS1*,

SALL2, and *RARG*. While 73% of expressed orthologs show conserved changes in expression between the two species, another 27% show species-specific behavior.

In order to determine the roles of differentially expressed genes during DE differentiation, we ran GO analysis on each cluster and summarized the biological processes of these genes by their early or late expression profiles. The early groupings of genes were significantly associated with DNA replication (JDR1, $p < 1x10^{-43}$), cellular lipid catabolic process ($p < 1x10^{-8}$)/response to the extracellular stimulus ($p < 1x10^{-5}$) (JDR2) and DNA repair (JDR9, $p < 1x10^{-11}$) in both human and rat (Figure 2.5B). However, genes in cluster JDR3 were involved in basic cellular processes, such as actin cytoskeleton organization ($p < 1x10^{-13}$) and cell projection assembly ($p < 1x10^{-12}$) (Figure 2.5C). We found genes in cluster JDR6 reached to the highest level during the intermediate stage of DE differentiation and Brachyury T marked this cluster as a mesendodermal cluster (Figure 2.5A). As expected, genes in JDR6 were significantly associated with canonical Wnt signaling pathway ($p < 1x10^{-8}$), mesenchyme development ($p < 1x10^{-7}$) and mesoderm development ($p < 1x10^{-7}$) (Figure 2.5C). We also found three clusters that show reverse expression patterns in two species (Figure 2.5A). Genes in cluster JDR4 shown increased expression in rat but decreased in human were enriched in ncRNA metabolic process ($p < 1x10^{-9}$) while genes in cluster JDR5 and JDR12 that were down-regulated during rat differentiation but up-regulated in human were highly involved in cellular protein catabolic process ($p < 1x10^{-15}$), response to oxidative stress ($p < 1x10^{-12}$), autophagy ($p < 1x10^{-11}$) and neutrophil degranulation ($p < 1x10^{-8}$) (Figure 2.5D).

To further understand transcriptional control during DE differentiation in human and rat, we identified cCREs that showed differentially chromatin accessibility during DE differentiation in two species (Figure 2.5E) and found 18,942 (74% out of 25,599) regions that formed 15 clusters with distinct accessibility patterns (Figure 2.5E, Table 2.9). We further selected clusters

based on their associated genes that were shared in human and rat. Corresponding to the gene expression patterns (Figure 2.5A), we found that stem cell specific clusters enriched with pluripotent markers while three clusters were enriched with endodermal markers. This resulted in a loss of 3,856 (20%) regions while a gain of 4,506 (24%) regions during DE differentiation that were shared between two species. We paired DNA accessibility data with the gene expression data by separating chromatin accessibility clusters into three groups: early, late, and species-specific. The clusters in the early group were classified by the highest accessibility in the first half of each time-course while the late group was classified by the highest accessibility in the later stages of each time-course. The species-specific group is defined by differing accessibility trends between the rat and human in a cluster. There are 7,689 (41%) regions in six early clusters. Specifically, three of these clusters showed the early expression profiles in both rat and human. There are 4,506 (24%) regions in three late clusters, and collectively show the late expression patterns in both species. In addition, we also found 6,747 (35%) regions in six species-specific clusters. Clusters in the early group included regions surrounding *MYCN* and *TCF7L1*, which marked the stemness and early differentiation processes. Regions that are in the late clusters include *FOXA2*, *FOXA1*, and *GSC* which marked the final stage of DE differentiation in both species. Corresponding to the gene expression profiles (Figure 2.5A), we found one intermediate cluster, JDA1, reached to the highest accessible level at day 2 for human and day 4 for rat DE differentiation and also marked by the chromatin region activity around Brachyury T. Therefore, we treated this cluster as the mesendodermal cluster. We also found a set of TFs associated with cell cycle regulation and biosynthetic processes were enriched in species-specific clusters. These TFs included *ETS1*, *NRSF/REST*, and *ID3*. Interestingly, we found CTCF, which is involved in gene regulation by mediating the formation of chromatin loops [73–75], in cluster JDA6 that had early activation in rat but late activity in human (Figure 2.5E). Genes within this cluster were significantly enriched in cellular response to growth

factor stimulus and apoptotic signaling pathway, indicating different roles of CTCF in mediating the formation of transcription units in two species. While 65% of alignable cCRE show conserved changes in chromatin accessibility between the two species, another 35% show species-specific behavior.

In order to further characterize the TF regions in each of the groups of clusters we analyzed the GO terms for the clusters in each group. The early grouping of accessible regions were significantly associated with regulation of neuron differentiation ($p < 1x10^{-25}$) and developmental growth ($p < 1x10^{-18}$)(Figure 2.5F), which indicates the regulation of early differentiation. Cluster JDA11,14,15 were highly enriched with endodermal markers that showed active chromatin accessibility at the end of DE differentiation for both species (Figure 2.5E). Regions in JDA15 are associated with wounding ($p < 1x10^{-16}$) and muscle structure development ($p < 1x10^{-10}$) (Figure 2.5G). As for the mesendodermal cluster, JDA1, regions are involved in mesenchyme development ($p < 1x10^{-20}$), Wnt signaling pathway ($p < 1x10^{-14}$) and pattern specification ($p < 1x10^{-18}$), which correspond well with the intermediate cluster JDR6 in gene expression profiles (Figure 2.5A). Interestingly, the species-specific cluster regions are highly enriched in gene regulation processes in response to the external stimulus. For example, JDA7 is associated with mRNA processing ($p < 1x10^{-16}$) and histone modification ($p < 1x10^{-15}$) and JDA8 is associated with negative regulation of gene expression ($p < 1x10^{-17}$) and Wnt signaling pathway ($p < 1x10^{-9}$)(Figure 2.5H).

These results indicate that there are groups of genes and regulatory elements with conserved transcriptional patterns across species. Chromatin accessibility is complementary with gene expression to identify conserved regulatory modules during DE differentiation in two species. However, species-specific patterns may reflect the difference in growth condition and following generation of transcripts during DE differentiation.

2.3.6 Integrative analysis of NPC Differentiation

We then compared the relative conservation of gene expression and cCRE clusters between rat and human during NPC differentiation. We analyzed Early, Late, and species-specific RNA and chromatin clusters from Figures 2.4 to identify significant enrichments of shared genes and cCREs (Figure 2.6A; see methods). We found that 2,309 cCRE (out of 7,400 in Figure 2.4E) were within 20kb of 1,530 differentially expressed genes (out of 3,562 in Figure 2.4A). We detected seven sets of significantly enriched interactions, five of which were conserved, three with late RNA and late ATAC, one with early RNA and early ATAC, one with early RNA and late ATAC (discordant; Table 2.12). Overall, we found 175 (8%) cCREs for 145 genes in these five conserved clusters including a total 14 TF cCREs associated with 11 TFs. TFs in the early enriched NPC group include *GBX2*. TFs with shared late NPC affiliations include *ID1* (JNR8, JNA13), *JUNB* (JNR8, JNA13), *EGR1* (JNR8, JNA13), and *KLF16* (JNR17, JNA12) and are very strong candidates for conserved temporal regulation of differentiation in both species. The remaining two enriched associations showed species-specific differences. TFs in species-specific affiliated clusters include *PBX1* (JNR3, JNA3), *ATF3* (JNR3, JNA3), and *RARG* (JNR3, JNA3). These TFs may differ in the temporal aspects of their regulation during NPC differentiation in the two species.

We then mined the cCREs in the enriched clusters for footprints using Wellington and motifs using homer (Figure 2.6B). The most enriched footprints are SP5, KLF, SP1, and CTCF. These motifs are found in significant overlapping most clusters. Footprint enrichments for SOX2, IRF3, PBX1, and TGIF1 are mainly enriched in the Early cluster overlap and are potential candidate regulators for early NPC differentiation in both species. By contrast, footprint enrichments including EGR1, TCF4, and ZIC3 are mainly found in Late cluster overlaps, which suggest that these TFs are likely candidate regulators for late NPC differentiation in both

species. JUND and KLF10 footprint enrichments are found in the discordant cluster overlap. MAZ has footprint enrichment mainly in the species-specific cluster overlap, which suggests that MAZ may play a different role in NPC regulation in each species.

2.3.7 Integrative analysis of DE Differentiation

We investigated the TFs that drive DE differentiation in both species using the same approach that we applied to NPC differentiation. We found that 10,148 cCRE (out of 18,942 in Figure 2.5E) were within 20kb of 5,322 differentially expressed genes (out of 7,320 in Figure 2.5A). We detected 41 sets of significantly enriched interactions between RNA and chromatin (see methods). 27 sets of interactions were conserved, including six with early RNA and early ATAC, seven with late RNA and late ATAC, seven with early RNA and late ATAC and seven with late RNA and early ATAC (Table 2.13). Overall, we found 2,214 (22%) cCREs for 1,548 genes in these conserved clusters, including a total 166 TF cCREs associated with 90 TFs. Some TFs with shared affiliations in early DE clusters include *SOX2* (JDR9, JDA9; JDR9, JDA13), *ETV1* (JDR1, JDA5), *IRF1* (JDR1, JDA5), *SOX11* (JNR8, JNA5), *FOXP1* (JDR9, JDA11), *GBX2* (JDR9, JDA13), *NANOG* (JDR9,JDA13), *CIR1* (JDR10, JDA5), and *LEF1* (JDR9, JDA13). Some of the TFs with shared late DE affiliations include *FOXA2* (JDR11, JDA11), *GATA4* (JDR11, JDA11), *EGR1* (JDR3, JDA11), *PITX2* (JDR11, JDA11), *JUND* (JDR3, JDA11), *TGIF1* (JDR3, JDA11), *FOXA1* (JDR6, JDA11), *ATF3* (JDR3, JDA11), *PRDM1* (JDR11, JDA11), and *TEAD2* (JNR6, JDA15). Some of the TFs with early RNA and late ATAC include *MYC* (JDR8, JDA11) and *GBX2* (JDR9, JDA14). Some of the TFs with late RNA and early ATAC include *TGIF1* (JDR3, JDA5), *T* (JDR6, JDA13), *JUN* (JDR3, JDA5), and *ZIC3* (JDR7, JDA9). These may be genes that are primed at different times before expression or are partially controlled by active repression by a repressive TF. The remaining 14 significantly enriched interactions are species-specific. TFs in species-specific affiliated clusters include

MEIS2 (JDR5, JDA8), *NRSF/REST* (JDR5, JDA8), *RARG* (JDR4, JDA2), *ID3* (JDR4, JDA2), *ETS1* (JDR5, JDA8, JDA11), *E2F1* (JDR4, JDA2), *TCF7L1* (JDR5, JDA15), *T* (JDR6, JDA1), and *FOXP1* (JDR9, JDA2). These TFs may differ in the temporal aspects of their regulation during DE differentiation in the two species.

We also mined the ATAC regions with significant overlap for footprints (Figure 2.6B). The most enriched footprints include SP1, MAZ, and TFE3. These motifs are found in the majority of clusters and are general candidate regulators for both mesendoderm and DE differentiation of both species. Footprint enrichments including IRF3, E2F3, and E2F7 are mainly enriched in the conserved early clusters, making these TFs candidate regulators for early DE differentiation in both species. Footprint enrichments including EGR1, bHLHE40, TCF4, and ZIC3 are mainly enriched in late cluster overlaps. These TFs are strong candidate regulators for late DE differentiation in both species. ETV1, MEF2A, TCF12, and TGIF1 are some motifs found in clusters with significant overlapping clusters with discordant Early and Late affiliations. Footprint enrichments for GATA4/6, IRF1, NANOG, and PRDM1 are enriched in species-specific cluster overlaps, suggesting that these TFs are important candidate regulators for DE differentiation, but they differ in their temporal regulation in each species. The role of these TFs in DE differentiation are most likely not as evolutionary conserved as the TFs that share an early or late association between rat and human. Overall, we recovered many more significantly enriched interactions in DE differentiation than NPC differentiation (Figure 2.7).

2.4 Discussion

We used RNA and ATAC sequencing in a daily time-course to map gene and cCRE dynamics during differentiation into ectodermal NPCs and DE in rat and human. The comprehensive gene expression and chromatin accessibility changes allowed us to identify

conserved regulatory modules of lineage-specific regulators in both species. In particular, we were able to connect gene expression changes with associated changes of open chromatin regions during differentiation. We also found as expected that pluripotent gene expression was downregulated with the loss of chromatin accessibility. In addition, we identified substantial species-specific changes with instances of species-specific trajectories in rat and human during the same cell type differentiation. We found putative TF binding using open chromatin footprinting enriched in different sets of genes with conserved or species-specific changes. This is to our knowledge the first systematic comparison of *in vitro* differentiation from ESCs to ectodermal NPCs as well as to DE in rat and human.

We find many cCRE similarities between NPC and DE in both species. For instance, the neuroectodermal marker, GBX2 is in significantly enriched early gene interactions in both NPC and DE time-courses in both species [63]. EGR1 is found in significantly enriched late gene interactions in both NPC and DE time-courses in both species (Figures 2.2 and 2.3). Candidate regulators that increase in both species most likely have conserved function and temporal regulation. However, it is interesting to examine the genes that differed in expression over each time-course, as these genes may not be as conserved across rat and human. Genes like PBX1, ATF3, and RARG have species-specific expression and accessibility, suggesting they are required for different stages of NPC differentiation in the two species. Taken together, these candidate regulators are less evolutionarily conserved between species. Further experiments are required to understand the exact function and conservation of these candidate regulators in NPC differentiation between rat and human.

The conserved TFs FOXA2 and SOX17 were both up-regulated during DE differentiation, with relatively higher expression in rat than in human. CXCR4 is another key marker of mammalian endoderm differentiation [9,14,21] and was upregulated in both species with

higher expression in human. We also found that both MIXL1 and EOMES expression starts early during DE differentiation, which has also been demonstrated in other organisms [32,76,77]. We also found that GATA4/6 and GSC are actively expressed in our time-course. Furthermore, we found that T, which is a marker of mesendoderm, was expressed highly in the middle of the time-course. This indicates that differentiating cells experienced the proper intermediate stage during endoderm development in both species. However, we found that some genes showed unexpected expression differences between species. The first is OCT4 (POU5F1), which is one of the key TFs defining the pluripotent cell fate in stem cells, that stayed highly expressed in later DE differentiation in human but not in rat. Previous studies have also found high expression of OCT4 at the end of human DE differentiation and they suggested this is possibly the effects of culture conditions [37]. We also detected human-specific expression of NANOG in late DE stage that agrees with studies where OCT4 and NANOG induce DE markers during endoderm differentiation in human [31,38]. However, at least one endoderm single-cell study has shown the co-expression of OCT4 and endoderm markers in human DE [55]. Further experiments are required to understand the exact function of OCT4 during DE differentiation as it may be one of the more important human-specific DE regulatory changes. In addition, we also detected expression of FOXA1 in rat but not in human during DE development. As a core member TF of the FOXA family, expression of FOXA1 and 2 have been detected in both mouse and Xenopus endoderm. Although a study showed that FOXA1 is the direct target of EOMES in human DE and knockdown of EOMES decreases FOXA1 expression [31], our results indicate that FOXA1 may not be not an essential TF during human DE development as the equivalent function may be taken over by FOXA2.

Of the most enriched footprints shared between both NPC and DE differentiations SP5, KLF, SP1, and CTCF are needed in many cell types and for a broad range of cell functions [78,79]. SP5 has been found to interact with brachyury/T during development [80]. Footprints for IRF3 are

found in early significant cluster affiliations for both NPC and DE differentiations. This may be due to the exiting of the pluripotent state and into a more differentiated state, as pluripotent ESCs do not utilize IRF3 [81] and it is a regulator of cell proliferation. Footprints for SOX2 are found in early NPC significant cluster overlaps due to its ability to regulate both pluripotent ESCs and multipotent NPCs [82]. Footprints for TGIF1 and PBX1 were also found in early NPC cluster overlaps. TGIF1 [83] and PKNOX, a co-factor of PBX1, regulate neural cell fate specification [69]. Footprints for TEAD, EGR1, TCF4, and ZIC3 are found in late significant cluster affiliations for both NPC and DE differentiations. ZIC3 and EGR1 regulate the transition out of naïve pluripotency [84,85] and TCF4 has been shown to regulate cell survival and neuronal differentiation [86]. Whereas members of the TEAD family are needed for the regulation of differentiation into both endoderm and neuronal lineages [87,88]. Footprint enrichments for MAZ are also shared in NPC and DE differentiations. MAZ is part of the most enriched footprints in DE differentiation whereas it shows species-specific enrichment in NPC differentiation. MAZ has been shown to mediate promoter activity during neuronal differentiation [70] and SP1 and MAZ may share the same cis-regulatory elements [89]. Further experiments are required to understand the exact function and conservation of these candidate regulators in NPC and DE differentiation.

One of the most striking results of our comparative analysis of NPC and DE differentiation is that we find considerably more conservation during DE differentiation than NPC differentiation (Figure 2.7). This could be a product of the *in vitro* system or inherent biological differences. One possible reason is the differences between differentiation protocols for each species. Both NPC differentiation protocols concluded after eight days but used different inhibitors. For human NPC differentiation, three inhibitors were added to the media (ALK2/3, ALK5, and GSK3) while only one was used in rat (ALK2/3) with the addition of the growth factor FGF2. For DE differentiation, the 5-day human protocol was performed using an established kit while in rat the 7-day rat protocol used a plethora of growth factors and inhibitors such as

Activin. The differentiation of both these cell-types took equally as long or longer in rat, which is a key difference from *in vivo* where the formation of both these cell-types occurs on a faster scale in rat [90,91]. Variations in differentiation efficiency for each cell-type between the two species is also a possibility. Another class of differences could be biological. For example, our results could be impacted by different starting ESC states between rat and human. Human H1 ESCs are in a primed state (post-implantation; epiblast cells) while mouse ESCs are in a naive state (preimplantation blastocyst; inner cell mass) [92]. The rat DAc8 ES cells are thought to be in-between the naive state of the mouse ESCs and primed state of the H1 ESCs. Rat ESCs are cultured in similar conditions to naive mouse ESCs, with medium containing two inhibitors (2i), but express Cdx2, which is a trophoblast marker not expressed in naive mouse ESCs [93]. Another possible explanation is the differing endpoints between DE and NPC differentiation. The complexity of neuroectoderm and NPC differentiation *in vivo* is extremely difficult to recapitulate *in vitro*. These cells could have positional identity *in vivo* such that their fate is influenced by environmental cues from neighboring cells, which can only be partially recapitulated *in vitro*. Rodents and humans have a large gestational time gap in neuroepithelial cell formation. In rodents these cells can be distinguished after about 7 days gestation but are only distinguishable near the end of the third week of gestation in humans. Numerous NPC subtypes exist, many of which can be recapitulated *in vitro* [90]. Formation of DE *in vivo* occurs much earlier in development than the formation of NPCs. DE develops from epiblast cells that transition from the most anterior region of the primitive streak, which marks the beginning of gastrulation [91]. Previous studies have shown that there is a common pathway for inducing DE formation that is shared between many species [94]. This similarity of early development of DE between species paired with the complexity of

later formation of NPCs and developmental time differences between species may account for the striking differences in conservation seen between NPC and DE differentiations.

We used transcriptome and cCRE changes during the differentiation of rat and human ESCs to map expression changes and cCRE dynamics during differentiation into ectodermal NPCs and DE lineages. Our gene expression time-courses allowed us to carefully define temporal profiles of expression in connection with the timing of early differentiation in the two germ layers.

We have discovered evolutionarily conserved and species specific *cis*-regulatory elements by comparing conserved DNA between rodents and humans whose lineages diverged about 91 million years ago [5]. While mice are often used in comparative genomic analyses with human [6], rats have been less intensively studied using functional genomics for comparative purposes. In particular, mice and rats diverged around 12 million years ago, there are likely to be very interesting and novel evolutionary changes found between rat and human that are not found in mouse [95], [3]. A comparative analysis of GRNs between the three species would be an important next step to understand how the changes that we see in gene expression and regulation during differentiation can occur while conserving the overall process of DE and NPC differentiation.

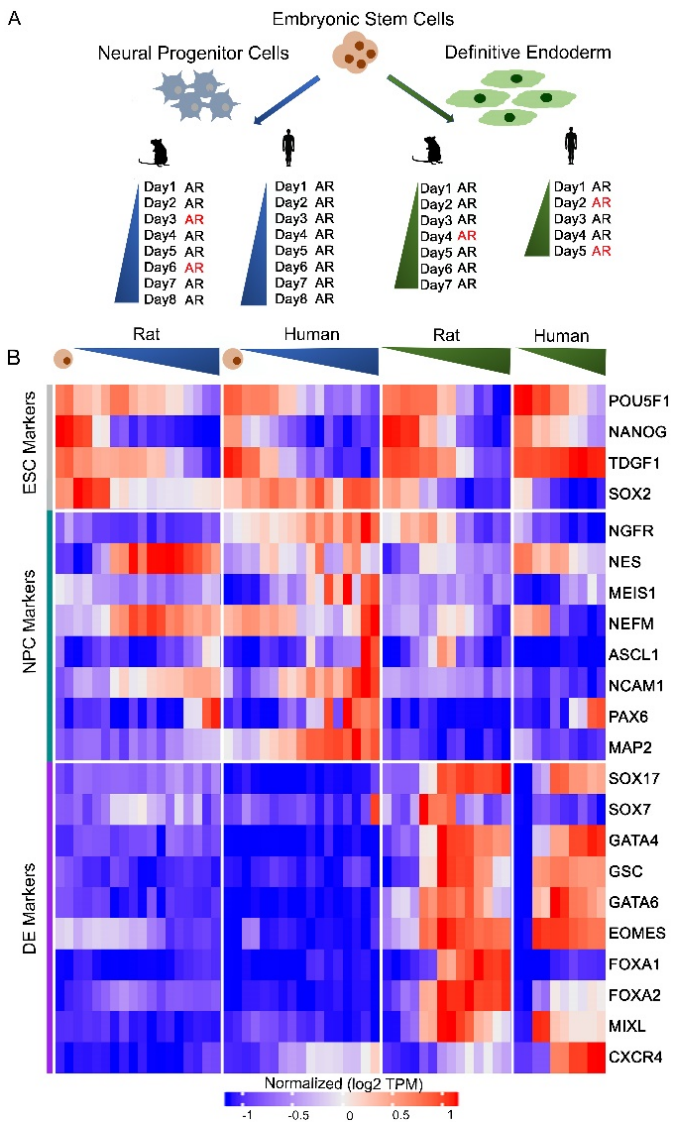

Figure 2.1 Generation of Definitive Endoderm and Neural Progenitor Cells in Two Species

A) Schematic representation of study design. Color indicates differentiation cell lineage. Letters next to differentiation days (AR) indicate RNA-seq (R) and ATAC-seq (A) timepoints. Red color indicates change in media composition. B) Heatmap of marker genes for each differentiation lineage and ESC markers. Color increasing bar on top of the heatmap shows progression of time course. Color indicates differentiation lineage. Colors to the left of the

heatmap refer to the lineage groupings. Grey: ESC markers, teal: NPC markers, and purple: DE markers. Transcripts per million (TPM) values were log transformed and row-mean normalized for all genes.

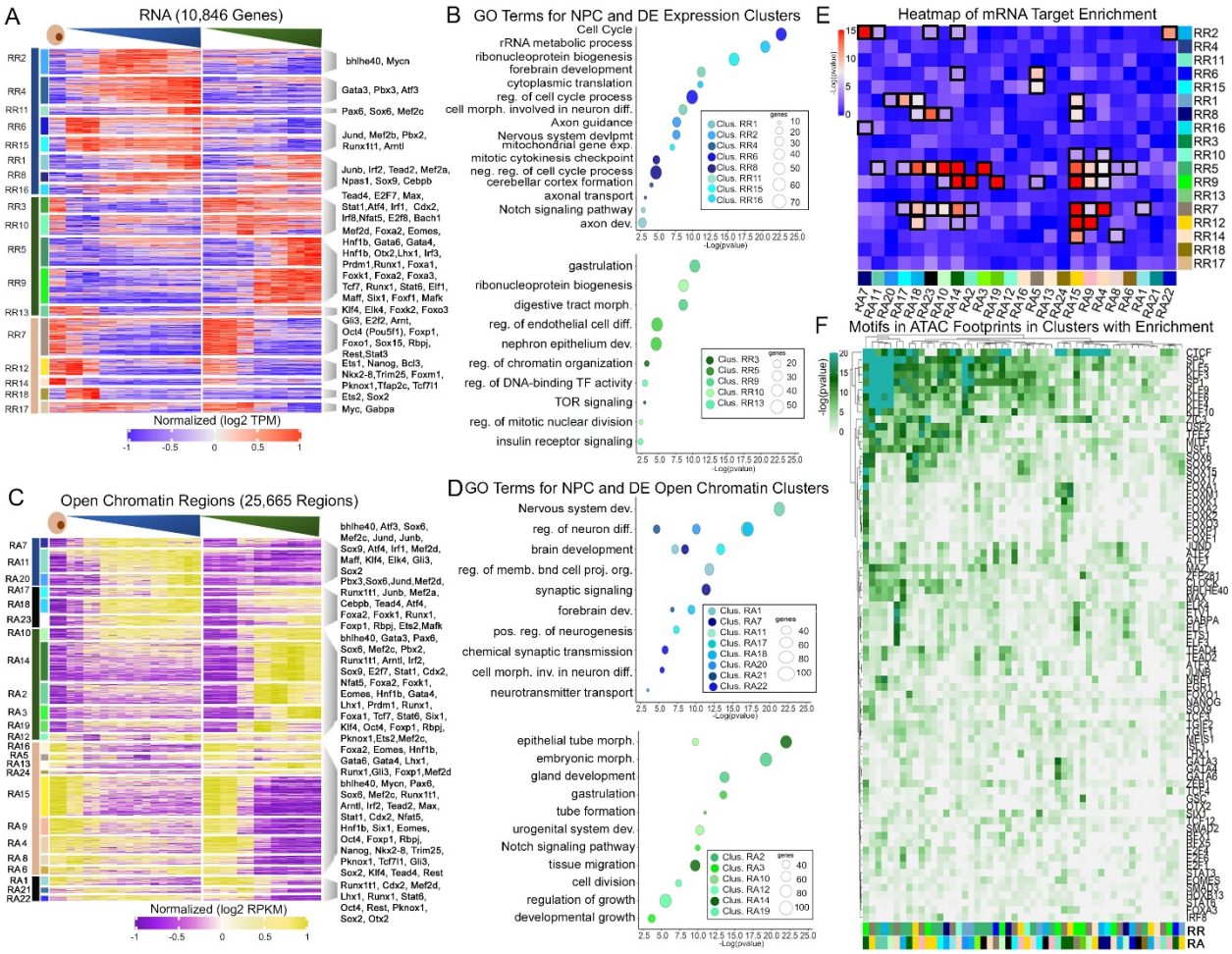

Figure 2.2 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC and DE in Rat

A) Heatmap of differentially expressed genes in NPC and DE differentiation in rat. Hierarchical clustering was used where dark blue is low expression and red is high expression TPM values were log transformed and row-mean normalized. Colors on the far left correspond to NPC (Blue), DE (Green), or ESC (Tan). The second row annotations color bar are to delineate individual clusters named on the left. B) GO terms for NPC and DE gene expression regions in clusters. C) Heatmap of differential cCRE regions in NPC and DE differentiation in rat. Hierarchical clustering was used where purple is less accessible and yellow is more accessible DNA regions Colors on the far left correspond to NPC (Blue), DE (Green), ESC

(Tan), or undefined (Black). The second row annotations color bar are to delineate individual clusters named on the left. RPKM values were log transformed and row-mean normalized. D) GO terms for NPC and DE cCRE in clusters E) Heatmap of X^2 p-values on a contingency table of clusters of accessible regions and expression clusters with applied P-value of 10^{-4}. The significant cluster overlaps are in a black box. F) Motif enrichments in ATAC footprints for regions with significant overlap between clusters.

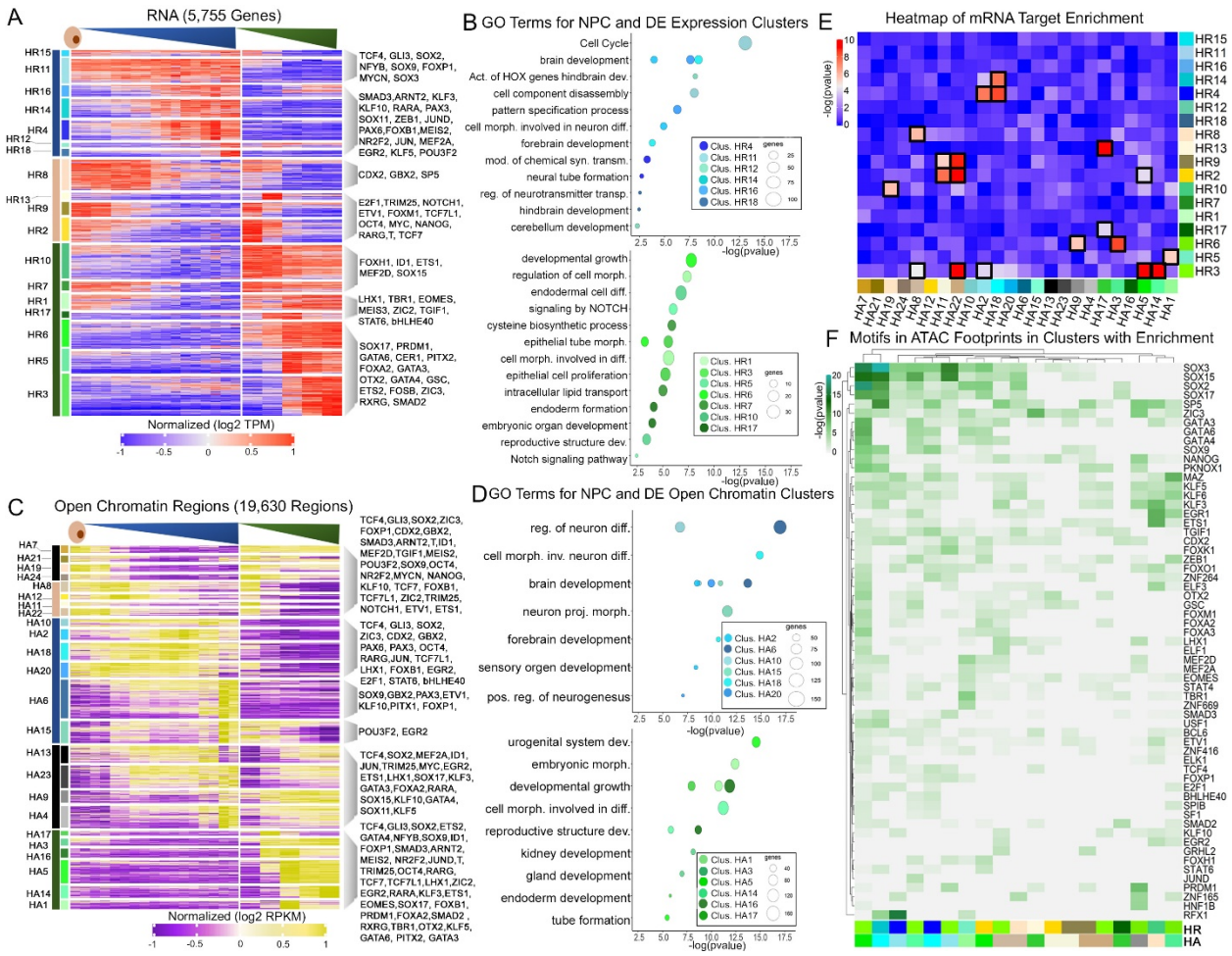

Figure 2.3 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC and DE in Human

A) Heatmap of 5,755 differentially expressed genes (alpha = 0.05, p < 0.05) in NPC and DE differentiation in human. Hierarchical clustering was used where dark blue is low expression and red is high expression. Transcripts per million (TPM) values were log transformed and row-mean normalized. Colors on the far left correspond to NPC (Blue), DE (Green), or ESC (Tan). The second row annotations color bar are to delineate individual clusters named on the left.. Select TFs are labeled. B) GO terms for NPC and DE related gene expression clusters. The GO terms for clusters that increase over the NPC time course: blue. The GO terms for clusters that increase over the DE time course: green. C) Heatmap of 19,630 differential cCRE in NPC and

DE differentiation in human. Hierarchical clustering was used where purple is less accessible and yellow is more accessible DNA regions Colors on the far left correspond to NPC (Blue), DE (Green), ESC (Tan), or undefined (Black). The second row annotations color bar are to delineate individual clusters named on the left. Values were log transformed and row-mean normalized for all regions. D) GO terms for NPC and DE related cCRE clusters. E) Heatmap of X^2 p-values on a contingency table of clusters of accessible regions and expression clusters with applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[18 × 24]). The significant cluster overlaps are surrounded by a black box. F) ATAC footprints in regions of significant overlap between clusters. The color bars are the same in A, C, and E for each RNA and ATAC cluster.

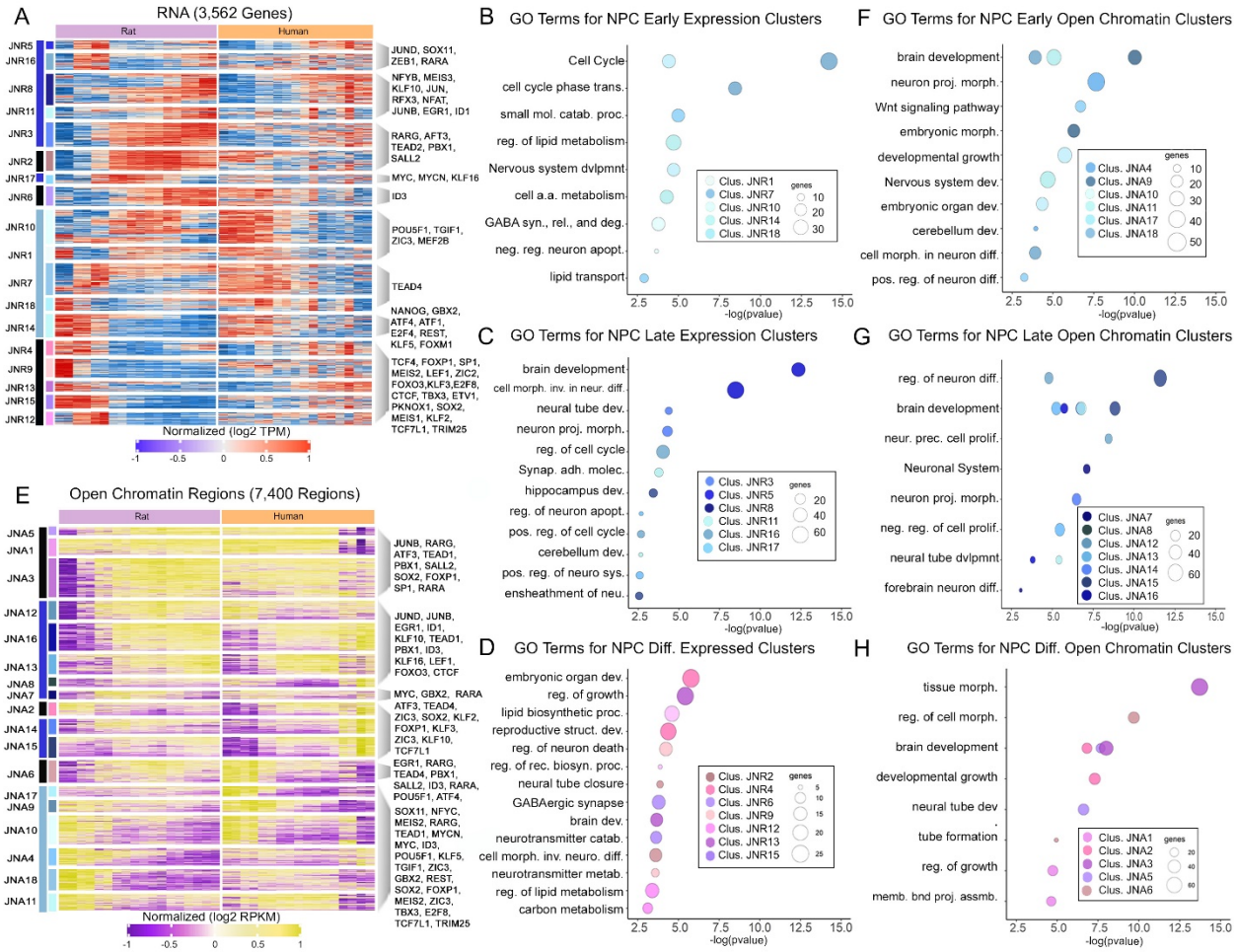

Figure 2.4 Distinct Transcriptional and Chromatin Accessibility Trajectories for NPC Between Rat and Human

A) Heatmap of 3,562 differentially expressed genes (alpha = 0.05, $p < 0.05$) in NPC differentiation in rat and human. Hierarchical clustering was used to generate eighteen expression clusters denoted by the adjacent number. Transcripts per million (TPM) values were log transformed and row-mean normalized for all genes. Colors on the far left correspond to the group (early [light blue], late [dark blue], or species-specific [black]) the cluster is associated with. The second row annotations color bar are to delineate individual clusters named on the left. Select TFs are labeled. B) GO terms for Early grouped gene expression clusters. The GO terms for clusters of genes that increase early in both rat and human NPC differentiation are in

light blue. GO terms are determined using metascape. C) GO terms for Late grouped gene expression clusters. The GO terms from genes that increase later in the NPC time-course for both rat and human are in dark blue. D) GO terms for species-specific grouped gene expression clusters. These genes showed different expression profiles in rat and human. GO terms were obtained from genes that differ during differentiation of the two species. E) Heatmap of 7,400 differential cCRE (alpha = 0.05, FDR = 0.05, $p < 0.05$) in NPC differentiation in rat and human. Hierarchical clustering was used to generate 18 DNA region clusters denoted by the adjacent number. RPKM (Reads Per Kilobase Million) values were log transformed and row-mean normalized for all regions. Colors on the far left correspond to the group (early [light blue], late [dark blue], or species-specific [black]) the cluster is associated with. The second row annotations color bar are to delineate individual clusters named on the left. Select TFs are labeled. F) GO terms for cCRE that were accessible early on in NPC differentiation in both rat and human. The GO terms for clusters that increase over the NPC time-course are in light blue. G) GO terms for cCRE that were accessible later on in NPC differentiation for both rat and human. The GO terms for these regions are shades of dark blue. H) GO terms for cCRE that have species-specific accessibility in NPC differentiation between rat and human. These regions are shades of purple and pink.

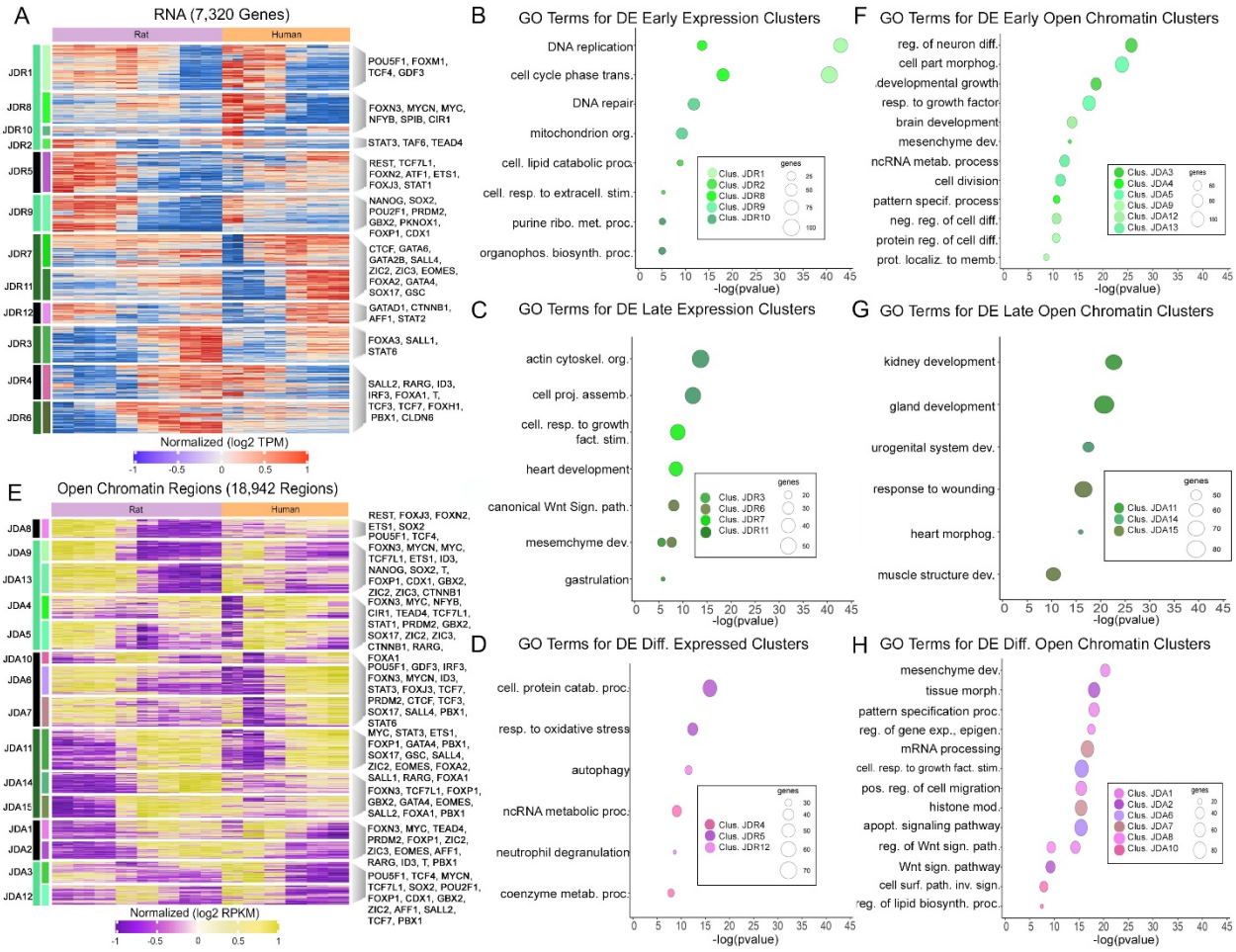

Figure 2.5 Distinct Transcriptional and Chromatin Accessibility Trajectories for DE Between Rat and Human

A) Heatmap of 7,320 differentially expressed genes (alpha = 0.05, $p < 0.05$) in DE differentiation in rat and human. Hierarchical clustering was used to generate 12 expression clusters denoted by the adjacent number. Transcripts per million (TPM) values were log transformed and row-mean normalized for all genes. Colors on the far left correspond to the group (early [light green], late [dark green], or species-specific [black]) the cluster is associated with. The second color bar is for each individual cluster. Select TFs are labeled. B) GO terms for Early grouped gene expression clusters. The GO terms for clusters of genes that increase early in both rat and human DE differentiation are in light green. GO terms are determined using

metascape. C) GO terms for Late grouped gene expression clusters. The GO terms from genes that increase later in the DE time-course for both rat and human are in dark green. D) GO terms for species-specific grouped gene expression clusters. These genes showed different expression profiles in rat and human. GO terms were obtained from genes that differ during differentiation of the two species. E) Heatmap of 18,942 differential cCRE (alpha = 0.05, $p < 0.05$) in DE differentiation in rat and human. Hierarchical clustering was used to generate 15 DNA region clusters denoted by the adjacent number. RPKM (Reads Per Kilobase Million) values were log transformed and row-mean normalized for all regions. Colors on the far left correspond to the group (early [light green], late [dark green], or species-specific [black]) the cluster is associated with. The second-row annotations color bar are to delineate individual clusters named on the left. Select TFs are labeled. F) GO terms for cCRE that were accessible early on in DE differentiation in both rat and human. The GO terms for clusters that increase over the DE time-course are in light green. G) GO terms for cCRE that were accessible later on in DE differentiation for both rat and human. The GO terms for these regions are shades of dark green. H) GO terms for cCRE that have species-specific accessibility in DE differentiation between rat and human. These regions are shades of purple and pink.

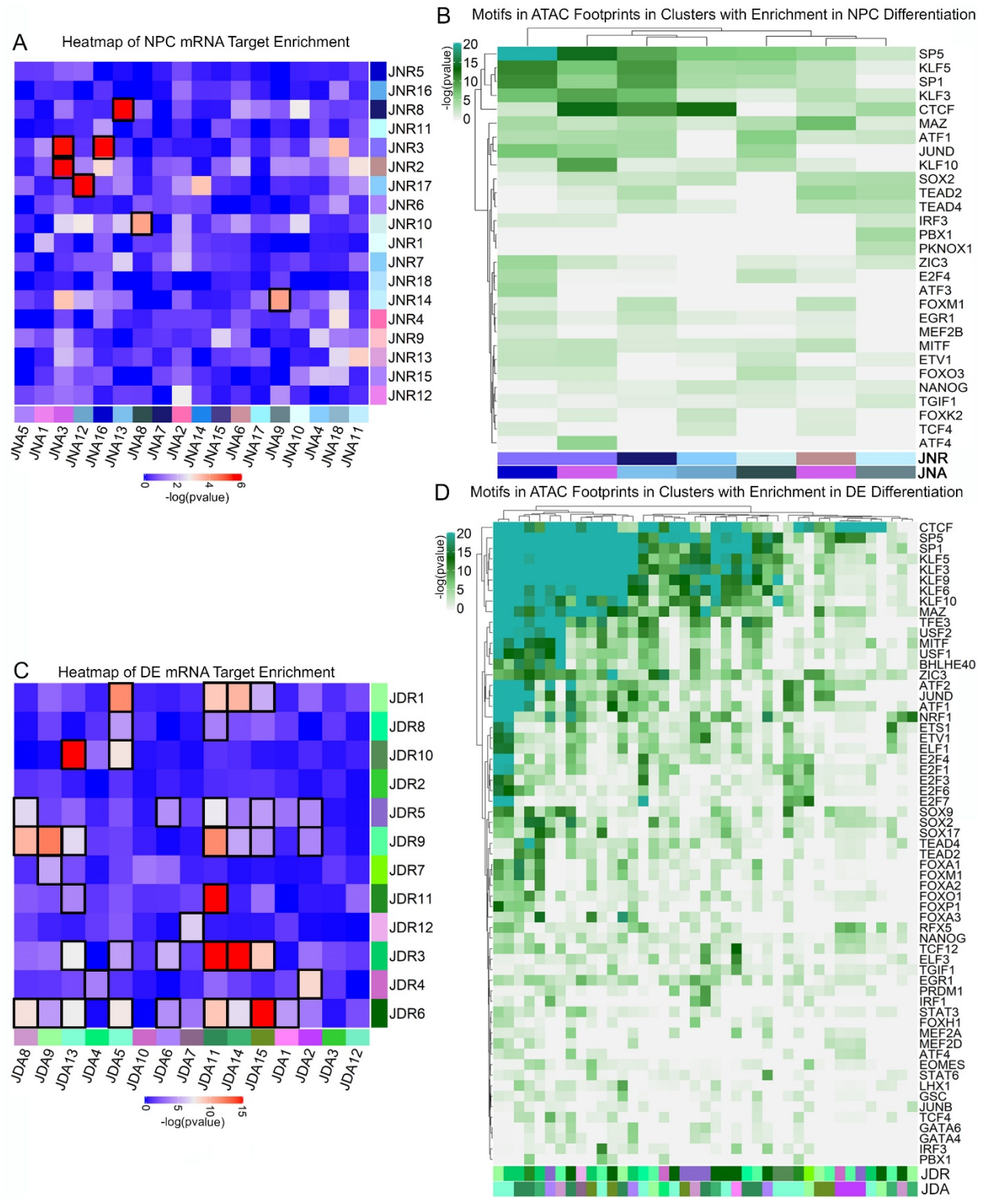

Figure 2.6 Integrative analysis of NPC and DE Differentiation

A) Heatmap of X^2 p-values on a contingency table of clusters of accessible regions and expression clusters with applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[18 × 18]). The significant cluster overlaps are surrounded by a black box. The colors next to each cluster match the color bar from Figure 4A and 4E, and depict the affiliation with the early NPC group (light blue), late NPC group (dark blue), or species-specific group (pink). B) ATAC footprints in regions of significant overlap between clusters. The color bars are the same in Figure 4A, C, and E for each RNA and ATAC cluster. C) Heatmap of X^2 p-values on a contingency table of clusters of accessible regions and expression clusters with applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[12 × 15]). The significant cluster overlaps are surrounded by a black box. The colors next to each cluster match the color bar from Figure 5A and 5E, and depict the affiliation with the early DE group (light green), late DE group (dark green), or species-specific group (pink). D) ATAC footprints in regions of significant overlap between clusters. The color bars are the same in Figure 2.5A, C, and E for each RNA and ATAC cluster.

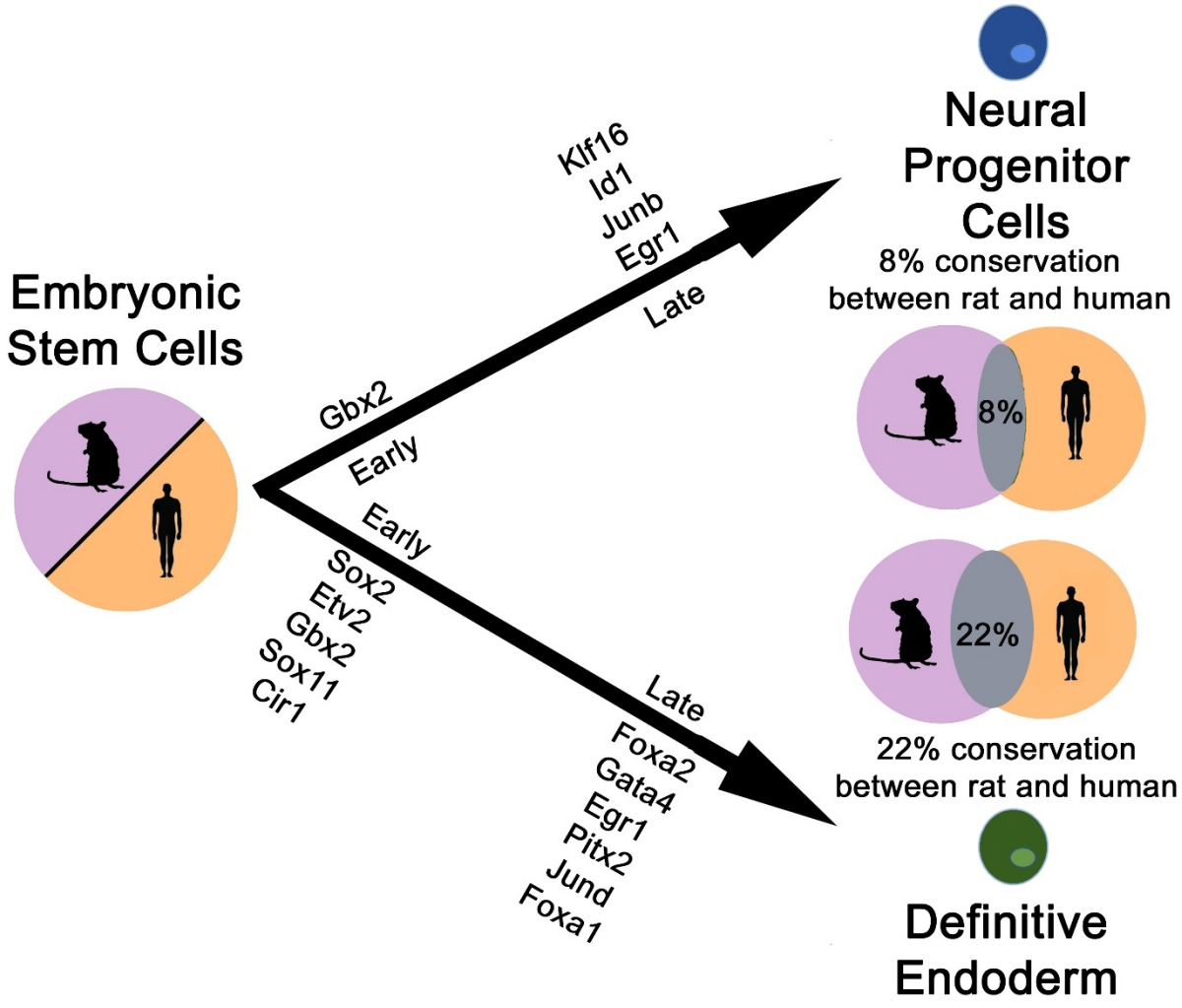

Figure 2.7 Summary Figure

Genes listed under the Early and Late NPC and DE arrows are some of the conserved genes we found during each stage of differentiation in rat and human. Overall, we find more conservation in DE differentiation than NPC differentiation.

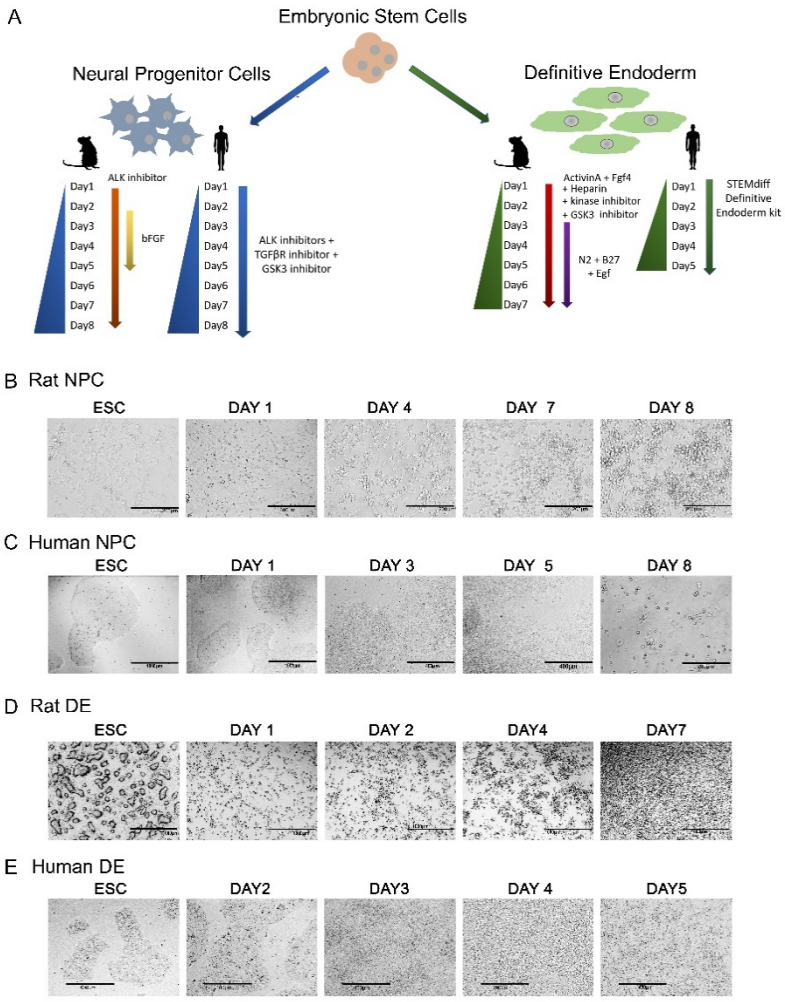

Figure S2.1 Related to Figure 2.1. Monolayer differentiation of embryonic stem cells into definitive endoderm and neural progenitor cells

A) Schematic representation of study design. Color indicates differentiation cell lineage. Human or rat symbols denote species. Arrows indicate media composition. B) Brightfield images of eight day time-course of rat NPCs by using adapted protocol [23]. C) Brightfield images of eight-day time-course of human NPCs by using adapted protocol [28]. D) Brightfield images of seven day time-course of rat DE differentiation by using adapted protocol [14]. E) Brightfield images during five-day time-course of human DE differentiation by using STEMdiff Definitive endoderm kit

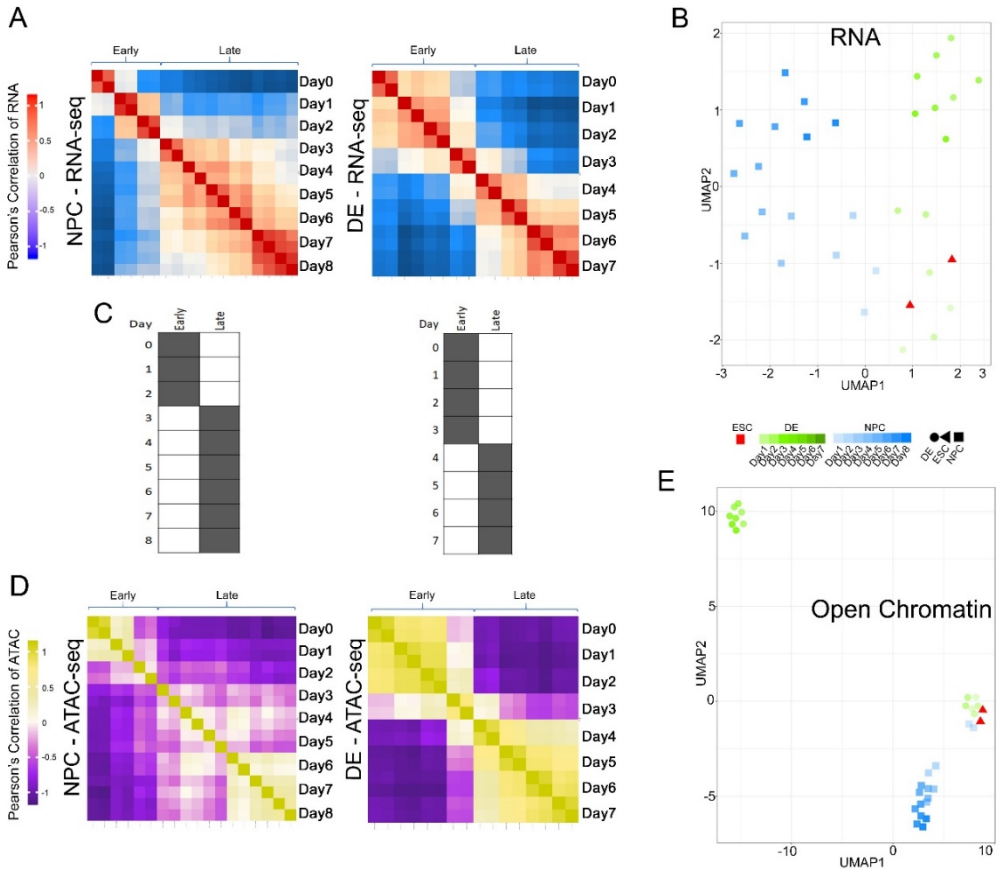

Figure S2.2 Related to Figure 2.2, correlation and visualization of differential ATAC and RNA profiles over rat NPC and DE differentiation time-courses

A) Pearson correlation of expression profiles over rat NPC and DE differentiation time-courses. The samples are separated into early and late expression profiles for each time-course. B) UMAP representation of differential expression profiles of rat NPC and DE time-courses annotated by cell-types. Color darkens as time-course progresses. C) Temporal staging of time points for each cell type was based on the clustering of RNA-seq and ATAC-seq data.

Pearson correlation of open chromatin profiles over rat NPC and DE differentiation time-courses. The samples are separated into early and late expression profiles for each time-course. D) UMAP representation of differential open chromatin profiles of rat NPC and DE time-courses annotated by cell-types

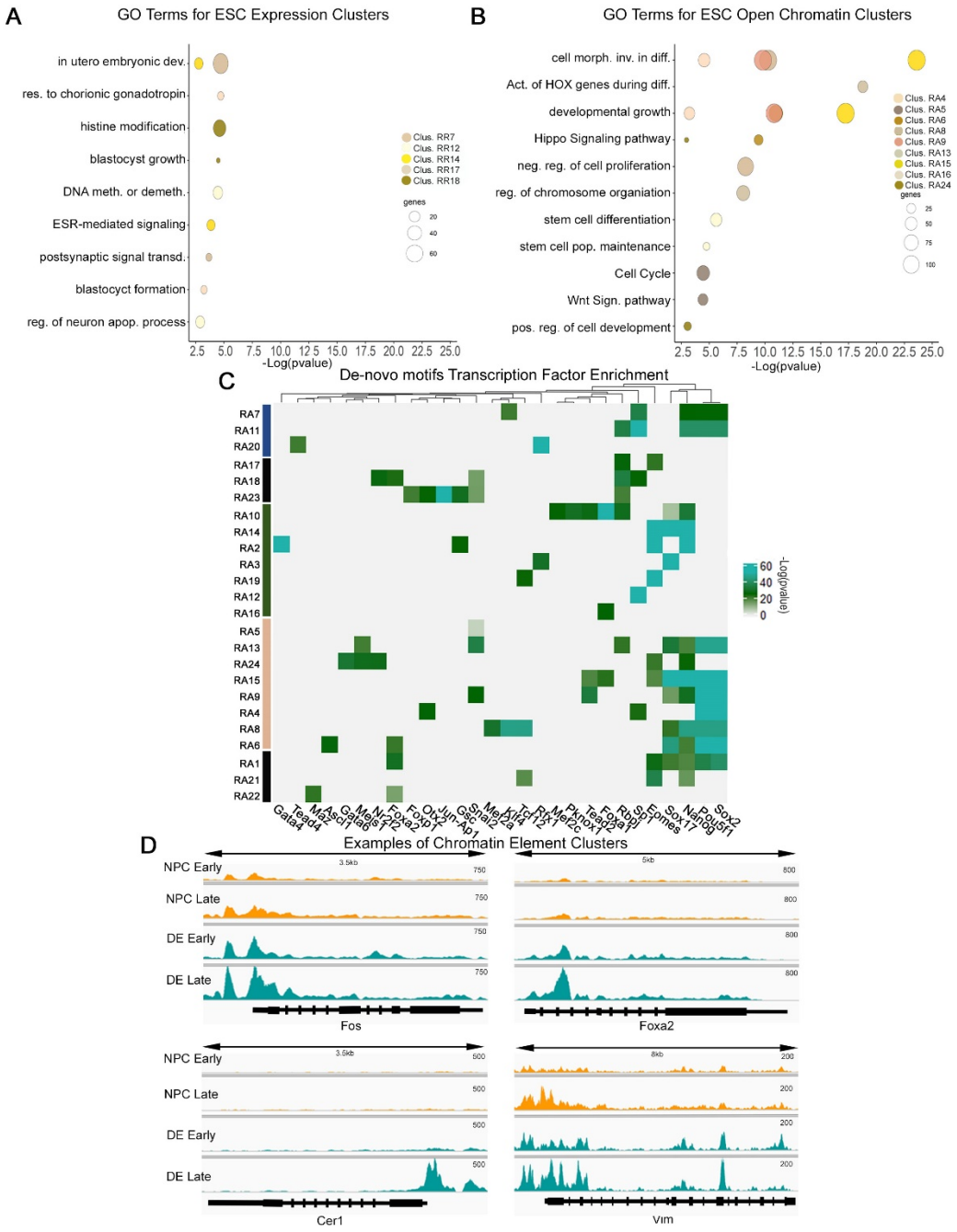

Figure S2.3 Related to Figure 2.2, gene ontology and motif analysis of differential open chromatin regions in clusters

A) Gene ontology (GO) enrichment analyses of differentially expressed rat genes in ESC category. B) GO enrichment analyses of differential open chromatin rat regions in ESC category

Motif enrichment between differential open chromatin regions in clusters. Clusters on the left are in the same order as the differential heatmap in Figure 2.2C. C) Examples of four open chromatin elements in rat visualized in Integrative Genome Viewer (IGV) that increase in either NPC (orange) or DE (teal) differentiation. Mapped ATAC reads are separated into early and late groupings from Figure S2.2E.

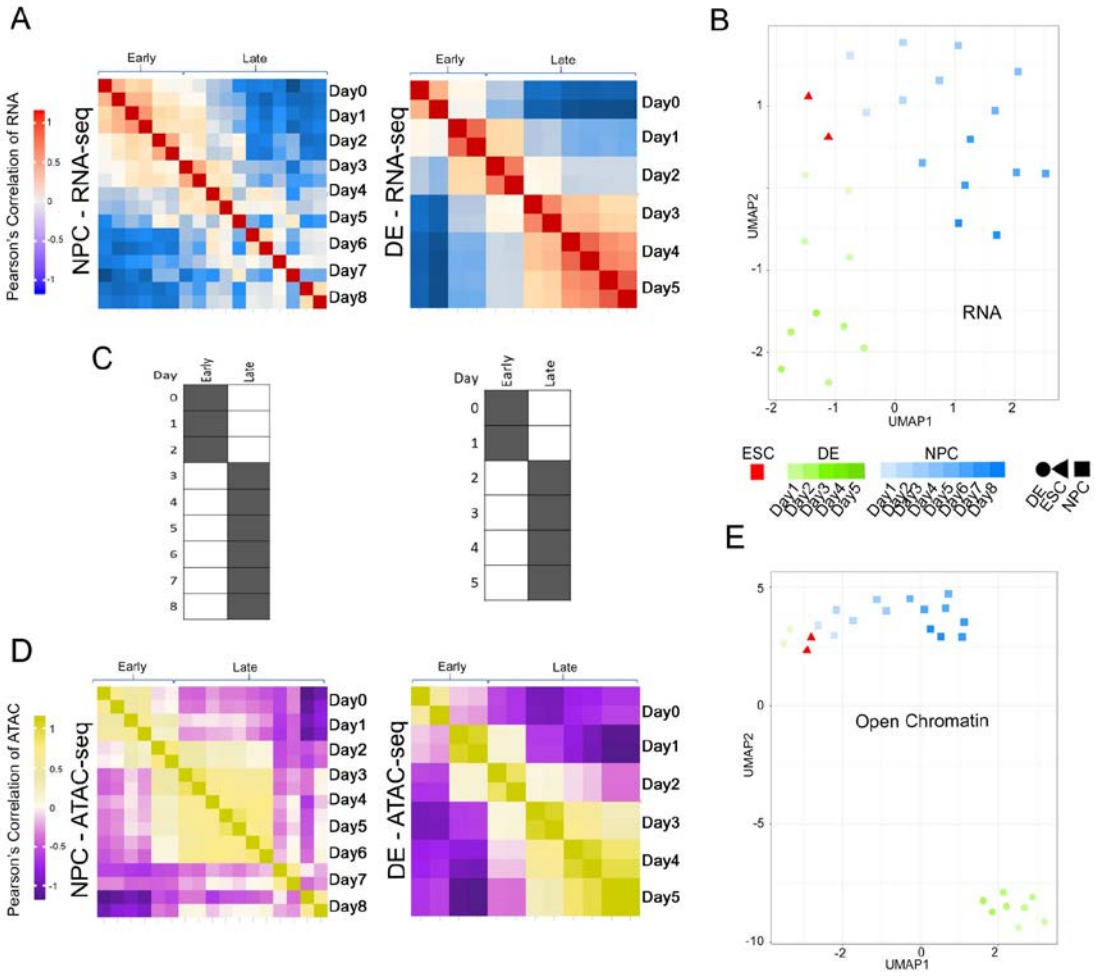

Figure S2.4 Related to Figure 2.3, correlation and visualization of differential ATAC and RNA profiles over human NPC and DE differentiation time-courses

A) Pearson correlation of expression profiles over humanNPC and DE differentiation time-courses. The samples are separated into early and late expression profiles for each time-

course. B) UMAP representation of differential expression profiles of human at NPC and DE time-courses annotated by cell-types. Color darkens as time-course progresses. Temporal staging of time points for each cell type was based on the clustering of RNA-seq and ATAC-seq data. C) Pearson correlation of open chromatin profiles over human NPC and DE differentiation time-courses. The samples are separated into early and late expression profiles for each time-course. D) UMAP representation of differential open chromatin profiles of human NPC and DE time-courses annotated by cell-types

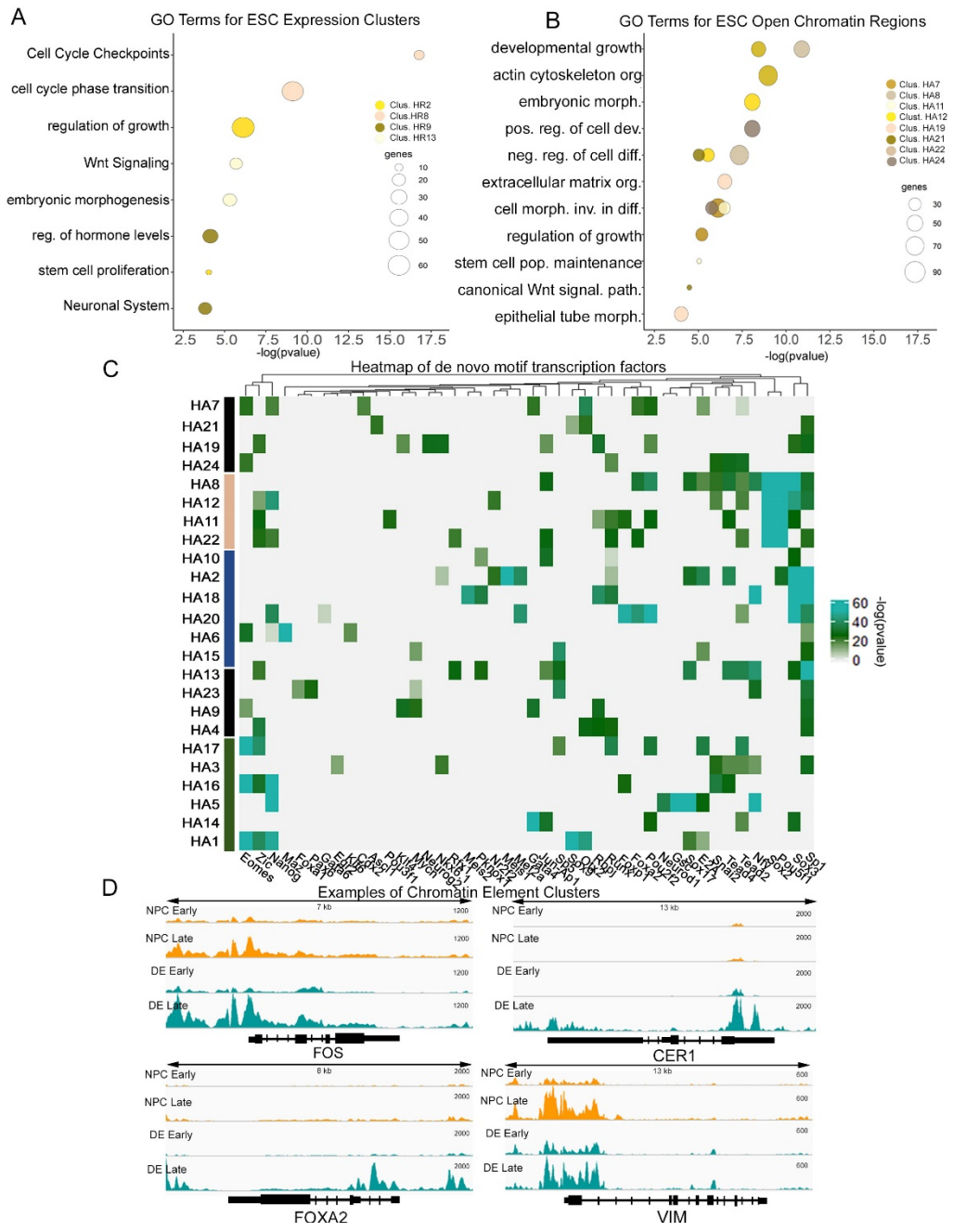

Figure S2.5 Related to Figure 2.3, gene ontology and motif analysis of differential open chromatin regions in clusters

A) GO enrichment analyses of differentially expressed human genes in ESC category. B) GO enrichment analyses of differential open chromatin human regions in ESC category. C)

Motif enrichment between differential open chromatin regions in clusters. Clusters on the left are in the same order as the differential heatmap in Figure 2.3C. D) Examples of four human open chromatin elements visualized in Integrative Genome Viewer (IGV) that increase in either NPC (orange) or DE (teal) differentiation. Mapped ATAC reads are separated into early and late groupings from Figure S2.4E.

Statistics for RNA and ATAC experiments							
			RNA-seq		ATAC-seq		
Timepoint	Species	Replicate	Mapped Reads	Genes > 1TPM	Mapped Reads	Homer Total Peaks	IDR-passing Total peaks
ESC	rat	Rep1	6.2M	11,245	37.6M	363,643	123,289
		Rep2	6.2M	11,213	49.0M	348,975	
NPC_Day1	rat	Rep1	9.8M	11,299	23.1M	47,677	13,042
		Rep2	9.3M	11,407	24.7M	59,720	
NPC_Day2	rat	Rep1	6.7M	11,439	35.0M	20,603	3,227
		Rep2	10.7M	11,465	23.4M	26,427	
NPC_Day3	rat	Rep1	7.2M	11,633	15.8M	49,803	6,452
		Rep2	10.5M	11,549	12.1M	87,091	
NPC_Day4	rat	Rep1	7.8M	10,807	29.1M	54,983	14,300
		Rep2	5.0M	10,995	17.1M	58,065	
NPC_Day5	rat	Rep1	7.3M	10,590	17.8M	43,418	4,300
		Rep2	10.2M	11,022	17.5M	26,591	
NPC_Day6	rat	Rep1	8.3M	10,283	26.4M	78,267	20,833
		Rep2	10.2M	10,588	19.3M	67,612	
NPC_Day7	rat	Rep1	19.2M	11,614	28.1M	87,006	18,708
		Rep2	22.8M	11,567	29.1M	74,759	
NPC_Day8	rat	Rep1	21.2M	11,538	24.6M	75,012	24,159
		Rep2	17.0M	11,473	23.3M	70,405	
DE_Day1	rat	Rep1	7.2M	12,053	55.5M	211,508	91,742
		Rep2	7.7M	11,650	43.1M	244,500	
DE_Day2	rat	Rep1	7.7M	11,997	47.8M	266,103	122,465
		Rep2	9.0M	11,900	45.5M	285,468	
DE_Day3	rat	Rep1	8.7M	11,850	55.3M	162,161	71,448
		Rep2	7.2M	11,900	50.5M	168,315	
DE_Day4	rat	Rep1	4.2M	11,770	68.4M	201,537	91,774
		Rep2	7.1M	11,793	56.3M	248,981	
DE_Day5	rat	Rep1	6.4M	11,206	62.7M	241,442	113,170
		Rep2	7.8M	11,269	32.8M	256,181	
DE_Day6	rat	Rep1	6.2M	11,629	48.6M	259,781	125,133
		Rep2	6.7M	11,663	48.6M	247,749	
DE_Day7	rat	Rep1	7.0M	11,561	46.5M	194,079	91,938
		Rep2	8.0M	11,601	46.3M	239,870	
ESC	Human	Rep1	7.1M	13,338	12.5M	161,388	34,217
		Rep2	17.2M	13,136	19.6M	142,540	
NPC_Day1	Human	Rep1	10.7M	13,000	19.6M	202,251	56,291
		Rep2	17.5M	13,259	15.6M	218,580	
NPC_Day2	Human	Rep1	14.9M	12,082	23.3M	262,432	100,204
		Rep2	13.3M	12,163	21.2M	262,432	
NPC_Day3	Human	Rep1	13.1M	12,783	27.6M	305,299	100,605
		Rep2	12.0M	12,500	23.9M	260,423	
NPC_Day4	Human	Rep1	14.8M	13,131	21.7M	219,505	--
NPC_Day5	Human	Rep1	13.7M	12,665	17.9M	262,553	73,072
		Rep2	13.0M	11,796	18.8M	207,578	
NPC_Day6	Human	Rep1	13.1M	13,332	18.2M	247,329	68,798
		Rep2	17.2M	12,842	21.9M	209,962	
NPC_Day7	Human	Rep1	17.9M	12,945	16.8M	188,416	64,496
		Rep2	11.0M	11,604	20.0M	200,339	
NPC_Day8	Human	Rep1	13.6M	13,262	11.7M	179,426	54,324
		Rep2	13.4M	13,318	29.3M	194,218	
DE_Day1	Human	Rep1	9.2M	13,633	62.3M	405,457	205,433
		Rep2	10.1M	13,551	73.9M	419,697	
DE_Day2	Human	Rep1	11.1M	13,586	65.5M	261,612	209,755
		Rep2	10.9M	13,757	75.2M	239,697	
DE_Day3	Human	Rep1	8.7M	13,995	67.6M	383,702	208,755
		Rep2	10.1M	14,415	84.5M	409,826	
DE_Day4	Human	Rep1	10.4M	14,091	70.4M	376,628	187,183
		Rep2	11.7M	14,316	83.2M	415,654	
DE_Day5	Human	Rep1	11.6M	14,029	92.6M	369,550	182,606
		Rep2	14.1M	13,947	70.4M	427,796	

Table 2.1 Statistics for RNA and ATAC experiments

name	RR2	RR4	RR11	RR6	RR15	RR1	RR8	RR16	RR3	RR10	RR5	RR9	RR13	RR7	RR12	RR14	RR18	RR17
RA7	0.0	0.0	0.7	0.0	0.5	0.1	0.0	0.0	0.3	0.2	0.3	0.0	0.3	0.0	0.0	0.0	0.5	0.0
RA11	0.0	0.2	0.1	0.2	0.0	1.0	0.0	0.0	1.0	0.1	0.0	0.0	0.9	0.0	0.0	0.0	0.1	0.7
RA20	0.4	0.0	0.0	0.7	0.0	0.0	0.7	0.3	0.8	0.0	0.0	0.6	1.0	0.1	0.4	1.0	0.9	0.2
RA17	0.2	0.6	0.0	0.6	0.0	0.0	0.6	0.2	0.7	0.1	0.0	0.3	0.9	0.0	0.8	0.2	0.5	0.0
RA18	0.0	0.1	0.6	0.2	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
RA23	0.0	0.1	0.5	0.0	0.5	0.1	0.0	0.1	0.0	0.7	0.0	0.7	1.0	0.0	0.0	0.0	0.0	0.0
RA10	0.4	0.6	0.7	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.4	0.0	0.0	0.2	0.2	0.0
RA14	0.0	0.4	0.1	0.0	0.0	0.0	0.1	0.2	0.3	0.9	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
RA2	0.0	0.0	0.3	0.3	0.2	0.9	0.4	0.1	0.2	0.9	0.6	0.0	0.4	0.0	0.0	0.2	0.2	0.1
RA3	1.0	0.6	0.3	0.1	0.1	0.5	0.3	0.3	0.1	0.4	0.0	0.2	0.2	0.0	0.3	0.1	0.4	0.3
RA19	0.6	0.3	0.1	0.0	0.7	0.6	0.0	0.0	0.2	0.0	0.1	0.0	0.9	0.0	0.0	0.2	0.4	0.8
RA12	0.8	0.1	0.4	1.0	0.0	1.0	0.8	0.8	0.0	1.0	0.3	0.1	0.9	0.0	0.4	0.7	0.3	0.7
RA16	0.1	0.3	0.3	0.2	0.0	0.2	0.2	0.6	1.0	0.2	0.3	0.0	0.1	0.0	0.1	1.0	0.4	0.2
RA5	0.6	0.0	1.0	0.0	0.0	0.2	0.2	0.6	0.6	0.5	0.0	0.0	0.1	0.6	1.0	1.0	0.2	0.5
RA13	0.1	0.6	0.9	0.2	0.5	0.2	0.3	0.0	0.4	0.9	0.0	0.9	0.2	0.3	1.0	0.0	0.3	0.0
RA24	0.6	0.7	0.1	1.0	0.4	0.7	0.8	0.8	0.1	0.3	0.3	0.1	0.5	0.6	0.7	0.8	0.9	1.0
RA15	0.0	0.1	0.1	0.0	0.3	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0
RA9	0.6	0.5	0.1	0.0	0.7	0.0	0.7	0.1	0.6	0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.5	0.1
RA4	0.4	0.1	0.8	0.2	0.5	0.0	0.0	0.1	0.4	0.0	0.0	0.0	0.0	0.0	0.5	0.5	0.0	0.0
RA8	1.0	0.2	0.2	0.0	0.3	0.4	0.1	0.4	0.7	0.6	0.0	0.0	0.2	0.1	0.0	0.0	0.2	0.6
RA6	0.3	0.1	0.0	1.0	0.1	0.3	0.2	0.1	0.4	0.2	0.0	0.0	0.3	0.0	0.0	0.4	0.4	0.5
RA1	0.4	0.4	0.9	0.2	1.0	0.0	0.1	1.0	0.0	0.6	0.0	0.5	0.2	0.0	0.4	0.2	0.2	0.0
RA21	0.5	0.4	0.8	1.0	0.9	0.4	0.6	0.3	1.0	0.0	0.0	0.1	1.0	0.0	1.0	1.0	0.4	0.0
RA22	0.0	0.1	1.0	0.5	1.0	0.0	0.3	0.8	1.0	0.3	0.0	0.0	0.4	0.2	0.5	0.1	0.1	0.1

Table 2.10 Correlation of mRNA-ATAC clusters in rat NPC and DE comparison analysis

name	HR15	HR11	HR16	HR14	HR4	HR12	HR18	HR8	HR13	HR9	HR2	HR10	HR7	HR1	HR17	HR6	HR5	HR3
HA7	0.8	0.1	0.1	0.2	0.1	0.4	1.0	0.2	0.5	1.0	0.7	0.7	0.1	0.0	1.0	0.7	1.0	0.1
HA21	0.3	0.3	1.0	0.7	0.0	0.5	0.9	1.0	0.7	0.1	0.5	0.7	0.0	0.3	1.0	0.2	0.5	0.0
HA19	0.4	1.0	0.7	0.1	0.2	0.8	0.7	0.4	0.4	0.5	0.0	0.0	0.6	0.4	0.6	0.9	0.4	0.3
HA24	0.3	0.1	0.2	0.7	0.6	0.5	1.0	0.2	0.1	0.8	0.0	0.2	1.0	0.1	0.5	0.1	0.5	0.0
HA8	0.3	0.4	0.2	0.1	0.2	0.5	1.0	0.0	0.7	0.0	0.2	0.0	0.4	0.1	0.9	0.1	0.0	0.0
HA12	0.5	1.0	0.5	0.0	0.1	1.0	0.6	0.0	1.0	0.5	0.8	0.5	1.0	0.9	0.1	0.9	0.0	0.1
HA11	0.1	0.2	0.1	1.0	0.4	0.2	0.6	0.0	1.0	0.0	0.0	0.2	0.1	0.2	0.1	0.0	0.5	0.1
HA22	0.6	0.8	0.5	0.3	0.5	0.2	0.2	0.0	0.7	0.0	0.0	0.0	0.9	0.7	0.3	0.0	0.0	0.0
HA10	0.2	0.2	0.4	0.3	0.3	1.0	0.6	0.1	0.9	1.0	0.4	0.0	0.1	0.4	0.6	0.8	0.6	0.2
HA2	0.4	0.3	0.3	0.0	0.0	0.3	0.2	0.3	0.8	0.0	0.0	1.0	0.4	0.5	0.8	0.5	0.1	0.0
HA18	0.9	0.5	0.1	0.0	0.0	0.2	0.1	0.3	0.8	0.2	0.4	0.4	0.2	0.6	1.0	0.1	0.6	0.0
HA20	0.9	0.6	0.1	0.7	0.0	0.8	0.4	0.0	0.3	0.3	0.1	0.1	0.0	0.1	0.1	0.3	0.0	0.0
HA6	0.0	0.2	1.0	0.2	0.2	1.0	0.0	0.3	0.5	0.9	0.1	1.0	0.5	0.2	0.2	0.1	0.1	0.0
HA15	0.1	0.4	0.1	0.4	1.0	0.1	0.0	0.0	0.1	0.0	0.1	0.1	0.1	0.3	0.8	0.1	0.4	0.1
HA13	0.1	0.8	0.0	0.4	0.0	0.1	1.0	0.7	0.1	0.1	0.1	0.0	0.0	0.4	0.9	0.0	0.2	0.9
HA23	0.5	0.1	0.2	0.6	0.0	0.9	0.0	0.1	1.0	0.1	0.1	0.0	0.1	0.3	0.3	0.9	0.1	0.0
HA9	0.2	0.7	0.9	0.1	0.4	0.1	1.0	0.1	0.9	0.2	0.6	0.1	0.3	0.1	0.8	0.0	1.0	0.2
HA4	0.1	0.4	0.1	0.1	0.1	0.8	1.0	0.6	1.0	1.0	0.2	0.3	0.7	0.3	0.7	0.2	0.0	0.6
HA17	0.5	0.9	0.0	0.9	0.9	0.4	0.7	0.5	0.0	0.1	0.2	0.1	1.0	0.3	0.0	0.8	0.2	0.0
HA3	0.7	0.0	0.5	0.0	0.0	0.3	0.3	0.0	0.7	0.2	0.2	1.0	0.5	0.7	0.3	0.0	0.0	0.1
HA16	0.2	1.0	1.0	0.8	1.0	0.2	0.0	0.2	0.0	0.4	0.2	0.0	0.9	0.1	0.3	0.2	0.0	0.6
HA5	0.3	0.3	0.6	0.1	0.0	0.4	0.1	0.2	0.0	0.0	0.0	0.0	0.4	0.3	0.0	0.1	0.0	0.0
HA14	0.0	0.4	0.2	0.2	0.1	0.3	0.3	0.0	0.1	0.0	0.0	0.2	1.0	0.1	0.6	0.3	0.5	0.0
HA1	1.0	0.5	0.7	0.6	0.2	0.4	0.7	0.3	0.7	0.8	0.3	0.0	0.0	0.3	0.9	0.1	0.0	0.0

Table 2.11 Correlation of mRNA-ATAC clusters in human NPC and DE comparison Analysis

name	JNR5	JNR16	JNR8	JNR11	JNR3	JNR2	JNR17	JNR6	JNR10	JNR1	JNR7	JNR18	JNR14	JNR4	JNR9	JNR13	JNR15	JNR12
JNA5	0.4	1.0	0.4	1.0	0.4	0.8	0.1	1.0	0.6	1.0	0.3	1.0	0.3	1.0	0.1	1.0	1.0	0.3
JNA1	0.3	0.5	0.9	0.9	0.0	0.1	0.7	0.3	1.0	0.0	1.0	0.8	0.5	1.0	0.2	0.7	0.4	0.1
JNA3	0.1	0.5	0.1	0.6	0.0	0.0	0.3	0.1	0.0	0.4	0.4	0.6	0.0	0.2	0.0	0.0	0.1	0.1
JNA12	0.1	1.0	0.5	0.5	0.4	0.0	0.0	0.7	0.0	1.0	0.2	1.0	0.0	1.0	0.4	0.1	0.2	0.1
JNA16	1.0	0.4	0.6	0.0	0.0	0.0	0.2	0.6	0.1	0.1	0.2	0.1	0.0	0.1	0.7	0.7	0.2	0.8
JNA13	0.9	1.0	0.0	1.0	0.5	0.0	0.0	0.9	0.0	1.0	0.0	1.0	0.1	0.6	1.0	0.4	1.0	0.3
JNA8	1.0	1.0	0.1	0.5	1.0	0.8	0.7	0.2	0.0	0.3	0.7	1.0	1.0	0.7	0.5	0.2	1.0	0.3
JNA7	0.7	0.9	1.0	1.0	0.6	0.1	0.3	0.3	0.9	0.4	0.2	0.6	1.0	0.3	0.4	0.7	0.3	1.0
JNA2	0.2	0.1	0.2	0.5	0.8	0.1	0.0	0.2	0.0	0.0	0.0	1.0	0.1	0.3	0.5	0.2	0.1	0.0
JNA14	0.7	0.5	0.8	1.0	0.7	0.9	0.0	0.2	0.5	0.3	0.2	1.0	0.8	0.3	0.4	1.0	0.5	0.3
JNA15	0.8	0.5	0.8	0.2	0.9	0.1	0.4	1.0	0.1	0.4	0.3	0.6	0.7	0.5	0.0	0.4	0.7	0.8
JNA6	0.6	0.2	0.0	0.5	0.1	0.0	0.5	1.0	0.0	0.5	0.4	0.6	0.4	0.1	0.1	0.2	1.0	0.5
JNA17	0.9	0.7	0.1	1.0	0.6	0.9	1.0	0.8	0.3	0.9	0.8	0.9	0.1	0.5	1.0	0.1	0.8	0.9
JNA9	1.0	1.0	0.1	0.6	0.2	0.0	0.5	1.0	0.1	1.0	0.5	0.3	0.0	0.5	0.6	1.0	1.0	0.0
JNA10	0.7	0.7	0.0	0.9	0.5	0.0	0.6	0.8	0.0	1.0	0.3	1.0	0.3	0.1	0.8	0.7	0.0	0.6
JNA4	0.6	0.9	0.2	0.2	0.0	0.1	0.6	0.2	0.5	0.7	0.6	1.0	0.1	0.3	0.0	0.7	0.0	1.0
JNA18	0.7	0.0	0.3	0.8	0.0	0.0	0.1	0.0	0.6	0.7	0.6	0.9	0.0	0.0	0.1	0.0	0.0	0.7
JNA11	0.4	0.6	0.3	0.2	0.1	0.0	0.0	0.5	0.3	0.9	0.1	0.8	0.6	0.4	0.1	0.0	0.6	0.5

Table 2.12 Correlation of mRNA-ATAC clusters in rat and human NPC differentiation

name	JDR1	JDR8	JDR10	JDR2	JDR5	JDR9	JDR7	JDR11	JDR12	JDR3	JDR4	JDR6
JDA8	0.3	0.7	1.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
JDA9	0.0	0.0	0.8	0.1	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.0
JDA13	0.0	0.8	0.0	0.4	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0
JDA4	0.3	0.1	0.0	1.0	0.3	0.1	0.7	0.1	0.1	1.0	0.0	0.8
JDA5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
JDA10	0.3	0.1	0.6	0.7	0.5	0.4	0.0	0.1	0.3	0.1	0.6	1.0
JDA6	0.4	1.0	0.5	0.3	0.0	0.9	0.0	0.1	0.8	0.0	0.9	0.0
JDA7	0.2	0.6	0.8	0.6	0.0	0.5	0.3	0.0	0.0	0.0	0.3	0.0
JDA11	0.0	0.0	0.1	0.2	0.0	0.0	0.2	0.0	0.4	0.0	0.0	0.0
JDA14	0.0	0.0	0.2	0.1	0.0	0.0	0.7	0.3	0.5	0.0	0.0	0.0
JDA15	0.0	0.0	0.6	0.1	0.0	0.0	0.5	0.0	0.5	0.0	0.2	0.0
JDA1	0.4	0.1	0.6	0.3	0.0	0.3	0.7	1.0	0.1	0.8	0.0	0.0
JDA2	0.0	1.0	0.2	0.7	0.0	0.0	0.7	0.5	0.8	0.0	0.0	0.0
JDA3	0.2	0.1	0.6	0.2	0.6	0.2	0.1	0.5	0.9	0.0	0.1	0.6
JDA12	0.0	0.7	0.3	0.8	0.8	0.7	0.3	0.0	1.0	0.1	1.0	0.0

Table 2.13 Correlation of mRNA-ATAC clusters in rat and human DE differentiation

2.7 Methods

2.7.1 Embryonic stem cell maintenance

H1 (male) human embryonic stem cells (ESCs) were obtained from WiCell and maintained in STEMCELL Technologies TeSR-E8 medium on growth factor reduced (GFR) Corning Matrigel. Cells were routinely passaged every 2-3 days with 0.5mM EDTA in dPBS. DAc8 (male) rat embryonic stem cells were purchased from Rat Resource and Research Center (RRRC), University of Missouri. Cells were first maintained on mouse embryonic fibroblast (MEF) feeders plated on 0.1% gelatin. MEF medium was made with GMEM, 10% FBS,1% GlutaMAX and 1% Pen/Strep. Rat ES cells were cultured in a 1:1 mixture of DMEM/F12 and Neurobasal medium supplemented with N2, HEPES (1M), B27, GlutaMax-I, Insulin, CHIR99021 (3mM), PD0325901 (1mM), 2ß-ME, Y-27632 (5mM), and hLIF(10ug/ml). MEF cells were plated at least one day before plating rat ES cells. Rat ES cells were passaged every 4-6 days with Accutase.

2.7.2 Definitive endoderm (DE) differentiation on monolayer *in vitro*

Human ES cells were differentiated using the STEMdiff Definitive Endoderm Kit (TeSR-E8 Optimized) from STEMCELL Technologies [29]. Rat ES cells were differentiated following an optimized mouse protocol [12–14]. For rat, cells were first transferred from MEFs to gelatin coated plate and they were seeded in rat ES medium for 4-6 hours. Then the medium was changed to the NDiff N2B27 basal medium supplemented with PD0325901 (1uM), CHIR99021 (3uM) and mLIF (100U/ml) (in 50ml, 49.939 ml of NDiff, 1ul mLIF, 10ul PD03, 50ul CHIR) on 0.1% gelatin. After 2-3 days, medium was changed to NDiff medium supplemented with Activin A, Fgf4, Heparin, kinase inhibitor (PI103), and CHIR99021 (in 50ml, 49.88ml of NDiff medium, 10ul of 100ug/ml ActivinA, 5ul of 100ug/ml Fgf4, 50ul of 1mg/ml of Heparin, 5ul of 1mM PI103, 50ul of

3uM CHIR). After 2 days of differentiation, medium was changed to DMEM/F12 with N2, B27-VA, L-glutamine, 2ß-ME, BSA, ActivinA, Fgf4, Heparin, Egf, PI103 and CHIR99021 (in 50ml,48.8295ml of DMEM/F12,250ul of N2, 500ul of B27-VA, 250uL-glutamine, 50ul 1000x 2ß-ME, 0.025g BSA, 10ul 100ug/ml ActivinA, 5ul 100ug/ml Fgf4, 50ul 1mg/ml Heparin, 0.5ul100ug/ml Egf, 5ul 1mM PI103 and 50ul of 3uM CHIR) for 5 days.

2.7.3 Neural Progenitor Cell (NPC) Differentiation on monolayer *in vitro*

Human ES cells were differentiated using an adapted, previously established protocol [28]. Briefly, cells were passaged and plated on matrigel and allowed to reach 30-40% confluence. After which media was changed to NPC medium (1:1 IMDM/F12 supplemented with NEAA, N2, B27, PSA, ALK2 and ALK3 inhibitor [0.2uM LDN193189], ALK5 inhibitor [10uM SB431542], and GSK3 Inhibitor [3uM CHIR99021]) for 8 days. Rat ESCs were also differentiated using a modified, previously published protocol [23]. For rat, cells were first transferred from MEFs to gelatin coated plate and they were seeded in rat ESC medium for 4-6 hours before changing to 1:1 IMDM/F12 differentiation medium supplemented with NEAA, PSA, N2, B27, LDN193189 (200 nM, Sigma-Aldrich), with FGF2 (20 ng/ml, Sigma-Aldrich) added to the media from day 2-5.

2.7.4 RNA-seq library construction

Total RNA was extracted once a day during differentiation using the RNeasy kit (QIAGEN). RNA was converted to cDNA using the SmartSeq2 protocol [96]. Libraries were constructed by using the Nextera DNA Flex Library Prep or Illumina DNA Prep kit (Illumina). Libraries were quality-controlled prior to sequencing using Agilent 2100 Bioanalyzer profiles and normalized using the KAPA Library Quantification Kit (Illumina). The libraries were sequenced using Illumina NextSeq500 platform and sequenced to a depth of around 10 million reads per sample.

2.7.5 ATAC-seq library construction

ATAC-seq samples were collected from the same pool of cells collected daily for RNA-seq and flash frozen based on the omni-ATAC protocol [97]. Around 75,000 cells were used per replicate and libraries were size-selected between 150-500bp using electrophoresis as done previously [98]. In short, gels were viewed and cut between 150-500bp and DNA was extracted from gel pieces using QIAquick Gel Extraction Kit (QIAGEN). Libraries were normalized using the KAPA Library Quantification Kit (Illumina). The libraries were sequenced using Illumina NextSeq500 platform and sequenced to a depth of around 60-100million reads per sample.

2.7.6 Gene expression analysis

Raw reads were mapped to hg38 (human), and rn6 (rat) using STAR (version2.5.1b) [99] using defaults except with a maximum of 10 mismatches per pair, a ratio of mismatches to read length of 0.07, and a maximum of 10 multiple alignments. Quantitation was performed using RSEM (version1.2.31) [100] with the defaults, and results were output in transcripts per million (TPM) and counts. Batch effects between species were corrected by using limma removebatcheffect [101]. Clustering of differentially expressed genes across the time-course was done by using maSigPro [44] with alpha of 0.05 for multiple hypothesis testing and a false discovery rate of 0.05%. Gene ontology analysis was done by using Metascape [102].

2.7.7 ATAC-seq data processing and analysis

Raw reads were mapped to hg38 (human) and rn6 (rat) using bowtie2 [100,103].Reads mapped to ChrM were discarded and PCR duplicates were removed by using Picard [104]. HOMER/4.7 [105] was used to call candidate regulatory elements (cRE). It was first used for calling 150 bp narrow peaks and then 500 bp broad peaks. Then narrow and broad peaks were

merged into a single peak list and they were further filtered by overlapping with ENCODE "blacklist" regions. Peaks that were shown in both replicates were considered as reproducible cRE. Reads coverage was calculated using bedtools [106] for each region and they were further normalized by the size of the library and the size of peaks. Normalized coverages in inter-species comparisons were further batch corrected for the technical batch effects by using limma removebatcheffect [101]. The differential cCRE regions through differentiation time-course were identified by using maSigPro [44] with alpha=0.05 and FDR<0.05.4.6.7 de novo motif enrichment analysis denovo motif calling was performed using HOMER/4.7 [105] using masked genomes for each species. The size of motif was set as "-len 6, 8, 10, 12, 15, 120" with at most 3 mismatches.

2.7.8 Contingency Tables and X2 test

RNA and ATAC gene overlaps were counted for each cluster pair and a $\chi 2$ test was used to determine significance for each cluster overlap. To study the enrichment of accessibility regions in different mRNA clusters for Figure 2E and Figure 3E an applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[18 × 24]) was used (Table S4 and S7). Figure 6A overlaps had an applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[18 × 18]; Table S12) and Figure 6C overlaps had an applied P-value of 10^{-4} (Bonferroni corrected P-value: 0.05/[12 × 15]; Table S13).

2.7.9 Footprint Calling and GRN Construction

ATAC-seq reads from duplicates were merged to achieve at least 100-120 million reads for footprints calling. Footprints were called in chromatin regions from clusters with significant overlap with mRNA clusters by using Wellington algorithm from pyDNase [107] with parameters "-

A -fp 6, 31, 1 -sh 7, 36, 4 -fdr limit -2", 1%FDR. Then motifs were scanned within footprints regions by using homer [105].

2.6.10 Inter-species analysis

Inter-species pairwise comparisons were performed by aligned identified cCRE between species in a reciprocal manner using UCSC liftOver [108] on genomic assemblies in two species. Each of the species was used as an anchor species and the cCRE were mapped to the other 2 species with 50% minimum map ratio. Regions failing to be mapped in any of the other genomes were considered as unaligned regions. To identify conserved regions between two species, regions having orthologous regions overlapped in the second species with at least 1 bp were collected. Final pair wise conserved regions were confirmed by doing this comparison reciprocally between species.

2.8 References

1. Davidson, E. H. et al. A genomic regulatory network for development. Science 295, 1669–1678 (2002).

2. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4936–4942 (2005).

3. Hardison, R. C. & Taylor, J. Genomic approaches towards finding cis-regulatory modules in animals. *Nat. Rev. Genet.* **13**, 469–483 (2012).

4. Berthelot, C., Villar, D., Horvath, J. E., Odom, D. T. & Flicek, P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat Ecol Evol* **2**, 152–163 (2018).

5. Jørgensen, F. G. *et al.* Comparative analysis of protein coding sequences from human, mouse and the domesticated pig. *BMC Biol.* **3**, 2 (2005).

6. Stergachis, A. B. *et al.* Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* **515**, 365–370 (2014).

7. Benitah, S. A. & Frye, M. Stem cells in ectodermal development. *Journal of Molecular Medicine* vol. 90 783–790 (2012).

8. Houston, D. W. Vertebrate Axial Patterning: From Egg to Asymmetry. *Adv. Exp. Med. Biol.* **953**, 209–306 (2017).

9. D'Amour, K. A. *et al.* Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat. Biotechnol.* **23**, 1534–1541 (2005).

10. D'Amour, K. A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat. Biotechnol.* **24**, 1392–1401 (2006).

11. Wang, A. *et al.* Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).

12. Kim, P. T. W. *et al.* Differentiation of mouse embryonic stem cells into endoderm without embryoid body formation. *PLoS One* **5**, e14146 (2010).

13. Yasunaga, M. *et al.* Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat. Biotechnol.* **23**, 1542–1550 (2005).

14. Mfopou, J. K. *et al.* Efficient definitive endoderm induction from mouse embryonic stem cell adherent cultures: a rapid screening model for differentiation studies. *Stem Cell Res.* **12**, 166–177 (2014).

15. Kanton, S. *et al.* Organoid single-cell genomic atlas uncovers human-specific features of brain development. *Nature* **574**, 418–422 (2019).

16. Bunina, D. *et al.* Genomic Rewiring of SOX2 Chromatin Interaction Network during Differentiation of ESCs to Postmitotic Neurons. *Cell Syst* **10**, 480–494.e8 (2020).

17. Shang, Z. *et al.* Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation. *Gigascience* **7**, (2018).

18. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304.e18 (2018).

19. Zhang, S. *et al.* Open chromatin dynamics reveals stage-specific transcriptional networks in hiPSC-based neurodevelopmental model. *Stem Cell Res.* **29**, 88–98 (2018).

20. Forrest, M. P. *et al.* Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. *Cell Stem Cell* **21**, 305–318.e8 (2017).

21. Morrison, G., Scognamiglio, R., Trumpp, A. & Smith, A. Convergence of cMyc and β-catenin on Tcf7l1 enables endoderm specification. *EMBO J.* **35**, 356–368 (2016).

22. Carpentier, A. *et al.* Hepatic differentiation of human pluripotent stem cells in miniaturized format suitable for high-throughput screen. *Stem Cell Res.* **16**, 640–650 (2016).

23. Alsanie, W. F. *et al.* Specification of murine ground state pluripotent stem cells to regional neuronal populations. *Sci. Rep.* **7**, 16001 (2017).

24. Yap, M. S. *et al.* Neural Differentiation of Human Pluripotent Stem Cells for Nontherapeutic Applications: Toxicology, Pharmacology, and In Vitro Disease Modeling. *Stem Cells Int.* **2015**, 105172 (2015).

25. Muratore, C. R., Srikanth, P., Callahan, D. G. & Young-Pearse, T. L. Comparison and optimization of hiPSC forebrain cortical differentiation protocols. *PLoS One* **9**, e105807 (2014).

26. Milani, P. *et al.* Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Sci. Rep.* **6**, 25474 (2016).

27. Lamas, N. J. *et al.* Neurotrophic requirements of human motor neurons defined using amplified and purified stem cell-derived cultures. *PLoS One* **9**, e110324 (2014).

28. Porterfield, V. Neural Progenitor Cell Derivation Methodologies for Drug Discovery Applications. *Assay Drug Dev. Technol.* **18**, 89–95 (2020).

29. Ramme, A. P., Faust, D., Koenig, L. & Marx, U. Generation of four integration-free iPSC lines from related human donors. *Stem Cell Res.* **41**, 101615 (2019).

30. Woodland, H. R. & Zorn, A. M. The core endodermal gene network of vertebrates: combining developmental precision with evolutionary flexibility. *Bioessays* **30**, 757–765 (2008).

31. Teo, A. K. K. *et al.* Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* **25**, 238–250 (2011).

32. Ryan, K., Garrett, N., Mitchell, A. & Gurdon, J. B. Eomesodermin, a key early gene in Xenopus mesoderm differentiation. *Cell* **87**, 989–1000 (1996).

33. Conlon, F. L., Fairclough, L., Price, B. M. J., Casey, E. S. & Smith, J. C. Determinants of T box protein specificity. *Development* **128**, 3749–3758 (2001).

34. Dwyer, D. *The Pharmacology of Neurogenesis and Neuroenhancement*. (Elsevier, 2007).

35. Veronika Fedorova, Tereza Vanova, Lina Elrefae, Jakub Pospisil, Martina Petrasova, Veronika Kolajova, Zuzana Hudacova, Jana Baniariova, Martin Barak, Lucie Peskova, Tomas Barta, Marketa Kaucka, Michael Killinger, Josef Vecera, Ondrej Bernatik, Lukas Cajanek, Hana Hribkova, Dasa Bohaciakova. Differentiation of neural rosettes from human pluripotent stem cells in vitro is sequentially regulated on a molecular level and accomplished by the mechanism reminiscent of secondary neurulation. *Stem Cell Res.* **40**, 101563 (2019).

36. Siller, R. *et al.* Development of a rapid screen for the endodermal differentiation potential of human pluripotent stem cell lines. *Sci. Rep.* **6**, 1–14 (2016).

37. Lu, J. *et al.* Single-cell RNA sequencing reveals metallothionein heterogeneity during hESC differentiation to definitive endoderm. *Stem Cell Res.* **28**, 48–55 (2018).

38. Wang, Z., Oron, E., Nelson, B., Razis, S. & Ivanova, N. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440–454 (2012).

39. Sandberg, C. J., Vik-Mo, E. O., Behnan, J., Helseth, E. & Langmoen, I. A. Transcriptional profiling of adult neural stem-like cells from the human brain. *PLoS One* **9**, e114739 (2014).

40. Abernathy, D. G. *et al.* MicroRNAs Induce a Permissive Chromatin Environment that Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts. *Cell Stem Cell* **21**, 332–348.e9 (2017).

41. Jørgensen, O. S. Neural cell adhesion molecule (NCAM) as a quantitative marker in synaptic remodeling. *Neurochem. Res.* **20**, 533–547 (1995).

42. Tang, B. L. The Potential of Targeting Brain Pathology with Ascl1/Mash1. *Cells* **6**, (2017).

43. Owa, T. *et al.* Meis1 Coordinates Cerebellar Granule Cell Development by Regulating Pax6 Transcription, BMP Signaling and Atoh1 Degradation. *J. Neurosci.* **38**, 1277–1294 (2018).

44. Conesa, A., Nueda, M. J., Ferrer, A. & Talón, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102 (2006).

45. Zhu, B., Carmichael, R. E., Valois, L. S., Wilkinson, K. A. & Henley, J. M. The transcription factor MEF2A plays a key role in the differentiation/maturation of rat neural stem cells into neurons. *Biochemical and Biophysical Research Communications* vol. 500 645–649 (2018).

46. Cho, E.-G. *et al.* MEF2C Enhances Dopaminergic Neuron Differentiation of Human Embryonic Stem Cells in a Parkinsonian Rat Model. *PLoS One* **6**, e24027 (2011).

47. Dietrich, J.-B. The MEF2 family and the brain: from molecules to memory. *Cell Tissue Res.* **352**, 179–190 (2013).

48. YAP/TAZ enhance mammalian embryonic neural stem cell characteristics in a Tead-dependent manner. *Biochem. Biophys. Res. Commun.* **458**, 110–116 (2015).

49. Shalizi, A. *et al.* A calcium-regulated MEF2 sumoylation switch controls postsynaptic differentiation. *Science* **311**, 1012–1017 (2006).

50. Hamilton, K. A. *et al.* Mice lacking the transcriptional regulator Bhlhe40 have enhanced neuronal excitability and impaired synaptic plasticity in the hippocampus. *PLoS One* **13**, e0196223 (2018).

51. Zhang, S. & Cui, W. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* **6**, 305–311 (2014).

52. M. Uittenbogaard, A. C. Expression of the bHLH transcription factor Tcf 12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. *Brain Res. Gene Expr. Patterns* **1**, 115 (2002).

53. Simone Mesman, M. P. S. Tcf12 Is Involved in Early Cell-Fate Determination and Subset Specification of Midbrain Dopamine Neurons. *Front. Mol. Neurosci.* **10**, (2017).

54. Pierfelice, T. J., Schreck, K. C., Eberhart, C. G. & Gaiano, N. Notch, neural stem cells, and brain tumors. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 367–375 (2008).

55. Chu, L.-F. *et al.* Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).

56. Boareto, M., Iber, D. & Taylor, V. Differential interactions between Notch and ID factors control neurogenesis by modulating Hes factor autoregulation. *Development* **144**, 3465–3474 (2017).

57. David Wotton, K. T. Functions of TGIF Homeodomain Proteins and Their Roles in Normal Brain Development and Holoprosencephaly. *Am. J. Med. Genet. C Semin. Med. Genet.* **178**, 128 (2018).

58. Moore, S. *et al.* Distinct Regulatory Mechanisms Act to Establish and Maintain Pax3 Expression in the Developing Neural Tube. *PLoS Genet.* **9**, e1003811 (2013).

59. Ziller, M. J. *et al.* Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* **518**, 355 (2015).

60. Sinner, D., Rankin, S., Lee, M. & Zorn, A. M. Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes. *Development* **131**, 3069–3080 (2004).

61. Shimosato, D., Shiki, M. & Niwa, H. Extra-embryonic endoderm cells derived from ES cells induced by GATA factors acquire the character of XEN cells. *BMC Dev. Biol.* **7**, 80 (2007).

62. Mesman, S., Bakker, R. & Smidt, M. P. Tcf4 is required for correct brain development during embryogenesis. *Mol. Cell. Neurosci.* **106**, 103502 (2020).

63. Martinez-Barbera, J. P. *et al.* Regionalisation of anterior neuroectoderm and its competence in responding to forebrain and midbrain inducing activities depend on mutual antagonism between OTX2 and GBX2. *Development* **128**, 4789–4800 (2001).

64. Sherwood, R. I., Chen, T.-Y. A. & Melton, D. A. Transcriptional dynamics of endodermal organ formation. *Dev. Dyn.* **238**, 29–42 (2009).

65. Dollé, P. Developmental expression of retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7**, e006 (2009).

66. Symmank, J., Gölling, V., Gerstmann, K. & Zimmer, G. The Transcription Factor LHX1 Regulates the Survival and Directed Migration of POA-derived Cortical Interneurons. *Cereb. Cortex* **29**, 1644–1658 (2019).

67. Scott, C. E. *et al.* SOX9 induces and maintains neural stem cells. *Nat. Neurosci.* **13**, 1181–1189 (2010).

68. Faucourt, M., Houliston, E., Besnardeau, L., Kimelman, D. & Lepage, T. The pitx2 homeobox protein is required early for endoderm formation and nodal signaling. *Dev. Biol.* **229**, 287–306 (2001).

69. Golonzhka, O. *et al.* Pbx regulates patterning of the cerebral cortex in progenitors and postmitotic neurons. *Neuron* **88**, 1192 (2015).

70. Okamoto, S., Sherman, K., Bai, G. & Lipton, S. A. Effect of the ubiquitous transcription factors, SP1 and MAZ, on NMDA receptor subunit type 1 (NR1) expression during neuronal differentiation. *Brain Res. Mol. Brain Res.* **107**, (2002).

71. Faial, T. *et al.* Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142**, 2121–2135 (2015).

72. Turner, D. A., Rué, P., Mackenzie, J. P., Davies, E. & Martinez Arias, A. Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biol.* **12**, 63 (2014).

73. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).

74. Stadhouders, R., Filion, G. J. & Graf, T. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature* **569**, 345–354 (2019).

75. Zhao, L. *et al.* Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat. Commun.* **10**, 3640 (2019).

76. Lemaire, P., Darras, S., Caillol, D. & Kodjabachian, L. A role for the vegetally expressed Xenopus gene Mix.1 in endoderm formation and in the restriction of mesoderm to the marginal zone. *Development* **125**, 2371–2380 (1998).

77. Colas, A. *et al.* Mix.1/2-dependent control of FGF availability during gastrulation is essential for pronephros development in Xenopus. *Dev. Biol.* **320**, 351–365 (2008).

78. Arzate-Mejía, R. G., Recillas-Targa, F. & Corces, V. G. Developing in 3D: the role of CTCF in cell differentiation. *Development* **145**, (2018).

79. Kaczynski, J., Cook, T. & Urrutia, R. Sp1- and Krüppel-like transcription factors. *Genome Biol.* **4**, 206 (2003).

80. Harrison, S. M., Houzelstein, D., Dunwoodie, S. L. & Beddington, R. S. Sp5, a new member of the Sp1 family, is dynamically expressed during development and genetically interacts with Brachyury. *Dev. Biol.* **227**, 358–372 (2000).

81. Eggenberger, J., Blanco-Melo, D., Panis, M., Brennand, K. J. & tenOever, B. R. Type I interferon response impairs differentiation potential of pluripotent stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 1384–1393 (2019).

82. Lodato, M. A. *et al.* SOX2 co-occupies distal enhancer elements with distinct POU factors in ESCs and NPCs to specify cell state. *PLoS Genet.* **9**, e1003288 (2013).

83. van de Leemput, J. *et al.* CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. *Neuron* **83**, 51–68 (2014).

84. Yang, S.-H. *et al.* ZIC3 Controls the Transition from Naive to Primed Pluripotency. *Cell Rep.* **27**, 3215–3227.e6 (2019).

85. Kalkan, T. *et al.* Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **144**, 1221–1234 (2017).

86. Forrest, M. P., Waite, A. J., Martin-Rendon, E. & Blake, D. J. Knockdown of human TCF4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. *PLoS One* **8**, e73169 (2013).

87. Mukhtar, T. *et al.* Tead transcription factors differentially regulate cortical development. *Sci. Rep.* **10**, 4625 (2020).

88. Cebola, I. *et al.* TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat. Cell Biol.* **17**, 615–626 (2015).

89. Song, J., Ugai, H., Kanazawa, I., Sun, K. & Yokoyama, K. K. Independent repression of a GC-rich housekeeping gene by Sp1 and MAZ involves the same cis-elements. *J. Biol. Chem.* **276**, 19897–19904 (2001).

90. Guillaume, D. J. & Zhang, S.-C. Human embryonic stem cells: a potential source of transplantable neural progenitor cells. *Neurosurg. Focus* **24**, E3 (2008).

91. Murry, C. E. & Keller, G. Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell* **132**, 661–680 (2008).

92. Takahashi, S., Kobayashi, S. & Hiratani, I. Epigenetic differences between naïve and primed pluripotent stem cells. *Cell. Mol. Life Sci.* **75**, 1191–1203 (2018).

93. Hong, J. *et al.* A focused microarray for screening rat embryonic stem cell lines. *Stem Cells Dev.* **22**, 431–443 (2013).

94. Yiangou, L., Ross, A. D. B., Goh, K. J. & Vallier, L. Human Pluripotent Stem Cell-Derived Endoderm for Modeling Development and Clinical Applications. *Cell Stem Cell* **22**, 485–499 (2018).

95. Ramsdell, C. M. *et al.* Comparative genome mapping of the deer mouse (Peromyscus maniculatus) reveals greater similarity to rat (Rattus norvegicus) than to the lab mouse (Mus musculus). *BMC Evol. Biol.* **8**, 65 (2008).

96. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

97. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

98. Ramirez, R. N. *et al.* Dynamic Gene Regulatory Networks of Human Myeloid Differentiation. *Cell Syst* **4**, 416–429.e3 (2017).

99. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

100. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

101. Smyth, G. K. limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 doi:10.1007/0-387-29362-0_23.

102. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

103. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

104. Website. http://broadinstitute.github.io/picard.

105. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* vol. 38 576–589 (2010).

106. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

107. Piper, J. *et al.* Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Research* vol. 42 11272–11272 (2014).

108. Kent, W. J. The Human Genome Browser at UCSC. *Genome Research* vol. 12 996–1006 (2002).

Chapter 3

Single-Cell Transcriptomics of hypothalamic CRH-expressing neurons informs mechanisms by which Early-Life Adversity sculpts stress-responses enduringly

Notes: (1) Dr. Annabel Short and I equally contributed to the material in this chapter. I built the single cell libraries, sequenced samples, and analyzed the single cell data. She designed and performed the experiments. She provided interpretation of the results.

(2) Aidan L Pham assisted on the experiments and mouse upkeep.

(3) Matthew T Birnie provided interpretation of the results

(4) Jessica L Bolton provided interpretation of the results

(5) Dr. Ali Mortazavi and Dr. Tallie Z. Baram conceived the idea and provided continued support and guidance throughout the project.

Chapter 3

Single-Cell Transcriptomics of hypothalamic CRH-expressing neurons informs mechanisms by which Early-Life Adversity sculpts stress-responses enduringly

3.1 Abstract

Mental and cognitive health, as well as vulnerability to neuropsychiatric disorders, involve the interplay of genes with the environment during sensitive developmental periods. Genetic and environmental factors contribute to the development and maturation of neurons, synapses, and the resulting brain circuits. Within the hypothalamus, early-life experiences cause changes in the number of excitatory synapses onto corticotropin-releasing hormone (CRH)-expressing neurons in the paraventricular nucleus (PVN). Further, such synaptic changes suffice to induce enduring epigenomic changes within these cells, influencing programs of gene expression. However, the epigenetic mechanisms by which early-life experiences orchestrate transcriptional programs within individual CRH-expressing neurons, with enduring functional consequences, remain unknown. We utilize animal models of an impoverished environment and unpredictable maternal care (in a limited bedding and nesting [LBN] paradigm), which provokes major alterations in cognitive and emotional outcomes. We are focused on the change in gene expression profiles of stress-sensitive CRH-neurons in the PVN following early-life adversity (ELA). Because of the known heterogeneity of CRH-expressing neuronal populations, we are using single-cell RNA sequencing (RNA-seq) to determine differential gene expression associated with early-life adversity and establish the upstream mechanisms and downstream consequences. We found that CRH-expressing populations of neurons in the PVN can be clustered by both their gene expression profiles as well as by their neurotransmitter phenotype. ELA significantly modified gene expression programs in some neuronal clusters more than others, reducing programs of neuronal development and differentiation and enhancing gene

families involved in responses to stress and inflammation. The use of single-cell transcriptomics revealed that ELA impacts gene expression profiles in a cell-type specific manner, with distinctive influences on different clusters and subpopulations of CRH neurons.

3.2 Introduction

Development of the postnatal brain and brain circuits are regulated by both genetics and environmental stimuli through epigenetic mechanisms [1]. During the critical period of brain development, environmental stimuli have increased impact on brain and circuit development [2,3]. In humans, ELA such as poverty, parental loss, abuse, and neglect, is associated with enduring consequences including impairments in cognition and emotional regulation [4–9]. ELA has been associated with adverse long-term changes in the regulation of the stress response [3,10,11], however, the mechanistic underpinnings of how these experiences are encoded are yet to be elucidated.

Stress-sensitive corticotropin-releasing hormone (CRH)-expressing neurons residing in the hypothalamus are candidate mediators of the long-lasting effects of ELA, because they contribute to both behavioral and hormonal responses to stress [12]. CRH is the key central regulator of the hormonal stress response, which is quickly secreted by neurons of the hypothalamic paraventricular nucleus (PVN), causing subsequent secretion of adrenocorticotropic hormone (ACTH) and glucocorticoids [12–15]. During development, the stress sensitive CRH neurons of the PVN are particularly sensitive to changes in the environment. First, CRH expression levels themselves are modulated by experiences including stress [16,17] and augmented maternal care, which causes a repression in CRH expression. Environment, such as augmented sensory signals from the mother also leads to a reduction in excitatory synapses to the CRH cells within the hypothalamus [18,19], while an unpredictable early-life

environment causes enhanced excitatory glutamatergic transmission in the same cell population [20,21].

The regulation of the stress response pathway is critical during development to ensure proper levels of glucocorticoids. High levels of glucocorticoids during critical periods may adversely affect brain development [12,22,23]. ELA alters gene expression in the brain, resulting in abnormal responses to stress that persist into adulthood, including altered CRH synthesis and secretion [12,24–34]. However, the mechanisms by which ELA orchestrates transcriptional programs within individual CRH-expressing neurons, with enduring functional consequences, remain unknown. Specifically, it is unknown whether ELA influences gene expression programs in the cells, whether these are specific to sub-populations of CRH-expressing neurons and whether the transcriptional changes are associated with a stress-sensitive phenotype.

The population of CRH positive PVN neurons has traditionally been divided into three major subpopulations; parvocellular cells, magnocellular neurons expressing vasopressin and oxytocin neurons, and autonomic CRH neurons. The parvocellular neurons have axons projecting to the median eminence. In response to stress, parvocellular CRH neurons secrete Crh into the pituitary portal circulation [35,36]. The vasopressinergic and oxytocinergic neurons send projections to the posterior pituitary and release these peptides to the peripheral circulation [35,37]. The autonomic neurons have projections to the brainstem and spinal cord [38,39]. These neurons are involved in regulating the connection between the sympathetic nervous system and the adrenal medulla [35]. CRH positive neurons in the PVN express different genes and neurotransmitters based on their location and function [40–43]. By investigating the transcriptome of individual neurons we can start to understand the complexity of the fundamental building blocks of neural networks and how they might be influenced by ELA. We

further gain a deeper understanding of how this complex system is impacted by early-life stress, contributing to enduring phenotypic changes.

Here we employed a model of ELA in mice, the limited bedding/nesting (LBN) model, which recapitulates the effects of adverse environments on brain development [44,45]. The LBN model is a well-characterized, preclinical model of ELA [46] that mimics an impoverished environment, inducing unpredictable maternal care, during a critical period of brain development [45,47,48]. This induces enduring long-term consequences [45] including reduced reward seeking [49–52], impairments in hippocampal memory [46,53–57], and development of fear inhibition pathways [58,59]. We focused on the change in gene expression profile of CRH-neurons in the hypothalamic PVN following ELA. Due to the known heterogeneity of CRH-expressing neuronal populations, we employed single cell RNA sequencing (scRNA-seq) to determine differential gene expression associated with ELA and establish the upstream mechanisms and downstream consequences.

3.3 Results

3.3.1 CRH-expressing neurons in the developing mouse hypothalamus belong to distinct populations with unique gene-expression profiles

To profile changes between an early life stress model and Control CRH+ neurons from the hypothalamic paraventricular nucleus (PVN) we used the Limited Bedding and Nesting (LBN) model (Fig. 3.1A). We collected PVN cells from male mice between Postnatal Day (PND)10-12 for fluorescence activated cell sorting (FACS) for both Control (CTL) (Fig. 3.1-A) and LBN (Fig. 3.1-A2) mice to capture gene expression changes that occur in response to ELA. Briefly, single cell suspensions of a heterogenous cell population were generated from PVN punches were FACS sorted and SmartSeq2 [60] was performed on single cells (Fig. 3.1B; Methods) to achieve deep sequencing. In total, 602 cells were profiled from 88 animals (Table

3.1). This low yield may result some biases in cell heterogeneity but these biases should be similar for both CTL and ELA conditions, so we do not expect it to significantly influence the results of this study. We used the Uniform Manifold Approximation and Projection (UMAP) using the number of nearest neighbors dimensional reduction technique to cluster cells based on their cellular identity 511 cells expressed at least 1000 genes expressed in at least 4 CTL cells or 4 LBN cells (Fig. 3.2A-D). Age, weight, or batch effects do not drive the clusters (Fig. 3.2A-C). Since our dissociation method included small margins of cells outside the PVN, we looked at the expression of *Pgr15l*, which is not normally expressed in the PVN (Fig. 3.2D) [42] and excluded 81 cells based on *Pgr15l* expression. From the remaining 430 cells, we classified cell types by cell type specific markers; *C1qa* and *Csf1r* (microglia, Fig. 3.3A), *Olig1* and *Olig2* (oligodendrocyte, Fig. 3.3B), *Apoe* and *Vim* astrocyte, Fig. 3.3C), *Flt1* and *Cldn5* (endothelial cells, Fig. 3.3D), and *Snap25* (neurons, Fig. 3.3E). We chose to move forward with clusters 2-5 due to their expression of *Snap25*, a neuronal marker, and *Crh* (Fig. 3.4). A subset of 254 cells from clusters 2-5 was sub-clustered to identify sub-types of *Crh* positive PVN neurons (Fig. 3.5A). The *Crh* positive neurons were split into 5 sub-populations or clusters (Fig. 3.5B). The PVN is known to contain many distinct subdivisions [39,61]. Cells from both ELA condition and CTL condition are equally distributed between the 5 clusters ($X^2_{(4,\ N=254)}$ =7.42, p=0.11) (Fig. 3.6A). In cluster 1 there are 53 cells from the ELA condition and 27 cells from the CTL condition, in cluster 2 there are 23 cells from the ELA condition and 25 cells from the CTL condition, in cluster 3 there are 26 cells from the ELA condition and 21 cells from the CTL condition, in cluster 4 there are 32 cells from the ELA condition and 13 cells from the CTL condition, in cluster 5 there are 19 cells from the ELA condition and 16 cells from the CTL condition (Fig. 3.6A). Cells do not cluster based on mouse weight (Fig. 3.6B), batch (Fig. 3.6C), or age (Fig. 3.6D).

As neuronal subtypes can be distinguished by their major neurotransmitter, we further analyzed the data based on the expression of glutamate and GABA markers. Cells expressing *Slc17a6* (above 1 transcripts per million [TPM]), which encodes the glutamate transporter VGLUT2, were classified as glutamatergic (Fig. 3.7A), and the presence of *Gad2* (Fig. 3.7B) and/or *Slc32a1* (Fig. 3.7C), the GABA synthetic enzyme Glutamate decarboxylase 2 (GAD2) and transporter (VAGT) respectively, was used to identify GABAergic cells. As reported previously [62,63], the majority of cells are glutamatergic, however, there is a small subset of cells which express GABAergic markers, GAD2 and VGAT (Fig. 3.7D). There is some overlap of clustering based on neurotransmitter expression and the clustering, as seen on the UMAP. Cells in clusters 2 and 5 have the majority of the GABAergic marker expression whereas the majority of cells in other clusters have expression of the glutamatergic marker, *Slc17a6* (Fig. 3.7F).The expression of GABAergic markers may be transient in the cells which also express the glutamate marker, *Slc17a6* [64], so we classified the cells with both glutamatergic markers and GABAergic markers as glutamatergic (light green, Fig. 3.7D).

To determine if neuronal subtypes were altered by adverse early-life experiences, we compared the number of ELA and CTL cells within each neurotransmitter sub-cluster. Cells from both Glutamatergic and GABAergic categories are equally distributed between ELA and CTL cells ($X^2_{(1, N=254)} = 2.10$, $p = 0.15$) (Fig. 3.7E). Clusters were at least somewhat biased by whether the cells expressed Glutamatergic and GABAergic markers, as cells in clusters 2 and 5 have the majority of the GABAergic marker expression whereas the majority of cells in other clusters have expression of the glutamatergic marker, Vglut2 ($X^2_{(4, N=254)} = 52$, $p < 0.00001$) (Fig. 3.7F).

The glutamatergic cells of the PVN can be further defined into neuronal subtypes based on neuropeptide or other marker expression [42]. As described previously [40]**,** we noticed that the

glutamatergic expressing cells could be further sub-clustered by expression of *Ntng1* (Fig. 3.8A) or Avp (Fig. 3.8B). Arginine vasopressin (AVP) is secreted by parvocellular and magnocellular neurons of hypothalamus, and acts to control ACTH secretion from the pituitary. Like CRH, AVP is a key regulator of the HPA axis and these hormones work together to stimulate the secretion of ACTH during times of stress [65]. Netrin G1 (*Ntng1*) is associated with fear and anxiety-like behaviors, as knockout *Ntng1* mice display less fear and anxiety-like behaviors than CTL mice [66]. *Ntng1* has also been shown to be expressed in specific CRH positive neurons in the PVN [40]. We further sub-clustered the glutamatergic, CRH neurons into *Avp* positive (dark green), *Ntng1* positive (light green), or *Avp* negative and *Ntng1* negative (turquoise) glutamatergic neurons, GABAergic neurons are in blue and cells with no expression of glutamatergic or GABAergic markers are shown in grey (Fig 3.8C). Seurat cluster identities are shown by the colored boundaries of each cell cluster (Fig 3.8C). The data suggests that *Crh* positive neurons of the PVN can be clustered based on expression of specific neurotransmitters and peptides.

3.3.2 Early-life adversity significantly impacts gene expression programs in hypothalamic CRH cells

Clustering did not separate ELA from CTL conditions; therefore, we first measured the global differential expression of the 153 ELA and 101 CTL cells. This analysis revealed marked differences in transcriptomic profiles between CTL and ELA cells. We found 28 genes with increased expression in CTL cells, or downregulated in ELA, passing a FDR filter of less than 0.1, including; *Cbx3, Tatdn1, Gnl3l, Irf2bp3, Rbm47*, and *Cox7c* (Fig. 3.9B, Table 3.2). Genes upregulated in neurons from ELA mice include; *Stmn1, Rpl10, Hsp90ab1, Vamp2, and Actg1* (Fig. 3.9B, Table 3.2).

All the genes passing the FDR<0.1 filter for differential expression are shown on the heatmap in Fig. 3.10. The heatmap is quantile and row normalized so the expression values are between -1 and 1, which highlights which cells have the highest expression and which cells have the lowest expression of a gene, respectively. CTL expression and ELA expression of each gene are separated. The top 10 genes on the heatmap have noticeably higher expression in the ELA condition. Whereas the next 14 genes have higher expression in cells from CTL mice. The heatmap highlights that these genes are clearly upregulated in the majority of cells in their respective condition. The data further demonstrates that the genes found to be differentially expressed in the ELA condition are downregulated in CTL cells and the genes that are upregulated in the CTL condition are downregulated in the ELA condition.

In order to further classify the genes that were upregulated in each condition we looked at pathway analysis for all genes passing the FDR<0.1 filter (Fig 11A and 11B). The pathway analysis for genes higher in the CTL condition suggest that these cells are undergoing normal cell development and function (Fig. 3.11A). For example, the *Chl1* interactions pathway where *Chl1* promotes neuron survival and includes genes including *Itga1*. The processing of *Smdt1* pathway is involved in regulation of mitochondrial calcium uniporter (MCU) complex and includes the gene *Micu3*.

The pathways found for genes higher in the ELA condition suggest significant perturbations following adversity during a sensitive developmental period. Compellingly, many neuronal type pathways are highlighted, the top highlighted pathway is axon guidance which includes genes including *Scn3b*, *Rpl39*, *Hsp90ab1*, and *Rpl10* (Fig. 3.11B). Other interesting pathways are Translocation of Glut4 to the plasma membrane with genes including *Actg1* and *Vamp2*. This pathway is interesting because it further suggests increased neuronal activity when compared to the CTL condition. Glut4 is a major transporter that mediates glucose

uptake. Glut4 may provide a mechanism through which neurons rapidly increase their glucose utilization during increases in neuronal activity or under conditions of cell stress. The genes found to be differentially expressed in a global analysis of all PVN neurons from either the ELA or CTL condition, as well as the pathways these genes are involved in, suggest that neurons in the CTL condition are undergoing normal neuronal development, whereas neurons from the ELA condition have either accelerated their maturation or otherwise have abnormal neuronal development and maturation. To investigate this further we classified neurons into sub-clusters by their expression.

3.3.3 Enrichment of genes in expression defined clusters suggest differing subpopulations of CRH cells

As there are specific neuronal subtypes of CRH positive neurons in the PVN that are also functionally different [40,61,63], we identified genes that were specific to each cluster (Table 3.3). The top 12 differentially expressed genes from each cluster are shown in a heatmap (Fig. 3.12). The columns are cells split by their cluster affiliation, and labeled with whether the cells are from the ELA or CTL conditions, and the age of the mice at collection. The rows are the top 12 genes that are split by the genes enriched in each cluster.

The top genes enriched in cluster 1 include *Nnat, Npm1*, but the majority of the top enriched genes in cluster 1 are *Rps* and *Rpl* genes including *Rps3a1, Rps27, Rps27a, Rps20, Rpl10, Rpl27a, Rpl13a*, and *Rpl7* (Fig.12). Ribosomes, which are RNA and protein structures that organize and catalyze translation, have two subunits: a ribosomal protein large subunit (Rpl) and a ribosomal protein small subunit (Rps) [67].

Cluster 2 has enriched genes including; *Dusp3, Gsk3b, Slc39a1, Sec23ip, Inpp4a,* and *Actb* (Fig.12). These genes are involved in many neuronal and cellular functions, including;

glutamate excitotoxicity and cell death, glycogen synthesis activation, activation of neuronal mitochondrial, axon guidance, and redox pathways, and protection of neurons from degeneration [68–71].

Cluster 3 has enriched genes including; *Junb, Npff, Dlk1, Snhg11, Snd1,* and *Ece1* (Fig. 3.12). Genes found to be associated with cellular functions like cell proliferation, differentiation, and neuronal cell death, synaptic drive, postnatal development of hypothalamus, and cell proliferation and angiogenesis [72–74].

Ddah1, Tcf7l2, Zic2, Wnt3, Ntng1, and *Prox1* are among the genes enriched in cluster 4 (Fig. 3.12). These genes are known to be involved in neuronal identity and circuitry, axonal navigation, neuronal migration, and neurogenesis [75–77].

Cluster 5 has enriched genes such as; *Tubb2a, Tubb2b, Vamp1, Rgs10*, and *Nefl* (Fig. 3.12). Genes enriched in this cluster were representative of axon development, microtubule development, and synaptic transmission [78–80].

Taken together, the genes enriched in each cluster suggest that each cluster may have a slightly different function, consistent with the notion that these clusters may be different sub-populations of PVN CRH neurons. To investigate the effect that the ELA condition has on these potential sub-populations of PVN CRH neurons we performed differential expression analysis on ELA and CTL mice within each cluster.

3.3.4 ELA impacts gene expression programs differentially in distinct expression defined subpopulations of CRH cells

Using the expression defined clustering in Fig. 3.5B we determined the differential expression between ELA and CTL conditions within each cluster of cells (Fig. 3.13A-E).

Interestingly, when using an FDR<0.1 cut-off for significant enrichment, we find enriched genes in clusters 1 (Fig. 3.13A), 2 (Fig. 3.13B), and 4 (Fig. 3.13D), but not in clusters 3 (Fig. 3.13C) and 5 (Fig. 3.13E). Because the number of cells in each cluster did not correlate with the number of differentially expressed genes, we concluded that the differential expression was not based on the number of cells in each cluster (Fig. 3.6A).

In cluster 1, we found 7 genes enriched in the ELA condition, including; *Stmn1, Nnat, Rpl10, Vamp2, Hsbp1, Sez6l2,* and *Calr*. Only 1 gene was differentially upregulated in the CTL condition, *Igf2bp3*, for the same cluster (Fig. 3.13A). For cluster 2, there were 3 enriched genes in the CTL condition (*Dctn4, Cds1, Wwc2*) and none in the ELA condition (Fig. 3.13B). These genes did not show up in the global analysis, demonstrating the increased sensitivity provided by the single-cell and cluster approaches. Cluster 4 has 2 enriched genes in the CTL condition (*Cbx3 and Ndufa8*), and none in the ELA condition (Fig. 3.13D). In addition to genes that were shared with the differential analysis of ELA with all cells, there were genes that were specifically enriched in cluster 1 following ELA *Nnat, Hsbp1, Sez6l2*, and *Calr*. The enriched genes from cluster 1 suggest cells from the ELA condition in cluster 1 may be undergoing neuronal ER stress and are trying to mitigate neural damage caused by stress [81–85].

3.3.5 ELA impacts gene expression programs differentially in distinct neurotransmitter expression defined subpopulations of CRH cells

As described above, we sub-clustered the cells based on whether or not they expressed *Slc17a6* (VGLUT2), a glutamatergic marker, or *Slc32a1* (VGAT) and/or *Gad2* without VGLUT2, and are GABAergic (Fig. 3.7D). In order to determine the effect of early life experiences on the neurotransmitter defined sub-clusters, we performed differential expression between neurons from the ELA and CTL conditions within each cluster (Fig. 3.14A). Differentially expressed

genes (that pass an FDR<0.1 filter) between CTL and ELA were found to be neuronal subtype specific (Fig. 3.14A-D, Table 3.4). There were no differentially expressed genes that passed the FDR<0.1 filter in the GABAergic sub-cluster or the sub-cluster composed of cells expressing neither neurotransmitter marker (Fig. 3.14C-D).

We found 23 differentially expressed genes enriched in the CTL condition passing the FDR<0.1 filter in glutamatergic cells (Fig.14B). 14 of these genes also were differentially expressed in the global CTL condition from section 3.3.2. Some of these shared genes include *Cbx3, Cox7c, Gnl3l, Gabra2, Inpp4a*, and *Tatdn1*. It is interesting that these genes are enriched in the CTL condition (and downregulated in the ELA condition) which since these genes are involved in differentiation, maturation, and possibly normal cellular function of CRH positive PVN neurons [86–89].

Some genes that did not show up in the global analysis enriched in the CTL condition but were apparent in the neurotransmitter-defined cluster analyses are *Nrxn2, Cycs, Sec23ip, Sap18b, Pin1,* and *Ndufa8*. Overall, since these genes are downregulated in the ELA condition, suggests that neurons in the ELA condition may have altered energy production or are not able to properly maintain excitability or synaptic transmission [90–92]. The majority of the genes upregulated in the CTL condition suggest that these cells are undergoing normal development and maturation.

We found 16 differentially expressed genes enriched in the ELA condition passing the FDR<0.1 filter in glutamatergic cells (Fig. 3.14B). We found 9 of these genes also were differentially expressed in the global ELA condition from section 3.3.2. Some of these shared genes include; *Actg1, Eif1, Hsp90ab1, Psma6, Rps13, Rpl10,* and *Stmn1*. Taken together, the genes with shared enrichment in PVN neurons and the glutamatergic PVN neurons of the ELA

condition suggest the cells have increased protein synthesis as well as removal of damaged proteins [89,93], suggesting that these neurons may be undergoing stress.

There are also some unique genes found enriched in the glutamatergic sub-clusters in the ELA condition that were not found to be significant in the global differential expression, including; *Dynll1* and *Sec24b*. These genes suggest the glutamatergic specific ELA neurons may have dysfunctional transcription and translation mechanisms [94–96].

To further investigate the differentially expressed genes in both ELA and CTL conditions in the glutamatergic sub-cluster, GO analysis using metascape [97] and pathway analysis with Reactome [98] were performed (Fig. 3.15A-D). The genes enriched in the CTL condition (Fig. 3.15A) includes GO terms including mitochondrial ATP synthesis coupled electron transport, which include *Cox7c, Cycs, Ndufa8, Ubb*, and *Pin1*. Positive regulation of intrinsic apoptotic signaling pathway includes the genes *Skil, Ubb*, and *Gsdme*. Finally, regulation of protein binding which includes the genes *Gnl3l* and *Hypk*. The CTL condition pathways (Fig. 3.15B) include RHO GTPases Activate NADPH Oxidases which includes *Pin1*. The citric acid (TCA) cycle and respiratory electron transport including genes *Ndufa8* and *Cox7c*. These GO terms and pathways further suggest that the neurons in the CTL condition are functioning well, with normal cell functions highlighted.

The GO terms (Fig. 3.15C) and pathway analysis (Fig. 3.15D) of the genes enriched in the ELA condition suggest that these cells are maturing faster than the cells in the CTL condition. GO terms include positive regulation of cellular component movement which include the genes *Actg1*, *Atp5a1*, and *Stmn1* (Fig. 3.15C). This GO term is a parent term of regulation of axon guidance and regulation of actin filament-based movement. Another interesting GO term is cell morphogenesis involved in neuron differentiation which includes *Hsp90ab1*, *Stmn1*, and *Sec24b*. The pathways (Fig. 3.15D) include Regulation of expression of SLITs and ROBOs

which has been shown to be involved in neuronal axon guidance [99]. Genes involved in this pathway include *Rsp24*, *Rpl36a*, *Rpl10*, *Rps13*, and *Psma6*. Other interesting pathways are Nervous system development which includes the genes *Actg1*, *Rpl39*, *Hsp90ab1*, and *Rpl27a*. These genes, GO terms, and pathways suggest that neurons from the ELA condition are actively maturing and migrating, possibly at a faster rate than the CTL cells.

3.3.6 ELA impacts gene expression programs differentially in distinct neurotransmitter and marker expression defined subpopulations of CRH cells

Given the significant changes observed in the glutamatergic cells, we analyzed the *Avp* or *Ntng1* sub-clusters of these cells (Fig. 3.8C) and examined the differential expression between neurons from the ELA and CTL conditions (Fig. 3.16A-B). Enriched genes that passed the FDR<0.1 filter were found in the CTL condition but not the ELA condition in the glutamatergic *Avp* positive sub-cluster. These 5 genes enriched in the CTL cluster included; *Tatdn1, Ndufa8, Cbx3, Rap2b*, and *Bc1*. All of these genes are also found in either the enriched global CTL condition or glutamatergic CTL condition except for *Rap2b*. *Rap2B* is a pro-survival p53 target gene following DNA damage [100]. Enrichment of this gene in glutamatergic cells in the CTL condition suggests that even in the CTL condition there are some cellular stressors during the sensitive developmental period.

There were 4 enriched genes that passed the FDR<0.1 filter in the CTL condition in the glutamatergic *Ntng* positive sub-cluster (Fig. 3.16B). These genes included *Ubb, Ndufa8, Cox7c*, and *Pold4*. Pold4 is a novel gene that was not discovered in any of the previous sub-cluster analyses. *Pold4* encodes a subunit of DNA polymerase delta which has polymerase activity, exonuclease activity, and plays a critical role in DNA replication and repair. The upregulation of this gene suggests that these cells are undergoing processes requiring DNA

repair. A more likely explanation is that the ELA condition represses DNA repair, thus leading to enrichment of this gene in CTL cells.

Next, we investigated differentially expressed ELA genes in the glutamatergic *Ntng1* positive sub-cluster *Stmn1*, *Rpl10*, and *Actg1* were found enriched in the ELA condition (Fig. 3.16B) in the *Ntng1* positive sub-cluster and were also enriched in the other sub-clusters (Global, Glutamatergic, and Glutamatergic *Ntng1* positive sub-clusters). By investigating the differential expression of individual sub-clusters we can start to understand that cells that express different neurotransmitters are also functionally different from each other and may be involved in different functions in the PVN during ELA.

All the genes passing the FDR<0.1 filter for differential expression are shown on all the CRH positive PVN neurons in the heatmap in Fig. 3.17 and all the glutamatergic cells in Fig. 3.18. The heatmap is quantile and row normalized so the expression values are between -1 and 1. CTL expression and ELA expression of each gene are separated. The top 12 genes on the heatmap have enrichment in the ELA condition. Whereas the following 19 genes have higher expression in cells from the CTL condition. The heatmap shows the clear enrichment of differentially expressed genes from the two conditions.

3.4 Discussion

In this study we provide the first comparison of single cell transcriptomics analysis of CRH positive PVN neurons between ELA and CTL conditions. We found multiple sub-types of CRH neurons within the PVN that were differentially affected by ELA and differing transcriptomic profiles with varied experience during sensitive developmental periods. Genes altered with experience are associated with neuronal differentiation and activation in response to the environment. Although the heterogeneity of the CRH positive PVN neuron population is known,

to date[40], little is known about the changes in gene expression that occur in this population of neurons in response to ELA.

3.4.1 Adversity during sensitive periods enriches genes associated with neuronal maturation, differentiation, and stress

Based on a comparison of gene expression profiles of PVN CRH positive neurons from the ELA condition and the CTL condition as well as pathway analysis of all enriched genes, our results suggest that ELA leads to accelerated neuronal differentiation. Control cells are engaged in growth and express high levels of *Cbx3*, possibly involved in promoting early neural differentiation and inhibiting other fates [101]. *Cbx3,* a gene important for neural differentiation, was also significantly upregulated in CTL cells. *Cbx3* increases expression of neuronal genes and inhibits expression of genes specific to other fates [101]. This upregulation of *Cbx3* in CTL cells and the concurrent downregulation in ELA cells, suggests altered neuronal maturation of CRH cells in the PVN following ELA. The *Tatdn1* long non coding RNA (lncRNA) plays a role in cell proliferation, adhesion, and migration [102], and its downregulation in ELA cells suggests that neurons from the ELA condition may be inhibited in these processes. *Gnl3l* is expressed in developing neurons, increasing in expression as maturation occurs and is involved in proliferation [103,104]. *Cbx3*, *Tatdn1,* and *Gnl3l* are upregulated in CTL conditions when compared to the ELA condition, which suggests that ELA alters maturation of these Crh positive PVN neurons. These results are comparable with previous studies, where long term stress has been shown to cause dendritic simplification, cell death, and decreased neurogenesis in adults[105,106,107,108]. ELA shortens the process of neural development, thought to manifest as developmental delay [109].

Additionally, cells from the ELA condition are enriched for genes that play a role in neuronal maturation, differentiation, and mitigating neuronal stress. Suggesting that neurons from the ELA condition are undergoing more stress due to increases in neuronal activity which may be causing a faster rate of neuronal maturation. *Stmn1*, part of the Stathmin family of neurotubule regulators, has been implicated in controlling HPA-axis activation [110,111]. Variations in this gene can also cause changes in microtubule stability, synaptic plasticity and memory [110,111]. *Hsp90ab1* functions as a heat shock gene which suggests that these cells are undergoing more stress than the CTL cells [112]**.** Vesicle-associated membrane protein 2 (*Vamp2* or *Syb2*), is one of the most common synaptic vesicle proteins in the brain and is very important in Ca2+-triggered vesicle exocytosis [80]. Actin gamma 1 (*Actg1*) is involved in axon elongation and plasticity [113]. Genes involved in synaptic vesicles, axon elongation, and microtubule stability are increased in cells in the ELA condition compared to CTL, suggesting that these cells may be maturing faster than in the CTL condition. Previous studies have shown that, in the hippocampus, ELA led to earlier expression of genes associated with neural maturation and decreased expression of markers of cell proliferation[109]. Changes of expression of these markers suggests that ELA, in the hippocampus, causing neurons to prematurely switch from growth to neural maturation.3.4.2 Clustering of neurons identified sub-populations of neurons

Clustering using Seurat identified interesting differences in gene expression profiles of CRH positive PVN neurons, suggesting different neuronal populations. This is consistent with the literature suggesting that there are different sub-populations of PVN neurons in the hypothalamus, many of which express CRH [40–43]. To identify these populations, we examined changes in gene expression between the different clusters of neurons.

Cluster 1 was characterized by genes encoding ribosomal proteins. The process of translation is heavily controlled and regulated, and many Rpl genes also control cell proliferation, cell cycle arrest, and cell death [67].

Genes enriched in cluster 2 were representative of glutamate excitotoxicity and cell death, glycogen synthesis activation, activation of neuronal mitochondrial, axon guidance, and redox pathways, and protection of neurons from neuronal degeneration. Genes in this cluster included *Dusp3*, or *Vhr*, increases in *Dusp3* phosphatase activity leads to glutamate excitotoxicity and cell death caused by decreases in ERK activation [68]. *Gsk3b* was originally known to be insulin sensitive and involved in glycogen synthesis activation, but more recently it has been shown to regulate the activation of neuronal mitochondrial and redox pathways [71]. *Slc39a1* expression protects neurons from neuronal degeneration due to Zinc entry, excessive Zinc entry can cause apoptosis [70]. *Actb* encodes beta-actin, which plays an important role in growth cone mediated axon guidance in neuronal development [69].

Cluster 3 included expression of genes associated with cell proliferation, differentiation, and neuronal cell death, as well as genes associated with synaptic drive, postnatal development of hypothalamus, and cell proliferation and angiogenesis. Expression of genes in this cluster include *Junb* which is essential for neuronal differentiation, and the *Jun* transcription factor family is important in cell proliferation, differentiation, and neuronal cell death [72]. *Npff* has been shown to decrease synaptic dive to magnocellular vasopressin-secreting PVN neurons, as vasopressin release is blunted by *Npff*. In parvocellular PVN neurons, *Npff* presynaptically increases the excitability of these neurons [73]. *Dlk1* has been shown to play a role in postnatal development of hypothalamus. It is expressed in PVN dendrites and neuronal cell bodies of vasopressin and oxytocin PVN neurons in both adult and postnatal mice [74]. *Snhg11* and *Snd1* both are oncogenes and play a role in cell proliferation and angiogenesis [114,115].

Genes upregulated in Cluster 4 were primarily transcription factors. With genes involved in neuronal identity and circuitry, axonal navigation, neuronal migration, and neurogenesis. *Tcf7l2*, is a transcription factor in the *Tcf/Lef* family, which activates *Wnt* target genes and plays key roles in development. More recently, *Tcf7l2* has been shown to mediate establishment of neuronal identity and circuitry in the developing brain [76]. In later neuron maturation, *Zic2* controls the axonal navigation, but is also important for neuronal migration during earlier stages of neuro-development and the dorsal neural tube [77]. *Prox1*, a homeobox transcription repressor, promotes neurogenesis and has shown to regulate the transition of NPCs from a self-renewal state into neuronal differentiation [75].

Cluster 5 genes were representative of axon development, microtubule development, and synaptic transmission. Genes enriched in this cluster include the tubulin family of genes, including *Tubb2a* and *Tubb2b*, these are required for microtubule development which is essential for many cellular processes, especially during neural development [78]. *Vamp1* is involved in the presynaptic docking of vesicles, and is required for Ca2+-triggered synaptic transmission [80]. *Nefl*, neurofilament light, is involved in axon formation, *Nefl* mutants fail to develop proper axons during development, which leads to myelination and conduction defects [79]. The genes enriched in each cluster or potential sub-population of PVN neurons suggest that the cells in each cluster may have a slightly different function within the PVN. Further suggesting that these clusters may represent different sub-populations of PVN neurons.

3.4.3 ELA differentially affects neuronal subtypes of CRH expressing cells in the PVN

To investigate differences between the clusters, and sub-populations, we used differential expression analysis between the ELA and CTL conditions within each cluster.

Cluster 1, 2 and 4 were the only clusters that had differentially expressed enriched genes in either the ELA or CTL conditions.

Cluster 1 had some very interesting genes enriched in the ELA condition which suggested that these cells may be undergoing neuronal ER stress and are trying to mitigate neural damage caused by stress. For example, *Nnat* is normally expressed in the embryo during neurogenesis, but it is also expressed in the dendritic endoplasmic reticulum (ER) to regulate modulation of intracellular Calcium [116]. *Hsbp1* is a small heat shock protein which is neuroprotective in various neuronal diseases and after injury [82]. *Hsbp1* mutations have been shown to affect the stability and remodeling of microtubules, and disrupt cytoskeletal transport leading to neurodegeneration [83]. Seizure 6-like protein, *Sez6l2*, is part of the Sez6 family of proteins which, in mice, control synaptic connectivity and motor coordination [81,85]. Alzheimer's disease linked APP cleaving enzyme 1 (BACE1) has been shown to regulate members of the Sez6 family [81]. Calreticulin, *Calr*, is an ER chaperone, and in neurons, is located in axons in response to ER stress and calcium release. *Calr* has been shown to increase its expression in response to axonal injury in order to regulate axonal retraction and promote axonal regeneration [84].

Only three genes were enriched in the CTL condition in cluster 2 so it is hard to draw any robust conclusions. *Dctn4*, encodes Dynactin Subunit 4, the multi-subunit complex, Dynactin, is required for most cellular processes powered by the microtubule-based motor, dynein [117]. The protein family, Ww-and-C2-domain-containing (Wwc) is involved in cell differentiation, proliferation, regulation of organ growth, and are upstream of the hippo signaling pathway. More specifically, *Wwc2* is involved in embryogenesis and angiogenesis [118]. These two genes are consistent with augmented cell growth, including neurite extension and connectivity in the CTL condition. By contrast, *Cds1* is involved in the DNA damage checkpoint

in the cell cycle and is activated in p53-dependent response to DNA damage and cell cycle arrest [119].

3.4.4 ELA differentially affects cells which express different neurotransmitters

As the clusters were not representative of early-life experience, we further investigated differences between ELA and CTL conditions within cells that were expressing glutamatergic markers and GABAergic markers. We found many interesting differentially expressed genes between the ELA and CTL conditions in glutamatergic cells.

Upregulated in the control condition are genes including *Cox7c*, or cytochrome c oxidase (COX) catalyzes the electron transfer from reduced cytochrome c to oxygen and is a component of the mitochondrial respiratory chain [86]. Also upregulated is NADH dehydrogenase ubiquinone 1 α subcomplex (*Ndufa8*). *Ndufa8* is heavily involved in mitochondrial oxidative phosphorylation, and is found to be expressed in the hypothalamus [90]. *Cycs* is the somatic isoform of cytochrome C, which has been shown to have expression responsiveness to changes in neural activity. Together these genes support the notion of the use of this energetic pathway which is efficient and promotes growth, in the CTL condition. *Gabra2* is the alpha-2 subunit of the GABAA receptor (GABAAR), which in the brain regulates the fast inhibitory actions of GABA [87,88]. *Inpp4a* protects neurons from excitotoxic cell death, which is the process of apoptosis or neuronal damage by over activation of glutamate receptors [89].

Upregulated in the ELA condition of glutamatergic cells, *Eif1* function is involved in the initiation of translation, and is very important in all cells [89]. *Psma6* helps to form an essential component of the 20S core proteasome complex. This complex is involved in the degradation of proteins. *Psma6* also is involved in the formation of the 26S proteasome and thus ATP-dependent degradation of proteins and removes misfolded or damaged proteins [93]. *Rbs13*

encodes a ribosomal protein component of the 40S subunit which plays an important role in translation [93]. Also upregulated in the ELA condition is *Dynll1* is a regulator of axonal mitochondria anchorage [94], and play a role in complexes involving; myosin, neuronal nitric oxide synthase (nNOS), apoptotic factors, and transcription. Mutations in the *Dynll1* gene have been known to cause abnormal neurogenesis and apoptosis [95]. *Sec24b* is involved in protein transport from the ER to the Golgi apparatus. Mice with mutations in *Sec24b* exhibit neural tube closure defects resulting from decreased ER trafficking [96]. *Rpl10*, which is enriched in the ELA condition in glutamatergic cells, is a component of the 60S ribosomal subunit and is required for joining the 60S and 40S subunits in translation initiation. Mutations in *Rpl10* decrease a cells ability to perform translation, and mutations in this gene have been implicated in autism, suggesting *Rpl10* plays a role in brain development [120]. *Ttc3* is a novel gene that is not found to be enriched in the overall populations of cells from the ELA condition. *Ttc3* promotes neuronal actin polymerization through signaling pathways including RhoA among others. During neuronal differentiation, overexpression of *Ttc3* can prevent neurite extension and disrupt Golgi compactness [121].

In these cells, individual enriched genes as well as pathway and GO analysis suggests that neurons in the ELA condition may have altered energy production or are not able to properly maintain excitability or synaptic transmission. Whereas, the majority of genes enriched in the CTL condition point to normal development and maturation, and the use of the mitochondrial redox and TCA cycle for efficient energy.

3.4.5 Conclusion

In conclusion, ELA causes altered gene expression patterns in CRH expressing PVN neurons, which are associated with cellular processes such as ER stress, altered energy

production, and protection from injury. This may be an indication of abnormal increased number of excitatory synapses observed neuroanatomical in these cells after ELA. It is unclear whether these consequences will be sustained long term or if they are transient. It would be interesting to determine whether these changes are robust and sustained. Interestingly, enduring changes in the transcriptome of neurons in the hippocampus after ELA have been reported [24]. Further studies need to be conducted in order to answer these questions. In addition to sub-clusters of PVN CRH neurons, we also found populations of cells without CRH expression, and with markers for other non-neuronal cell types that might be interesting to investigate in the future. This work highlights the known heterogeneity of the CRH cells within PVN, and we have discovered that sub-clusters of these cells are particularly sensitive to adversity early in life. Understanding which cell types undergo transcriptional programming in response to early environmental signals, and how these experiences are encoded transcriptionally, is vital for identifying novel interventions to mitigate the enduring effects on brain development.

3.5 Figures

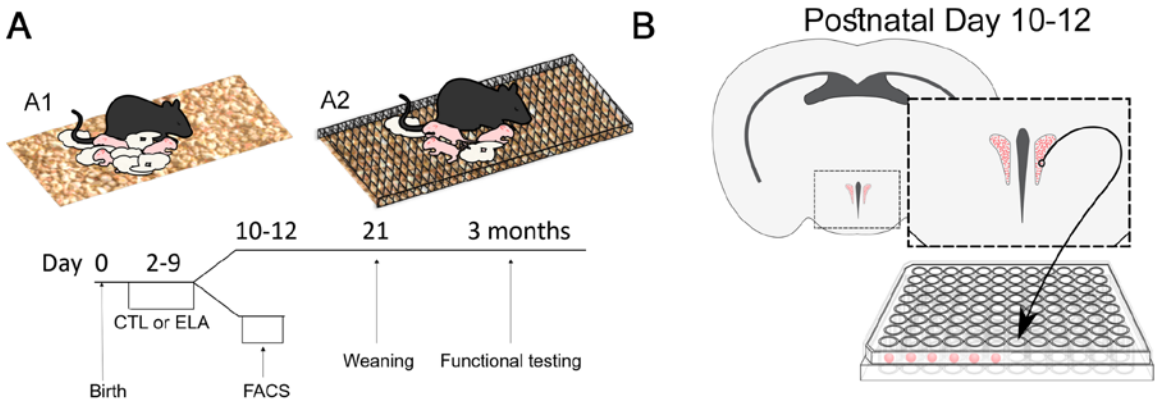

Figure 3.1 Limited bedding and nesting (LBN) model for ELA and experimental timeline.

A) A1 is the CTL model and A2 is the LBN model for ELA. Below A1 and A2 is the postnatal day timeline for ELA exposure and subsequent preparation of single cell suspension of PVN cells for FACS. B) Schematic of CRH positive PVN cells, in red, are isolated and FACS sorted into 96 well plates.

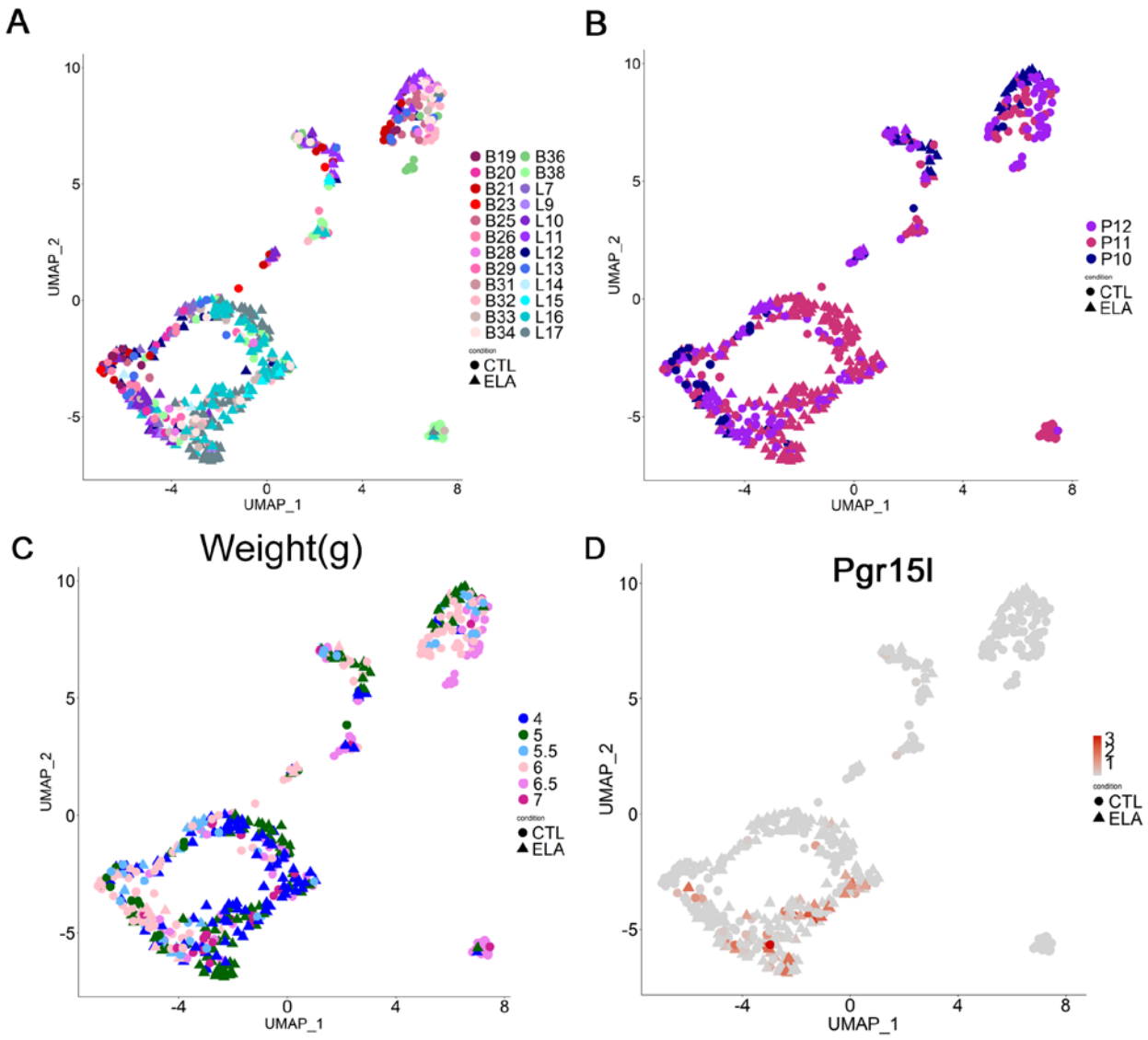

Figure 3.2 Characterization of batch, age, weight, and non-PVN marker *Pgr15l* expression of 511 cells

A) UMAP characterization of 24 batches across 511 cells to determine little or no batch effect between sorts. B) UMAP characterization of age the cells were harvested between postnatal day 10 - 12. C) UMAP characterization of weight in grams of mice on the day cells were harvested. D) Heatmap overlay of non-PVN *Pgr15l* expression over the UMAP of 511 cells to determine cell filter for *Pgr15l* expression. All UMAPs were generated using Seurat v3.

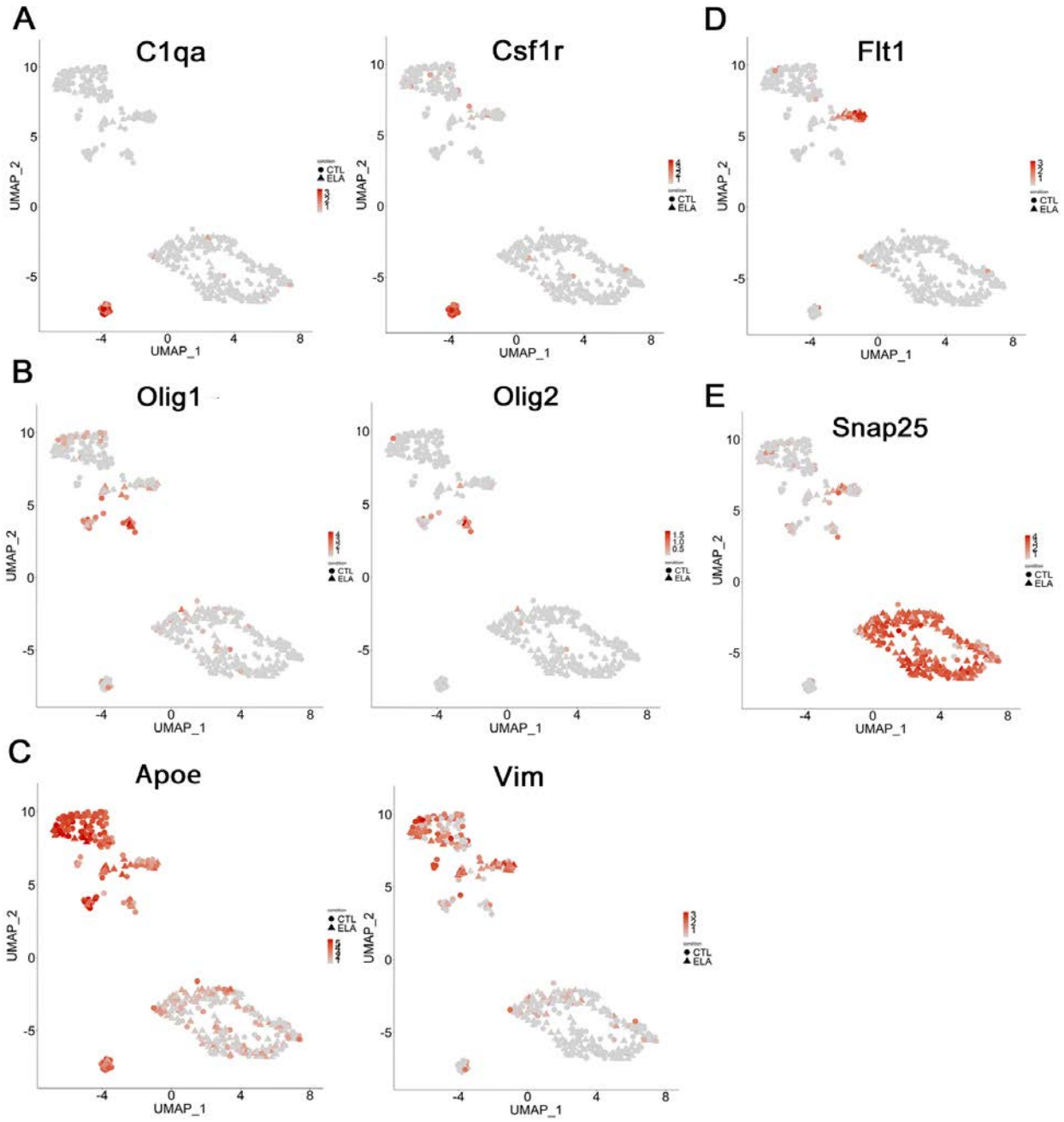

Figure 3.3 Classification of cell types in 430 cells by cell-type specific markers

A) UMAP heatmap of microglia markers, C1qa and Csf1r. B) UMAP heatmap of oligodendrocyte markers Olig1 and Olig2. C) UMAP heatmap of astrocyte markers Apoe and Vim. D) UMAP heatmap of the endothelial marker Flt1. D) UMAP heatmap of the neuronal marker Snap25. All UMAPs were generated using Seurat v3.

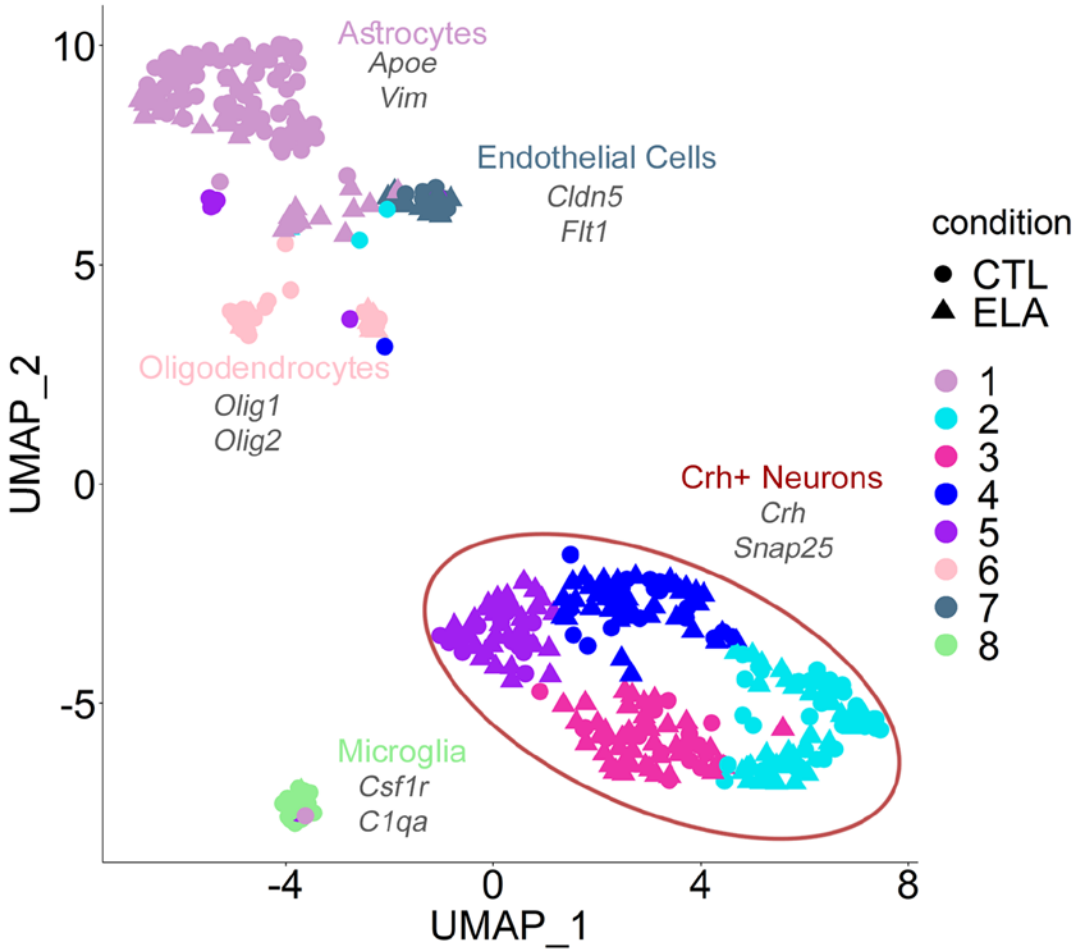

Figure 3.4 Summary of marker gene expression and *Crh* positive PVN neurons in 430 cells.

The marker gene expression from Figure 3.3 summarized on one UMAP. The *Crh* positive Snap25 positive will be sub-clustered and further analyzed in the following figures. All UMAPs were generated using Seurat v3.

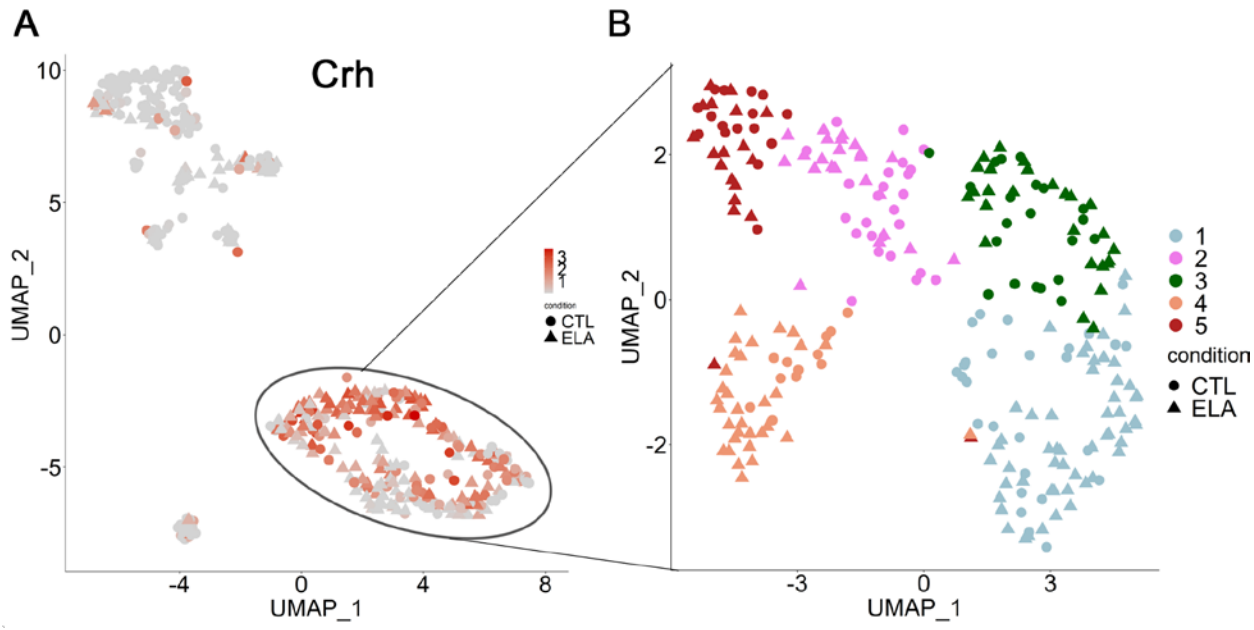

Figure 3.5 ***Crh*** **expression in 430 cells and sub-clustering of 254** ***Crh*** **positive PVN neurons.**

A) UMAP heatmap of *Crh* expression in 430 cells. The circled area are the 254 cells to be sub-clustered for further analysis. B) Sub-clustering of 254 *Crh* positive, *Snap25* positive PVN neurons. Cells clustered into 5 distinct clusters using Seurat v3.

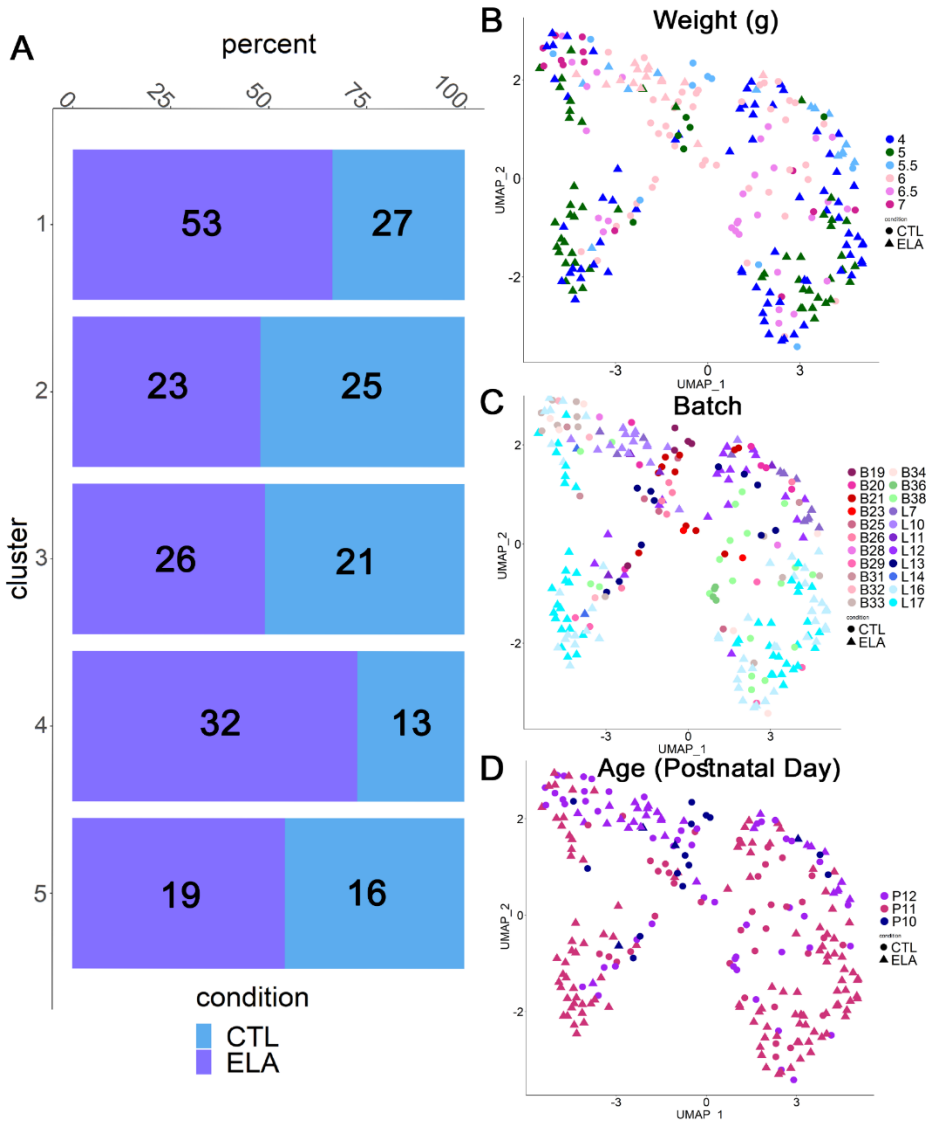

Figure 3.6 Percentage of cells from ELA condition and CTL condition in each cluster and characterization of weight, batch, and age over 254 *Crh* positive PVN neuronal cells.

A) Stacked bar plot of percentage of cells from the ELA condition and CTL condition in each cluster. Numbers on the bar represent the number of genes in each condition. B) UMAP characterization of weight in grams of mice on the day cells were harvested. C) UMAP characterization of batches across 254 cells. D) UMAP characterization of age at time of harvest.

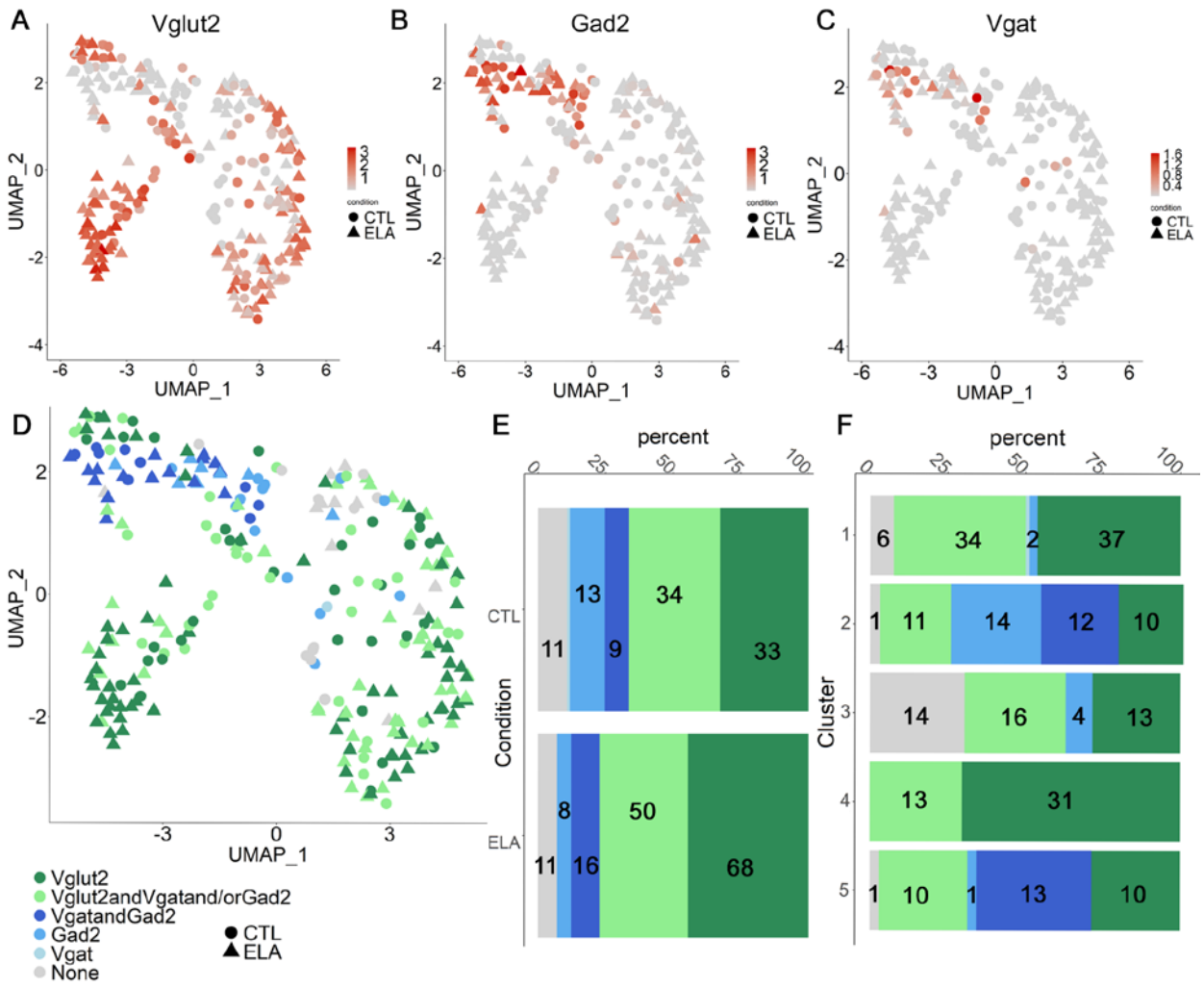

Figure 3.7 Characterization of glutamatergic and GABAergic neurotransmitters.

A) UMAP heatmap of glutamatergic marker, Vglut2. B) UMAP heatmap of GABAergic marker, GAD2. C) UMAP heatmap GABAergic marker, VGAT. D) Classification of neurotransmitter clusters based on expression profiles from Fig. 3.7A-C. E) Percentage of Glutamatergic and GABAergic cells in the ELA and CTL conditions. The numbers in each category are the numbers of cells in each category. F) Percentage of Glutamatergic and GABAergic cells in each expression defined Seurat cluster. The numbers in each category are the numbers of cells in each category.

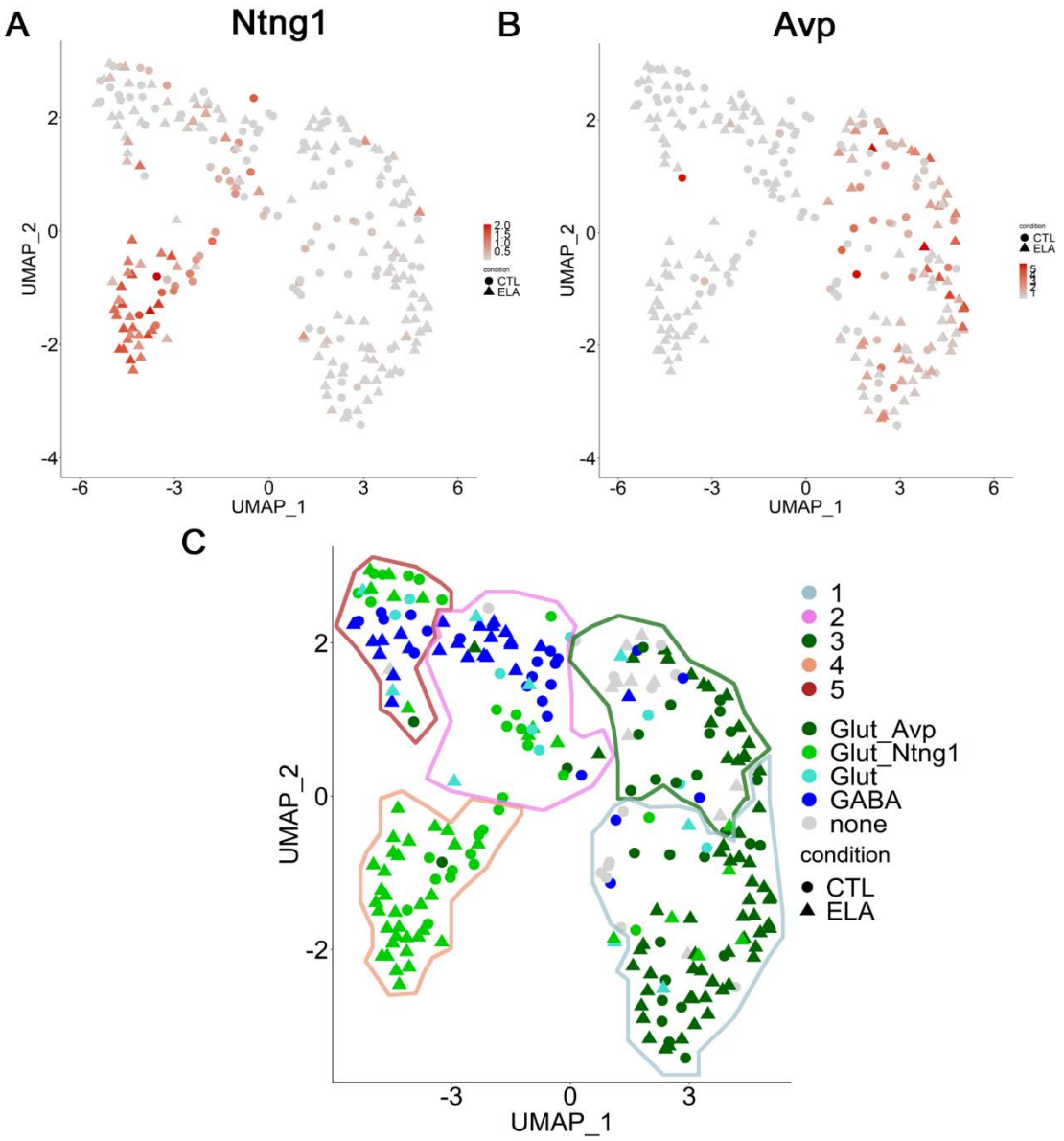

Figure 3.8 Further classification of glutamatergic, Vglut2 expressing, cells based on *Avp* expression or *Ntng1* expression.

A) UMAP heatmap of *Ntng1* expression in 254 *Crh* positive PVN neurons. B) UMAP heatmap of Avp expression in 254 Crh positive PVN neurons. C) Further Classification of glutamatergic cells based on expression profiles of Ntng1 and Avp. Seurat custer borders are drawn out to distinguish the difference between seurat clusters and neurotransmitter and marker-based clusters.

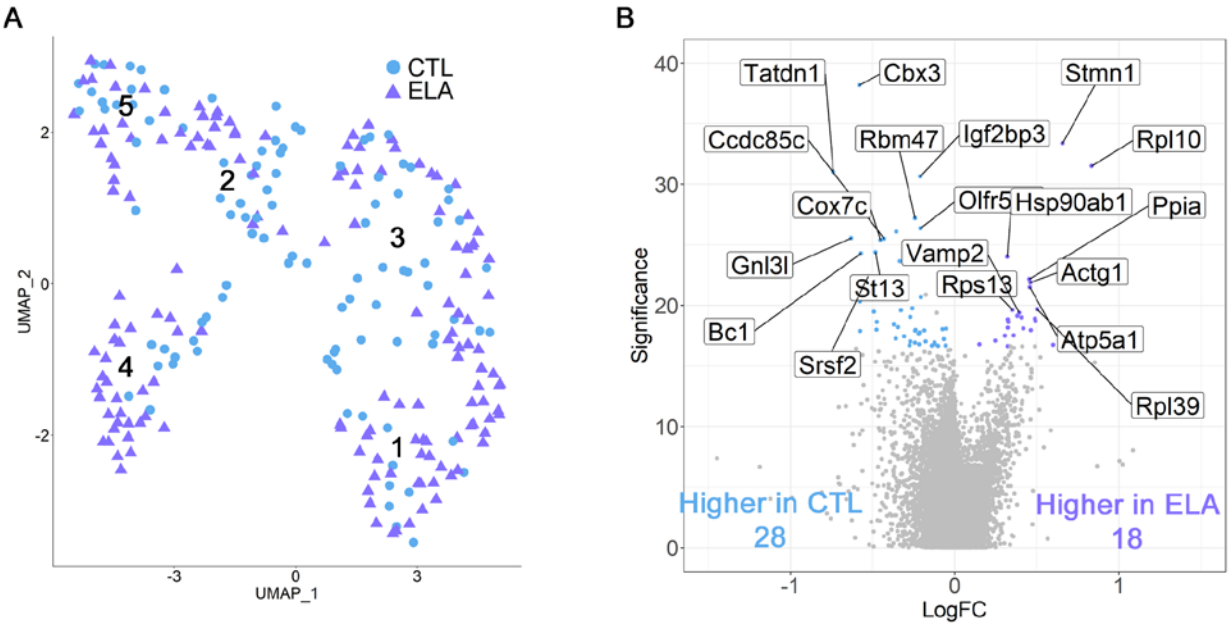

Figure 3.9 Global differential expression analysis of 254 *Crh* positive PVN neurons between the ELA and CTL condition.

A) Distribution of cells from the ELA and CTL condition represented via UMAP. The numbers are Seurat cluster numbers from Fig. 3.5B. B) Volcano plot of differential expression between the ELA and CTL conditions. There were 18 differentially expressed genes enriched in the ELA condition and 28 in the CTL condition passing an FDR<0.1 filter. Some of the top enriched genes are listed for both the CTL and ELA conditions.

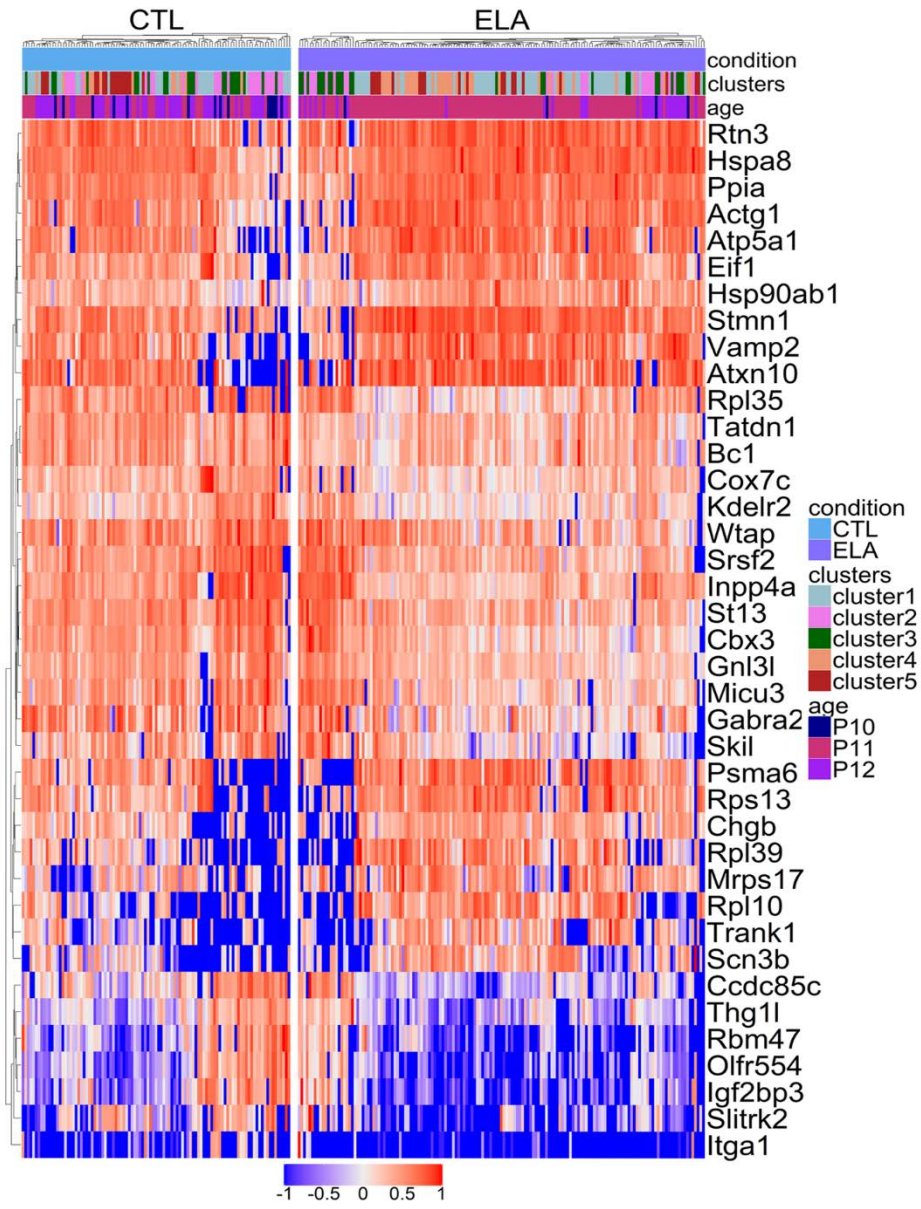

Figure 3.10 Heatmap of all 46 globally differentially expressed genes between the ELA and CTL conditions.

Row normalized expression values (TPMs) are shown on the color bar at the bottom of the heatmap. Cells are columns and rows are genes. Cells are split by whether they are from the CTL or ELA conditions. Cluster and age identities are also displayed in color bars at the top of the heatmap.

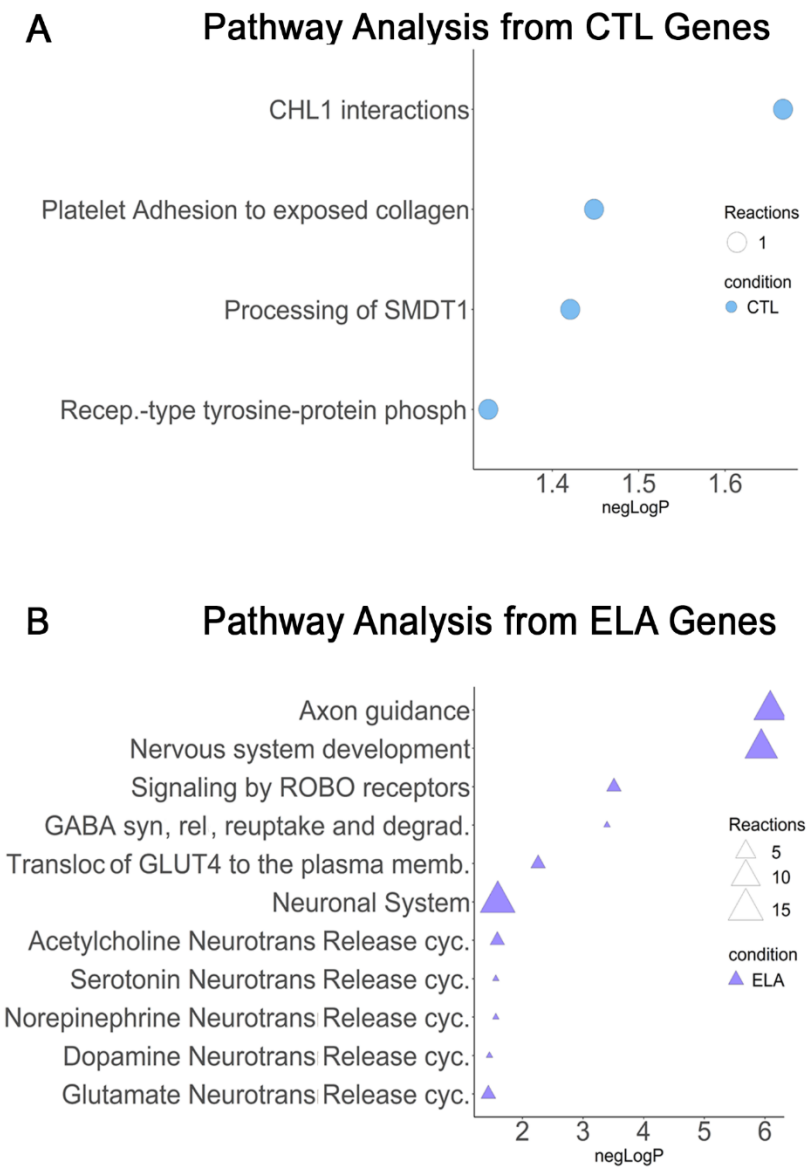

Figure 3.11 Pathway analysis of differentially expressed genes in the ELA and CTL conditions using Reactome.

A) Genes enriched in the CTL condition were analyzed using Reactome. B) Genes enriched in the ELA condition were analyzed using Reactome. Only pathways passing an FDR<0.05 filter were considered.

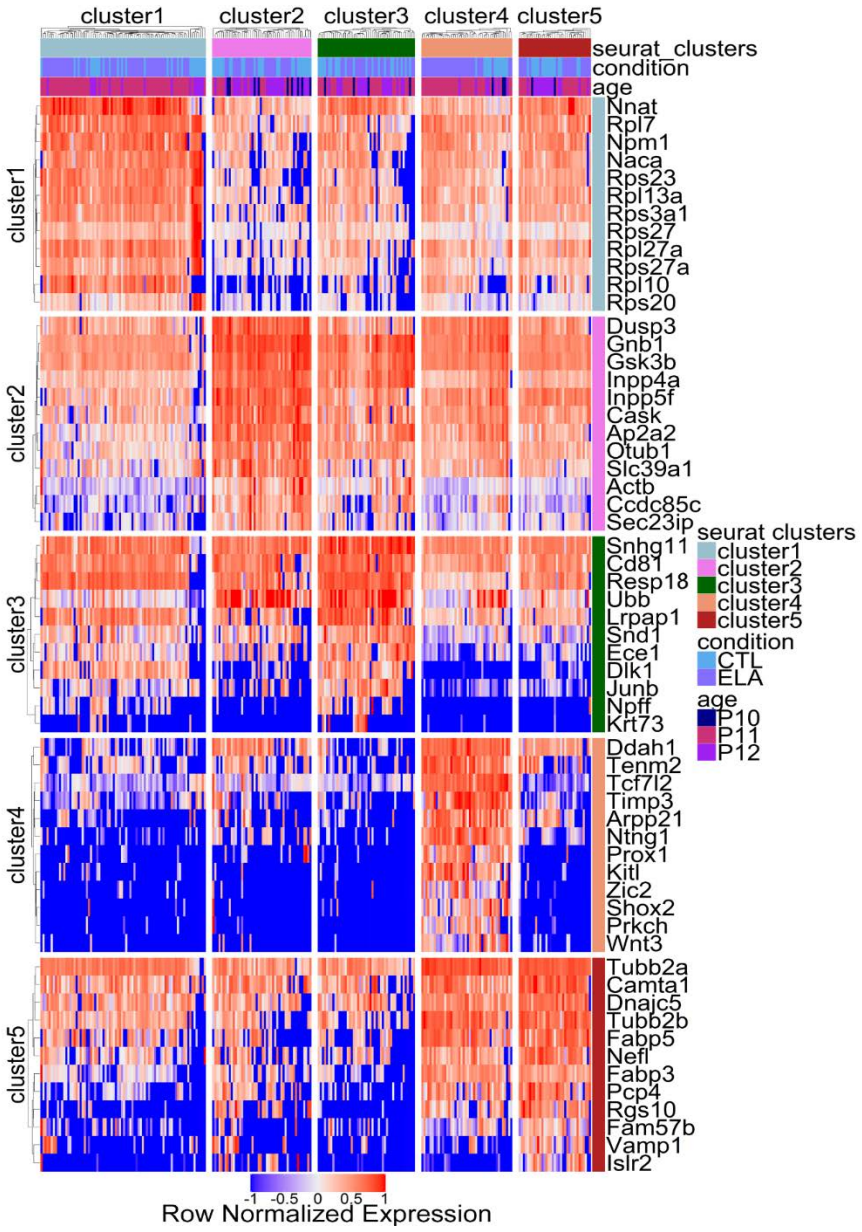

Figure 3.12 Heatmap of top 12 enriched genes from each Seurat cluster using differential expression analysis.

The top 12 differentially expressed genes from each cluster are listed. The columns are cells split by their cluster affiliation, and labeled with whether the cells are from the ELA or CTL conditions, and the age of the mice at collection. The rows are the top 12 genes that are split by the genes enriched in each cluster. The expression of each gene is row normalized TPMs.

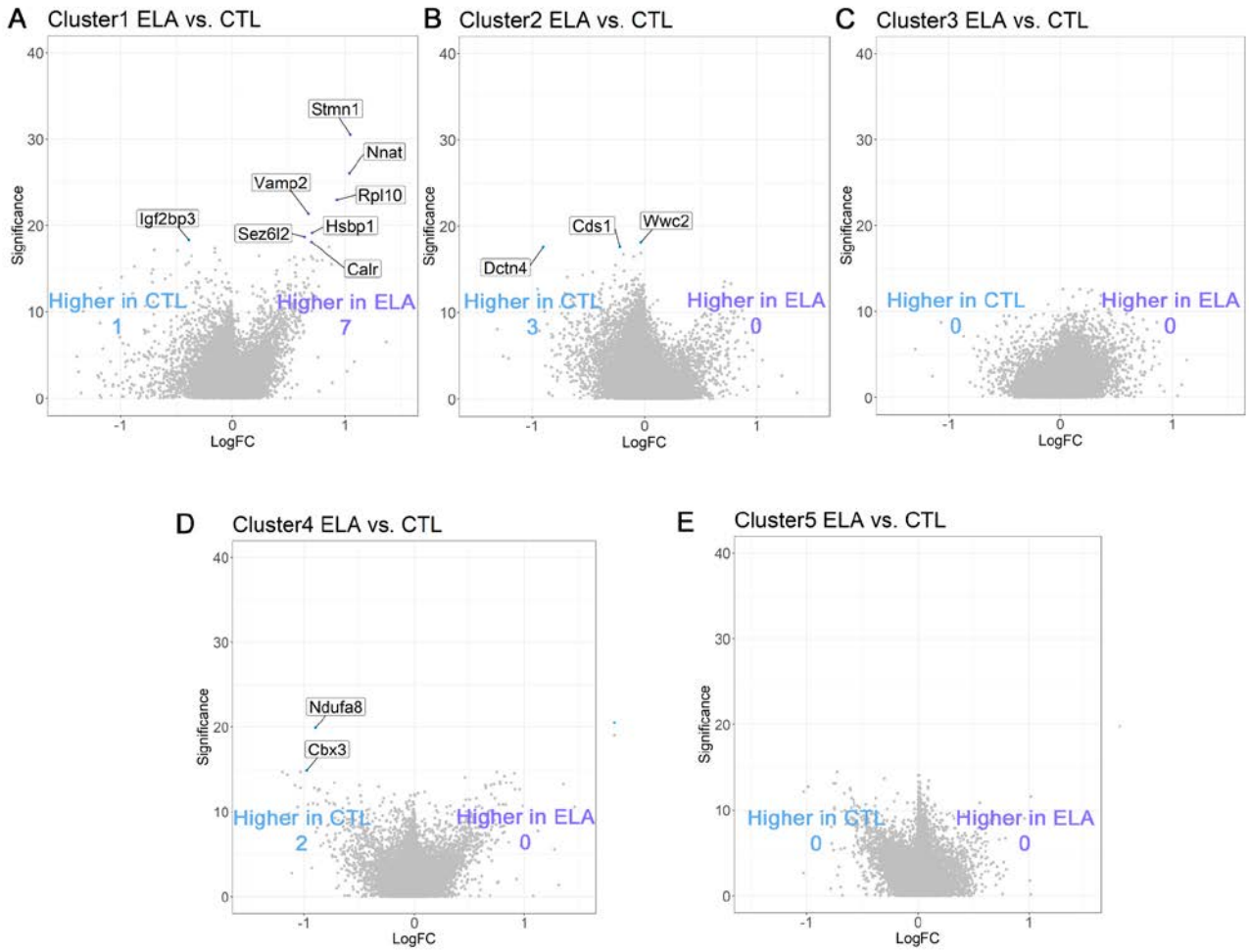

Figure 3.13 Differential expression analysis in each Seurat cluster between the ELA and CTL condition.

A) Volcano plot showing differential expression between the ELA and CTL conditions from cells in cluster 1. There were 7 differentially expressed genes enriched in the ELA condition and 1 in the CTL condition. B) Volcano plot showing differential expression between the ELA and CTL conditions from cells in cluster 2. There were 0 differentially expressed genes enriched in the ELA condition and 3 in the CTL condition. C) Volcano plot showing differential expression between the ELA and CTL conditions from cells in cluster 3. No differentially expressed genes were found. D) Volcano plot showing differential expression between the ELA

and CTL conditions from cells in cluster 4. There were 0 differentially expressed genes enriched in the ELA condition and 2 in the CTL condition. E) Volcano plot showing differential expression between the ELA and CTLconditions from cells in cluster 5. No differentially expressed genes were found. All enriched genes had to pass an FDR<0.1 filter. The enriched genes are listed for both the CTL and ELA conditions on each Volcano plot.

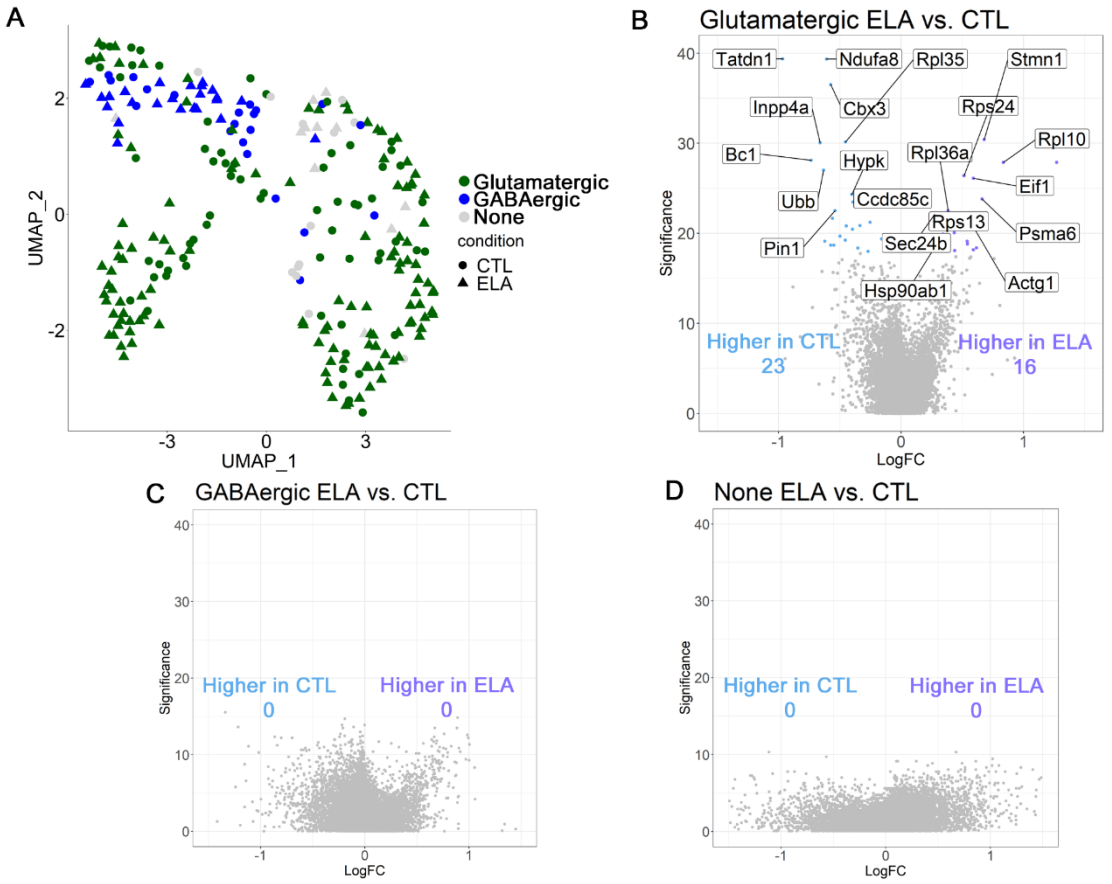

Figure 3.14 Differential expression of distinct neurotransmitter expression defined subpopulations of CRH cells.

A) Characterization of glutamatergic and GABAergic neurotransmitters. B) Volcano plot representing differential expression between the ELA and CTL conditions from cells in the glutamatergic cluster. There were 16 differentially expressed genes enriched in the ELA condition and 23 in the CTL condition. C) Volcano plot of differential expression between the ELA and CTL conditions from cells in the GABAergic cluster. No differentially expressed genes were found. D) Volcano plot of differential expression between the ELA and CTL conditions from cells in the cluster with cells expressing neither neurotransmitter. No differentially expressed genes were found. All enriched genes had to pass an FDR<0.1 filter. Some of the top enriched genes are listed for both the CTL and ELA conditions on each Volcano plot.

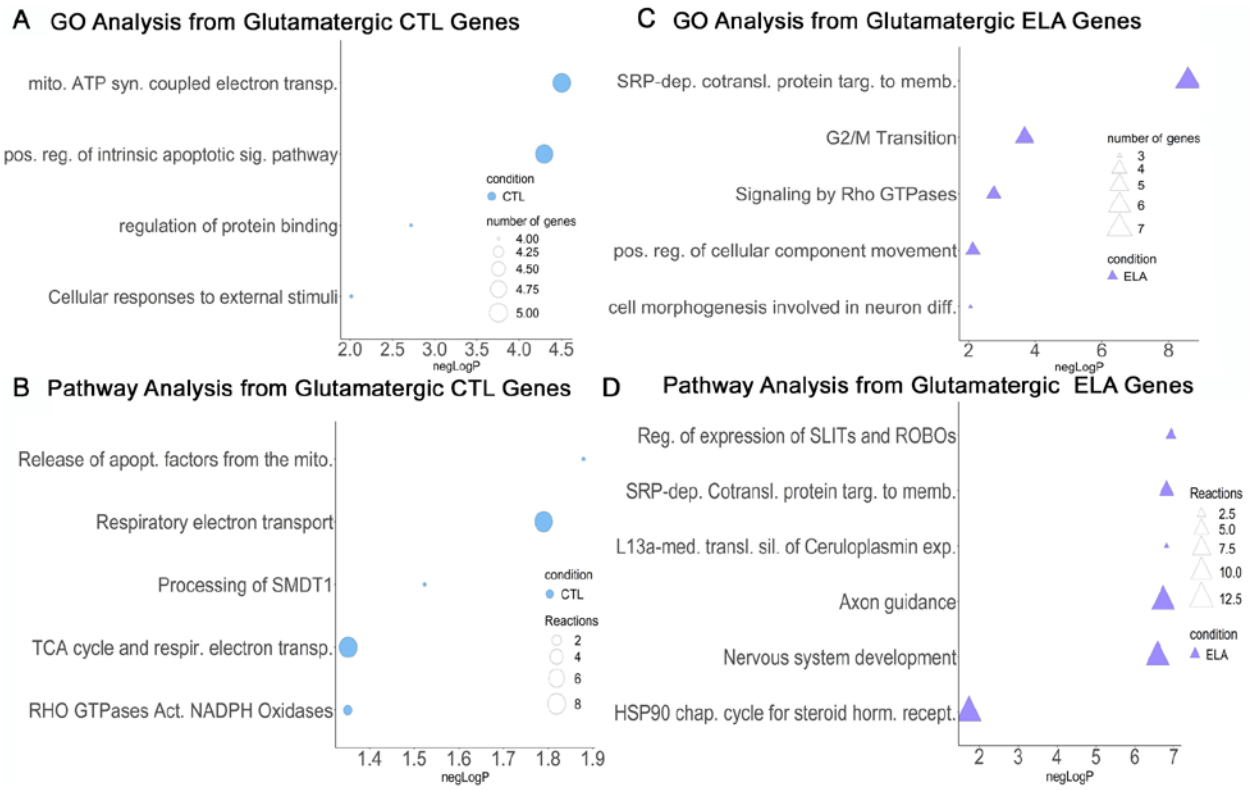

Figure 3.15 Gene ontology (GO) and pathway analysis of differentially expressed genes between the glutamatergic ELA and CTL conditions.

A) GO analysis on genes enriched in the glutamatergic CTL condition were analyzed using metascape. B) Pathway analysis on genes enriched in the glutamatergic CTL condition were analyzed using Reactome. C) GO analysis on genes enriched in the glutamatergic ELA condition were analyzed using metascape. D) Pathway analysis on genes enriched in the glutamatergic ELA condition were analyzed using Reactome. Only GO terms passing a p-value<0.05 and pathways passing an FDR<0.05 filter were considered.

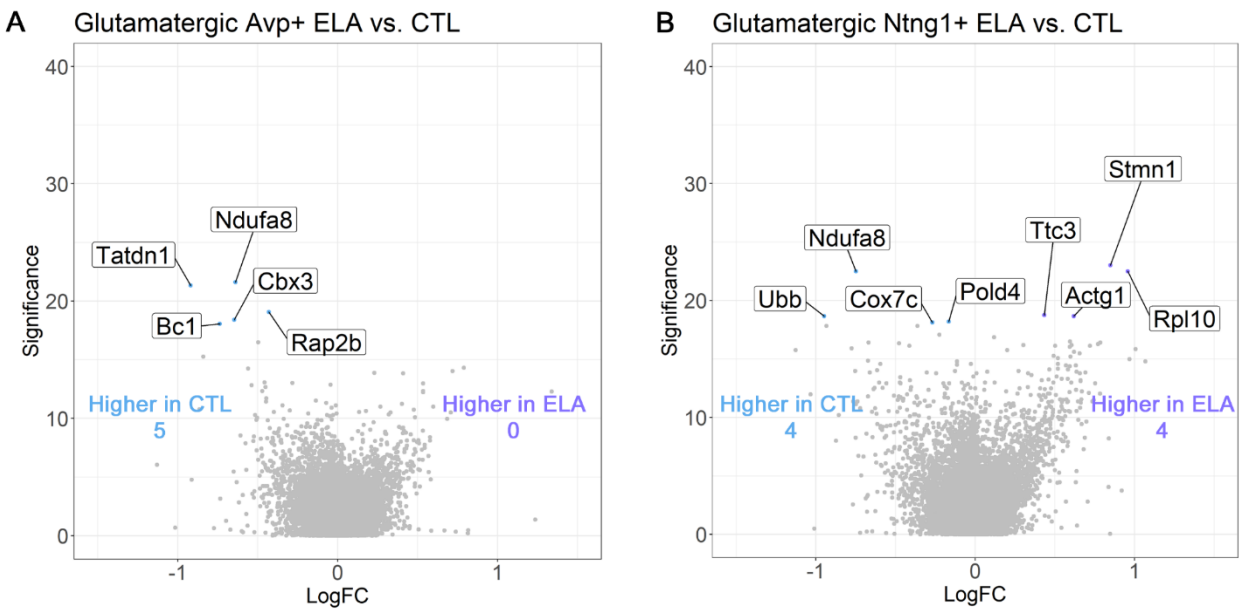

Figure 3.16 Differential expression of distinct glutamatergic marker expression defined subpopulations of CRH cells.

A) Volcano plot of differential expression between the ELA and CTL conditions from cells in the glutamatergic *Avp*+ cluster. There were 0 differentially expressed genes enriched in the ELA condition and 5 in the CTL condition. B) Volcano plot of differential expression between the ELA and CTL conditions from cells in the glutamatergic *Ntng1*+ cluster. There were 4 differentially expressed genes enriched in the ELA condition and 4 in the CTL condition. All enriched genes had to pass an FDR<0.1 filter. The enriched genes are listed for both the CTL and ELA conditions on each Volcano plot.

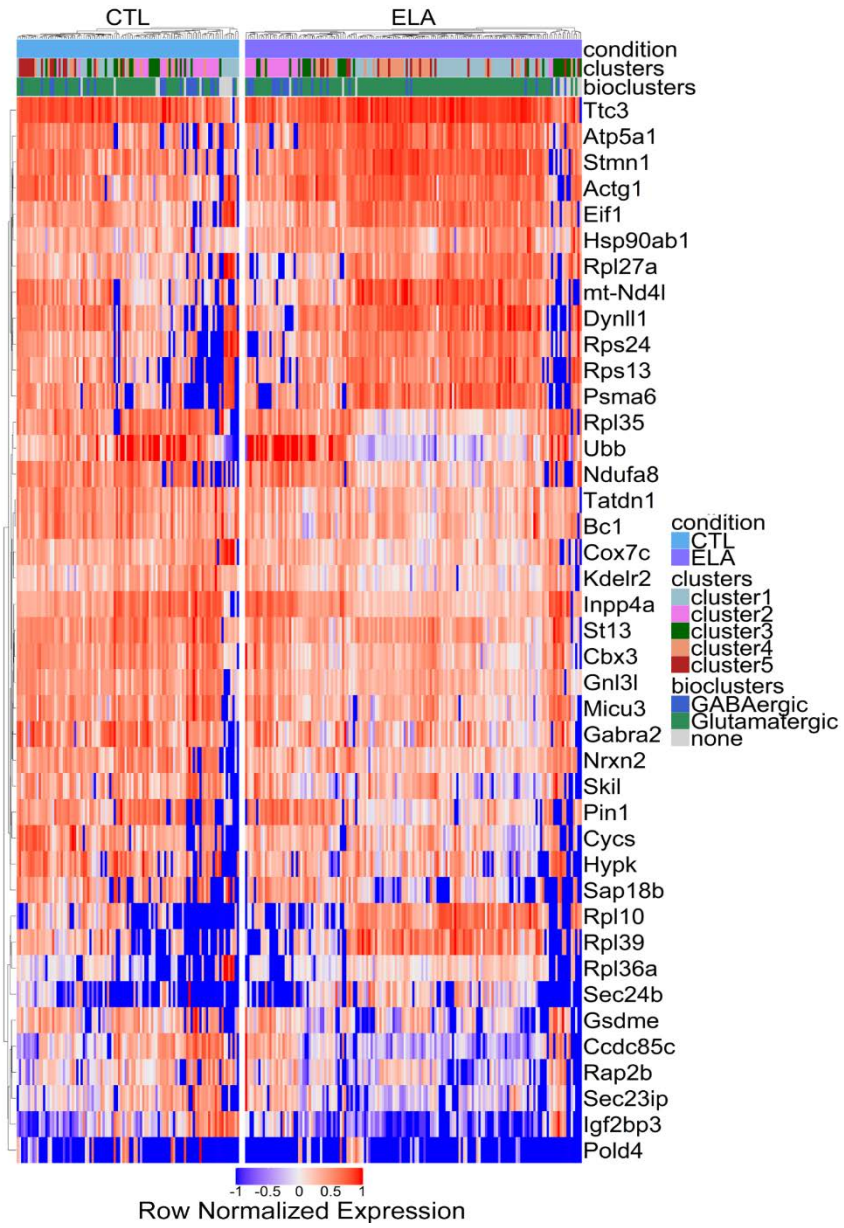

Figure 3.17 Heatmap of all 41 glutamatergic and biomarker differentially expressed genes between the ELA and CTL conditions.

Row normalized expression values (TPMs) are shown on the color bar at the bottom of the heatmap. Cells are columns and rows are genes. Cells are split by whether they are from the CTL or ELA conditions. Cluster and bio-cluster identities are also displayed in color bars at the top of the heatmap.

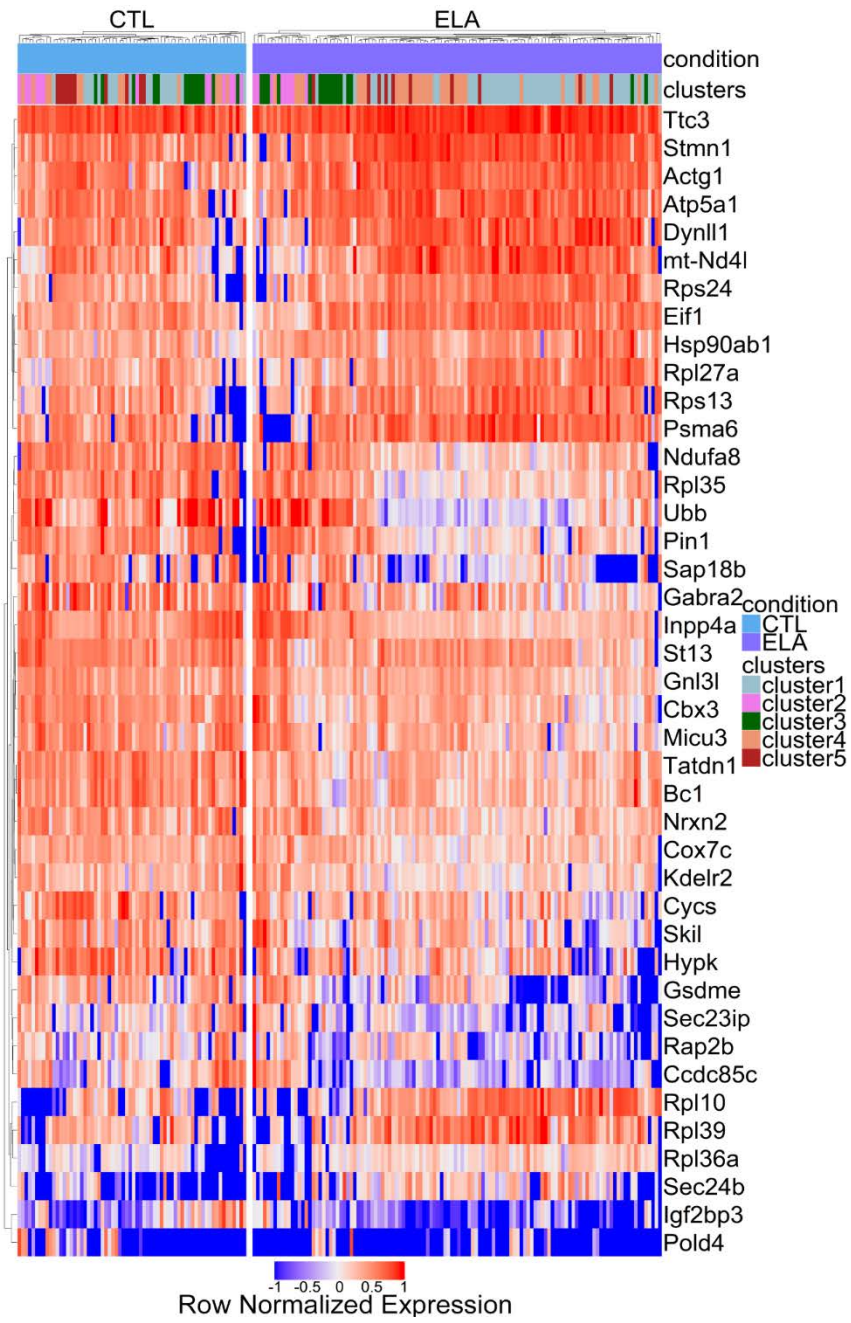

Figure 3.18 Heatmap of all 41 glutamatergic and biomarker differentially expressed genes between the ELA and CTL conditions in glutamatergic cells. Row normalized expression values (TPMs) are shown on the color bar at the bottom of the heatmap. Cells are columns and rows are genes. Cells are split by whether they are from the CTL or ELA conditions. Cluster identities are also displayed in color bars at the top of the heatmap.

3.6 Tables

Batch ID	Condition	Postnatal Day	# of Animals	Av. Weight (g)	# of cells
Batch-19	Ctrl	10	2	5.35	9
Batch-20	Ctrl	12	2	5.7	8
Batch-21	Ctrl	12	3	6.13	26
Batch-23	Ctrl	11	5	5.78	3
Batch-25	Ctrl	11	3	7.2	21
Batch-26	Ctrl	10	3	5.1	14
Batch-28	Ctrl	12	5	6.22	9
Batch-29	Ctrl	12	3	4.6	13
Batch-31	Ctrl	10	3	6.4	3
Batch-32	Ctrl	12	4	6	23
Batch-33	Ctrl	12	5	6.7	22
Batch-34	Ctl	12	4	5.3	21
Batch-36	Ctl	12	4	6.6	15
Batch-38	Ctrl	11	4	5.2	51
Batch-L7	LBN	12	4	5.45	15
Batch-L9	LBN	12	4	5.2	2
Batch-L10	LBN	12	4	5.9	25
Batch-L11	LBN	10	5	5.14	35
Batch-L12	LBN	11	5	4.1	27
Batch-L13	LBN	11	4	5.7	21
Batch-L14	LBN	12	3	5.5	4
Batch-L15	LBN	12	1	3.7	5
Batch-L16	LBN	11	5	3.8	70
Batch-L17	LBN	11	3	5.1	69

Table 3.1 FACS Stats, 602 cells were profiled from 88 animals

ensembl	gene	cluster
ENSMUSG00000029836.15	Cbx3	CTL
ENSMUSG00000050891.9	Tatdn1	CTL
ENSMUSG00000029814.10	Igf2bp3	CTL
ENSMUSG00000070780.11	Rbm47	CTL
ENSMUSG00000073971.3	Olfr554	CTL
ENSMUSG00000115725.1	4930572G02Rik	CTL
ENSMUSG00000025266.11	Gnl3l	CTL
ENSMUSG00000084883.2	Ccdc85c	CTL
ENSMUSG00000017778.14	Cox7c	CTL
ENSMUSG00000022403.15	St13	CTL
ENSMUSG00000115783.1	Bc1	CTL
ENSMUSG00000111806.1	Gm47403	CTL
ENSMUSG00000034120.18	Srsf2	CTL
ENSMUSG00000107023.2	Gm42715	CTL
ENSMUSG00000102718.1	Gm37761	CTL
ENSMUSG00000079111.3	Kdelr2	CTL
ENSMUSG00000011254.16	Thg1l	CTL
ENSMUSG00000062997.6	Rpl35	CTL
ENSMUSG00000039478.15	Micu3	CTL
ENSMUSG00000107927.1	Gm44090	CTL
ENSMUSG00000036790.5	Slitrk2	CTL
ENSMUSG00000026113.17	Inpp4a	CTL
ENSMUSG00000060475.13	Wtap	CTL
ENSMUSG00000027660.16	Skil	CTL
ENSMUSG00000097408.1	Gm26831	CTL
ENSMUSG00000042284.10	Itga1	CTL
ENSMUSG00000000560.9	Gabra2	CTL
ENSMUSG00000102700.1	Gm38312	CTL
ENSMUSG00000028832.11	Stmn1	ELA
ENSMUSG00000058443.5	Rpl10	ELA
ENSMUSG00000023944.14	Hsp90ab1	ELA
ENSMUSG00000071866.12	Ppia	ELA
ENSMUSG00000062825.15	Actg1	ELA
ENSMUSG00000025428.15	Atp5a1	ELA
ENSMUSG00000079641.3	Rpl39	ELA
ENSMUSG00000090862.3	Rps13	ELA
ENSMUSG00000020894.16	Vamp2	ELA
ENSMUSG00000034211.14	Mrps17	ELA
ENSMUSG00000016541.10	Atxn10	ELA
ENSMUSG00000015656.17	Hspa8	ELA
ENSMUSG00000062296.8	Trank1	ELA
ENSMUSG00000049281.16	Scn3b	ELA
ENSMUSG00000024758.15	Rtn3	ELA
ENSMUSG00000035530.13	Eif1	ELA
ENSMUSG00000027350.8	Chgb	ELA
ENSMUSG00000021024.14	Psma6	ELA

Table 3.2 Top differentially expressed genes between ELA and CTL

ensembl	gene	cluster	ensembl	gene	cluster	ensembl	gene	cluster
ENSMUSG00000049517.8	Rps23	1	ENSMUSG00000044349.15	Snhg11	3	ENSMUSG00000030844.11	Rgs10	5
ENSMUSG00000090733.6	Rps27	1	ENSMUSG00000033061.15	Resp18	3	ENSMUSG00000051243.14	Islr2	5
ENSMUSG00000096921.1	Gm26822	1	ENSMUSG00000023052.8	Npff	3	ENSMUSG00000045136.6	Tubb2b	5
ENSMUSG00000020460.15	Rps27a	1	ENSMUSG00000001424.14	Snd1	3	ENSMUSG00000022055.7	Nefl	5
ENSMUSG00000067786.16	Nnat	1	ENSMUSG00000019505.7	Ubb	3	ENSMUSG00000090223.2	Pcp4	5
ENSMUSG00000074129.14	Rpl13a	1	ENSMUSG00000057530.14	Ece1	3	ENSMUSG00000058672.7	Tubb2a	5
ENSMUSG00000058443.5	Rpl10	1	ENSMUSG00000087679.2	Tmem250	3	ENSMUSG00000030337.16	Vamp1	5
ENSMUSG00000028081.6	Rps3a1	1	ENSMUSG00000052837.6	Junb	3	ENSMUSG00000058966.13	Fam57b	5
ENSMUSG00000046364.14	Rpl27a	1	ENSMUSG00000029103.16	Lrpap1	3	ENSMUSG00000027533.10	Fabp5	5
ENSMUSG00000028234.6	Rps20	1	ENSMUSG00000111017.1	Gm48314	3	ENSMUSG00000054013.6	Tmem179	5
ENSMUSG00000061315.14	Naca	1	ENSMUSG00000115410.1	2810457G06Rik	3	ENSMUSG00000028773.8	Fabp3	5
ENSMUSG00000075015.3	Gm10801	1	ENSMUSG00000063661.6	Krt73	3	ENSMUSG00000014592.20	Camta1	5
ENSMUSG00000057113.13	Npm1	1	ENSMUSG00000037706.17	Cd81	3	ENSMUSG00000000826.16	Dnajc5	5
ENSMUSG00000043716.13	Rpl7	1	ENSMUSG00000040856.17	Dlk1	3	ENSMUSG00000020601.8	Trib2	5
ENSMUSG00000086324.8	Gm15564	1	ENSMUSG00000024030.7	Abcg1	3	ENSMUSG00000000563.17	Atp5f1	5
ENSMUSG00000067288.13	Rps28	1	ENSMUSG00000109722.1	Gm10578	3	ENSMUSG00000052301.14	Doc2a	5
ENSMUSG00000037742.14	Eef1a1	1	ENSMUSG00000040412.16	5330417C22Rik	3	ENSMUSG00000031367.15	Ap1s2	5
ENSMUSG00000053332.13	Gas5	1	ENSMUSG00000109320.1	Gm44679	3	ENSMUSG00000043419.11	Rnf227	5
ENSMUSG00000058600.13	Rpl30	1	ENSMUSG00000026113.17	Inpp4a	3	ENSMUSG00000030310.10	Slc6a1	5
ENSMUSG00000075014.1	Gm10800	1	ENSMUSG00000104467.1	Gm37660	3	ENSMUSG00000066705.7	Fxyd6	5
ENSMUSG00000035530.13	Eif1	1	ENSMUSG00000086847.1	Tbx3os2	3	ENSMUSG00000048040.8	Arxes2	5
ENSMUSG00000087002.1	Gm16277	1	ENSMUSG00000057465.5	Saa2	3	ENSMUSG00000038500.15	Prr3	5
ENSMUSG00000046330.10	Rpl37a	1	ENSMUSG00000085696.8	Hoxaas3	3	ENSMUSG00000008658.16	Rbfox1	5
ENSMUSG00000021685.11	Otp	1	ENSMUSG00000020423.6	Btg2	3	ENSMUSG00000031438.11	Rnf128	5
ENSMUSG00000105702.1	Gm43780	1	ENSMUSG00000037727.6	Avp	3	ENSMUSG00000026260.12	Ndufa10	5
ENSMUSG00000029810.15	Tmem176b	1	ENSMUSG00000054580.14	Pla2r1	3	ENSMUSG00000027581.12	Stmn3	5
ENSMUSG00000055322.15	Tns1	1	ENSMUSG00000003411.10	Rab3b	3	ENSMUSG00000072235.6	Tuba1a	5
ENSMUSG00000097312.1	Gm26870	1	ENSMUSG00000021337.8	Scgn	3	ENSMUSG00000060279.14	Ap2a1	5
ENSMUSG00000085014.1	Gm13490	1	ENSMUSG00000023034.7	Nr4a1	3	ENSMUSG00000029674.13	Limk1	5
ENSMUSG00000061477.4	Rps7	1	ENSMUSG00000089706.7	B230216N24Rik	3	ENSMUSG00000026443.3	Lrrn2	5
ENSMUSG00000032265.14	Tent5a	1	ENSMUSG00000048292.3	Olfr1417	3	ENSMUSG00000093497.1	Gm20713	5
ENSMUSG00000073940.3	Hbb	1	ENSMUSG00000047535.11	Olfr67	3	ENSMUSG00000030806.6	Stx1b	5
ENSMUSG00000045534.4	Kcna5	1	ENSMUSG00000087083.2	Gm15083	3	ENSMUSG00000051323.16	Pcdh19	5
ENSMUSG00000060703.12	Cd302	1	ENSMUSG00000112376.1	Gm40649	3	ENSMUSG00000094002.3	Gm9866	5
ENSMUSG00000015619.10	Gata3	1	ENSMUSG00000032726.12	Bmp8a	3	ENSMUSG00000062380.4	Tubb3	5
ENSMUSG00000032353.9	Tmed3	1	ENSMUSG00000110523.1	C230057M02Rik	3	ENSMUSG00000020366.18	Mapk9	5
ENSMUSG00000049775.16	Tmsb4x	1	ENSMUSG00000107573.2	Olfr1389	3	ENSMUSG00000048332.13	Lhfp	5
ENSMUSG00000095186.1	Gm10718	1	ENSMUSG00000118045.1	AC131709.1	3	ENSMUSG00000043531.16	Sorcs1	5
ENSMUSG00000040856.17	Dlk1	1	ENSMUSG00000093629.1	Prox2os	3	ENSMUSG00000022193.7	Psmb5	5
ENSMUSG00000047215.14	Rpl9	1	ENSMUSG00000030245.10	Golt1b	3	ENSMUSG00000025889.13	Snca	5
ENSMUSG00000050711.7	Scg2	1	ENSMUSG00000102009.6	4933400F21Rik	3	ENSMUSG00000117361.1	AC122818	5
ENSMUSG00000029359.13	Tesc	1	ENSMUSG00000103625.1	Gm37357	3	ENSMUSG00000079355.5	Ackr4	5
ENSMUSG00000078974.10	Sec61g	1	ENSMUSG00000103459.1	Gm38096	3	ENSMUSG00000037185.9	Krt80	5
ENSMUSG00000049751.6	Rpl36al	1	ENSMUSG00000085318.1	Gm12524	3	ENSMUSG00000032946.16	Rasgrp2	5
ENSMUSG00000021917.16	Spcs1	1	ENSMUSG00000108137.1	Gm44053	3	ENSMUSG00000039485.8	Tspyl4	5
ENSMUSG00000021660.14	Btf3	1	ENSMUSG00000086486.1	Ift88os	3	ENSMUSG00000063374.4	Olfr763	5
ENSMUSG00000031320.9	Rps4x	1	ENSMUSG00000018476.7	Kdm6b	3	ENSMUSG00000110839.1	Gm47800	5
ENSMUSG00000065947.3	mt-Nd4l	1	ENSMUSG00000109879.1	Gm10997	3	ENSMUSG00000025132.13	Arhgdia	5
ENSMUSG00000052305.6	Hbb	1	ENSMUSG00000073119.3	Speer4a	3	ENSMUSG00000037225.13	Fgf2	5
ENSMUSG00000041453.12	Rpl21	1	ENSMUSG00000107667.2	C530044C16Rik	3	ENSMUSG00000023906.3	Cldn6	5
ENSMUSG00000033685.13	Ucp2	1	ENSMUSG00000048173.5	Olfr1324	3	ENSMUSG00000020496.9	Rnf187	5
ENSMUSG00000020048.13	Hsp90b1	1	ENSMUSG00000106651.1	Gm42608	3	ENSMUSG00000028159.14	Dapp1	5
ENSMUSG00000021248.9	Tmed10	1	ENSMUSG00000073414.8	Mpig6b	3	ENSMUSG00000019943.10	Atp2b1	5
ENSMUSG00000049115.15	Agtr1a	1	ENSMUSG00000035227.7	Spcs2	3	ENSMUSG00000109770.1	Gm30085	5
ENSMUSG00000060036.14	Rpl3	1	ENSMUSG00000115430.1	Gm41300	3	ENSMUSG00000064254.6	Ethe1	5
ENSMUSG00000039114.16	Nrn1	1	ENSMUSG00000116618.1	Gm49719	3	ENSMUSG00000074354.11	Arhgap20c	5
ENSMUSG00000028673.10	Fuca1	1	ENSMUSG00000036855.3	Gjd4	3	ENSMUSG00000090936.2	Gm17705	5
ENSMUSG00000024431.15	Nr3c1	1	ENSMUSG00000111570.1	Gm32335	3	ENSMUSG00000050248.11	Evc2	5
ENSMUSG00000037727.6	Avp	1	ENSMUSG00000101840.1	Gm28294	3	ENSMUSG00000029086.15	Prom1	5
ENSMUSG00000037072.15	Selenof	1	ENSMUSG00000036109.17	Mbnl3	3	ENSMUSG00000020905.11	Usp43	5

ensembl	gene	cluster	ensembl	gene	cluster
ENSMUSG00000029064.15	Gnb1	2	ENSMUSG00000024985.19	Tcf7l2	4
ENSMUSG00000042105.18	Inpp5f	2	ENSMUSG00000059857.15	Ntng1	4
ENSMUSG00000026113.17	Inpp4a	2	ENSMUSG00000019966.18	Kitl	4
ENSMUSG00000002957.11	Ap2a2	2	ENSMUSG00000010175.13	Prox1	4
ENSMUSG00000029580.14	Actb	2	ENSMUSG00000020044.13	Timp3	4
ENSMUSG00000052310.9	Slc39a1	2	ENSMUSG00000061524.8	Zic2	4
ENSMUSG00000084883.2	Ccdc85c	2	ENSMUSG00000028194.15	Ddah1	4
ENSMUSG00000003518.13	Dusp3	2	ENSMUSG00000032503.18	Arpp21	4
ENSMUSG00000037754.13	Ppp1r16b	2	ENSMUSG00000027833.16	Shox2	4
ENSMUSG00000055319.8	Sec23ip	2	ENSMUSG00000000125.5	Wnt3	4
ENSMUSG00000031012.17	Cask	2	ENSMUSG00000021108.18	Prkch	4
ENSMUSG00000038141.11	Tmem181a	2	ENSMUSG00000085438.1	1700020I14Rik	4
ENSMUSG00000024767.11	Otub1	2	ENSMUSG00000049336.16	Tenm2	4
ENSMUSG00000034361.10	Cpne2	2	ENSMUSG00000041272.11	Tox	4
ENSMUSG00000040459.11	Arglu1	2	ENSMUSG00000042834.15	Nrep	4
ENSMUSG00000022812.14	Gsk3b	2	ENSMUSG00000021259.5	Cyp46a1	4
ENSMUSG00000070576.4	Mn1	2	ENSMUSG00000030729.17	Pgm2l1	4
ENSMUSG00000002341.8	Ncan	2	ENSMUSG00000040612.13	Ildr2	4
ENSMUSG00000069456.4	Rdh16	2	ENSMUSG00000032936.13	Camkv	4
ENSMUSG00000020661.15	Dnmt3a	2	ENSMUSG00000046178.3	Nxph1	4
ENSMUSG00000026643.16	Nmt2	2	ENSMUSG00000027495.4	Fam210b	4
ENSMUSG00000033066.15	Gas7	2	ENSMUSG00000035864.14	Syt1	4
ENSMUSG00000017188.3	Coa3	2	ENSMUSG00000050315.14	Synpo2	4
ENSMUSG00000096768.7	Gm47283	2	ENSMUSG00000024109.18	Nrxn1	4
ENSMUSG00000041852.15	Tcf20	2	ENSMUSG00000027577.14	Chrna4	4
ENSMUSG00000034120.18	Srsf2	2	ENSMUSG00000053025.13	Sv2b	4
ENSMUSG00000007415.15	Gatad1	2	ENSMUSG00000026688.5	Mgst3	4
ENSMUSG00000043311.7	D17H6S53E	2	ENSMUSG00000024501.20	Dpysl3	4
ENSMUSG00000057132.16	Rpgrip1	2	ENSMUSG00000040339.10	Fam102b	4
ENSMUSG00000031997.9	Trpc6	2	ENSMUSG00000033060.15	Lmo7	4
ENSMUSG00000048915.12	Efna5	2	ENSMUSG00000032368.14	Zic1	4
ENSMUSG00000022048.8	Dpysl2	2	ENSMUSG00000078816.9	Prkcg	4
ENSMUSG00000094145.2	Vmn2r20	2	ENSMUSG00000030500.6	Slc17a6	4
ENSMUSG00000039542.16	Ncam1	2	ENSMUSG00000058672.7	Tubb2a	4
ENSMUSG00000007476.18	Lrrc8a	2	ENSMUSG00000042804.13	Gpr153	4
ENSMUSG00000075199.3	Olfr52	2	ENSMUSG00000023192.12	Grm2	4
ENSMUSG00000002846.9	Timmdc1	2	ENSMUSG00000032238.17	Rora	4
ENSMUSG00000038658.15	Ric1	2	ENSMUSG00000032549.7	Rab6b	4
ENSMUSG00000028030.12	Tbck	2	ENSMUSG00000019230.14	Lhx9	4
ENSMUSG00000115570.1	Gm48986	2	ENSMUSG00000002107.18	Celf2	4
ENSMUSG00000113200.1	Gm48632	2	ENSMUSG00000020431.5	Adcy1	4
ENSMUSG00000067199.4	Frat1	2	ENSMUSG00000021719.9	Rgs7bp	4
ENSMUSG00000116048.1	2-Sep	2	ENSMUSG00000036564.17	Ndrg4	4
ENSMUSG00000029755.10	Dlx5	2	ENSMUSG00000031561.16	Tenm3	4
ENSMUSG00000097617.2	Gm10687	2	ENSMUSG00000044847.13	Lsm11	4
ENSMUSG00000027678.17	Ncoa3	2	ENSMUSG00000005089.15	Slc1a2	4
ENSMUSG00000112304.1	Gm35206	2	ENSMUSG00000020121.15	Srgap1	4
ENSMUSG00000033720.12	Sfxn5	2	ENSMUSG00000055022.14	Cntn1	4
ENSMUSG00000110120.1	Gm45251	2	ENSMUSG00000007207.10	Stx1a	4
ENSMUSG00000065954.11	Tacc1	2	ENSMUSG00000044117.12	2900011O08Rik	4
ENSMUSG00000111740.1	Gm49783	2	ENSMUSG00000030557.17	Mef2a	4
ENSMUSG00000030539.13	Sema4b	2	ENSMUSG00000034613.12	Ppm1h	4
ENSMUSG00000022822.16	Abcc5	2	ENSMUSG00000074607.11	Tox2	4
ENSMUSG00000042506.15	Usp22	2	ENSMUSG00000029168.14	Dpysl5	4
ENSMUSG00000029432.12	Nipsnap2	2	ENSMUSG00000072235.6	Tuba1a	4
ENSMUSG00000021913.8	Ogdhl	2	ENSMUSG00000031137.17	Fgf13	4
ENSMUSG00000059654.7	Reg1	2	ENSMUSG00000038065.13	Mturn	4
ENSMUSG00000047466.10	8030462N17	2	ENSMUSG00000051359.15	Ncald	4
ENSMUSG00000029516.19	Cit	2	ENSMUSG00000020635.8	Fkbp1b	4
ENSMUSG00000021135.8	Slc10a1	2	ENSMUSG00000026585.13	Kifap3	4

Table 3.3 Top differentially expressed genes between Seurat clusters

ensembl	gene	cluster
ENSMUSG00000115783.1	Bc1	CTL
ENSMUSG00000029836.15	Cbx3	CTL
ENSMUSG00000084883.2	Ccdc85c	CTL
ENSMUSG00000027245.11	Hypk	CTL
ENSMUSG00000026113.17	Inpp4a	CTL
ENSMUSG00000026895.7	Ndufa8	CTL
ENSMUSG00000032171.7	Pin1	CTL
ENSMUSG00000062997.6	Rpl35	CTL
ENSMUSG00000050891.9	Tatdn1	CTL
ENSMUSG00000019505.7	Ubb	CTL
ENSMUSG00000039478.15	Micu3	CTL
ENSMUSG00000017778.14	Cox7c	CTL
ENSMUSG00000027660.16	Skil	CTL
ENSMUSG00000029821.15	Gsdme	CTL
ENSMUSG00000033768.17	Nrxn2	CTL
ENSMUSG00000022403.15	St13	CTL
ENSMUSG00000029814.10	Igf2bp3	CTL
ENSMUSG00000025266.11	Gnl3l	CTL
ENSMUSG00000079111.3	Kdelr2	CTL
ENSMUSG00000000560.9	Gabra2	CTL
ENSMUSG00000063694.5	Cycs	CTL
ENSMUSG00000055319.8	Sec23ip	CTL
ENSMUSG00000061104.5	Sap18b	CTL
ENSMUSG00000062825.15	Actg1	ELA
ENSMUSG00000035530.13	Eif1	ELA
ENSMUSG00000023944.14	Hsp90ab1	ELA
ENSMUSG00000021024.14	Psma6	ELA
ENSMUSG00000058443.5	Rpl10	ELA
ENSMUSG00000079435.9	Rpl36a	ELA
ENSMUSG00000090862.3	Rps13	ELA
ENSMUSG00000025290.17	Rps24	ELA
ENSMUSG00000001052.15	Sec24b	ELA
ENSMUSG00000028832.11	Stmn1	ELA
ENSMUSG00000096921.1	Gm26822	ELA
ENSMUSG00000009013.5	Dynll1	ELA
ENSMUSG00000079641.3	Rpl39	ELA
ENSMUSG00000065947.3	mt-Nd4l	ELA
ENSMUSG00000046364.14	Rpl27a	ELA
ENSMUSG00000025428.15	Atp5a1	ELA

Table 3.4 Top differentially expressed genes in glutamatergic cells between ELA and CTL

3.7 Methods

3.7.1 Animals

Dams were Crh-IRES-Cre +/+ and were paired with Ai14 tdTomato males both on a C57Bl6 background. The resulting offspring were Crh-IRES-Cre; Ai14 tdTomato and were used for all experiments. Animals were housed in a 12-hour light cycle and provided *ad libitum* food and water. All experiments were carried out in accordance with the Institutional Animal Care and Use Committee at the University of California, Irvine, and were consistent with Federal guidelines.

3.7.2 Limited Bedding and Nesting (LBN).

The simulated poverty limited bedding and nesting paradigm (LBN) consisted of limiting nesting and bedding materials in cages between P2-P9 as described previously. For the LBN group, a plastic-coated mesh platform was placed ~2.5cm above the floor of a standard cage. Cobb bedding was reduced to cover the cage floor sparsely, and one-half of a single nestlet was provided for nesting material on the platform. CTL dams and litters resided in standard cages containing ample cob bedding and one whole nestlet for nesting. This paradigm causes maternal care to be fragmented and unpredictable [48,52], provoking chronic stress in the pups [55]. CTL and experimental cages were undisturbed during P2–P9, housed in temperature-controlled rooms (22°C). For RNA sequencing pups remained on the LBN paradigm until tissue was collected P10-P12. For experiments in adulthood, experimental groups were transferred to standard cages on P10 and were weaned on P21. Animals were housed by sex with littermates.

3.7.3 Single cell preparation

Dissection. P10-P12 pups were killed via decapitation and brains were removed immediately on ice. The brain was trimmed to a smaller block containing hypothalamus and was placed into a slush of EBSS. The block was then sliced on a vibratome to obtain at 1.5mm slice posterior of the anterior commissure. The slice was then trimmed under a dissection microscope to remove thalamus and cortex.

Dissociation. The trimmed slices were placed immediately into papain and gently triturated ~15 times with a pipette to break up tissue before being placed on a heated (37C) orbital shaker for 30 minutes. Digested tissue was then homogenized by triturating with pipette ~ 50 times until there were no visible tissue chunks. Homogenate was spun at 500 x g for 15 minutes at room temperature. Supernatant was removed and cells were resuspended in 500uL 2% FBS in PBS.

FACS. Samples were run on the aria fusion. Immediately prior to sort, cell suspension was run through a 70uM filter and washed with 500uL 2% FBS in PBS. Cells were sorted into 8 well strip tubes and immediately spun down at 4C and frozen on dry ice.

3.7.4 RNAseq pipeline

Tdtomato positive hypothalamic cells from both ELA and CTL animals that are Crh positive are processed using the SmartSeq2 RNA sequencing protocol [60], Illumina Library prep and sequenced using the Illumina NextSeq500 sequencer. FASTQ files were mapped and quantified using kallisto [122]. The resulting TPM matrix was quantile normalized before using the R package Seurat to cluster and run non-linear dimensional reduction [123]. ComplexHeatmap

was used for heatmap generation [124]. Metascape [97] was used to determine gene ontology, and reactome [98] was used to determine pathways.

3.8 References

1. Yasui, D., Peedicayil, J. & Grayson, D. R. *Neuropsychiatric Disorders and Epigenetics*. (Academic Press, 2016).

2. Eagleson, K. L., Villaneuva, M., Southern, R. M. & Levitt, P. Proteomic and mitochondrial adaptations to early-life stress are distinct in juveniles and adults. *Neurobiol Stress* **13**, 100251 (2020).

3. Bale, T. L. *et al.* Early life programming and neurodevelopmental disorders. *Biol. Psychiatry* **68**, 314–319 (2010).

4. Kaplan, G. A. *et al.* Childhood socioeconomic position and cognitive function in adulthood. *Int. J. Epidemiol.* **30**, 256–263 (2001).

5. Nelson, C. A., 3rd *et al.* Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. *Science* **318**, 1937–1940 (2007).

6. Heim, C., Newport, D. J., Mletzko, T., Miller, A. H. & Nemeroff, C. B. The link between childhood trauma and depression: insights from HPA axis studies in humans. *Psychoneuroendocrinology* **33**, 693–710 (2008).

7. Klengel, T. & Binder, E. B. Epigenetics of Stress-Related Psychiatric Disorders and Gene × Environment Interactions. *Neuron* **86**, 1343–1357 (2015).

8. Pratchett, L. C. & Yehuda, R. Foundations of posttraumatic stress disorder: does early life trauma lead to adult posttraumatic stress disorder? *Dev. Psychopathol.* **23**, 477–491 (2011).

9. Short, A. K. & Baram, T. Z. Early-life adversity and neurological disease: age-old questions and novel answers. *Nat. Rev. Neurol.* **15**, 657–669 (2019).

10. Lupien, S. J., McEwen, B. S., Gunnar, M. R. & Heim, C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat. Rev. Neurosci.* **10**, 434–445 (2009).

11. Chen, Y. & Baram, T. Z. Toward Understanding How Early-Life Stress Reprograms Cognitive and Emotional Brain Networks. *Neuropsychopharmacology* **41**, 197–206 (2016).

12. Brunson, K. L., Avishai-Eliner, S., Hatalski, C. G. & Baram, T. Z. Neurobiology of the stress response early in life: evolution of a concept and the role of corticotropin releasing hormone. *Mol. Psychiatry* **6**, 647–656 (2001).

13. de Kloet, E. R., Joëls, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nat. Rev. Neurosci.* **6**, 463–475 (2005).

14. Joëls, M. & Baram, T. Z. The neuro-symphony of stress. *Nat. Rev. Neurosci.* **10**, 459–466 (2009).

15. Ulrich-Lai, Y. M. & Herman, J. P. Neural regulation of endocrine and autonomic stress responses. *Nat. Rev. Neurosci.* **10**, 397–409 (2009).

16. Hatalski, C. G., Guirguis, C. & Baram, T. Z. Corticotropin releasing factor mRNA expression in the hypothalamic paraventricular nucleus and the central nucleus of the amygdala is modulated by repeated acute stress in the immature rat. *J. Neuroendocrinol.* **10**, 663–669 (1998).

17. Fenoglio, K. A., Brunson, K. L., Avishai-Eliner, S., Chen, Y. & Baram, T. Z. Region-specific onset of handling-induced changes in corticotropin-releasing factor and glucocorticoid receptor expression. *Endocrinology* **145**, 2702–2706 (2004).

18. Singh-Taylor, A. *et al.* NRSF-dependent epigenetic mechanisms contribute to programming of stress-sensitive neurons by neonatal experience, promoting resilience. *Mol. Psychiatry* **23**, 648–657 (2018).

19. Korosi, A. *et al.* Early-life experience reduces excitation to stress-responsive hypothalamic neurons and reprograms the expression of corticotropin-releasing hormone. *J. Neurosci.* **30**, 703–713 (2010).

20. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice & Committee on Applying Neurobiological and Socio-Behavioral Sciences from Prenatal Through Early Childhood Development: A Health Equity Approach. *Vibrant and Healthy Kids: Aligning Science, Practice, and Policy to Advance Health Equity*. (National Academies Press, 2019).

21. Gunn, B. G. *et al.* Dysfunctional astrocytic and synaptic regulation of hypothalamic glutamatergic transmission in a mouse model of early-life adversity: relevance to neurosteroids and programming of the stress response. *J. Neurosci.* **33**, 19534–19554 (2013).

22. Kapoor, A., Dunn, E., Kostaki, A., Andrews, M. H. & Matthews, S. G. Fetal programming of hypothalamo-pituitary-adrenal function: prenatal stress and glucocorticoids. *J. Physiol.* **572**, 31–44 (2006).

23. Champagne, D. L., de Kloet, E. R. & Joëls, M. Fundamental aspects of the impact of glucocorticoids on the (immature) brain. *Semin. Fetal Neonatal Med.* **14**, 136–142 (2009).

24. Bolton, J. L. *et al.* Unexpected Transcriptional Programs Contribute to Hippocampal Memory Deficits and Neuronal Stunting after Early-Life Adversity. *Cell Rep.* **33**, 108511 (2020).

25. Bale, T. L. Epigenetic and transgenerational reprogramming of brain development. *Nat. Rev. Neurosci.* **16**, 332–344 (2015).

26. Hunter, R. G. & McEwen, B. S. Stress and anxiety across the lifespan: structural plasticity and epigenetic regulation. *Epigenomics* **5**, 177–194 (2013).

27. Suderman, M. *et al.* Conserved epigenetic sensitivity to early life experience in the rat and human hippocampus. *Proc. Natl. Acad. Sci. U. S. A.* **109 Suppl 2**, 17266–17272 (2012).

28. Deussing, J. M. & Jakovcevski, M. Histone Modifications in Major Depressive Disorder and Related Rodent Models. *Adv. Exp. Med. Biol.* **978**, 169–183 (2017).

29. Gray, J. D. *et al.* Translational profiling of stress-induced neuroplasticity in the CA3 pyramidal neurons of BDNF Val66Met mice. *Mol. Psychiatry* **23**, 904–913 (2018).

30. Lucassen, P. J. *et al.* Perinatal programming of adult hippocampal structure and function; emerging roles of stress, nutrition and epigenetics. *Trends Neurosci.* **36**, 621–631 (2013).

31. McClelland, S., Korosi, A., Cope, J., Ivy, A. & Baram, T. Z. Emerging roles of epigenetic mechanisms in the enduring effects of early-life stress and experience on learning and memory. *Neurobiol. Learn. Mem.* **96**, 79–88 (2011).

32. Nestler, E. J. Epigenetic mechanisms of depression. *JAMA Psychiatry* **71**, 454–456 (2014).

33. Peña, C. J. *et al.* Early life stress confers lifelong stress susceptibility in mice via ventral tegmental area OTX2. *Science* **356**, 1185–1188 (2017).

34. Szyf, M., Tang, Y.-Y., Hill, K. G. & Musci, R. The dynamic epigenome and its implications for behavioral interventions: a role for epigenetics to inform disorder prevention and health promotion. *Transl. Behav. Med.* **6**, 55–62 (2016).

35. Aguilera, G. & Liu, Y. The molecular physiology of CRH neurons. *Front. Neuroendocrinol.* **33**, 67–84 (2012).

36. Swanson, L. W., Sawchenko, P. E., Lind, R. W. & Rho, J. H. The CRH motoneuron: differential peptide regulation in neurons with possible synaptic, paracrine, and endocrine outputs. *Ann. N. Y. Acad. Sci.* **512**, 12–23 (1987).

37. Carter, D. A. & Lightman, S. L. Diurnal pattern of stress-evoked neurohypophyseal hormone secretion: sexual dimorphism in rats. *Neurosci. Lett.* **71**, 252–255 (1986).

38. Palkovits, M. Afferents onto Neuroendocrine Cells. in *Current Topics in Neuroendocrinology* 197–222 (Springer Berlin Heidelberg, 1986).

39. Swanson, L. W. & Kuypers, H. G. The paraventricular nucleus of the hypothalamus: cytoarchitectonic subdivisions and organization of projections to the pituitary, dorsal vagal complex, and spinal cord as demonstrated by retrograde fluorescence double-labeling methods. *J. Comp. Neurol.* **194**, 555–570 (1980).

40. Xu, S. *et al.* Behavioral state coding by molecularly defined paraventricular hypothalamic cell type ensembles. *Science* **370**, (2020).

41. Romanov, R. A. *et al.* Molecular design of hypothalamus development. *Nature* **582**, 246–252 (2020).

42. Romanov, R. A., Alpár, A., Hökfelt, T. & Harkany, T. Molecular diversity of corticotropin-releasing hormone mRNA-containing neurons in the hypothalamus. *J. Endocrinol.* **232**, R161–R172 (2017).

43. Chen, R., Wu, X., Jiang, L. & Zhang, Y. Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Rep.* **18**, 3227–3241 (2017).

44. Nishi, M., Horii-Hayashi, N. & Sasagawa, T. Effects of early life adverse experiences on the brain: implications from maternal separation models in rodents. *Front. Neurosci.* **8**, 166 (2014).

45. Walker, C.-D. *et al.* Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* **20**, 421–448 (2017).

46. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* **149**, 4892–4900 (2008).

47. Molet, J., Maras, P. M., Avishai-Eliner, S. & Baram, T. Z. Naturalistic rodent models of chronic early-life stress. *Dev. Psychobiol.* **56**, 1675–1688 (2014).

48. Ivy, A. S., Brunson, K. L., Sandman, C. & Baram, T. Z. Dysfunctional nurturing behavior in rat dams with limited access to nesting material: a clinically relevant model for early-life stress. *Neuroscience* **154**, 1132–1142 (2008).

49. Bolton, J. L. *et al.* Anhedonia Following Early-Life Adversity Involves Aberrant Interaction of Reward and Anxiety Circuits and Is Reversed by Partial Silencing of Amygdala Corticotropin-Releasing Hormone Gene. *Biol. Psychiatry* **83**, 137–147 (2018).

50. Levis, S. C. *et al.* On the early life origins of vulnerability to opioid addiction. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0628-5.

51. Bolton, J. L. *et al.* Early-life adversity facilitates acquisition of cocaine self-administration and induces persistent anhedonia. *Neurobiol Stress* **8**, 57–67 (2018).

52. Molet, J. *et al.* Fragmentation and high entropy of neonatal experience predict adolescent emotional outcome. *Transl. Psychiatry* **6**, e702 (2016).

53. Naninck, E. F. G. *et al.* Chronic early life stress alters developmental and adult neurogenesis and impairs cognitive function in mice. *Hippocampus* **25**, 309–328 (2015).

54. Bath, K. G. *et al.* Early life stress leads to developmental and sex selective effects on performance in a novel object placement task. *Neurobiol Stress* **7**, 57–67 (2017).

55. Brunson, K. L. *et al.* Mechanisms of late-onset cognitive decline after early-life stress. *J. Neurosci.* **25**, 9328–9338 (2005).

56. Molet, J. *et al.* MRI uncovers disrupted hippocampal microstructure that underlies memory impairments after early-life adversity. *Hippocampus* **26**, 1618–1632 (2016).

57. Ivy, A. S. *et al.* Hippocampal dysfunction and cognitive impairments provoked by chronic early-life stress involve excessive activation of CRH receptors. *J. Neurosci.* **30**, 13005–13015 (2010).

58. Bath, K., Manzano-Nieves, G. & Goodwill, H. Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Horm. Behav.* **82**, 64–71 (2016).

59. Daviu, N. *et al.* Paraventricular nucleus CRH neurons encode stress controllability and regulate defensive behavior selection. *Nat. Neurosci.* **23**, 398–410 (2020).

60. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

61. Wamsteeker Cusulin, J. I., Füzesi, T., Watts, A. G. & Bains, J. S. Characterization of corticotropin-releasing hormone neurons in the paraventricular nucleus of the hypothalamus of Crh-IRES-Cre mutant mice. *PLoS One* **8**, e64943 (2013).

62. Ziegler, D. R., Cullinan, W. E. & Herman, J. P. Distribution of vesicular glutamate transporter mRNA in rat hypothalamus. *J. Comp. Neurol.* **448**, 217–229 (2002).

63. Füzesi, T., Daviu, N., Wamsteeker Cusulin, J. I., Bonin, R. P. & Bains, J. S. Hypothalamic CRH neurons orchestrate complex behaviours after stress. *Nat. Commun.* **7**, 11937 (2016).

64. Kim, D. W. *et al.* The cellular and molecular landscape of hypothalamic patterning and differentiation from embryonic to late postnatal development. *Nat. Commun.* **11**, 4360 (2020).

65. Yadawa, A. K. & Chaturvedi, C. M. Expression of stress hormones AVP and CRH in the hypothalamus of Mus musculus following water and food deprivation. *Gen. Comp. Endocrinol.* **239**, 13–20 (2016).

66. Zhang, Q. *et al.* Netrin-G1 regulates fear-like and anxiety-like behaviors in dissociable neural circuits. *Sci. Rep.* **6**, 28750 (2016).

67. Nosrati, N., Kapoor, N. R. & Kumar, V. Combinatorial action of transcription factors orchestrates cell cycle-dependent expression of the ribosomal protein genes and ribosome biogenesis. *FEBS J.* **281**, 2339–2352 (2014).

68. Song, H., Kim, W., Kim, S.-H. & Kim, K.-T. VRK3-mediated nuclear localization of HSP70 prevents glutamate excitotoxicity-induced apoptosis and Aβ accumulation via enhancement of ERK phosphatase VHR activity. *Sci. Rep.* **6**, 38452 (2016).

69. Xie, X., Jankauskas, R., Mazari, A. M. A., Drou, N. & Percipalle, P. β-actin regulates a heterochromatin landscape essential for optimal induction of neuronal programs during direct reprograming. *PLoS Genet.* **14**, e1007846 (2018).

70. Qian, J. *et al.* Knockout of Zn transporters Zip-1 and Zip-3 attenuates seizure-induced CA1 neurodegeneration. *J. Neurosci.* **31**, 97–104 (2011).

71. Martin, S. A. *et al.* GSK3β Regulates Brain Energy Metabolism. *Cell Reports* vol. 23 1922–1931.e4 (2018).

72. Schlingensiepen, K.-H. *et al.* The role of Jun transcription factor expression and phosphorylation in neuronal differentiation, neuronal cell death, and plastic adaptationsin vivo. *Cellular and Molecular Neurobiology* vol. 14 487–505 (1994).

73. Jhamandas, J. H. & Goncharuk, V. Role of neuropeptide FF in central cardiovascular and neuroendocrine regulation. *Front. Endocrinol.* **4**, 8 (2013).

74. Villanueva, C., Jacquier, S. & de Roux, N. DLK1 is a somato-dendritic protein expressed in hypothalamic arginine-vasopressin and oxytocin neurons. *PLoS One* **7**, e36134 (2012).

75. Kaltezioti, V. *et al.* Prox1 regulates the notch1-mediated inhibition of neurogenesis. *PLoS Biol.* **8**, e1000565 (2010).

76. Lee, M. *et al.* Tcf7l2 plays crucial roles in forebrain development through regulation of thalamic and habenular neuron identity and connectivity. *Dev. Biol.* **424**, 62–76 (2017).

77. Murillo, B., Ruiz-Reig, N., Herrera, M., Fairén, A. & Herrera, E. Zic2 Controls the Migration of Specific Neuronal Populations in the Developing Forebrain. *J. Neurosci.* **35**, 11266–11280 (2015).

78. Bittermann, E. *et al.* Differential requirements of tubulin genes in mammalian forebrain development. *PLoS Genet.* **15**, e1008243 (2019).

79. Didonna, A. & Opal, P. The role of neurofilament aggregation in neurodegeneration: lessons from rare inherited neurological disorders. *Mol. Neurodegener.* **14**, 19 (2019).

80. Liu, Y., Sugiura, Y. & Lin, W. The role of synaptobrevin1/VAMP1 in Ca2+-triggered neurotransmitter release at the mouse neuromuscular junction. *J. Physiol.* **589**, 1603–1618 (2011).

81. Pigoni, M. *et al.* Seizure protein 6 and its homolog seizure 6-like protein are physiological substrates of BACE1 in neurons. *Mol. Neurodegener.* **11**, 67 (2016).

82. Srivastava, A. K. *et al.* Mutant HSPB1 overexpression in neurons is sufficient to cause age-related motor neuronopathy in mice. *Neurobiol. Dis.* **47**, 163–173 (2012).

83. Weeks, S. D. *et al.* Characterization of human small heat shock protein HSPB1 α-crystallin domain localized mutants associated with hereditary motor neuron diseases. *Sci. Rep.* **8**, 688 (2018).

84. Pacheco, A., Merianda, T. T., Twiss, J. L. & Gallo, G. Mechanism and role of the intra-axonal Calreticulin translation in response to axonal injury. *Exp. Neurol.* **323**, 113072 (2020).

85. Pigoni, M., Gunnersen, J. M. & Lichtenthaler, S. F. Seizure-6 proteins highlight BACE1 functions in neurobiology. *Oncotarget* **8**, 7214–7215 (2017).

86. Timón-Gómez, A. *et al.* Mitochondrial cytochrome c oxidase biogenesis: Recent developments. *Semin. Cell Dev. Biol.* **76**, 163–178 (2018).

87. Jacob, T. C., Moss, S. J. & Jurd, R. GABA(A) receptor trafficking and its role in the dynamic modulation of neuronal inhibition. *Nat. Rev. Neurosci.* **9**, 331–343 (2008).

88. Gonzalez-Nunez, V. Role of , GABA receptor alpha-2 subunit, in CNS development. *Biochem Biophys Rep* **3**, 190–201 (2015).

89. Sasaki, J. *et al.* The PtdIns(3,4)P(2) phosphatase INPP4A is a suppressor of excitotoxic neuronal death. *Nature* **465**, 497–501 (2010).

90. Liu, T. *et al.* Functional human antibody CDR fusions as long-acting therapeutic endocrine agonists. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1356–1361 (2015).

91. El Boustani, M. *et al.* A Guide to PIN1 Function and Mutations Across Cancers. *Front. Pharmacol.* **9**, 1477 (2018).

92. Delgado, J. Y. & Owens, G. C. The cytochrome c gene proximal enhancer drives activity-dependent reporter gene expression in hippocampal neurons. *Front. Mol. Neurosci.* **5**, 31 (2012).

93. Malygin, A. A., Parakhnevitch, N. M., Ivanov, A. V., Eperon, I. C. & Karpova, G. G. Human ribosomal protein S13 regulates expression of its own gene at the splicing step by a feedback mechanism. *Nucleic Acids Res.* **35**, 6414–6423 (2007).

94. van Oterendorp, C. *et al.* The expression of dynein light chain DYNLL1 (LC8-1) is persistently downregulated in glaucomatous rat retinal ganglion cells. *Exp. Eye Res.* **92**, 138–146 (2011).

95. King, S. M. Dynein-independent functions of DYNLL1/LC8: redox state sensing and transcriptional control. *Sci. Signal.* **1**, e51 (2008).

96. Baines, A. C., Adams, E. J., Zhang, B. & Ginsburg, D. Disruption of the Sec24d gene results in early embryonic lethality in the mouse. *PLoS One* **8**, e61114 (2013).

97. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

98. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

99. Tong, M., Jun, T., Nie, Y., Hao, J. & Fan, D. The Role of the Slit/Robo Signaling Pathway. *J. Cancer* **10**, 2694–2705 (2019).

100. Qu, D. *et al.* Structure, functional regulation and signaling properties of Rap2B. *Oncol. Lett.* **11**, 2339–2346 (2016).

101. Huang, C. *et al.* Cbx3 maintains lineage specificity during neural differentiation. *Genes Dev.* **31**, 241–246 (2017).

102. Zequn, N. *et al.* The role and potential mechanisms of LncRNA-TATDN1 on metastasis and invasion of non-small cell lung cancer. *Oncotarget* **7**, 18219–18228 (2016).

103. Yasumoto, H., Meng, L., Lin, T., Zhu, Q. & Tsai, R. Y. L. GNL3L inhibits activity of estrogen-related receptor gamma by competing for coactivator binding. *J. Cell Sci.* **120**, 2532–2543 (2007).

104. Meng, L., Hsu, J. K. & Tsai, R. Y. L. GNL3L depletion destabilizes MDM2 and induces p53-dependent G2/M arrest. *Oncogene* **30**, 1716–1726 (2011).

105. Gould E., Woolley C.S., McEwen B.S. The hippocampal formation: morphological changes induced by thyroid, gonadal and adrenal hormones. *Psychoneuroendocrinology* **16**, 67- 84 (1991)

106. Radley JJ, Sisti HM, Hao J, Rocher AB, McCall T, Hof PR, McEwen BS, Morrison JH Chronic behavioral stress induces apical dendritic reorganization in pyramidal neurons of the medial prefrontal cortex. *Neuroscience* **125**, 1-6 (2004)

107. Wood G.E., Young L.T., Reagan L.P., Chen B., McEwen B.S. Stress-induced structural remodeling in hippocampus: prevention by lithium treatment. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3973-3978 (2004)

108. Hajszan T., Dow A, Warner-Schmidt J.L., Szigeti-Buck K., Sallam N.L., Parducz A., Leranth C., Duman R.S. Remodeling of hippocampal spine synapses in the rat learned helplessness model of depression. *Biological psychiatry* **65**, 392-400 (2009)

109. Bath, K. G., Manzano-Nieves, G., & Goodwill, H. Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Hormones and Behavior*, **82**, 64–71 (2016)

110. Uchida, S. *et al.* Learning-induced and stathmin-dependent changes in microtubule stability are critical for memory and disrupted in ageing. *Nat. Commun.* **5**, 4389 (2014).

111. Brocke, B. *et al.* Stathmin, a gene regulating neural plasticity, affects fear and anxiety processing in humans. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **153B**, 243–251 (2010).

112. Chen, B., Piel, W. H., Gui, L., Bruford, E. & Monteiro, A. The HSP90 family of genes in the human genome: insights into their divergence and evolution. *Genomics* **86**, 627–637 (2005).

113. Moradi, M. *et al.* Differential roles of α-, β-, and γ-actin in axon growth and collateral branch formation in motoneurons. *J. Cell Biol.* **216**, 793–814 (2017).

114. Chidambaranathan-Reghupaty, S., Mendoza, R., Fisher, P. B. & Sarkar, D. The multifaceted oncogene SND1 in cancer: focus on hepatocellular carcinoma. *Hepatoma Res* **4**, (2018).

115. Xu, W. *et al.* Circulating lncRNA SNHG11 as a novel biomarker for early diagnosis and prognosis of colorectal cancer. *Int. J. Cancer* **146**, 2901–2912 (2020).

116. Oyang, E. L., Davidson, B. C., Lee, W. & Poon, M. M. Functional characterization of the dendritically localized mRNA neuronatin in hippocampal neurons. *PLoS One* **6**, e24879 (2011).

117. Karki, S., Tokito, M. K. & Holzbaur, E. L. A dynactin subunit with a highly conserved cysteine-rich motif interacts directly with Arp1. *J. Biol. Chem.* **275**, 4834–4839 (2000).

118. Hermann, A. *et al.* The Hippo pathway component Wwc2 is a key regulator of embryonic development and angiogenesis in mice. *Cell Death Dis.* **12**, 117 (2021).

119. Rhind, N. & Russell, P. Chk1 and Cds1: linchpins of the DNA damage and replication checkpoint pathways. *J. Cell Sci.* **113 (Pt 22)**, 3889–3896 (2000).

120. Klauck, S. M. *et al.* Mutations in the ribosomal protein gene RPL10 suggest a novel modulating disease mechanism for autism. *Mol. Psychiatry* **11**, 1073–1084 (2006).

121. Berto, G. E. *et al.* The DCR protein TTC3 affects differentiation and Golgi compactness in neurons through specific actin-regulating pathways. *PLoS One* **9**, e93721 (2014).

122. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

123. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

124. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

Chapter 4

Methylation signatures of diverse early life experiences in Human Infants: a novel, within-subject approach

Notes:

(1) Dr. Noriko Kamei and I equally contributed to the material in this chapter. She designed and performed experiments, as well as provided suggestions on interpreting the results. I analyzed all of the data and interpreted the results.

(2) Cassie McGill performed quality checks and sequenced experimental samples.

(3) Melanie Oakes provided guidance for sequencing on NovaSeq platform

(4) Xinyi Ma performed some quality checks and sequenced experimental samples.

(5) Dr. Hal S. Stern provided valuable suggestions on statistical analysis

(6) Dr. Tallie Z. Baram and Dr. Ali Mortazavi conceived the idea and provided continuedsupport and guidance throughout the project.

Chapter 4

Methylation signatures of diverse early life experiences in Human Infants: a novel, within-subject approach

4.1 Abstract

Early life adversity (ELA) is one of the largest environmental risk factors for abnormal brain circuit development and the development of psychiatric diseases. The brain is particularly sensitive to environmental stimuli during early brain development which, in humans, probably centers around the first two years of life [1]. The development and maturation of neurons and neural connections depend on both genetic and environmental inputs. Mechanisms encoded by underlying neural epigenetic pathways established during ELA contribute to later pathologies found in adults [2,3]. It is largely unknown how much epigenetic mechanisms such as DNA methylation contribute and whether there are epigenetic markers that can be used as a predictor of adverse effects due to ELA.

The overall goal of this study is to establish a marker for the impact of diverse experiences on the developing human infant. We use an intra-individual approach to test DNA methylation changes that occur between birth and the first year of life in humans over a spectrum of experiences. We employed a non-invasive buccal swab collection method on 77 individuals to investigate intra-individual DNA methylation changes as measured using reduced representation bisulfite sequencing (RRBS) in human babies in the same individuals at birth and after 1 year. We compare these changes in methylation to a signature of ELA called unpredictability and found that samples may have changes in methylation within an individual that correlate with this score. These preliminary results will need to be validated using the methylene of more individuals along with their unpredictability scores.

4.2 Introduction

Early life adversity (ELA) is a blanket term that includes adverse experiences in early life that includes abuse, neglect, malnourishment, poverty, and inconsistent maternal and caregiver inputs [3–5]. ELA experiences pose a major threat to an individual's mental development with potential long-term consequences for both mental and overall health throughout life [6,7]. Although ELA is present in all societies and social classes, the epigenetic encoding and consequences of ELA as well as its long-term effects are largely unknown and understudied [4]. There are many reasons for this, including the complexity of early brain development and the moral implications of studying human children undergoing adversity without intervention. For this reason, many ELA studies are conducted in model organisms such as rodents and non-human primates [8–10]. Animal models have allowed many advancements in the field, but there are some drawbacks to using animal models. Brain development occurs on a vastly different timescale in humans when compared to rodents [11,12]. Non-human primate brain development also differs in developmental time scale compared to humans [10,12]. It is possible to use specific milestones of brain development that can be compared between species. However, due to species-specific differences in brain development, it can be challenging to study global brain development across species [11]. Furthermore, studies using model organisms typically use two very distinct groups of individuals: an ELA group with typically identical genotypes that is exposed to a substantial amount of ELA in an otherwise completely controlled environment, and a normal development, or control, group with the same genotype and environment. However, human experiences usually occur over a spectrum of experiences, some of which contribute to ELA with many confounding environmental and genetic factors. It is important to study the epigenetic effects of ELA in humans to better understand the interaction between the spectrum of experiences with genetics and the environment.

Chromatin modifications have been shown to be dynamically regulated in the developing brain [7,13–15]. DNA methylation changes provided epigenetic markers of ELA in rats between a baseline control of normal maternal nurturing and an ELA model of adverse life experiences [16]. In rats, the ELA experiences showed increased methylation in genes associated with growth and metabolism, suggesting decreased expression of these genes. The authors also found a decrease in DNA methylation, which suggests increased expression, in genes associated with development, inflammation, and apoptosis. These changes suggest that peripheral cells in rats experiencing ELA have abnormal development and differentiation timelines, and they are experiencing more cellular stressors [16]. With differential methylation in so many interesting gene families in a model system, we hypothesize that similar predictive epigenetic markers could be found in humans.

Early life parental sensory signals can shape the epigenetic programming of brain development, which can promote vulnerability or resilience to disorders presenting later in life [17]. In rodent models of ELA, strong ELA paradigms combined with identical genotypes, and controlled environments allow for comparisons to typical maternal inputs under similar controlled conditions. However, in humans, genetic variability, diverse environments, and quality of the infant-caregiver relationship add many variables to studying ELA [17]. The unpredictability measurement, or entropy score, is based on defined mathematical characteristics of patterns of behavior to provide a prediction of developmental outcomes [17–19]. An individual's behavior can be characterized by observing and recording the actions of the individual interacting with their environment. Entropy measures the predictability of a single random variable having a random probability distribution with some uncertainty, and the behavioral entropy rate is an expansion of this concept to a group of random variables defined on a common probability space over time [18]**.** This allows the degree of predictability of maternal sensory signals (or caregivers) for each individual to be quantified and scored [18,19].

The overall goal of this study is to determine if ELA during a sensitive period of brain development in humans would lead to a detectable DNA methylation signature between a spectrum of ELA experiences. Here, we employ a preliminary analysis of a novel method of investigating intra-individual DNA methylation differences between human early life experiences, or unpredictability. We aim to determine a robust, cost effective method of obtaining and analyzing DNA methylation data from human buccal swab samples. We used the same reduced representation bisulfite sequencing (RRBS) method previously used in a rat study of ELA[16] to process 154 buccal-swab samples from 77 individuals obtained after birth (newborn, sample A) and after one year of life (one-year-old, sample B). While this is a preliminary analysis that will need to be expanded, we have found evidence of a similar epigenetic signature in a spectrum of human early life experiences as previously seen in rats with controlled genetic, environmental, and ELA conditions [16].

4.3 Results

4.3.1 Experimental set-up and analysis of RRBS samples between 77 individuals

We used non-invasive buccal swabs to obtain a mix of epithelial and white blood cell (WBC) DNA from individual human babies twice, once as a newborn (A) and once after one year of life (B) (Fig. 4.1). We used a non-invasive method to collect the samples because it is less stressful than a needle prick used to obtain a blood sample, and a buccal swab is a routine procedure [20]. Using a non-invasive method of sample collection is especially important since the stress caused due to the sample collection may influence the methylation profile of the sample [16,20,21]. We used intra-individual methylation signatures between newborn and one-year-old samples from the same individual within a spectrum of early life experiences [16].

We used RRBS to assess DNA methylation (Fig. 4.1, and Fig. 4.2). RRBS uses bisulfite conversion to differentiate unmethylated cytosines (Cs) from methylated Cs [16,22]. The unmethylated Cs are converted to uracils (Us), which are then converted to thymines (Ts) in PCR amplification. These converted Cs allow the accurate differentiation of methylated and unmethylated Cs [22] (Fig. 4.2). RRBS samples were sequenced using either the NextSeq500 (NextSeq) and NovaSeq6000 (NovaSeq) Illumina sequencing platforms with libraries sequenced to at depth of at least 10 million mapped reads (Fig. 4.1).

After sequencing, Illumina adapters were trimmed from reads using TrimGalore! [23] and mapped to a bisulfite-converted reference genome using Bismark and bowtie2 (Fig. 4.3) [24]. We detected an average of 1,632,737 CpGs in both samples of the same individual. We performed differential methylation analysis (using methylKit) on sites with a minimum coverage (methylated + unmethylated samples) of 10 reads in the newborn (A) and one-year-old (B) samples for each individual.

4.3.2 Inter-individual analysis separated samples based on age in 34 individuals (68 samples) sequenced on the NextSeq platform

First, we analyzed the samples sequenced using the NextSeq platform and identified a total of 21,748 significantly differentially methylated sites (DMSs) in at least one individual (Fig. 4.4). We concatenated DMSs shared in at least 1 individual into 100 base pair (bp) regions, a process called tiling (Fig. 4.3). These sites were then tiled into 13,503 differentially methylated regions (DMRs) (Fig. 4.4).

DNA methylation levels of DMRs in newborn (A) and one-year-old (B) samples were analyzed across individuals, some of which have unpredictability scores (Table 4.1), using dimensional reduction methods. Similar to what was observed in rats between high and low

unpredictability groups [16], inter-individual principal component analysis (PCA) on the percentage of DNA methylation of 13,503 DMRs in both the A and B samples from each individual showed clear differences between age groups using the first two principal components (PCs) (Fig. 4.5) (21.9% of the variances explained). We then colored the samples on the same PCA plot by their unpredictability scores (Table 4.1) and found that the samples did not separate by their unpredictability scores (Fig. 4.6) in the first two PCs. This suggests that buccal swab methods between individual A and B samples indicate the epigenetic signature of age, but not life experience.

4.3.3 Intra-individual changes in methylation may be able to distinguish early-life experience in individuals sequenced on the NextSeq platform

In order to determine the epigenetic changes occurring during different experiences we wanted to compare intra-individual (or within an individual) differences in the dataset. In order to accomplish this, we compared the delta methylation occurring within an individual between the newborn and one-year-old time-points. Similar to Jiang et al, we calculated intra-individual fold changes in the epigenetic signature within the same individual, or delta methylation [16]. To calculate this, the A and B samples for each individual were used to determine the delta-methylation within an individual ($\log_2(B/A)$).

We analyzed the intra-individual changes for each individual using PCA and determined that individuals could be separated by unpredictability score. PC1 and 2 include up to 21.1% of the variances. PC2 could be separating the samples by their unpredictability score (Fig. 4.7), although imperfectly. however, since we do not have scores for every individual, we can only predict which PC is able to separate the samples based on their life experience. Taken

together, more samples are needed to determine if delta-methylation can truly separate individuals based on their unpredictability scores.

4.3.4 Inter-individual analysis separated samples based on age in all 77 individuals (154 samples)

Using the NextSeq platform we had to sequence the samples on average 2.6 times before we achieved the mapping read depth of 10 million mapped reads (Table 4.2). Due to this we decided to use a newer sequencing platform, the NovaSeq6000, to sequence additional samples from 43 individuals and reduce the number of times each sample needs to be sequenced.

The samples sequenced on the NextSeq platform were sequenced with a length of 43bps and the samples sequenced on the NovaSeq platform were sequenced to a length of 100bps. The NextSeq platform is the first benchtop sequencing platform from Illumina and can generate up to 500 million reads per sequencing run and the NovaSeq platform is the most powerful Illumina platform on the market and can generate up to 20 billion reads per sequencing run [25,26].

A total of 77 individuals, or 154 samples, were sequenced on either the NextSeq or NovaSeq platforms. We Identified a total of 44,542 significant DMSs in at least one individual. We tiled 9,726 DMSs shared in at least 2 individuals into 4,821 DMRs (Figure 4.8). When combining two different sample preparation methods, we need to take batch effects into consideration. We analyzed A and B samples for each individual by PCA. We found that samples separate by age in PC1 (Fig. 4.9A) and by sequencing method on PC2 with 34.1% of the variance explained by both PCs (Fig. 4.9B). As in the individual analysis of samples

sequenced exclusively on the NextSeq platform, the individual A and B samples did not separate by unpredictability scores (Fig. 4.10).

4.3.5 Intra-individual methylation may be able to distinguish early-life experience in all individuals

The age and sequencing method effects vanished after investigating the fold change within an individual (delta-methylation). Intra-individual analysis does not show differences between the two sequencing platforms, as individuals sequenced from each platform are intermixed (Fig. 4.11). PC1 seems to reasonably separate the samples by their unpredictability score, illustrated by the green line (Fig 4.12). Most of the high unpredictability scored samples are on the right side of 0 on PC1, and most of the individuals with low unpredictability scores are on the left of 0 on PC1 (14.7% explained variance).

Now that we have determined that PC1 can separate levels of delta-methylation between individuals based on their early life experiences. We set the cutoff threshold for significance at 0.015 and -0.015 (Fig. 4.13). We determined that the highest weights that pass the significant threshold included 1,129 DMRs which are associated with 918 genes. The most negative weights accounted for 555 DMRs which are associated with 417 genes (Fig. 4.13). We found that the top weighted DMRs were more methylated in the B samples with a higher unpredictability score and the lowest weighted DMRs were less methylated in B samples with higher unpredictability scores. Genes that have increased DNA methylation generally have decreased gene expression.

We next wanted to investigate the genes with the most positive and most negative PC1 weights. Genes found to be associated with DMRs with the most positive weights (suggesting they have decreased expression) are enriched in GO terms such as cell junction organization

(which included 48/918 annotated, or 5.3%, of genes such as *CTNNA1, KIFC3, SMAD3,* and *ACTN4*), actin filament-based process (which included 82/918 annotated, or 9%, of genes including *SMAD6, EGFR, PLXNB1,* and *WNT3*), cell morphogenesis involved in differentiation (which included 125/918 annotated, or 13.6%, of genes such as *NOTCH3, KLF4, CARM1, SEPTIN9, PBX1,* and *TCF3*), and positive regulation of GTPase activity (which included 48/918 annotated, or 5.3%, of genes such as *LIMS1, FGFR3, TNNT3*, and *MTCH1*) (Fig. 4.14A).

We analyzed the significant GO terms with genes associated with DMRs from the most negative PC1 weights (suggesting increased expression) (Fig. 4.14B). Enriched GO terms included negative regulation of cell proliferation (which included 28/417 annotated, or 6.7%, of genes such as *ALOX5, NOTCH1, OSM, SOX9,* and *NKX2-8*), cellular response to growth factor stimulus (which included 25/417 annotated, or 6%, of genes including *PAX9, TB1, WNT10A,* and *PRDM14*), positive regulation of cell motility (which included 26/417 annotated, or 6.2%, of genes such as *RHOB, NOS3, PECAM1,* and *S100P*), and regulation of system process (which included 26/417 annotated, or 6.2%, of genes such as *GRK2, P2RX1, PDGFB, TBX2,* and *CUX2*).

Taken together, early life experiences cause decreases in methylation (or increases in expression) in genes associated with cell proliferation, energy regulation, as well as growth and increases in methylation (decreases in expression) in genes associated with differentiation, neuronal projection, actin filament processes, and cell junction organization. The genes associated with decreased methylation are involved in neuronal differentiation and growth, suggesting that neuronal-associated genes have increased expression in individuals with high unpredictability scores

4.4 Discussion

In summary, we sequenced RRBS samples collected using buccal swabs from 77 individuals, 18 of which have known unpredictability scores. We used two different sequencing methods; 34 individuals were sequenced using the NextSeq500 platform, 43 individuals were sequenced using the NovaSeq6000 platform.

Previous studies investigating genome-wide DNA methylation over time have found that there is a large variability in methylation between subjects, this is especially true in human studies with a lot of genetic and environmental variability [27,28]. However, this effect has also been seen in genetically similar, environmentally controlled animal models [16], suggesting that an individual's DNA methylation profile may be highly sensitive to the individual's age at collection. Epigenetic changes due to age and the high level of variability between individuals, make it difficult to assess changes that occur due to environmental paradigms and caregiver-child interactions.

In order to investigate changes due to environmental stressors it is important to remove inter-individual differences and focus on changes that occur within an individual over time. To overcome these hurdles, we investigated a novel method of investigating the epigenetic change that occurs within an individual between two time points during a very sensitive developmental time period. The delta-methylation between birth and the first year of life covers a very sensitive developmental time in human development [3,7,29]. By investigating the intra-individual differences in DNA methylation between timepoints taken at the newborn stage and after the first year of life, we exclude inter-individual differences, including age and sequencing method, and focus on the differences that occur within an individual during this very sensitive time in brain development.

As expected [16,30,31], when the samples are analyzed together, from all 77 individuals, individual A and B samples are separated by age, and not unpredictability score. Samples are also separated by sequencing method on PC2 (Fig. 4.9B). Using delta-methylation analysis of intra-individual differences mitigate the effects of age. Delta-methylation also mitigates the effect of sequencing method (Fig. 4.11), because an individual's samples were sequenced using the same platform. We found that the individuals could be separated by their unpredictability scores by using delta-methylation (the change between the two samples obtained from the same individual)[16]. These mitigations remove large inter-individual variances. More samples are needed to determine if the samples can be robustly separated in this manner. Taken together, the data suggests that individuals sequenced using either of these two sequencing platforms can be separated by unpredictability scores using delta-methylation and intra-individual analysis. However, the NovaSeq platform is much more efficient at sequencing samples to the required read depth.

PCA analysis of intra-individual methylation separated individuals classified by high and low unpredictability groups via PC1 (Fig. 4.12). PC1 identified the highest and lowest weighted genes associated with an individual's unpredictability score that were differentially methylated in the one-year-old stage compared with the same individuals in the newborn stage. These genes may be important in predicting genes involved in early life experiences and their possible long-term consequences. GO analyses of the top contributing PC1 genes for individuals in the high unpredictability group found a striking enrichment of increased methylation (suggesting reduced expression) in genes involved in cell junction organization, actin filament-based process, and differentiation. Decreased methylation, suggesting overexpression, of gene families associated with cell proliferation, growth factor stimulus, and cell motility. How an individual's experiences affect these changes remains unclear.

In conclusion, when comparing inter-individual differences between both A and B samples in each individual the epigenetic signatures in each sample separate by age. In order to separate individuals by their early life experiences, we determined the fold change between the A and B samples in each individual using delta-methylation. Using delta-methylation we were able to separate the individuals by experience and found that there were distinct epigenetic changes between experiences. These changes in an individual's epigenetic signatures may be able to predict future outcomes in mental health. However, more data and analysis are needed to solidify these results.

4.5 Figures

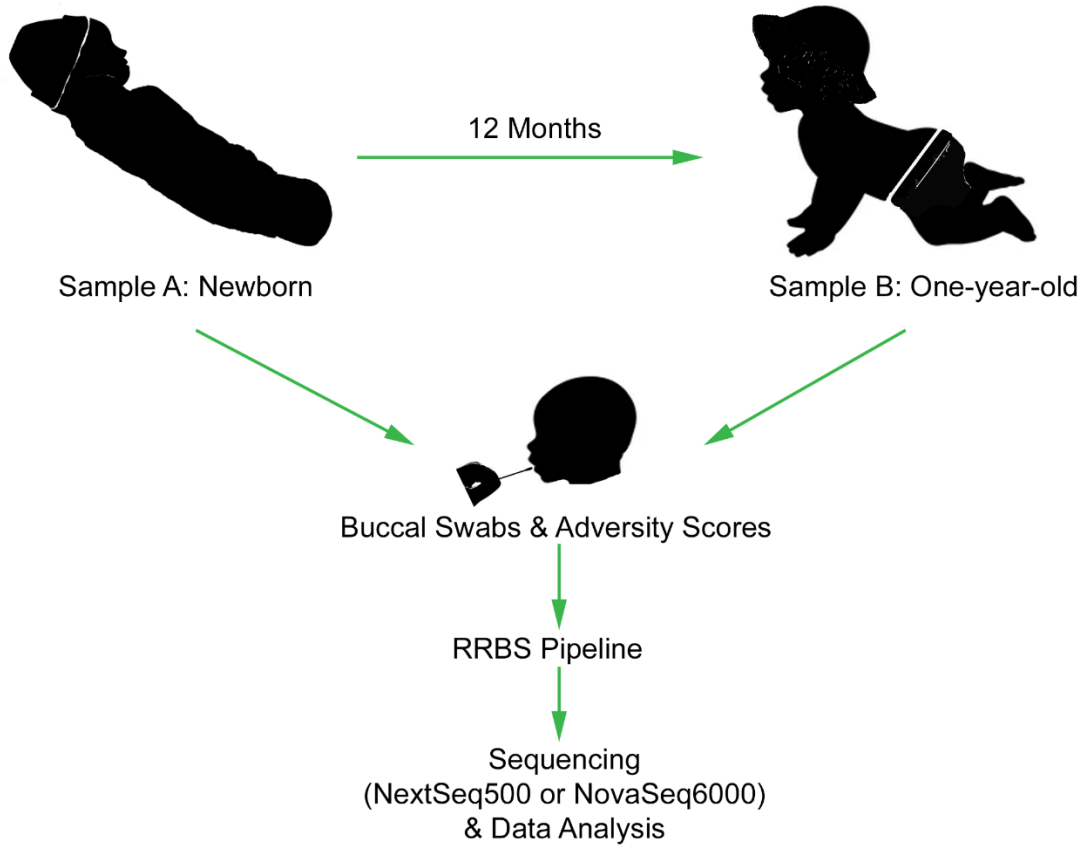

Figure 4.1 Sample collection and experimental timeline

Buccal swab samples were taken as a newborn, sample A, and 12 months later, sample B. A spectrum of ELA conditions measured via an unpredictability score which is calculated using surveys and in-lab observations. Buccal swabs were then processed using the RRBS pipeline and sequenced.

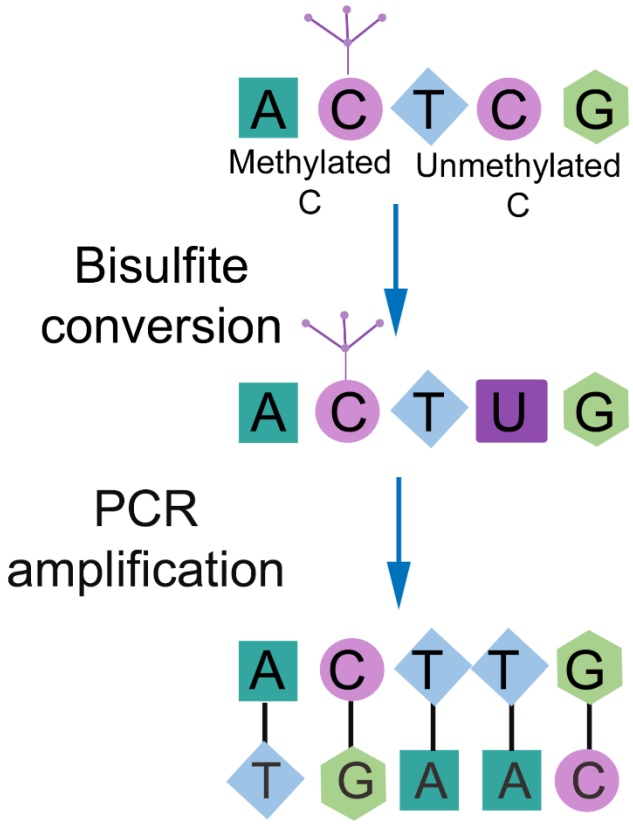

Figure 4.2 Bisulfite conversion and PCR amplification in RRBS

Unmethylated C's are converted to U's during bisulfite conversion after which the converted DNA was amplified via PCR, converting Us to Ts. The converted DNA sequences were then sequenced using Illumina technology.

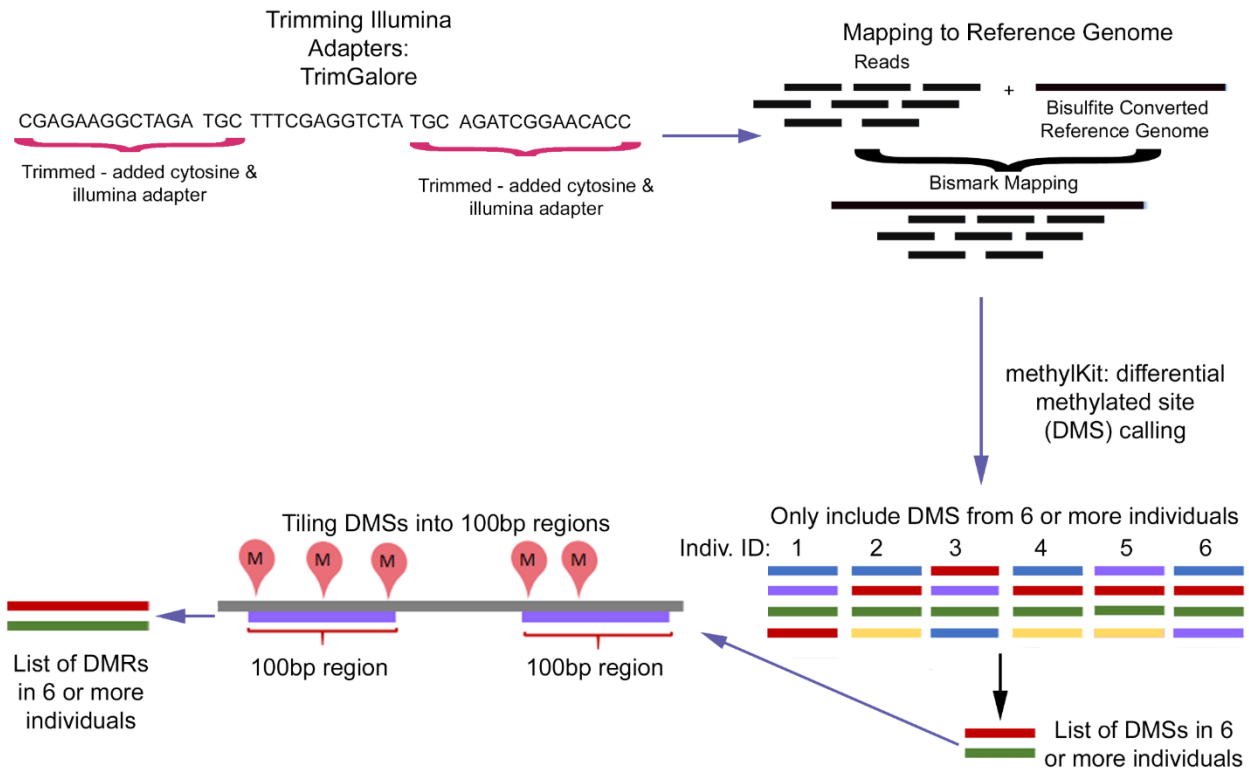

Figure 4.3 Computational pipeline for analyzing RRBS data

Sequencing reads were adapter trimmed using TrimGalore! Samples sequenced using the NovaSeq6000 platform were also trimmed to a bp length of 43 to match sequencing reads from the NextSeq500. After trimming the reads were mapped to a bisulfite-converted reference genome using Bismark. Differential methylation was determined using methylKit between an individual's A and B samples. These differentially methylated sites (DMSs) are then compared between individuals. Only sites that were found in over 5 individuals were tiled into 100bp regions. These regions were used for downstream analysis.

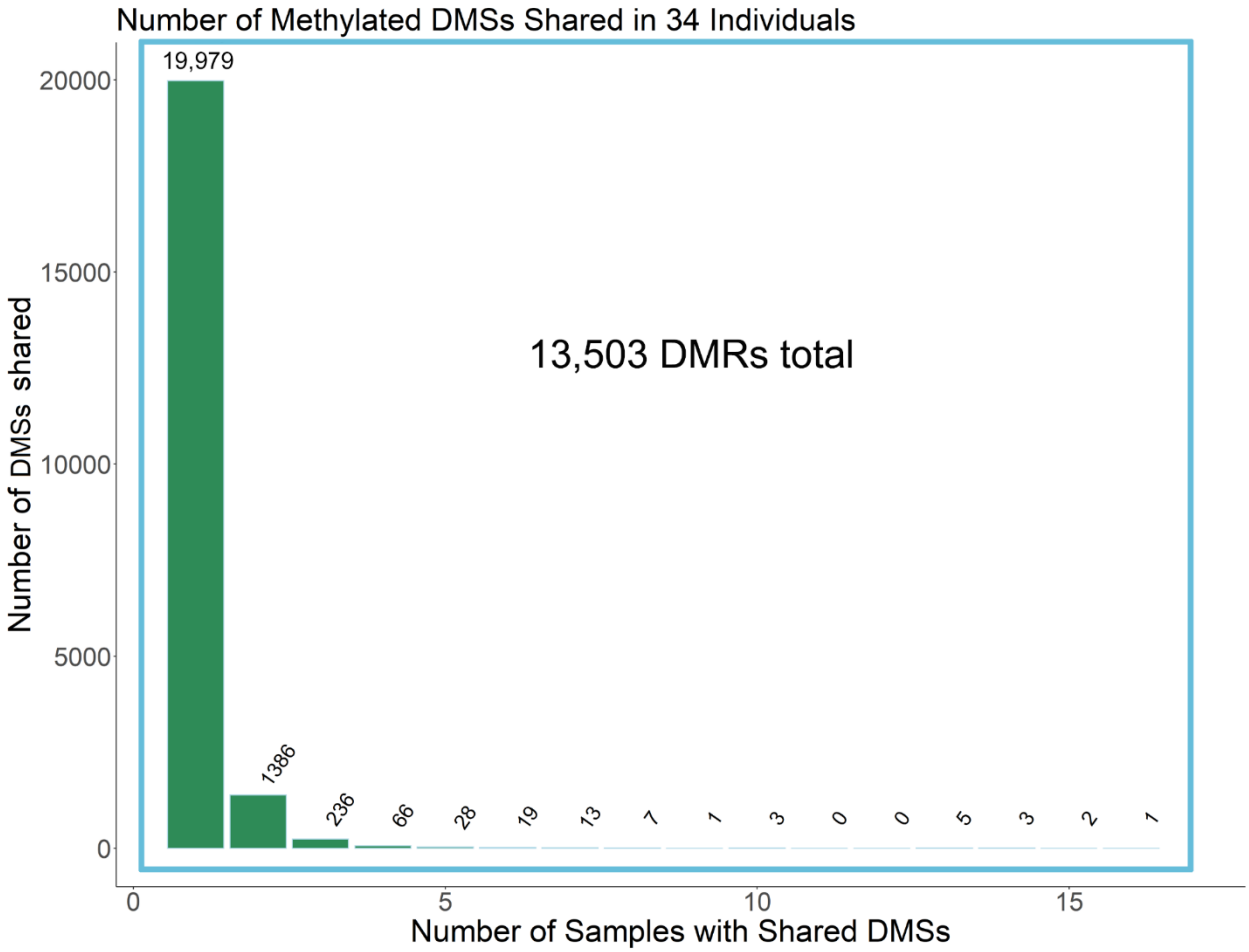

Figure 4.4 Number of shared DMSs found in 34 individuals with samples sequenced using the NextSeq500 platform

A total of 21,748 DMSs were found in at least one individual in a total of 34 individuals. These DMSs were then tiled into 13,503 regions of 100 base pairs. The regions were then used to combine the samples into a matrix.

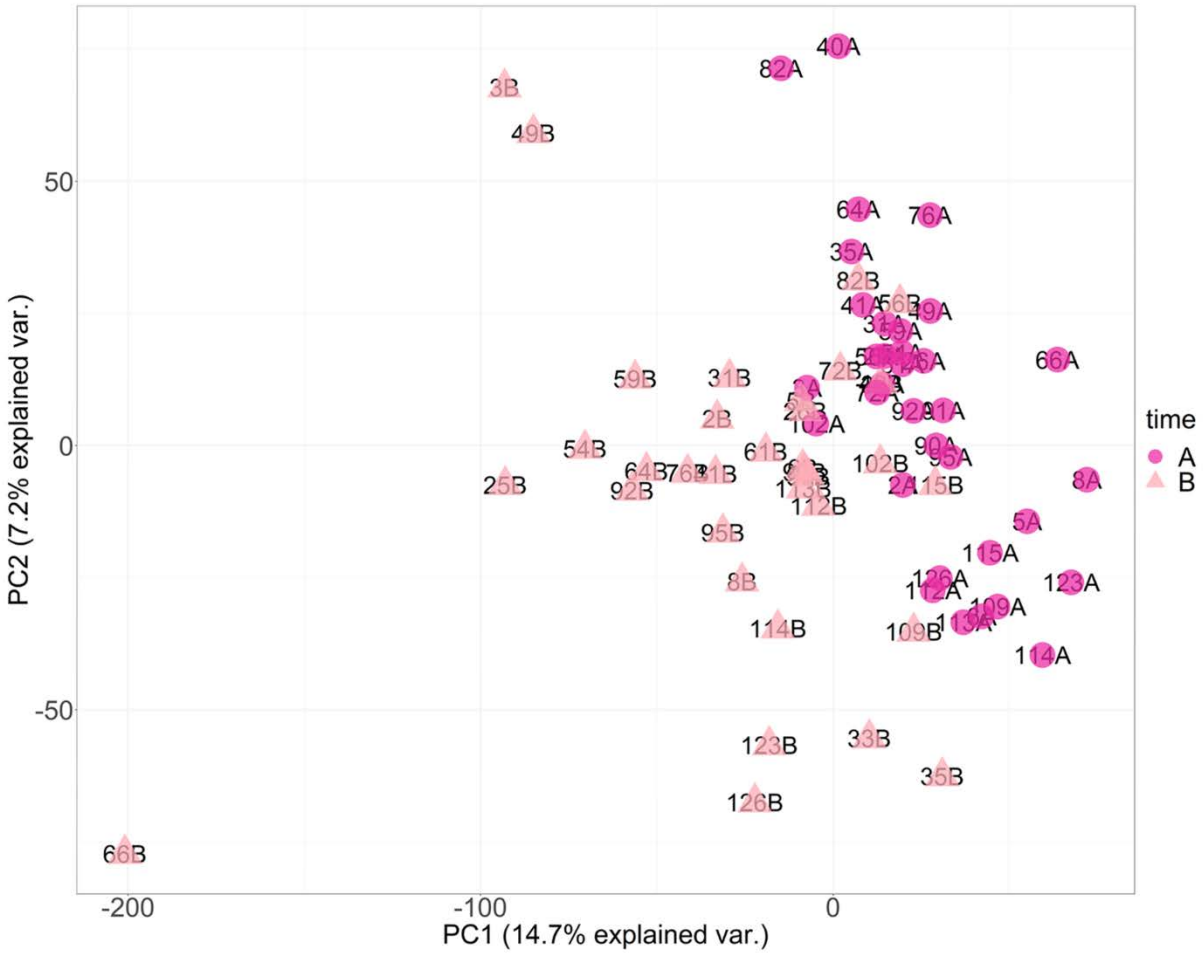

Figure 4.5 Separation by age when using methylation levels of 13,503 significant DMRs.

Focusing on these 13,503 DMRs in 68 samples separate by age on a PCA of methylation profiles. Individual samples are labeled by age; dark pink: A, or newborn samples and, light pink: B, or one-year-old samples.

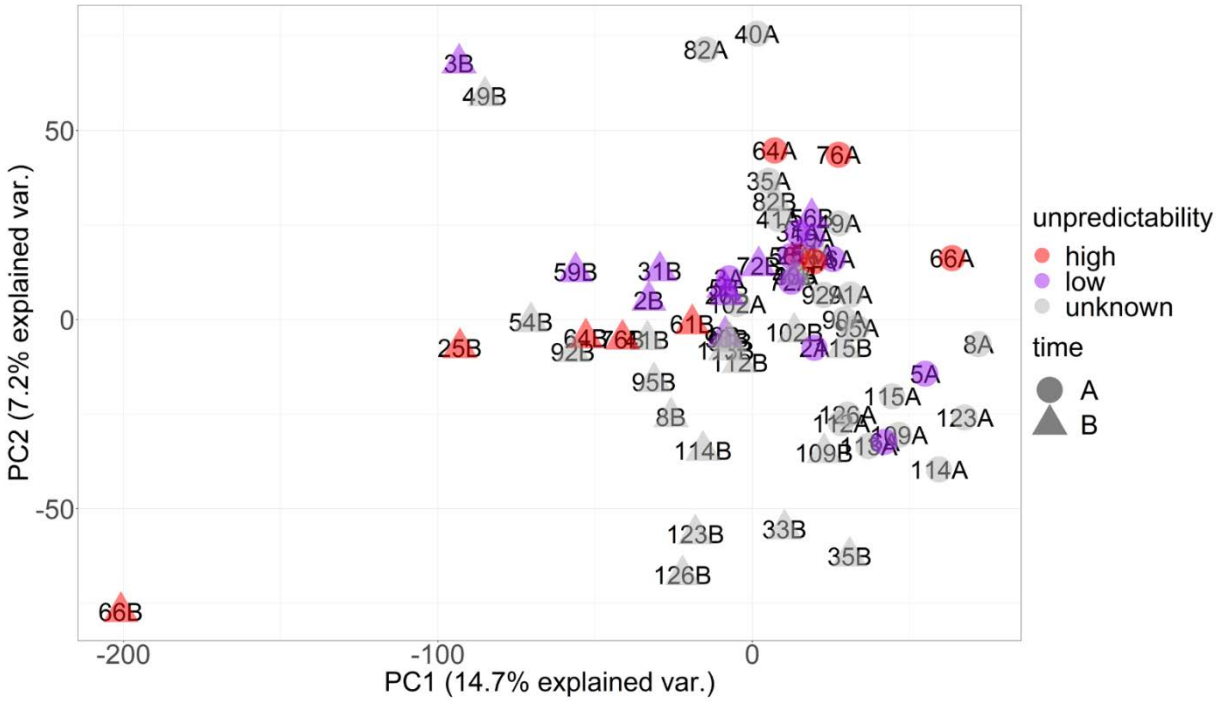

Figure 4.6 PCA of samples focusing on the same 13,503 DMRs used previously, colored by high and low unpredictability.

PCA of 68 samples labeled by unpredictability scores. Samples do not separate by unpredictability.

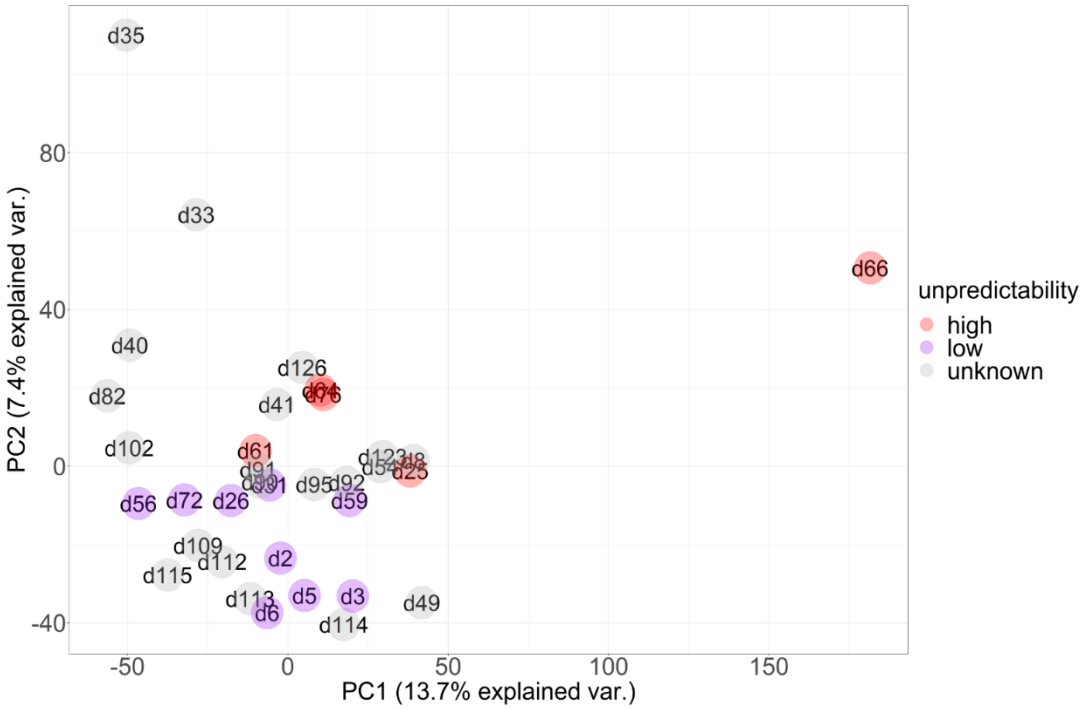

Figure 4.7 PCA of delta methylation between B (one-year-old) and A (newborn) (log2(B/A))

PCA of delta methylation focusing on the same 13,503 DMRs. Samples are labeled by high or low unpredictability scores.

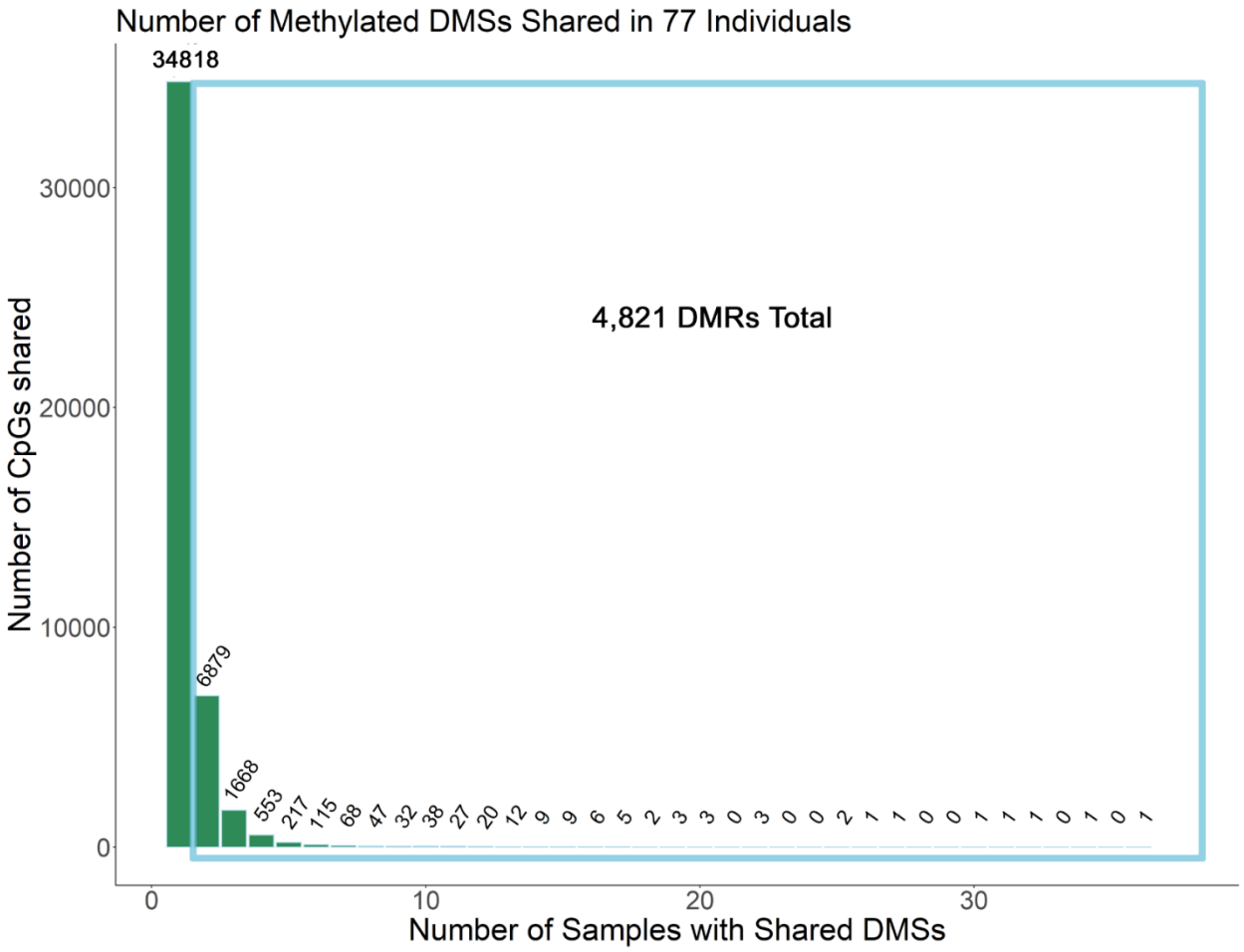

Figure 4.8 Number of shared DMSs found in 77 individuals with samples sequenced using the NextSeq500 and NovaSeq6000 platforms

A histogram of 44,542 DMSs was found in at least one individual in a total of 77 individuals. We consolidated 9,726 DMSs found in at least 2 individuals into 4,821 DMRs tiled into 100 base pairs. The regions were then used to combine the samples into a matrix.

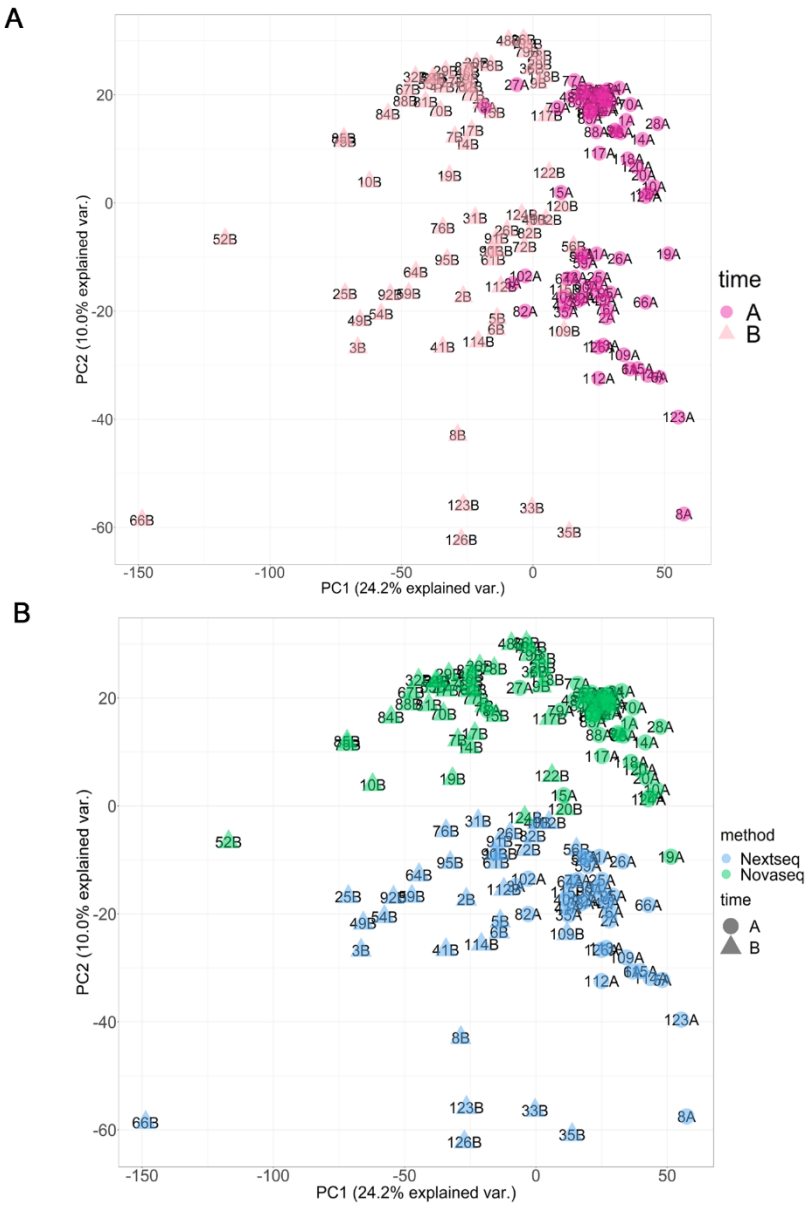

Figure 4.9 Separation by age and sequencing method when using methylation levels of 4,821 significant DMRs.

A) PCA of methylation profiles focusing on these 4,821 DMRs in 154 samples separate by age. Individual samples are labeled by age; dark pink: A, or newborn samples and, light pink: B, or one-year-old samples. B) PCA of the same profiles labeled by sequencing method. Blue (NextSeq500). Green (NovaSeq6000).

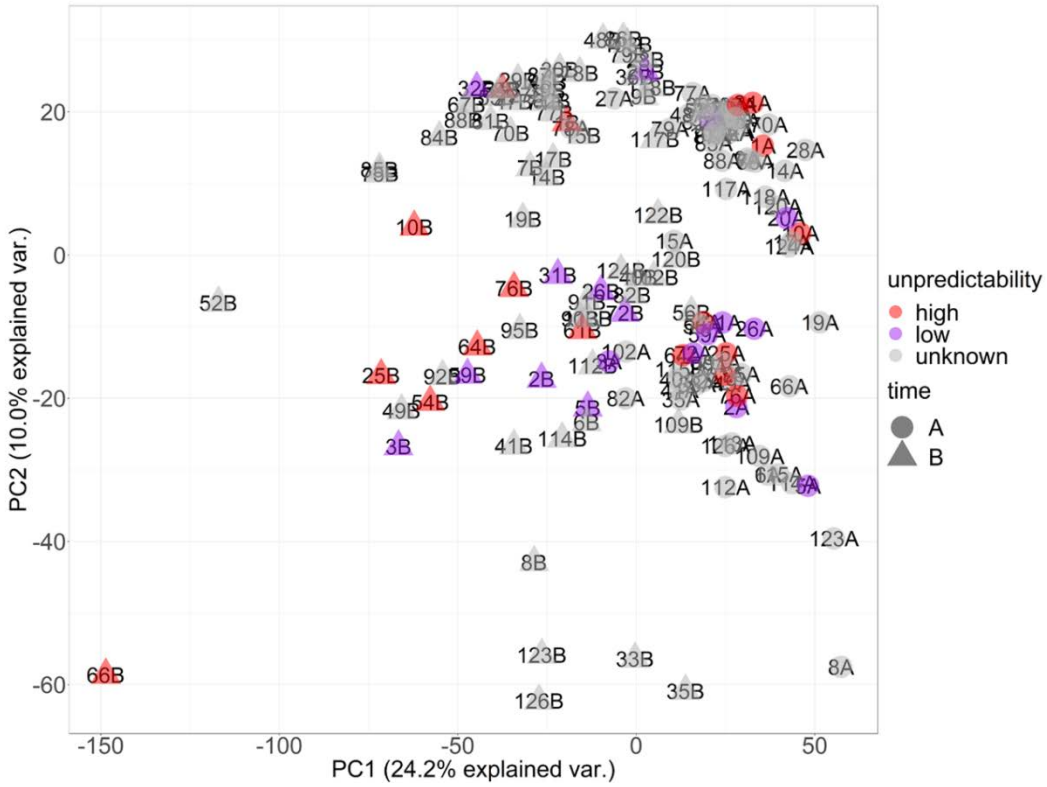

Figure 4.10 PCA of samples focusing on the same 4,821 DMRs used previously, colored by high and low unpredictability.

PCA of 154 samples labeled by unpredictability scores. Samples do not separate by unpredictability.

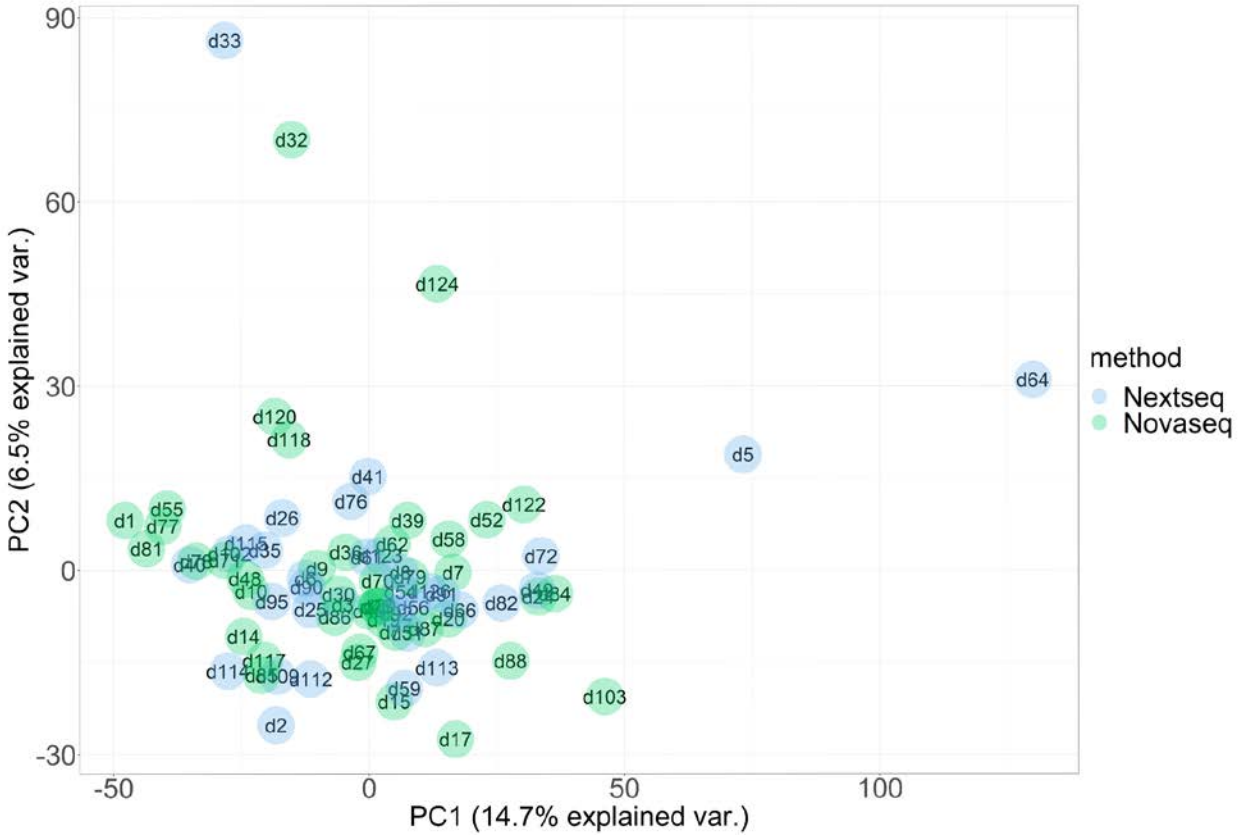

Figure 4.11 PCA of delta methylation between B and A samples colored by sequencing method

PCA of delta methylation focusing on 4,821 DMRs. Samples are labeled by whether the individuals were sequenced using the NextSeq500 (blue) or NovaSeq6000 (green) platforms..

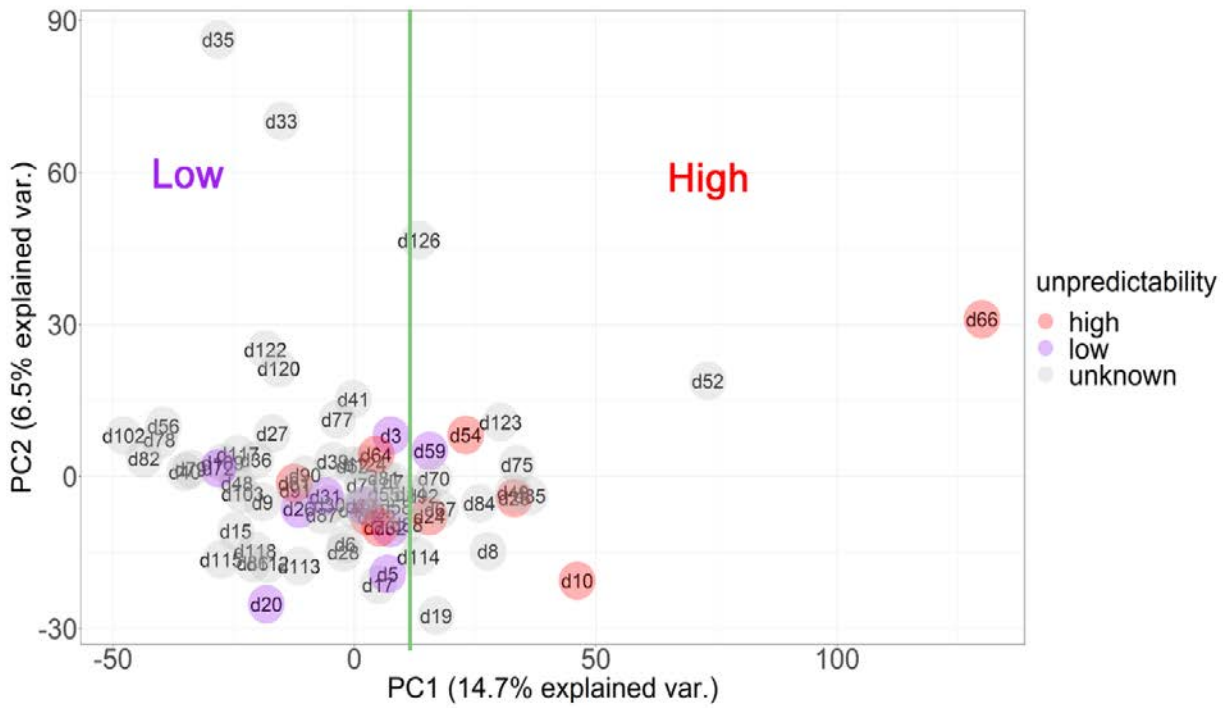

Figure 4.12 PCA of delta methylation between B and A samples colored by unpredictability score

PCA of delta methylation focusing on the same 4,821 DMRs. Samples are labeled by high or low unpredictability scores. Green line separates unpredictability scores in PC1.

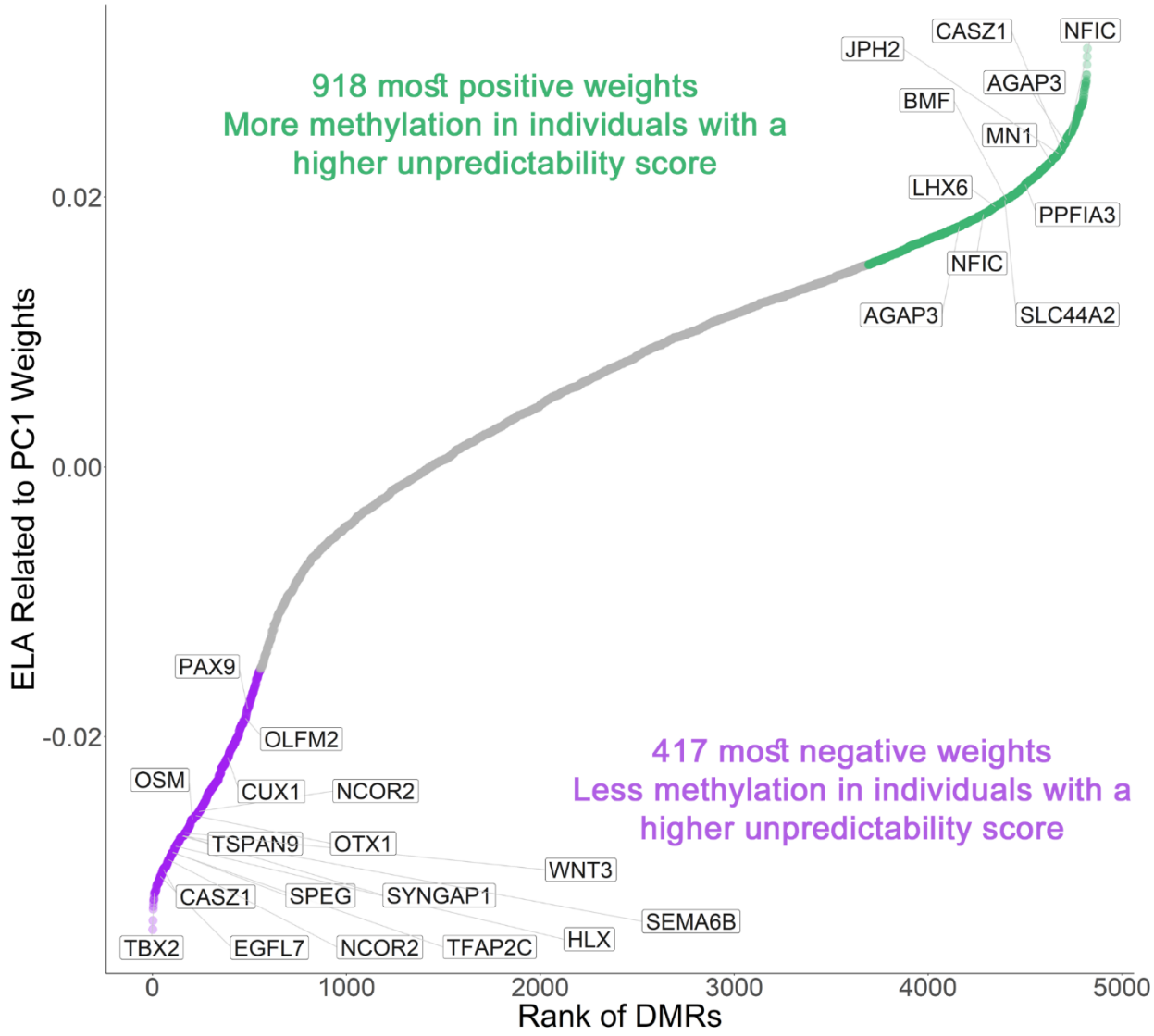

Figure 4.13 Expected consequences of unpredictability-related DMRs

PC1 weight analysis that separates individuals that scored in the high unpredictability group and the low unpredictability group. Out of the 918 most positive weights, we found 24 genes overlapping with the rat ELA methylation paper [16]. Out of 417 most negative weights, we found 29 genes overlapping with the rat ELA paper [16]. Labeled genes are some of the overlapping one-to-one orthologs found in both rat and human PCs separating ELA.

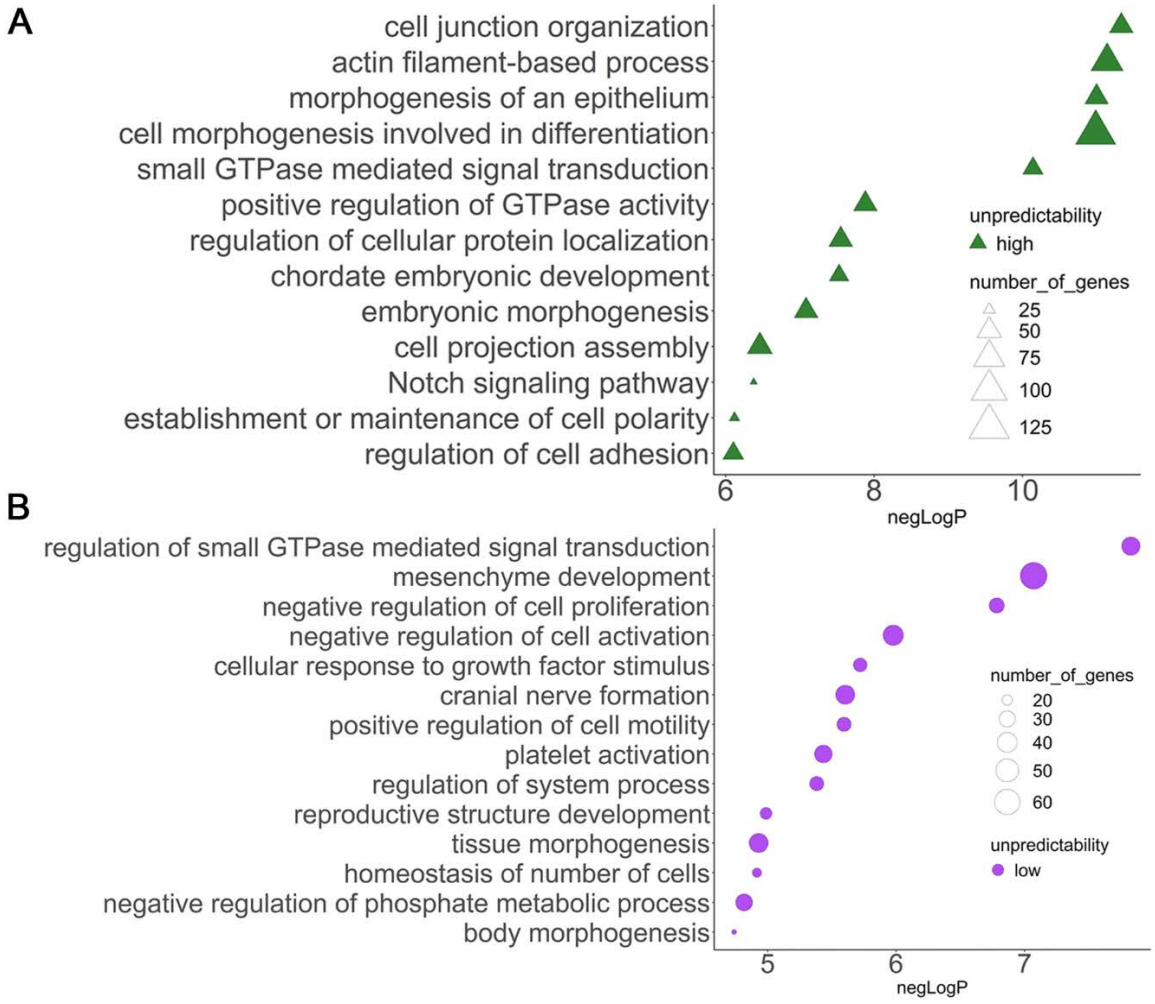

Figure 4.14 GO analysis of top and bottom PC1 weights

A) GO terms of 918 genes associated with most positively weighted DMRs (green). B) GO terms of 417 genes associated with most negatively weighted DMRs (purple).

4.6 Tables

Individual	Unpredictability score	Category	Sequencing method
1	1.095519996	High	NovaSeq
2	0.687346189	Low	NextSeq
3	0.694968214	Low	NextSeq
5	0.182986409	Low	NextSeq
6		Unknown	NextSeq
7		Unknown	NovaSeq
8		Unknown	NextSeq
9		Unknown	NovaSeq
10	1.006394521	High	NovaSeq
14		Unknown	NovaSeq
15		Unknown	NovaSeq
17		Unknown	NovaSeq
19		Unknown	NovaSeq
20	0.763673634	Low	NovaSeq
24	1.042096825	High	NovaSeq
25	1.104387948	High	NextSeq
26	0.744199101	Low	NextSeq
27		Unknown	NovaSeq
28		Unknown	NovaSeq
29		Unknown	NovaSeq
30		Unknown	NovaSeq
31	0.856400897	Low	NextSeq
32	0.84812394	Low	NovaSeq
33		Unknown	NextSeq
35		Unknown	NextSeq
36		Unknown	NovaSeq
39		Unknown	NovaSeq
40		Unknown	NextSeq
41		Unknown	NextSeq
46		Unknown	NovaSeq
47		Unknown	NovaSeq
48		Unknown	NovaSeq
49		Unknown	NextSeq
52		Unknown	NovaSeq
54	1.252115733	High	NextSeq
55		Unknown	NovaSeq
56		Unknown	NextSeq

Individual	Unpredictability score	Category	Sequencing method
58		Unknown	NovaSeq
59	0.8051204	Low	NextSeq
61	1.177409808	High	NextSeq
62		Unknown	NovaSeq
64	1.040572921	High	NextSeq
66	1.16943665	High	NextSeq
67		Unknown	NovaSeq
70		Unknown	NovaSeq
71		Unknown	NovaSeq
72	0.763285066	Low	NextSeq
75		Unknown	NovaSeq
76	1.031929041	High	NextSeq
77		Unknown	Novaseq
78		Unknown	Novaseq
79		Unknown	Novaseq
81		Unknown	Novaseq
82		Unknown	Nextseq
84		Unknown	Novaseq
85		Unknown	Novaseq
86		Unknown	Novaseq
87		Unknown	Novaseq
88		Unknown	Novaseq
90		Unknown	Nextseq
91		Unknown	Nextseq
92		Unknown	Nextseq
95		Unknown	Nextseq
102		Unknown	Nextseq
103		Unknown	Novaseq
109		Unknown	Nextseq
112		Unknown	Nextseq
113		Unknown	Nextseq
114		Unknown	Nextseq
115		Unknown	Nextseq
117		Unknown	Novaseq
118		Unknown	Novaseq
120		Unknown	Novaseq
122		Unknown	Novaseq
123		Unknown	Nextseq
124		Unknown	Novaseq
126		Unknown	Nextseq

Table 4.1 Unpredictability scores, categories and sequencing methods for each individual

Sample A	# times sequenced	Sample B	# times sequenced	Sequencing method
1	1	1	1	NovaSeq
2	3	2	3	NextSeq
3	2	3	1	NextSeq
5	3	5	2	NextSeq
6	2	6	3	NextSeq
7	1	7	1	NovaSeq
8	1	8	6	NextSeq
9	1	9	1	NovaSeq
10	1	10	1	NovaSeq
14	1	14	1	NovaSeq
15	1	15	1	NovaSeq
17	1	17	1	NovaSeq
19	1	19	1	NovaSeq
20	1	20	1	NovaSeq
24	1	24	1	NovaSeq
25	3	25	3	NextSeq
26	3	26	4	NextSeq
27	1	27	1	NovaSeq
28	1	28	1	NovaSeq
29	1	29	1	NovaSeq
30	1	30	1	NovaSeq
31	3	31	3	NextSeq
32	1	32	1	NovaSeq
33	4	33	5	NextSeq
35	2	35	4	NextSeq
36	1	36	1	NovaSeq
39	1	39	1	NovaSeq
40	2	40	4	NextSeq
41	2	41	4	NextSeq
46	1	46	1	NovaSeq
47	1	47	2	NovaSeq
48	2	48	1	NovaSeq
49	2	49	2	NextSeq
52	1	52	1	NovaSeq
54	4	54	4	NextSeq
55	2	55	1	NovaSeq
56	3	56	3	NextSeq
58	1	58	1	NovaSeq
59	3	59	3	NextSeq
61	4	61	3	NextSeq
62	1	62	1	NovaSeq
64	2	64	3	NextSeq
66	1	66	3	NextSeq
67	1	67	2	NovaSeq
70	1	70	1	NovaSeq
71	1	71	1	NovaSeq
72	3	72	3	NextSeq
75	1	75	1	NovaSeq
76	3	76	3	NextSeq
77	1	77	2	Novaseq
78	1	78	1	Novaseq
79	1	79	1	Novaseq
81	1	81	1	Novaseq
82	3	82	4	Nextseq
84	1	84	1	Novaseq
85	1	85	1	Novaseq
86	1	86	1	Novaseq
87	1	87	1	Novaseq
88	1	88	1	Novaseq
90	1	90	3	Nextseq
91	1	91	1	Nextseq
92	1	92	3	Nextseq
95	1	95	1	Nextseq
102	2	102	1	Nextseq
103	1	103	1	Novaseq
109	2	109	2	Nextseq
112	1	112	1	Nextseq
113	3	113	1	Nextseq
114	2	114	5	Nextseq
115	1	115	5	Nextseq
117	1	117	1	Novaseq
118	1	118	1	Novaseq
120	2	120	1	Novaseq
122	1	122	1	Novaseq
123	1	123	3	Nextseq
124	1	124	1	Novaseq
126	1	126	1	NextSeq

Table 4.2 Number of times each sample was sequenced

4.7 Methods

4.7.1 Collection of buccal swabs

The buccal swab/saliva was collected using saliva sponges with OG-250 (DNA Genotek. Inc.) from each human baby on one week (newborn) and one year after birth following the manufacturer's recommendation. The institutional IRB approved the study protocol, and the written informed consent was obtained from mothers.

4.7.2 Isolation and quantification of DNA for making RRBS libraries from human buccal swab samples

Each buccal swab/saliva sample in OG-250 was incubated in a 50 °C water bath for 1 hour. Then all liquid was collected from each sample. We added reagent prepIT•L2P (DNA Genotek, Inc.) at 1/25 of the sample volume. Each sample was then incubated for 10 min on ice, and centrifuged at max for 10 min, to isolate crude solution containing DNA and eliminate all precipitate. DNA was prepared using the Quick-gDNA™ MiniPrep kit (Zymo Research) following the manufacturer's protocol. The quantity of double-stranded DNA was analyzed using Qubit, and RRBS Libraries were prepared from 200 ng of genomic DNA digested with Msp I, and then extracted with DNA Clean & Concentrator™-5 kit (Zymo Research). Fragments were ligated to pre-annealed adapters containing 5'-methylcytosine instead of cytosine according to Illumina's specified guidelines (www.illumina.com). Adaptor-ligated fragments were then bisulfite-treated using the EZ DNA Methylation-Lightning™ Kit (Zymo Research). Preparative-scale PCR was performed, and the resulting products were purified with DNA Clean & Concentrator for sequencing. Amplified RRBS libraries were quantified and qualified by Qubit, Bioanalyzer (Agilent), and Kapa library quant (Kapa systems), and then sequenced on the Illumina NextSeq 500 as well as NovaSeq 6000 platforms.

4.7.3 RRBS data processing and detection of differentially methylated regions (DMRs)

The Illumina adapters were trimmed using Trim Galore v0.6.6 [23] with Cutadapt v2.10 [32] and FastQC v0.11.9 [33] to apply quality and adapter trimming to Fastq files for RRBS-type libraries using the parameters '--fastqc --stringency 5 --rrbs --length 30 --non_directional'. Reads were mapped to the human genome (hg38) we used Bismark v0.22.3 [24] in '--non_directional' mode. Each read is aligned to a pre-converted form of the reference genome using Bowtie2 v8.4.0 [34]. CpG sites were called using Bismark methylation extractor to extract methylation information from the Bismark alignment output [24]. Single CpG sites with more than ten reads coverage were kept for DMS calling. Differential methylation sites (DMSs) were first called using MethyKit (R version 4.0.2) with a false discovery rate (FDR) less than 0.05. Python v3.8.0 was used to 'tile' DMSs falling within 100 base pairs. Sites were merged as differentially methylated regions (DMRs).

4.7.4 DNA methylation percentage calculation and Delta methylation

The methylation percentage was calculated as the ratio of the methylated read counts over the sum of both methylated and unmethylated read counts for a single DMS or DMR. Delta methylation was calculated by the log2 transformation of the ratio of methylation level in the one-year-old sample (B) and the methylation level in the newborn sample (A). Increased methylation in B is shown as a positive value while decreased methylation in B is shown as a negative value.

4.7.5 PCA and K-Means clustering

Principle component analysis (PCA) was performed by using the prcomp function using R version 4.0.2. PCA figures were made using ggplot2 in R 4.0.2 [35]. A DNA methylation heatmap was generated with the ComplexHeatmap function in R 4.0.2 [36].

4.7.6 Gene association analysis

Genes associated with DMRs were identified using Homer v4.11 [37]. For subsequent analyses, annotations were kept if DMRs were located within 20kb of transcription start site (TSS) and DMRs were located in gene exons or introns. Gene ontology analysis was performed using metascape [38] with p-value less than 0.05.

4.8 References

1. Nelson, C. A., 3rd *et al.* Cognitive recovery in socially deprived young children: the Bucharest Early Intervention Project. *Science* **318**, 1937–1940 (2007).

2. Lux, V. Epigenetic Programming Effects of Early Life Stress: A Dual-Activation Hypothesis. *Curr. Genomics* **19**, 638–652 (2018).

3. Short, A. K. & Baram, T. Z. Early-life adversity and neurological disease: age-old questions and novel answers. *Nat. Rev. Neurol.* **15**, 657–669 (2019).

4. Ramo-Fernández, L. *et al.* The effects of childhood maltreatment on epigenetic regulation of stress-response associated genes: an intergenerational approach. *Sci. Rep.* **9**, 1–12 (2019).

5. Eriksson, M., Räikkönen, K. & Eriksson, J. G. Early life stress and later health outcomes--findings from the Helsinki Birth Cohort Study. *Am. J. Hum. Biol.* **26**, 111–116 (2014).

6. Heim, C. & Nemeroff, C. B. The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biol. Psychiatry* **49**, 1023–1039 (2001).

7. Bale, T. L. *et al.* Early life programming and neurodevelopmental disorders. *Biol. Psychiatry* **68**, 314–319 (2010).

8. Walker, C.-D. *et al.* Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* **20**, 421–448 (2017).

9. Nishi, M., Horii-Hayashi, N. & Sasagawa, T. Effects of early life adverse experiences on the brain: implications from maternal separation models in rodents. *Front. Neurosci.* **8**, 166 (2014).

10. Wakeford, A. G. P., Morin, E. L., Bramlett, S. N., Howell, L. L. & Sanchez, M. M. A review of nonhuman primate models of early life stress and adolescent drug abuse. *Neurobiol Stress* **9**, 188–198 (2018).

11. Birnie, M. T. *et al.* Plasticity of the Reward Circuitry After Early-Life Adversity: Mechanisms and Significance. *Biol. Psychiatry* **87**, 875–884 (2020).

12. Avishai-Eliner, S. Stressed-out, or in (utero)? *Trends in Neurosciences* vol. 25 518–524 (2002).

13. Peña, C. J. & Nestler, E. J. Progress in Epigenetics of Depression. *Prog. Mol. Biol. Transl. Sci.* **157**, 41–66 (2018).

14. Singh-Taylor, A. *et al.* NRSF-dependent epigenetic mechanisms contribute to programming of stress-sensitive neurons by neonatal experience, promoting resilience. *Mol. Psychiatry* **23**, 648–657 (2018).

15. McClelland, S., Korosi, A., Cope, J., Ivy, A. & Baram, T. Z. Emerging roles of epigenetic mechanisms in the enduring effects of early-life stress and experience on learning and memory. *Neurobiology of Learning and Memory* vol. 96 79–88 (2011).

16. Jiang, S. *et al.* Intra-individual methylomics detects the impact of early-life adversity. *Life Sci Alliance* **2**, (2019).

17. Baram, T. Z. *et al.* Fragmentation and unpredictability of early-life experience in mental disorders. *Am. J. Psychiatry* **169**, 907–915 (2012).

18. Vegetabile, B. G., Stout-Oswald, S. A., Davis, E. P., Baram, T. Z. & Stern, H. S. Estimating the Entropy Rate of Finite Markov Chains With Application to Behavior Studies. *Journal of Educational and Behavioral Statistics* vol. 44 282–308 (2019).

19. Davis, E. P. *et al.* Exposure to unpredictable maternal sensory signals influences cognitive development across species. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10390–10395 (2017).

20. Said, M. *et al.* Genomics in premature infants: a non-invasive strategy to obtain high-quality DNA. *Sci. Rep.* **4**, 4286 (2014).

21. Lowe, R. *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics* **8**, 445–454 (2013).

22. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).

23. Krueger, F. Trim Galore! http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

24. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

25. Welcome to immense discovery power. https://www.illumina.com/systems/sequencing-platforms/novaseq.html.

26. NextSeq Series Specifications. https://www.illumina.com/systems/sequencing-platforms/nextseq/specifications.html.

27. Wang, D. *et al.* Individual variation and longitudinal pattern of genome-wide DNA methylation from birth to the first two years of life. *Epigenetics* **7**, 594–605 (2012).

28. Schroeder, J. W. *et al.* Neonatal DNA methylation patterns associate with gestational age. *Epigenetics* **6**, 1498–1504 (2011).

29. Vaiserman, A. M. Epigenetic programming by early-life stress: Evidence from human populations. *Dev. Dyn.* **244**, 254–265 (2015).

30. Reizel, Y. *et al.* Postnatal DNA demethylation and its role in tissue maturation. *Nat. Commun.* **9**, 2040 (2018).

31. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).

32. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* vol. 17 10 (2011).

33. Andrews, S. FastQC: a quality control tool for high throughput sequence data. *FastQC: a quality control tool for high throughput sequence data* http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).

34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

35. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer Science & Business Media, 2009).

36. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

37. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell* vol. 38 576–589 (2010).

38. Zhou, Y. *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

Chapter 5

Explanations/Hypotheses and Future Directions

Chapter 5

Explanations/Hypotheses and Future Directions

5.1 Advancements in ATAC-mRNA integrative analysis for inter-species comparison

In Chapter 2, we used a daily time-course of RNA and ATAC sequencing to map gene expression and open chromatin dynamics during differentiation into both neural progenitor cells (NPCs) and definitive endoderm (DE) in rat and human. We identify conserved regulatory modules using gene expression and chromatin accessibility changes of lineage-specific regulators in both species. Using this method, we were able to connect gene expression changes with associated changes of open chromatin regions during differentiation.

We correlate clusters of mRNAs with mean expression and ATAC clusters with similar mean open chromatin dynamics to identify potential candidate cis-regulatory elements (cCRMs). One assumption for the correlation analysis is that the changes in the expression of mRNA clusters in respective lineages are mainly driven by chromatin dynamics. Changes in transcription factor (TF) binding site accessibility near the promoter and enhancer cis-regulatory modules (CRMs) of other TFs collectively form gene regulatory networks (GRNs) [1,2].

We can predict potential binding sites of TFs to CRMs using the deeply sequenced ATAC data to determine footprints to build GRNs, which we investigated in Chapter 2.We investigated differentially expressed genes that also have differential open chromatin accessibility surrounding a TFs transcriptional start site (TSS). We investigated whether differentially expressed TFs may be regulated by motifs (TF binding sites) within open chromatin regions or, even better, footprints within open chromatin regions. This study concluded that 8% of candidate cis-regulatory elements with significant transcriptional overlaps are shown to be

conserved in NPC differentiation whereas 22% of significantly overlapping candidate cis-regulatory are shown to be conserved in DE differentiation between rat and human

There is a lack of software or packages that can determine links between footprinting data and TF expression data in bulk data. There are software packages that have been developed for single-cell (sc)RNA and scATAC-seq methods, but they have not been investigated using bulk data with much smaller sample sizes [3]. Developing software for predicting GRN links from time-course bulk RNA and ATAC data with data from multiple time points may be a way to get around the smaller sample sizes normally associated with bulk experiments.

Predictions are not enough on their own, so the interesting interactions which are found in Chapter sections 2.3.6 and 2.3.7, including genes such as *GBX2, EGR1,* and *PBX1* for NPC differentiation and *ETV1, GBX2,* and *TGIF1* for DE differentiation, can be investigated further using either perturbations or other sequencing methods such as ChIP-seq to determine higher resolution relationships between TFs, cREs, and gene expression. Changes in gene expression are also regulated by other transcriptional regulatory mechanisms such as methylation. We can use perturbation, ChIP-seq, and methylation data, or a combination of the listed methods to find regulators of expression during differentiation. We can use the collective data to build GRNs or to validate GRN interactions we obtained from RNA and ATAC data.

As mentioned previously, integrating scRNA-seq and scATAC-seq data can be used to build predictive GRNs. By conducting time course experiments similar to what we did in Chapter 2, we can collect scRNA and scATAC daily time points in order to further increase the sample size and the ability to predict GRN regulatory links. Since scRNA and scATAC-seq software packages have already been built [3–5], we can use single-cell data and single-cell GRNs to validate the predictive GRN models from bulk data.

Rodents, especially rats and mice, are the most utilized model organisms for biomedical research, particularly in neuroscience [6,7]. Until recently mice were predominantly used for research, however, rats are becoming increasingly popular, [7–9]. Rats are genetically different from mice, and using rats as a model organism can have clear advantages over mice [6]. Because of this recent shift to using rat as a model it is becoming increasingly important to determine the differences between humans and rats. However, it is also important to determine the differences between rats and mice. In Chapter 2, we used modified mouse differentiation protocols to differentiate ESCs into NPCs or DE in rat. Since differentiation protocols have already been validated in mice, it is important to also prove similar protocols in rat. By building GRN models in all three species we can determine the similarities and differences between the species. By comparing GRNs in rat to mouse and human we can further validate rat as a good model organism to be used in conjunction with mice to use as human analogs for research.

5.2 Improvements to comparisons between early life adversity and control conditions using single-cell transcriptomics

In Chapter 3, we set out to determine the effect of early life adversity (ELA) relative to a well nurtured control (CTL) on the transcriptome of Crh positive neurons in the paraventricular nucleus (PVN). We used the LBN model in Crh-IRES-Cre; Ai14 tdTomato mice to visualize and fluorescence-activated cell sort (FACS) Crh positive neurons from the PVN. We found that CRH expressing PVN neurons from the ELA condition have altered gene expression. The genes with altered expression are associated with cellular processes such as ER stress, altered energy production, and protection from injury. This may be an indication of abnormal increased number of excitatory synapses in these cells after ELA. Although the cells were FACS sorted for Crh expression, we found cells with little to no Crh expression that also had markers for other non-

neuronal cell types. These cells would be interesting to investigate in the future to determine their role in the PVN.

We decided to use SmartSeq2 as the method of scRNA-seq to achieve deep full-length sequencing of transcripts each cell. There are numerous benefits and drawbacks to using this method. Some benefits include recovery of full-length cDNAs. Cells are also sequenced deeper since each cell is built manually or with an automated liquid handling platform, and multiplexing is relatively limited as SmartSeq2 uses a dual-indexing method with only 96 combinations of barcodes [10,11]. Deeper sequencing allows for TF and low abundance transcript expression to be assessed with greater accuracy, but the drawback is the manual labor needed to build RNA from each cell into cDNA libraries individually and a relatively low multiplexing ability does not allow screening of large numbers of cells at a time [10,11]. This method makes it easy to pair with cell collection methods that are more difficult, especially in the case of sequencing primary single cells, including neurons.

However, since this method has the major drawback of increased manual labor and low multiplexing it is much harder to get a lot of cells using this method. To combat this we could FACS the cells from both ELA and control conditions for Crh expression and then use one of the scRNA-seq methods that allow for higher throughput, such as SPLiT-seq [12], 10x, or drop-seq [13]. By having cells from multiple methods, we can take advantage of the deeper sequencing we achieved using SmartSeq2, and the higher throughput of combinatorial barcoding methods. We would be able to determine how sub-populations of Crh positive PVN neurons differ between ELA and CTL in a larger population of cells. Having more cells from each condition would hopefully allow us to evaluate a larger number of differences between these two conditions.

Models of ELA induce enduring long-term behavioral consequences [14]. Some of these consequences include diminished reward seeking [15–18], hippocampal memory deficits [19–23], and

augmented fear inhibition pathways [24] in adult rats and mice. One of the cellular functions that has been shown to be impaired as a result of ELA is the increase in excitatory drive on Crh positive PVN neurons [25]. It is still unclear how this process are encoded in these Crh positive neurons and if these transcriptional changes have a functional output. However, past literature has found that Crh expression has a significant role in the dysfunction of this neural circuit. Learning and memory was shown to be impaired by adding Crh directly into immature rat brains, and when the Crh pathway was inhibited these learning deficits were prevented [26]. To investigate the transcriptional change and how these transcriptional changes affect cellular activity and behavior we can use an unlearned advancing threat experiment to test if ELA affects escape behavior and anticipation of an advancing threat in adult mice [27]. We hypothesize that ELA will affect the response to the unlearned advancing threat and anticipation of a threat. If this turns out to be the case we can test if optogenetic silencing of Crh positive PVN neurons, rescues the phenotype.

5.3 Expanding on comparisons between a spectrum of early life adversity using methylomics in humans

In Chapter 4, we used buccal swab collections soon after birth and after one year of life in a spectrum of ELA experiences in 128 human individuals. We determined intra-individual differences instead of focusing on inter-individual differences that distinguish age and developmental differences between the samples [28]. By using the intra-individual differences, which investigate the differences between paired samples from the same individual before and after experiences they endured during their first year of life, we were able to distinguish between a spectrum of ELA experiences. Because of the complexity of inputs determining ELA in humans (such as living conditions, siblings, food intake, and parent involvement, etc.) a scoring system was used to evaluate each environmental factor. We were able to distinguish

differences between individuals (intra-individual analysis) over a spectrum using principal component analysis (PCA).

PCA is a dimensionality-reduction method used with large data sets. PCA tries to capture most of the information in the data set by transforming a large set of variables into smaller ones. Correlations found using PCA can be used to develop the formula to generate ELA scores from the multitude of quality-of-life events and experiences. Unfortunately, in the quest for simplicity, reducing the data set comes at the expense of accuracy, especially when using large data sets with many input variables. With a data set as complex as the one in Chapter 4, applying machine learning algorithms, such as support vector networks and logistic regression, may be a better way to determine the differences in the data set and ELA scores. Methods such as support vector machines and logistic regression may deliver more accurate results and will be able to predict possible methylation markers with more accuracy than using a simple linear dimensional reduction method such as PCA. It is important to be able to determine possible methylation changes between a diverse set of ELA experiences that could influence downstream gene expression and affect brain development in humans, as changes in methylation of DNA regions may be able to be used to predict if ELA will have lasting effects on an individual. Models can be refined using principal component regression (PCR), where results are fed back into the model training set.

Once a set of possible predictive methylation changes in specific DNA regions have been identified, it would be important to understand how these changes affect gene expression of target genes. For any future collections, we could collect samples for RNA sequencing and DNA methylation at the same time to examine how DNA methylation regulates gene expression. We could collect both RNA sequencing samples and DNA methylation samples from a buccal swab. This method may prove slightly more challenging, as RNA is unstable and should be

further processed to preserve RNA integrity. However, RNA can be preserved using methods such as fixation of cells, or cells frozen on dry ice. Using the buccal swab method to collect peripheral RNA may also not be as informative as the collection site is removed from the brain or any other tissue that would be affected by ELA. To combat this, it may be interesting to investigate single cell methylation and transcriptome changes that occur in the brains of a rodent model of ELA, using similar collection methods as Chapter 3, but collecting DNA methylation and RNA-sequencing data from the same pool of cells. We can then compare the DNA methylation changes we see in humans to changes we see in rodent brains, specifically the PVN Crh expressing neurons.

By following the individuals throughout multiple stages of their life, data from later stages of life can eventually be added to the models. There is hope that the research will lead to prediction and, ultimately, treatment of negative long term adverse effects that result from ELA.

In conclusion, epigenetic changes are often seen at the level of individual genes, but conservation seems to be at a higher level. For example, networks of genes and neural circuits respond to differentiation and/or environmental cues.

5.4 References

1. Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).

2. Levine, M. & Davidson, E. H. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4936–4942 (2005).

3. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biology* vol. 21 (2020).

4. Jansen, C. *et al.* Building gene regulatory networks from scATAC-seq and scRNA-seq using Linked Self Organizing Maps. *PLoS Comput. Biol.* **15**, e1006555 (2019).

5. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361–376.e17 (2019).

6. Ellenbroek, B. & Youn, J. Rodent models in neuroscience research: is it a rat race? *Dis. Model. Mech.* **9**, 1079–1087 (2016).

7. Suckow, M. A., Claire Hankenson, F., Wilson, R. P. & Foley, P. L. *The Laboratory Rat*. (Academic Press, 2019).

8. Homberg, J. R., Wöhr, M. & Alenina, N. Comeback of the Rat in Biomedical Research. *ACS Chem. Neurosci.* **8**, 900–903 (2017).

9. Brenowitz, E. A. & Zakon, H. H. Emerging from the bottleneck: benefits of the comparative approach to modern neuroscience. *Trends Neurosci.* **38**, 273–278 (2015).

10. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

11. Wang, X., He, Y., Zhang, Q., Ren, X. & Zhang, Z. Direct Comparative Analysis of 10X Genomics Chromium and Smart-seq2. *Cold Spring Harbor Laboratory* 615013 (2019) doi:10.1101/615013.

12. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).

13. Zhang, X. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol. Cell* **73**, 130–142.e5 (2019).

14. Walker, C.-D. *et al.* Chronic early life stress induced by limited bedding and nesting (LBN) material in rodents: critical considerations of methodology, outcomes and translational potential. *Stress* **20**, 421–448 (2017).

15. Bolton, J. L. *et al.* Anhedonia Following Early-Life Adversity Involves Aberrant Interaction of Reward and Anxiety Circuits and Is Reversed by Partial Silencing of Amygdala Corticotropin-Releasing Hormone Gene. *Biol. Psychiatry* **83**, 137–147 (2018).

16. Levis, S. C. *et al.* On the early life origins of vulnerability to opioid addiction. *Mol. Psychiatry* (2019) doi:10.1038/s41380-019-0628-5.

17. Bolton, J. L. *et al.* Early-life adversity facilitates acquisition of cocaine self-administration and induces persistent anhedonia. *Neurobiol Stress* **8**, 57–67 (2018).

18. Molet, J. *et al.* Fragmentation and high entropy of neonatal experience predict adolescent emotional outcome. *Transl. Psychiatry* **6**, e702 (2016).

19. Naninck, E. F. G. *et al.* Chronic early life stress alters developmental and adult neurogenesis and impairs cognitive function in mice. *Hippocampus* **25**, 309–328 (2015).

20. Bath, K. G. *et al.* Early life stress leads to developmental and sex selective effects on performance in a novel object placement task. *Neurobiol Stress* **7**, 57–67 (2017).

21. Brunson, K. L. *et al.* Mechanisms of late-onset cognitive decline after early-life stress. *J. Neurosci.* **25**, 9328–9338 (2005).

22. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* **149**, 4892–4900 (2008).

23. Molet, J. *et al.* MRI uncovers disrupted hippocampal microstructure that underlies memory impairments after early-life adversity. *Hippocampus* **26**, 1618–1632 (2016).

24. Bath, K., Manzano-Nieves, G. & Goodwill, H. Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Horm. Behav.* **82**, 64–71 (2016).

25. Gunn, B. G. *et al.* Dysfunctional astrocytic and synaptic regulation of hypothalamic glutamatergic transmission in a mouse model of early-life adversity: relevance to neurosteroids and programming of the stress response. *J. Neurosci.* **33**, 19534–19554 (2013).

26. Short, A. K., Maras, P. M., Pham, A. L., Ivy, A. S. & Baram, T. Z. Blocking CRH receptors in adults mitigates age-related memory impairments provoked by early-life adversity. *Neuropsychopharmacology* **45**, 515–523 (2020).

27. Daviu, N. *et al.* Paraventricular nucleus CRH neurons encode stress controllability and regulate defensive behavior selection. *Nat. Neurosci.* **23**, 398–410 (2020).

28. Jiang, S. *et al.* Intra-individual methylomics detects the impact of early-life adversity. *Life Sci Alliance* **2**, (2019).