

Identification of Novel Factors Involved in the Development of the Nervous System

By

Adib Moghimy

Supervisor: Dr. Kambiz N. Alavian

This thesis is submitted to the Department of Biology as a partial fulfillment of the BSc program
in Biology

October 2019

Department of Biology

Faculty of Science

Baha'i Institute for Higher Education (BIHE)

© Adib Moghimy, Tehran, Iran, 2019

Summary

Neural development begins by the formation of neural ectoderm. Four genes, HESX1, OTX2, GBX2, and SIX3, play key roles in this formation. The neural development process continues by the differentiation of neural ectoderm to the neural plate, leading to the development of neural tube and neural crest cells, which give rise to the central and peripheral nervous system, respectively. These processes are mainly regulated by a cascade of developmental transcription factors. Phylogenetic profiling is a method that detects the evolutionary patterns of genes. The main premise of this method is that the genes that have evolved together are likely to share common functional or structural features. In this work, I hypothesize that the four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) are evolutionarily correlated with each other, and have similar or correlated evolutionary patterns to that of the genes involved in neural development. In this work, I first determined the correlation amongst the phylogenetic profiles of the four neural ectoderm factors and identified, in a genome-wide study, the genes that are highly correlated with these factors in order. The study led to identification of 35 genes that are correlated to four neural ectoderm factors, suggesting novel interactions or molecular players in the development of the nervous system. We then examined the spatial and temporal expression patterns of the neural ectoderm factors and their putative interacting partners in the developing mouse brain, using Allen Brain Atlas. Our result show that from these 35 genes, POU4F1, PAX7, DMBX1, PITX2, OTP, PHOX2B and ARX have the most similar expression patterns to those of the four neural ectoderm genes. These results suggest novel players in the pathways that are responsible for the development of the

three primary vesicles of the brain. The results also might lead to the discovery of non-canonical functions for the four neural ectoderm genes.

Table of Contents

Summary.....	ii
Table of Contents	iv
List of Figures	vi
List of Tables	vi
List of Abbreviations.....	vii
Acknowledgement.....	xi
Chapter 1: Introduction	1
Stem Cells and Pluripotency	1
Pluripotent Genes of ESCs.....	3
Formation of the Neural Tube by Primary Neurulation.....	4
Development of the Central Nervous System (Brain)	7
Development of the Peripheral Nervous System	8
Neural Ectoderm Genes.....	8
Gbx2.....	9
Hesx1	9
OTX2	10
Six3	11
Sox1	11
Genes of Neural Development.....	12
ISL2	12
ARX	12
CRX.....	13
EMX1.....	14
GSX2.....	14
OTP	14
PAX7.....	15
PHOX2 (PHOX2A & PHOX2B).....	15
POU4F1	15
RAX	16
VAX1	16
PITX (PITX2, PITX3).....	16

Phylogenetic Profiling	17
Overall Hypotheses and Aims	18
Chapter 2: Materials & Methods	19
Selection of Pluripotent Genes and Neural Ectoderm Genes	19
Constructions of Phylogenetic Profiles	19
Allen Brain Atlas	20
Chapter 3: Results	21
Pluripotency and Neural Ectoderm Genes.....	21
Correlated Genes with HESX1, OTX2, GBX2, and SIX3.....	22
Correlations between the 35 Genes that Correlated with HESX1, OTX2, GBX2 and SIX3	23
Correlation Matrix of 35 Random Genes and Four Neural Ectoderm Genes.....	25
Expression of Genes during Developing Mouse Brain.....	27
Similar Expression between Genes	34
Chapter 4: Discussion	37
References:	43

List of Figures

Figure 1	2
Figure 2	3
Figure 3.....	6
Figure 4.....	8
Figure 5.....	22
Figure 6.....	24
Figure 7.....	26
Figure 8.....	27
Figure 9.....	28
Figure 10.....	29
Figure 11.....	30
Figure 12.....	32
Figure 13.....	34

List of Tables

Table 1	35
Table 2	35
Table 3	36
Table 4	36

List of Abbreviations

A

ARX: Aristaless-related homeobox

ALX1: Aristaless-like homeobox 1

ALX4: Aristaless-like homeobox 4

B

BCAP31: B cell receptor associated protein 31

C

CRX: Cone-Rod Homeobox

D

DPRX: Divergent-paired related homeobox

DMBX1: Diencephalon/mesencephalon homeobox 1

DUXA: Double homeobox A

DUX1: Double homeobox 1

DUX5: Double homeobox 5

E

EMX1: Empty spiracles homeobox 1

F

G

GBX2: Gastrulation brain homeobox 2

GSX2: The genetic screened homeobox 2

GID4: Glucose-induced degradation protein 4 homolog

H

Hesx1: HESX Homeobox 1

HOXC11: Homeobox C11

HOXC13: Homeobox C13

HOXD13: Homeobox D13

HHEX: Hematopoietically expressed homeobox

I

ISL2: Insulin related protein 2

ISX: Intestine specific homeobox

J

K

KLF4: Krüppel-like factor 4

L

M

MPPE1: Metallophosphoesterase 1

N

NANOG: Nanog Homeobox

O

OCT4: Octamer-binding transcription factor 4

OTX2: Orthodenticle homeobox 2

OTP: Orthopedia

P

PAX7: Paired box 7

PHOX2A: Paired like homeobox 2A

PHOX2B: Paired like homeobox 2B

POU4F1: POU class 4 homeobox 1 (POU4F1)

PITX1: Paired-like homeodomain 1

PITX2: Paired-like homeodomain 2

PITX3: Paired-like homeodomain 3

PROP1: Paired like homeodomain factor 1

Q

R

RAX: Retina and anterior neural fold homeobox

RAX2: Retina and anterior neural fold homeobox 2

Rex1 (Zfp42): Zinc finger protein 42

S

SII4: Sal-like protein 4

SIX3: Sine oculis-related homeobox 3

SOX1: a member of the SOXB1 subgroup of the high mobility box (HGM-box) family

SLC15A5: Solute carrier family 15 member 5

SPTB: Spectrin beta, erythrocytic

SHOX: Short stature homeobox

T

U

V

VAX1: Ventral anterior homeobox 1

W

Z

Acknowledgement

Special thanks to Dr. Alavian who suggested the topic and supported me during this thesis.

Chapter 1: Introduction

Stem Cells and Pluripotency

In mammals after the fertilization, the zygote undergoes cell division, cleavage, and differentiate to form morula and then blastocyst. The early blastocyst has two kinds of cells, the trophectoderm (TE) that surrounds a fluid-filled cavity and inner cell mass (ICM). Later, the TE will implant to the uterine wall and develop to some parts of the placenta. In the late blastocyst, ICM divided into two groups of cells: The epiblast that gives rise to the embryo and the primitive endoderm (figure 1). The embryonic stem cells (ESCs) are derived from the ICM (Rossant and Tam, 2009) (Bell and Watson, 2013). ESCs are pluripotent means that they can transform into all three primary layers of the embryo (endoderm, mesoderm, and ectoderm) and produce all the cell types of the body (figure 2). Also, they can proliferate indefinitely without differentiation (self-renewal) (Mullen and Wrana, 2017). Many transcription factors and genes are responsible for maintaining the pluripotency and self-renewal of stem cells. In the next section, we will examine briefly five of these transcription factors including Oct4, Nanog, Klf4, Sox2, and Rex1.

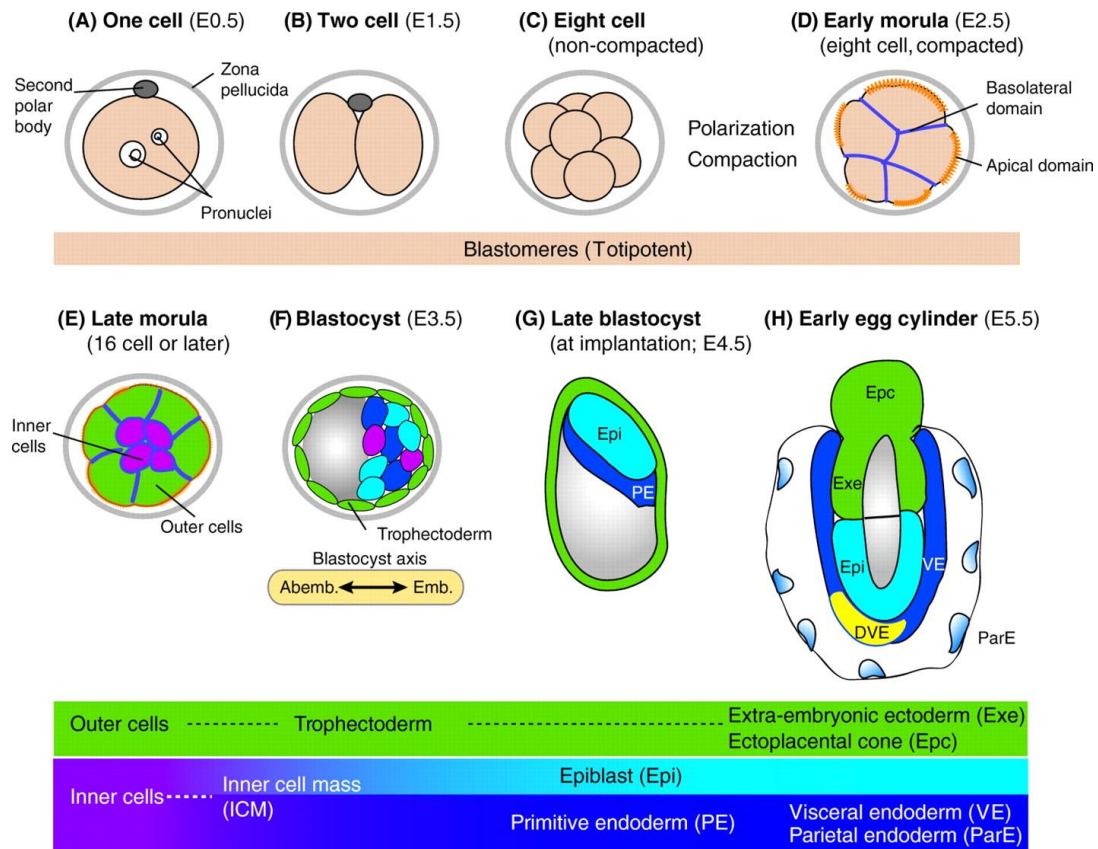


Figure 1 (Rossant and Tam, 2009). (A-H) Schematics of different stages of developing mouse embryo from fertilization (E0.5) to early egg cylinder (E5.5). The colored bars illustrate the progressive allocation of totipotent blastomeres to outer and inner cells and the trophectoderm and inner cell mass lineages. The different cell types of the embryo are color-coded.

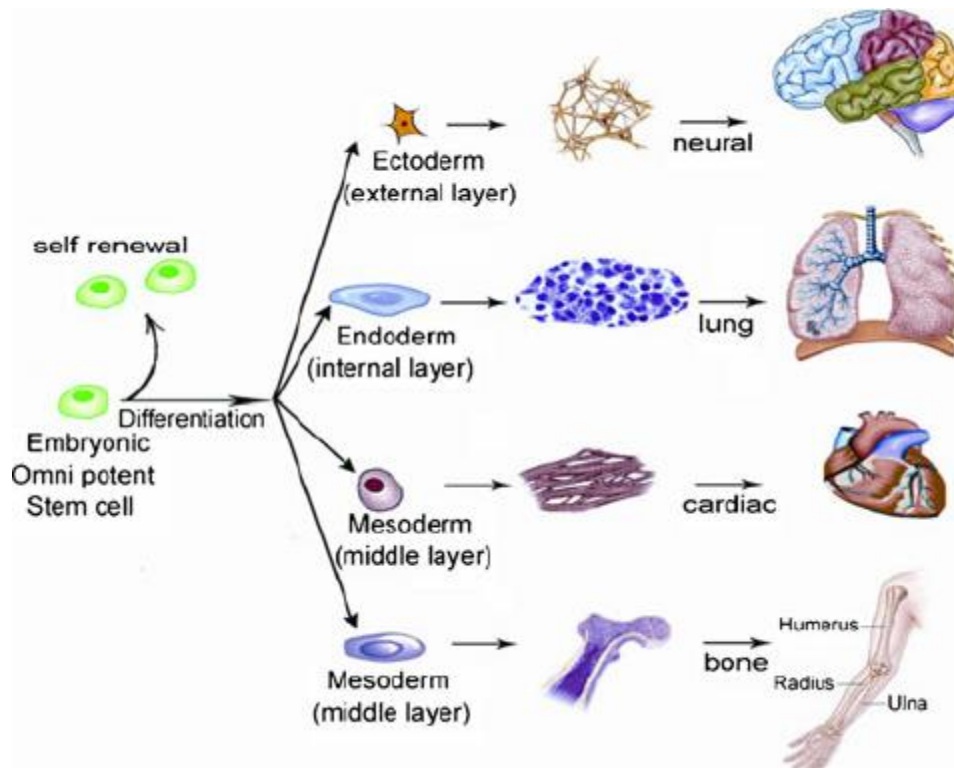


Figure 2 (Vira *et al.*, 2012). Pluripotency of ESCs: ESCs can differentiate to all cell types of the body. In the early stages of developing embryo, ESCs differentiate into three primary layers of the embryo (Ectoderm or external layer, endoderm or internal layer and mesoderm or middle layer). All the different tissues and cells come from these three layers.

Pluripotent Genes of ESCs

NANOG is a transcription factor that plays a key role in the self-renewal and pluripotency of ESCs. This transcription factor suppresses the activators and activates the repressors of differentiation. The NANOG gene located on chromosome 12 and its protein has 305 amino acids and 3 functional domains. These domains are the N-terminal domain with 94 amino acid; the homeodomain with 60 amino acid; and the C-terminal domain with 151 amino acids. Its mRNA can only be detected in the epiblast (Gawlik-Rzemieniewska and Bednarek, 2016). The octamer-

binding transcription factor 4 (OCT4) is a transcription factor that belongs to the POU family. This transcription factor maintains ESCs on the pluripotent and self-renewal stage. This transcription factor can bind to DNA with other transcription factors such as Sox2. This binding promotes cell reprogramming of ESCs (Yesudhas *et al.*, 2016). The Krüppel-like factor 4 (KLF4) is a transcription factor that plays key roles in the regulation of proliferation, differentiation, and pluripotency of ESCs. In the ESCs, KLF4 activates telomerase reverse transcriptase (TERT) expression. So, it maintains the ESCs in the self-renewal and pluripotent stages (Hsieh *et al.*, 2017). Rex1 (Zfp42) protein is a member of the zinc finger protein family. Both Oct4 and Nanog target Rex1. These two genes (Oct4 and Nanog) bind to their binding sites in the promoter of Rex1 and act as positive regulators of Rex1 gene. In the vitro, the expression of Rex1 is used as a marker of undifferentiated ESCs (Lee, Lu and Gudas, 2010). Sll4 is a member of the spalt-like (SALL) gene family. This gene has a function in maintaining ESCs in the pluripotent and self-renewal stage. In late mice blastocysts, the expression of Sall4 is observed in ICM and trophectoderm (Tatetsu *et al.*, 2016).

Formation of the Neural Tube by Primary Neurulation

The epiblast is transformed into three primary layers of the developing embryo. These three layers from inner to outer included endoderm, mesoderm and ectoderm (Figure 2). The neural tube originates from ectoderm through two major ways: primary and secondary neurulation. The ectoderm is divided into three groups of cells during primary neurulation. The first group is the neural tube that is internally positioned. Later, the brain and spinal cord originate from the neural tube. The second group includes the epidermis of the skin that is

externally positioned. The third group is neural crest cells that are positioned between the other two groups. The neural crest cells connect the neural tube and epidermis. Later, they will migrate to elsewhere. The peripheral neurons, glia, and other cells types originate from neural crest cells (Gilbert, 2000b).

The primary neurulation is a complex process and begins in the 3rd week after fertilization. First, the neural plate is derived from ectoderm by overlying ectoderm cells. The neural plate is in the central part of the ectoderm and other parts of ectoderm will be epidermis and neural crest cells. In this stage, there is no neural development. In other words, the default state of the ectoderm is neural (not epidermal). In the central part of ectoderm, the epidermis formation is inhibited and the neural plate is derived from this part (Sadler, 2005).

The cells of the neural plate are elongated in contrast to epidermis cells. The formation of hinge regions bends the neural plate. The medial hinge point (MHP) cells are located at the midline of the neural plate. The notochord is located beneath the MHP cells and these cells become anchored to it. This anchor forms a hinge. Then, the notochord causes the MHP cells to decrease their height. After that, first, the dorsolateral hinge points (DLHPs) regions and second the paired neural folds are formed. The DLHPs are two hinge regions that connect the neural plate to other reminders of the ectoderm. At the dorsal midline, the cells of two folds merge and the neural tube closes at the dorsal midline. The cells at the junction of these two folds generate the neural crest cells (figure 3). Many factors such as genetic and environmental factors have the function in human neural tube closure (Gilbert, 2000b).

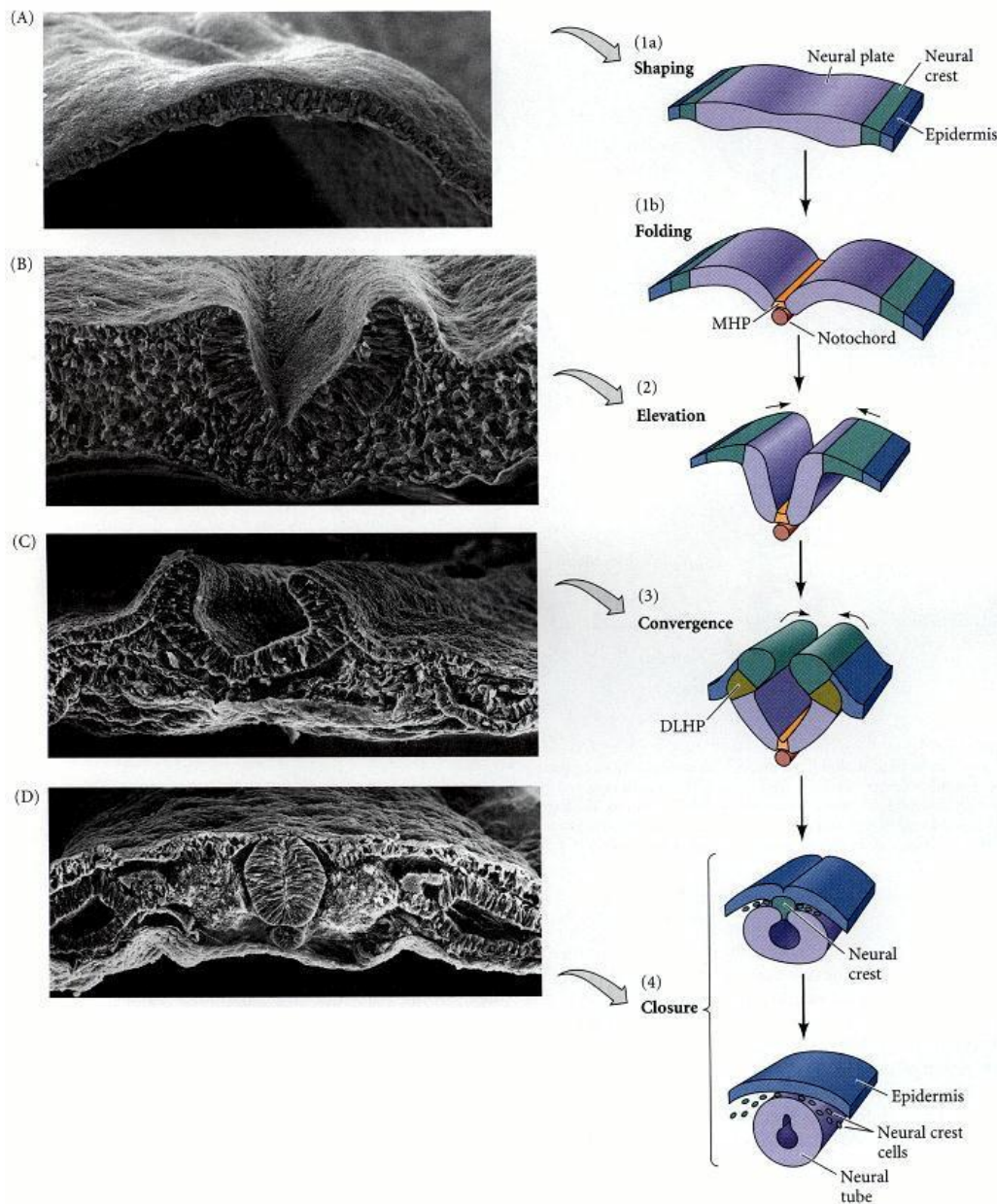


Figure 3 (Gilbert, 2000b). Different stages of neural tube formation in the chick embryo. (A, 1) Progenitor cells of the neural tube are elongated in the dorsal region. The medial neural hinge point (MHP) cells change their shape by anchoring to the notochord, the epidermal cells move towards the center and folding begins. (B, 1) The

epidermis continues moving toward the dorsal midline and neural folds are elevated. (C, 3) The dorsolateral hinge point (DLHP) cells are wedge-shaped, epidermis move toward center and convergence of neural folds occurs. (D, 4) The formation of the neural tube completes. The neural crest cells separate from the epidermis and disperse.

Development of the Central Nervous System (Brain)

After the completion of the neural tube, the neural progenitor cells located in the center of the neural tube and adjacent to its hollow center. This region is called “ventricular zone” (VZ). The most neural progenitor cells of the rostral region will become the brain (Stiles and Jernigan, 2010). First, they become three primary vesicles. These vesicles include forebrain (prosencephalon), midbrain (mesencephalon), and hindbrain (rhombencephalon). Then, these three vesicles become into five secondary vesicles (Gilbert, 2000a).

The most anterior of these three vesicles is prosencephalon. Later this vesicle will subdivide into two regions: anterior (telencephalon) and caudal (diencephalon) regions. Also, the ventral diencephalon has a function in the formation of the pituitary gland. Telencephalon will form the cerebral hemispheres, the cerebral cortex, hippocampus, basal ganglia, basal forebrain nuclei, and olfactory bulb. The diencephalon will form the thalamus, hypothalamus, and retina. In contrast to prosencephalon, the mesencephalon does not become divided and its lumen will become the cerebral aqueduct. The rhombencephalon will divide into two regions: the posterior (myelencephalon) and the anterior (metencephalon). The metencephalon eventually gives rise to the cerebellum and pons and myelencephalon gives rise to medulla (Gilbert, 2000a), (Purves D, Augustine GJ, Fitzpatrick D, 2001), (Larkin and Ansorge, 2000).

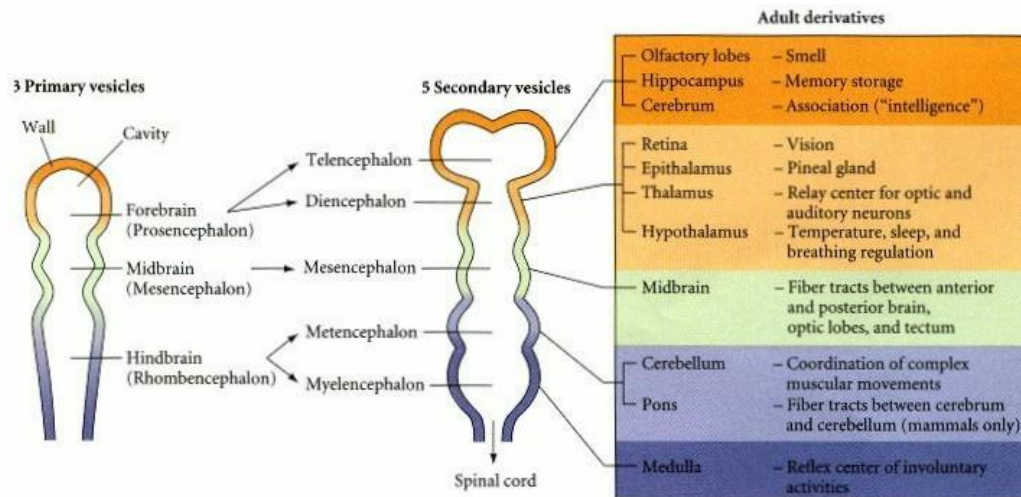


Figure 4 (Gilbert, 2000a). Primary and secondary vesicles of human brain development. The three primary brain vesicles are subdivided to 5 secondary vesicles during development. At the right is a list of the names and functions of adult derivatives of the 5 secondary vesicles.

Development of the Peripheral Nervous System

The peripheral nervous system develops from neural crest cells. First, the neural crest cells are located in the roof of the neural tube. Then, they migrate into the surrounding tissues of the neuroepithelium. They migrate in three different ways. The two ways of these three generate dorsal root ganglia and autonomic ganglia (Catala and Kubis, 2013).

Neural Ectoderm Genes

Here, the five genes of neural ectoderm will be explained. These genes included Gbx2, Hesx1, Sox1, Otx2, and Six3.

Gbx2

Gastrulation brain homeobox 2 (Gbx2), a member of the homeobox family gene, existed in mammalian, avian, amphibian and fish species. *Drosophila unplugged* gene (involved in tracheal branching) is related to Gbx genes. The Gbx2 is expressed in midbrain-hindbrain boundary (local signaling center which is positioned between the presumptive midbrain and hindbrain) and plays a key role in the early development of the brains of vertebrates. Mutation of Gbx2 in the mice causes a lack of the anterior hindbrain. The Groucho interaction partners Engrailed, Pax2/5/8, and Tcf/Lef have the most important role in the further development of the midbrain-hindbrain boundary (MHB). Groucho genes play important roles in many developmental processes such as neurogenesis. Gbx2 has engrailed homology region 1 (eh1)-like motifs that are highly conserved through evolution. Groucho proteins have WD40 domains (located at C termini) that can interact with eh-1 motifs (Heimbucher *et al.*, 2007).

Hesx1

Hesx1 is a homeobox transcription factor that is identified in embryonic stem cells. This gene can be identified from 7.5 to 8.5 days post coitum (d.p.c) in the anterior forebrain. This gene is a member of the Paired (prd) class that encodes a transcription repressor for cell differentiation and proliferation and organ commitment. The human gene of Hsex1 has four exons that encode a protein with 185 amino acid. There are two highly conserved domains in this protein. The first domain is engrailed homology (eh1) domain that is in the N-terminal and has 7 amino acid. This eh1 domain binds to the transducing-like enhancer of split-1 (TLE1) co-repressor. The second domain is a C-terminal homeodomain with transcriptional repressor activity. The Hesx1 can be

detected from 8.5 to 13.5 days d.p.c in the Rathke's pouch, an invagination of oral ectoderm that becomes anterior pituitary primordium during development. Prop1 is one of the downstream targets of Hesx1. In the cell differentiation of pituitary, when Hesx1 expression disappears the expression of Prop1 rises. Also, Hesx1 plays important roles in the development of the hypothalamus, optic nerve and posterior pituitary (Fang *et al.*, 2016), (Newbern *et al.*, 2013).

OTX2

Orthodenticle homeobox 2 (OTX2), a transcription factor, plays important roles in many neural developments such as the brain, cerebellum, pineal gland and eye development. Past studies showed that complete deletion of mouse Otx2 (the mouse homolog of human OTX2) causes deletion of both forebrain and midbrain. This gene positioned on chromosome 14q23.1 and has five exons, but only three of them will be translated. OTX has a highly conserved homeodomain. In the early development of the embryo and during gastrulation, Otx2 specifies the anterior neuroectoderm. Also, when the formation of the eye begins, the Otx2 expression decreases. Also, Otx2 is required for the development of the presumptive retinal pigment epithelium (RPE) and neural retina. Past studies show that Otx2 protein exists in bipolar and ganglion cells. It can regulate the expression of Crx gene in the outer nuclear layer. Crx regulates the expression of photoreceptor function genes such as opsins. Both Otx2 and Crx bind to the same DNA-binding sequences. Further researches show that OTX2 is also playing key roles in the development of retinal bipolar cells because deletion or ablation of OTX2 in retinal progenitor cells causes a decrease in rod photoreceptor cells and bipolar cells (Kaur *et al.*, 2015), (Boulay *et al.*, 2017), (Abdalla-Elsayed *et al.*, 2017), (Henderson *et al.*, 2009).

Six3

SIX homeobox 3 (SIX 3) is a transcription factor with a six-domain and a homeobox domain for DNA binding. This gene is expressed in the central nervous system (CNS) and is necessary for the development of CNS. So, its mutation causes many neural developmental disorders such as holoprosencephaly, aprosencephaly, and atelencephaly (Yu *et al.*, 2017). There are two DNA bindings regions in the amino acid sequence of the SIX3: the SIX domain and the homeodomain. SIX3 has sequence homology with the mouse Six3 and Optix, a drosophila gene that expresses in eye primordial. In mouse, Six3 is expressed in the hypothalamus, pituitary, retina, and lens of developing eyes (Pasquier *et al.*, 2005).

The heterozygous mutation of SIX3 causes holoprosencephaly (HPE), the most common developmental anomaly of the human forebrain. There is an interaction between Six3 and members of the Groucho co-receptor that this interaction is important for the development of vertebrate forebrain. Wnt1 gene is one of the genes that is a direct target of Six3. Wnts cause maintenance and the specification of anterior forebrain. Also, Six3 can interact with Geminin. Geminin regulates the proliferation and differentiation of neuroepithelial cells (Domené *et al.*, 2008).

Sox1

SOX1, a transcription factor, is a member of the SOXB1 subgroup of the high mobility box (HGM-box) family. Transcription factors of this subgroup expressed in the early development of the embryo and have a key role in the development of CNS. Also, they have functions in the

maintenance of neural stem cells. SOX2 is another member of SOXB1 subgroup that has similarities in sequence, expression pattern and overexpression phenotype with SOX1. Both SOX1 and SOX2 have a common ancestor, and they were resulted by duplication during evolution (Ahmad *et al.*, 2017). Previous studies showed that Nestin is at the downstream of Sox1. Sox1 can bind to Nestin and active the expression of this gene. The expression of both Sox1 and Nestin is necessary for the development of CNS (Feng *et al.*, 2014). Also, Sox1 plays a role in the development of the lens of mice eyes by regulating the γ -crystallin genes (Nishiguchi *et al.*, 1998).

Genes of Neural Development

In the below, the genes that play key roles in the neural development are explained.

ISL2

Isl2 is a LIM homeodomain with a transcriptional activity that has functions in the development and differentiation of visceral motor neurons of the spinal cord. It also expressed in post-mitotic cells of both inner and outer retina. By continuing development, Isl2 expressed only in the ganglion cell layer (GCL). It expressed approximately 40% of all retinal ganglion cells (RGCs) (Triplett *et al.*, 2014).

ARX

An aristaless-related homeobox (ARX) is a homeodomain transcription factor that believes to play important roles in the development of interneurons. Many disorders of the nervous system such as lissencephaly, movement disorders, intellectual disabilities and epilepsy caused by a mutation in ARX. The research on *Arx* of *Mus musculus* (house mouse) shows that

ARX has an important function to migrate interneurons in the dorsal direction into the cortex, but the role of ARX in the specification of interneurons is very restricted. Arx also can repressor the expression of many genes such as *Shox2*, *Ebf3*, and *Lmol* by binding to their 5' promoter regions. The N-terminal octapeptide domain of Arx is responsible for repressor activity and is similar to the Engrailed homology repressor domain (eh-1), and recruit Groucho/transducing-like enhancer split co-factor proteins (TLE1-4) (Marsh *et al.*, 2016), (Sunnen *et al.*, 2014), (Nasrallah *et al.*, 2012).

CRX

Cone-Rod Homeobox (CRX) is a transcription factor that plays an important role in the development of retina, Genesis, and maturation of photoreceptors. This gene is expressed in the precursors of post-mitotic retinal photoreceptors. Animal studies show that Crx has an important role in the genesis of cones (cone cells) and rod cells (Collin *et al.*, 2016). CRX has three exons that encode 299 amino acid polypeptide that is very similar to sequences of OTX family proteins. So, it belongs to OTX family genes. Similar to OTX family genes, CRX has two highly conserved sequences: a WSP motif (13 amino acid) and OTX tail (12 amino acid). During the development of the retina, CRX activates many genes of photoreceptors such as opsins, arrestin, and β -phosphodiesterase. In the section of OTX2, I mentioned the relation between OTX2 and CRX during retinal development. Studies show that CRX is expressed in the pineal gland and plays as a transcription for pineal genes (Sohocki *et al.*, 1998).

EMX1

EMX1 is a transcription factor with homeodomain that is expressed in the cortical neural progenitors and developing forebrain. This gene is expressed in the most neurons of all cortical layers during different stages of proliferation, migration, differentiation, and maturation. It also is expressed in the olfactory bulb during its development, and after birth, it is present in some mitral cells of the olfactory system (Briata *et al.*, 1996).

GSX2

The genetic screened homeobox 2 (Gsx2) is a transcription factor that has important roles in the development of many nervous system parts. For example, this transcription factor is necessary for the regional specification of the telencephalon and the development of the olfactory bulb. In the ventral ventricular zone of the telencephalon, Gsx2 is expressed and makes early striatal progenitors. So, it plays an important role in the development of striatal neurons. In the early development of telencephalon, the expression of the Gsx2 and Pax6 (Paired Type Homeobox 6) identifies the pallium and subpallium (Faedo *et al.*, 2017).

OTP

Orthopedia (OTP) is a homeobox gene that is involved in the development of the hypothalamus. This gene is expressed in hypothalamic neural precursors that develop to paraventricular (PVN), supraoptic (SON) and anterior periventricular (aPV) nuclei. Also, its expression is continued in the mature hypothalamic neurons of mouse and zebrafish (Amir-Zilberstein *et al.*, 2012), (Acampora *et al.*, 1999).

PAX7

Paired box 7 (PAX7) is a transcription factor of Pax family that plays important roles in the developing of the nervous system. Pax7 is closely related to Pax3, and both of them have similar targets and are expressed in the dorsal neural tube. In the development of chicken spinal cord, overexpression of Pax7 increases the expression of Pax3, but overexpression of Pax3 inhibits the expression of Pax7. Previous studies show that Pax7 has a function in morphogenesis and cell differentiation of developing spinal cord (Luan *et al.*, 2013), (Lin *et al.*, 2016).

PHOX2 (PHOX2A & PHOX2B)

PHOX2A and PHOX2B are homeodomain transcription factors that have an important role in the development of the nervous system. These two transcription factors are paralogue to each other and they regulate the expression of tyrosine hydroxylase and dopamine- β -hydroxylase (enzymes of noradrenaline biosynthesis). The expression of these two enzymes causes specification of the neurotransmitter phenotype by directing the precursor's neurons to their terminal noradrenergic differentiation. During sympathetic neural differentiation, PHOX2A and PHOX2B exit neural progenitors from the cell cycle and begin their differentiation (Di Lascio *et al.*, 2016).

POU4F1

POU class 4 homeobox 1 (POU4F1) is a homeodomain transcription factor that is responsible for the development of the peripheral nervous system. This gene is expressed in neural crest and differentiates them into precursor cells of the peripheral nervous system.

POU4F1 also is expressed in other parts of the nervous system such as sensory ganglia and some parts of the central nervous system (Hohenauer *et al.*, 2013).

RAX

Retina and anterior neural fold homeobox (RAX) is homeodomain that is responsible for eye development. It is expressed in the precursor cells of the retina and differentiates the retina. In mice, mutation of this gene causes brain malformation. In human, its mutation causes anophthalmia, microphthalmia, and eye colomba. This transcription factor activates transcription of eye-specific genes such as Pax6 (Abouzeid *et al.*, 2012), (London *et al.*, 2009).

VAX1

Ventral anterior homeobox 1 (VAX1) is a transcription factor for the development of basal forebrain and optic nerve. Vax1 is a member of the *Emx* and *Not* gene family. VAX1 (in humans) has two isoforms. The first isoform has 3 exons that encode a polypeptide with 334 amino acid. The second isoform has 4 exons with a 186 amino acid polypeptide. The mutation in VAX1 causes eye diseases in humans (Slavotinek *et al.*, 2012), (Hallonet *et al.*, 1999).

PITX (PITX2, PITX3)

Paired-like homeodomain 2 (PITX2) is a transcription factor responsible for the development of many organs such as the heart, eyes, and pituitary. It has four isoforms including PITX2A, PITX2B, PITX2C, and PITX2D, but the expression patterns of these isoforms have not been

understood yet. In mice, Pitx2 has three isoforms including Pitx2a, Pitx2b, and Pitx2c. These isoforms are expressed in the forebrain, midbrain and spinal cord of CNS (Martin *et al.*, 2004), (Waite and Martin, 2015). Paired-like homeodomain 3 (PITX3) is also a transcription factor and belongs to the bicoid-like subgroup. This transcription factor is responsible for the development of midbrain in human (Sakazume *et al.*, 2007).

Phylogenetic Profiling

Phylogenetic profiling is a method that compares genome sequences between different species or inside of an organism. The basic assumption of this method is that functionally related proteins have evolved from a common ancestor. In other words, analyzing the correlation of evolutionarily patterns of different proteins can predict the interactions, direct and indirect functional links between these proteins. So, this method can be used to identify new protein complexes, expand common signaling pathways and analyze non-coding elements. There are two groups of prediction algorithms in phylogenetic profiling based on the occurrence of information across different species. Methods such as Hamming distance, Pearson correlation coefficient, and Fisher's exact tests are the examples of the first group that phylogenetic profiles between proteins compared by a "co-occurrence" method. Although the co-occurrence-based methods are very time-efficient and used for predicting functional linkages, they aren't used for the effect of phylogenetic bias or the non-independence of the profile values. Methods such as collapsing of the subtree, tree-kernel and maximum likelihood are the examples of the second group that used "model-based" methods. These methods use the phylogenetic tree to correlate the evolutionarily processes and non-independence of the profile values. PrePhyloPro (PPP) is a

method and an online tool that uses multiple co-occurrence measures. Then, it determines potential linkages. PrePhyloPro was constructed with a phylogenetic profile of 972 different species to identify human whole proteome functional lineages (Niu *et al.*, 2017).

Overall Hypotheses and Aims

Phylogenetic profiling is used to detect correlation of evolutionarily patterns. These evolutionarily correlated genes are likely to have functional links (direct or indirect) or structural interactions. Here, we hypothesize that the major genes (HESX1, OTX2, GBX2, and SIX3) involved in the development of neural ectoderm share common evolutionarily patterns, due to their role in a common developmental pathway. The common evolutionarily patterns could be used to identify novel genes involved in this process or in other functions shared by these functionally-related factors. We will utilize phylogenetic profiling and will determine the spatial and temporal expression patterns of these genes in embryonic and postnatal mouse brains to establish functional connections. This study is likely to result in identification of novel pathways involved in the development of different layers of the central nervous system. The results might also suggest non-canonical functions for the four neural ectoderm genes.

Chapter 2: Materials & Methods

Selection of Pluripotent Genes and Neural Ectoderm Genes

The five pluripotent genes included OCT4, NANOG, SALL4, REX1, and KLF4 and five neural ectoderm genes included HESX1, OTX2, GBX2, SIX3, and SOX1 were selected. In the introduction section, their functions were explained in summary.

Constructions of Phylogenetic Profiles

The PrePhyloPro software available at <http://prephylopro.org/phyloprofile/> was used to construct the phylogenetic profiles. For the first step, the 5 genes of pluripotency (OCT4, NANOG, SALL4, REX1, and KLF4) and 5 genes of neural ectoderm (HESX1, OTX2, GBX2, SIX3, and SOX1) were uploaded to the software. For this construction, we set the algorithm parameters of the PPP as default (Linkage number: 20, Blast Evalue: 0.001, Organism: Homo sapiens). The output files are "Correlation Matrix" table and plot, profile plot, and predicted linkages. For this construction, we only used the "Correlation Matrix" plot.

In the second step, only four of neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) were uploaded to PrePhyloPro (Linkage number: 50, Blast Evalue: 0.001, Organism: Homo sapiens). Then, we only used the predicted linkages of the output files. In predicted linkages, we searched for the genes that correlated to all four neural ectoderm genes. But, we only found 35 genes that correlated to 3 of four of them. We used these 35 genes for the next construction of phylogenetic profiles.

In the third step, we uploaded four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) and 35 genes of the previous construction to the software (Linkage number: 20, Blast Evalue: 0.001, Organism: Homo sapiens). Then, we used the "Correlation Matrix" plot of output files. This

is positive control of PrePhyloPro. We chose randomly 35 genes from the whole human genome for a negative group. There are 20127 genes in the hsa_wholeGenomeAnno.csv available at <https://github.com/YulongNiu/PhyloProfileWeb/blob/master/readme.md>. We wrote bellow python code to generate 35 random numbers between 1 and 20127. Then, we uploaded four neural ectoderm genes with these 35 randomly genes to the PrePhyloPro. We used the "Correlation Matrix" plot of output files. This is a negative control of PrePhyloPro.

```
import numpy as np
np.random.seed(42)
a = np.random.randint(1, 20127, 35)
print(a)
```

```
[15796  861  5391 11965 11285  6266 16851  4427 14424 11364 16024  8323
 1686   770  2434  5312  5052  6421 17569 19770  6397  8667 18943 18432
 2748  190 19119  3006  1900  1268 17913 11395  3557  3891  8839]
```

Allen Brain Atlas

The expression of 4 neural ectoderm genes and 35 genes that correlated to them were studied during developing mouse brain. For this purpose, the ISH data (In Situ Hybridization) existed in Allen brain atlas was used. The names of genes were searched at <http://developingmouse.brain-map.org/> (option: Gene Search). The output result was an expression summary graph for each gene. Also, the 3D file exists for each gene that is opened with a desktop app (Brain Explorer 2) and shows 3D expression in sagittal, coronal and horizontal sections in all ages (E11.5, E13.5, E15.5, E18.5, P4, P14, P28, and P56). It is hard to show all three sections here, so I used just sagittal sections of some genes in E11.5, E13.5, E15.5, and E18.5 ages.

Chapter 3: Results

Pluripotency and Neural Ectoderm Genes

The PPP can determine potential linkages of all human genes. This method and an online tool is based on that functionally related proteins have evolved from a common ancestor. The interaction between these proteins can be determined with the correlation of evolutionarily patterns. Here, the different genes were entered to PPP in purpose to determine their correlations. If the correlation between two genes was high, it is likely that these two have evolved from a common ancestor and are functionally related to each other.

The 5 pluripotency and 5 neural ectoderm genes were inputted to PPP. The PPP shows that the correlations between these two groups are not high, but the correlations between the 5 neural ectoderm genes are approximately high. The PPP also groups genes based on their evolutionary. Four genes of neural ectoderm (HESX1, OTX2, GBX2, and SIX3) are grouped in one cluster (figure 5). So, SOX1 is separated from other neural ectoderm genes because of its correlation and evolution with others. We used the other four genes (HESX1, OTX2, GBX2, and SIX3) to input to PPP for further exploration.

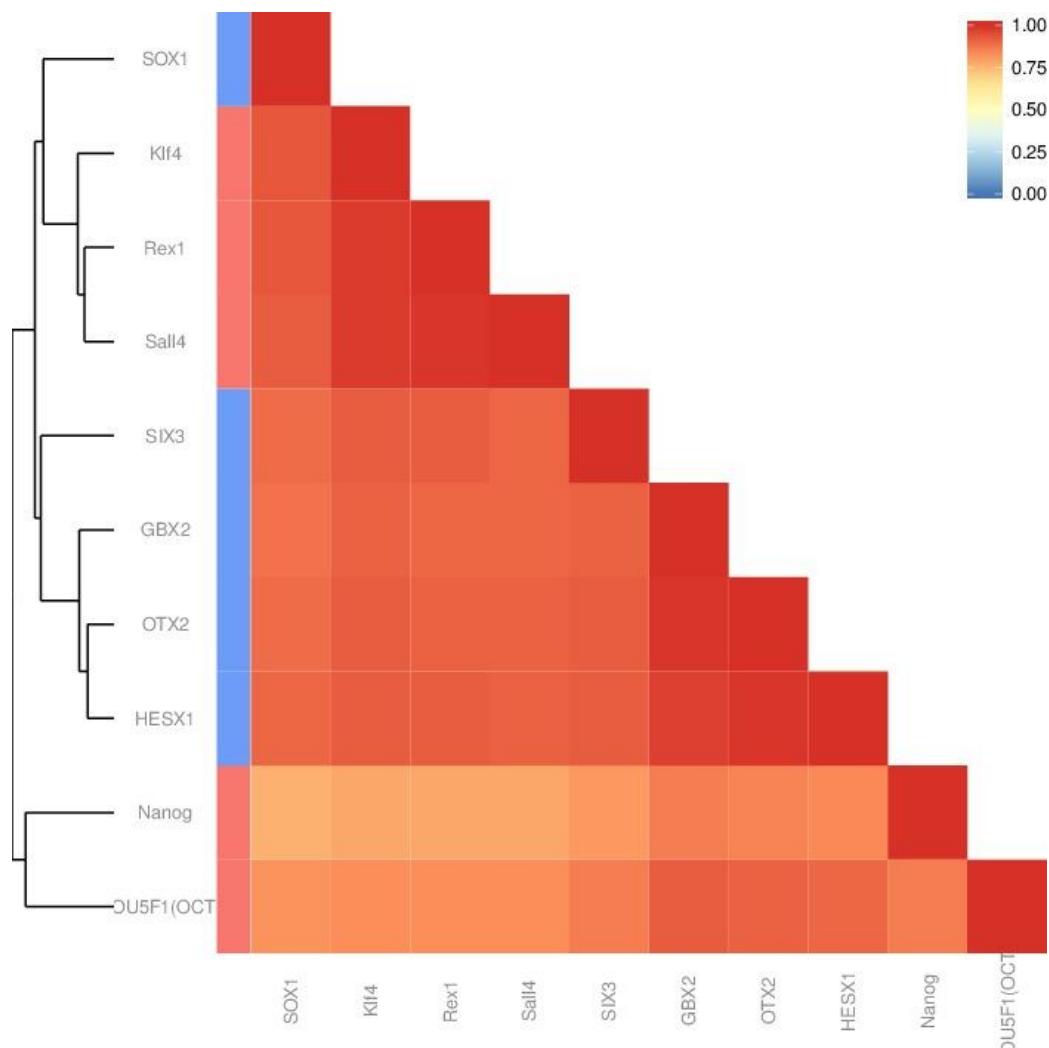


Figure 5. The correlation matrix of the ten genes: 5 pluripotent genes included OCT4, NANOG, SALL4, REX1 and KLF4, and 5 neural ectoderm genes included HESX1, OTX2, GBX2, SIX3, and SOX1. The legend (upper right) shows different color from dark blue to red based different correlations from no to high correlation. PPP can cluster genes based on their correlations. This clustering is shown in the left of the genes.

Correlated Genes with HESX1, OTX2, GBX2, and SIX3

The four neural ectoderm genes were inputted to PPP. PPP can detect other genes that have correlations with these four genes. The resulted genes are many. So, the genes were selected that correlate with all four genes. No gene correlated with four genes, but 35 genes

were correlated with three genes of four neural ectoderm genes. These 35 genes included SLC15A5, MPPE1, POU4F1, GID4, BCAP31, SPTB, DPRX, ISL2, PAX7, PROP1, CRX, HOXD13, DMBX1, ALX1, PITX1, PITX2, OTP, PITX3, SHOX, DUXA, VAX1, EMX1, ISX, PHOX2A, PHOX2B, DUX1, DUX5, RAX, RAX2, HOXC13, HHEX, ALX4, ARX, GSX2 and HOXC11.

Correlations between the 35 Genes that Correlated with HESX1, OTX2, GBX2 and SIX3

The 35 genes that PPP found and four neural ectoderm genes were inputted to PPP to study the correlations that existed between them and cluster them based on these correlations. The high correlation between two genes shows that they are functionally and evolutionarily related to each other. Also, PPP grouped these 35 genes based on this information (figure 6). These clusters included:

- HOXC11 & GSX2
- ARX & ALX4 & HHEX & HOXC13
- RAX & RAX2
- DUX5 & DUX1
- PHOX2A & PHOX2B
- ISX & EMX1 & VAX1
- PITX3 & OTP
- PITX1 & PITX2

Correlation Matrix of 35 Random Genes and Four Neural Ectoderm Genes

The previous correlation matrix plots (figure 5 and 6) only shows high correlations (orange and red colors). To test that PrePhyloPro can detect all correlations (low, high and without correlations), we randomly choose 35 genes. They were inputted to PrePhyloPro with four neural ectoderm genes. The output is a correlation matrix plot with all different correlations from zero to high (color ranges from dark blue to dark red, figure 7). Also, the dendrogram is different between the two plots (figure 6 and 7). The correlation matrix plot of random genes (figure 7) acts as a negative control for the performance of PrePhyloPro.

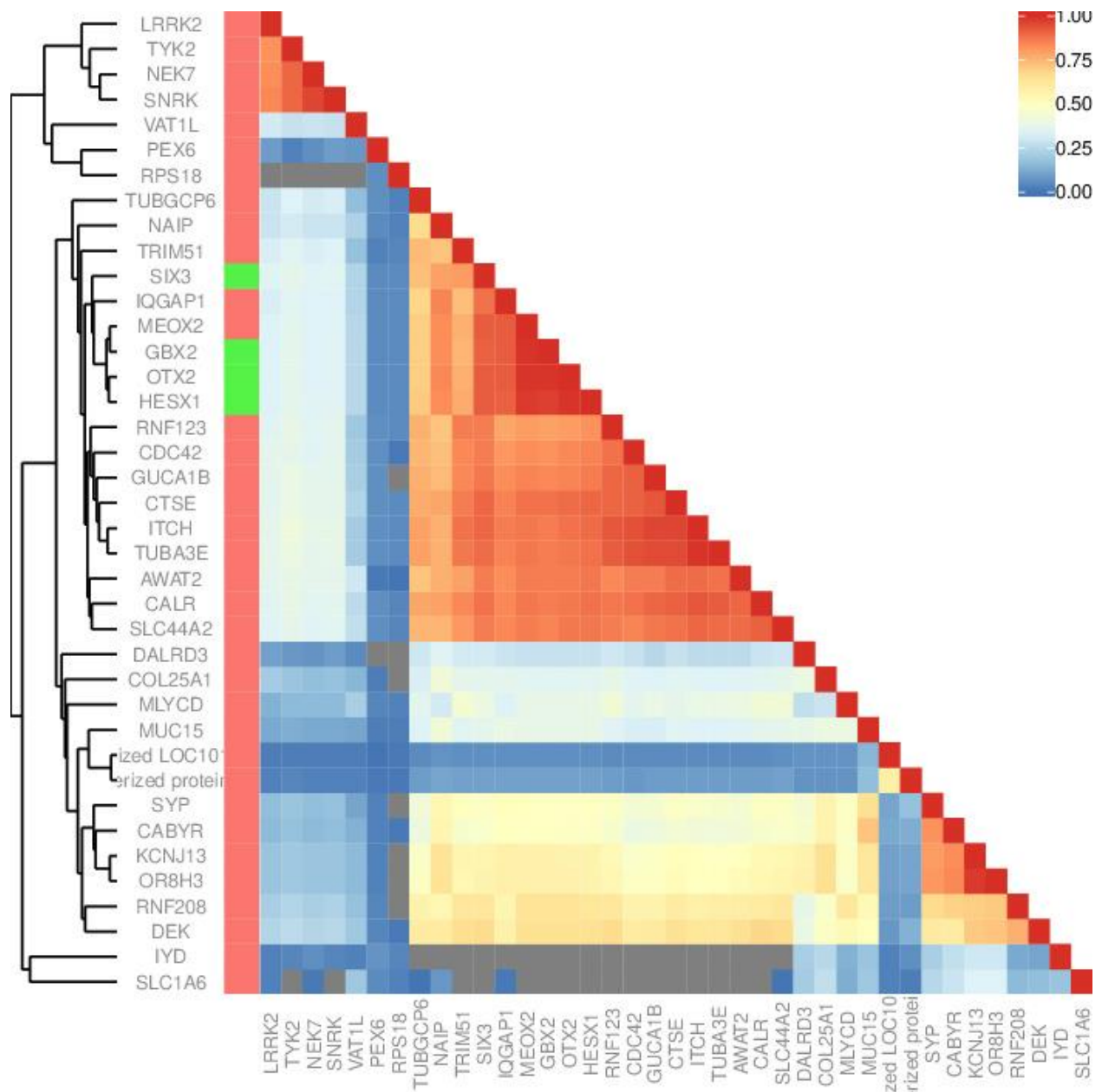


Figure 7. The correlation matrix four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) and 35 random genes that correlated with them. The legend (upper right) shows different color from dark blue to red based different correlations from no to high correlation. PPP can cluster genes based on their correlations. This clustering is shown in the left of the genes.

Expression of Genes during Developing Mouse Brain

DEVELOPING MOUSE BRAIN ATLAS section of Allen brain atlas contains the expression of the most genes during developing mouse brain. Here, the expression of four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) act as the positive control group (figure 8). The result shows that there is no expression for HESX1 (figure 8.a). The expression of the 35 genes correlated with four neural ectoderm genes act as the experimental or test group. There was no result for some genes such as SLC15A5, MPPE1, GID4, BCAP31, SPTB, DPRX, SHOX, DUXA, DUX1, DUX5, and RAX2. The expression of the remaining genes shows in figures 9 & 10 & 11. The pattern expression of both groups is compared to each other. If the pattern of a gene from the experimental group similar to a positive group, it is very likely that they have similar or correlated functions or perhaps they have interaction in that age and anatomic region.

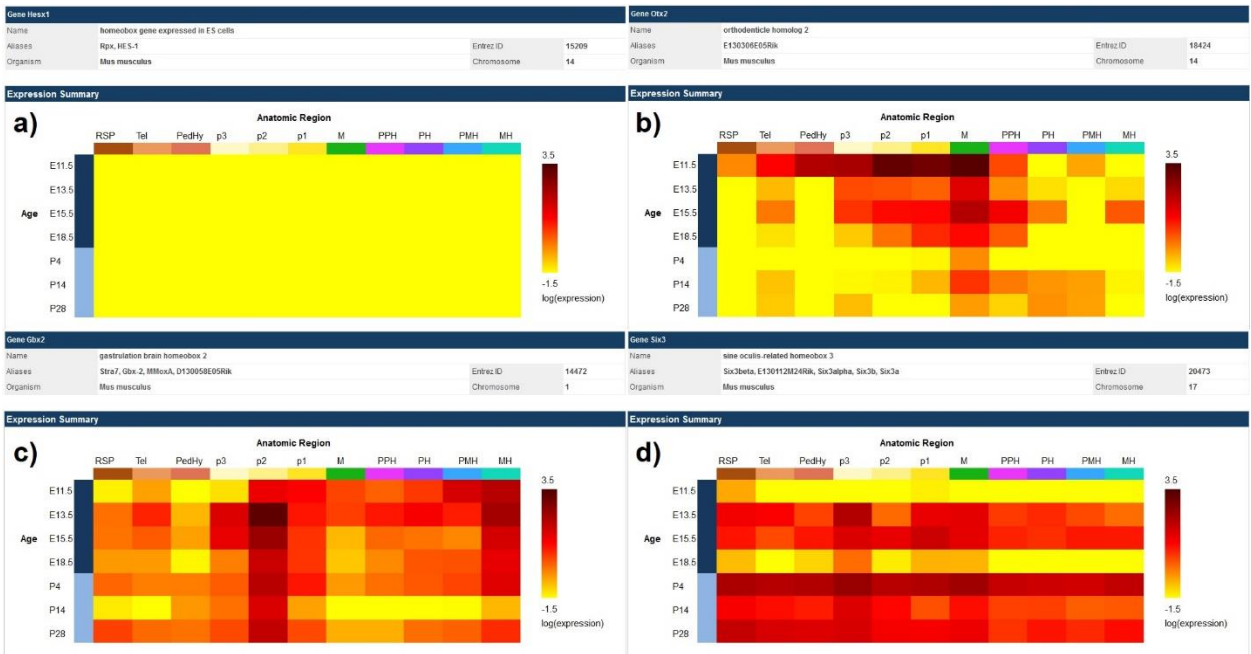


Figure 8. The raw expressions of the four neural ectoderm genes included: HESX1 (a), OTX2 (b), GBX2 (c) and SIX3 (d) in the different anatomic regions and ages of the developing mouse brain. The legend (upper right) shows different color from yellow to dark red based different expressions from no to high expression.

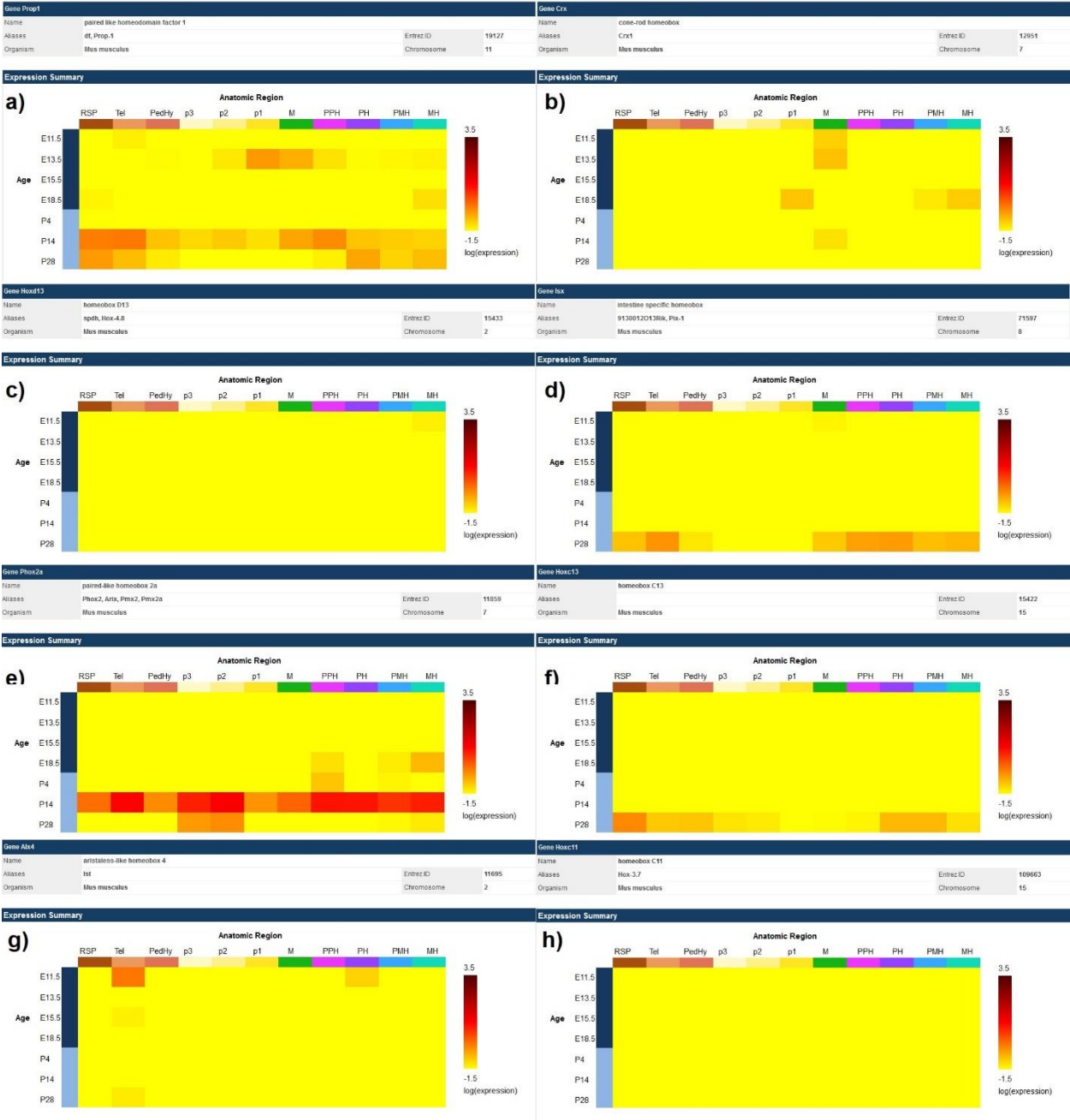


Figure 9. The raw expressions of PROP1 (a), CRX (b), HOXD13 (c), ISX (d), PHOX2A (e), HOXC13 (f), ALX4 (g) and HOXC11 (h) in different anatomic regions of developing mouse brain. The legend (upper right) shows different color from yellow to dark red based different expressions from no to high expression.

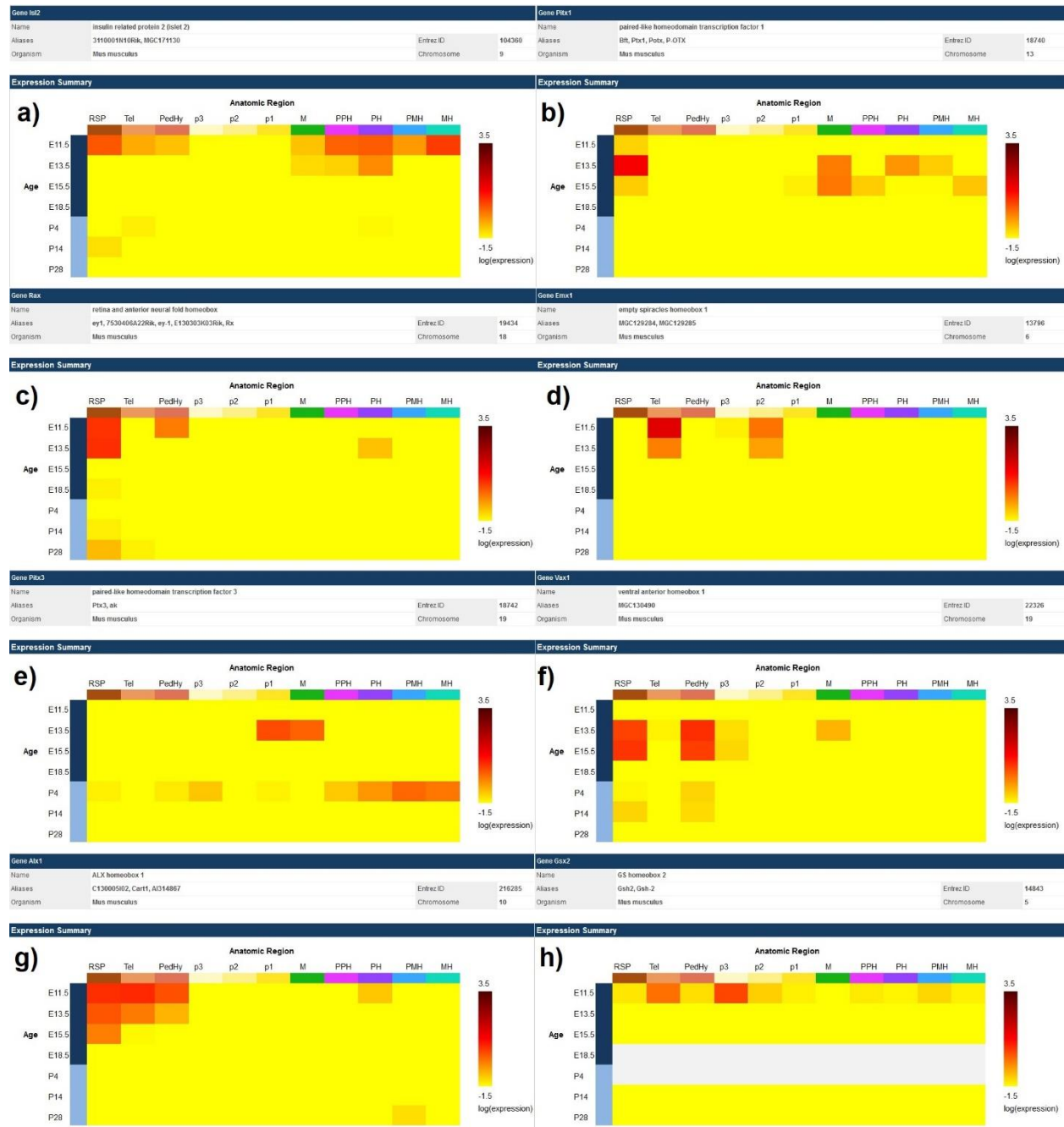


Figure 10. The raw expressions of ISL2 (a), PITX1 (b), RAX (c), EMX1 (d), PITX3 (e), VAX1 (f), ALX1 (g) and GSX2 (h) in different anatomic regions of developing mouse brain. The legend (upper right) shows different color from yellow to dark red based different expressions from no to high expression.

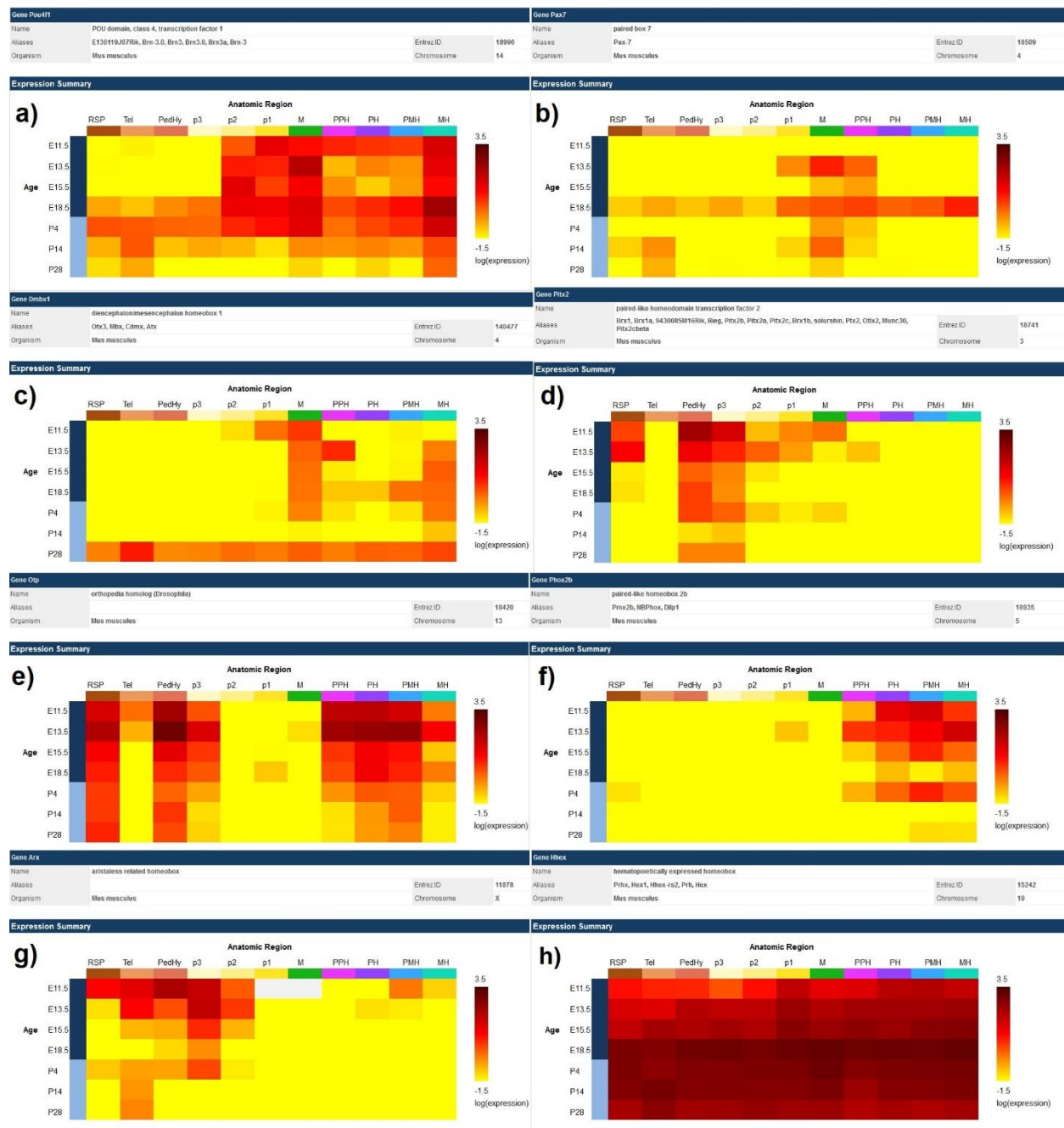


Figure 11. The raw expressions of POU4F1 (a), PAX7 (b), DMBX1 (c), PITX2 (d), OTP (e), PHOX2B (f), ARX (g) and HHX (h) in different anatomic regions of developing mouse brain. The legend (upper right) shows different color from yellow to dark red based different expressions from no to high expression.

The sagittal sections of ISH give us a high quality to compare the pattern expression between different genes. First, the sagittal sections of expression of 3 neural ectoderm genes (OTX2, GBX2, and SIX3) were studied (figure 12). Both OTX2 and GBX2 have an expression in E113.5, E13.5, E15.5 and E18.5 (figure 8, 12), but SIX3 is expressed in E13.5 and E15.5. Pay attention that sagittal section (figure 12.c) shows no expression for SIX3 in E15.5, but SIX3 has an expression in E15.5 according to figure 8.d. The sagittal section of the gene expressions of figure 11 (except HHEX) is shown in figure 13. HHEX is a ubiquitous gene because it is expressed everywhere.

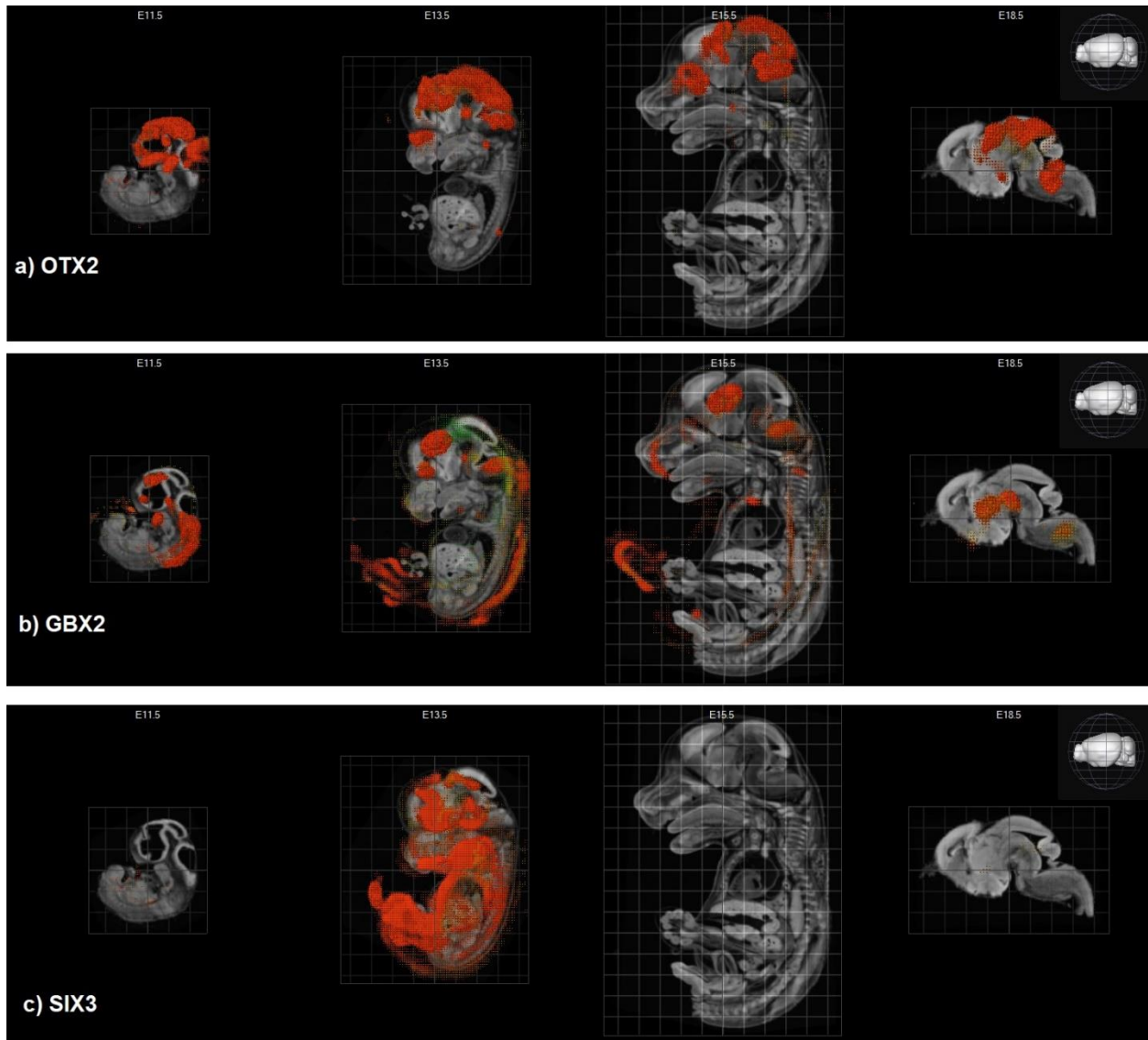
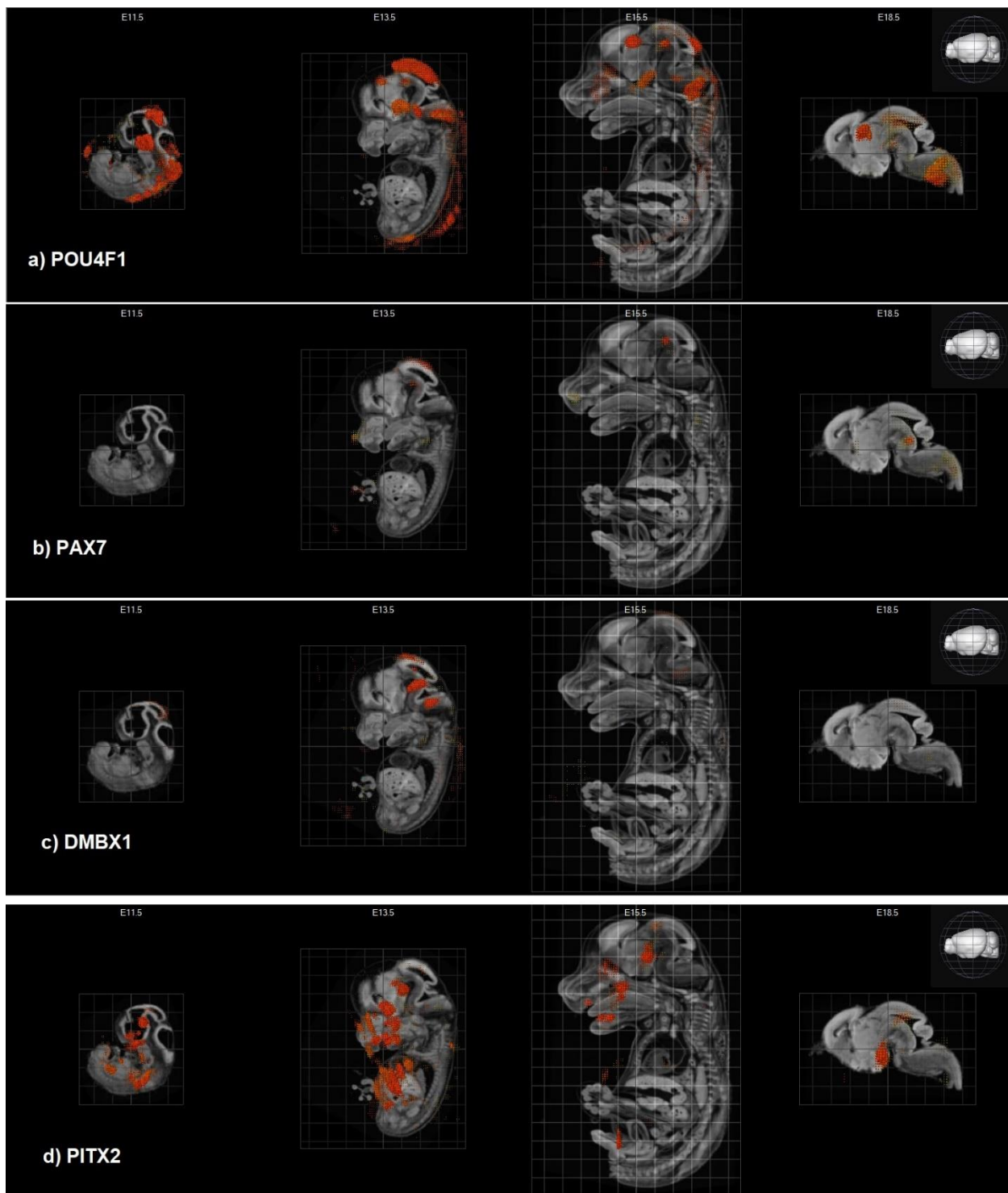


Figure 12. The sagittal sections of the developing mouse brain in the E11.5, E13.5, E15.5, and E18.5 ages. These sections show the ISH of OTX2 (a), GBX2 (b) and SIX3(c).



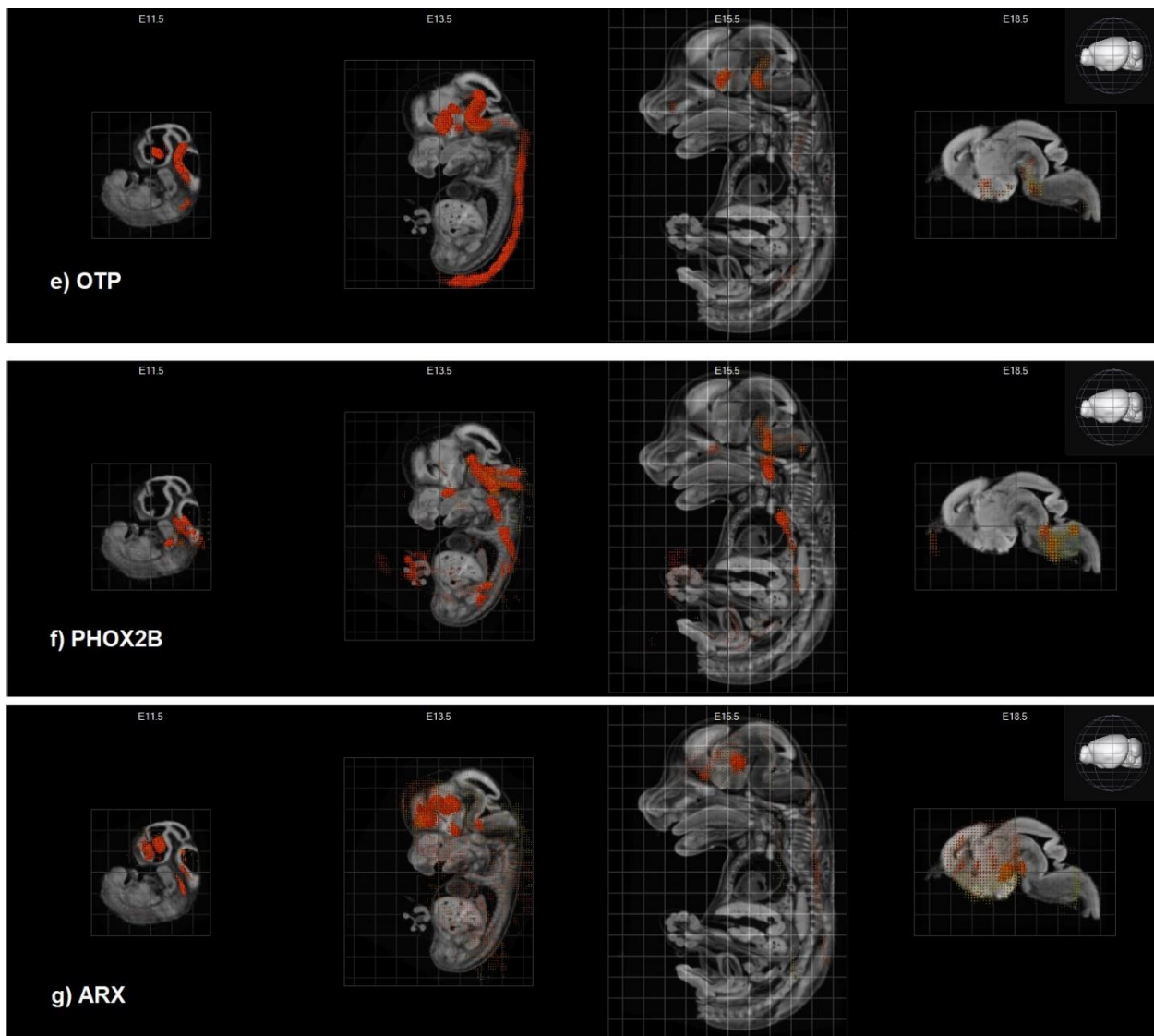


Figure 13. The sagittal sections of the developing mouse brain in the E11.5, E13.5, E15.5, and E18.5 ages. These sections show the ISH of POU4F1 (a), PAX7 (b), DMBX1 (c), PITX2 (d), OTP (e), PHOX2B (f) and ARX (g).

Similar Expression between Genes

According to figures 8-13, we created four tables (tables 1-4) that shows the expression of the genes in different ages and anatomic regions. With this table, we can compare genes

expression easily and find genes expressed at the same time and anatomic regions with each other.

Gene/Anatomic Region	RSP	Tel	PedHy	p3	p2	p1	M	PPH	PH	PMH	MH
OTX2											
GBX2											
SIX3											
POU4F1											
PAX7											
DMBX1											
PITX2											
OTP											
PHOX2B											
ARX											

Table 1. Expression of genes during age E11.5 and anatomic regions. RSP: rostral secondary prosencephalon, Tel: telencephalic vesicle, PHy: peduncular (caudal) hypothalamus, p3: prosomere, p2: prosomere 2, p1: prosomere 1, M: midbrain, PPH: prepontine hindbrain, PH: pontine hindbrain, PMH: pontomedullary hindbrain, MH: medullary hindbrain.

Gene/Anatomic Region	RSP	Tel	PedHy	p3	p2	p1	M	PPH	PH	PMH	MH
OTX2											
GBX2											
SIX3											
POU4F1											
PAX7											
DMBX1											
PITX2											
OTP											
PHOX2B											
ARX											

Table 2. Expression of genes during age E13.5 and anatomic regions. RSP: rostral secondary prosencephalon, Tel: telencephalic vesicle, PHy: peduncular (caudal) hypothalamus, p3: prosomere, p2: prosomere 2, p1: prosomere 1, M: midbrain, PPH: prepontine hindbrain, PH: pontine hindbrain, PMH: pontomedullary hindbrain, MH: medullary hindbrain.

Gene/Anatomic Region	RSP	Tel	PedHy	p3	p2	p1	M	PPH	PH	PMH	MH
OTX2											
GBX2											
SIX3											
POU4F1											
PAX7											
DMBX1											
PITX2											
OTP											
PHOX2B											
ARX											

Table 3. Expression of genes during age E15.5 and anatomic regions. RSP: rostral secondary prosencephalon, Tel: telencephalic vesicle, PHy: peduncular (caudal) hypothalamus, p3: prosomere, p2: prosomere 2, p1: prosomere 1, M: midbrain, PPH: prepontine hindbrain, PH: pontine hindbrain, PMH: pontomedullary hindbrain, MH: medullary hindbrain.

Gene/Anatomic Region	RSP	Tel	PedHy	p3	p2	p1	M	PPH	PH	PMH	MH
OTX2											
GBX2											
SIX3											
POU4F1											
PAX7											
DMBX1											
PITX2											
OTP											
PHOX2B											
ARX											

Table 4. Expression of genes during age E18.5 and anatomic regions. RSP: rostral secondary prosencephalon, Tel: telencephalic vesicle, PHy: peduncular (caudal) hypothalamus, p3: prosomere, p2: prosomere 2, p1: prosomere 1, M: midbrain, PPH: prepontine hindbrain, PH: pontine hindbrain, PMH: pontomedullary hindbrain, MH: medullary hindbrain.

Chapter 4: Discussion

The nervous system is the most complex system of the human body. Also, its development is very complex and some of its mechanisms are unclear yet. Many genes, proteins, transcription factors, and mechanisms take part in neural development and differentiation. The more information about these mechanisms and processes can help scientist to find better treatments for the neural disorders (especially those with fetal origin). Here, the two key methods that are used are PPP and Allen brain atlas. PPP can find genes that have similar evolutions and interactions. The expressions of different genes exist in Allen brain atlas during different embryonic ages. In vivo or vitro experiments are very expensive and time-consuming, and sometimes they reached no results. Software like PPP can help scientists to do more study before their real experiments and direct their experiments in a more specific way.

The pluripotent genes are the first genes expressed in the embryo. Their function is to maintain stem cells in self-renewal and proliferate indefinitely stages. Also, these genes suppress differentiation in the stem cells. In contrast, the neural ectoderm genes have the opposite function. They differentiate stem cells to neural ectoderm. These two groups of genes (pluripotent and neural ectoderm genes) suppress the expression of each other. As figure 5 shows the correlation between these two groups is not high. So, it is expected that the interactions and functional linkage between these two groups are low. The interesting thing is that PPP separates SOX1 (a neural ectoderm gene) from the other four neural ectoderm genes and clusters this neural ectoderm gene with KLF4, REX1, and SALL4 (pluripotent genes). The four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) are in one cluster.

The four neural ectoderm genes (HESX1, OTX2, GBX2, and SIX3) are responsible to differentiate the stem cells into neural ectoderm. They are early developing genes. Also, they are expressed in the developing mouse brain. But, our data show no expression of HESX1 (figure 8.a) during E11.5 to P28. Probably, this is because of technical problems, for example, the probe of this gene doesn't work. Our data shows the expression for the other three neural ectoderm genes (figure 8). OTX2 is expressed in most anatomic regions during E11.5, but its expression is restricted in some anatomic regions and midbrain in later days. After birth, it continues its expression in the midbrain (figure 8.b). GBX2 almost maintains its expression pattern in all ages (It is expressed in p2, p1, and MH, figure 8.c). SIX3 is expressed in all anatomic regions in ages E13.5, E15.5, P4, P14, and P28, but there is no expression for it in ages E11.5 and E18.5 (figure 8.d) and this data is strange. It is very rare in the development that a gene is express during a specific time, and suppress in other times. Probably, technical problems cause no expression for SIX3 in ages E11.5 and E18.5.

Our data also shows that no expression for some correlated genes of four neural ectoderm genes. These genes included HOXD13 (figure 9.c) and HOXC11 (figure 9.h). Probably, this no expression is because of technical problems, or perhaps these two genes have no expression in the developing mouse brain. Also, the expression of other some genes is very low (figure 9). According to the expressions data (table 1-4), we can divide the remaining genes into three categories:

- Early developing genes: genes that expressed in the early days and suppress after (DMBX1, PITX2, and PHOX2B)
- Late developing genes: genes that expressed in the late days of embryonic (PAX7 or ARX)

- Genes that expressed approximately during all ages

Previous researches showed a sequence similarity between DMBX1 and OTX genes. DMBX1 (OTX3) represses the OTX2 activity (Wong *et al.*, 2010), (Kimura *et al.*, 2005), (Gogoi *et al.*, 2002). OTX2 and GBX2 are necessary for brain structures, but DMBX1 is not (Ohtoshi and Behringer, 2004). OTX2 and GBX2 are necessary for midbrain and hindbrain development, respectively. Our result suggests that DMBX1 is probably the downstream target of OTX2 and GBX2 in the early developments of midbrain and hindbrain, respectively (table 1, 2, 3, figure 13.c). After birth, the expression of DMBX1 is decreased in midbrain and hindbrain, but it is expressed almost in all anatomic region in age P28 (figure 9.c). Probably, its function or signaling pathway is changed after birth.

PAX7 is a key gene for midbrain development. Its expression is restricted during brain development. Also, OTX2 is responsible to maintain the expression of PAX7 (Di Giovannantonio *et al.*, 2014), (Zemke *et al.*, 2015). Also, our data suggest (figure 11.b, table 1-4) that PAX7 has functional linkage with OTX2 during the development of midbrain. This gene is expressed in the midbrain in ages E13.5 and E18.5 as OTX2 (figure 11.b). After birth, its expression is decreased in the midbrain.

OTP is an important gene for brain progenitor generation. Also, it is specifically important for the development of hypothalamus (Curt, Yaghmaeian Salmani and Thor, 2019). Our results show that it is expressed widely in most anatomic regions and all ages (figure 11.e). So, it is expected that this gene potentially has the general functions for brain development. Interestingly, almost its expression pattern doesn't change between different ages. So, OTP

probably maintains its functions between different ages. Also, it has an expression in hindbrain during ages E11.5 to E18.5 as GBX2 (table 1-4). Perhaps, these two genes have interactions or functional linkage during the development of hindbrain.

In the early days of the embryo, PITX2 is expressed in RSP, PedHy, and p3. Also, it has the expression in midbrain on ages E11.5 and E13.5, but no expression on ages E15.5 and E18.5. Probably, in the early ages, PITX2 has interactions with OTX2 and GBX2, but in later ages, PITX2 loss its function in the developing midbrain, but maintains its expression in RSP and PedHy. Previous researches show that PITX2 is necessary for pituitary development (Stallings and Ellsworth, 2018).

ARX is necessary for forebrain patterning and growth (Lim *et al.*, 2019). Our data shows its expression in the forebrain in the early days (figure 13.g). ARX maintains its expression in RSP, Tel, PedHy, p3, and p2 of all ages. Probably, it has a role in the developing of these regions. Also, in the early days of the embryo, ARX has no expression in midbrain and hindbrain, but our data shows it is expressed in midbrain and hindbrain in age E18.5 (Table 4). So, ARX potentially acts as a late developmental factor and has functional linkage with OTX2 and GBX2 in the midbrain and hindbrain development, respectively.

PHOX2B is expressed in the hindbrain in all ages except E18.5. It maintains its expression in hindbrain. So, probably, it has no functional change during brain development. This gene has co-expression with OTX2, GBX2, and SIX3 in hindbrain in the early days of the embryo. Probably, it is potentially a gene for developing hindbrain and have interaction with GBX2.

POU4F1 has an expression in midbrain and hindbrain in all ages. Also, its expression pattern is very similar to GBX2 and OTX2 in both midbrain and hindbrain. Previous studies show that this gene identifies the first differentiated neurons of the midbrain. Also, it has expression only in post-mitotic neurons (Smith, Dawson and Latchman, 1997). Our data shows the expression of POU4F1 is midbrain in all ages (figure 11.a). As this gene is expressed from early days in the midbrain, probably it has functional linkage with OTX2 during the development of the embryo. Other studies pointed out about the expression of POU4F1 in hindbrain sensory ganglia (Lanier *et al.*, 2009). Also, we showed here (figure 11.a) POU4F1 has an expression in hindbrain in ages E11.5 to E18.5 as GBX2. Potentially, POU4F1 has functional linkage with GBX2 during hindbrain development.

The expression pattern of each gene can give us other information. For example, in the early days of the embryo, the expression of PAX7 is limited to M and PPH (table 2). But, in the late days, its expression is reached to p1, PH, PMH, and MH too. Its regional identity is changed. Also, the sagittal section of developing brain during ages E11.5, E13.5 and E15.5 can be used to compare the expression of different genes in the spinal cord. For example, in the E13.5, GBX2, OTP, and POU4f1 are expressed in the spinal cord (figures 12 and 13). Probably, these genes have interaction during the development of the spinal cord.

Many aspects of neural differentiation are unclear yet. Here, the expressions of the genes were studied just in the developing mouse brain. Further studies can study them on the developing spinal cord or peripheral nervous system. Also, here we focused on the ages of E11.5 to E18.5. In the future, the expression of these genes can be studied after birth in detail. The mechanisms of how these genes develop the neural system can be examined. Also, PPP finds

many genes that correlate with the neural ectoderm genes. Here, we only study 35 of them that correlated with three of four neural ectoderm genes. Further studies can examine remaining genes, genes that correlated with two or one of these four neural ectoderm genes. Also, further studies can use other methods to examine the correlation between genes. They can examine the sequences of them and compare. Also, other databases existed. For example, single-cell RNA sequencing data can be used in further studies.

The mechanism and function of some genes are clearer during neural development. Further studies can determine the evolutionarily correlation or expression pattern between these genes and the 35 novel genes to further investigate functional and mechanistic interactions.

References:

- Abdalla-Elsayed, M. E. A. *et al.* (2017) 'Heterozygous mutation in OTX2 associated with early-onset retinal dystrophy with atypical maculopathy', *Molecular Vision*.
- Abouzeid, H. *et al.* (2012) 'RAX and anophthalmia in humans: Evidence of brain anomalies', *Molecular Vision*.
- Acampora, D. *et al.* (1999) 'Progressive impairment of developing neuroendocrine cell lineages in the hypothalamus of mice lacking the Orthopedia gene', *Genes and Development*. doi: 10.1101/gad.13.21.2787.
- Ahmad, A. *et al.* (2017) 'Expression of a SOX1 overlapping transcript in neural differentiation and cancer models', *Cellular and Molecular Life Sciences*. doi: 10.1007/s00018-017-2580-3.
- Amir-Zilberstein, L. *et al.* (2012) 'Homeodomain Protein Otp and Activity-Dependent Splicing Modulate Neuronal Adaptation to Stress', *Neuron*. doi: 10.1016/j.neuron.2011.11.019.
- Bell, C. E. and Watson, A. J. (2013) 'p38 MAPK Regulates Cavitation and Tight Junction Function in the Mouse Blastocyst', *PLoS ONE*. doi: 10.1371/journal.pone.0059528.
- Boulay, G. *et al.* (2017) 'OTX2 activity at distal regulatory elements shapes the chromatin landscape of group 3 medulloblastoma', *Cancer Discovery*. doi: 10.1158/2159-8290.CD-16-0844.
- Briata, P. *et al.* (1996) 'EMX1 homeoprotein is expressed in cell nuclei of the developing cerebral cortex and in the axons of the olfactory sensory neurons', *Mechanisms of Development*. doi: 10.1016/0925-4773(96)00544-8.
- Catala, M. and Kubis, N. (2013) 'Gross anatomy and development of the peripheral nervous system', in *Handbook of Clinical Neurology*. doi: 10.1016/B978-0-444-52902-2.00003-5.
- Collin, J. *et al.* (2016) 'Using Zinc Finger Nuclease Technology to Generate CRX-Reporter Human Embryonic Stem Cells as a Tool to Identify and Study the Emergence of Photoreceptors Precursors during Pluripotent Stem Cell Differentiation', *Stem Cells*. doi: 10.1002/stem.2240.
- Curt, J. R., Yaghmaeian Salmani, B. and Thor, S. (2019) 'Anterior CNS expansion driven by brain transcription factors', *eLife*. doi: 10.7554/elife.45274.
- Domené, S. *et al.* (2008) 'Mutations in the human SIX3 gene in holoprosencephaly are loss of function', *Human Molecular Genetics*. doi: 10.1093/hmg/ddn294.
- Faedo, A. *et al.* (2017) 'Differentiation of human telencephalic progenitor cells into MSNs by inducible expression of Gsx2 and Ebf1', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1611473114.
- Fang, Q. *et al.* (2016) 'HESX1 mutations in patients with congenital hypopituitarism: variable phenotypes with the same genotype', *Clinical Endocrinology*. doi: 10.1111/cen.13067.
- Feng, N. *et al.* (2014) 'Generation of highly purified neural stem cells from human adipose-derived mesenchymal stem cells by Sox1 activation', *Stem Cells and Development*. doi: 10.1089/scd.2013.0263.
- Gawlik-Rzemieniewska, N. and Bednarek, I. (2016) 'The role of NANOG transcriptional factor in the development of malignant phenotype of cancer cells', *Cancer Biology and Therapy*. doi:

10.1080/15384047.2015.1121348.

Gilbert, S. F. (2000a) 'Differentiation of the Neural Tube (6th Edition)', pp. 1–4.

Gilbert, S. F. (2000b) 'Formation of the Neural Tube', in *Developmental Biology*.

Di Giovannantonio, L. G. *et al.* (2014) 'Otx2 cell-autonomously determines dorsal mesencephalon versus cerebellum fate independently of isthmus organizing activity', *Development (Cambridge)*. doi: 10.1242/dev.102954.

Gogoi, R. N. *et al.* (2002) 'The paired-type homeobox gene Dmbx1 marks the midbrain and pretectum', *Mechanisms of Development*. doi: 10.1016/S0925-4773(02)00067-9.

Hallonet, M. *et al.* (1999) 'Vax1, a novel homeobox-containing gene, directs development of the basal forebrain and visual system', *Genes and Development*. doi: 10.1101/gad.13.23.3106.

Heimbucher, T. *et al.* (2007) 'Gbx2 and Otx2 Interact with the WD40 Domain of Groucho/Tle Corepressors', *Molecular and Cellular Biology*. doi: 10.1128/mcb.00811-06.

Henderson, R. H. *et al.* (2009) 'A rare de novo nonsense mutation in OTX2 causes early onset retinal dystrophy and pituitary dysfunction.', *Molecular vision*.

Hohenauer, T. *et al.* (2013) 'The neural crest transcription factor Brn3a is expressed in melanoma and required for cell cycle progression and survival', *EMBO Molecular Medicine*. doi: 10.1002/emmm.201201862.

Hsieh, M. H. *et al.* (2017) 'PARP1 controls KLF4-mediated telomerase expression in stem cells and cancer cells', *Nucleic Acids Research*. doi: 10.1093/nar/gkx683.

Kaur, R. *et al.* (2015) 'OTX2 exhibits cell-context-dependent effects on cellular and molecular properties of human embryonic neural precursors and medulloblastoma cells', *DMM Disease Models and Mechanisms*. doi: 10.1242/dmm.020594.

Kimura, K. *et al.* (2005) 'Functional analysis of transcriptional repressor Otx3/Dmbx1', *FEBS Letters*. doi: 10.1016/j.febslet.2005.04.042.

Lanier, J. *et al.* (2009) 'Brn3a regulates the transition from neurogenesis to terminal differentiation and represses non-neural gene expression in the trigeminal ganglion', *Developmental Dynamics*. doi: 10.1002/dvdy.22145.

Larkin, S. and Ansorge, O. (2000) *Development And Microscopic Anatomy Of The Pituitary Gland, Endotext*.

Di Lascio, S. *et al.* (2016) 'PHOX2A and PHOX2B are differentially regulated during retinoic acid-driven differentiation of SK-N-BE(2)C neuroblastoma cell line', *Experimental Cell Research*. doi: 10.1016/j.yexcr.2016.02.014.

Lee, M. Y., Lu, A. and Gudas, L. J. (2010) 'Transcriptional regulation of Rex1 (zfp42) in normal prostate epithelial cells and prostate cancer cells', *Journal of Cellular Physiology*. doi: 10.1002/jcp.22071.

Lim, Y. *et al.* (2019) 'Arx Expression Suppresses Ventralization of the Developing Dorsal Forebrain', *Scientific Reports*. doi: 10.1038/s41598-018-36194-6.

Lin, J. *et al.* (2016) 'Pax3 and Pax7 interact reciprocally and regulate the expression of cadherin-7 through inducing neuron differentiation in the developing chicken spinal cord', *Journal of Comparative*

Neurology. doi: 10.1002/cne.23885.

London, N. J. S. *et al.* (2009) 'Sequence alterations in RX in patients with microphthalmia, anophthalmia, and coloboma', *Molecular Vision*.

Luan, Z. *et al.* (2013) 'SUMOylation of Pax7 is essential for neural crest and muscle development', *Cellular and Molecular Life Sciences*. doi: 10.1007/s00018-012-1220-1.

Marsh, E. D. *et al.* (2016) 'Developmental interneuron subtype deficits after targeted loss of Arx', *BMC Neuroscience*. doi: 10.1186/s12868-016-0265-8.

Martin, D. M. *et al.* (2004) 'PITX2 is required for normal development of neurons in the mouse subthalamic nucleus and midbrain', *Developmental Biology*. doi: 10.1016/j.ydbio.2003.10.035.

Mullen, A. C. and Wrana, J. L. (2017) 'TGF- β family signaling in embryonic and somatic stem-cell renewal and differentiation', *Cold Spring Harbor Perspectives in Biology*. doi: 10.1101/cshperspect.a022186.

Nasrallah, M. P. *et al.* (2012) 'Differential effects of a polyalanine tract expansion in Arx on neural development and gene expression', *Human Molecular Genetics*. doi: 10.1093/hmg/ddr538.

Newbern, K. *et al.* (2013) 'Identification of HESX1 mutations in Kallmann syndrome', *Fertility and Sterility*. doi: 10.1016/j.fertnstert.2013.01.149.

Nishiguchi, S. *et al.* (1998) 'Sox1 directly regulates the γ -crystallin genes and is essential for lens development in mice', *Genes and Development*. doi: 10.1101/gad.12.6.776.

Niu, Y. *et al.* (2017) 'PrePhyloPro: Phylogenetic profile-based prediction of whole proteome linkages', *PeerJ*. doi: 10.7717/peerj.3712.

Ohtoshi, A. and Behringer, R. R. (2004) 'Neonatal Lethality, Dwarfism, and Abnormal Brain Development in Dmbx1 Mutant Mice', *Molecular and Cellular Biology*. doi: 10.1128/mcb.24.17.7548-7558.2004.

Pasquier, L. *et al.* (2005) 'First occurrence of aprosencephaly/atelencephaly and holoprosencephaly in a family with a SIX3 gene mutation and phenotype/genotype correlation in our series of SIX3 mutations.', *Journal of medical genetics*. doi: 10.1136/jmg.2004.023416.

Purves D, Augustine GJ, Fitzpatrick D, *et al.* (2001) 'Formation of the Major Brain Subdivisions', (Box C), pp. 1–2.

Rossant, J. and Tam, P. P. L. (2009) 'Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse', *Development*. doi: 10.1242/dev.017178.

Sadler, T. W. (2005) 'Embryology of neural tube development', in *American Journal of Medical Genetics - Seminars in Medical Genetics*. doi: 10.1002/ajmg.c.30049.

Sakazume, S. *et al.* (2007) 'Functional analysis of human mutations in homeodomain transcription factor PITX3', *BMC Molecular Biology*. doi: 10.1186/1471-2199-8-84.

Slavotinek, A. M. *et al.* (2012) 'VAX1 mutation associated with microphthalmia, corpus callosum agenesis, and orofacial clefting: The first description of a VAX1 phenotype in humans', *Human Mutation*. doi: 10.1002/humu.21658.

Smith, M. D., Dawson, S. J. and Latchman, D. S. (1997) 'The Brn-3a transcription factor induces neuronal process outgrowth and the coordinate expression of genes encoding synaptic proteins.', *Molecular and Cellular Biology*. doi: 10.1128/mcb.17.1.345.

- Sohocki, M. M. *et al.* (1998) 'A range of clinical phenotypes associated with mutations in CRX, a photoreceptor transcription-factor gene', *American Journal of Human Genetics*. doi: 10.1086/302101.
- Stallings, C. E. and Ellsworth, B. S. (2018) 'Premature expression of FOXO1 in developing mouse pituitary results in anterior lobe hypoplasia', *Endocrinology*. doi: 10.1210/en.2018-00107.
- Stiles, J. and Jernigan, T. L. (2010) 'The basics of brain development', *Neuropsychology Review*. doi: 10.1007/s11065-010-9148-4.
- Sunnen, C. N. *et al.* (2014) 'Arx is required for specification of the Zona Incerta and reticular nucleus of the thalamus', *Journal of Neuropathology and Experimental Neurology*. doi: 10.1097/NEN.0000000000000048.
- Tatetsu, H. *et al.* (2016) 'SALL4, the missing link between stem cells, development and cancer', *Gene*. doi: 10.1016/j.gene.2016.02.019.
- Triplett, J. W. *et al.* (2014) 'Dendritic and axonal targeting patterns of a genetically-specified class of retinal ganglion cells that participate in image-forming circuits', *Neural Development*. doi: 10.1186/1749-8104-9-2.
- Vira, D. *et al.* (2012) 'Cancer stem cells, microRNAs, and therapeutic strategies including natural products', *Cancer and Metastasis Reviews*. doi: 10.1007/s10555-012-9382-8.
- Waite, M. R. and Martin, D. M. (2015) 'Axial level-specific regulation of neuronal development: Lessons from PITX2', *Journal of Neuroscience Research*. doi: 10.1002/jnr.23471.
- Wong, L. *et al.* (2010) 'Duplicate dmbx1 genes regulate progenitor cell cycle and differentiation during zebrafish midbrain and retinal development', *BMC Developmental Biology*. doi: 10.1186/1471-213X-10-100.
- Yesudhas, D. *et al.* (2016) 'Structural mechanism behind distinct efficiency of Oct4/Sox2 proteins in differentially spaced DNA complexes', *PLoS ONE*. doi: 10.1371/journal.pone.0147240.
- Yu, Z. *et al.* (2017) 'SIX3, a tumor suppressor, inhibits astrocytoma tumorigenesis by transcriptional repression of AURKA/B', *Journal of Hematology and Oncology*. doi: 10.1186/s13045-017-0483-2.
- Zemke, M. *et al.* (2015) 'Loss of Ezh2 promotes a midbrain-to-forebrain identity switch by direct gene derepression and Wnt-dependent regulation', *BMC Biology*. doi: 10.1186/s12915-015-0210-9.