

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**THE ROLE OF ETS PROTEINS IN DROSOPHILA LINEAGE
FORMATION**

PINAR MUTLU
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

GEBZE
2023

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**THE ROLE OF ETS PROTEINS IN
DROSOPHILA LINEAGE FORMATION**

PINAR MUTLU
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

THESIS SUPERVISOR
PROF. DR. İŞİL KURNAZ

GEBZE
2023

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ETS PROTEİNLERİNİN DROSOPHİLA SOY
OLUŞUMUNDAKİ ROLÜ

PINAR MUTLU
YÜKSEK LİSANS TEZİ
MOLEKÜLER BİYOLOJİ VE GENETİK ANABİLİM DALI

DANIŞMANI
PROF. DR. IŞIL KURNAZ

GEBZE
2023



YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 19/01/2023 tarih ve 2023/04 sayılı kararıyla oluşturulan jüri tarafından 24/01/2023 tarihinde tez savunma sınavı yapılan Pınar MUTLU'nun tez çalışması Moleküler Biyoloji ve Genetik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Prof. Dr. Işıl KURNAZ

ÜYE

:Prof. Dr. Tunahan Çakır

ÜYE

:Dr. Öğr. Üyesi Neşe AYŞİT

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

...../...../..... tarih ve/..... sayılı kararı.

SUMMARY

ETS transcription factor family is one of the largest families and plays role in different processes such as apoptosis, nervous system development, cell proliferation etc. This family is divided into subfamilies according to similarities of conserved domains.

Drosophila larval brain contains different types of neuroblasts (NB) which can be distinguished from each other such as Type II NB, antennal lobe NB, mushroom body NB, optic lobe NB. During development, neuron and glia formation are tightly regulated by TFs both temporally and spatially.

Pnt, is used as a marker for Type II NB and also plays a role in glia formation. The antagonist of Pnt, Aop plays role in cell fate determination. According to the interfamilial relationship of ETS TFs like in Pnt and Aop, ETS genes may play a role in lineage formation. This thesis focused on genes regulated by the ETS TFs in neuroglial lineage formation. Neuron and glia formation was evaluated by comparing either neuron or glia samples with Type II NB, antennal lobe NB or mushroom body NB samples that were obtained from published *Drosophila* larva transcriptome data. Nervous system related genes were selected from genes that were found both in differentiated neuron and glia in comparison to NBs and promoter analysis was performed for ETS TFs. Thereafter, gene regulatory networks were created to illuminate the potential role of ETS TFs in the regulation of selected genes during nervous system development. These results may contribute to the understanding and potential use of ETS TFs in neuroregenerative treatments.

Key Words: ETS Family, Transcription Factor, *Drosophila melanogaster*, Lineage Formation.

ÖZET

ETS transkripsiyon faktörü ailesi en büyük transkripsiyon faktörü ailelerinden biridir ve apoptoz, sinir sistemi gelişimi, hücre çoğalması vb. gibi farklı süreçlerde rol oynar. Bu aile, evrimsel korunmuşluğu olan bölgelerinin benzerliklerine göre alt ailelere ayrılmıştır. Bu TF ailesinin bütün üyelerinin rolleri tam olarak bilinmemekle birlikte, aktivatör veya baskılayıcılar rolüne sahiptir.

Drosophila larva beyni, Tip II NB, anten lobu NB, mantar gövdesi NB, optik lob NB gibi ayırt edilebilen farklı nöroblast türleri içermektedir. Gelişim sırasında, bu nöroblastlardan nöron ve glia oluşumu ve nöroblastların yenilenmesi, hem zamansal hem de mekansal olarak transkripsiyon faktörleri tarafından sıkı bir şekilde düzenlenmektedir.

ETS genlerinden biri olan Pnt, Tip II NB için belirteç olarak kullanılır ve embriyoda glia oluşumunda rol almaktadır. Pnt'nin antagonisti olan Aop, hücre kaderinin belirlenmesinde rol oynamaktadır. Pnt ve Aop arasındaki ilişkide olduğu gibi ETS transkripsiyon faktörlerinin aileler arası ilişkisine göre, ETS genleri soy oluşumunda rol oynayabilir. Bu tez, nöroglial soy oluşumunda ETS transkripsiyon faktörleri tarafından düzenlenen genlere odaklanmıştır. Bu analizlerde, yayınlanmış *Drosophila* larva verilerinden elde edilen Tip II NB, anten lob NB veya mantar gövdesi NB örnekleri ile nöron veya glia örnekleri karşılaştırılarak değerlendirilmiştir. Sinir sistemi ile ilgili genler, hem nöron hem de glia karşılaştırmasında bulunan genler arasından seçilerek ETS transkripsiyon faktörleri için promotör analizi yapılarak belirlendi. Daha sonra, sinir sistemi gelişimi sırasında seçilen genlerin düzenlenmesinde ETS TF'lerin potansiyel rolünü aydınlatmak için gen düzenleyici ağlar oluşturulmuştur. Bu sonuçlar, ETS TF'lerin nörorejeneratif tedavilerde anlaşılmasına ve potansiyel kullanımına katkıda bulunabilir.

Anahtar Kelimeler: ETS Ailesi, Transkripsiyon Faktörleri, *Drosophila melanogaster*, Soy Oluşumu.

ACKNOWLEDGEMENTS

Firstly, I would like to thank Prof. Dr. Işıl AKSAN KURNAZ for her supervision, support and advice during my master study. I feel so lucky to be her student.

I would like to thank Ekin SONMEZ for her guidance, teaching me patiently and always supporting me, Yiğit Koray BABAL for his help in my all questions and his support. I would also like to thank AxanLab members, my teammates Büşra Nur ÇİÇEK, Betül ERTURAL, Mustafa Doğukan METİNER, Atılay İLGÜN, Sedef YUSUFOĞULLARI and Ebru KÖLEOĞLU for all their supports, friendship, being a family and good memories.

I would like to thank to all my close friends Fadime ÇİVİLİ, Sevim ÖZKUNT, Nil OĞUŞ, Melek GÖÇMEN, Buse AKSU, Berk İLERİ and Batuhan ŞEVİKER for their endless support, motivation, and friendship.

Finally, I especially thank to my mother Meliha MUTLU, my father Zekeriya MUTLU, and my brother Özgün MUTLU for all their supports, patience, and their endless love. Thank you.

TABLE of CONTENTS

SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
LIST of ABBREVIATIONS and ACRONYMS	ix
LIST of FIGURES	x
LIST of TABLES	xiii
1. INTRODUCTION	1
1.1. Drosophila Central Nervous System (CNS) Development	1
1.2. Drosophila Larval Neuroblasts	2
1.3. ETS Transcription Factor Family	5
2. MATERIALS and METHODS	7
2.1. Methods	7
2.1.1. GALAXY and data processing	7
2.1.2. Differentially Gene Expression Analysis	9
2.1.3. Gene Set Enrichment Analysis	10
2.1.4. Gene Regulatory Network Construction	10
3. RESULTS	12
3.1. Differential Gene Expression Analysis	12
3.2. Glia and Neuron Lineage Formation from Type II NB	15
3.2.1. Gene Regulatory Network in Glia Lineage from Type II NB	20
3.2.2. Gene Regulatory Network in Neuron Lineage from Type II NB	23
3.3. Glia and Neuron Lineage Formation from Antennal Lobe (AL) NB	25
3.4. Glia and Neuron Lineage Formation from Mushroom Body (MB) NB	31
REFERENCES	47
BIOGRAPHY	51

LIST of ABBREVIATIONS and ACRONYMS

<u>Abbreviations</u> <u>and Acronyms</u>	<u>Explanations</u>
AL	: Antennal Lobe
AOP	: Anterior Open
BAM	: Binary Alignment Map
DEG	: Differentially Expressed Genes
Eip74EF	: Ecdysone-induced protein 74EF
ETS	: E-twenty six Transformation Specific
Ets65A	: Ets at 65A
GEO	: Gene Expression Omnibus
GMC	: Ganglion Mother Cell
GO	: Gene Ontology
INP	: Intermediate Neural Progenitor
MB	: Mushroom Body
NB	: Neuroblast
NSC	: Neural Stem Cells
p-adjusted	: Adjusted p-value
PCA	: Principal Component Analysis
PEA3	: Polyoma Enhancer Activator
PNT	: Pointed
RNA-seq	: Ribonucleic Acid Sequencing
TF	: Transcription Factor
TSS	: Transcription Start Site
VNC	: Ventral Nerve Cord
vst	: Variance Stabilizing Transformation

LIST of FIGURES

<u>Figure No:</u>	<u>Page</u>
1.1: Drosophila larval brain anatomy. NB, neuroblast; VNC, ventral nerve cord.	2
1.2: Neurogenesis patterns of Drosophila neuroblasts (NBs).	4
1.3: The subfamilies of ETS transcription factor family in Drosophila melanogaster.	6
2.1: Methodology of transcriptome analysis of Drosophila larval dataset.	9
3.1: Transcriptome analysis of the Type II NB, AL NB, MB NB, neuron and glia samples.	14
3.2: Glia and neuron lineage formation from Type II NB.	17
3.3: Glia and Neuron Lineage Formation from Type II NB.	19
3.4: The gene regulatory network of glia and neuron lineage formation from Type II NB	20
3.5: The gene regulatory network which was obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using glia vs TypeII NB dataset.	23
3.6: The gene regulatory network which was obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using neuron vs TypeII NB dataset.	26
3.7: PCA analysis of the samples' normalized gene expression of glia and neuron lineage formation from AL NB.	27
3.8: Up and down regulated genes was shown in volcano plot in neuronal lineage formation from AL NB.	27
3.9: In glial lineage formation from AL NB, up and down regulated genes shown in volcano plot.	28
3.10: The expression levels of the ETS genes in both neuron and glial lineage formation from AL NB.	28

3.11:	Glia and Neuron Lineage Formation from AL NB.	29
3.12:	The gene regulatory network of common genes according to Glia vs AL NB dataset.	30
3.13:	The gene regulatory network of common genes according to neuron compared to AL NB dataset.	30
3.14:	The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using glia vs AL NB dataset.	31
3.15:	The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes of neuron compared to AL NB.	32
3.16:	PCA analysis of the samples' normalized gene expression of glia and neuron lineage formation from MB NB.	33
3.17:	Up and down regulated genes was shown in volcano plot in neuronal lineage formation from MB NB.	33
3.18:	In glial lineage formation from MB NB, up and down regulated genes shown in volcano plot.	34
3.19:	The expression levels of the ETS genes in both neuron and glial lineage formation from MB NB.	34
3.20:	Glia and Neuron Lineage Formation from MB NB.	35
3.21:	The gene regulatory network of common genes according to Glia vs MB NB dataset.	36
3.22:	The gene regulatory network of common genes according to neuron compared to MB NB dataset.	37
3.23:	The gene regulatory network which obtained from promoter analysis results for AOP, Ets65A, Eip74EF, Ets65A and PNT transcription factors of nervous system related genes of common genes by using glia vs MB NB dataset.	38
3.24:	The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription	39

factors of nervous system related genes of common genes of neuron compared to MB NB.

- | | | |
|-------|---|----|
| 3.25: | The gene regulatory network of genes that positively regulated by Ets TFs in glia vs Type II NB dataset. | 42 |
| 3.26: | The gene regulatory network of genes that positively regulated by Ets TFs in neuron was compared to Type II NB dataset. | 43 |
| 3.27: | The gene regulatory network of genes that positively regulated by Ets TFs in glia vs AL NB dataset. | 44 |
| 3.28: | The gene regulatory network of genes that positively regulated by Ets TFs in neuron was compared to AL NB dataset. | 45 |
| 3.29: | The gene regulatory network of genes that positively regulated by Ets TFs in glia was compared to MB NB dataset. | 46 |
| 3.30: | The gene regulatory network of genes that positively regulated by Ets TFs in neuron was compare MB NB dataset. | 47 |

LIST of TABLES

<u>Table No:</u>	<u>Page</u>
1.1: The sample information of GSE71104 dataset.	8

1. INTRODUCTION

1.1. *Drosophila* Central Nervous System (CNS) Development

Drosophila is a commonly used organism in research because of its short developmental time as well as for the broad similarity of developmental programs to those in mammalian organisms. Fly completes its development from embryo to first instar larva about one day and in about 10 days larval growth and metamorphosis occurs [1].

The *Drosophila* central nervous system (CNS) is formed by the assembly of the segments called neuromere [2]. This CNS can be divided into 2 parts: brain and ventral nerve cord (VNC). Each part contains different number of neuromeres and homology between neuromeres can be understood by looking at the gene expression profile of the neural stem cells, or neuroblasts (NBs), in each neuromer [1],[3]. The combination of the transcription factors (TFs) generate distinct neuroblasts and neuronal identities.

In *Drosophila* development, neurogenesis occurs in 3 stages: (i) embryonic stage, (ii) larval stage and (iii) pupal stage [4], [5]. NBs in embryo divide asymmetrically and can give different cell types including (a) another NB, (b) a ganglion mother cell (GMC) small intermediate progenitors that produce neuron and glia, (c) intermediate neural precursor (INP), (d) neuron or (e) glia [1],[4],[6]. In the embryonic stage, neuroblasts are formed and produce neuron and glia mostly for larval CNS. At the end of embryogenesis, neuroblasts enter the quiescent state, except mushroom body (MB) neuroblasts, but are activated again in larval stage [5]. NBs should produce different types of neurons for neuronal diversity. For this purpose, temporal expression of Hunchback (Hb), Kruppel (Kr), Pdm and Castor (Cas) transcription factors provide production of different types of neuron or glia [7]–[9]. In larval and pupal stages, NBs divide hundreds of times to produce a variety of neurons and glia. Most of the adults' neurons and glia are produced at these stages.

1.2. Drosophila Larval Neuroblasts

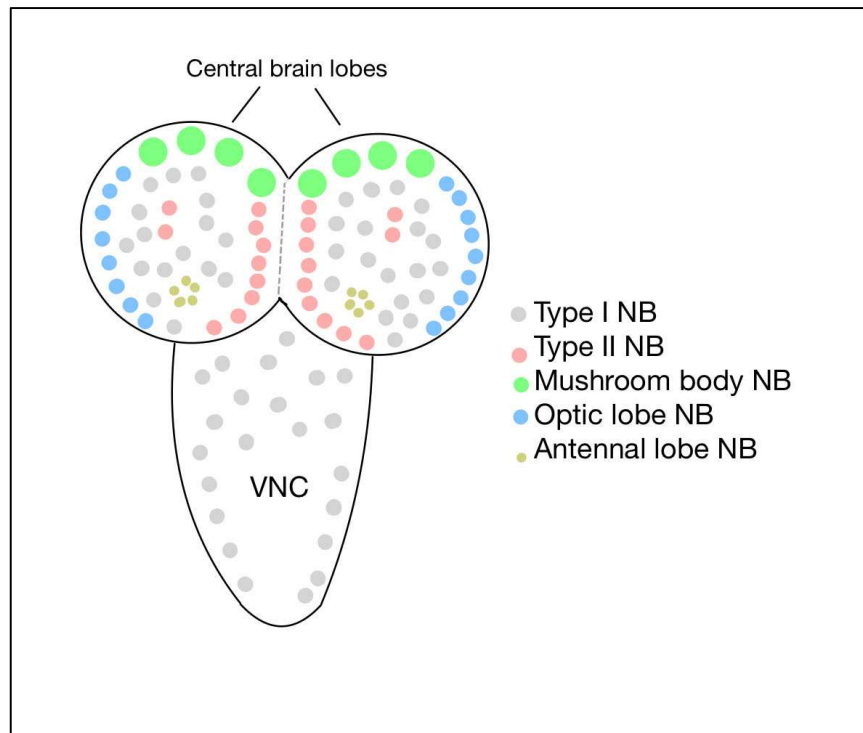


Figure 1.1: Drosophila larval brain anatomy. NB, neuroblast; VNC, ventral nerve cord.

The Drosophila larval brain contains two central brain lobes and ventral nerve cord (VNC). The larval brain of Drosophila is used to examine neuroblasts lineage formation because there are different types of neuroblasts and they can be distinguished from each other according to their location. In the Drosophila larva brain, there are 5 main neuroblast types as Type I NBs, Type II NBs, Mushroom Body (MB) NBs, Optic Lobe NBs and Antennal Lobe (AL) NBs (Figure 1.1, modified from [10]).

Type I NBs undergo asymmetric cell division and produce GMCs which give rise to neuron or mature neuron. Most of the brain NBs divide according to Type I manner [1],[11] (Figure 1.2 A). Type II NBs divide asymmetrically and produce intermediate neural progenitors (INPs) and which first divide symmetrically to increase their number and then divide asymmetrically to produce GMCs that generate neurons or glial cells [4],[12]. The fact that INPs divide symmetrically and increase their number allows Type II neuroblasts to produce more neurons than Type I neuroblasts [13] (Figure 1.2 B). Type II NBs lack the expression of a proneural gene,

asense, and thus called PAN (posterior *Asense* negative) neuroblasts. Type II NBs also show expression of the Ets transcription factor *Pointed* which is a master regulator of Type II NBs [11],[14].

Mushroom bodies are the olfactory learning and memory center of *Drosophila* [15],[16]. Mushroom body neuroblasts (MB NBs) produce Kenyon cells and can be distinguished from other NBs by the expression of *eyeless* (*ey*) gene [15],[17]. The MB NBs divide asymmetrically and produce two GMCs, one of which produces either a neuron or a glia with no sister fate diversification (Figure 1.2 C). Antennal lobe is the primary olfactory center that takes odor information from the antenna and transfers this to relevant areas in the brain [16],[18]. Antennal lobe neuroblast (AL NB) lineage needs the proper expression of the *empty spiracles* (*ems*) for the normal function [19]. AL NBs divide until pupation and there are binary sister fate decisions unlike MB NBs (Figure 1.2 D).

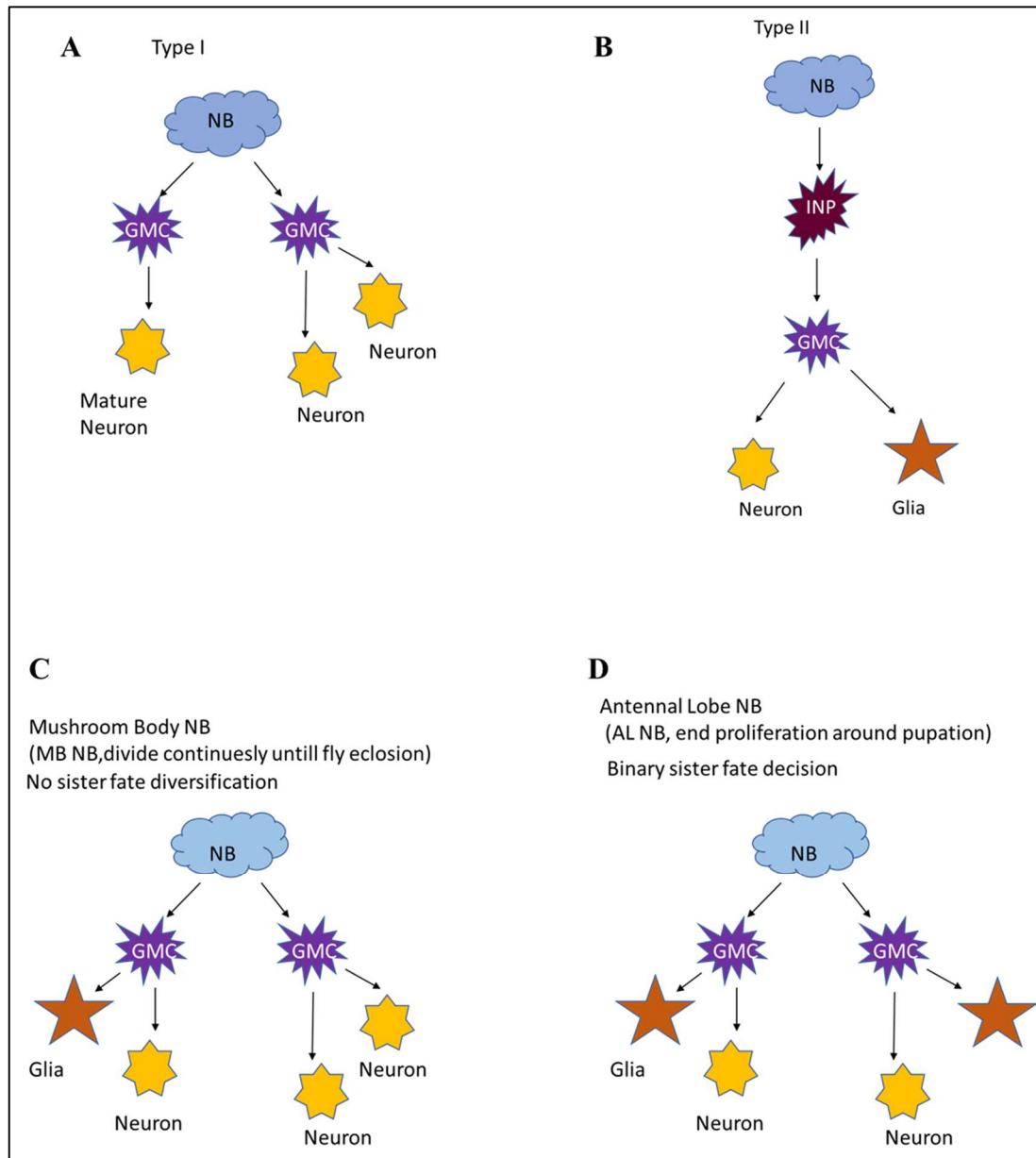


Figure 1.2: Neurogenesis patterns of *Drosophila* neuroblasts (NBs). A. Type I lineages, one progenitor produce ganglion mother cells (GMCs) which either divide once to produce two neurons or directly differentiate into a single mature neuron. B. Type II lineages, intermediate neural progenitors (INPs) are produced by asymmetric cell division and then GMCs produced from INPs to produce an independent series of offspring. C. The MB NBs yield only three major classes of MB neurons, with no sister fate diversification in the paired neurons made by one GMC. D. The AL NBs can generate ~40 neuron types from a single hemi lineage, as their GMCs make daughter cells with distinct A/B fates due to a Notch-mediated binary sister fate decision.

1.3. ETS Transcription Factor Family

E-twenty-six (ETS) transcription factor family is one of the largest transcription factor families. The ETS family is defined according to their DNA binding domains, which is called the ETS binding domain and it is about 85 amino acids [20]. The core motif of this ETS DNA binding domain is 5'-GGA (A / T)-3' and core DNA binding domain contains the helix-turn-helix motif [20],[21]. Although many other ETS domain proteins have been identified from various organisms according to similar items on ETS domain, this transcription factor family is unique to the metazoan lineage [22]. This family plays an important role in the nervous system development, apoptosis, proliferation and cell migration [23],[24].

This large family is divided into subfamilies according to their conserved domain similarities such as ETS, ERG, TEL, PEA3 (Figure 1.3, modified from Ekin Sonmez- personal communication). One of the conserved domains is the POINTED domain, which derives its name from the Pnt protein that was first identified in *Drosophila*. This domain has a role in homo oligomerization, heterodimerization and repression of transcription [22]. Ets65A is the member of the ERG subfamily and it is a RNA polymerase II specific transcription factor. It plays a role in cell differentiation and regulation of transcription [25]. Another RNA polymerase II specific ETS transcription factor is Eip74EF and it has a role in cell differentiation and autophagy [25],[26]. Aop plays role as a negative regulator of photoreceptor development and it is antagonist to Pnt [27]. Ets96B is found in dopaminergic neurons and plays role in dopamine signaling. It also plays roles in regulation of circadian cycle and cell differentiation [28].

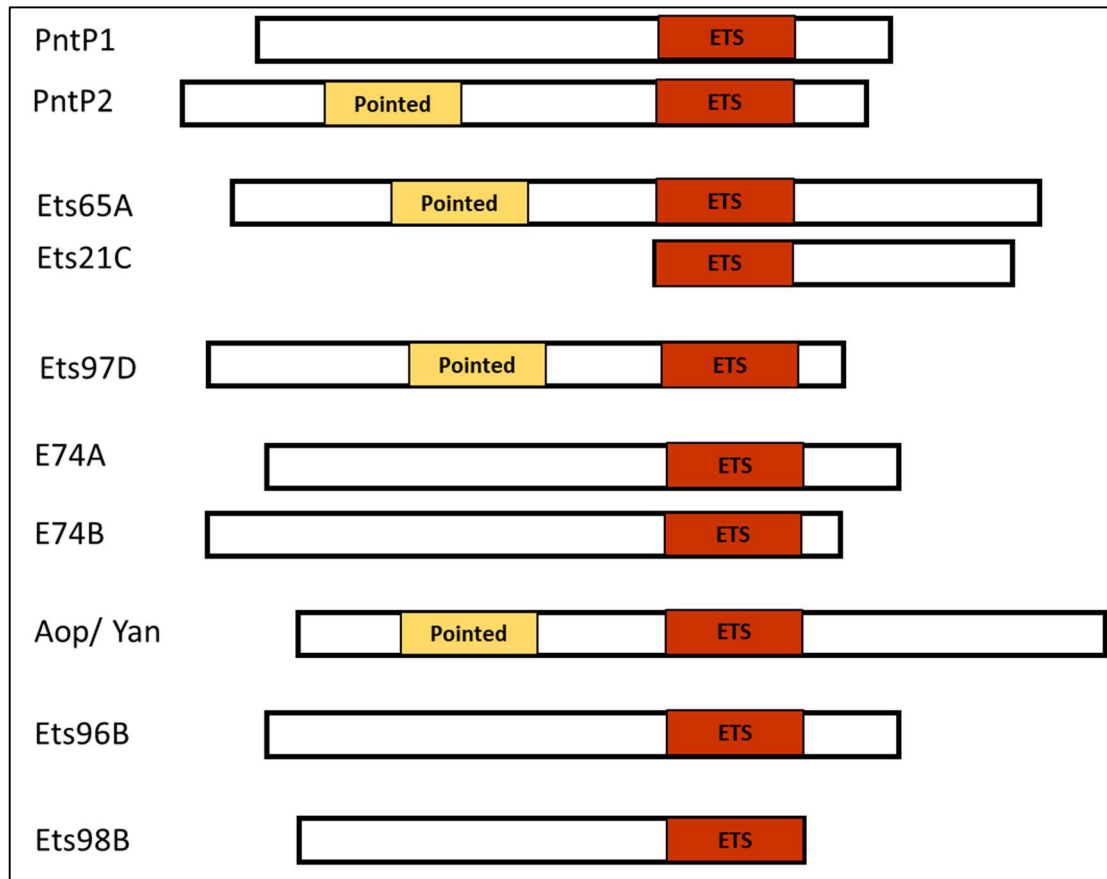


Figure 1.3: The subfamilies of ETS transcription factor family in *Drosophila melanogaster*.

Purpose of Thesis

Transcription factors play important role in the formation of different neuroblast lineage. The role of ETS transcription factors in the lineage formation is not fully known. However, Pnt, one of the ETS TFs member, has important role in Type II NB characteristic and also play role in glia formation in embryo and its antagonist Aop play role in cell fate determination. According to interfamilial relationship of ETS like on Pnt and Aop, the regulation of ETS transcription factors and the change of target genes have an effect on lineage formation. In this study, published RNAseq data which contain specifically labelled Type II NB, AL NB, MB NB, neuron, and glia data were analyzed to generate gene regulatory networks that shows the potential targets and the role of the ETS transcription factor in lineage formation.

2. MATERIALS and METHODS

2.1. Methods

RNA-seq datasets which are related with *Drosophila* lineage formation were obtained from Gene Expression Omnibus (GEO) platform and GSE71104 [16] dataset was selected for this study. In this dataset neuroblast lineages are labelled by using recombinase-dependent NB drivers such that all NBs are selected by chimeric *dpnEE* promoter and then each lineage is selected from all NBs by using exogenous transcriptional activator cassette.

2.1.1. GALAXY and Data Processing

GSE71104 dataset contains transcriptome data of mushroom body neuroblast, antennal lobe neuroblast, Type II neuroblast, non-selective neuroblasts (3 biological replicates for each neuroblast), neuron and glia (2 biological replicates for neuron and glia samples) which were labelled from *Drosophila* larval brain. The information related to samples in data is shown in Table 1. Preprocessing of the raw data was performed by using the GALAXY platform [29]. The reads in raw data were loaded to GALAXY platform by using Faster Download and Extract Reads in FASTQ format tool from NCBI the Short Read Archive (SRA).

Table 1.1: The sample information of GSE71104 dataset.

SRR number	Age	Cell type	Source
SRR2107300	50hr after larva hatch	neuroblast	Brain
SRR2107301	50hr after larva hatch	neuroblast	Brain
SRR2107302	50hr after larva hatch	neuroblast	Brain
SRR2107303	50hr after larva hatch	neuroblast	Antennal lobe
SRR2107304	50hr after larva hatch	neuroblast	Antennal lobe
SRR2107305	50hr after larva hatch	neuroblast	Antennal lobe
SRR2107306	wandering larva	glia	Brain
SRR2107307	wandering larva	glia	Brain
SRR2107308	50hr after larva hatch	neuroblast	Mushroom body_ 2hr wait
SRR2107309	50hr after larva hatch	neuroblast	Mushroom body_ 2hr wait
SRR2107310	50hr after larva hatch	neuroblast	Mushroom body_ 2hr wait
SRR2107311	50hr after larva hatch	neuroblast	Mushroom body_ 3 wash
SRR2107312	50hr after larva hatch	neuroblast	Mushroom body_ 3 wash
SRR2107313	50hr after larva hatch	neuroblast	Mushroom body_ 3 wash
SRR2107314	50hr after larva hatch	neuroblast	Mushroom body
SRR2107315	50hr after larva hatch	neuroblast	Mushroom body
SRR2107316	50hr after larva hatch	neuroblast	Mushroom body
SRR2107317	wandering larva	neuron	Brain
SRR2107318	wandering larva	neuron	Brain
SRR2107319	50hr after larva hatch	neuroblast	Brain
SRR2107320	50hr after larva hatch	neuroblast	Brain
SRR2107321	50hr after larva hatch	neuroblast	Brain

The reads may contain adapter sequences and there might be some low quality parts in sequences. These parts in reads should be eliminated for reliable further analysis. To check the quality of the reads FASTQC tool was used. The low quality reads and adapter sequences were eliminated by using TrimGalore! as cut-off 28. After elimination of low quality parts, quality control of reads had been repeated by using FASTQC tool to be sure all samples sequences have good quality. To find the location of the reads in reference organism's genome, alignment of the reads should be performed against reference genome. There are different tools to perform this processes such as TopHat2 and GSNAP but these programs use large memory especially when working large genome data and the process can take a long time. For these reasons, we have aligned the reads based on the reference genome, dmel_BGDP5, using the HISAT2 tool which works faster and requires less memory. BAM (binary alignment map) file was obtained as result of the HISAT2 tool and it was used as input in featureCounts tool which is used to count reads that matched the reference genome. As a result, count data of each samples was combined and gene

expression matrix was obtained. The flow chart of our methodology of study is shown in Figure 2.1 (Sonmez et.al, in preparation).

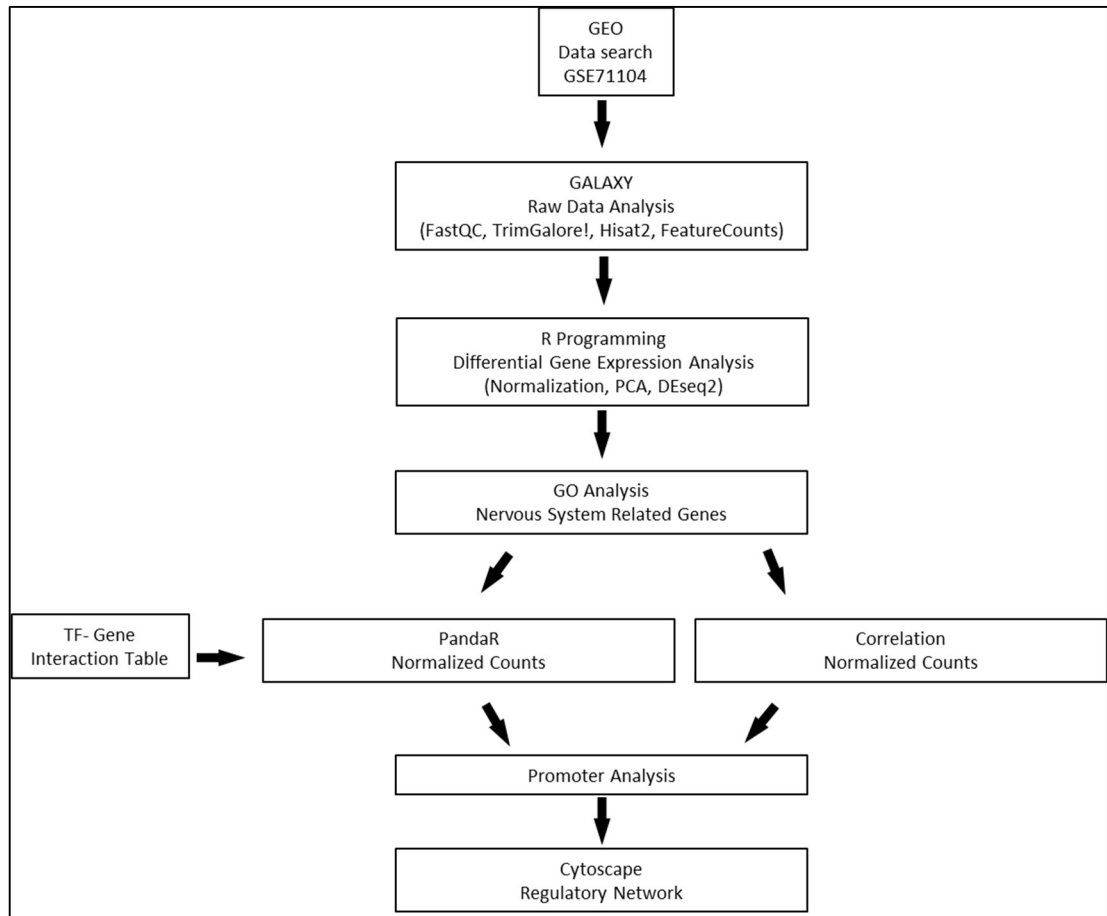


Figure 2.1: Methodology of transcriptome analysis of Drosophila larval dataset.

2.1.2. Differential Gene Expression Analysis

Downstream analysis of this data was performed by using R programming language. The count data matrix, which was obtained from GALAXY platform as a FeatureCounts was uploaded to R programming. The range of the raw counts is large so it should be transformed and fitted to a more eligible range for more consistent analysis. So, the normalization of the data was performed by using variance stabilizing transformation (vst) which assures that variance is constant in the range of mean values. After normalization of data, principal component analysis (PCA) was performed to check the distribution of samples and whether there were outlier samples. R package DEseq2 which used raw counts of samples was used for

differentially gene expression analysis for each neuroblast sample versus neuron and glia samples. Differentially expressed genes (DEG) were obtained by using Benjamini & Hochberg corrected p-value < 0.05 significance level (p-adjusted). The up regulated differentially expressed genes were determined by using Log2 fold change > 1 and down regulated DEGs were determined by Log2 fold change < -1 . The visualization of the differentially expressed genes in each analysis condition (such as glia compared to Type II NB, neuron compared to TypeII NB) was performed by using volcano plot in EnhancedVolcano R package [30].

2.1.3. Gene Set Enrichment (GSE) Analysis

The intersection of glia and neuron datasets for each neuroblast comparison was examined by using the venn package of R [31]. The regulation of the common genes that are obtained from intersection of glia and neuron datasets may play an important role in lineage formation, therefore these genes were used for further analysis. Gene set enrichment analysis is used to examine classes of over-represented genes in large groups, such as the biological processes they are involved in, or the signalling pathway and cellular components they are found in. Gene ontology (GO) enrichment analysis of selected common genes were performed by using clusterProfiler R package [32] to find the biological processes in which these common genes are involved. Nervous system related GO terms were filtered by using “neur”, “nervous”, “glia”, “glio” terms. The results graph was visualized by using the ggplot2 R package [33].

2.1.4. Gene Regulatory Network Construction

Gene regulatory networks show the interaction between transcription factors (TFs) and genes which have binding sites for that transcription factor. In gene regulatory networks, there are nodes and edges. Nodes are used to represent genes and transcription factors while edges are used to represent interaction between genes and transcription factors.

To examine the binding site of the transcription factor for the selected promoter sequences, a genome-wide promoter analysis for *Drosophila* was first performed

using the FindM tool [34]. For this purpose, the letter probability matrix of transcription factors taken from the CIS-BP database [35] was used as input. To prevent random and irrelevant binding probability of TF on sequences, sequences between -1000 to 100 TSS (transcription start site) and a p-value threshold of 0.001 for the binding motif filtration were used and promoter sequences were determined. The resulting output file was then converted into a readable TF-gene interaction table using the R program.

To create gene regulatory network of common genes and selected ETS transcription factors (TF)- Aop, Eip74EF, Ets96B, Ets65A, Pnt- in each neuron and glia dataset R package PANDA was used [36]. The PANDA algorithm uses correlation coefficients to create a connection, network, between transcription factor (TF) and gene. TF-gene interaction table and expression matrix were used in the PANDA algorithm. The correlation analysis was performed by using `corr()` function in R to detect the relationship between TF and genes and results were combined with the PANDA algorithm output for creating an edge-weighted gene regulatory network for selected transcription factors. The correlation values were filtered as absolute correlation > 0.75 . The resulting network table was uploaded to Cytoscape for visualization [37].

Then to create a more specific gene regulatory network, promoter analysis of nervous system related genes that were selected from common genes was performed. For this purpose, TxDb annotation package of R [38] was used for promoter sequences and -1000 to 100bp TSS was selected as binding region for transcription factors. For the motif of transcription factors, Aop, Eip74EF, Ets65A, Ets96B, Pnt, MotifDB package was used in R [39]. PANDA and promoter analysis results were used to create gene regulatory networks and these networks were visualized using Cytoscape.

3. RESULTS

Different types of neurons come together to form a brain. The production of this diversity of neurons is provided by neural stem cells (NSCs), which are called neuroblast in *Drosophila*. In this study, we focused on the role of the ETS transcription factors on the formation of neurons and glia from different neuroblast lineages as antennal lobe (AL) neuroblast, mushroom body (MB) neuroblast and Type II neuroblast.

3.1. Differential Gene Expression Analysis

In the first step of analysis, outlier samples and the distribution of samples in data should be checked and outlier samples should be eliminated for consistent analysis. For this purpose, all samples in the data (ie Type II NB, AL NB, MB NB, neuron and glia data) were analyzed using principal component analysis (PCA; Figure 3.1). In this particular dataset, there are 3 different types of mushroom body neuroblast samples because the researchers have tried different conditions such as 2 hours prolonged incubation time and adding 3 more washing steps in the cleaning step of experimental procedure. We chose not to include these additional conditions for further analysis because these conditions were shown to have no effect on the results when compared to normal mushroom body samples. Principal component analysis (PCA) showed that neuron and glia samples were clearly separated as distinct clusters from each other as well as neuroblast samples (Figure 3.1 A). However, the neuroblast samples clustered together in contrast to neuron and glia samples. This could be partly explained by overall similarity in transcriptional profiles of all undifferentiated neuroblasts as compared to differentiated mature cells.

When we repeated the PCA analysis with only neuroblast samples, it was observed that Type II NB, MB NB and AL NB were all separate from each other as distinct clusters, and there was no outlier group (Figure 3.1 B). All neuroblast samples (showed as blue dots and referred to as NB in Figure 3.1 B) which are a mixture of Type I NB and Type II NB, clustered together with Type II NB (purple dots) in PCA plot. After this PCA analysis, we determined that the sample data could be directly used for further analysis without extra filtering process. Next, the expression levels

of ETS genes and marker genes were examined with a heatmap using normalized counts (Figure 3.1 C). In the experimental step of this dataset neuroblasts of larvae were selected by using *dpn* promoter, which validates our methodology since *dpn* gene is expressed in all neuroblast samples.

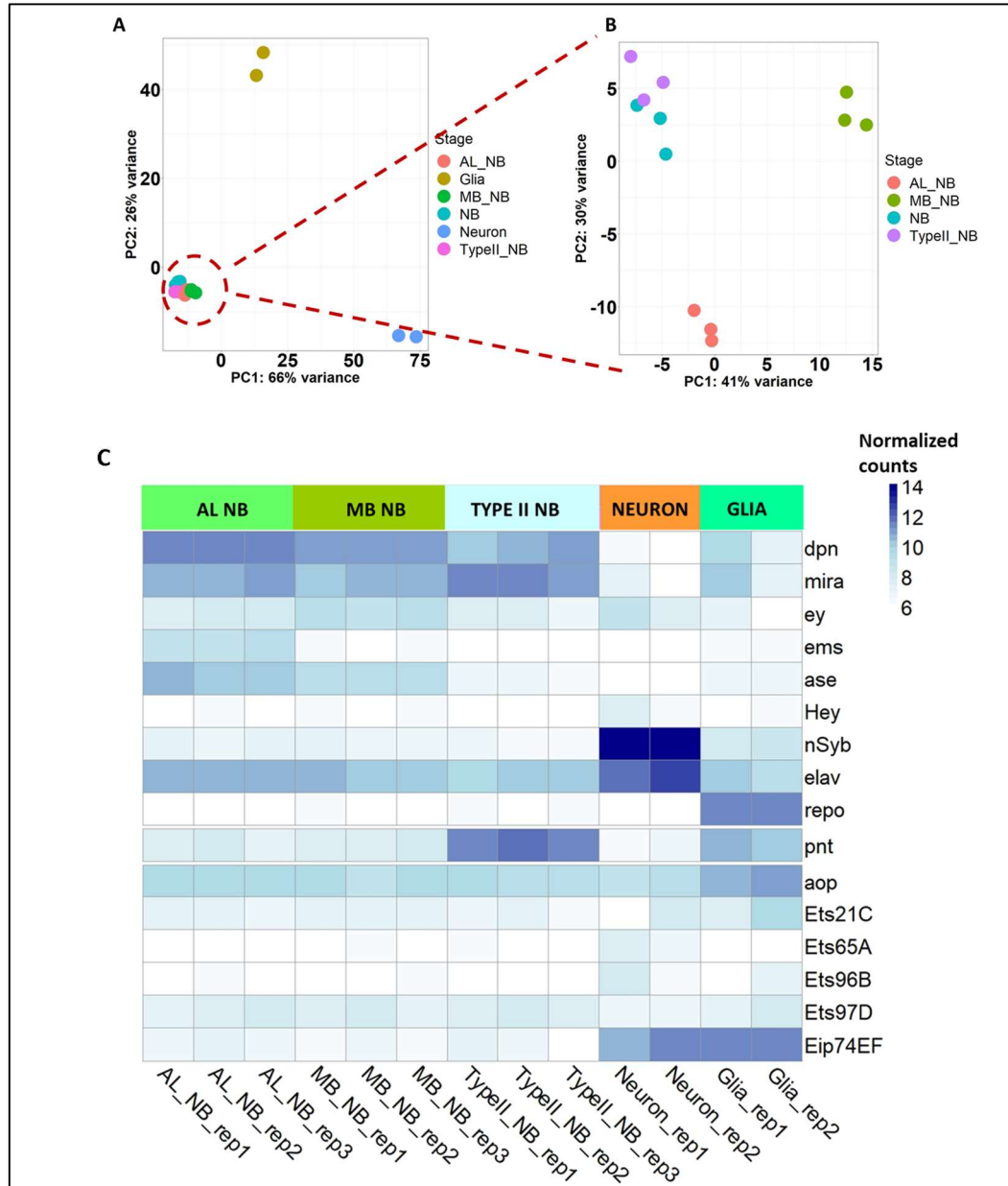


Figure 3.1: Transcriptome analysis of the Type II NB, AL NB, MB NB, neuron and glia samples. A. The distribution of the samples was shown in the PCA plot by using normalized gene expression. B. Each neuroblast type is distinguished from each other in the neuroblast cluster from (A) by using principal component analysis of normalized gene expression. C. The heatmap of expression of ETS genes and marker genes for all neuroblast, AL NB, MB NB, neuron and glia.

mira (miranda) is a cortical scaffold protein that is asymmetrically localized during neuroblast cell division, and it is used as a marker for Type I neuroblast [40]. The *eyeless* (*ey*) gene is a transcription factor that is mostly expressed in mushroom body (MB) NB and is commonly used as MB NB marker while another transcription factor, *empty spiracles* (*ems*) is expressed in antennal lobe (AL) NB and used as AL NB marker. A member of the *achate-scute* complex transcription factor family,

asense (ase) is a proneural gene and a Type I NB marker, and not surprisingly its expression was observed in both AL NB and MB NB. The gene Hairy/E(spl)-related with YRPW motif (Hey) is another key transcription factor that is used as a new-born neuron marker, while neuronal Synaptobrevin (nSyb), a SNAP receptor protein involved in vesicular fusion, is a mature neuron marker [41]. The embryonic lethal abnormal vision (elav) is a general neuron marker and is expressed in all developmental stages, while reverse polarity (repo) is a homeodomain transcription factor expressed in glia and is therefore used as glia marker [42].

Pointed (pnt) is a member of the Ets gene superfamily that plays a role in formation of glia and correlates with repo but it is also used as Type II NB markers.

Anterior open (aop) was found to exhibit high expression in glia although have expression in all samples. Ets21C is expressed in both glia and neuron but inconsistent in neuron. Ets65A is expressed in neuron but not glia while Ets96B is expressed in neuron and heterogeneous expression in glia. Ets97D is expressed almost all samples and Eip74EF is highly expressed in neuron and glia samples.

When we looked at the consistency of the ETS genes, we selected aop, Ets65A, Ets96B, Eip74EF and pnt for further analysis.

3.2. Glia and Neuron Lineage Formation from Type II NB

To examine role of the Ets genes in lineage formation from Type II NB, differentially expressed genes for both glia compared Type II NB and neuron compared Type II NB were analyzed. Firstly, the distribution of the samples was performed by PCA analysis and result showed that samples were discriminated each other (Figure 3.2 A). 2326 genes were upregulated while 1831 genes were downregulated, when neuron was compared to Type II NB (Figure 3.2 B). 2454 genes were upregulated and 1566 genes were downregulated while glia formation from Type II NB (Figure 3.2 C). When the expression levels of Ets genes were examined in neuron and glia comparison data, it was observed that the expressions of aop, Eip74EF, Ets96B and pnt changed significantly (Figure 3.2 D).

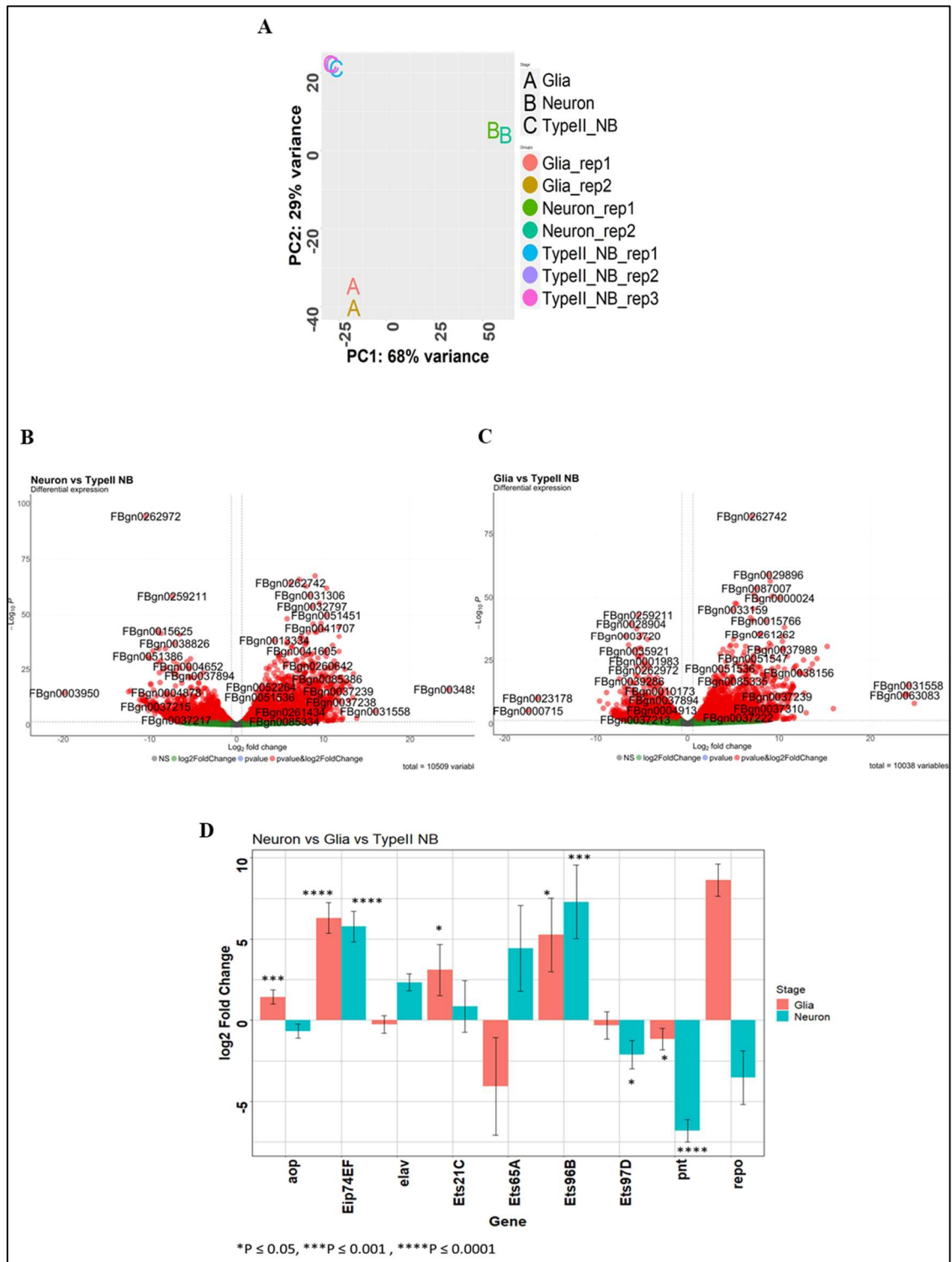


Figure 3.2: Glia and neuron lineage formation from Type II NB. A. PCA analysis of the samples' normalized gene expression was performed to examine the distribution of samples. B. Up and down regulated genes was shown in volcano plot in neuronal lineage formation from Type II NB. C. In glial lineage formation, up and down regulated genes shown in volcano plot. D. The expression levels of the ETS genes in both neuron and glial lineage formation from Type II NB data.

To examine the regulation between neuronal and glial lineage formation, we compared these two datasets and we obtained 1504 genes were found common in both dataset while 1531 genes found only in glia and 1712 genes found only in neuron datasets (Figure 3.3 A). Gene Ontology (GO) analysis revealed that common genes are related to both neurogenesis and gliogenesis related process while only glia genes play role in just gliogenesis in nervous system (Figure 3.3 B). The common genes play role both in neuronal and glial processes so we also examined that how ETS transcriptional factors regulate these common genes and lead them to the formation of neuron or glia lineage. For this purpose, we obtained gene regulatory network from common genes (Figure 3.4). In glia data, Aop, Eip74EF, and Ets96B upregulated while Ets65A and Pnt downregulated. We observed that while Aop and Ets96B TFs regulate a gene cluster, Eip74EF and Pnt TFs regulate another gene cluster together. A large group of genes is also regulated by 5 ETS transcription factors, while Ets65A is included in this group by regulating a small number of genes (Figure 3.4 A). In neuron data, Pnt and Aop downregulated while other ETS genes upregulated. We observed that Pnt and Eip74EF regulate a smaller gene cluster together, except for the gene cluster in which all transcription factors play a role in regulation. We also found that, unlike the glia data, Ets65A regulates more genes and this might be because Ets65A is upregulated in neuron data (Figure 3.4 B). To be more specific, we selected nervous system related genes from common genes and performed promoter analysis to detect that those genes have binding region for our ETS transcription factors and created gene regulatory network (Figure 3.5-3.6).

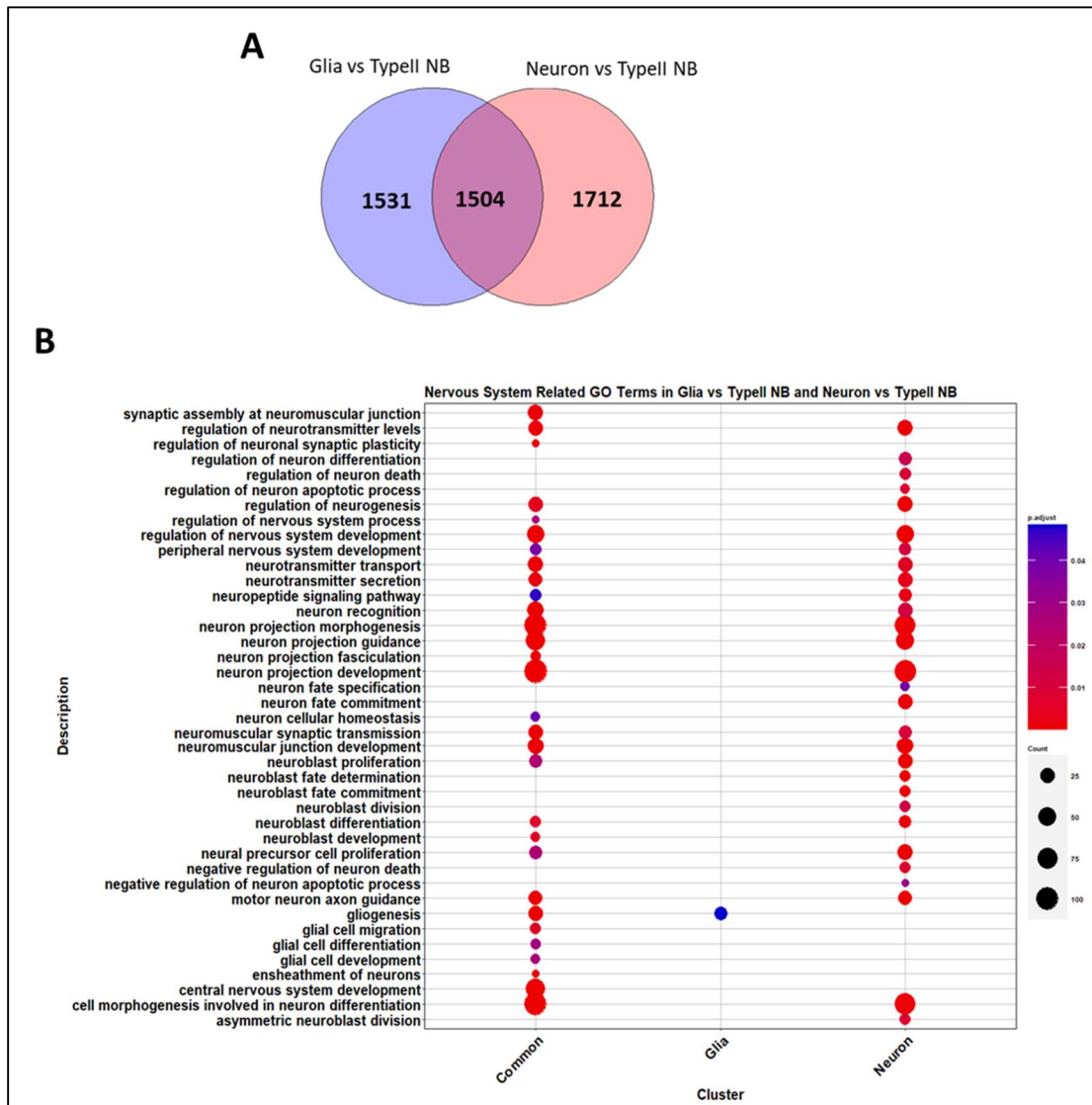


Figure 3.3: Glia and Neuron Lineage Formation from Type II NB. A. The number of differentially expressed genes in glia, neuron and intersection of them when compare to Type II neuroblast (p-adjusted < 0.05). B. Nervous system related GO terms in neuron, glia and intersection of both when neuron and glia originated from Type II NB.

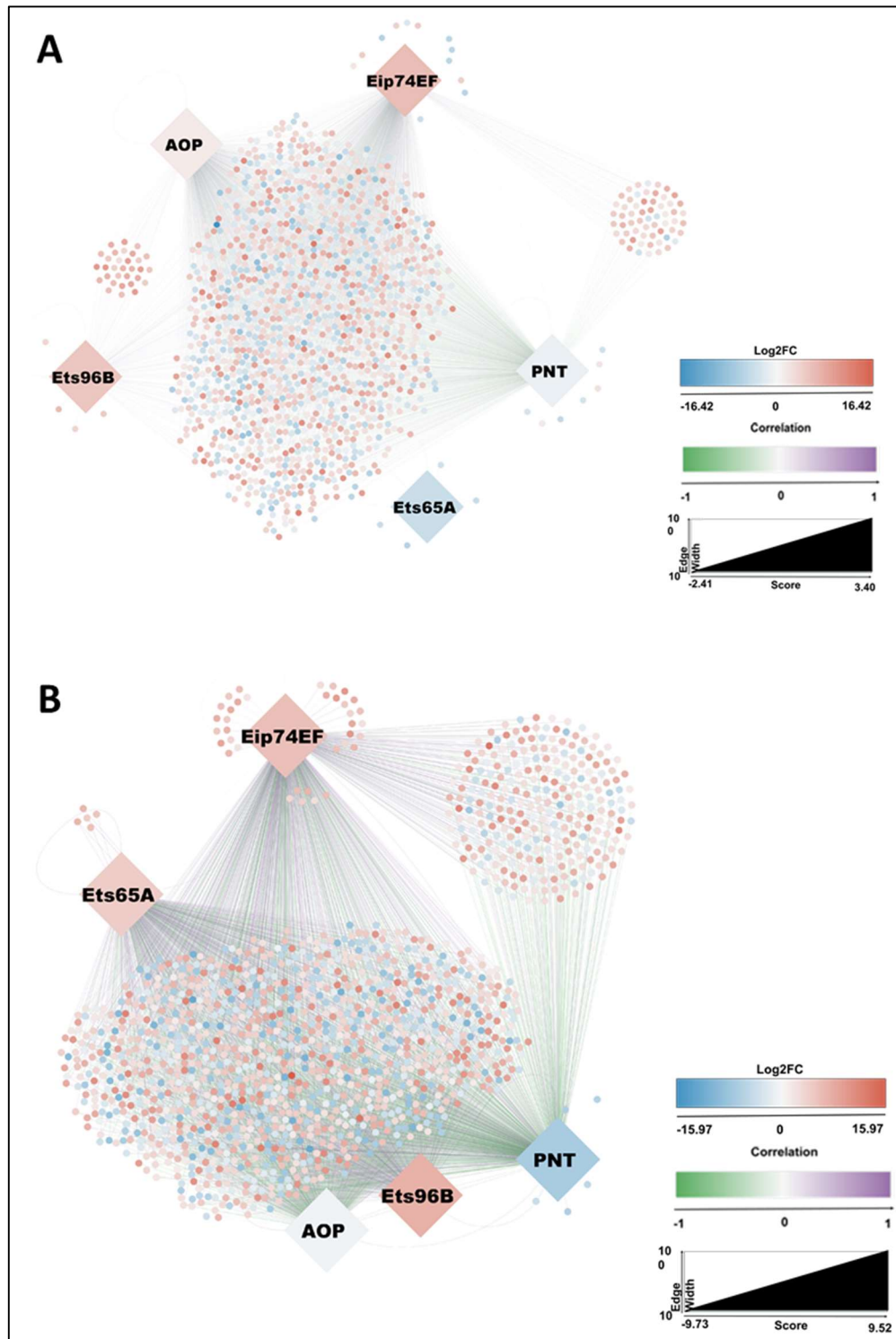


Figure 3.4: The gene regulatory network of glia and neuron lineage formation from Type II NB A. The gene regulatory network of common genes according to Glia vs Type II NB dataset. B. The gene regulatory network of common genes according to Neuron vs Type II NB dataset.

3.2.1. Gene Regulatory Network in Glia Lineage from Type II NB

In glia compared to Type II NB data, we found that only four TFs – Pnt, Aop, Eip74EF and Ets96B- regulate nervous system related genes and there is no potential interaction with other TFs for Ets96B (Figure 3.5). Among these TFs, Aop and Ets96B, especially Ets96B, regulated genes most of the genes alone while Pnt and Eip74EF regulated a gene cluster together. Pnt has 15 targets and 5 of them controlled by only Pnt. Pnt is downregulated in glial formation from Type II NB and its 4 targets also downregulated and positively correlated with Pnt while 11 of the targets are negatively correlated with Pnt as upregulated. Specific Pnt targets are mostly related with motor neuron axon guidance such as Ptp99A and side genes [43]–[45]. Although they play role in same process, Ptp99A is upregulated while side is downregulated. Tbh and spas genes are also specific targets of Pnt and these genes found in both glia compared to Type II NB and neuron compare to Type II NB data. Tbh (Tyramine β hydroxylase) is play role in olfactory learning and memory while spas have role in axon transport and severing of microtubule. They are up regulated in both dataset but only regulator of the Tbh changes and is controlled in the neuron data by Eip74EF. Gad1 is upregulated and play role in production of the nervous system specific glutamic acid decarboxylase which convert to glutamate to inhibitory neurotransmitter GABA.

Aop have 12 specific targets and only 3 of them is downregulated while others upregulated. The downregulated genes, Dbx take part in neuron differentiation and mostly found in GABAergic neurons and GlyRS (Glycyl-tRNA synthetase) play role in dendrite formation. These two genes also found in neuron formation data but they under the control of different transcription factors. The most of the upregulated targets of the Aop related with neurotransmitter secretion, neuromuscular synaptic transmission processes. One of them is Plc21C (Phospholipase C) which codes the phospholipase C and this gene is necessary for the regulation of neuron cellular homeostasis. slo (slowpoke) is another upregulated gene which under the control of Aop in glia data, this gene provide the production of the α subunit of calcium-activated potassium channel. Therefore, this gene has important role in communication between neurons because it controls the release of neurotransmitter at the synapses. This gene is also found in neuron formation data as upregulated but

regulated by Ets96B. When we looked at the genes that Aop and Pnt regulate together, there are only 2 genes, Syt7 and wcd, and they are both downregulated so they positively correlated with Pnt while negatively correlated with Aop. Syt7 is another gene that regulate neurotransmitter release that provide synthesis of calcium-dependent phospholipid binding proteins. wcd play role in neuroblast proliferation and stem cell maintenance. It does this process by regulating the size of the cell through ribosome synthesis. This gene also found in neuron formation data as downregulated and it is under the control of same transcription factor, Aop and Pnt.

Eip74EF has 6 specific target and only one Snap25 is upregulated and positively correlated. Snap25 (Synaptosomal- associated protein 25kDa) provide synthesis of a part of SNARE protein complexes which is important to recognition of the neurotransmitters by correct membrane receptors. The downregulated genes Pdf (pigment-dispersing factor) provides the production of a neuropeptide that triggers intracellular signaling. Cdk5alpha is another downregulated genes and by making a complex with Cdk5, it play important role in the regulation of synaptic homeostasis and axon pattern in post-mitotic neurons. ths (thisbe) gene code Fibroblast growth factor (FGF) and it can act as chemoattractant to provide cell migration and play role in glial cell development. Syx6 is the gene that regulated by both Eip74EF and Aop transcription factors. It belongs to SNARE proteins family like Snap25 and required for binding activity of SNARE proteins. It also have role in retrograde transport and synaptic vesicle docking. When we looked the genes that controlled both Pnt and Eip74EF, there are 8 genes and only the gene produce protein for neuronal development Cbp53E is downregulated, others are upregulated. Pvf1, Timp and dap genes found in neuron data but they are regulated by different transcription factors and dap downregulated in neuron data while others expression patterns did not change. dap gene inhibits activity of protein kinase and when dap is upregulated cells arrest in G1/G0 phase and post-mitotic cells enter the differentiation stage. Pvf1 produce the PDGF- and VEGF- related factors and regulated cell migration by binding of this factor to the receptor tyrosine kinase and activated it. Timp is another upregulated genes that regulated by Pnt and Eip74EF and have role in extracellular matrix organization.

Ets96B regulated 13 genes and all of them are the specific targets of it. 5 of them is downregulated while 8 genes upregulated. Rib and Cirl genes found in neuron data and upregulated also but rib is regulated by Eip74EF while Cirl is regulated by

Ets65A in neuron data. Cirl is G-protein coupled receptor and play role in tactile and auditory stimuli perception. Downregulated genes mainly related with neuroblast development. klu is one of the downregulated genes and play role in characterization of the secondary precursor cells in neuroblast lineages. Pon (partner of numb) is other downregulated genes and worked with numb which is important for asymmetric cell division of neuroblast.

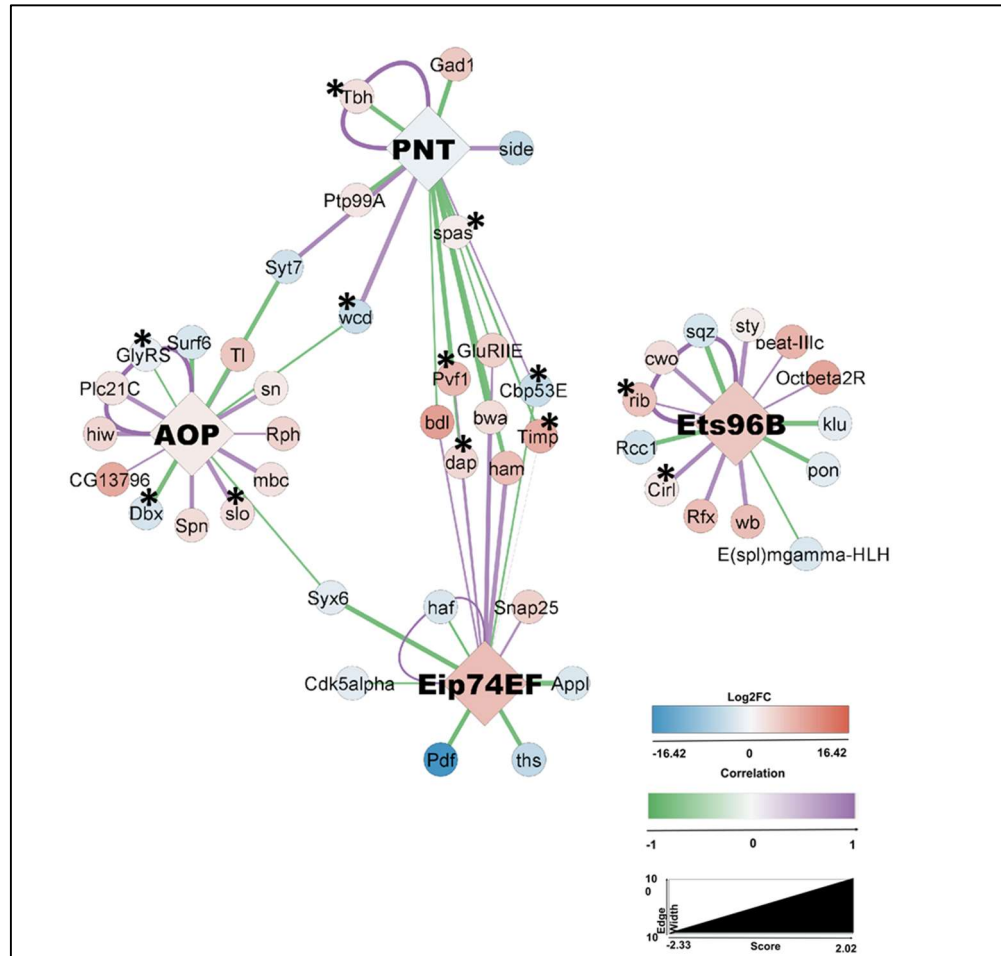


Figure 3.5: The gene regulatory network which was obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using glia vs TypeII NB dataset. (* shows the genes that are found in both glia and neuron datasets.)

3.2.2. Gene Regulatory Network in Neuron Lineage from Type II NB

In the network which obtain from neuron compared to Type II NB data, five TFs play role in the regulation of the nervous system related genes differ from the glia compare to Type II NB data (Figure 3.6).

Pnt is downregulated similar to glia data and has 14 unique target. The only 4 of the targets positively correlated with Pnt. When we compare to glia data, Pnt regulate more genes in neuronal lineage formation. Dbx which take part in neuron differentiation is one of the specific targets of the Pnt and it is upregulated. This gene is regulated by the Aop and downregulated in glia dataset. The changing of the transcription factors which regulate this gene effect the changing the expression level and may crucial role in formation of the neuron. nvy is transcriptional repressor and it is upregulated gene that take part in axon guidance process.

While Aop is upregulated in glia data, it is downregulated in neuron formation data. Although the number of specific targets is same in both data, Aop regulates different genes in neuron data. The fact that Ets65A started to regulate genes by being included in the neuronal lineage formation network may have caused to change the expression of Aop. Aop and Ets65A regulate together the gene which play role in recognition of the synaptic targets, CG5758. Although Syx1A and Syx18 is member of the SNARE proteins, they have different expression profile. Syx1A which play role in neurotransmitter release and neuronal differentiation is upregulated while Syx18 gene which provide the coding of neuromuscular synaptic transmission receptor is downregulated. The gene that have role in axon transport, spas, is unique target of the Aop and upregulate whereas this gene under the control of the Pnt in glial data with similar expression profile. Aop and Pnt regulates a gene, wcd, together in both glia and neuron data and this genes that have role in neuroblast proliferation downregulated in both data.

In neuron lineage formation network, the specific targets of Eip74EF is increase, it regulates 13 genes. 5 of the genes downregulated while 8 genes upregulated and expression profile of the Eip74EF is similar to glia data. There are 4 unique target which found in glia formation also. Although expression profile rib did not change, it is the unique target of Ets96B in glia data. The Pvf1 gene involved in cell migration is regulated by Pnt and Eip74EF in glia data, while here it is the specific

target of Eip74EF and upregulated in both data. The another gene that found in both data is GlyRS, it is the specific target of the Aop in glia data. Although this gene regulated by different transcription factors, it is downregulated in both network. The Tbh, which is involved in learning and memory, is upregulated in both data sets, but Pnt regulates specifically this gene in the glia data. Ser (serrate) which is involved in wing development, is upregulated by Eip74EF and Aop.

Ets65A is a transcription factor that involved in neuron lineage formation. It is upregulated and have 11 specific targets. Ets65A have mainly negative correlation with its unique targets. galectin and aPKC are the positively correlated genes with Ets65A and galectin take part recognition of synaptic targets while aPKC is have role in self renewal of neuroblast and neuroblast proliferation. There are 2 genes, Cirl and dap, that found in both glia and neuron data as well as specific target of Ets65A in neuron data. Cirl is regulated by Ets96B and is upregulated in glia network while dap is regulated by Pnt and Eip74EF together in glia network. The unique target of Ets65A, dap is downregulated in neuron while it is upregulating in glia data. The downregulated genes mostly related with neuron development processes. For example mas gene play role in axon development and neuron projection fasciculation while btd take part in neuron differentiation. FER is another downregulated genes which is tyrosine kinase and have role in regulation of cell adhesion and adherent junction so it involved in axon guidance. Invadolysin, which is involved in mitosis and cell migration, is regulated by Eip74EF and Ets65A and has a positive correlation with both.

Ets96B regulate 13 genes specifically and there are no collaboration with other ETS TFs like in glia network. Ets96B is upregulated in here also, and mostly have positive correlation with its targets. Only spin and loco is downregulated genes which both are related with glial cell differentiation process. loco is also important for asymmetric neuroblast division. There are 3 upregulated genes which found also in glia network. Timp and Cbp53E is regulated by Pnt and Eip74EF in glia data but Cbp53E which take part in regulation of neuron projection development is downregulated in glia network while Timp is upregulated in both network. Another gene which found in glia network is slo, and it is upregulated in both network whereas it is specific target of the Aop in glia network. The most of the upregulated targets are related with learning and memory processes such as Galphaq, Alk and Rdl. Cals is one of the upregulated genes have role in synaptic transmission regulation.

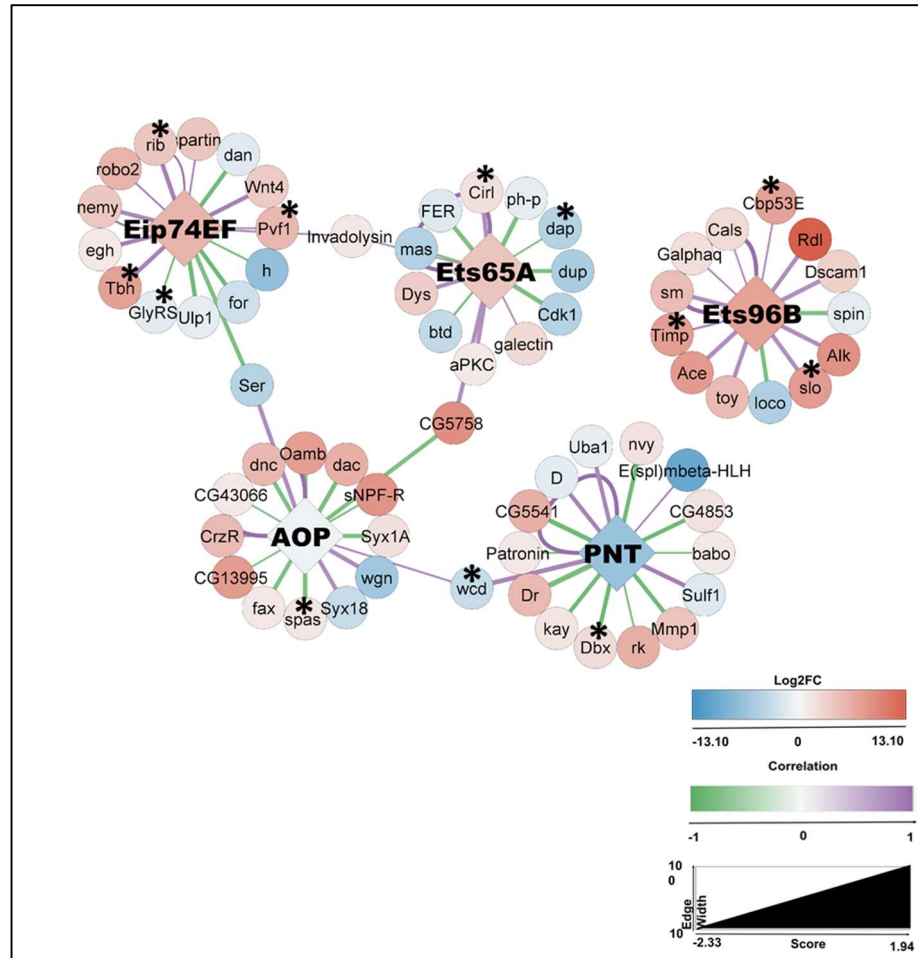


Figure 3.6: The gene regulatory network which was obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using neuron vs TypeII NB dataset. (* shows the genes that are found in both glia and neuron datasets.)

3.3. Glia and Neuron Lineage Formation from Antennal Lobe (AL) NB

In the lineage formation from Antennal lobe (AL) NB, differentially expressed genes for both glia compared AL NB and neuron compared AL NB were analyzed. Firstly, the distribution of the samples performed by PCA analysis and result showed that samples were discriminated each other (Figure 3.7).

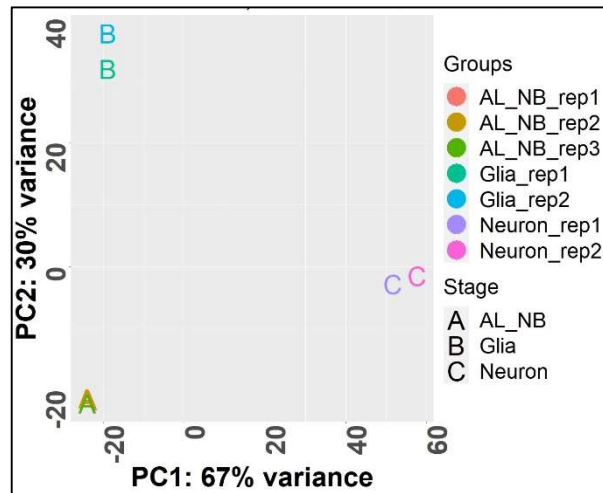


Figure 3.7: PCA analysis of the samples' normalized gene expression of glia and neuron lineage formation from AL NB.

2119 genes upregulated while 1939 genes downregulated, when neuron compared to AL NB (Figure 3.8). 2287 genes upregulated and 1669 genes downregulated while glia formation from AL NB (Figure 3.9). The expression of aop, Eip74EF and pnt genes significantly changes in both glia compare to AL NB and neuron compare to AL NB data (Figure 3.10).

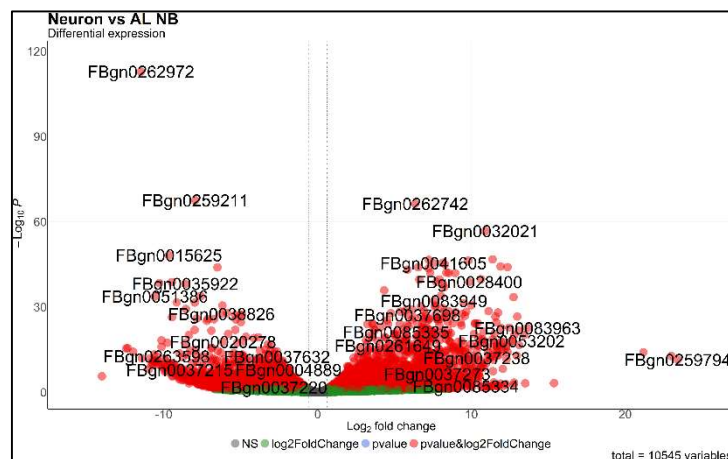


Figure 3.8: Up and down regulated genes was shown in volcano plot in neuronal lineage formation from AL NB.

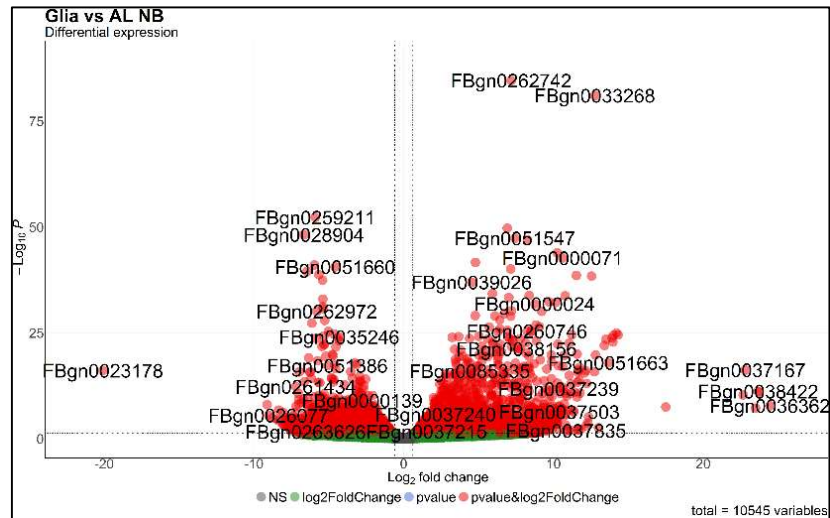


Figure 3.9: In glial lineage formation from AL NB, up and down regulated genes shown in volcano plot.

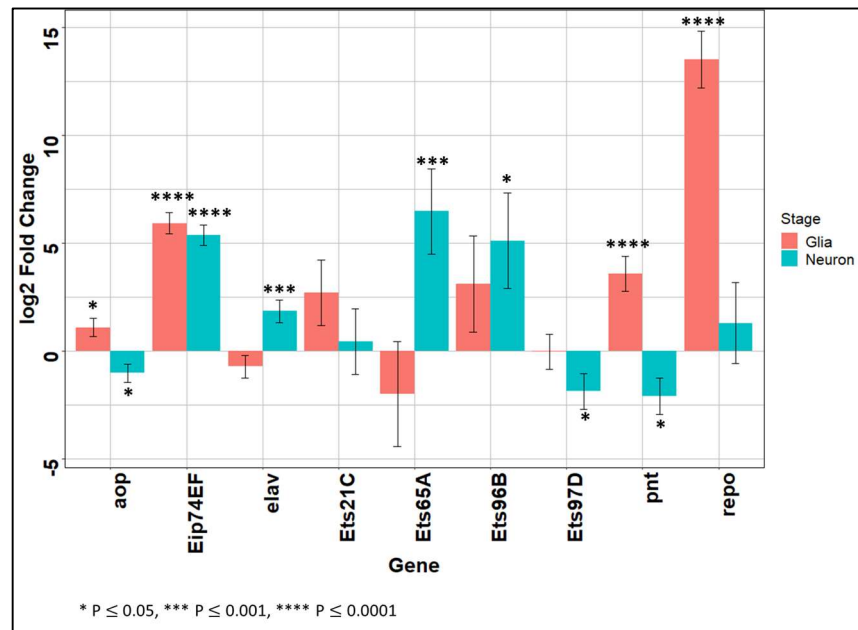


Figure 3.10: The expression levels of the ETS genes in both neuron and glial lineage formation from AL NB.

The comparison of the glia and neuron data give us 1352 genes were found in both dataset while 1539 genes found only in glia and 1721 genes found only in neuron datasets (Figure 3.11 A). Gene Ontology (GO) analysis revealed that common genes are related both neurogenesis and gliogenesis related process similar to lineage formation from Type II NB, however only glia genes play role in neuron remodelling in addition to gliogenesis (Figure 3.11 B).

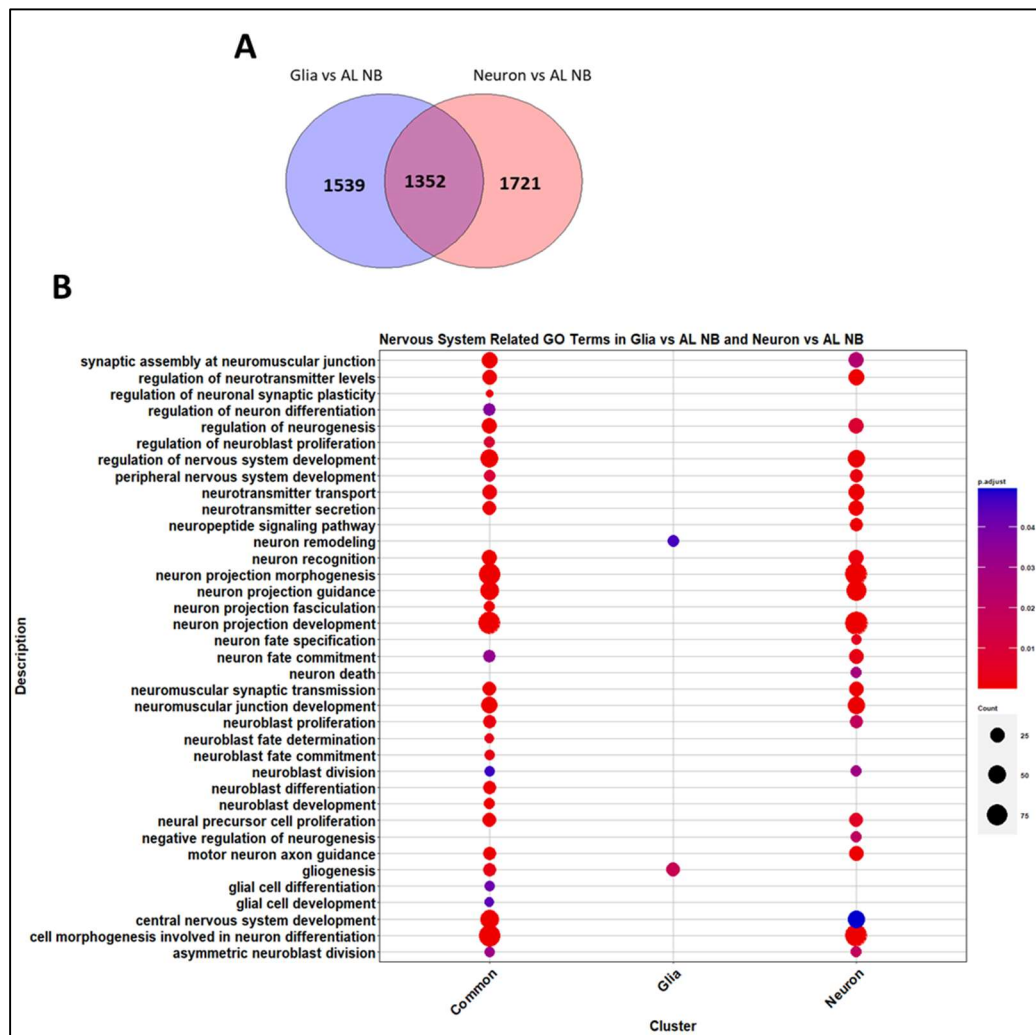


Figure 3.11: Glia and Neuron Lineage Formation from AL NB. A. The number of differentially expressed genes in glia, neuron and intersection of them when compare to AL neuroblast (p-adjusted < 0.05). B. Nervous system related GO terms in neuron, glia and intersection of both when neuron and glia originated from AL NB.

To examine the role of the ETS transcription factors, gene regulatory network from the common genes which contain genes that play role both neuronal and glial processes were created (Figure 3.12, 3.13). When we looked the glia compare to AL NB data, Aop, Eip74EF, Ets96B and Pnt upregulated and only Ets65A downregulated. All 5 transcription factors regulates a most of the genes together (Figure 3.12). However, Pnt and Aop downregulated while Eip74EF, Ets65A and Ets96B downregulated in neuron compare to AL NB data. Like in glia data, all TFs regulate most of the genes together and there is no significant cluster specific to TFs (Figure 3.13).

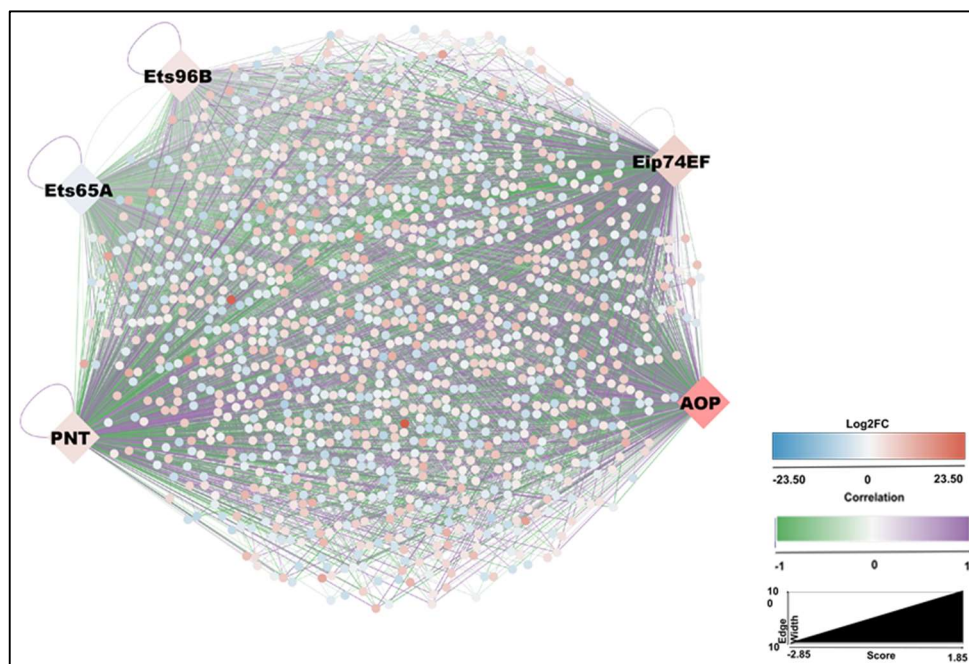


Figure 3.12: The gene regulatory network of common genes according to Glia vs AL NB dataset.

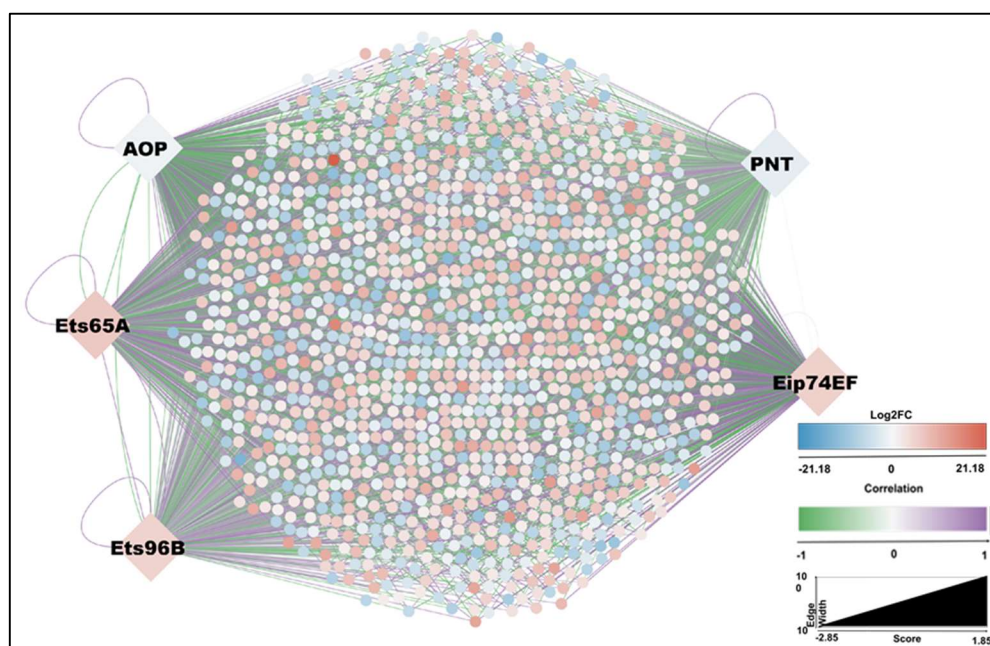


Figure 3.13: The gene regulatory network of common genes according to neuron compared to AL NB dataset.

To distinguish specific genes which are regulated by selected ETS TFs, we selected nervous system related genes from common genes and performed promoter analysis to detect that those genes have binding regions for our ETS transcription

factors and created a gene regulatory network (Figure 3.14, 3.15). In both glia compared to AL NB network (Figure 3.14) and neuron compared to AL NB network (Figure 3.15), same 15 genes are regulated by the same TFs.

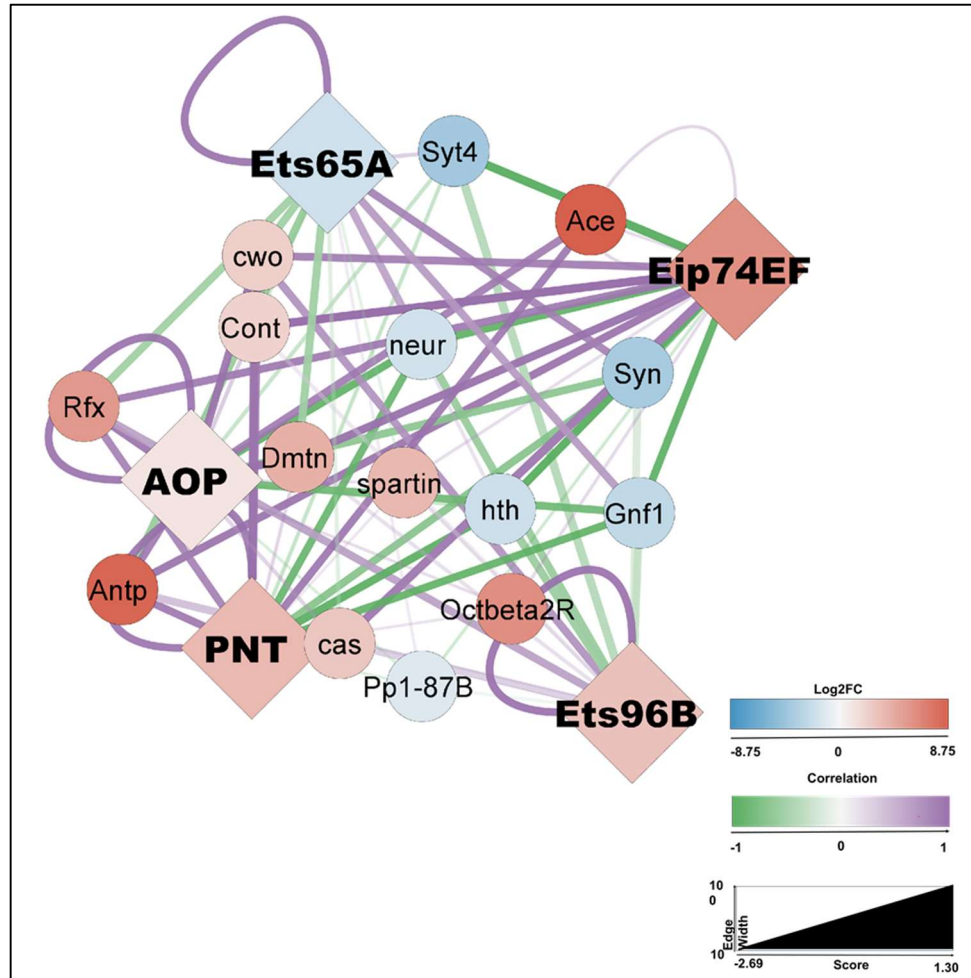


Figure 3.14: The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes by using glia vs AL NB dataset.

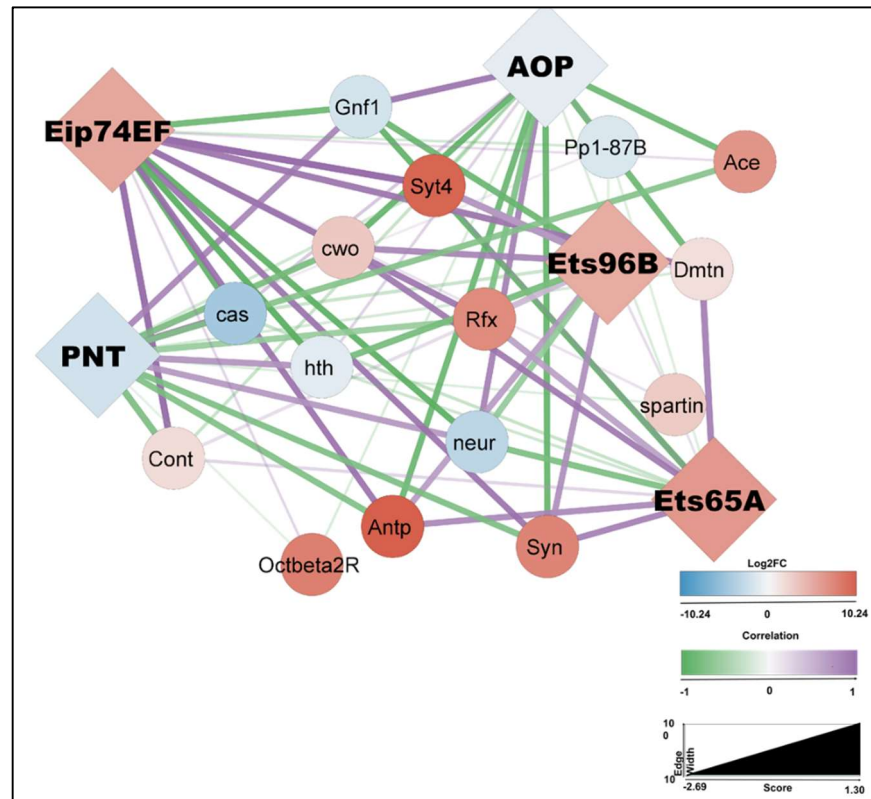


Figure 3.15: The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes of neuron compared to AL NB.

Among these genes, only Syt4, Syn and cas genes expression were changed. Syt4 gene has role in regulation of synaptic plasticity and it is upregulated in neuron network while downregulated in glia formation [46]. Syn gene is also upregulated in neuron formation while downregulated in glia formation and this gene play role in olfactory learning which is quite associated with antennal lobe characteristic [47]. The other gene, cas, plays a role in glial cell development so it is upregulated in the glia network while downregulated in the neuron network [48].

3.4. Glia and Neuron Lineage Formation from Mushroom Body (MB) NB

To examine the lineage formation from mushroom body neuroblast, differentially expression gene analysis was performed for both glia compared MB NB and neuron compared MB NB. Firstly, the distribution of the samples performed by PCA analysis and result showed that there are no outliers and samples were discriminated each other (Figure 3.16). 2189 genes upregulated while 2046 genes

downregulated, when neuron compared to MB NB (Figure 3.17). 2241 genes upregulated and 1789 genes downregulated while glia formation from MB NB (Figure 3.18). The expression of Eip74EF and pnt genes significantly changes in both glia compare to MB NB and neuron compare to MB NB data (Figure 3.19).

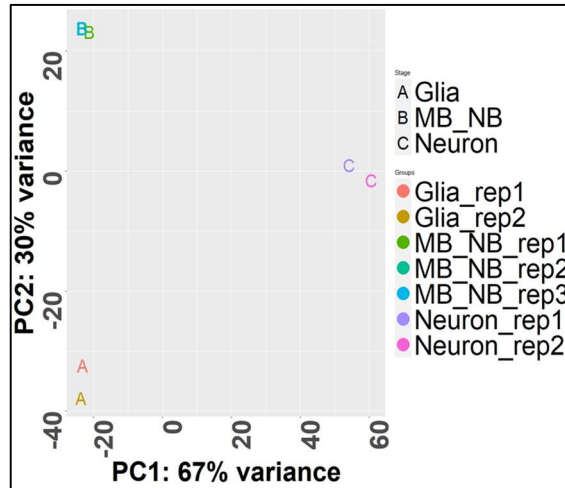


Figure 3.16: PCA analysis of the samples' normalized gene expression of glia and neuron lineage formation from MB NB.

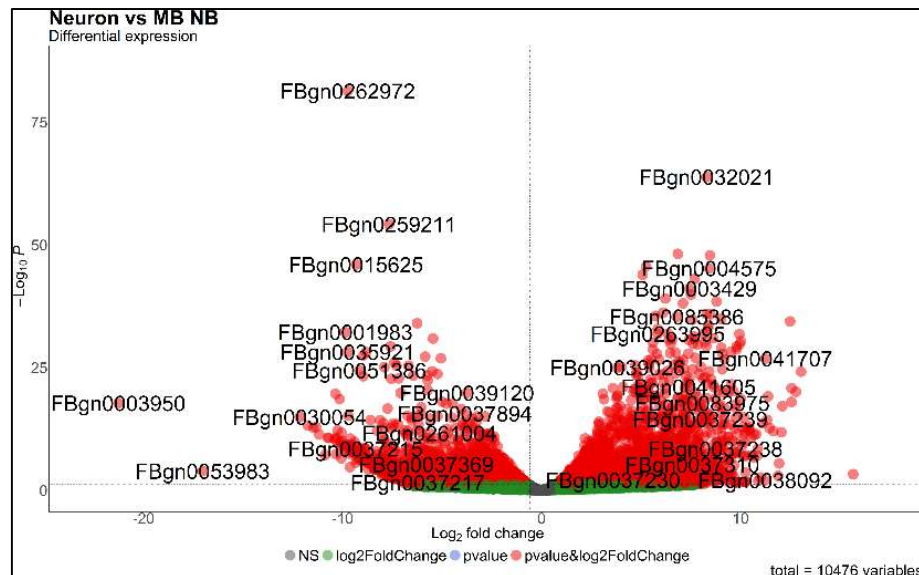


Figure 3.17: Up and down regulated genes was shown in volcano plot in neuronal lineage formation from MB NB.

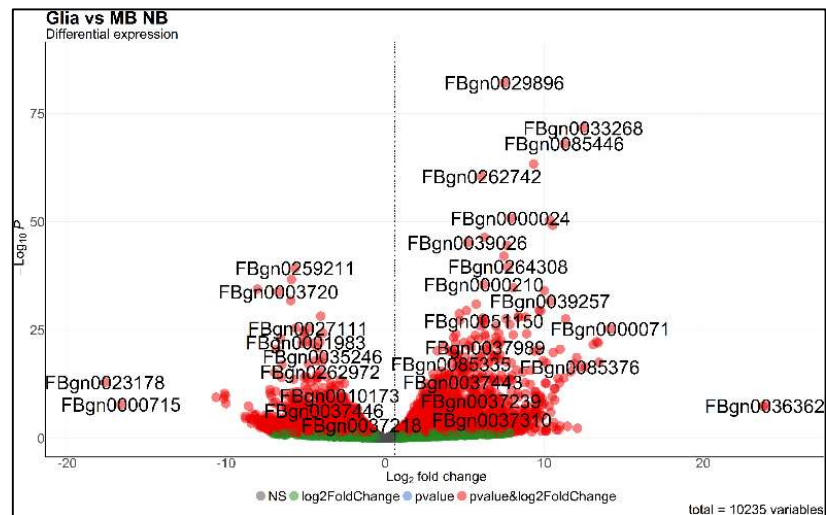


Figure 3.18: In glial lineage formation from MB NB, up and down regulated genes shown in volcano plot.

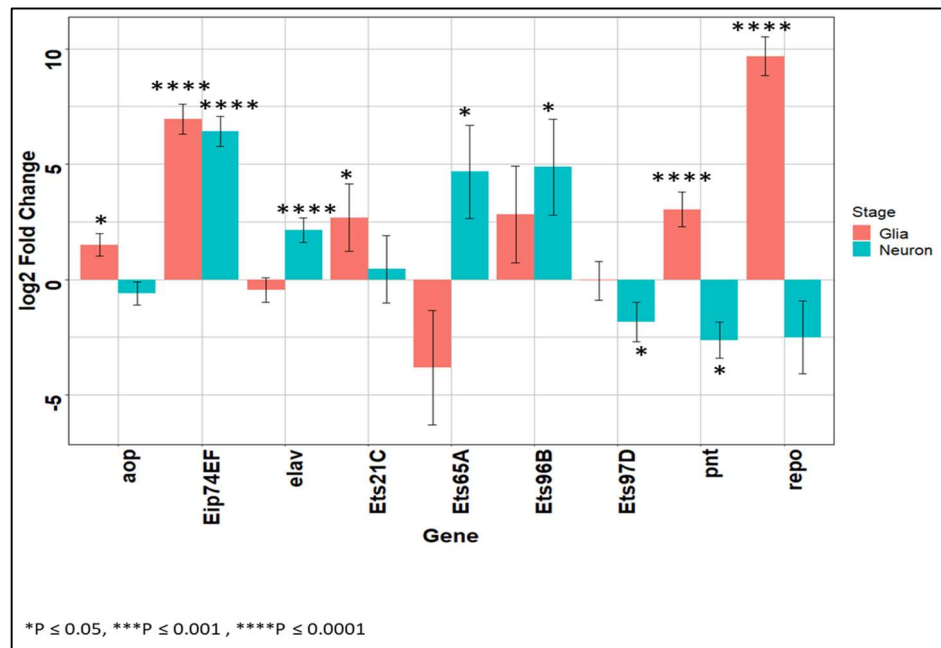


Figure 3.19: The expression levels of the ETS genes in both neuron and glial lineage formation from MB NB.

When we compared the glia and neuron data, 1488 genes were found in both dataset while 1495 genes found only in glia and 1746 genes found only in neuron datasets (Figure 10A). Gene Ontology (GO) analysis revealed that common genes are related both neurogenesis and gliogenesis related process similar to lineage formation from Type II NB and AL NB, while only neuron genes related with neurogenesis and there are no enriched nervous system related process for only glia genes (Figure 10B).

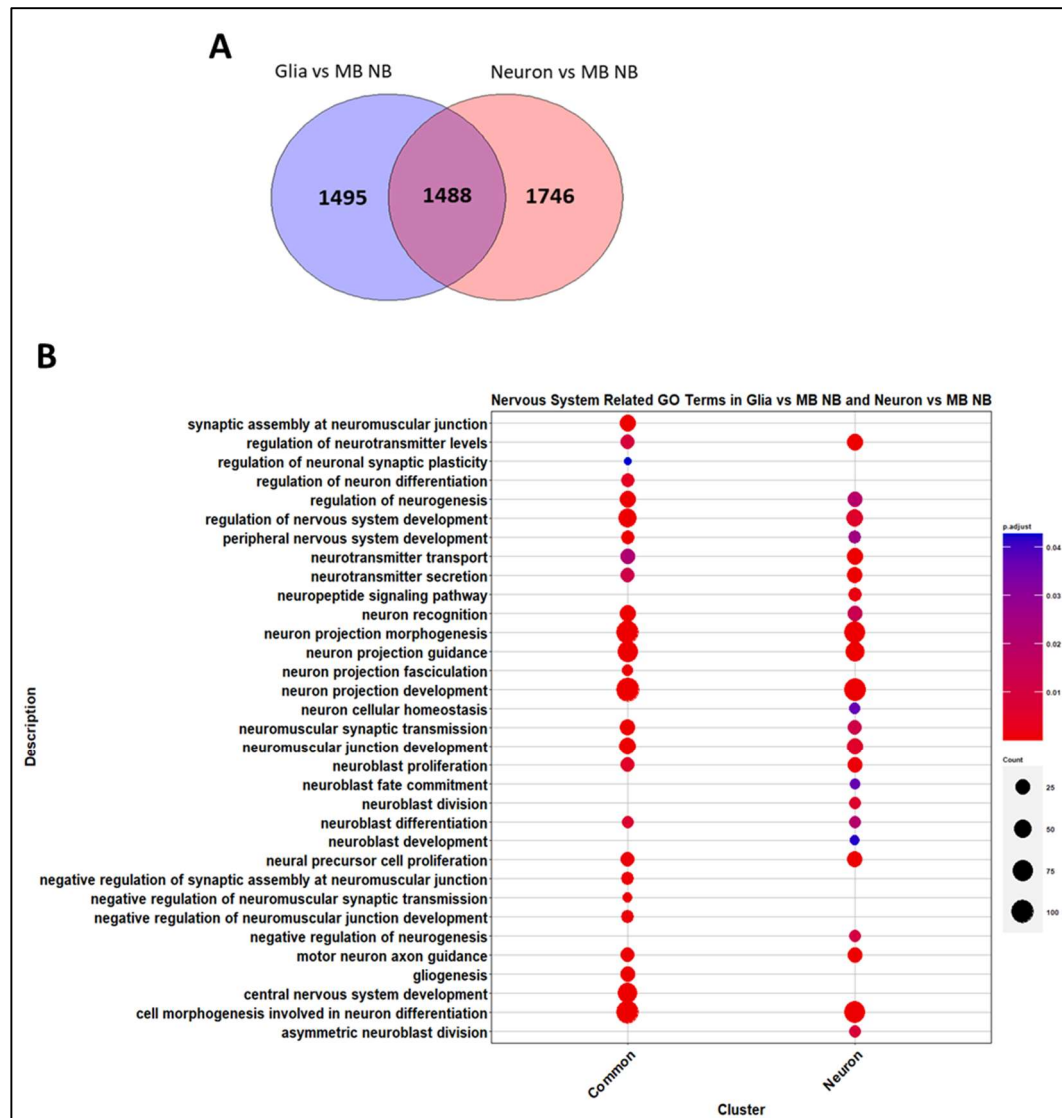


Figure 3.20: Glia and Neuron Lineage Formation from MB NB. A. The number of differentially expressed genes in glia, neuron and intersection of them when compare to AL neuroblast (p-adjusted < 0.05). B. Nervous system related GO terms in neuron, glia and intersection of both when neuron and glia originated from MB NB.

The common genes that contain genes that play role both neuronal and glial processes were taken and created gene regulatory network for examining the role of the ETS transcription factors in lineage formation from MB NB (Figure 3.21, 3.22). Aop, Eip74EF, Ets96B and Pnt upregulated and only Ets65A downregulated in the glia compare to MB NB data which is similar to glia formation from AL NB network (Figure 3.21). These 5 transcription factors regulates most of the genes together. However, Pnt and Aop downregulated while Eip74EF, Ets65A and Ets96B

downregulated in neuron compare to MB NB data (Figure 3.22). Similar to glia data, all TFs regulate the genes together and there is no significant cluster specific to TFs.

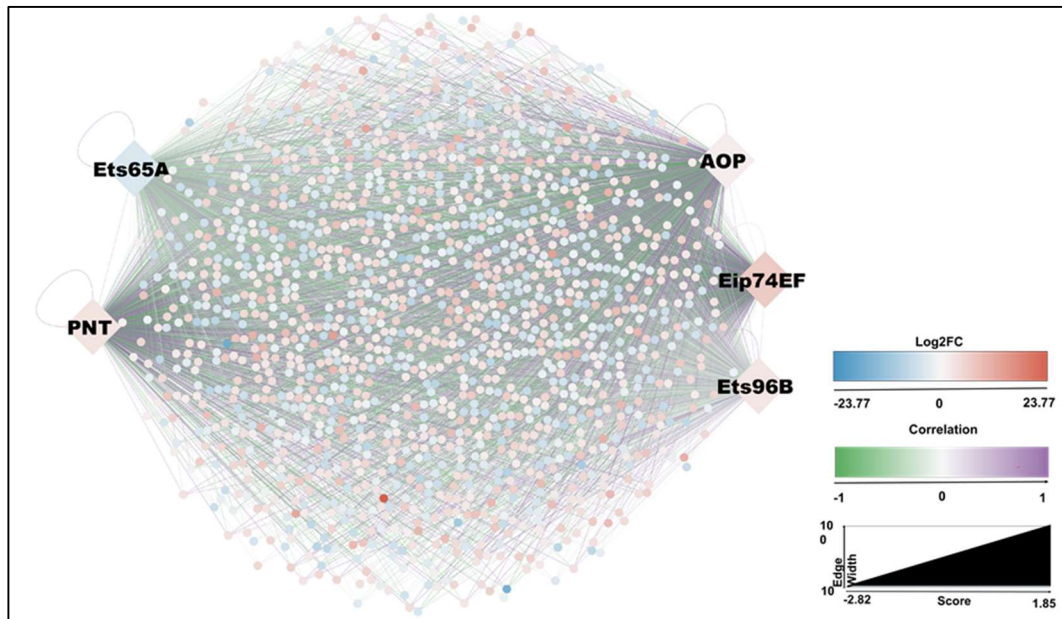


Figure 3.21: The gene regulatory network of common genes according to Glia vs MB NB dataset.

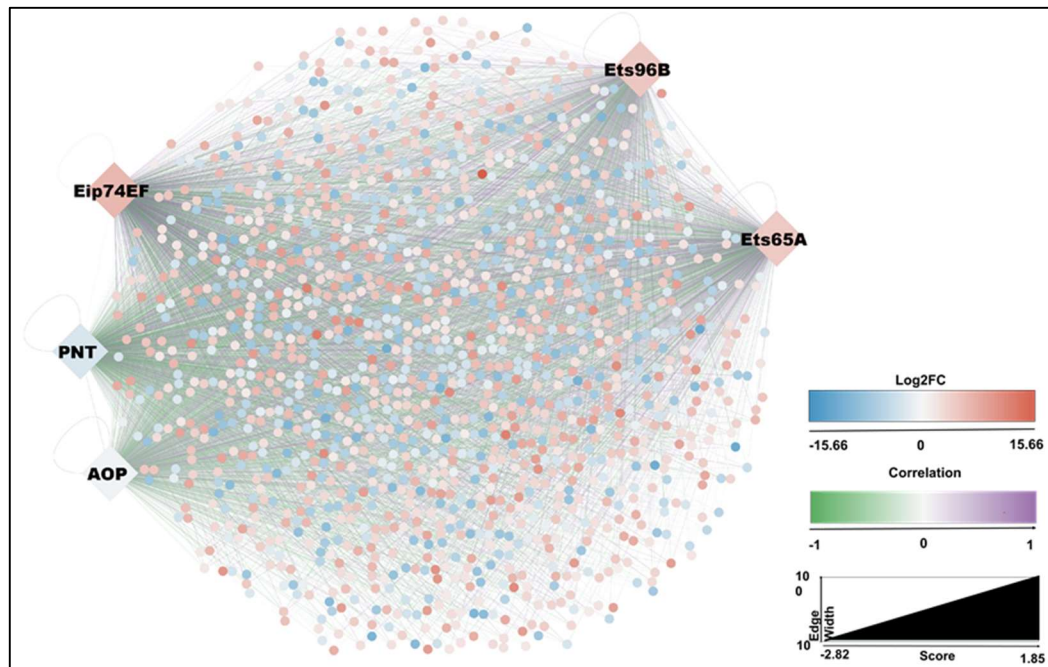


Figure 3.22: The gene regulatory network of common genes according to neuron compared to MB NB dataset.

To distinguish specific genes which are regulated by selected ETS TFs, we selected nervous system related genes from common genes and performed promoter analysis to detect that those genes have binding regions for our ETS transcription factors and created a gene regulatory network (Figure 3.23, 3.24). In both glia compared to MB NB network (Figure 3.23) and neuron compared to MB NB network (Figure 3.24), same 15 genes are regulated by the same TFs. Among these genes, Nlg3, Syn and dpr11 genes have different expression profile in glia and neuron network. Nlg3 gene plays role in formation of synapse and synaptic transmission by providing coding of synaptic adhesion molecule and it is upregulated in neuron network while downregulated in glia formation. Syn which have role in olfactory learning is also upregulated in neuron formation while downregulated in glia formation. The dpr11 plays a role in neuromuscular junction development regulation and axon guidance so it is upregulated in the neuron network while downregulated in the glia network. There are 9 genes -Syn, Gnf1, Antp, Dmtn, spartin, neur, Rfx, Cont, cwo- are found in both AL NB networks and MB NB networks with same expression patterns.

The genes only found in MB NB networks are Itpr, Nlg3, opa and dpr11. Itpr is needed to production of intracellular ligand gated calcium channel which is important for signalling and it also play role in mushroom body development [49] so this gene is upregulated in both neuron and glia formation from MB NB. opa is one of the gene that responsible from temporal patterning and it indicates that the neuron stem cell is aging [50]. This gene is downregulated in both neuron and glia compared to MB NB networks.

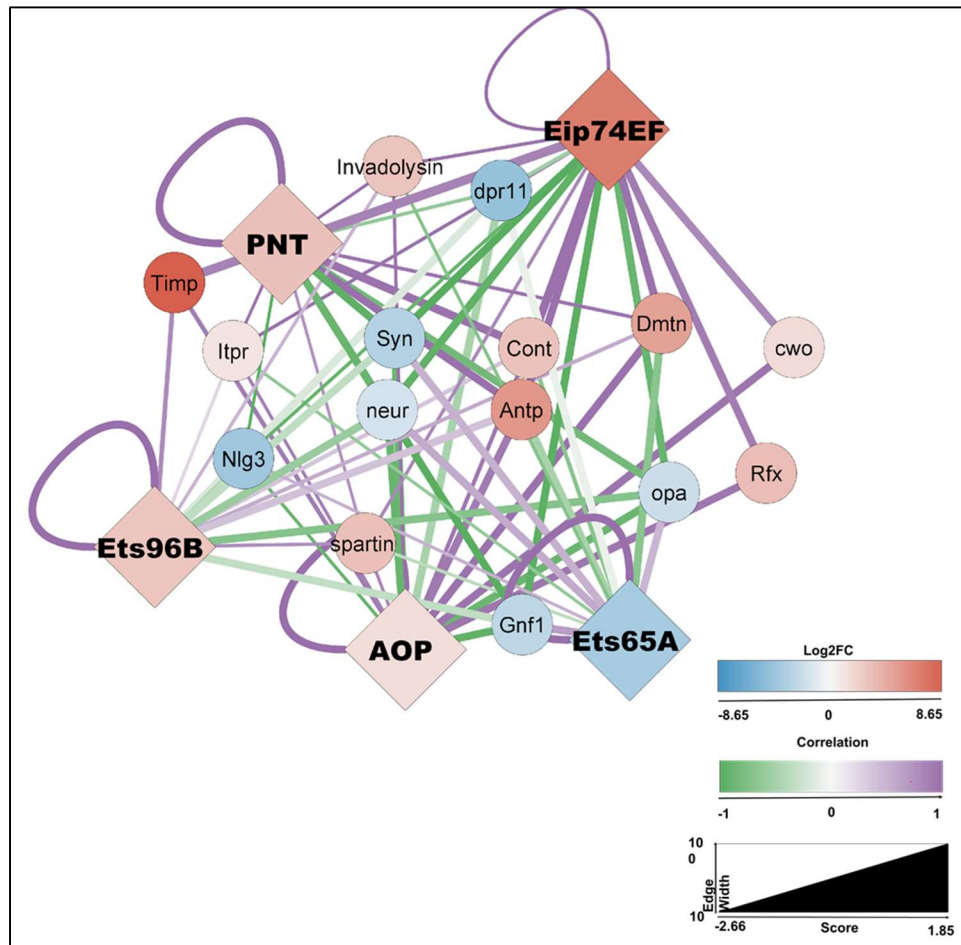


Figure 3.23: The gene regulatory network which obtained from promoter analysis results for AOP, Ets65A, Eip74EF, Ets65A and PNT transcription factors of nervous system related genes of common genes by using glia vs MB NB dataset.

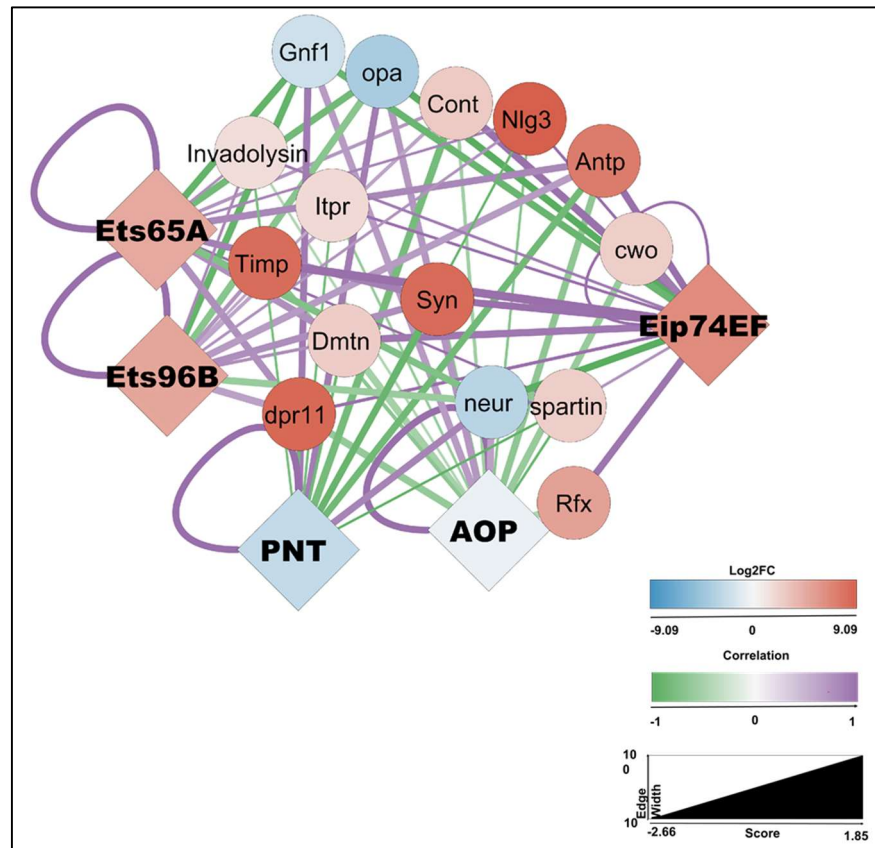


Figure 3.24: The gene regulatory network which obtained from promoter analysis results for Aop, Ets65A, Eip74EF, Ets65A and Pnt transcription factors of nervous system related genes of common genes of neuron compared to MB NB.

4. DISCUSSION and CONCLUSION

ETS transcription factor family have different functions and play role in different processes such as Ets65A have role in cell differentiation, Pnt take part in cell migration and cell differentiation. Two ETS members, Aop and Pnt work as antagonistically and based on this interfamilial relationship, ETS genes may play role in lineage formation.

In this thesis, we analyzed lineage formation from 3 different types of neuroblast, Type II NB, AL NB and MB NB, by using previously published *Drosophila* larva data. After differentially expressed genes analysis and promoter analysis of common genes that found in both neuron and glia data, gene regulatory networks were obtained for each neuroblast and lineage types. In lineage formation from Type II NB, Ets65A more specific to neuron lineage formation because when we compare subnetworks, Ets65A found only in neuron network and it regulates genes that play role in neuroblast proliferation like atypical protein kinase C (aPKC) and buttonhead (btd) and also regulates genes take part in neurogenesis like mas that have role in axon development, FER which play role in axon guidance and galectin that have role in synaptic target recognition. When we compare glia and neuron subnetworks, there are genes that found at both networks and although their expression patterns similar, their regulators change except for genes play role in neuron differentiation dap, Cbp53E, Dbx. The expression pattern and regulators of these 3 genes are changed. There is only one gene that expression pattern and regulator did not change which name is wicked (wcd) that regulates cell size for stem cell maintenance by ribosome synthesis. Also, the expression of the Aop decrease in neuron so it may important for glia formation.

In lineage formation from antennal lobe neuroblast, same genes regulated by the same transcription factors in both glia and neuron networks. Only 3 genes expression were changed. Syt4 that have role in synaptic plasticity regulation, Syn which take part in olfactory learning were both related with neurogenesis so it is upregulated in neuron. cas which is the another gene that expression was changed and it play role in glial cell development so it was upregulated in glia network. Although they are under the control of the same transcription factors, the change in the expression levels of transcription factors as Aop and Pnt upregulated in glia and

downregulated in glia and Ets65A is upregulated in neuron provide the change in expression levels of these genes.

In lineage formation from MB NB, same transcription factors regulate same genes for both glia and neuron networks similar to AL NB networks. In here, again there are 3 genes, Nlg3, Syn and dpr11 which expression levels are changed while other genes have similar expression patterns for both networks. These three genes have role in neurogenesis so they are upregulated in neuron network while downregulated in glia network. When we compare AL NB networks and MB NB networks, there are 9 genes that found in both lineage formation networks. These genes have role in neuroblast development, learning and memory. Syn has role in olfactory learning, memory and neur has role in long-term memory which have similar role reason finding of these genes in both AL NB and MB NB analysis and they might be separated in the adult.

When we examined the lineage formation from all three NB, there are only two gene that found in all three networks which are cwo and spartin has been identified. cwo (clockwork orange) is important role in the regulation of transcription of the rhythmic genes which keeps circadian time. This gene is upregulated and under the control of Ets96B in glia formation from Type II NB. In ALNB networks, it is also upregulated and controlled by the all 5 transcription factor in both neuron and glia lineage formation networks while it is regulated by only Aop and Eip74EF in MB NB networks for both and glia and neuron formation. spartin is found in organelle membranes and can associated with different molecules like lipid droplets, mitochondria. By interacting with mitochondria, it play role in the regulation of the mithondrial calcium homeostasis [51]. This gene is regulated by the Eip74EF in neuron lineage formation from Type II NB while it is under the control of all five ETS transcription factor in both glia and neuron formation although it is upregulated in all networks.

Transcription factors can be activators that promote gene expression or repressors that inhibit the gene expression. The Ets transcription factors Ets65A, Ets96B, Eip74EF, Pnt are activators and Aop is repressor. According to this information, genes with positive correlation with Ets TFs were selected for the interpretation of the networks obtained as a result of the analysis (Figure 3.25 – 3.30). As an example, if Pnt was downregulated, we selected downregulated genes because

Pnt is activator. However, if Aop was upregulated, the downregulated genes were selected because Aop is repressor.

In lineage formation form Type II NB, Ets96B regulated genes that mostly related with neuronal process like axon guidance, synaptic processes in glia dataset (Figure 3.25). However, different transcription factors are involved in the regulation of neuronal processes in the neuron dataset. For example, Eip74EF regulates genes involved in axon guidance, while genes involved in synaptic processes are regulated by Ets96B and Ets65A (Figure 3.26).

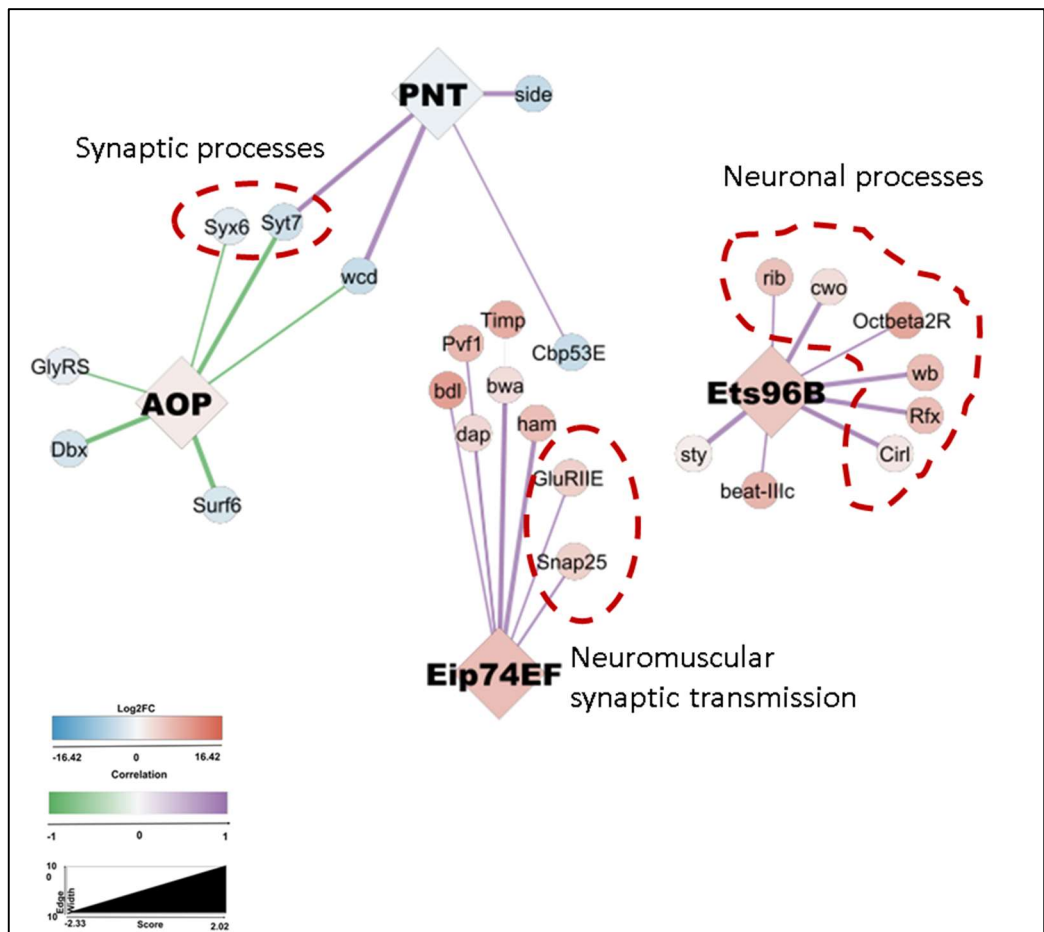


Figure 3.25: The gene regulatory network of genes that positively regulated by Ets TFs in glia vs Type II NB dataset.

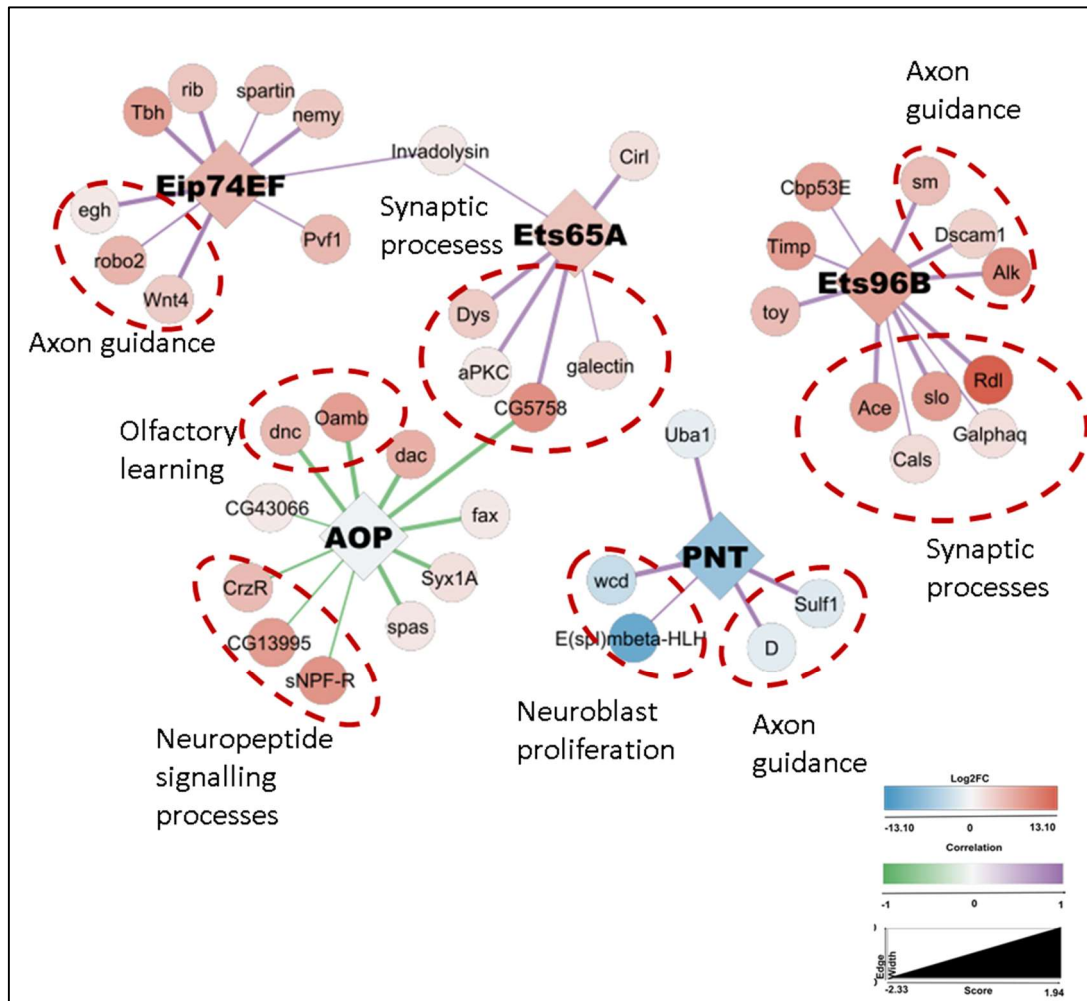


Figure 3.26: The gene regulatory network of genes that positively regulated by Ets TFs in neuron compared to Type II NB.

In lineage formation form AL NB, synaptic process genes were regulated by Eip74EF and Pnt, and neuronal process genes were regulated by Ets96B, Eip74EF and Pnt in glia dataset (Figure 3.27). However, the genes involved in these processes are regulated in the neuron network by Aop, Eip74EF, Ets96B and Ets65A (Figure 3.28).

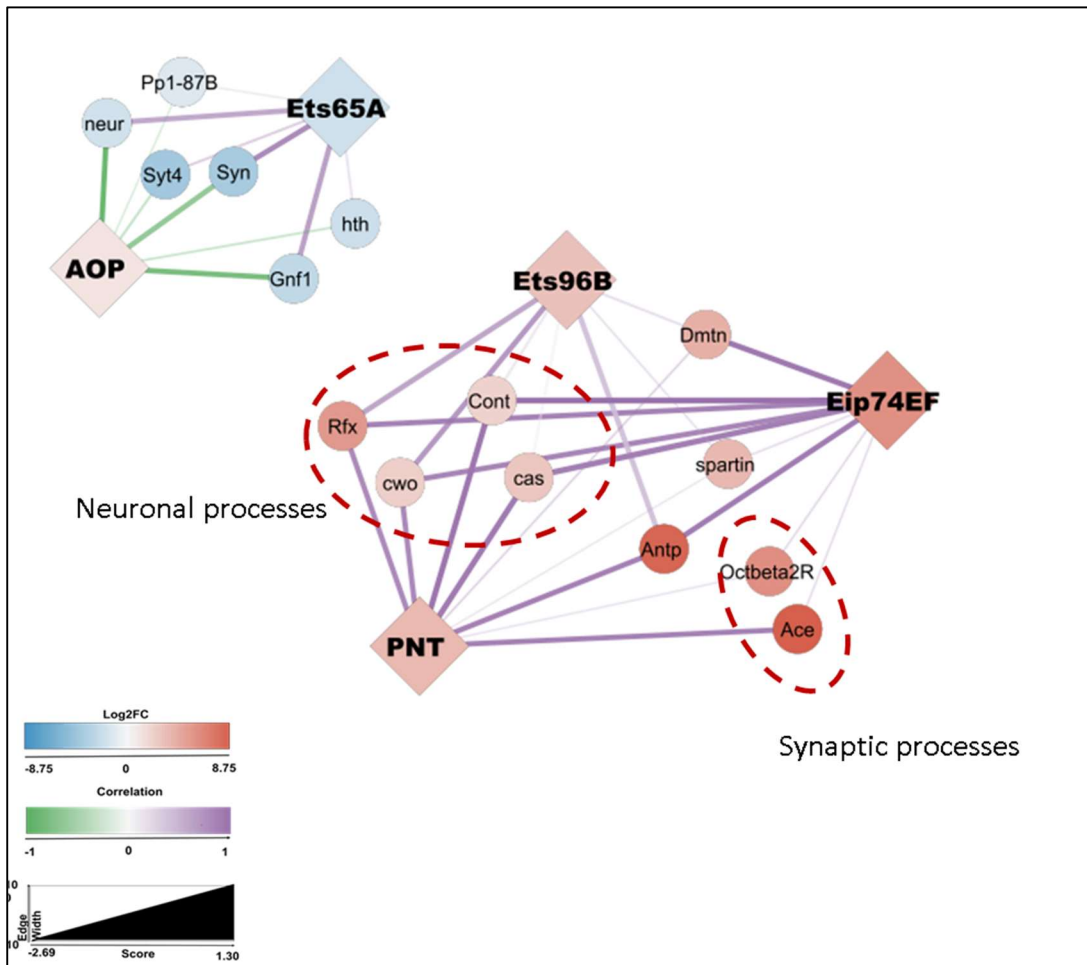


Figure 3.27: The gene regulatory network of genes that positively regulated by Ets TFs in glia vs AL NB dataset.

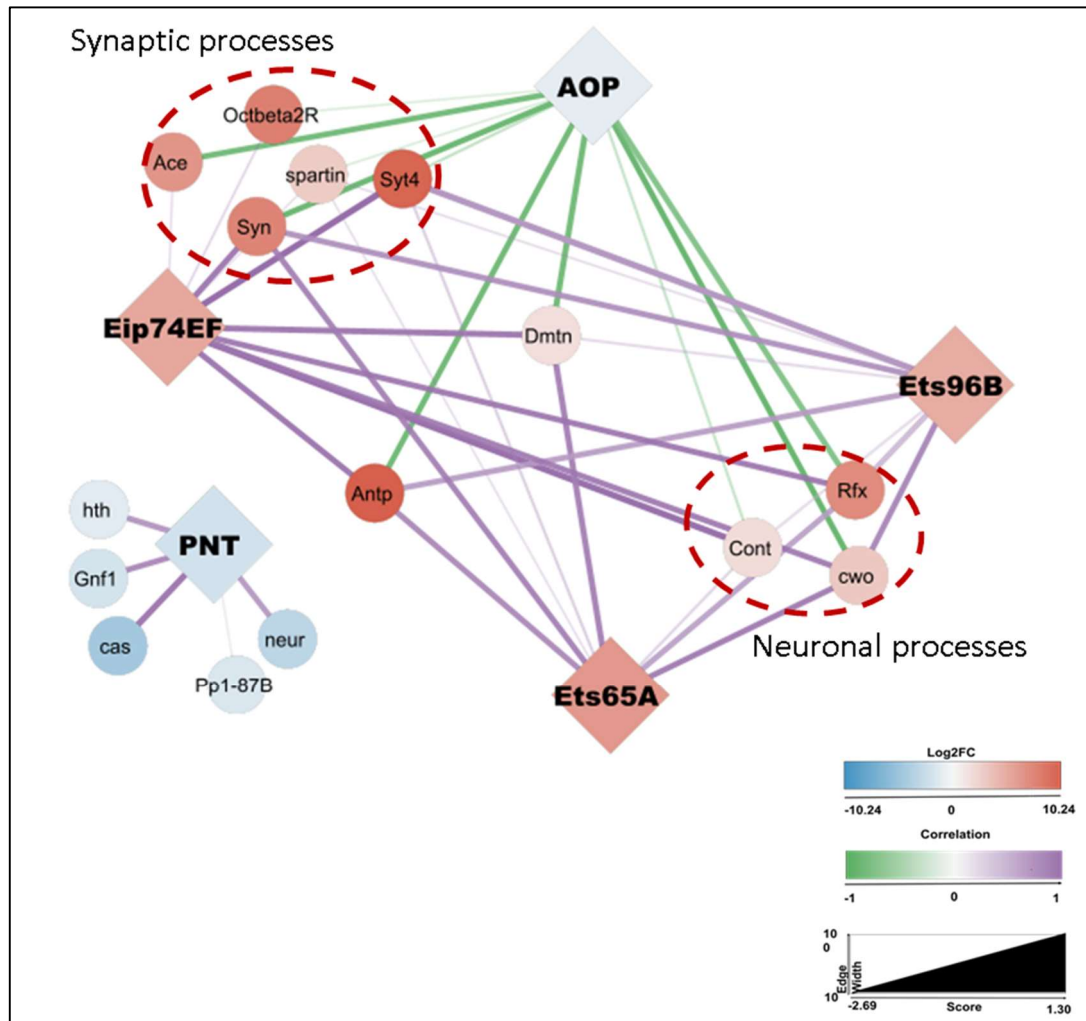


Figure 3.28: The gene regulatory network of genes that positively regulated by Ets TFs in neuron was compared to AL NB.

In lineage formation from MB NB, neuronal process genes were regulated by Eip74EF, and neuromuscular junction development genes were regulated by Aop and Ets65A in glia dataset (Figure 3.29). In neuron network, Eip74EF and Aop were regulated neuronal processes genes and other 4 TFs which are Aop, Eip74EF, Ets96B and Ets65A were regulated neuromuscular junction development, synaptic assembly processes genes (Figure 3.30).

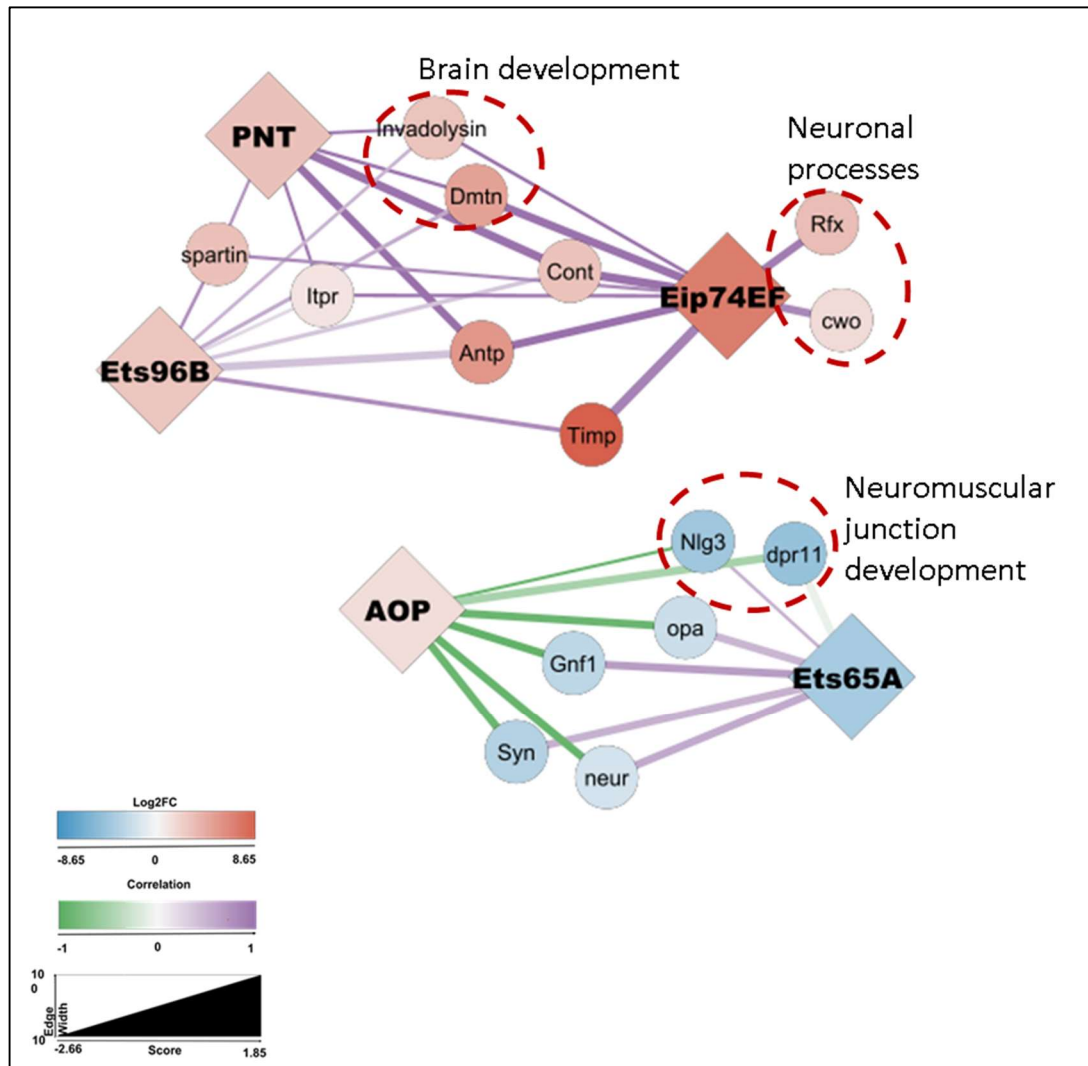


Figure 3.29: The gene regulatory network of genes that positively regulated by Ets TFs in glia was compared to MB NB dataset.

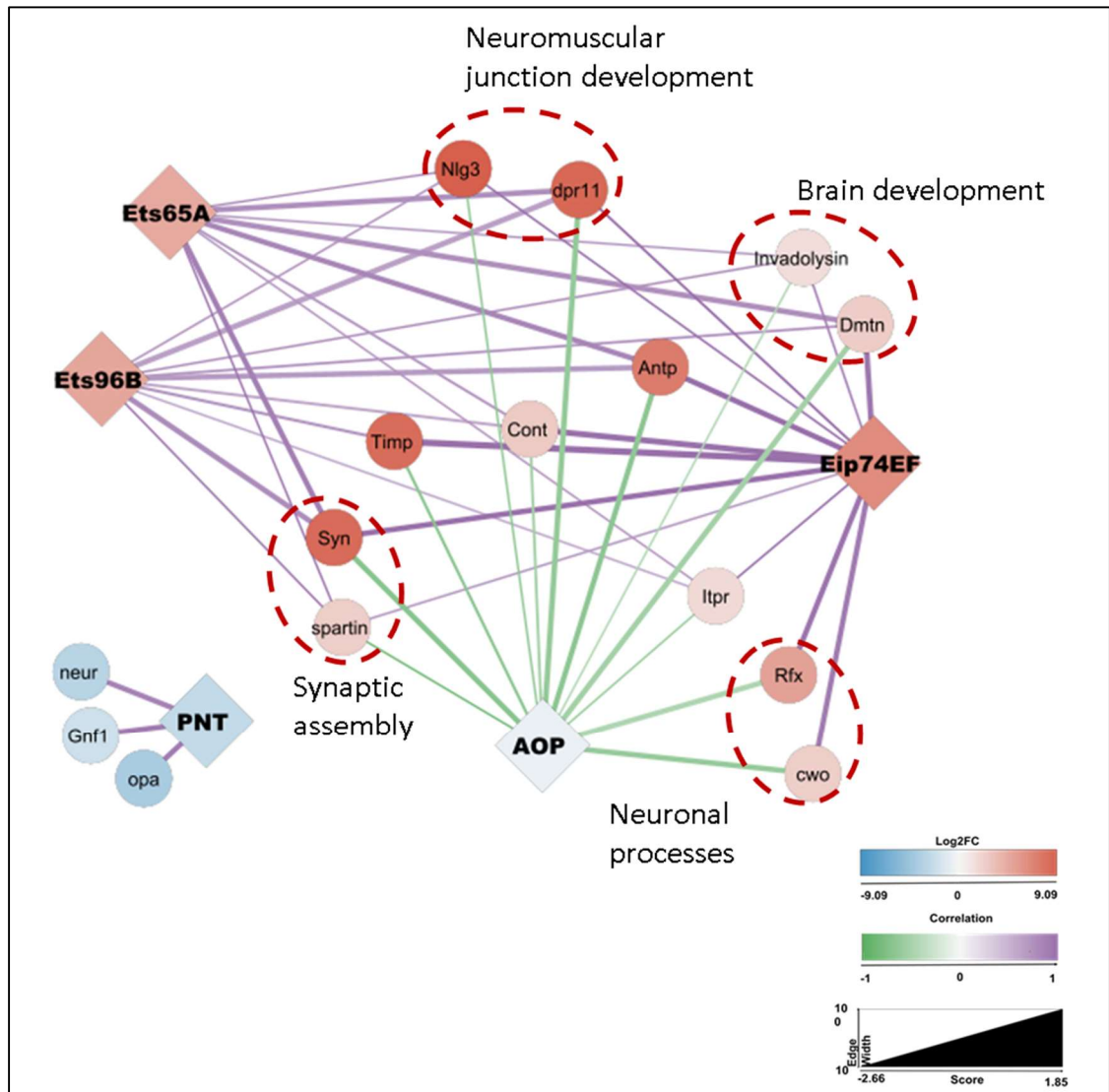


Figure 3.30: The gene regulatory network of genes that positively regulated by Ets TFs in neuron was compared to MB NB.

REFERENCES

- [1] Crews ST., (2019), "Drosophila Embryonic CNS Development: Neurogenesis, Gliogenesis, Cell Fate, and Differentiation", *Genetics*, 213 (4),1111.
- [2] Niven JE, Graham CM, Burrows M., (2007), "Diversity and Evolution of the Insect Ventral Nerve Cord", 53, 253–71.
- [3] Urbach R, Jussen D, Technau GM., (2016), "Gene expression profiles uncover individual identities of gnathal neuroblasts and serial homologies in the embryonic CNS of Drosophila", *Development*, 143 (8),1290–301.
- [4] Li G, Hidalgo A., (2020), "Adult Neurogenesis in the Drosophila Brain: The Evidence and the Void", *Int J Mol Sci*, 21 (18),6653.
- [5] Yasugi T, Nishimura T., (2016), "Temporal regulation of the generation of neuronal diversity in Drosophila", *Dev Growth Differ*, 58 (1),73–87.
- [6] Sada A, Tumber T., (2013), "New Insights into Mechanisms of Stem Cell Daughter Fate Determination in Regenerative Tissues", *Int Rev Cell Mol Biol*, 1 300,1–50.
- [7] Kambadur R, Koizumi K, Stivers C, Nagle J, Poole SJ, Odenwald WF., (1998), "Regulation of POU genes by castor andhunchback establishes layered compartments in the Drosophila CNS", *Genes Dev*, 12 (2), 246–60.
- [8] Brody T, Odenwald WF., (2000), "Programmed transformations in neuroblast gene expression during Drosophila CNS lineage development", *Dev Biol*, 226 (1), 34–44.
- [9] Isshiki T, Pearson B, Holbrook S, Doe CQ., (2001), "Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny", *Cell*, 106 (4), 511–21.
- [10] Baral SS, Lieux ME, DiMario PJ., (2019), "Nucleolar stress in Drosophila neuroblasts, a model for human ribosomopathies", *bioRxiv*, 704734.
- [11] Homem CCF, Knoblich JA., (2012), "Drosophila neuroblasts: A model for stem cell biology", *Development Cambridge*,139 (23), 4297–310.
- [12] Homem CCF, Knoblich JA., (2012), "Drosophila neuroblasts: a model for stem cell biology, *Development*, 139 (23):4297–311.
- [13] Izergina N, Balmer J, Bello B, Reichert H., (2009), "Postembryonic development of transit amplifying neuroblast lineages in the Drosophila brain", *Neural Dev*, 4 (1), 1–13.
- [14] Chen R, Deng X, Zhu S., (2022), "The Ets protein Pointed P1 represses Asense expression in type II neuroblasts by activating Tailless", *PLOS Genet*, 18

(1),e1009928.

- [15] Kunz T, Kraft KF, Technau GM, Urbach R., (2012), "Origin of *Drosophila* mushroom body neuroblasts and generation of divergent embryonic lineages", *Development*, 139 (14), 2510–22.
- [16] Yang CP, Fu CC, Sugino K, Liu Z, Ren Q, Liu LY, Yao X, Lee LP, Lee T, (2016), "Transcriptomes of lineage-specific *Drosophila* neuroblasts profiled by genetic targeting and robotic sorting", *Development*, 143 (3), 411–21.
- [17] Sipe CW, Siergrist SE, (2017), "Eyeless uncouples mushroom body neuroblast proliferation from dietary amino acids in *Drosophila*", *Elife*, 6.
- [18] Lai SL, Awasaki T, Ito K, Lee T., (2008), "Clonal analysis of *Drosophila* antennal lobe neurons: diverse neuronal architectures in the lateral neuroblast lineage", *Development*, 135 (17), 2883–93.
- [19] Das A, Sen S, Lichtneckert R, Okada R, Ito K, Rodrigues V, (2008), "*Drosophila* olfactory local interneurons and projection neurons derive from a common neuroblast lineage specified by the empty spiracles gene", *Neural Dev*, 3 (1), 1–17.
- [20] Gutierrez-Hartmann A, Duval DL, Bradford AP., (2007), "ETS transcription factors in endocrine systems", *Trends Endocrinol Metab.*, 18 (4), 150–8.
- [21] Chotteau-Lelièvre A, Desbiens X, Pelczar H, Defossez PA, De Launoit Y., (1997), "Differential expression patterns of the PEA3 group transcription factors through murine embryonic development", *Oncogene*, 15 (8), 937–52.
- [22] Sharrocks AD., (2001), "The ETS-domain transcription factor family", *Nat Rev Mol Cell Biol.*, 2 (11), 827–37.
- [23] Vivekanand P., (2018), "Lessons from *Drosophila* Pointed, an ETS family transcription factor and key nuclear effector of the RTK signaling pathway", *Genesis*, 56 (11–12).
- [24] Livet J, Sigrist M, Stroebel S, De Paola V, Price SR, Henderson CE, Jessel TM, Arber S, (2002), "ETS gene *Pea3* controls the central position and terminal arborization of specific motor neuron pools", *Neuron*, 35 (5), 877–92.
- [25] Gaudet P, Livstone MS, Lewis SE, Thomas PD., (2011), "Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium", *Brief Bioinform*, 12 (5), 449–62.
- [26] Lee CY, Baehrecke EH., (2001), "Steroid regulation of autophagic programmed cell death during development", *Development*, 128 (8), 1443–5.
- [27] Lai ZC, Rubin GM., (1992), "Negative control of photoreceptor development in *Drosophila* by the product of the *yan* gene, an ETS domain protein", *Cell*, 70 (4), 609–20.

- [28] Williams MJ, Klockars A, Eriksson A, Voisin S, Dnyansagar R, Wiemerslage L, (2016), "The Drosophila ETV5 Homologue Ets96B: Molecular Link between Obesity and Bipolar Disorder", PLoS Genet, 12 (6), 1–33.
- [29] Web 1, (2022), <https://usegalaxy.org/>, (Date accessed: 05/01/2022)
- [30] Web 2, (2022), <https://github.com/kevinblighe/EnhancedVolcano>, (Date accessed: 15/03/2022)
- [31] Web 3, (2022), <https://github.com/dusadrian/venn>, (Date accessed: 10/03/2022)
- [32] Web 4, (2022), <https://github.com/YuLab-SMU/clusterProfiler>, (Date accessed: 15/03/2022)
- [33] Wickham H. ggplot2 Elegant Graphics for Data Analysis. Use R! Ser. 2016;211.
- [34] Web 5, (2022), <https://ccg.epfl.ch/ssa/findm.php>, (Date accessed: 25/01/2022)
- [35] Web 6, (2022), <http://cisbp.cibr.utoronto.ca/index.php>, (Date accessed: 25/01/2022)
- [36] Web 7, (2022), www.bioconductor.org/packages/release/bioc/html/pandaR.html, (Date accessed: 19/06/2022)
- [37] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al., (2003), "Cytoscape: a software environment for integrated models of biomolecular interaction networks", Genome Res., 13 (11), 2498–504.
- [38] Web 8, (2022), <http://www.bioconductor.org/packages/release/data/annotation/html/TxDb.Dmelanogaster.UCSC.dm6.ensGene.html>, (Date accessed: 19/06/2022)
- [39] Web 9, (2022), <http://www.bioconductor.org/packages/release/bioc/html/MotifDb.html>, (Date accessed: 19/06/2022)
- [40] Mollinari C, Lange B, González C., (2002), "Miranda, a protein involved in neuroblast asymmetric division, is associated with embryonic centrosomes of Drosophila melanogaster", Biol Cell, 94 (1), 1–13.
- [41] Dillon N, Cocanougher B, Sood C, Yuan X, Kohn AB, Moroz LL, (2022), "Single cell RNA-seq analysis reveals temporally-regulated and quiescence-regulated gene expression in Drosophila larval neuroblasts", Neural Dev., 17 (1), 1–18.
- [42] Berger C, Renner S, Lüer K, Technau GM., (2007), "The commonly used

marker ELAV is transiently expressed in neuroblasts and glial cells in the *Drosophila* embryonic CNS", *Dev Dyn.*, 236 (12), 3562–8.

- [43] Siebert M, Banovic D, Goellner B, Aberle H., (2009), "Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields", *Genes Dev.*, 23 (9), 1052–62.
- [44] Meyer F, Aberle H., (2006), "At the next stop sign turn right: the metalloprotease Tolloid-related 1 controls defasciculation of motor axons in *Drosophila*", *Development*, 133 (20), 4035–44.
- [45] Sink H, Rehm EJ, Lee R, Bulls YM, Goodman CS., (2001), "Sidestep encodes a target-derived attractant essential for motor axon guidance in *Drosophila*", *Cell*, 105 (1), 57–67
- [46] Barber CF, Jorquera RA, Melom JE, Littleton JT., (2009), "Postsynaptic regulation of synaptic plasticity by synaptotagmin 4 requires both C2 domains", *J Cell Biol.*, 187 (2), 295–310.
- [47] Godenschwege TA, Reisch D, Diegelmann S, Eberle K, Funk N, Heisenberg M, (2004), "Flies lacking all synapsins are unexpectedly healthy but are impaired in complex behaviour", *Eur J Neuroscience*, 20 (3), 611–22.
- [48] von Hilchen CM, Beckervordersandforth RM, Rickert C, Technau GM, Altenhein B., (2008), "Identity, origin, and migration of peripheral glial cells in the *Drosophila* embryo", *Mech Dev.*, 125 (3–4), 337–52.
- [49] Kobayashi M, Michaut L, Ino A, Honjo K, Nakajima T, Maruyama Y, (2006), "Differential microarray analysis of *Drosophila* mushroom body transcripts using chemical ablation", *Proc Natl Acad Sci U S A*, 103 (39), 14417–22.
- [50] Abdusselamoglu MD, Eroglu E, Burkard TR, Knoblich JA., (2019), "The transcription factor odd-paired regulates temporal identity in transit-amplifying neural progenitors via an incoherent feed-forward loop", *Elife*, 8.
- [51] Nahm M, Lee MJ, Parkinson W, Lee M, Kim H, Kim YJ, Kim S, Cho YS, Min BM, Bae YC, Broadie K, Lee S, (2013), "Spartin Regulates Synaptic Growth and Neuronal Survival by Inhibiting BMP-Mediated Microtubule Stabilization", *Neuron*, 77 (4), 680–95.

BIOGRAPHY

Pınar MUTLU achieved her bachelor's degree from Department of Molecular Biology and Genetics, Gebze Technical University in 2020. She has continued master's degree program in the Department of Molecular Biology and Genetics at Molecular Neurobiology Laboratory, Graduate School of Natural and Applied Sciences, Gebze Technical University, since 2020 fall. She presented poster presentation in 18th National Neuroscience Congress in 2020 namely "ETS Transcription Factor Family in Drosophila Neurogenesis" and in 4th Congress of Cell Death Research Society- Turkiye with International Participation in 2022 with same name.