



HAL
open science

From molecular to functional diversity : a genome wide in vivo study of the *Drosophila melanogaster* neurons, glia and macrophages

Rosy Sakr

► **To cite this version:**

Rosy Sakr. From molecular to functional diversity : a genome wide in vivo study of the *Drosophila melanogaster* neurons, glia and macrophages. Agricultural sciences. Université de Strasbourg, 2021. English. NNT : 2021STRAJ075 . tel-04588154

HAL Id: tel-04588154

<https://theses.hal.science/tel-04588154>

Submitted on 26 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITÉ DE STRASBOURG

*ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ
INSTITUT DE GÉNÉTIQUE ET DE BIOLOGIE MOLÉCULAIRE ET CELLULAIRE*

THÈSE

présentée par :

SAKR ROSY

soutenue le : **15 décembre 2021**

pour obtenir le grade de: **Docteur de l'Université de Strasbourg**

Mention: Sciences de la Vie et de la Santé

Discipline/ Spécialité: Génétique, Biologie moléculaire et Cellulaire

**DE LA DIVERSITÉ MOLÉCULAIRE À LA DIVERSITÉ FONCTIONNELLE : UNE
ÉTUDE *IN VIVO* À L'ÉCHELLE DU GÉNOME DES NEURONES, DES CELLULES
GLIALES ET DES MACROPHAGES DE *DROSOPHILA MELANOGASTER***

THÈSE dirigée par:

Mme. GIANGRANDE Angela

Directeur de recherche, IGBMC, Illkirch, France

RAPPORTEURS EXTERNES:

Mme. GÖTZ Magdalena

Directeur de recherche, LMY, Munich, Allemagne

M. SOUTHALL Tony

Directeur de recherche, Imperial College London,
London, Royaume-Uni

EXAMINATEUR INTERNE:

M. TORA László

Directeur de recherche, IGBMC, Illkirch, France

ACKNOWLEDGEMENTS

This work would not have been possible without the participation of many people who made this journey truly life changing.

I would like to express my sincere gratitude to my advisor Dr. Angela Giangrande. I thank you for giving me the opportunity of joining your lab. Your continuous guidance, motivation and support allowed me to reach my goals. You have been an inspiration to me and out of all the things you taught me the one that will stay with me forever is your work ethics and utter dedication.

I would like to express my special appreciation to the jury members and invited experts: Prof. Magdalena Götz, Dr. László Tora, Dr. Tony Southall, Dr. Bassem Hassan and Dr. Pierre Cattenoz, for accepting to take part in my thesis defense and reading and judging my work.

To my teammates, you have become my second family. You have supported me throughout this journey from the day I walked in the lab as a young M1 student. I will forever cherish all the happy memories. Dr. Pierre Cattenoz, I cannot express my gratitude for your guidance, your patience and your advice. Thank you for introducing me to the world of bioinformatics and all its potential, for always answering my many questions and for helping me develop my critical thinking. Mrs. Claude Delaporte, I am grateful for your help and advice and for being my fly whisperer, always knowing what the problem was and how to solve it. Most importantly, I thank you for your emotional support and motivation throughout my time in the lab. Mrs. Céline Riet, I thank you deeply for your help with experiments and logistics and all for your time spent ordering and reordering my flies. Mrs. Holy Andriamampianina, thank you for providing the food to my flies and always adapting to their many needs. Mr. Thomas Boutet, I am grateful for your help,

your enthusiasm and for always being ready for the next sample. Dr. Sara Monticelli and Dr. Alexia Pavlidaki, I thank you for your friendship that I know will last beyond our days in the lab. Thank you for the coffee breaks, the constant support, the advice and the laughter.

To all my friends, thank you for supporting me throughout this journey. I want to especially thank Judy, for helping me understand the physics behind microscopy which turned out to be important after all, Narimanne for sharing the dream of science with me and Tarek for motivating me to always do better. I would also like to thank Dr. Mireille Kallassy who believed in me very early on and helped me find the path that led me here.

Finally, I dedicate this work for my family. To my parents, you always encouraged me to follow my dreams and did everything in your power to help me reach them. Without your support and many sacrifices none of this would have been possible. I hope I made you proud. To my sisters, thank you for being the wildest humans in my life. You continuously inspire and motivate me while always being there for me when I need you. To my partner Jeremy, I cannot express how grateful I am to have you in my life. Thank you for tolerating my crankiness and unavailability during the writing of this thesis, thank you for constantly supporting me and for always finding a way to cheer me up.

I also thank the University of Strasbourg and the “Fondation pour la Recherche Médicale en France” (FRM) for supporting my PhD studies.

TABLE OF CONTENTS

TABLE OF CONTENTS.....	5
ABBREVIATIONS	9
RÉSUMÉ	13
SUMMARY.....	19
INTRODUCTION	25
I- Cell identity:.....	25
II- The regulation of gene expression in Eukaryotes:	26
1. Cis-Acting Regulatory Elements.....	26
2. Transcription factors	27
3. Epigenetic modifications.....	30
III- The model of <i>Drosophila melanogaster</i> :	60
1. The <i>Drosophila</i> life cycle.....	60
2. The Gal4-UAS system	62
3. The <i>Drosophila</i> nervous system.....	63
4. The immune system of <i>Drosophila melanogaster</i>	75
AIMS.....	83
1. Is a cell’s identity reflected in its expression profile?	84
2. Is a cell’s identity reflected in its epigenetic profile?.....	84
3. Do all cells within a cell type share a common identity?.....	84
CHAPTER I.....	87
I- Introduction:.....	87
II- Materials and Methods:.....	90
1. Fly strains and genetics	90
2. FACS sorting of hemocytes, glia and neurons	90
3. RNA extraction and sequencing.....	92
4. Data analysis	92
5. Quantitative PCR.....	93
6. Histone extraction and western blot.....	94
III- Results:	95
1. Impact of development and function on transcriptional landscapes	95

2. Cell-specific molecular signatures define embryonic glia, neurons and hemocytes.....	99
3. Hemocytes acquire new functions over time	103
4. Glia and hemocytes express specific scavenger receptors	106
5. Stage-specific transcriptional landscapes.....	111
IV- Discussion:	117
1. Cell identity is defined by a relatively low number of transcripts and is heavily impacted by temporal cues	117
2. Neurons, hemocytes and glia, differences and commonalities	120
V- Supplementary data:.....	124
CHAPTER II.....	131
I- Introduction:.....	131
II- Material and methods:.....	133
1. Fly strains and genetics	133
2. FACS sorting and CUT&RUN	133
3. FACS sorting and ChIP-seq.....	138
4. CUT&RUN pipeline	140
5. ChIP-seq analysis pipeline	141
6. Data analysis and visualization	141
7. Immunolabeling and BBB permeability assay.....	142
III- Result:.....	143
1. Correlation between H3K4me3, H3K9ac and H3K27me3	143
2. H3K9ac, H3K4me3 and H3K27me3 peak at specific genes	146
3. H3K4me3 and H3K9ac peaks are found around the TSS.....	149
4. Colocalization of H3K4me3 and H3K9ac identifies genes expressed at high levels. .	151
5. H3K9ac is linked to cell-specific functions	153
6. The evolution of the three marks through development	156
7. Comparing CUT&RUN and ChIP-seq data	162
8. Preliminary data on Gcn5 knock-down in glia.....	164
IV- Discussion:	171
1. Correlation between expression levels and histone marks.....	171
2. H3K9ac reflects cell identity.....	172
3. Modulation of H3K4me3, H3K9ac and H3K27me3 profiles through development ...	173
V- Supplementary data:.....	175

CHAPTER III	179
Manuscript: Temporal specificity and heterogeneity of <i>Drosophila</i> immune cells.....	181
DISCUSSION AND PERSPECTIVES	183
1. Cell identity at the transcriptional and epigenetic levels	183
2. The evolution of cell identity through development	187
3. Cell identity on the single cell level.....	188
REFERENCES:	193

ABBREVIATIONS

A

A/P: Anterior/Posterior

AbdA: Abdominal- A

Ac: Acetylation

Acetyl-CoA: Acetyl-Coenzyme A

AD: Activation Domain

Ada: Transcriptional ADAptor

AG: Astrocyte-like Glia

Ascl1: Achaete-scute complex-like 1

ATAC: Ada Two A Containing complex

ATP: Adenosine Triphosphate

B

BBB: Blood-Brain Barrier

bHLH: Basic Helix-loop-Helix

BRD: Bromodomain-Containing

Brm: Brahma

bZIP: Leucine zipper

C

CBP: CREB Binding Protein

CG: Cortex Glia

CHD: Chromodomain Helicase DNA-binding

ChIP-seq: Chromatin Immuno-Precipitation and sequencing

CNS: Central Nervous System

COPMPAS: Complex Of Proteins Associated with Set 1

CoREST: Co- Repressor for Element- 1- Silencing Transcription factor

CRC: Chromatin Remodeling Complex

CRE: Cis-acting Regulatory Elements

CREB: cAMP Response Element-Binding

CUT&RUN: Cleavage Under Targets & Release Using Nuclease

D

DBD: DNA Binding Domain

DNA: Deoxyribonucleic Acid

E

E(z): Enhancer of zeste

E16: Embryonic stage 16

ECM: Extra-Cellular Matrix

EED: Embryonic Ectoderm Development

EG: Ensheathing Glia

Elav: Embryonic lethal abnormal vision

ELP3: Elongation Protein 3

ESC: Embryonic Stem Cell

Esc: Extra sex combs

EZH2: Enhancer of Zeste Homolog 2

F

FACS: Fluorescence Activated Cell Sorting

G

GB: Glioblast

Gcm/Glide: Glial cells missing/glial cell deficient

Gcn5: General control nonderepressible 5

GFP: Green Fluorescent Protein

GMC: Ganglion Mother Cell

GNAT: Gcn5-related histone N-acetyltransferases

GTF: General Transcription Factor

H

HAT: Histone Acetyltransferase

HBO1: Histone acetyltransferase Binding to ORC-1

HDAC: Histone Deacetylase

HMT: Histone Methyltransferase

Hox: Homeobox

HP1: Heterochromatin Protein 1

I

IP: Immunoprecipitation

ISN: Intersegmental Nerve

ISWI: Imitation Switch

J

JAK: Janus kinase

JHDM2: Jumonji domain-containing Histone Demethylase-2

K

K: Lysine

KAT: Lysine Acetyltransferase

KD: Knock-Down

KDM: Lysine Demethylase

KMT: Lysine Methyltransferase

L

L3: Third instar larva

Lid: Little imaginal discs

LOF: Loss Of Function

LSD1: Lysine Specific Demethylase 1

Lz: Lozange

M

MB: Mushroom body

me: Methyl

MG: Midline Glia

Mid: Midline

MLL: Mixed- Lineage Leukemia

MOF: Male absent On the First

MORF: MOZ-Related Factor

MOZ: Monocytic leukemia zinc finger protein

mRNA: messenger Ribonucleic Acid

mRNA-seq: RNA sequencing

MYST: MOZ/YBF2/SAS2/Tip60 NCoR

N

NAD: Nicotinamide Adenine Dinucleotide

NB: Neuroblast

NGB: Neuroglioblasts

Nrx IV: Neurexin IV

NS: Nervous System

NSC: Neural Stem Cell

NuRD: Nucleosome Remodeling and Deacetylase

NURF: Nucleosome Remodeling Factor

P

P300: Protein of 300 kDa

Pc: Polycomb

PCAF: p300/CBP-Associated Factor

PcG: Polycomb group

PEV: Position-Effect Variegation	SIRT: Sirtuin type protein
PH3: Phospho-Histone H3-Ser10	SN: Segmental Nerve
PHD: Plant Homeodomain	SPG: Subperineural Glia
PIC: Preinitiation Complex	Srp: Serpent
PIWI: P- element induced Wimpy Testis	STAT: Signal Transducer and Activator of Transcription
PNG : Perineural Glia	Su(var): Suppressor of variegation
PNS: Peripheral Nervous System	Su(z)12/SUZ12: Suppressor of zeste 12
Pnt: Pointed	SWI/SNF: Switch/ Sucrose Nonfermenting
PPO: ProPhenolOxydases	
PRC1: Polycomb Repressor Complex 1	T
PRC2: Polycomb Repressor Complex 2	TBP: TATA binding protein
PRMT: Protein Arginin Methyltransferase	TF: Transcription factor
R	Tip60: HIV- 1 Tat interacting protein
R: Arginine	Trr: Trithorax-related
Repo: Reversed polarity	Trx: Trithorax
RNA Pol II: RNA Polymerase II	TrxG: Trithorax Group
RNA: Ribonucleic Acid	TSS: Transcription Start Site
RNAi: RNA interference	Ttk: Tramtrack
ROS: Reactive Oxygen Species	U
Rpd3: Reduced potassium dependency 3	UAS: Upstream activating sequence
RT-qPCR / qPCR: Reverse Transcription-quantitative polymerase chain reaction	Ub: Ubiquitin
S	Ubx: Ubithorax
SAGA: Spt- Ada- Gcn5- acetyl-transferase	Ush: U-shaped
Salm: Spalt major	V
Sc: Scute	VNC: Ventral nerve cord
ScRNAseq: Single cell RNA sequencing	Y
Sin3: Switch independent3	YBF2/SAS3: Yeast binding factor 2/Something about silencing 3
Sir2: Silent information regulator 2	Z
	Zn: zinc

RÉSUMÉ

L'une des questions biologiques les plus importantes de notre époque est: **comment nos cellules acquièrent-elles leur identité fonctionnelle?**

Mon projet vise à étudier le lien entre identité, fonction et profil moléculaire des cellules. Les cellules acquièrent des profils d'expression génique spécifiques qui leur permettent de devenir des unités fonctionnelles dans l'organisme. Il est devenu de plus en plus évident que les modifications épigénétiques impriment de véritables signatures sur les différentes cellules. Le rôle de chaque signature épigénétique et leur relation avec la fonction cellulaire ne sont pas encore établis. En fait, l'étude de ces connexions nécessite des analyses *in vivo*, un facteur limitant chez les organismes complexes tels que les vertébrés. J'ai donc choisi *Drosophila melanogaster*, un modèle simple, dont la plupart des processus biologiques sont conservés dans l'évolution. Cela permet l'exploration de mécanismes physiologiques et pathologiques *in vivo* grâce à un grand nombre d'outils génétiques et moléculaires sophistiqués. J'ai notamment mis en place une analyse comparative entre les cellules gliales, les hémocytes et les neurones. Le choix de ces cellules repose sur deux facteurs. Premièrement, les cellules gliales et les hémocytes proviennent d'origines différentes (ectoderme pour les cellules gliales, mésoderme pour les hémocytes) mais une fonction commune, la phagocytose et l'immunité innée. En fait, les hémocytes sont l'équivalent des macrophages. Ils constituent les cellules phagocytaires circulantes ou associées aux tissus et composent donc la première ligne de défense contre les agents pathogènes. Cependant, ces cellules sont exclues du système nerveux qui est isolé du reste de l'organisme par la barrière hémato-encéphalique (BHE). Ainsi, les cellules gliales représentent les cellules immunitaires du système nerveux. C'est la seule population phagocytaire du système nerveux et ces cellules correspondent

à la microglie des vertébrés, dont les défauts sont associés à de nombreuses pathologies humaines telles que l'épilepsie, la schizophrénie la sclérose en plaques et les cancers du cerveau. Deuxièmement, alors que les hémocytes et les cellules gliales ont des origines différentes mais partagent une fonction commune, les neurones et les cellules gliales ont des fonctions très différentes mais émergent du même précurseur et interagissent tout au long du développement pour assurer le bon fonctionnement du système nerveux. Ainsi, l'étude de l'aspect moléculaire de l'identité de ces cellules permettra de comprendre le lien entre les signatures moléculaires des cellules et leurs origines et fonctions.

Pour décrypter ces mécanismes moléculaires complexes, j'ai divisé le travail en 3 points principaux.

1) Comprendre comment l'identité des cellules se reflète dans leur expression génique

J'ai effectué un séquençage d'ARNm messenger (mRNA-seq) sur des hémocytes, des cellules gliales et des neurones pour établir leurs transcriptomes. Afin de comprendre comment l'expression des gènes évolue au cours du développement, nous avons effectué les analyses à un stade précoce, l'embryon, et à un stade ultérieur, la larve. La comparaison des transcriptomes de différents types cellulaires, en plus du même type cellulaire à différents stades de développement m'a permis de tirer quatre conclusions principales :

- Les trois types cellulaires présentent une plus grande similitude au stade embryonnaire qu'au stade larvaire. Cela peut être dû à l'acquisition de fonctions plus définies, avec un niveau supplémentaire de spécialisation aux stades larvaires.

- Les cellules gliales restent beaucoup plus proche des neurones que des hémocytes. Ceci suggère l'importance de l'histoire d'une cellule et de son environnement immédiat, avec lequel la cellule interagit tout au long de sa vie.
- Les hémocytes sont les plus affectés par les changements dus au stade de développement car leur profil d'expression change de manière significative entre les deux stades. Encore une fois, l'environnement semble jouer un rôle clé dans la spécification, car les hémocytes sont les sentinelles de première ligne répondant aux défis internes et externes, tandis que le système nerveux est affecté moins directement.
- Les cellules gliales et les hémocytes expriment les gènes nécessaires à la phagocytose, mais chaque type cellulaire exprime un ensemble différent de récepteurs phagocytaires. Cela implique un potentiel phagocytaire distinct des deux types cellulaires.

2) Etablir le lien entre fonction cellulaire et modifications épigénétiques

J'ai contribué au développement d'une nouvelle technique d'analyse de la chromatine, fiable et sensible, appelée Cleavage Under Targets & Release Using Nuclease (CUT&RUN), en collaboration avec un étudiant de notre laboratoire, T. Boutet. Nous avons effectué une analyse CUT&RUN sur les neurones et les cellules gliales des deux stades de développement mentionnés ci-dessus. En utilisant cette technique, j'ai identifié la distribution de trois marques, H3K4me3, H3K9ac et H3K27me3 dans les neurones et les cellules gliales. Cela m'a permis de tirer quatre conclusions principales :

- H3K9ac seul n'indique pas une transcription active, cependant, sa combinaison avec H3K4me3 représente les niveaux d'expression les plus élevés. Cela indique que même si H3K9ac

peut ne pas être suffisant pour activer l'expression des gènes, il est nécessaire pour augmenter les niveaux d'expression des gènes présentant H3K4me3.

- La combinaison de H3K9ac et H3K4me3 se trouve sur des gènes impliqués dans des processus spécifiques aux cellules. Alors que H3K4me3 semble marquer la majorité des sites actifs et ne montre aucune spécificité entre les différents types cellulaires, H3K9ac se trouve principalement sur les gènes spécifiques des cellules dans les neurones embryonnaires et les cellules gliales. Ainsi, H3K9ac est une marque épigénétique spécifique à la cellule reflétant l'identité cellulaire.

- Il y a un changement majeur dans la signature H3K9ac entre les stades embryonnaire et larvaire. Les gènes spécifiques aux cellules présentant H3K4me3 et H3K9ac aux stades embryonnaires perdent H3K9ac aux stades larvaires. Cela suggère que H3K9ac n'est nécessaire qu'au début de la vie de la cellule.

- H3K9ac est directement lié à la fonction cellulaire. La modulation des niveaux de H3K9ac dans les cellules gliales spécifiquement a montré une dérégulation des gènes présentant H3K9ac et une perturbation de la fonction des cellules gliales. Par conséquent, ces résultats mettent en évidence le rôle de H3K9ac dans les processus spécifiques aux cellules.

3) Comprendre la diversité cellulaire au sein d'une même population

Alors que les sous-types de neurones et de cellules gliales ont été largement étudiés, la diversité des hémocytes est moins bien comprise. Les hémocytes sont principalement des cellules immunitaires, mais des études récentes mettent en évidence leurs nombreuses fonctions au-delà de l'immunité. Comme il est peu probable que tous les hémocytes soient capables de remplir les différentes fonctions, cela suggère une hétérogénéité au sein de la population d'hémocytes. Ainsi,

nous avons utilisé des technologies de séquençage et de bioinformatique de pointe pour effectuer, pour la première fois, le séquençage d'ARN messager unicellulaire d'hémocytes de drosophile.

- Nous avons établi le premier atlas des hémocytes de drosophile et identifié 14 sous-types d'hémocytes. Chaque sous-type a un profil d'expression génique spécifique reflétant leurs identités uniques.

- Nous avons évalué comment le profil d'expression de ces cellules est modulé lors d'un « challenge » immunitaire et avons constaté que les principaux marqueurs exprimés dans un sous-type restent présents, indiquant ainsi que les sous-types identifiés présentent une identité réelle et non un état transitoire de ces cellules capturées à un certain moment. Ces travaux ont montré la grande diversité moléculaire et fonctionnelle au sein d'un même type cellulaire.

L'identité fonctionnelle d'une cellule est un concept complexe qui dépend non seulement des composantes moléculaires contenues dans la cellule mais aussi de son environnement et de l'état et les besoins de l'organisme. Nos cellules sont capables de s'adapter de manière transcriptionnelle, épigénétique et fonctionnelle afin de s'adapter aux changements que subit l'organisme. L'incapacité de s'adapter déclenche de nombreuses dérégulations qui affectent l'homéostasie entre les tissus et entraîne l'apparition de pathologies humaines graves comme les maladies auto-immunes, la neurodégénérescence et le cancer. Par conséquent, il est impératif de comprendre comment une cellule acquiert, maintient et fait évoluer ces fonctions pour pouvoir étudier les potentielles dérégulations. Ainsi, mes travaux serviront à la fois à la communauté scientifique intéressée par l'aspect fondamental de la fonction des cellules, et à la communauté dédiée à l'étude des nombreuses maladies liées à la perturbation de l'identité cellulaire.

SUMMARY

One of the most important biological questions of our time is: **How do our cells acquire their functional identity?**

My project aims to study the connection between cell identity, function and molecular profile. Cells acquire specific gene expression profiles that allow them to become functional units in the organism. It has become increasingly evident that epigenetic modifications imprint true signatures on different cells. The role of each epigenetic signature and their relationship to cell function is not yet established, as investigating these connections requires *in vivo* studies, a limiting factor in complex organisms such as vertebrates. Therefore, I chose *Drosophila melanogaster*, a simple model, most of whose biological processes are conserved in evolution. This allows the exploration of physiological and pathological mechanisms *in vivo* through a large number of sophisticated genetic and molecular tools. In particular, I implemented a comparative analysis between glia, hemocytes and neurons. The choice of these cells is based on two factors. First, glia and hemocytes have different origins (ectoderm for glia, mesoderm for hemocytes) but a common function, phagocytosis and innate immunity. In fact, hemocytes are the equivalent of macrophages. They make up the circulating or tissue-associated phagocytic cells and therefore form the first line of defense against pathogens. However, these cells are excluded from the nervous system which is isolated from the rest of the body by the blood-brain barrier (BBB). Thus, glia represent the immune cells of the nervous system. It is the only phagocytic population of the nervous system and these cells correspond to vertebrate microglia, whose defects are associated with many human pathologies such as epilepsy, schizophrenia, multiple sclerosis and brain cancers. Second, while hemocytes and glia have different origins but share a common function, neurons and glia have

very different functions but emerge from the same precursor and interact throughout development to ensure the proper functioning of the nervous system. Thus, studying the molecular aspect of these cells' identities will allow us to understand the link between the molecular signatures of cells and their origins and functions.

To decipher these complex molecular mechanisms, I divided the work into 3 main points.

1) Understanding how cell identity is reflected in their gene expression

We performed messenger mRNA sequencing (mRNA-seq) on hemocytes, glia and neurons to establish their transcriptomes. In order to understand how gene expression changes during development, we performed the analyses at an early stage, the embryo, and at a later stage, the larva. Comparing the transcriptomes of different cell types in addition to the same cell type at different developmental stages, allowed me to draw four main conclusions:

- The three cell types show greater similarity in the embryonic stage than in the larva. This may be due to the acquisition of more defined functions, with an additional level of specialization at larval stages.
- Glia remain much closer to neurons than to hemocytes. This suggests the importance of a cell's history and its immediate environment, with which the cell interacts throughout its life.
- Hemocytes are most affected by the changes due to developmental stage as their expression profile shifts significantly between the two stages. Again, the environment appears to play a key role in the specification, as hemocytes are the frontline sentinels responding to internal and external challenges, while the nervous system is affected less directly.

- Glia and hemocytes express genes necessary for phagocytosis but each cell type expresses a distinct set of phagocytic receptors. This implies a distinct phagocytic potential of the two cell types.

2) Establishing the connection between cell function and epigenetic modifications

I contributed to the development of a new, reliable and sensitive chromatin analysis technique in our laboratory, called Cleavage Under Targets & Release Using Nuclease (CUT&RUN). We performed CUT&RUN analysis on neurons and glia from the two developmental stages mentioned above. Using this technique we identified the distribution of three marks, H3K4me3, H3K9ac and H3K27me3 in neurons and glia. This allowed me to draw three main conclusions:

- H3K9ac alone does not indicate active transcription, however, its combination with H3K4me3 shows the highest levels of expression. This indicates that while H3K9ac may not be sufficient to activate gene expression, it is needed to boost expression levels of genes presenting H3K4me3.

- The combination of H3K9ac and H3K4me3 is found on genes involved in cell-specific processes. While H3K4me3 seems to mark the majority of active sites and shows no specificity between the different cell types, H3K9ac is mainly found on cell-specific genes in embryonic neurons and glia. Thus, H3K9ac is a cell-specific epigenetic mark reflecting cell identity.

- There is a major change in the H3K9ac signature between embryonic and larval stages. Cell-specific genes presenting H3K4me3 and H3K9ac at embryonic stages lose H3K9ac at larval stages. This suggests that H3K9ac is only needed early in the life of the cell. 4) H3K9ac is directly linked to cell function. Modulating the levels of H3K9ac in glia specifically showed deregulation

in genes presenting H3K9ac and disruption of the function of glia. Therefore, these results highlight the role of H3K9ac in cell-specific processes.

3) Understanding cell diversity within the same population

While the subtypes of neurons and glia have been extensively studied, the diversity of hemocytes on the other hand is less well understood. Hemocytes are primarily immune cells however recent studies highlight their many functions beyond immunity. Since it is unlikely that all hemocytes are able to perform the different functions, this calls for heterogeneity within the hemocyte population. Through this, we have used cutting edge sequencing and bioinformatics technologies to perform, for the first time, single-cell messenger RNA sequencing of *Drosophila* hemocytes.

- We established the first atlas of *Drosophila* hemocytes and identified 14 subtypes of hemocytes. Each subtype has a specific gene expression profile reflecting their unique identities.
- We assessed how the expression profile of these cells is modulated upon immune challenge and found that the main markers expressed in a subtype remain presents thus indicating that the subtypes identified present a real identity and not a state of these cells captured at a certain time point. This work has shown the great molecular and functional diversity within the same cell type.

The functional identity of a cell is a complex concept that depends not only on the molecular components contained in the cell but also on its environment and the condition and needs of the organism. Our cells are able to adapt transcriptionally, epigenetically and functionally to accommodate changes in the body. The inability to adapt triggers many deregulations that affect tissue homeostasis and leads to the development of severe human pathologies such as autoimmune diseases, neurodegeneration and cancer. Therefore, it is imperative to understand how a cell

acquires, maintains these functions and evolves in order to be able to study potential deregulations. Thus, my work will serve both the scientific community interested in the fundamental aspect of cell function, and the community dedicated to the study of the many diseases linked to the disruption of cell identity.

INTRODUCTION

I- Cell identity:

One of the greatest mysteries in biology is how a single fertilized egg is able to give rise to a fully functional organism. To attempt to answer such complex questions, biology requires classification, organisms are broken up into organs, tissues and cells, the latter being smallest functional units of life. It is clear that the proper function of cells is crucial for the wellbeing of the organism which makes understanding how they function of uttermost importance. The human body comprises 37.2 trillion cells (Bianconi et al., 2013). This calls for a subdivision of cells into different cell types based on their identity to facilitate investigating them. Deep understanding of cell identity is not only relevant to the fundamental standpoint. Accurate identification of cell identity has several applications in research and clinics and raises new prospects for regenerative medicine. For instance, the goal of regenerative medicine is to drive the differentiation of pluripotent cells into a specific cell type and one crucial question is whether the cells produced have acquired the true desired identity (Abdolhosseini et al., 2019; Murry and Keller, 2008). But the question remains: what is a cell's identity? How do we define it? Is it a constant state that is acquired when a cell differentiates and remains the same throughout the cell's life?

Until recently, studying a cell's identity involved the characterization of its origin, its morphology and its function. The current advancement of genome-wide studies allows us to investigate these ancient questions from a new perspective. We are now able to study a population of cells, even a single cell, on different levels and achieve high resolution through generating immense amount of information that traditional techniques cannot attain. The purpose of my thesis was to investigate cell identity on the transcriptional and the epigenetic level.

II- The regulation of gene expression in Eukaryotes:

How can cells sharing the same genetic information present immense heterogeneity? The short answer to this question is: differential gene expression. Gene expression is the process by which the instructions in our DNA are converted into a functional product responsible for the phenotype of a cell. A gene is said to be active when it is transcribed and inactive when it is not. The activity of a cell is determined by the activation of specific sets of genes at a specific time and in specific quantities. This regulation is central both to tissue specific-gene expression and to the regulation of gene activity in response to specific stimuli. The regulation of gene expression is an intricate mechanism involving primarily the combined actions of multiple different transcriptional regulatory proteins. In addition, the packaging of DNA into chromatin and its modification reveals further levels of complexity in the control of gene expression (Cooper, 2000)

1. Cis-Acting Regulatory Elements

CREs are composed of typically non-coding DNA containing binding sites for transcription factors and other regulatory molecules that are needed to activate and sustain transcription (Ong and Corces, 2011). One main step in gene expression is the transcription of DNA into mRNA. The initiation of transcription requires the binding of RNA polymerase to the promoter of the gene which is one of the cis-acting regulatory elements (CREs). The best studied CREs are promoters and enhancers. The promoters are located near the transcription start site (TSS) and are required for transcription in eukaryotes since they carry the binding sites for general transcription factors. For instance, the TATA box is an A/T-rich sequence located approximately 30 nucleotides upstream of the TSS in mammals that serves as a binding sequence for the preinitiation complex (PIC) (Gershenzon et al., 2006; Smale and Kadonaga, 2003). While promoters direct gene

transcription in a position- and orientation-dependent manner, enhancers traditionally function independently of their position and orientation with respect to their target gene, as they can loop over long genomic ranges to engage distant promoters (Plank and Dean, 2014).

2. Transcription factors

Transcription factors (TFs) are the proteins that bind CREs and modulate expression. They bind specific DNA motifs via their DNA binding domains (DBD). The types of TFs, their DBD and their ability to bind specific DNA motifs define their range of action (Stegmaier et al., 2004; Vaquerizas et al., 2009).

2.1. The general transcription machinery:

In eukaryotes, transcription of mRNA starts with the assembly of the PIC at the promoter of the target gene. The PIC includes the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH in addition to RNA polymerase II (pol II). PIC formation usually begins with TFIID binding to the TATA box of the promoter by its TATA box binding protein (TBP) subunit. In some cases, the binding to DNA is assisted by a second factor, TFIIB. The next step is the separation of the two DNA strands by TFIID that contains two helicases. This action is modulated by TFIIE that is considered a loading factor for TFIIH. As for TFIIF, it has been suggested that one of its functions is to facilitate loading of RNAP II into the PIC. TFIIA is not required to assemble a transcriptionally active complex, however, it has been reported to stabilize TBP-DNA interactions (Conaway and Conaway, 1993; Eichner et al., 2010; Holstege et al., 1995; Lagrange et al., 1998; Roeder, 1996; Tirode et al., 1999). This assembly allows transcription at basal levels as transcription can be regulated by other transcription factors able to bind specific DNA sequences (**Figure 1**).

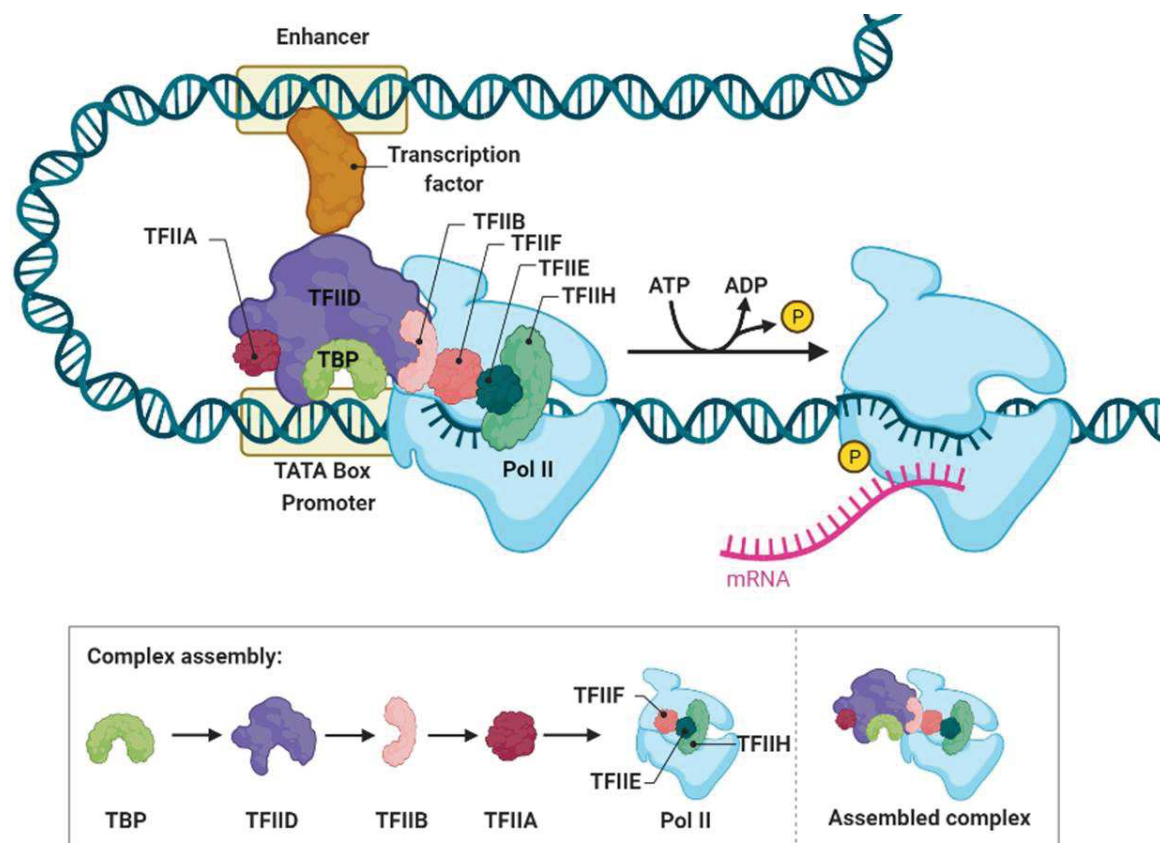


Figure 1: Preinitiation complex (PIC) assembly at the promoter.

The PIC comprises RNA polymerase II and 6 general transcription factors: TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH. TFIID binds the TATA box through its TATA Box binding Protein domain (TBP). Sequence specific transcription factors bind enhancer sequences and contribute to the activation of transcription.

(Modified from “Eukaryotic gene regulation-Transcriptional initiation” BioRender.com 2021).

2.2. Types of transcription factors:

TFs can be divided based on their DBD. The abundance of each type varies between organisms therefore I will mainly focus on the types abundant in metazoans.

2.2.1. Basic superfamily

TFs belonging to this superfamily are able to dimerize and bind DNA in a scissor-type. They belong to two main families: the leucine zipper (bZIP) and helix-loop-helix (bHLH) families. Members of this superfamily are composed of a basic α -helical DNA-contacting region, and a dimerization interface containing a leucine zipper or an HLH. Most members of this superfamily recognize palindromic binding sequences due to their dimerization (Amoutzias et al., 2008; Vinson et al., 1989). Among these TFs we find the bHLH *Drosophila* TF coded by *scute* (*sc*), a proneural gene of the achaete-scute complex (Ruiz-Gómez and Ghysen, 1993).

2.2.2. Zinc finger

This superfamily includes members of the C2H2 (Cys2-His2) zinc fingers. The classical C2H2 domain includes a β -hairpin, followed by an α -helix. The zinc finger structure is stabilized by the coordination of a zinc atom with two cysteine residues at one end of the β sheet and with two histidine residues at the α -helix C-terminus. Most C2H2-containing TFs possess multiple zinc fingers, each of which recognizes specific base pairs (Wolfe et al., 2000).

The GATA family TFs also belong to the zinc finger superfamily as they contain two adjacent zinc fingers. GATA TFs are essential regulators of the specification and differentiation of numerous tissues and their disruption is linked to developmental diseases (Tremblay et al., 2018). For instance, the *Drosophila* GATA factor Serpent (*Srp*) as well as its human ortholog GATA1 are crucial players in hematopoiesis (Ferreira et al., 2005; Rehorn et al., 1996).

2.2.3. The helix-turn-helix superfamily

The helix-turn-helix motif allows these proteins to bind the large groove of DNA. One major family belonging to the group in the homeodomain TFs. This family includes the Hox TFs which are well-known regulators of development (Maconochie et al., 1996). The *Drosophila* glia fate determinant Reverse polarity (Repo) comprises a homeodomain allowing it to bind specific DNA sequences (Halter et al., 1995).

In addition to their DBD, transcription factors also contain activation domains (ADs) that are less well characterized. The AD interacts with components of the preinitiation complex (PIC) to enhance recruitment and stabilization of the general factors at target promoters and regulate transcriptional output (Juven-Gershon and Kadonaga, 2010; Reiter et al., 2017).

3. Epigenetic modifications

The concept of epigenetics was first introduced by Conrad Waddington in 1942. He used the term “epigenotype” to describe the complex of developmental processes that lies between genotype and phenotype and connecting them to each other. Later on, the term “epigenetics”, has been defined as the study of heritable alterations in gene expression that are not caused by changes in DNA sequence (Ct and Morris, 2001; Waterland, 2006). As mentioned above, the regulation of gene expression is the basis of cell differentiation and development. This regulation, that establishes cell-specific gene expression profiles, is acquired during differentiation and must be maintained during cell division. Thus, daughter cells inherit not only the DNA information, but also stable epigenetic modifications. Recently, this definition has become broader and now includes both heritable changes in gene activity and expression, and stable, long-term, alterations in the transcriptional potential of a cell that are not necessarily heritable (Gibney and Nolan, 2010).

Regardless of the exact definition, epigenetic modifications described in today's literature mostly relate to chromatin remodeling, post-transcriptional histone modifications, DNA methylation and RNA-based mechanisms.

3.1. **Chromatin structure and function:**

In eukaryotes, DNA exists in the nucleus of the cell in the form of DNA-protein complex called chromatin. The immense amount of information carried within the DNA needs to be compacted enough to fit into the cell. Yet, despite this compaction, DNA must be rapidly accessible to permit its interaction with different machineries. The balance between compaction and accessibility is regulated at the level of the chromatin.

The principal protein components of the chromatin are the histones, and they are amongst the most highly conserved eukaryotic proteins known. While it was first believed that the purpose of histones was merely to package the DNA of eukaryotes, it was later discovered that archaeobacteria possess a simplified version of a histone used mainly to regulate gene expression (Dinger et al., 2000; Oudet et al., 1975; Ouzounis and Kyripides, 1996; Sandman et al., 1998). Therefore, as suggested by Felsenfeld and Groudine, chromatin might have evolved originally as a mechanism of gene expression regulation and that DNA packaging was an auxiliary benefit that was needed only for the more complex organisms (Felsenfeld and Groudine, 2003). Histones participate in the formation of the smallest subunit of chromatin, the nucleosome, that occurs every 200 +/- 40 base pairs (bp) all along the genome. This subunit consists of 146 bp of DNA wrapped around an octamer of two H3–H4 and two H2A–H2B histone dimers (**Figure 2**). Nucleosomes are separated from each other by 10–80 bp linker DNA associated with linker histone H1 (Arents et al., 1991; Van Holde et al., 1974; Kornberg, 1974; Luger et al., 1997; Wolffe and Matzke, 1999).

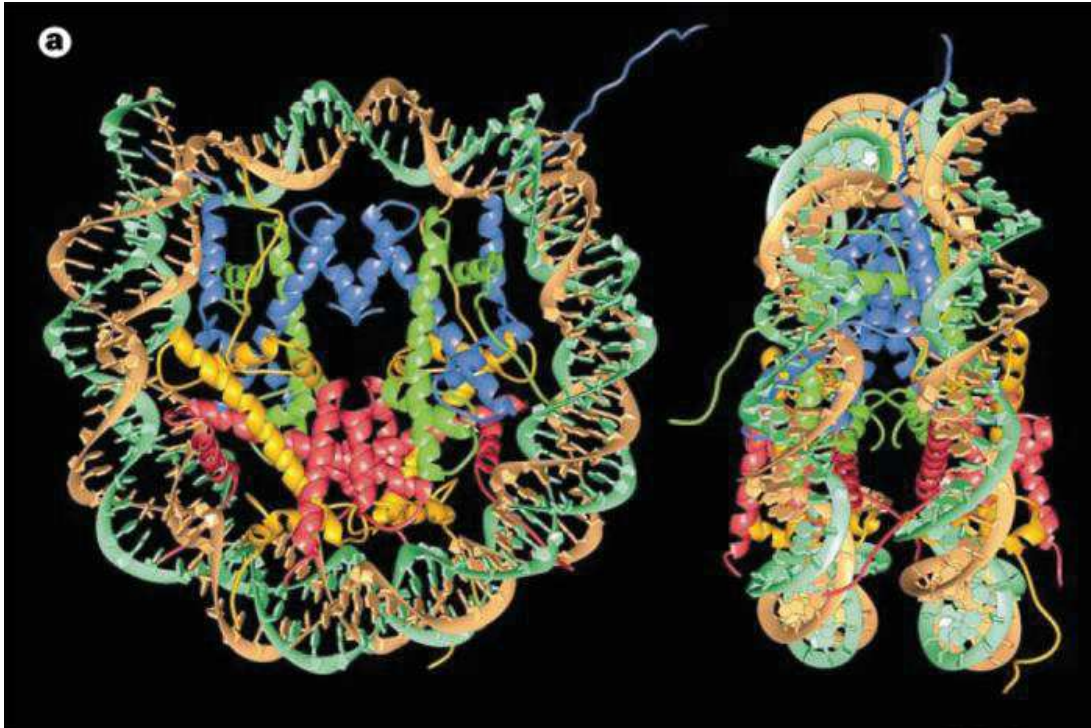


Figure 2: Nucleosome core particle.

Ribbon traces for the 146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone protein main chains (blue: H3; green: H4; yellow: H2A; red: H2B) (Luger et al., 1997).

The DNA-nucleosome complex forms the so called “beads on a string” structure of 11 nm diameter (Olins and Olins, 1974). The 11 nm chromatin fiber has been shown *in vitro* to form a higher-order fiber of 30 nm containing 6–11 nucleosomes per turn, which can be either a one-start solenoid or a two-start zigzag (Finch and Klug, 1976; Woodcock et al., 1984). The 30 nm fiber results in the compaction of DNA by 40 folds. However, this level of packaging is not sufficient to fit DNA into a nucleus. It is suggested that the looping of the 30 nm fiber further will lead to

the formation of an even higher order, the supercoil of 300 nm of diameter to finally create the chromatid (**Figure 3**) (Daban, 2000; Tremethick, 2007; Woodcock and Dimitrov, 2001).

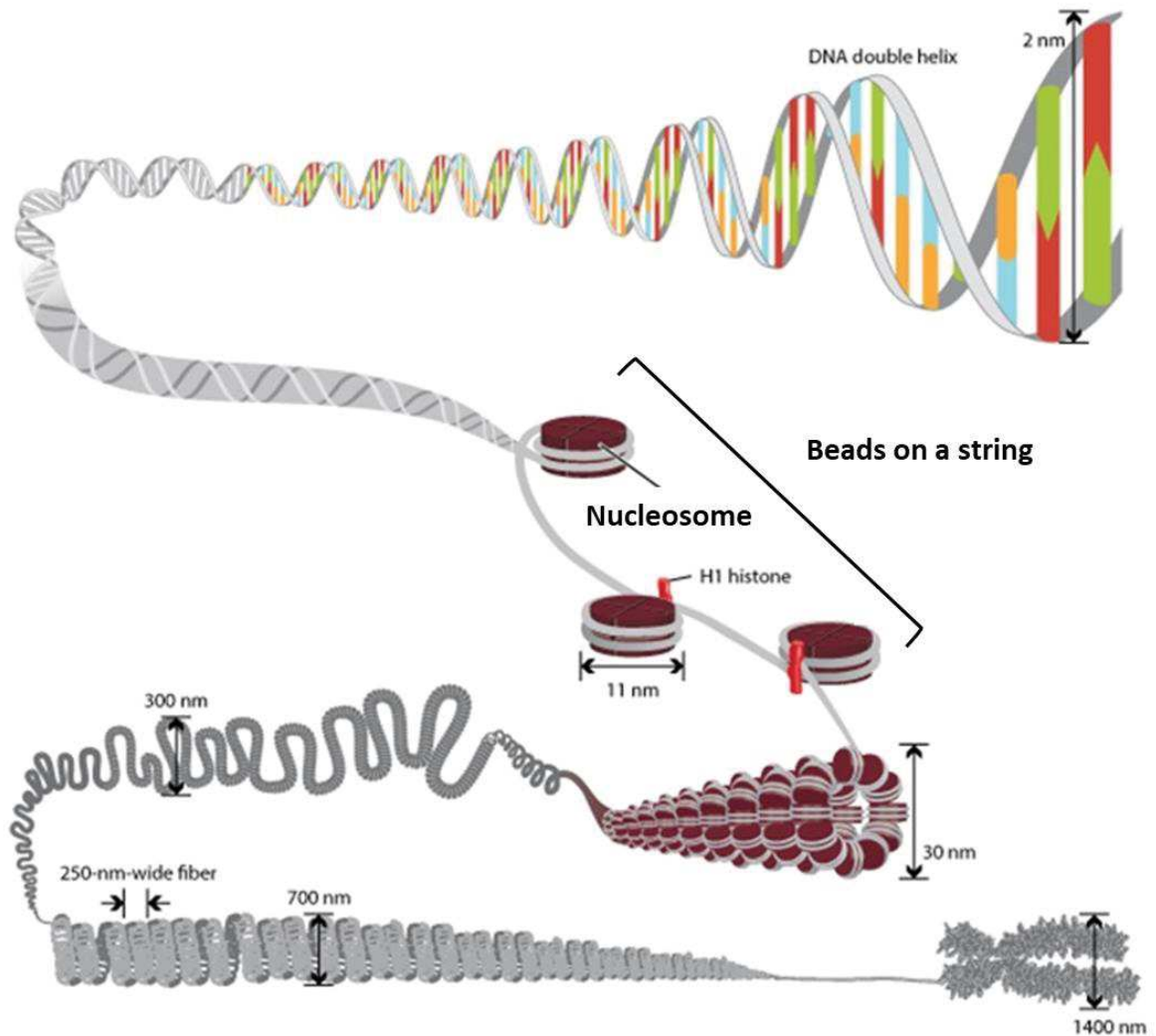


Figure 3: DNA packing in eukaryotic chromosomes.

The different levels of compaction of chromatin: the DNA double helix (2 nm diameter). The beads on a string fiber (11 nm diameter). Tight helical fiber (30 -nm diameter). Histones Supercoil (300 -nm diameter). (Modified from Nature Education: Eukaryotic Genome Complexity 2013).

Based on the level of condensation in the nucleus, chromatin can be divided mainly into two main types: euchromatin and heterochromatin. Euchromatin is decondensed and represents the transcriptionally active form of chromatin. On the other hand, heterochromatin represents regions that are condensed, and it can either be facultative or constitutive (Brown, 1966). Facultative heterochromatin corresponds to silenced euchromatin, while constitutive heterochromatin occurs primarily in large blocks near centromeres and telomeres and consists mostly of repetitive DNA sequences. Our understanding of the relationship between chromatin structure and gene expression regulation started with the discovery of the phenomenon of position-effect variegation (PEV) in *Drosophila* by H.J. Muller in 1930. This highlighted how changing the position of a gene with respect to heterochromatin-euchromatin boundaries can affect its expression and thus provided the first description of a phenomenon with an underlying epigenetic basis (Baker, 1968; LEWIS, 1950). Nowadays, two general ways in which chromatin structure can be altered are of main concern: nucleosome remodeling and post-translational histone modifications.

3.2. Chromatin remodeling:

It is evident that the degree of condensation of chromatin raises the issue of DNA inaccessibility. To counter this issue, chromatin can be remodeled. Adenosine triphosphate (ATP)-dependent chromatin remodeling complexes (CRCs) can alter chromatin structure by changing the location or conformation of nucleosomes. This change may liberate a segment of DNA by changing histone–DNA interactions (Hargreaves and Crabtree, 2011; Kingston and Narlikar, 1999; Workman and Kingston, 1998). Chromatin remodelers are helicase-like proteins that share conserved catalytic core domains and are able to couple ATP hydrolysis to chromatin remodeling. Their action is reversible, nucleosomes can be disassembled or assembled which means they can be involved in either activation or repression and this is largely determined the association of the catalytic cores with auxiliary domains and other subunits within the remodeling complexes (Becker and Workman, 2013; Clapier and Cairns, 2009; Narlikar et al., 2013).

The first ATP dependent chromatin remodeling complex ever identified is switch/ sucrose nonfermenting (SWI/SNF). It was identified in *S. cerevisiae* and considered a general activator of transcription (Laurent et al., 1991; Neigeborn and Carlson, 1984; Peterson and Herskowitz, 1992). One of the SWI/SNF complex subunits, SWI2/SNF2, bears similarity to helicases and was found to have DNA-stimulated ATPase activity (Laurent et al., 1993; Winston and Carlson, 1992). The SWI/SNF family is evolutionarily conserved, and homologous proteins were subsequently identified in flies, and mammals. In *Drosophila* the ortholog of SWI2/SNF2, Brahma (BRM), was found to be needed for the expression of homeotic genes further emphasizing its role in the regulation of gene expression (Elfring et al., 1994; Kennison and Tamkun, 1988). BRM is also part of SWI/SNF-type complexes in mammals, which play important roles in development and cellular homeostasis (Hargreaves and Crabtree, 2011). Several other enzymes were discovered

after that and today we distinguish four subfamilies of chromatin-remodeling enzymes based on their phylogenic relationship and their function: imitation switch (ISWI), chromodomain helicase DNA-binding (CHD), INO80 and SWI/SNF (Becker and Workman, 2013; Clapier and Cairns, 2009; Narlikar et al., 2013).

The function of these different complexes can be summarized with 3 main actions. 1) The assembly and organization of nucleosomes requires assembly remodelers, such as ISWI and CHD, first to assist the formation of the octameric nucleosome and second to space the nucleosomes at relatively fixed distances apart. This actions is usually linked to gene silencing (Corona et al., 1999; Fei et al., 2015; Lusser et al., 2005; Torigoe et al., 2011; Varga-Weisz et al., 1997). 2) Chromatin access is mostly performed by SWI/SNF and involves sliding nucleosomes along the DNA, removing nucleosome components or removing full nucleosomes to expose binding sites on gene promoters for example. This action is mainly linked to gene activation (Boeger et al., 2004; Clapier et al., 2017). 3) Histone replacement involves remodelers of the INO80 family that replace a specific histone with either canonical or a variant histone. This replacement can alter the biophysical properties of chromatin, and fine-tune transcriptional responses by affecting the binding of transcription factor and remodelers (Clapier et al., 2017; Mizuguchi et al., 2004; Ruhl et al., 2006). Together, remodeling complexes act through different mechanisms to affect nucleosome structure. They ensure dense nucleosome packaging and can also be recruited by transcription factors to facilitated access to DNA. Their selective association with cofactors allows specific regulation of genes (**Figure 4**).

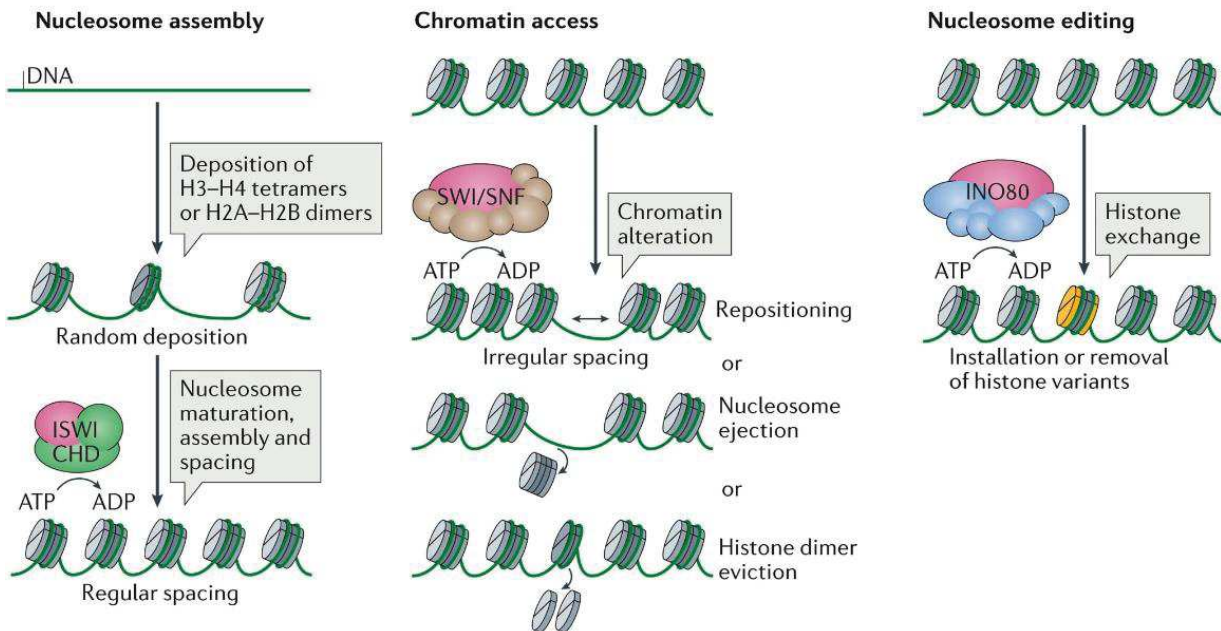


Figure 4: Simplified classification of chromatin remodelers.

Nucleosome assembly is performed by ISWI and CHD subfamily remodelers. They are involved in the deposition of histones, the maturation of nucleosomes and their spacing. Chromatin access is mediated primarily by SWI/SNF subfamily remodelers. They alter chromatin by repositioning nucleosomes, ejecting octamers or evicting histone dimers. Nucleosome editing is accomplished by remodelers of the INO80 subfamily. They modify the composition of nucleosome by exchanging canonical and variant histones (yellow). (Modified from Clapier et al., 2017).

3.3. **Histone modifications:**

While remodelers shuffle the chromatin structure, histones are not mere passive spectators. In fact, histones actively participate in gene expression regulation and are considered major players in this regulation.

In the 1960s while scientists mainly thought that histones were nothing but the glue that holds the DNA together, Vincent G. Allfrey discovered that these proteins are modified post-translationally by the introduction of methyl and acetyl groups. He predicted that these modifications might have a role in the regulation of gene expression (Allfrey et al., 1964). However, these observations remained correlative until other works in yeast proved that histones regulate gene expression. In addition, studies in algae and mammals identified histone modifying enzymes that were related by sequence homology to previously identified transcriptional regulators in yeast (Brownell and Allis, 1996; Han and Grunstein, 1988; Taunton et al., 1996). The mechanism by which histone modifications can regulate gene expression become a little clearer after solving the nucleosome structure. This structure demonstrated the existence of highly basic histone amino (N)-terminal tails that protrude from their own nucleosome. These tails can be modified and thus provide an exposed surface for potential interactions with other nucleosomes or other factors (Luger et al., 1997). The compaction of chromatin into higher-order structures is mediated by the N-terminal tails of histones H3 and H4 through binding to DNA and/or to acidic regions on the neighboring nucleosomes (Dorigo et al., 2003; Luger and Richmond, 1998). These exciting findings opened the doors to a cascade of discoveries of histone modifications that led to the concept of the “language of histone modifications” formed by combinations of different modifications on one or more histone tails (Strahl and Allis, 2000). Even though this concept is

still controversial, it's becoming more and more evident that histone modifications allow the fine tuning of chromatin structure and, subsequently, gene expression (Rando, 2012).

3.3.1. Histone acetylation:

The mechanisms in which histone modifications influence chromatin structure can be currently best explained with histone acetylation that is usually linked to transcriptional activation. Acetylation (ac) of histones involves adding an acetyl group (COCH₃) from acetyl coenzyme A to the positively charged lysine (K) which neutralizes this charge. Consequently, this weakens the interactions between histones and DNA and contributes to a more open chromatin state which in turn can enable the binding of transcription factors for example (**Figure 5**) (Bannister and Kouzarides, 2011). Beyond its action on chromatin structure, lysine acetylation can also be bound by factors such as bromodomain-containing (BRD) transcription factors that induce gene expression (Wang et al., 2007). A concrete example of the influence of acetylation is the histone H4 tail. The H4 tail can directly interact with the neighboring nucleosome (Luger et al., 1997). Lysine acetylation in the H4 tail, in particular the acetylation of the lysine 16 of histone 4 (H4K16ac) has been shown to disrupt interactions between nucleosomes and affect chromatin structure (Allahverdi et al., 2011; Shogren-Knaak et al., 2006). This is due to the neutralization of the positive charge and also to a decreased flexibility in the tail that inhibits interactions (Colleparado-Guevara et al., 2015) thus highlighting new perspectives on how acetylation can affect the chromatin structure (Nitsch et al., 2021).

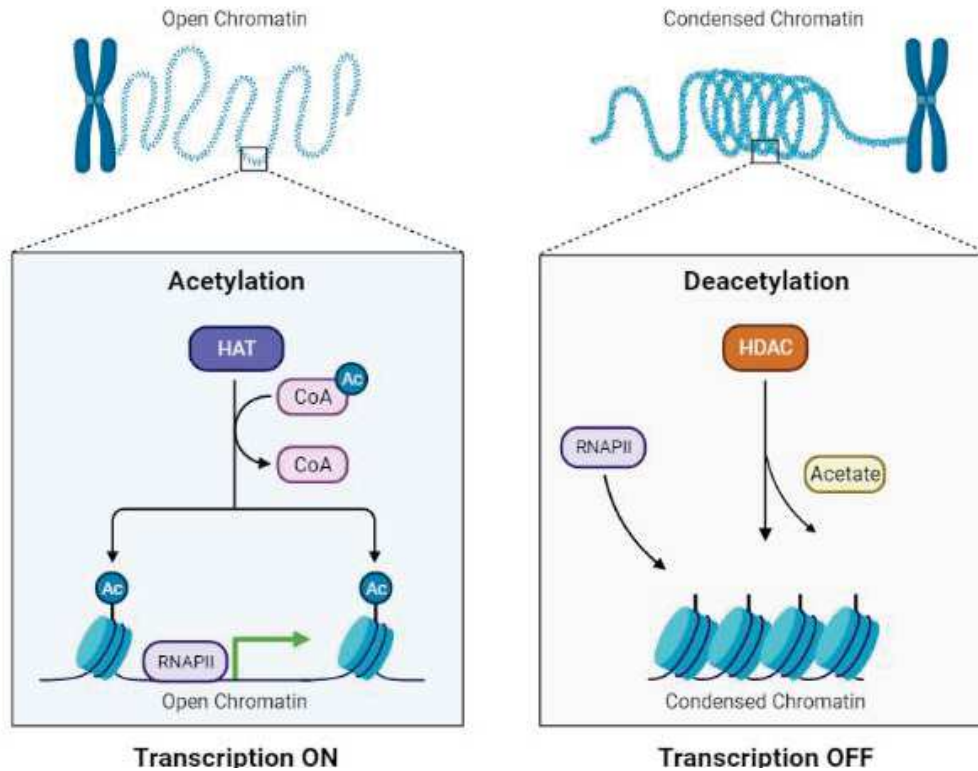


Figure 5: Regulation of chromatin accessibility by histone acetylation.

The histone acetyltransferase (HAT) adds an acetyl group to the histone residue from Co-enzyme A. This opens the chromatin structure, allowing the binding of RNA polymerase II (RNA Pol II). Deacetylation of the histone lysine residues by HDAC leads to condensed inaccessible chromatin. (Modified from “Epigenetics and Gene Expression” BioRender.com 2021).

a) Histone acetyltransferases:

Histone acetylation is a highly dynamic mechanism regulated by histone acetyltransferases (HATs) (Allis et al., 2007; Furdas et al., 2012) and histone deacetylases (HDACs). HATs are also called KATs (lysine acetyltransferases) since they are able to acetylate non-histone proteins (Allis et al., 2007). HATs are classified into two types, type A and type B, based on their localization within the cell. Type A HATs are nuclear and they act on histones already integrated within the nucleosome which means they can affect transcription (Roth et al., 2001). Type B HATs are localized in the cytoplasm and they are shown to acetylate the newly synthesized histones (Brownell and Allis, 1996; Garcea and Alberts, 1980; Sures and Gallwitz, 1980; Wiegand and Brutlag, 1981). Regardless of their localization, HATs are also divided into families based on their catalytic domain (**Table 1**).

Family	HAT	Organisms
GNAT	Gcn5	Sc, Dm, Hs
	PCAF	Hs
	Elp3	Sc, Dm, Hs
p300/CBP	CBP	Dm, Hs
	p300	Hs
MYST	Tip60	Dm, Hs
	Esa1	Sc
	Sas2	Sc
	Sas3	Sc
	MOZ	Hs
	MORF	Hs
	HBO1	Hs
	MOF	Dm, Hs

Table 1: Histone acetyltransferases.

Main histone acetyltransferases (HATs) in *Saccharomyces cerevisiae* (Sc), *Drosophila melanogaster* (Dm) and *Homo sapiens* (Hs) and the families they belong to.

(Marmorstein and Zhou, 2014; Simon et al., 2016)

The GNAT family

The GNAT (Gcn5-related histone N-acetyltransferases) family comprises HATs that show sequence conservation with general control non de-repressible 5 (Gcn5) within their catalytic domain, as they are generally characterized by the presence of a bromodomain (Neuwald and Landsman, 1997). Gcn5 is the first HAT discovered and is highly conserved from yeast to humans (Brownell et al., 1996; Candau et al., 1996; Kuo et al., 1996; Xu et al., 1998). While only one Gcn5 factor exists in yeast and *Drosophila*, vertebrates have a second factor, PCAF (p300/CBP-associated factor) (**Figure 6**) (Yang et al., 1996). However, Gcn5 on its own is not able to acetylate histones within nucleosomes. This action requires the association of Gcn5 with other subunit to form HAT complexes (Grant et al., 1997; Kuo et al., 1996). The two main HAT complexes involving Gcn5 are SAGA (Spt- Ada- Gcn5- acetyl- transferase) and ATAC (Ada two A containing complex) and both have been linked to the regulation of gene expression. SAGA and ATAC are two distinct complexes sharing some common factors including their common HAT module Gcn5/PCAF (Grant et al., 1997; Riss et al., 2015).

Gcn5 association with the members of the alteration/deficiency in activation (ADA) family, in particular Ada2 and Ada3, is needed for its activity (Rössl et al., 2019). In *Drosophila* and humans, Gcn5 is associated to two forms of Ada2, Ada2a that is exclusively found in ATAC and Ada2b that is specific to SAGA (Muratoglu et al., 2003; Spedale et al., 2012). ATAC has been shown to contain an additional HAT domain ATAC2 that *in vitro* is able to acetylate H3 and H4. *Drosophila* ATAC shows strong specificity for histone H4 in nucleosomal substrates, however the specificity of ATAC is still under debate (Guelman et al., 2009; Suganuma and Workman, 2008). SAGA, containing only Gcn5, seems to preferentially acetylate H3K9 and H3K14 (Grant et al., 1999; Helmlinger et al., 2021). All these findings suggest that these two complexes have distinct

functions. However, it is still unclear to which extent their functions rely on their shared HAT activities.

SAGA, ATAC and Gcn5 in particular are involved in many mechanisms in development and disease. In *Drosophila*, loss of Gcn5 causes oogenesis arrest at stage 5 and 6 and zygotic *gcn5* mutation block metamorphosis. Gcn5 is also necessary for cell proliferation in wing imaginal discs and its mutation induces abnormal pupae development (Carré et al., 2005). Interestingly, both loss of the SAGA-specific Ada2b and ATAC-specific Ada2a lead to developmental lethality indicating that the two HAT complexes are needed for the fly's development (Pankotai et al., 2005; Qi et al., 2004).

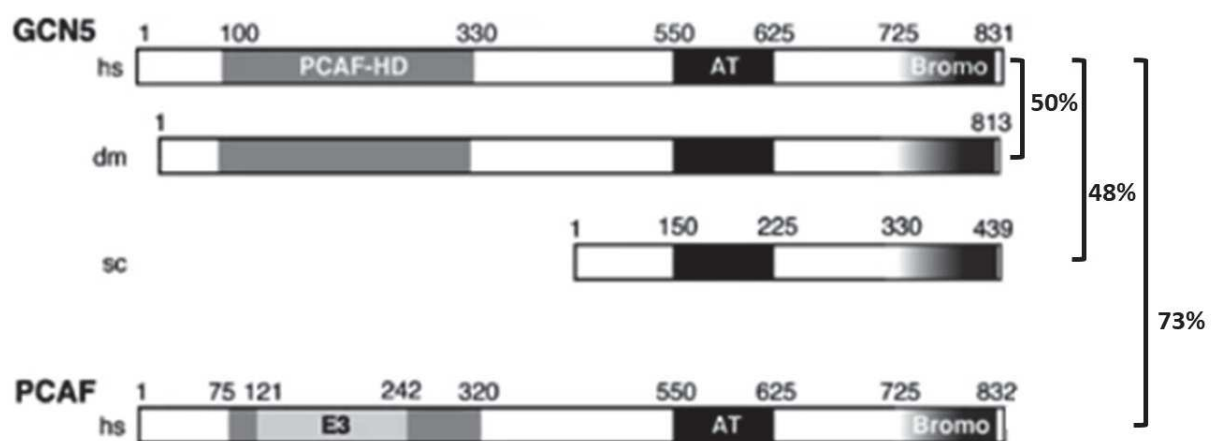


Figure 6: The overall structure of the GCN5 and PCAF enzymes in humans, *Drosophila* and yeast.

Representation of GCN5 and PCAF in humans (hs; *Homo sapiens*), *Drosophila melanogaster* (dm) and yeast (sc; *Saccharomyces cerevisiae*). The PCAF homology domain (PCAF-HD) is shown in grey, the AT domain is shown in black and the bromo domain (Bromo) is shaded. ubiquitin E3 ligase domain (E3) of PCAF is also indicated. The numbers over the boxes indicate amino-acid positions. The identity between the different factors is indicated in % on the right representing the pair wise comparisons. AT, acetyl transferase.

(Modified from Nagy and Tora, 2007)

Similarly in mammal, both SAGA and ATAC complexes are required for normal embryonic development, and Gcn5 deletion induces embryonic lethality (Bu et al., 2007). The role of these complexes goes beyond development. Studies have linked ATAC to different cancers such as lung, liver, breast, and prostate (Mi et al., 2017; Mustachio et al., 2020). In addition to cancer, SAGA is involved in the etiology of neurodegenerative diseases, such as spinal cerebellar ataxia type 7 characterized by cerebellar and retinal degeneration (Helmlinger et al., 2006; Palhan et al., 2005; Yoo et al., 2003).

Other members of the GNAT family include PCAF, HAT1 and ELP3. PCAF shares 73% homology with Gcn5 and the two HATs are incorporated in SAGA and ATAC in a mutually exclusive fashion. They are considered redundant even though some studies indicate synergistic and even distinct functions (Xu et al., 2000; Yamauchi et al., 2000). HAT1 belongs to the type B HATs. It is known to acetylate newly synthesized histones before their incorporation into nucleosomes and seems to be implicated in the telomerase regulation. Finally, ELP3 (Elongation Protein 3) is the catalytic histone acetyltransferase subunit of the RNA polymerase II elongator complex (Mersfelder and Parthun, 2008; Verreault, 2000; Wittschieben et al., 1999).

The P300/CBP family:

The cAMP response element-binding (CREB) protein binding protein (CBP) and p300 are ubiquitous and critical regulators of transcription. They can form larger protein complexes that serve as coactivators of different transcription factors. For instance, CBP binding to CREB has been shown to influence the expression of genes involved in circadian activity (Chatterjee et al., 2020). CBP and p300 can either utilize their intrinsic HAT activity or function as transcriptional

adaptors that link activators to the basal transcription machinery (Bannister and Kouzarides, 1996; Grant and Berger, 1999; Kimura et al., 2005; Shikama et al., 2003).

The MYST family

The MYST family is named after the founding members: MOZ, YBF2/SAS3, SAS2 and Tip60 that share a sequence similar to acetyl- CoA motif in an evolutionary conserved region, the MYST domain (Avvakumov and Côté, 2007; Pillus, 2008). MOZ (Monocytic leukemia zinc finger protein) is involved in development and tumorigenesis process (Yang and Ullah, 2007). Sas2 and Sas3 are involved in long- range gene repression and in maintaining the structure of the heterochromatin of telomeres in yeast (Carrozza et al., 2003). *Drosophila* MOF is involved in hyperacetylation of male X chromosome which leads to higher expression in order to compensate the second chromosome present in females (Kimura et al., 2005; Kind et al., 2008). Tip60 is implicated in DNA repair, cell cycle regulation and apoptosis (Sapountzi et al., 2006).

In addition to the three families, mentioned above, many other proteins with a HAT activity have been identified such as TFIID, the HAT catalytic subunit of TFIID is TBP- associated factor- 1 (Taf1, Taf250 in yeast) (Carrozza et al., 2003; Nagy and Tora, 2007).

b) Histone deacetylases:

As mentioned earlier, histone acetylation is a highly dynamic mechanism that regulates gene expression. Besides the HATs that catalyze the modification we find HDACs that remove acetylation. HDACs are divided into two families based on their deacetylase domain and their dependence on cofactors: the zinc dependent histone deacetylase family and the nicotinamide adenine dinucleotide (NAD) dependent sirtuin family (Yang and Seto, 2003, 2008). Members of the classical HDAC family fall into three different phylogenetic classes: class I (HDAC1, 2, 3, and

8), class II (HDAC4, 5, 6, 7, 9, and 10), and class IV (HDAC11). The Human class II enzymes are further divided into class IIa and class IIb, because HDACs of class IIa contain a long N- terminal and possess a low deacetylase activity (Yang and Seto, 2008) and according to their domain composition (**Table 2**).

Class	<i>S. cerevisiae</i>	<i>D. melanogaster</i>	<i>H. sapiens</i>
Class I	Rpd3	HDAC1	HDAC1
	Hos1	HDAC3	HDAC2
	Hos2		HDAC3
	Hos3		HDAC8
Class IIa	Hda1	HDAC4	HDAC4
			HDAC5
			HDAC7
			HDAC9
Class IIb		HDAC6	HDAC6
			HDAC10
Class III	Sir2	Sirt1	SIRT 1
	Hst1	Sirt2	SIRT 2
	Hst2	Sirt4	SIRT 3
	Hst3	Sirt6	SIRT 4
	Hst4	Sirt7	SIRT 5
			SIRT 6
			SIRT 7
Class IV		HDAC11	HDAC11

Table 2: Histone deacetylases.

Main histone deacetylases (HDACs) in *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens* and the classes they belong to.

(Feller et al., 2015; Seto and Yoshida, 2014)

Class I HDACs

The class I HDACs are closely related to the *S. cerevisiae* transcriptional regulator Rpd3 and present a highly conserved catalytic domain (Khochbin et al., 2001). They are ubiquitously expressed and mostly found in the nucleus. They show the strongest histone deacetylase activity (Seto and Yoshida, 2014) and are also able to deacetylate non-histone proteins such as TFs (Ito et al., 2002; Martínez-Balbás et al., 2000; Yao et al., 2001). The most studied class I HDACs are HDAC1 and HDAC2, only HDAC1 exists in *Drosophila*. They present the catalytic subunits of several conserved transcriptional repressor complexes such as Sin3 (switch independent3), NuRD (Nucleosome Remodeling and Deacetylase) and CoREST (co-repressor for element-1-silencing TF) (Knoepfler and Eisenman, 1999; You et al., 2001). In *Drosophila*, the Sin3 complex is found in non-condensed chromatin suggesting that it is required to repress specific genes within transcriptionally active chromatin domains (Pile and Wassarman, 2000). In addition, HDAC1 is important for cell proliferation during development and its disruption affects growth and development in the wing imaginal disc (Barnes et al., 2018; Pile and Wassarman, 2000; Swaminathan and Pile, 2010). In yeast and human, Sin3 can repress or activate gene expression. While its role in repression is mostly due to deacetylation, the mechanism for activation is less clear and may be HDAC independent (Baymaz et al., 2015; Kadamb et al., 2013).

The NuRD complex is a CHD-class complex (Chromodomain, Helicase, DNA binding domain). It is one of the only two known complexes coupling two independent chromatin-regulating activities: deacetylation through HDAC1/2 and ATP-remodeling through Tip60/p400. In *Drosophila melanogaster*, it was shown that Tramtrack (Ttk), a known neuronal repressor and glial fate promoter, can recruit NuRD (Reddy et al., 2010) suggesting that Ttk69 may use this mode of action to repress the neuronal genes (Murawsky et al., 2001).

The CoREST complex is a chromatin-modifying co-repressor complex, originally described as a corepressor of REST (RE1-silencing transcription factor). It regulates neuronal gene expression and neuronal stem cell fate (Andres et al., 1999; Dallman et al., 2004; Lakowski et al., 2006).

Class II HDACs

HDAC4, 5, 6, 7, 9 and 10 share similarities with the fission yeast Hda1 and belong to the Class II HDACs (Bjerling et al., 2002). They are commonly found in the nucleus and the cytoplasm and they are part of complexes that modulate the repressive effect of TFs such as Myocyte enhancing factor 2 (MEF2) (Khochbin et al., 2001; Yang and Grégoire, 2005).

Class III HDACs

The sirtuin proteins make up the class III HDACs, they share homology with the yeast transcriptional repressor silent information regulator 2 (Sir2) and they require NAD as a cofactor for their catalytic function. They are involved in establishing and maintaining the repressive structure of telomeres. In addition, Sir2, is involved in maintaining genome integrity and the process of DNA repair (Gottschling et al., 1990; Palladino et al., 1993; Tissenbaum and Guarente, 2001).

Class IV HDAC

Class IV (HDAC11) was the most recently discovered. It shows homology with the yeast Hos3 and shares a catalytic domain with class I and class II HDACs. However, it does not present enough similarity with other HDACs to be included in class I or class II nor with the sirtuin family (Gao et al., 2002). HDAC11 regulates immune activation and tumorigenesis, yet its biochemical function is largely unknown (Deubzer et al., 2013; Li et al., 2016; Sahakian et al., 2015).

The disruption of the balance between the activities of HATs and HDACs can result in the aberrant expression of genes that ultimately leads to the instability of chromatin structure and the onset of disease such as cancer and neurodegeneration (Di Gennaro et al., 2004; Vahid et al., 2015). Altered HAT-mediated pathways were found in numerous malignant tumors which raised the question of their involvement in the etiology of cancers (Lin et al., 1998; Marks et al., 2001). Acetylation can cause oncogenic transformation by: 1) abnormal over-recruitment of HDACs on tumor suppressor genes; 2) reduced activity of HATs that would lead to reduced activation of tumor repressors; 3) increased HAT activity at oncogenes. These deregulations can alter key physiological mechanisms such as proliferation and the cell cycle, differentiation and apoptosis (Di Cerbo and Schneider, 2013). Acute Promyelocytic Leukemia (APL) was the first model disease in which the involvement of HDACs was demonstrated. This form of Leukemia is characterized by the accumulation of immature cells (Olsson et al., 1996). Studies showed that combining the treatment of APL with HDAC inhibitors (HDACi) results in the re-initiation of the differentiation program of the leukemic cells, which then die by apoptosis (Minucci et al., 2001). HDACi mainly induce growth arrest, differentiation or apoptosis *in vitro* and *in vivo* which makes them a target for a general therapeutical strategy for cancer. Another example is neurodegenerative diseases. Histone hypoacetylation in the nervous system (NS) is a feature of multiple models of neurodegenerative diseases (Bennett et al., 2019; Konsoula and Barile, 2012; Saha and Pahan, 2006). In addition, accumulating evidence indicates that decreased histone acetylation is a negative determinant for neuronal survival (Lazo-Gómez et al., 2013; Valle et al., 2014). Therefore, HDACi were again considered as potential therapeutic targets for a wide range of neurodegenerative disorders, such as Alzheimer's disease, spinal muscular atrophy and amyotrophic lateral sclerosis (Avila et al., 2007; Benito et al., 2015; Majid et al., 2015; Rossaert et al., 2019; Sugai et al., 2004).

3.3.2. Histone methylation:

Unlike histone acetylation that is generally associated with active transcription, histone methylation has unique roles based on the methylated residue and the number of methyl groups. Histone methylation occurs most commonly on the arginine (R) or K residues of histones.

a) Arginine methylation:

In histones H3 and H4 arginine residues can be mono-, di-, or trimethylated. This methylation is catalyzed by protein arginine methyltransferases (PRMTs) and is involved in the regulation of gene expression. For instance, the di-methylation of H4R3 by PRMT1 is generally associated with active transcription, as it promotes histone acetylation through recruiting the p300/CBP-associated factor complex (Bedford and Clarke, 2009). Moreover, the CARM1 (coactivator-associated arginine methyltransferase) catalyzed methylation of H3R17, H3R26, and H3R42 has been associated with reduced binding of NuRD and thus activation of transcription (Yang and Bedford, 2013). On the other hand, H4R3 di-methylated by PRMT5 has been shown to repress transcription by antagonizing H3K4me₃, the hallmark of active transcription (Blanc and Richard, 2017; Guccione et al., 2007; Neault et al., 2012).

b) Lysine methylation:

Histone Lys methylation can also exist in one of three states: mono-, di- or tri-methylation. H3 can be methylated on the following residues: K4, K9, K27, K36 or K79. Unlike acetylation, methylation may characterize active or silent chromatin. In fact, mono- di- and tri-methylation of H3K4 and H3K79, as well as H3K36me₃ are linked to transcriptional activation, with H3K4me₃ being on of active promoters (Barski et al., 2007; Bernstein et al., 2002, 2005) while H3K36me₃

and K3K79me₃ are usually found over gene bodies (Bannister et al., 2005; Wang et al., 2008). H3K4 can also be monomethylated, this form is only found in enhancers (Heintzman et al., 2007).

H3K9 and H3K27 methylation are associated with gene repression but serve different functions (Kouzarides, 2007). H3K9me₃ is involved in constitutive heterochromatin, while H3K9me₂ is found on silent genes in euchromatin and H3K9me is linked to active transcription. H3K27me₃ is considered an easily reversible repression mark needed for dynamic regulation of genes and is most known for the silencing of Hox genes and the maintenance of inactive X chromosome in mammals while H3K27me is found on active gene bodies (Kouzarides, 2007; Peters et al., 2003; Rea et al., 2000; Rice et al., 2003; Trojer and Reinberg, 2007).

The impact of methylation on gene expression depends on the specificity and the number of added methyl groups, which in turn depends on the methylating enzyme and its cofactors. In the next section I will discuss the writers, erasers, readers and roles of three main histone methylations: H3K4me₃, H3K9me₃ and H3K27me₃.

H3K27 methylation

H3K27me₃ is a histone mark for Polycomb (Pc)-mediated genomic silencing and transcriptional repression and is associated with animal body patterning, X-chromosome inactivation, genomic imprinting, and stem cell maintenance (Moss and Wallrath, 2007; Plath et al., 2003; Sparmann and van Lohuizen, 2006). Polycomb group (PcG) proteins act in concert to epigenetically repress regulatory genes and enforce cell-type specific gene expression programs (Schwartz and Pirrotta, 2007). One of the complexes of PcG, PRC2 (Polycomb Repressive Complex 2), is responsible for the methylation of H3K27. This action is catalyzed by EZH2 (Enhancer-of-zest Homolog 2), known as E(z) in *Drosophila*, the catalytic subunit of PRC2

(Czermin et al., 2002; Müller et al., 2002). E(z) alone has no enzymatic activity, it requires incorporation into PRCs with other subunits to methylate H3K27. For instance, its association with Extra sex combs (Esc) (EED in mammals) and suppressor of zeste (Su(z)12) (SUZ12 in mammals) is needed to recruit the complex to nucleosomes (Cao and Zhang, 2004; Nekrasov et al., 2005).

Demethylation of H3K27me₂ and H3K27me₃ is mediated by Utx, in mammals KDM6, and is primarily linked to gene de-repression. In *Drosophila* Utx was found to suppress Notch-dependent and Retinoblastoma-dependent tumors. UTX was also found in complex with acetyltransferase CREB-binding protein (CBP) and chromatin remodeler Brm, where they antagonize Pc silencing (Herz et al., 2010).

H3K27me₃ is recognized by the chromodomain of Polycomb (Pc) of PRC1, an essential complex for PcG repression (Lee et al., 2007; Schwartz and Pirrotta, 2013; Shao et al., 1999). In mammals, PRC1 has been found to inhibit chromatin remodeling by SWI/SNF and restrict the access of RNA Pol II *in vitro* (King et al., 2002; Shao et al., 1999). PRC1 is also able to ubiquitinate H2A-K118 and H2A-K119 which seems indispensable for the silencing of target genes (Endoh et al., 2012; Gutiérrez et al., 2012; Wang et al., 2004). Even though the exact mechanism by which PRC1 silences gene expression is still not well understood, it has been considered as a potential therapeutic target due to its involvement in tumorigenesis, stem cell self-renewal and differentiation. Indeed, inhibiting PRC1 binding to H3K27me₃ derepresses PRC1 target genes and inhibits the proliferation of prostate cancers (Ren et al., 2015).

H3K4 methylation

H3K4 methylation is a key modification in gene activation. All three forms, mono, di and tri methylated H3K4, have been shown to differentially mark actively transcribing genes.

H3K4me1 is highly enriched at enhancers, H3K4me2 is highest toward the 5' end of transcribing genes, and H3K4me3 is the hallmark of active gene promoters as well as poised genes (see below) (Heintzman et al., 2007; Kim and Buratowski, 2009; Mikkelsen et al., 2007; Ruthenburg et al., 2007; Santos-Rosa et al., 2002). The strong correlation between H3K4 methylation and transcription has become evident. However, the exact mechanism by which lysine methyltransferase (KMTs) are recruited to target genes is still not understood, although it seems to be dependent on the action of specific TFs, such as p53, that recruit KMTs (Narayanan et al., 2007; Tang et al., 2013; Yokoyama et al., 2004).

KMTs are highly conserved from yeast to humans. In yeast, H3K4 methylations are carried out by Set1, which forms a complex known as COMPASS (complex of proteins associated with Set 1). In mammals, there are at least six SET1-related proteins that form COMPASS-like complexes with each having specific functions (Briggs et al., 2001; Miller et al., 2001; Shilatifard, 2012). In *Drosophila*, four SET domain-containing proteins, Trithorax (Trx), Trithorax-related (Trr), Set1, and Ash1, have been reported. Trx, Trr and Set1 are involved in H3K4 methylation (Eissenberg and Shilatifard, 2010) while Ash1, was later shown to methylate H3K36 (Tanaka et al., 2007; Yuan et al., 2011). dSet1 was identified based on sequence homology to the *S. cerevisiae* (Wu et al., 2008; Yokoyama et al., 2007). Trx counteracts the repression mediated by PcG and maintains activation of its target genes throughout development (Ringrose and Paro, 2004; Schuettengruber et al., 2007). Trr shares sequence similarity with Trx but was shown to function in the regulation of hormone-responsive gene expression (Sedkov et al., 2003). An interesting finding highlighting the importance of the relationship between the different histone tails showed that H3K4 methylation is dependent on the ubiquitylation of lysine 123 of H2B (Dehé et al., 2005; Dover et al., 2002; Sun and Allis, 2002). This happens through direct activation of H3K4 KMTs

by the H2B ubiquitin (Ub) (Holt et al., 2015; Kim et al., 2013; McGinty et al., 2008). However, the exact mechanism by which H2Bub activates COMPASS remains a mystery.

H3K4 demethylases (KDMs) were only identified relatively recently, the first being LSD1, Su(var)3-3 (Suppressor of variegation 3-3) in the fly, is able to demethylate H3K4me1 and H3K4me2 (Shi et al., 2004). In fact, Su(var) genes encode structural components of heterochromatin. Su(var)3-3 initiates and maintains heterochromatin formation during embryonic development, and establishes transcriptional silencing in the germline pole cells (Rudolph et al., 2007; Schotta et al., 2003). Interestingly, Su(var)3-3 was also found to associate with RPD3, heterochromatin protein 1 (HP1) and Su(var)3-9, which controls heterochromatin spreading in position-effect variegation (see below). In addition, Su(var)3-3 has also been identified as a protein in the Brahma (Brm) complex, involved in chromatin remodeling during development, as well as co-repressor complexes such as CoRest (Curtis et al., 2011; Lee et al., 2005; Tamkun, 1995). Moreover, Su(var)3-3 has been found to functionally interact with the histone KDM Little imaginal discs (Lid), JARID1 in humans, a KDM able to demethylase H3K4me2 and H3K4me3 (Christensen et al., 2007; Iwase et al., 2007; Di Stefano et al., 2011).

Unlike histone acetylation, the effect of methylation on chromatin is not based on a neutralized charge but on its ability to recruit “reader” factors that are able to modify chromatin and regulated gene expression. For instance, several ATP-dependent chromatin remodelers have been shown to interact with H3K4 methylation through their chromodomains such as CHD1 in humans or bromodomains such as BPTF, a member of the chromatin remodeling complex NURF (nucleosome remodeling factor) (Sims et al., 2005; Wysocka et al., 2006). Another way in which H3K4 methylation affects expression is by modulating the activity of histone modifying enzymes. For example, the SAGA complex is recruited to H3K4me2 or H3K4me3 sites via Sgf29 to

acetylate neighboring histones and thus modulating gene expression (Bian et al., 2011). Besides chromatin remodeling and histone modifiers, H3K4 methylation also influences preinitiation complex formation. In fact, H3K4me₃ was shown to recruit the general TFs TFIID that recognizes H3K4me₃ through its PHD domain-containing TAF3 subunit, resulting in more efficient preinitiation complex formation (Lauberth et al., 2013; Vermeulen et al., 2007).

Mutations in H3K4 methyltransferases highly increase the susceptibility to various cancers and these mutations have been found in mixed-lineage leukemia, acute lymphoblastic leukemia and acute myeloid leukemia which makes them important candidates for therapeutic approaches (Rao and Dou, 2015). Deregulation of H3K4 demethylation is also linked to disease. For instance, mutations in the corepressors belonging to the JARID1 family have been shown to be involved in pediatric acute leukemia, breast carcinomas and testicular cancer and X-linked mental retardation (Jensen et al., 2005; de Rooij et al., 2013; Yamane et al., 2007).

As mentioned above, H3K4me₃ is also a mark of poised genes. In these genes, both the activation mark H3K4me₃ and the repressive mark H3K27me₃ are found, they are said to be in a bivalent state. This state is important for the regulation of changes in gene expression from poised to active or inactive in embryonic stem cells (ESCs) and is responsible for differentiation into specific cell types (Barski et al., 2007; Mikkelsen et al., 2007; Vastenhouw and Schier, 2012).

H3K9 methylation

H3K9 methylation is a histone modification found on silenced genes and heterochromatin (Barski et al., 2007). Fission yeast have one H3K9 methyltransferase (Clr4/KMT1) that is responsible for all three states of H3K9 methylation (Nakayama et al., 2001). In mammalian cells, several H3K9 methyltransferases exist: SUV39H1, SUV39H2, SETDB1, G9a-GLP and the

PRDM family (Brower-Toland et al., 2009). SUV39H1 and SUV39H2 deposit H3K9me2 and H3K9me3 in constitutive heterochromatin, including the pericentromeric region (Lachner et al., 2001; Rea et al., 2000). G9a-GLP are responsible for H3K9me and H3K9me2 that repress gene expression in euchromatin. G9a-GLP is recruited to targets through direct interactions with different DNA-binding proteins (Duan et al., 2005; Nishida et al., 2007; Ogawa et al., 2002; Ueda et al., 2006) Once it methylates H3K9, G9a-GLP recruits additional dimers to spread H3K9me1 and H3K9me2 to neighboring nucleosomes (Tachibana et al., 2002). The G9a-GLP mediated gene repression is crucial in different biological processes, such as memory formation, immune responses and differentiation (Antignano et al., 2014; Gupta-Agarwal et al., 2012; Liu et al., 2015). Numerous studies have revealed that diverse biological processes are regulated by H3K9 methylation. However, only a few downstream target genes, and no upstream regulators of H3K9 methyltransferases, have been identified. Clarifying the detailed molecular mechanisms by which H3K9 methyltransferases mediate transcriptional repression will require a greater effort to identify upstream regulators that bring H3K9 methyltransferases to target genes. In the case of PRDM, it is still not known whether they regulate H3K9 methylation directly or indirectly.

In *Drosophila*, *Su(var)3-9*, along with *Su(var)25* and *Su(var)37* were identified as dominant modifiers of PEV. *Su(var)3-9* was shown to di- and trimethylate H3K9 at the chromocenter and is therefore considered as the heterochromatic H3K9 methyltransferase (Ebert et al., 2004; Schotta et al., 2003). *Su(var)25* codes for heterochromatin protein1 (HP1) that genetically and physically interacts with *Su(var)3-9* to mediate the formation of constitutive heterochromatin. HP1 is recruited to the sites containing H3K9me3 and in turn recruits *Su(var)3-9* that further methylates H3K9 thus mediating the spreading of heterochromatin. In fission yeast and plants, it has been found that the mechanism of heterochromatin formation involves the RNA

interference machinery (Grewal, 2010; Tamaru, 2010). However, the molecular mechanisms of H3K9me3 targeting and recruitment are not fully understood (Bannister et al., 2001; Canzio et al., 2013; Hiragami-Hamada et al., 2016; Lachner et al., 2001). Besides its interaction with Su(var)3-3, the H3K4 demethylase, Su(var)3-9 was also shown to associate with HDAC1, suggesting a mode of repression that start with deacetylation followed by the methylation of target residues (Czermin et al., 2001; Schotta et al., 2003).

The fly's Set1B mono- and dimethylates H3K9 on chromosome 4 where it is required for silencing of variegating transgenes. SETDB1 is critical for normal progression of oogenesis where it controls germline stem cell maintenance and differentiation (Clough et al., 2007, 2014; Wang et al., 2011).

Unlike in mice, depletion of G9 ortholog in *Drosophila* called dG9 or Egg, does not impact viability but was shown to mediate H3K9me3 in the germarium of the ovary and its loss of function (LOF) caused oogenesis arrests at very early stages (Clough et al., 2007; Seum et al., 2007). However, studies suggest that the apparent dispensability is an artifact of studying flies in the context of a laboratory. In other words, the optimal conditions of the lab shield the flies from much of the environmental related stress encountered in the wild. Indeed, it was reported that Egg plays a crucial role in acquiring tolerance to starvation stress by maintaining energy storage during starvation (An et al., 2017; Shimaji et al., 2017).

In mammals, three classes of KDMs present H3K9 demethylase activity: JHDM2 (jumonji domain-containing histone demethylase-2) /KDM3, JHDM3(JMJD2)/KDM4 and PHF8/KDM7. Two JHDM2 proteins have been shown to demethylate H3K9me1 and H3K9me2 and regulate hormone-dependent transcriptional activation. JHDM3 family proteins can demethylate H3K9me2

and H3K9me3 as well as H3K36me2 and H3K36me3 (Cloos et al., 2006; Okada et al., 2007; Yamane et al., 2006). PHF8 also demethylates H3K9me1 and H3K9me2 however it preferably demethylates regions containing H3K4me3 which might contribute to the mutually exclusive distribution of the three marks (Horton et al., 2010; Loenarz et al., 2010). In *Drosophila*, Kdm3 is able to demethylate H3K9me3 and is considered a suppressor of PEV necessary for heterochromatin regulation (Herz et al., 2014). Two members of the KDM4 family have been characterized, Kdm4A and Kdm4B. Unlike in mammals, the fly's Kdm4A is only able to demethylate H3K36 and not H3K9 (Klose et al., 2006; Lin et al., 2008).

Deregulation of H3K9 methylation is linked to neurodegenerative diseases such as Alzheimer disease, in a mouse model, and Huntington disease in patients (Ryu et al., 2006; Walker et al., 2013). In addition, similarly to HDACs, KMTs have also been linked to cancer through their repression of tumor repressor genes (Wozniak et al., 2007).

Methylation and acetylation are not the only post-translational histone modifications. For instance, histones can also be ubiquitinated, phosphorylated and sumoylated. In addition, Histone modifications are not the only epigenetics modifications modulation gene expression. DNA methylation is also a key player in the regulation of gene expression in mammals however its role in *Drosophila* is still controversial due to its very faint presence (Deshmukh et al., 2018; Guan et al., 2019; Lian et al., 2018; Takayama et al., 2014). Moreover, RNA-based mechanisms have been the center of many publications highlighting the extent of their implication in gene expression regulation. Non-coding RNAs can regulate expression at the level of the gene and the level of chromosome to control cell differentiation (Costa, 2008; Ghildiyal and Zamore, 2009; Moazed, 2009). On the other hand, there is whole world of gene expression regulation beyond transcription. For instance, chemical modifications of RNA, called Epi-transcriptomics, emerged as key players

in the post-transcriptional regulation of gene expression (Kumar and Mohapatra, 2021). However, these mechanisms are beyond the scope of my thesis and therefore will not be discussed further.

3.4. Cell-specific epigenetic signatures

Extensive studies on epigenetic modifications in stem cells and their role in cell fate determination highlight the powerful regulation mediated by these marks (Boland et al., 2014; Wu and Sun, 2006). However, the genome-wide chromatin patterns of differentiated cells and the extent to which the epigenome differs between distinct cell types, remain poorly defined. Recently, studies have highlighted the tight coordination between TFs and epigenetic marks that is able to either activate or silence specific genes, in distinct cell types. For example, Hey is a crucial transcription factor in mammalian heart development. Upon assessing the regulatory mechanisms by which it affects gene expression in a cell type specific manner, it was found that Hey represses genes via HDAC recruitment and histone deacetylation (Weber et al., 2003). Along the same line, another study on the mouse retina showed that H3K4me2 and H3K27me3 signatures are specific for functional groups of genes and can distinguish cell-specific genes from widespread transcripts (Popova et al., 2014). It seems evident that the interplay between TFs and epigenetic modifications controls the expression profile of differentiated cells. However, does this mean that different cell types present a specific epigenetic signature and that their identity might be encoded in their epigenome? It has become clear that in order to understand cell identity we must decipher the underlying epigenetics.

III- The model of *Drosophila melanogaster*:

It has been more than a century that the fruit fly was first employed by biologists. For the past 100 years it has proven to be a valuable asset in many different fields of biology. Starting with the white gene and genetics, the Hox genes and development, toll and immunity all the way to period and behavior, the fly has seen it all. *Drosophila* present 60% genome homology with humans, with less redundancy. In addition, 75% of the genes responsible for human diseases have orthologs in flies (Ugur et al., 2016). Therefore, the fly is a good model to study molecular mechanisms underlying development and disease. Beside the degree of conservation, its short life cycle (see below) and the availability of sophisticated genetic tools, *Drosophila* is a particularly good model for genome-wide studies because of its well characterized and small genome.

1. The *Drosophila* life cycle

The life cycle of *Drosophila* from egg fertilization to the adult, takes about 10 days at 25°C and is comprised of 4 developmental stages. Embryogenesis: at 25°C is completed in 24h after fertilization. This stage is subdivided into 17 embryonic stages, each stage corresponds to a principle developmental event. Larval stage: There are three larval stages (3 instars) which take altogether about 4 days. Pupal stage: During this stage the *Drosophila* undergoes metamorphosis, it takes 4 days to be completed. Adult life: Adult flies emerge after pupation is completed and they have a lifespan of around 30 days (**Figure 7**).

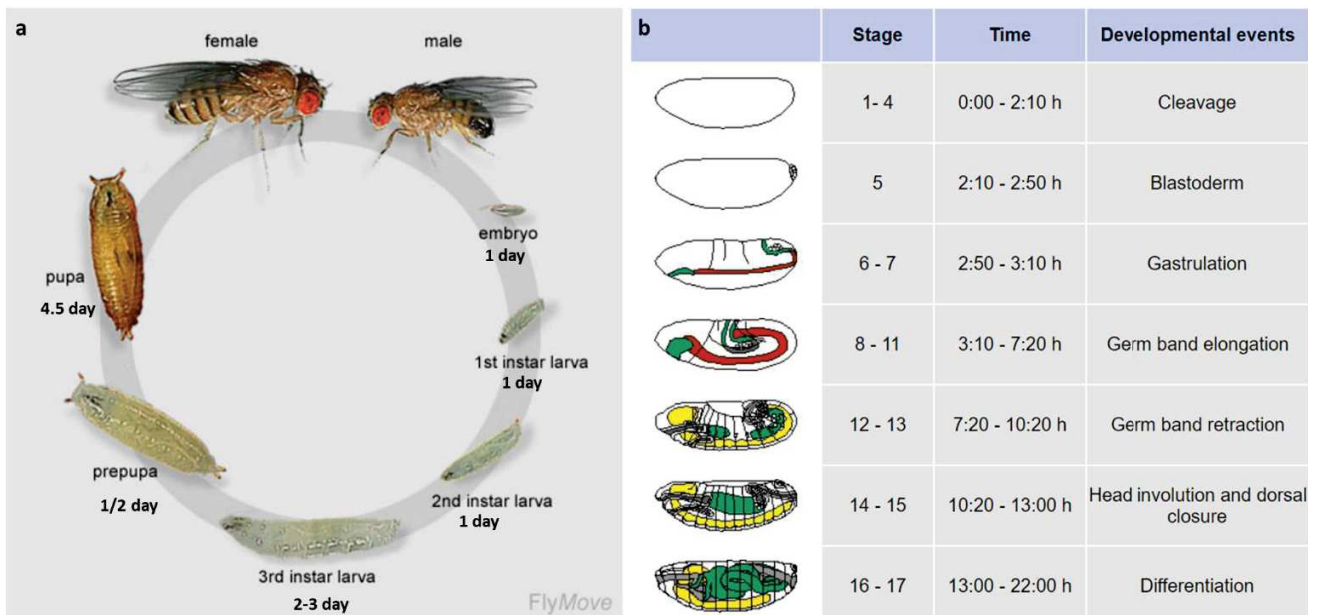


Figure 7: The life cycle of the *Drosophila melanogaster*.

(a) After fertilization females lay eggs that hatch in 22–24 hours at 25°C and the first instar larva emerges. After 25 hours, it molts into the second instar larva that molts into the third instar larva after about 24 hours. This is the longest of the larval forms and after 2 1/2 to 3 days the larva turns into a pupa. The pupal stage lasts 4 days during which metamorphosis allows the formation of the adult tissues.

(b) Embryonic development was subdivided into 17 stages by Volker Hartenstein and Jose Campos-Ortega based on main developmental event.

2. The Gal4-UAS system

The Gal4 transcriptional activator was borrowed from yeast to be used to regulate gene expression in *Drosophila* by inserting the upstream activating sequence (UAS) to which it binds next to a gene of interest. The Gal4 activator can be inserted next to a cell type or tissue specific gene to allow of expression of target in a conditional manner. This system can also be used to knock-down the expression of genes by inserting the UAS next to a sequencing coding the RNAi of the target gene (**Figure 8**) (Brand and Perrimon, 1993; Fischer et al., 1988).

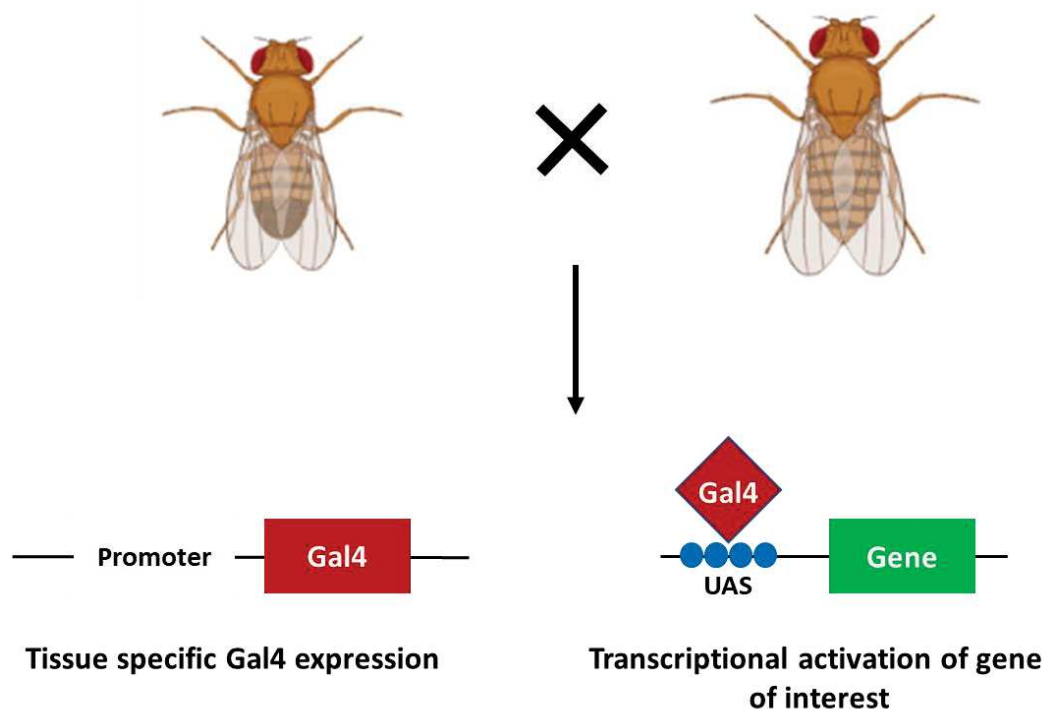


Figure 8: The Gal4-UAS system.

Crossing a tissue specific Gal4 driver with a UAS reporter line allows the expression of the gene of interest in the specific tissue only.

3. The *Drosophila* nervous system

3.1. The development of the *Drosophila* central nervous system

3.1.1. Key steps in the embryonic development of the central nervous system

The neural stem cell (NSC) the precursors of the central nervous system (CNS), arise from the neurogenic region of the ectoderm called neuroectoderm. The ventral neurogenic region gives rise to the neuroblasts of the ventral nerve cord (VNC) (the equivalent of the mammals' spinal cord) (Hartenstein and Campos-Ortega, 1984) and the pro-cephalic neurogenic region gives rise to the brain. The ventral neurogenic region is separated from the mesoderm by the mesoectoderm that generates neurons and glia of the midline. After gastrulation, the cells of the ventral neurogenic region swell and by embryonic stage 9 they start delaminating from the ectoderm. The first wave of delamination gives two rows of NSCs (SI) and the second wave (SII) fills the gap between the two previous rows. Between stage 9 and stage 13, NSCs undergo eight waves of mitosis to produce the intermediate precursor, the ganglion mother cells (GMCs) (Hartenstein et al., 1987). At stage 13 neuronal differentiation begins and a population of neurons lays down a scaffold of fibers on the dorsal surface of the CNS for axons to fasciculate along. Longitudinal fibers form the connectives and transversal fibers form two commissures in each segment that cross the midline while in contact with the midline glial cells. Axons that leave the CNS form an anterior fascicle and a posterior fascicle and sensory axons fasciculate with both tracts. At stage 14, the VNC starts to condense and by stage 17 it is fully condensed and the CNS is complete and isolated by the Blood-Brain Barrier (BBB) (**Figure 9**) (Hartenstein, 1993).

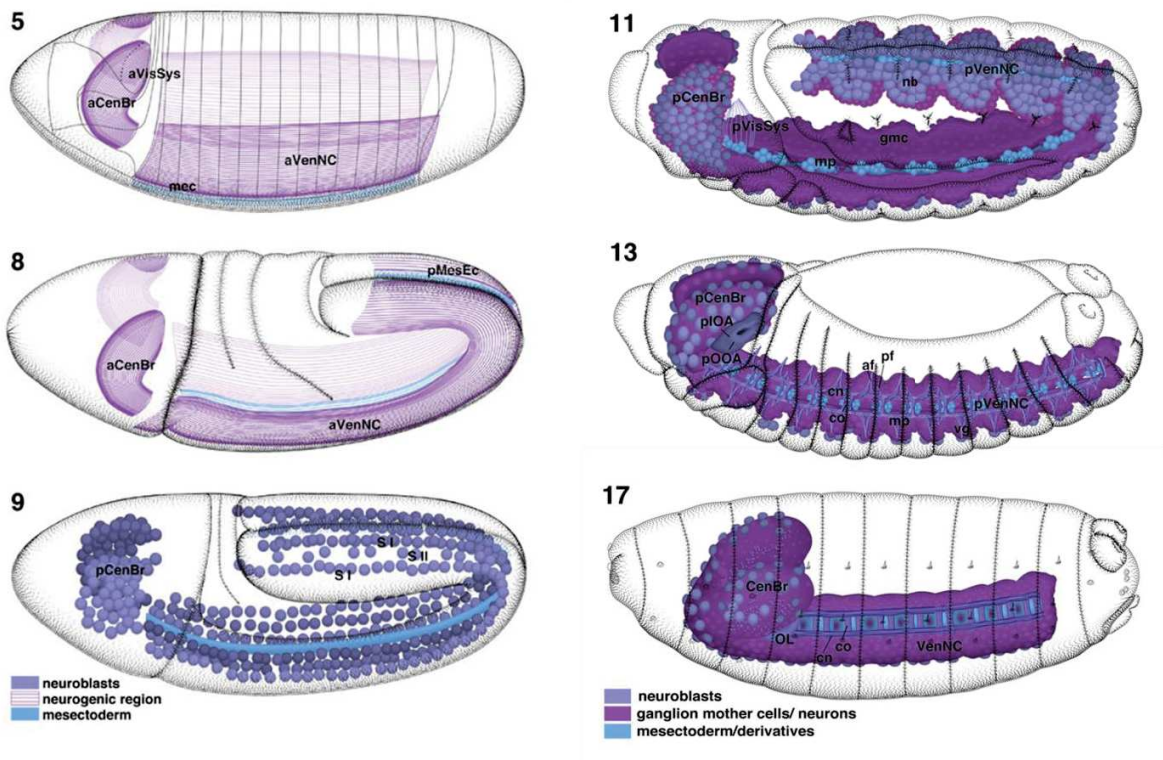


Figure 9: *Drosophila* embryonic central nervous system development

Lateral views of embryos from stage 5, stage 8, stage 9, stage 11, stage 13 and stage 17. CenBr: central brain, VisSyst: visual system. (Modified from the Atlas of *Drosophila* Development by Volker Hartenstein, 1993).

3.1.2. The development of NSCs:

Neurogenesis begins with the delamination of NSCs that start dividing to generate neurons and glia. The development of NSCs involves a defined sequence of expression of different genes. The proneural genes are essential to acquire the potential of becoming NSCs (Ghysen and Dambly-Chaudiere, 1989). Groups of cells in the neuroectoderm called proneural clusters express these proneural genes and the neurogenic genes prevent more than one cell in the cluster from forming a NSC in a process is called lateral inhibition (Simpson, 1997). The proneural genes code for basic Helix Loop Helix (bHLH) domain containing TFs (Campuzano et al., 1985; Villares and Cabrera, 1987) including the members of Achaete-Scute Complex: Achaete, Scute and Lethal of Scute (Martín-Bermudo et al., 1991). The neurogenic genes include the receptor Notch and its ligand Delta. Reinforcement processes that rely on feedback loops and patterning gene expression lead to increased expression of Delta in a cell of the proneural cluster, which triggers the activation of Notch in its neighboring cells. Notch represses the expression of the proneural genes in those cells. As a result, the cell within the proneural cluster that expresses the proneural genes to the highest level adopts the neuroblast fate, while the other cells adopt an epidermal fate (Cubas et al., 1991; Egger et al., 2008; Skeath and Carroll, 1992, 1994). NSCs undergo repeated asymmetric cell divisions leading to the generation of neurons and glia.

3.1.3. The different types of NSCs:

The NSCs of the *Drosophila* NS are divided into three different categories based on their lineage: neuroblasts (NBs) that are only able to generate neurons, glioblasts (GBs) that produce only glia and neuroglioblasts (NGBs) that give rise to both neurons and glia.

NBs go through a series of asymmetric divisions after which one NB and one GMC are generated. The GMC will divide one time and give rise to two neurons, while the NB continues to produce GMC and self-renew (Van De Bor and Giangrande, 2002; Bossing et al., 1996; Doe, 1992; Schmid and Tautz, 1997).

GBs divide symmetrically to generate glia that have proliferative potential.

The NGBs are further subdivided into two types: Type 1 NGBs that divide asymmetrically to produce one NB and one GB, which will then give neurons and glia. Type 2 NGBs also undergo asymmetric divisions, generating two daughter cells of distinct size and fate. The smaller being the GMC that is committed to the neuronal differentiation pathway and will divide one time to give rise to two neurons. The larger daughter cell retains NGB identity and continues to divide asymmetrically to generate one NGB and one GMC. The resulting NGB will generate another GMC that will eventually produce one neuron and one glia (**Figure 10**) (Altenhein et al., 2016).

The asymmetric division of NSCs depends on the asymmetrical localization of cell fate determinants as a cortical crescent during mitosis. In addition to the orientation, the asymmetry of the mitotic spindle itself ensures the exclusive segregation of the determinants to the smaller cell, *i.e.* the GMC (Jan and Jan, 1998).

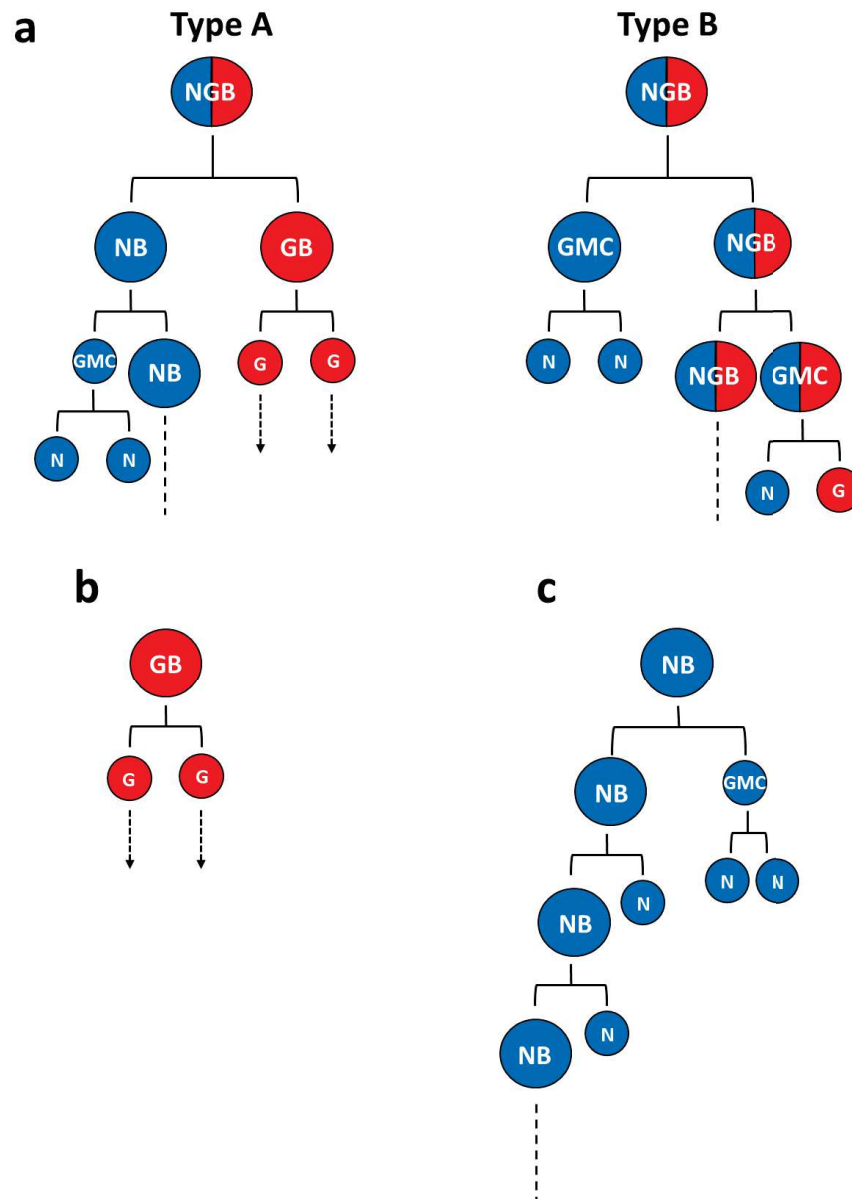


Figure 10: Neural stem cells in *Drosophila*.

(a) Neuroglioblast (NGB) Type 1 generate one glioblast (GB), which in turn gives rise to two glia (G), and one neuroblast (NB) which gives rise to another NB and one ganglion mother cell (GMC) that gives rise to two neurons (N). NGB Type 2 gives rise to one NGB and one GMC. The NGB generates one NGB and one GMC that will give rise to one neuron and one glia. The GMC gives rise to two neurons. (b) Pure GB that gives rise to two glia that can proliferate (c) NB gives rise to one NB and one GMC which will give rise to two neurons.

Modified from (Van De Bor and Giangrande, 2002).

3.1.4. Post-embryonic development of the CNS

Towards the end of embryogenesis, a subset of NSCs stops dividing and enters quiescence, whereas some others undergo apoptosis (Prokop and Technau, 1991; Truman, 1990; Truman and Bate, 1988).

The larval nervous system is generated during embryonic neurogenesis. From the first larval stage until pupal stage, a postembryonic period of neurogenesis is required to produce the majority of the adult CNS. 95% of the neurons in the adult brain are neurons generated during postembryonic neurogenesis (Dumstrei et al., 2003). This is accompanied by a new wave of neural proliferation that starts at the late first instar (Ito and Hotta, 1992). These larval NSCs correspond to the reactivated embryonic ones (Speder and Brand, 2018; Prokop and Technau, 1991; Sousa-Nunes et al., 2011; Truman, 1990). NSCs in the larval CNS divide asymmetrically to produce a pocket of progeny that remain immature until metamorphosis. The basic machinery involved in the asymmetric division of these larval neuroblasts appears to be conserved with embryonic neuroblasts (Chia et al., 2008; Egger et al., 2008).

3.2. **The cells of the *Drosophila* CNS:**

The nervous system of the *Drosophila* larva contains around 15,000 cells. This number increases massively to reach around 150 000 in the adult CNS (Ito et al., 1995; Jenett et al., 2012; Kremer et al., 2017). Besides the NSCs that were discussed in the previous section, the CNS of *Drosophila* contains two main cell types, neurons and glia. They both differentiate from the same precursor then take on very different morphologies and functions.

3.2.1. Neurons:

There are three major types of neurons in the nervous system of *Drosophila*: motoneurons, interneurons, neurosecretory neurons.

a) Motoneurons:

In the *Drosophila* VNC motoneurons are generated during embryogenesis by different NSCs that are not restricted to a particular sector, nor do they show a particular gene expression pattern. Each motoneuron extends axonal projection and innervate a particular target muscle. Motoneurons are further subdivided based on their choice of nerve root. In fact, the axons of motoneurons project through two main nerves: the intersegmental (ISN) and the segmental nerve (SN). Thus, intersegmental motoneurons innervate the internal muscles, while segmental motoneurons innervate the external set of muscles (Garces and Thor, 2006; Landgraf and Thor, 2006; Schmid et al., 1999).

b) Interneurons:

The interneurons extend axons within the CNS to innervate other neurons. For instance, they are the presynaptic element of motoneurons. There are 300 interneurons divided into two subclasses: intersegmental interneurons, that project axons that span more than one segment within the CNS and local interneurons whose axon projections terminate within their segment of origin in the CNS. Even though interneurons in the CNS outnumber motoneurons by 10-fold, less is known about the factors involved in their specification (Schmid et al., 1999).

c) Neurosensory neurons:

Neurosensory neurons extend axons either out into the periphery without extending to the body wall muscles or into the CNS to secrete neuropeptides and hormones (Schmid et al., 1999).

3.2.2. Glia:

The idea that glia are nothing but the glue that holds the brain together is long gone now that we have realized the immense diversity of these cells and how crucial their many functions are for the development and homeostasis of the NS.

The glial cells of the *Drosophila* CNS can be divided into two big groups based on their developmental origins and fate determinants. The midline glia that originate from the mesectoderm reside in the midline and ensheath the commissural fiber tracts (Bossing and Technau, 1994; Menne and Klämbt, 1994) and the lateral glia, that emerge from the neurectoderm and differentiate from the NSCs, occupy different positions around neurons and axons. The differentiation of midline glia requires the TF Single-minded (Nambu et al., 1991). The determination of lateral glia depends on the transient expression of a single TF, Glide/Gcm (Gcm from now on). In the embryo, *gcm* is expressed in NGBs and GBs to trigger the glial fate by activating the pan glial expression of TF Reverse polarity (Repo) and Ttk69 (Hosoya et al., 1995; Popkova et al., 2012; Vincent et al., 1996). In fact, *gcm* LOF causes the loss of glia differentiation since all precursors differentiate into neurons and, accordingly, ectopic expression of *gcm* is enough to induce the glial fate at the expense of the neuronal fate (Bernardoni et al., 1998; Hosoya et al., 1995; Miller et al., 1998).

The lateral glia fall into 4 groups based on their association with the CNS compartment:

a) Surface glia:

The CNS of the *Drosophila*, like that of mammals, is separated from the rest of the organism by the Blood-Brain Barrier (BBB) formed by the surface glia. The surface glia include two distinct subtypes: the perineural glia and the subperineural glia.

The **Perineural glia (PNG)** are thought to originate in the embryo however they have only been described starting from the larval life. PNG are found in the CNS and the peripheral nervous system (PNS). The ones of the CNS form the outermost layer on the surface of the CNS and are in contact with the hemolymph. They express transporter proteins required to shuttle nutrients and ions into the CNS. PNG are also involved in the deposition of the extracellular matrix (ECM). They are small cells that together with the extracellular matrix (ECM) make a mesh-like sheath surrounding the CNS and serving as a first barrier for large molecules. The PNG remain mitotically active and proliferate throughout the larval life (Awasaki et al., 2008; Altenhein et al., 2016; Brankatschk et al., 2014; Colonques et al., 2007; Limmer et al., 2014; Omoto et al., 2016; Yildirim et al., 2019).

The subperineural glia (SPG) are large flat cells formed during embryogenesis and remain until the end of larval stages. They connect to each other via septate junctions (SJ, the equivalent of the mammals' tight junctions). Together with the perineural glia they form the BBB that will seal the NS and control the diffusion of molecules. Moreover, SPG are also involved in the reactivation of NSCs in the larva in response to nutritional signals (Banerjee et al., 2008; von Hilchen et al., 2013; Omoto et al., 2016; Stork et al., 2008; Yildirim et al., 2019).

b) Cortex-associated glia:

Only one type of glia is cortex-associated, they are called Cell Body Glia (CG). They form a small population of cells that continue to proliferate post-embryonically and remain until adulthood. CG infiltrate the neuronal cell cortex and encapsulate neuronal cell bodies. CG associate closely with the SPG and are likely involved in the efficient transfer of nutrients from the hemolymph to neuronal cell bodies. They express the scavenger receptors *draper* (*drpr*) and *Nimrod C4* (*NimC4*) and have been reported to be the main phagocytic population in the embryo that is involved in the clearance of debris and apoptotic bodies (Freeman et al., 2003; Ito et al., 1995; Kurant et al., 2008; Peraanu et al., 2005; Younossi-Hartenstein et al., 2006).

c) Neuropile-associated glia:

The neuropile of the VNC is formed by connectives and segmentally repeated commissures. At the commissures, nerve roots carrying axons to and from the periphery enter the connectives. Glia associated to this structure are very heterogeneous, however they include two well defined subtypes. The **Ensheathing Glia (EG)** have flat cells bodies found on the surface of the neuropile enwrapping and separating it from the cortex without infiltrating within the neuropile. These cells do not divide during the larval life and they undergo apoptosis during metamorphosis to give way to adult ensheathing glia. The adult EG are thought to be the phagocytic cells of the adult CNS. The second well characterized subtype is the **Astrocyte-like Glia (AG)**. As their name suggests, they resemble the vertebrate astrocytes morphologically, molecularly and in some functional aspects. They are found between the EG but unlike the EG that remain on the surface of the neuropile, AG extend branched processes into the neuropil. They also do not proliferate during the larval stage but their size increases until they undergo apoptosis

during metamorphosis (Doherty et al., 2009; MacDonald et al., 2006; Peco et al., 2016; Watts et al., 2004; Ziegenfuss et al., 2008).

d) Glia of the peripheral nervous system:

Drosophila lacks myelin and the wrapping of axons is done by the glial cells themselves. In fact, the PNS of *Drosophila* contains a specific subtype of glia called **Wrapping Glia (WG)** that are associated to axons and wrap all sensory and motor axons. These cells do not divide during the larval life however they grow enormously in size. In fact, growth in *Drosophila* normally occurs during the larval stages, where the animal undergoes a 200-fold increase in body mass. This is accompanied by massive axon growth and since a few WG need to cover the entire nerve, they must grow as well (**Figure 11**) (Church and Robertson, 1966; von Hilchen et al., 2013; Schirmeier et al., 2016).

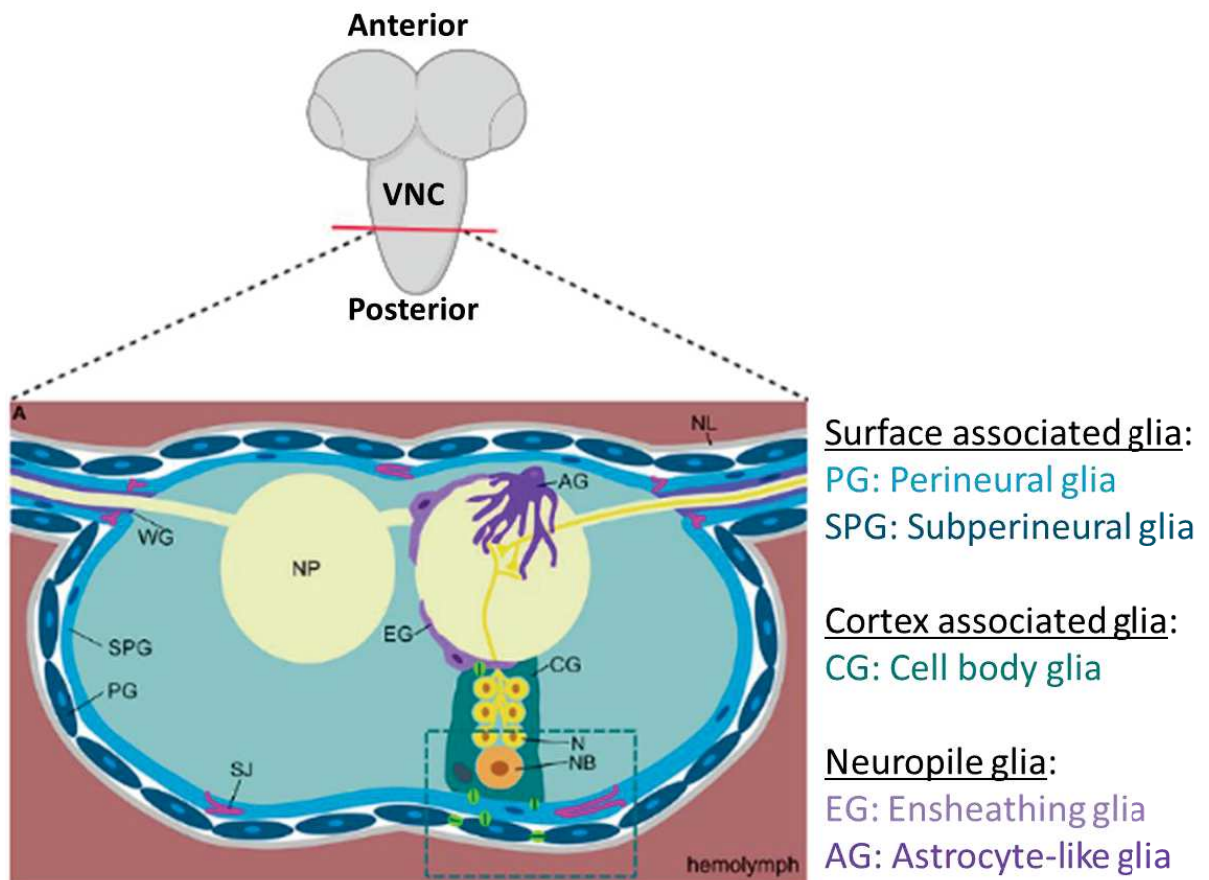


Figure 11: Schematic representation of the *Drosophila* glia subtypes.

Cross-section view of a *Drosophila* ventral nerve cord. NL: neural lamella, the extracellular matrix around the CNS. The outermost glial layer consists of perineural glial cells (PG). The subperineural glia (SPG) form septate junctions (SJ) and control paracellular transport. Neuronal cell bodies (N) and neuroblasts (NB) are surrounded by cortex glia (CG). The neuropil (NP) is covered by ensheathing glia (EG). Astrocytes (AG) invade the neuropil. In the peripheral nerves, wrapping glia (WG) ensheath axons.

(Modified from Limmer et al., 2014)

4. The immune system of *Drosophila melanogaster*

Drosophila has been considered a powerful model to study innate immunity because of the absence of adaptive immune response that could otherwise mask the innate immune system. The fly encounters many pathogens throughout its life starting from when it first starts interacting with its environment at early larval stages. The main immune response of *Drosophila* comprises humoral reactions of antimicrobial peptides and cellular reactions dependent on the hemocytes (the blood-cells) such as phagocytosis, melanization and encapsulation (Brennan and Anderson, 2004; Hoffmann, 2003; Hultmark, 2003; Imler and Bulet, 2005).

4.1. The *Drosophila* hemocytes and their functions:

The hemocytes of *Drosophila* can be divided into 3 distinct cell types that have specific morphologies and functions.

4.1.1. Plasmatocytes:

Plasmatocytes make up around 95% of hemocytes. They are professional phagocytes and are comparable to the mammalian macrophages. Plasmatocytes express scavenger receptors to recognize, engulf and destroy pathogens and virus infected cells (Lebestky et al., 2000; Lemaitre et al., 1996; Tepass et al., 1994). In the embryo, protected from environmental challenge by the eggshell and thus not prone to encounter pathogens, plasmatocytes are involved in the clearance of apoptotic bodies. This is mediated by the expression of the scavenger receptors including Croquemort (Crq), Draper (Drpr) and Nimrod C4 (NimC4) (Franc et al., 1996; Kurant et al., 2008; Manaka et al., 2004). Once the fly reaches the larval stage, it thrives in a micro-organism rich environment including pathogens. Thus, the plasmatocytes must rapidly recognize and destroy the threats. The phagocytosis of pathogens by plasmatocytes requires the expression of another set of

scavenger receptors, including members of the Nimrod family such as Eater, NimC1 and NimC2, in addition to members of the Scavenger Receptor (SR) family (Kocks et al., 2005; Kurucz et al., 2007; Rämét et al., 2002). Plasmatocytes also secrete small cationic peptides called Anti-Microbial Peptides (AMPs), which contribute to innate immune defense. These peptides disrupt microbial membranes and lead to their destruction (Imler and Bulet, 2005; Kragol et al., 2001; Rahnamaeian et al., 2015). The role of plasmatocytes goes beyond immunity and phagocytosis. They are necessary for the maintenance of the homeostasis of the organisms. The embryonic hemocytes secrete many components of the ECM that is necessary for tissue morphogenesis such as the ECM surrounding the CNS (Brown, 2011; Martinek et al., 2008). Plasmatocytes also play a crucial role in the regulation of stem cell activity as well as the regulation of glucose metabolism (Ayyaz et al., 2015; Van De Bor et al., 2015; Woodcock et al., 2015).

4.1.2. Crystal cells:

Crystal cells make up no more than 5% of the population of the hemocytes. They are larger in size than plasmatocytes and owe their name to the presence of cytoplasmic crystals composed of ProPhenolOxydases (PPO). These crystals are involved in an arthropod-specific defense mechanism called melanization that leads to the production of melanin and Reactive Oxygen Species (ROS). This mechanism plays an important role in the immune response since it prevents loss of hemocytes in wound sites, immobilizes pathogens, and potentially destroys them via ROS (Binggeli et al., 2014; Eleftherianos and Revenis, 2011; Jiravanichpaisal et al., 2006; Rizki et al., 1985).

4.1.3. Lamellocytes:

Lamellocytes are very large, flat and adhesive cells produced massively upon immune challenge. They contribute to melanization with the crystal cells and they mediate the encapsulation of pathogens or bodies that are too large to be phagocytosed by plasmatocytes. (Evans and Banerjee, 2003; Krzemien et al., 2010; Lanot et al., 2001; Sorrentino et al., 2002).

4.2. The development of the *Drosophila* immune system:

4.2.1. The hematopoietic waves:

Hemocytes are generated in two distinct waves of hematopoiesis during development. Each wave produces the three major types of hemocytes.

The first wave takes place during embryogenesis during which a population of hemocytes originates in the procephalic mesoderm then migrate and disperse throughout the embryo. These cells are long lived and they will remain through the larval life all the way to adult life (Tepass et al., 1994).

The second hematopoietic wave is activated at larval stages in a tissue called the lymph gland. The precursors of the lymph gland form in the dorsal thoracic mesoderm in the late embryo independently of the embryonic prohemocytes. The lymph gland grows and its cells proliferate during larval life until this tissue undergoes hystolysis at the onset of metamorphosis and releases a second pool of hemocytes into circulation (**Figure 12**) (Grigorian et al., 2011; Holz et al., 2003; Rizki and Rizki, 1978).

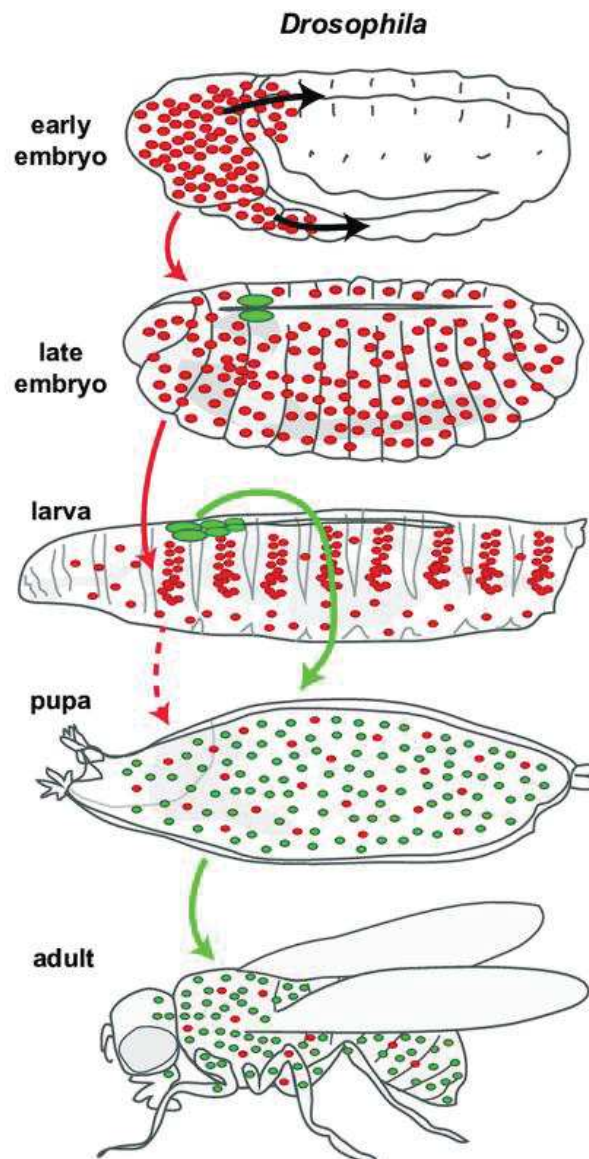


Figure 12: *Drosophila* hematopoiesis.

The first wave of hematopoiesis gives rise to the embryonic hemocytes (red), these cells represent the only circulating immune cell in the larva. The second wave takes place in the lymph gland (green) when hemocytes are released at the onset of metamorphosis. The two pools of hemocytes coexist in the adult.

(Modified from Wood and Martin, 2017).

4.2.2. Fate determination of hemocytes:

The fate of embryonic hemocytes is determined very early on in the blastoderm. This depends on the expression of the GATA factor *Srp* that is needed first within the mesoderm in the embryo, then for the maturation of hemocytes and later on in the lymph gland. Its role in the differentiation of hemocytes involves the activation of the expression of the TF U-Shaped (*Ush*) (Jung et al., 2005; Lebestky et al., 2000). The two factors induce *Gcm* and its homolog *Gcm2* expression, which in turn activate the expression of plasmatocyte-specific genes (Fossett and Schulz, 2001). Crystal cell differentiation requires Notch signaling and the expression of the transcription factor *Lozenge* (*Lz*), a Runt-domain protein that shares homology with mammalian RUNX proteins (de Bruijn and Speck, 2004; Lebestky et al., 2000; Rehorn et al., 1996; Waltzer et al., 2003). As for lamellocytes, they transdifferentiate from plasmatocytes upon immune challenge. The exact mechanism triggering trans-differentiation is still unknown, however, it has been shown that it is induced by the overexpression of *srp* and *charlatan* (*chn*) as well as by the repression of *ush*. In addition, some evidence point to a direct differentiation of lamellocytes from prohemocytes mediated by the overactivation of the Janus kinase (JAK)-signal transducers and activators of transcription (STAT) and Toll pathways (**Figure 13**) (Crozatier et al., 2004; Harrison et al., 1995; Hou et al., 2002; Stofanko et al., 2010).

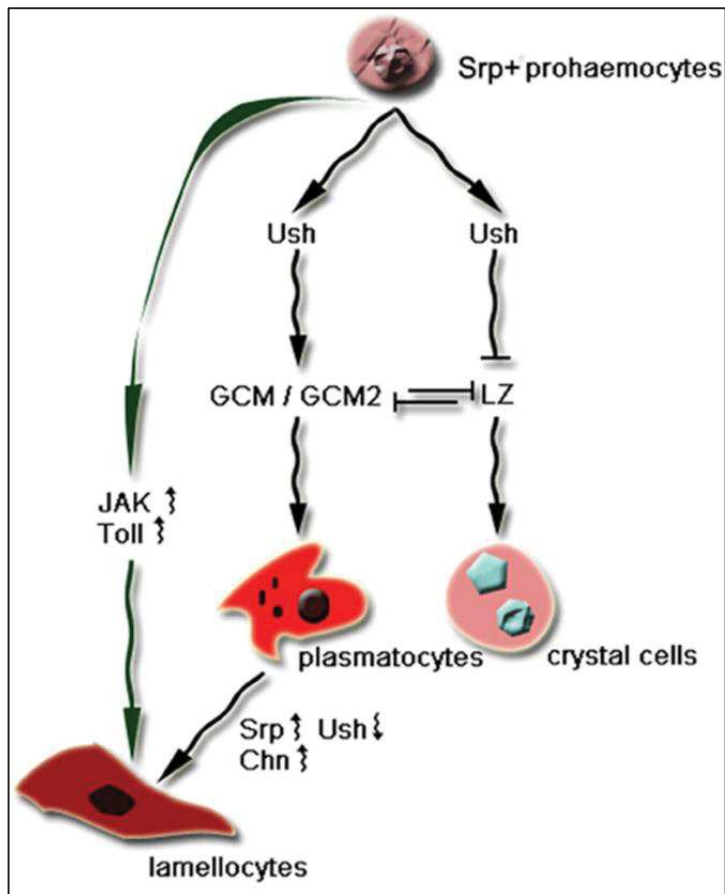


Figure 13: Fate determination of the *Drosophila* hemocytes.

Prohemocytes express the TF Serpent (Srp) which defines the hemocyte fate from embryogenesis. The Srp expressing prohemocytes turn on the transcription of U-shaped (Ush). Ush together with Srp activates the expression of Glial cell missing (Gcm) and its paralog Gcm2. Gcm and Gcm2 in turn activate the expression of plasmatocyte-specific genes and thus commits prohemocytes into plasmatocyte specification. The prohemocytes that express Lozenge (Lz) adopt the crystal cell fate by antagonizing Gcm and Gcm2 expression. Lamellocytes do not appear in normal circumstances but can be induced upon immune challenge. Studies have pointed out a direct differentiation of lamellocytes from plasmatocytes by the upregulation of Srp and Charlatan (Chn) and downregulation of Ush. However direct differentiation from the prohemocytes may also contribute to the total population of lamellocytes through the over-activation of JAK/STAT and Toll signaling.

(Modified from Wang et al., 2014)

AIMS

When a cell acquires its identity, it becomes a functional unit in the organism. Failure to accomplish this leads to the disruption of the homeostasis that manifests in the onset on disease, which makes understanding cell identity of uppermost importance. Therefore, I decided to explore cell identity from a molecular point of view starting from gene expression. The starting hypothesis was that a cell possesses a specific gene expression profile containing the information needed to unveil its characteristics and recapitulate its identity. However, the question remains: how does a cell express the defined set of genes needed at the right time? Gene expression is a tightly regulated mechanism and this regulation requires the combined actions of a set of transcription factors and the appropriate epigenetic profile.

Therefore, I explored the possibility of a cell-specific epigenetic signature that accompanies and partially explains the cell-specific gene expression profile. Finally, while the cell classification is needed to study cell functions, I wondered whether this classification leads to an underestimation of cell identity. In other words, do individual cells have their own identity beyond the one shared by the population of cells within the same type?

To detangle the many aspects of cell identity I employed the *Drosophila melanogaster* as a model organism. In particular, I investigated three cell types: neurons, glia and hemocytes. Neurons and glia differentiate from the same precursor yet take on very different morphologies and functions once differentiated. Therefore, these two cell types are the ideal model to study the impact of origin on cellular identity. In addition, while glia and hemocytes originate from different tissues they share the function of phagocytosis. Glia are the only phagocytic cells within the

isolated CNS and hemocytes are the phagocytic cells in the rest of the organism. Thus, glia and hemocytes represent a good example to investigate the impact of function on cell identity. Using these three cell types during my PhD I aimed to explore cell identity by addressing the following questions:

1. Is a cell's identity reflected in its expression profile?

The first aspect of cell identity I explored is the gene expression. Therefore, the transcriptomes of three different cell types were analyzed and I identified the expression profile characterizing each cell type. I also investigated the relation between their expression profile and their origin and function and studied the evolution of this profile through development.

2. Is a cell's identity reflected in its epigenetic profile?

The molecular identity of a cell is identified not only by its transcriptome but also but its epigenome. Therefore, I identified the correlation between the expression levels and the distribution of histone marks. I assessed whether different cells have specific epigenetic signatures regulating their expression profile and their identity. I also investigated whether this signature remains stable or changes through time. In addition, I studied the impact of histone marks on cell function.

3. Do all cells within a cell type share a common identity?

Dividing cells into different types is important to be able to study them. However, the current technology allows us to explore cell identity on a deeper level through assessing single cells. Therefore, I identified the expression profile of single cells to investigate their heterogeneity.

I examined whether cells belonging to the same type present different expression profiles and if so, whether this profile reflects distinct functions and identities. Finally, I assessed the degree to which the expression profile of a cell is impacted upon challenge.

CHAPTER I

Characterization of cell-specific gene expression profiles in the *Drosophila* neurons, glia and hemocytes

I- Introduction:

In multicellular organisms, different cell types emerge from the zygote to take on different characteristics and fulfill specific functions upon the acquisition of distinct gene expression profiles. The nature of cell identity represents a central question in biology, as the accurate definition of cell types has an impact on many areas of research and clinical applications, including regenerative medicine. Traditionally, cell identities have been defined by morphology, location, origin and function, relying on the expression levels of few proteins. While the combination of these parameters has produced valuable information, each of them on their own can induce to misleading conclusions. Cells in different organs can display similar functions and cells within the same organ can have different properties. Moreover, the *a priori* selection of proteins can introduce bias and neglect the wealth of molecular information intrinsic to each cell population. To ensure a more unbiased and quantitative analysis, I used a high throughput approach and a data-driven computational method. To analyze related and unrelated cell types, I chose *Drosophila* differentiated neurons, macrophages and glia, which display specific molecular signatures and cellular features according to their morphology, origin, location and function. Moreover, because functional plasticity and dynamic states may lead to a reductionist classification, I performed my analysis at two stages in the life cycle, the embryo and the larva.

Drosophila neurons and glia share the same origin as they differentiate from common precursors that derive from the ectoderm. Neurons represent a very specialized cell type: they sense, process and transmit information or secrete neuropeptides (Garces and Thor, 2006; Landgraf and Thor, 2006; Schmid et al., 1999). Glia are an extremely heterogeneous cell population that plays critical roles in the development, function, and maintenance of the nervous system. They control cell proliferation (Ebens et al., 1993), axonal and synapse development (Brink et al., 2012; Ou et al., 2014), neuronal insulation and survival (Shepherd, 2000; Stork et al., 2008; Volkenhoff et al., 2015). They also establish the blood-brain barrier (Abbott, 2005) and provide an immune function within the nervous system (Awasaki and Ito, 2004; Sonnenfeld and Jacobs, 1995; Watts et al., 2004). Macrophages, also called plasmatocytes, represent the vast majority of the *Drosophila* immune cells (or hemocytes) present outside the nervous system. They derive from the mesoderm and patrol the organism to ensure cellular immunity (Lebestky et al., 2000; Lemaitre et al., 1996; Tepass et al., 1994), hence sharing the function but not the origin with glial cells, while sharing neither the origin nor the function with neurons.

The analysis of the transcriptional landscape of the three cell types highlights numerous features. First, in the embryo, the expression profile of a cell is strongly affected by the origin of that cell, later on, cells become more affected by the function and acquire additional levels of specialization to accommodate to the changing needs of the organism. Typically, glia are more similar to neurons than to the hemocytes in the embryo. At larval stage glia diverge from neurons and start showing similarities with hemocytes. Moreover, fewer genes are expressed at larval stages compared to those observed in the embryo suggesting a higher specialization in the larval cells. In addition to cell-specific signatures, cells share significant stage-specific signatures. For example, embryonic hemocytes are closer to embryonic glia and neurons than to larval hemocytes.

Furthermore, cell identity involves stable as well as plastic features, the relative impact of whose depends on the type of the cell. The transcriptome of neurons remains more stable over time, that of hemocytes changes significantly in the larva, an open system that has to react to external stimuli such as pathogens. Finally, while larval glia and hemocytes share a common immune scavenger function, different scavenger receptors are upregulated in either cell type, highlighting specific properties and potential that likely reflects their different environment.

Altogether, this analysis makes it possible to disentangle the relative importance of origin, location, function and development in the definition of cell identity. The unbiased measurements of the whole transcriptomes pave the way for systematic identification of gene networks and expression programs, which will ultimately shed light on the biological mechanisms characterizing cell diversity in multicellular organisms.

II- Materials and Methods:

1. Fly strains and genetics

All flies were raised on standard media at 25°C. For the transcriptomes, hemocytes were collected from stage 16 (E16) *srp(hemo)Gal4/+; UAS-RFP/+* embryos obtained upon crossing *srp(hemo)Gal4* (gift from K. Brückner) (Brückner et al., 2004) and *UAS-RFP* flies (RRID:BDSC_8547). The wandering L3 (WL) hemocytes were collected from staged *HmlΔRFP/+* animals upon crossing *HmlΔRFP* (Makhijani et al., 2011) with *Oregon-R* flies (108-117 h After Egg Laying, h AEL). The transgenic line *repo-nRFP 43.1* on the III chromosome was used for the isolation of glia. The *repo-nRFP* line recapitulates the full *repo* expression pattern (Laneve et al., 2013). The *elav-nRFP 28.2* insertion on the III chromosome was used to purify embryonic neurons and *nSyb-Gal4/+; UAS-RFP/+* for the isolation of L3 neurons. The *Oregon-R* strain was used as the control for all experiments.

2. FACS sorting of hemocytes, glia and neurons

Embryonic cells were isolated as follows. Staged egg laying was carried out to produce E16 embryos. *srp(hemo)Gal4/+; UAS-RFP/+*, *repo-nRFP*, *elav-nRFP* and *Oregon-R* strains were amplified, several hundreds of flies of the different genotypes were then transferred into cages and raised on a yeast apple juice agar at 25°C. After a pre-lay period of 30 minutes, the agar plates and yeast were replaced with fresh plates and flies were left to lay for 3 hours at 25°C. Agar plates were then removed and kept at 25°C, embryos were collected 11 hours and 40 minutes AEL, when they reached stage 16. Embryos were then isolated from the medium and washed on a 100 μm mesh. The collected embryos were transferred into a cold solution of phosphate-buffered saline

(PBS) in a Dounce homogenizer on ice. The embryos were dissociated using the large clearance pestle then the small clearance pestle and then filtered through a 70 μm filter.

Third instar larval hemocytes were isolated as mentioned in (Cattenoz et al., 2020). Third instar neurons, glia and hemocytes were isolated from *elav-nRFP*, *repo-nRFP* and *Hml Δ RFP/+* larvae respectively. Staged lays of 3 h were carried out at 25°C and wandering larvae were collected 108–117 h AEL. For the isolation of hemocytes, larvae were bled in cold PBS containing PTU (Sigma-Aldrich P7629) to prevent hemocyte melanization (Lerner and Fitzpatrick, 1950), and filtered through a 70 μm filter. For the isolation of neurons and glia, larval brains were dissected in cold PBS on ice and transferred to tubes containing 0.5 μg of collagenase IV (Gibco, Invitrogen) in 220 μL of PBS. Brains were incubated at 37°C on a Thermomixer with shaking at 500 rpm for 20 minutes. Then the brains were dissociated by pipetting up and down with 10-gauge needles and syringes and filtered through a 70 μm filter.

The cells were sorted using FACS Aria II (BD Biosciences) at 4°C in three independent biological replicates for each genotype. Live cells were first selected based on the forward scatter and side scatter and only single cells were sorted according to the RFP signal. *Oregon-R* cells were used as a negative control to set the gate and sort the RFP positive cells (**Sup. Figure 1**), which were collected directly in TRI reagent (MRC) for RNA extraction. Around 100 000 cells were sorted for each replicate. The purity of the sorted populations was assessed by carrying out a post-sort step. The FACS sorter was set up to produce cell pools displaying at least 80% of purity on the post-sort analysis.

3. RNA extraction and sequencing

The sorted cells were homogenized in TRI reagent (MRC) for 5 minutes at room temperature (RT) to ensure complete dissociation of nucleoprotein complexes. 0.2 mL of chloroform was added to each sample followed by centrifugation at 12,000 g for 15 minutes at 4°C. The upper aqueous phase containing the RNA was collected and transferred to a new tube. 0.5 mL of 2-propanol were added, and the samples were incubated for 10 minutes at RT. The samples were then centrifuged at 12,000 g for 10 minutes to precipitate the RNA. The RNA pellet was then washed with 1 mL of 75% ethanol then precipitated again at 7,500 g for 5 minutes and air dried. 20 µL of RNase-free water was added to each sample before incubation at 55°C for 15 minutes. Single-end polyA⁺ RNA-Seq (mRNA-seq) libraries were prepared using the SMARTer (Takara) Low input RNA kit for Illumina sequencing. All samples were sequenced in 50-length Single-Read.

4. Data analysis

The data analysis was done using the GalaxEast platform, the Galaxy instance of east of France (<http://www.galaxeast.fr/>). First, FastQC (Babraham Bioinformatics) was used on the raw FastQ files to generate summary statistics and assess the quality of the data. The raw files were then converted to FastQ Sanger using FastQ Groomer for downstream analysis. Data were then mapped to the *Drosophila melanogaster* August 2014 Dm6 (BDGF release 6 + ISO1 MT/Dm6) reference genome using TopHat (Trapnell et al., 2009). The number of reads per annotated gene were counted using Htseq-Count (Anders et al., 2015) and the comparison and normalization of the data between the different cell types was done using Deseq2 (Anders and Huber, 2010). To compare E16 stage to L3 using Deseq2, I considered all cell types deriving from E16 as replicates

and compared to the cell deriving from L3 that were also considered replicates. Gene ontology studies were done using the Panther classification system version 16 (<http://pantherdb.org/>) (Mi et al., 2010; Thomas et al., 2003) for the identification of biological processes, Bonferroni correction was used for all analyses. Genes with a number of reads > 15 are considered significantly expressed. For the study of upregulated genes, only genes with adjusted p-value <0.05, fold change >1.5, expression level >15 were taken into account while genes with number of reads > 15 in all cell types and fold change < 1.5 are considered commonly expressed. The Venn diagram were generated using Venny 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/>) (Oliveros, J.C. (2007-2015)) and data representation was done using Microsoft Excel.

5. Quantitative PCR

For the comparison between the expression levels of scavenger receptors in hemocytes and glia, cells were isolated from *srp(hemo)3xmCherry* and *repo-nRFP* wandering third instar larvae, respectively. FACS sorting and RNA extraction were as describe previously. The extracted RNA was then treated with DNase I recombinant RNase free (Roche) and the reverse transcription was done using the Super-Script IV (Invitrogen) with random primers. The cycle program used for the reverse transcription is 65°C for 10 min, 55°C for 20 min, 80°C for 10 min. The qPCR was done using SYBR Green I Master (Roche). Actin5C (Act5C) and Ribosomal protein 49 (RP49) were used to normalize the data. The primers are listed in **supplementary table 1 (Sup. Table 1)**. The P-values and statistical test used are indicated in the figure legends.

For the quantitative PCR done on larval and embryonic tissues, whole larvae and whole embryos were crushed in cold PBS using a Dounce homogenizer, extracts were filtered first

through a cell strainer of 100 μm and RNA extraction, reverse transcription and qPCR were done as described.

6. Histone extraction and western blot

Histones from E16 embryos and wandering L3 *Oregon-R* were extracted as described in Abcam histone extraction protocol (www.abcam.com/protocols/histone-extraction-protocol-for-western-blot). Histone extracts were separated by 15% SDS-PAGE, transferred onto nitrocellulose membrane and probed with the primary antibodies mouse anti-UNC93-5.2.1 (1:5000) recognizing the *Drosophila* Histone 2A gamma variant, phosphorylated γH2AV , (Developmental Studies Hybridoma Bank (DSHB), Lake et al., 2013) and rabbit anti-H3 (1:5000) (Abcam # 1791) for normalization. Signal was detected with Pierce ECL western blotting substrate (Thermo Fisher Scientific, Waltham, MA) using appropriate HRP-conjugated secondary antibodies (1:10 000, Jackson).

III- Results:

1. Impact of development and function on transcriptional landscapes

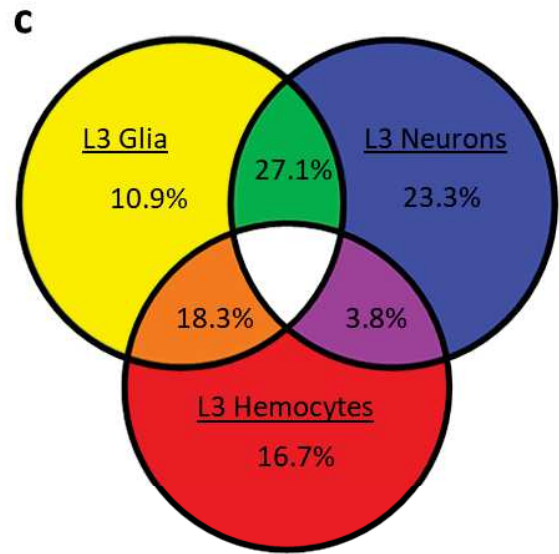
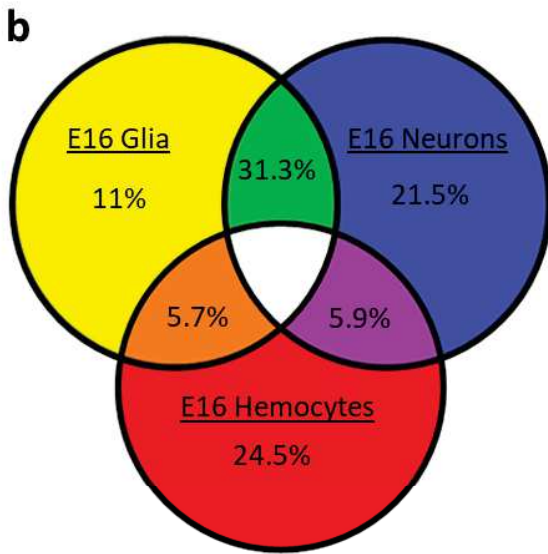
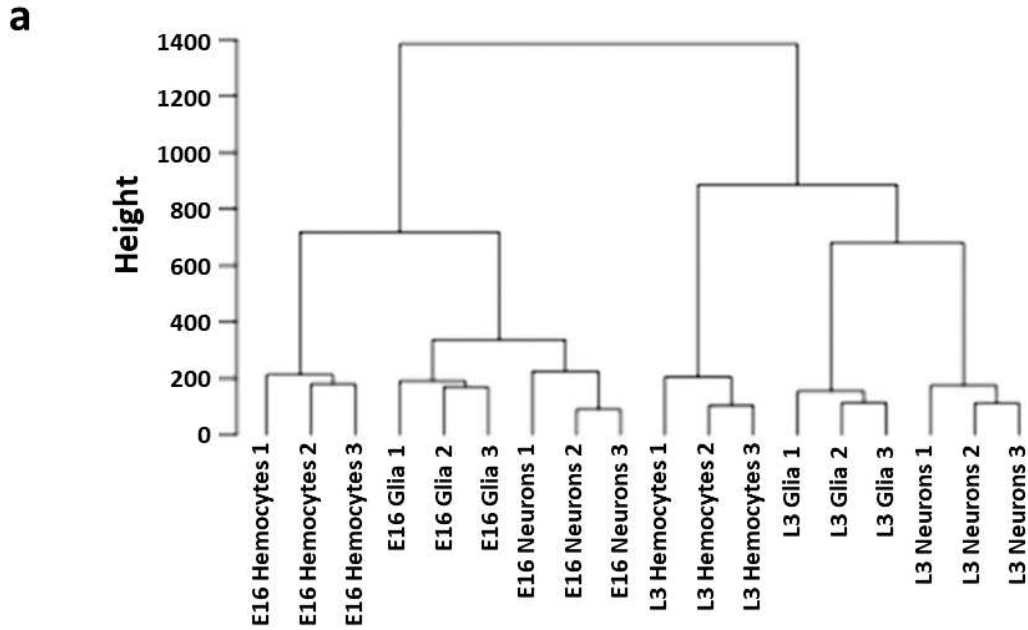
Our body is populated by a variety of cells that express specific genes according to their function, location and developmental stage. To assess the impact of these parameters on these molecular signatures, I compared the transcriptomes of embryonic neurons and glia, which share the same precursor in the neuroectoderm; those of embryonic hemocytes and glia, which share a common scavenging activity; and those neurons and hemocytes, which share neither origin nor function. I compared these three cell types in mature embryos (stage 16 or E16) and at the wandering third instar larval stage (L3). This allowed me to define how the expression profiles change due to the transition from a small, closed and immobile system to a bigger, open, system that is able to respond and adapt to the environment. In both stages, glia, hemocytes and neurons were purified by cell sorting from transgenic lines specific to each cell type. *repo-nRFP*, *srp(hemo)Gal4/+; UAS-RFP/+* and *elav-nRFP* were used respectively. The hierarchical clustering of the transcriptomes shows a good correlation between the biological replicates, thus highlighting a low biological variability (**Figure 14-a**). My first finding is that the developmental stage is a main factor of heterogeneity and seems more important than the cell type: the highest correlation is seen for samples of the same stage rather than the same type. Thus, each cell type is more similar to the two others at the same stage than to the same cell type at a different stage.

I next compared the number of differentially expressed genes to evaluate the commonalities and the differences existing between neurons, glia and hemocytes. I defined significantly upregulated genes based on three parameters: number of reads > 15, adjusted p-value < 0.05 and fold change > 1.5 between the number of reads between 2 samples. Upregulated genes

selected in each cell type were compared in the Venn diagram. The comparison of the cell-specific transcriptomes is shown in (**Figure 14-b**) for the embryo (E16), in (**Figure 14-c**) for the larva (L3). Clear trends emerge from this semi-quantitative analysis. The highest number of shared upregulated genes concerns embryonic neurons and glia (in green) (31.1% of all genes included in the diagram), only 5.7% of shared upregulated genes were found between embryonic hemocytes (in red) and glia. Interestingly, the shared upregulated transcripts between hemocytes and glia reaches 18.3% in the larval stage, suggesting that by L3 glia and hemocytes acquire common molecular features despite their distinct origin. The dendrogram also shows that glia and neurons cluster together, hence sharing the highest degree of similarity. Finally, the Venn diagram and the dendrogram indicate that glia and neurons maintain a high degree of similarity in the larva (27.1%), even though the distance between them increases and the distance between glia and hemocytes decreases, likely due to increased specificity (see below).

Interestingly, I found that the number of significantly expressed genes decreases in L3 compared to E16, and this is true for all cell types (transcripts with at least 15 reads, **Figure 14-d,e**). 10838 genes are significantly expressed in E16 glia, 11281 in E16 neurons and 9811 in E16 hemocytes. By L3, 9187 genes are expressed in glia, 9312 in neurons and 7828 in hemocytes, which means that at least 1500 genes are only expressed at E16 in each cell type (**Figure 14-d,e**). This is not due to a lower sequencing depth since the same number of cells was used for each sample and there are no differences between the numbers of reads mapped for embryonic and larval cells (**Sup. Figure 2**). It is not due to a low transcriptional activity in E16 either because calculating the quartiles and median of expression level between E16 and L3 cells showed that they express genes at similar levels.

In sum, results show that the developmental stage has a stronger impact on the transcriptome than the cell type. In addition, throughout development, glia and neurons display the strongest similarity, in line with their common origin and environment. Glia and hemocytes become more similar at L3 stage, due to their common scavenging function. Finally, all larval cells express fewer genes than the embryonic cells.



d

# of genes expressed in E16 glia	10838
# of genes expressed in E16 neurons	11281
# of genes expressed in E16 hemocytes	9811
# of genes of upregulated in E16 glia	3414
# of genes of upregulated in E16 neurons	4177
# of genes of upregulated in E16 hemocytes	2569

e

# of genes expressed in L3 glia	9187
# of genes expressed in L3 neurons	9312
# of genes expressed in L3 hemocytes	7828
# of genes of upregulated in L3 glia	4352
# of genes of upregulated in L3 neurons	4191
# of genes of upregulated in L3 hemocytes	2996

Figure 14: Correlation between the gene expression profiles of neurons, glia and hemocytes at E16 and L3

(a) Dendrogram showing the hierarchical clustering based on the distance between the sample. The height indicates the distance. Euclidean distance, Ward criterion were used for this representation. (b-c) Venn diagrams showing the distribution of differentially expressed genes between E16 neurons (blue), glia (yellow) and hemocytes (red) (b) and L3 neurons (blue), glia (yellow) and hemocytes (red) (c). Only upregulated genes are used for the analysis, no genes are common for all three, intersection (white) is empty. Only significantly upregulated genes in each condition (p-value < 0.05, Fold change > 1.5, expression level >15), identified with Desq2, were considered for this analysis. Since I only used upregulated genes, no genes are in common between all three and the intersection is empty (white). The percentage of genes in each category are presented on the diagram. Percentage was calculated based on all genes used for the analysis. (d,e) The total number of genes expressed and upregulated in E16 (d) and L3 (e).

2. Cell-specific molecular signatures define embryonic glia, neurons and hemocytes

While it is clear that neurons, glia and hemocytes are different cell types, the comparison of the transcriptomes indicates that almost half of the transcripts are common to neurons, hemocytes as well as glia, calling for more refined parameters defining cell identity. Since a significant fraction is upregulated only in one cell type or shared by two cell populations, I predicted that a pairwise comparison associated with quantitative and Gene Ontology (GO) analyses would allow the disclosure of cell-specific signatures and functions.

Comparing E16 neurons and glia shows that 1649 genes are upregulated in neurons (in blue), 992 in glia (in yellow) (**Figure 15–a**, upregulated genes were identified as in **Figure 14–c,d**). Keeping in mind that E16 glia express 10838 genes and E16 neurons express 11281 genes, this means that less than 15% of the expressed genes are specifically upregulated in one cell type compared to the other (the commonly expressed genes are shown in gray). Thus, the majority of the genes are expressed at similar levels between the different cell types, especially between

neurons and glia. Since upregulated genes depends on both fold change and p-value, some genes are more expressed in one cell types however are not considered significantly upregulated due to the high p-value. These genes are referred to as non-differentially expressed genes while the genes that show a fold change of < 1.5 between the different cell types will be referred to as commonly expressed genes. GO term analysis on commonly expressed genes (number of reads >15 in both cell types, fold change < 1.5) identifies biological processes found in all cells such as ncRNA processing, ribosome biogenesis and gene expression (**Sup. Figure 3-a**). On average, the commonly expressed transcripts are expressed at lower levels than the ones upregulated in a specific cell type (**Sup. Figure 3-b,c**). The transcripts upregulated in neurons are mostly involved in neuronal pathways, such as synaptic transmission, neuromuscular junction development and neuron development (**Figure 15-b**, “n” in the graphs indicates the number of genes in each group and “p” the adjusted p-value). The transcripts upregulated in glia are involved in the establishment of the Blood-Brain Barrier (BBB), axon ensheathment and the development of glial cells (**Figure 15-c**).

The comparative analysis between E16 glia and hemocytes reveals a much higher number of transcripts specifically upregulated in either cell type, compared to what was observed in the pairwise comparison between glia and neurons. This is in line with the Venn diagram in **Figure 14-b**. 1945 transcripts are upregulated in hemocytes (in red) and 2937 in glia (in yellow) (**Figure 15-d**). The GO term analysis on the transcripts upregulated in hemocytes shows very few terms related to immune functions. In fact, the only immune related GO term is ‘positive regulation of immune functions’, which comprises 39 genes (**Figure 15-e**). Among these genes, I found 7 members of the Immune Deficiency (IMD) pathway, an evolutionary conserved signaling pathway that activates the immune response through NF- κ B. These include *immune deficiency*,

Peptidoglycan recognition protein LC as well as *Kenny* and *I-kappaB kinase β* , the two subunits of the *Drosophila* IKK complex involved in the activation of the transcription factor *Relish* (Bandarra et al., 2014; Choe et al., 2002; Corbo and Levine, 1996; Gaudet et al., 2011; Lemaitre et al., 1995). This group also contains 14 members of the Toll pathway including *Toll*, *Dorsal-related immunity factor*, *spatzle*, *Spatzle-processing enzyme*, *persephone*, *Gram-negative bacteria-binding protein 1*, *Peptidoglycan-recognition protein SA*, *pelle*, and *cactus* (Gobert et al., 2003; Jang et al., 2006; Rutschmann et al., 2000; Weber et al., 2003). In the case of E16 glia, I found glial specific functions such as septate junction assembly (involved in the establishment of the BBB), glial cell differentiation and neural functions such as motoneuron axon guidance (Abbott, 2005; Barres, 2008) (**Figure 15-f**).

Altogether, these data reveal that only a minority of genes is cell-specific and embryonic glia are closer to neurons than to hemocytes. On the other hand, the genes that are specifically upregulated in E16 recapitulate the specific functions of neurons and glia, while few cell-specific terms are identified in hemocytes.

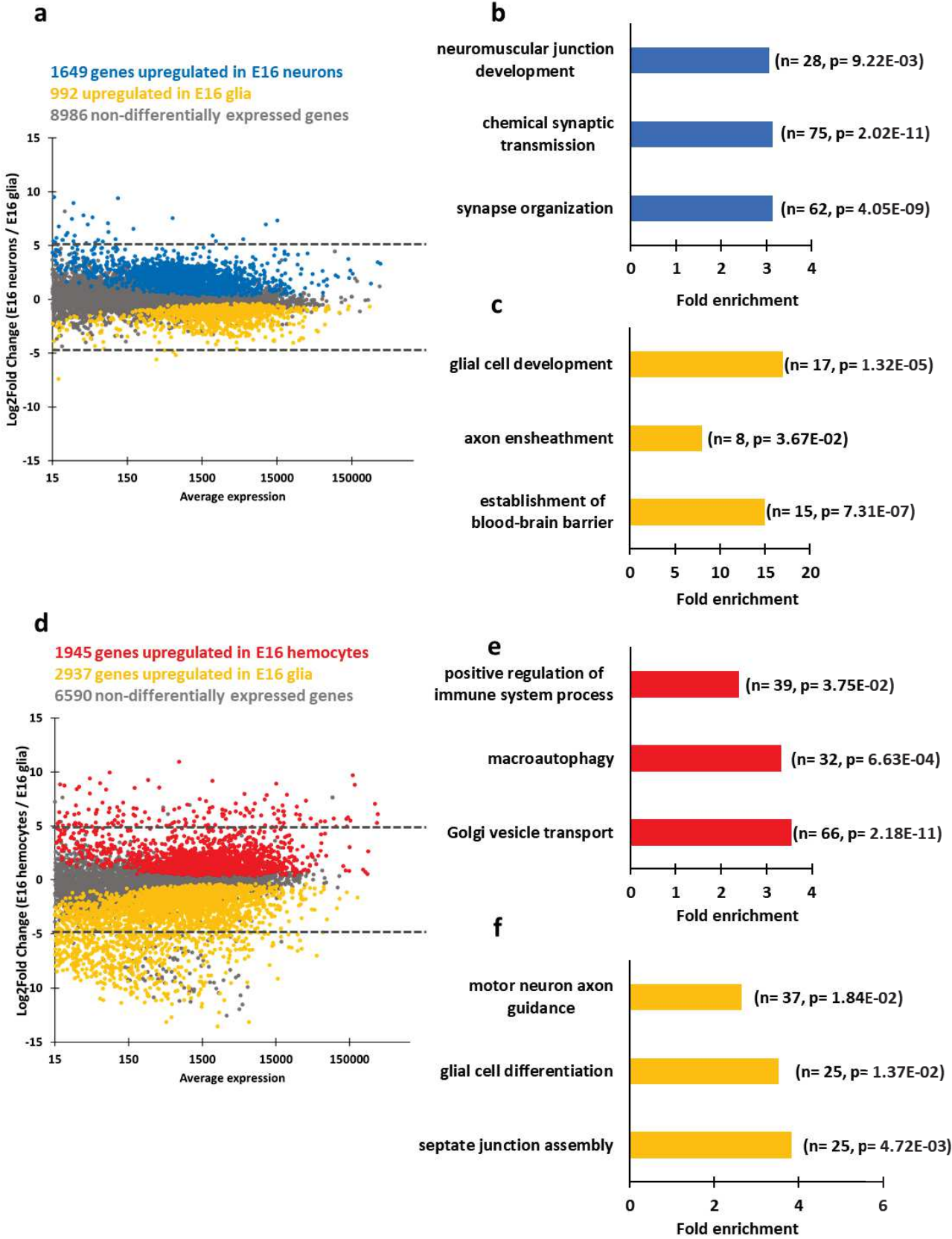


Figure 15: Gene ontology analysis on genes differentially expressed between E16 glia and neurons and E16 glia and hemocytes.

(a) The x-axis is the average gene expression levels ($n = 3$), and the y-axis is the \log_2 fold change E16 neurons/E16 glia. Genes significantly enriched in E16 neurons are shown in blue, genes significantly enriched in E16 glia are shown in yellow and non-differentially expressed genes are shown in gray. (b) Gene Ontology (GO) term enrichment analysis in E16 neurons (blue) and E16 glia (yellow). The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets. (c) Transcriptome comparison of E16 hemocytes and glia. Scatter plot as in (A), the y-axis is the \log_2 fold change E16 hemocytes/E16 glia. Genes significantly enriched in E16 hemocytes are shown in red, genes significantly enriched in E16 glia are shown in yellow and non-differentially expressed genes are shown in gray. (d) Gene Ontology (GO) term enrichment analysis in E16 hemocytes (red) and E16 glia (yellow) The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

3. Hemocytes acquire new functions over time

Cell functions adapt to changing environments during the lifecycle, typically embryos live in autarchy, whereas larvae function as open systems, due to feeding, respiration, access to pathogens and numerous other features. To evaluate the impact of the functional changes on the cell-specific transcriptional landscapes, I analyzed neurons, glia and hemocytes in the larva and performed pairwise comparisons to those described above.

Overall, cells become more different with time indicating increased specialization by the larval stage (compare the log scale $-15 +15$ in the embryo in **Figure 15** to the log scale $-20 +20$ in the larva in **Figure 16**). I also found more differentially expressed genes between the cell types. More specifically, 2500 genes are upregulated in L3 neurons compared to glia and 2279 genes upregulated in L3 glia compared to neurons (**Figure 16-a**). The increased number of the L3 specifically upregulated genes is in line with the increased difference between the two cell types observed in the dendrogram (**Figure 14-a**). This is even more interesting given that larval cells

express fewer genes than embryonic cells. GO term analysis on the genes upregulated in L3 neurons identifies neuronal functions and molecular cascades linked to signaling such as neuropeptide signaling pathway, neurotransmitter secretion and signal release from synapses (**Figure 16-b**). By comparison, E16 neurons are upregulated in transcripts linked to neuron development and neuromuscular junction development (**Figure 15-b**). In the case of larval glia, I found the GO term septate junction assembly also found in the embryo. In addition, terms related to the differentiation of glia and gliogenesis were also identified (**Figure 16-c**), likely due to the extensive glial proliferation undergoing during the larval life (Pereanu et al., 2005).

While the comparison between neurons and glia shows that the number of genes upregulated in a cell-specific manner increases by L3, the opposite happens upon comparing L3 glia and hemocytes. The number of upregulated genes in hemocytes compared to glia significantly decreases by L3 (1701 genes in red) (**Figure 16-d**), which fits with the increased number of shared transcripts observed in **Figure 14-b**. Compared to the embryo, I found many more terms involved in immune functions in L3 hemocytes: phagocytosis, positive regulation of antimicrobial humoral response and positive regulation of immune effector process (**compare Figure 15-e and Figure 16-e**). The clear shift in the expression profile of the hemocytes indicates a major change in the function of these cells during development. The genes upregulated in glia (2941 genes in yellow) are involved in glial functions such as synapse organization and axon guidance (**Figure 16-f**) a function that has already been investigated (Bittern et al., 2021).

In sum, the cell-specific features and specialization increase over time as in the larva hemocytes start expressing immune related functions. In addition, glia share more upregulated transcripts with hemocytes and less with neurons, compared to the embryonic glia.

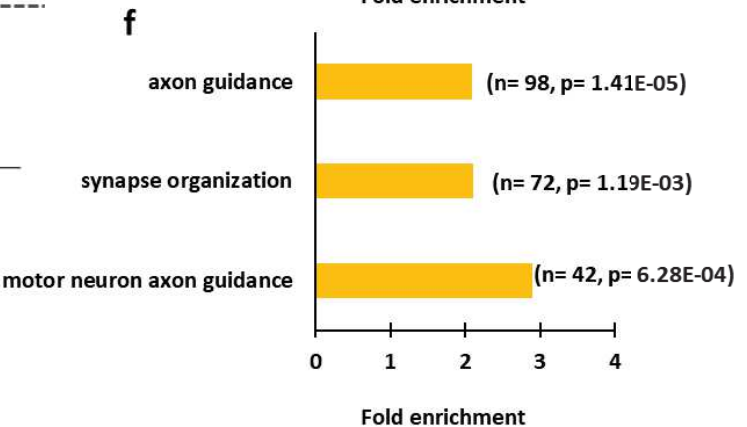
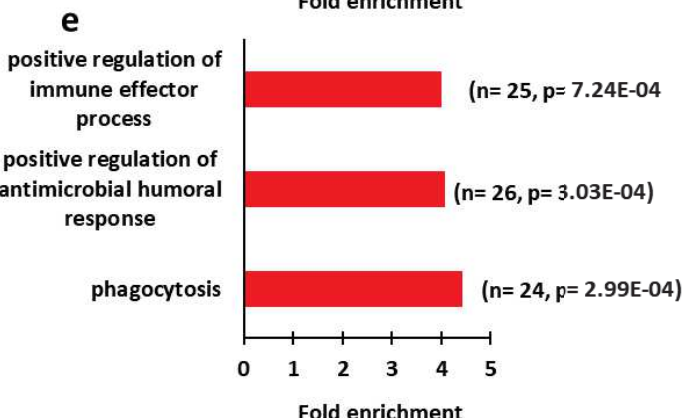
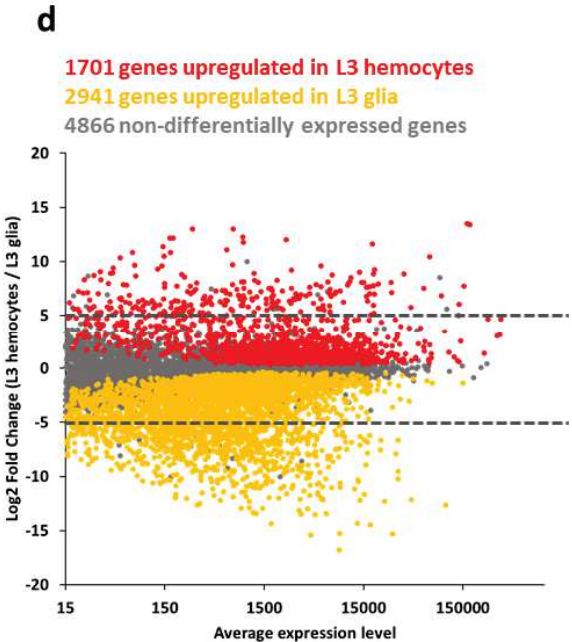
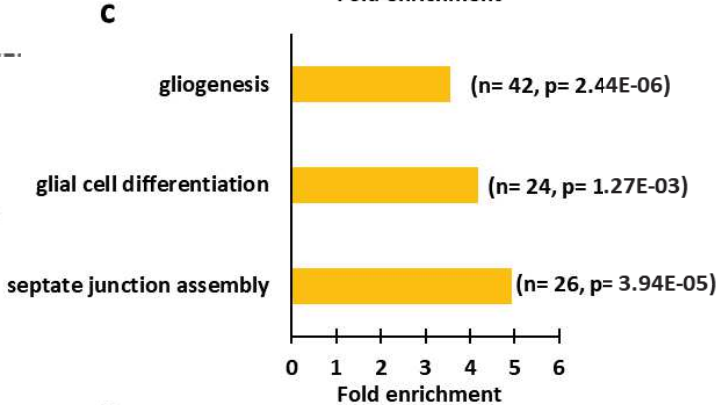
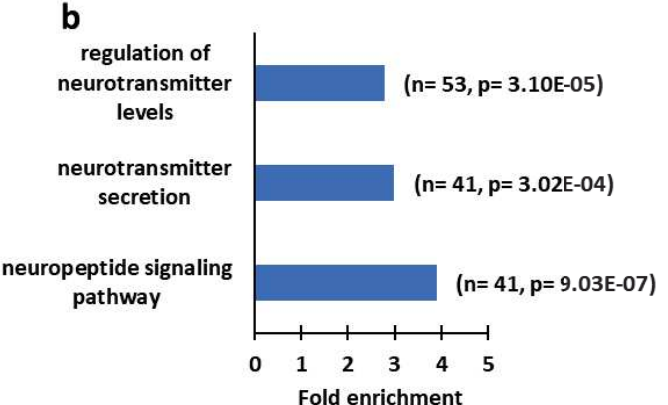
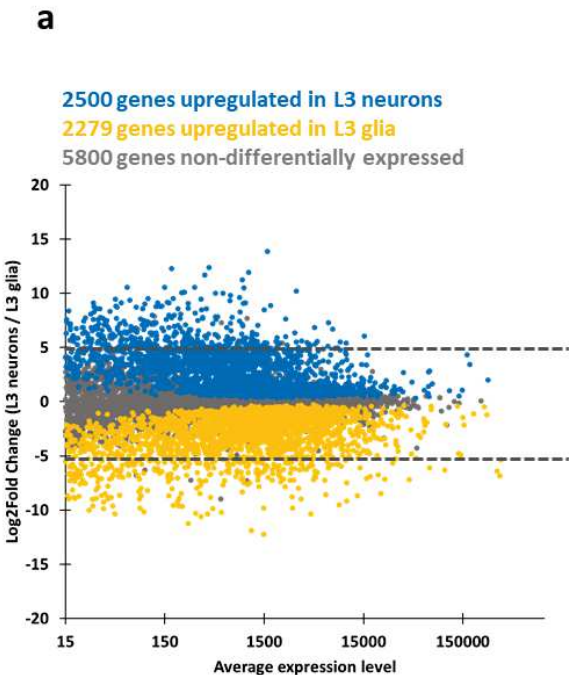


Figure 16: Gene ontology analysis on genes differentially expressed between L3 glia and neurons and L3 glia and hemocytes.

(a) The x-axis is the average gene expression levels ($n = 3$), and the y-axis is the log₂ fold change L3 neurons/ L3 glia. Genes significantly enriched in L3 neurons are shown in blue, genes significantly enriched in L3 glia are shown in yellow and non-differentially expressed genes are shown in gray. (b) Gene Ontology (GO) term enrichment analysis in L3 neurons (blue) and L3 glia (yellow). The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets. (c) Transcriptome comparison of L3 hemocytes and glia. Scatter plot as in (a), the y-axis is the log₂ fold change L3 hemocytes/L3 glia. Genes significantly enriched in L3 hemocytes are shown in red, genes significantly enriched in L3 glia are shown in yellow and non-differentially expressed genes are shown in gray. (d) Gene Ontology (GO) term enrichment analysis in L3 hemocytes (red) and L3 glia (yellow). The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

4. Glia and hemocytes express specific scavenger receptors

Plasmatocytes constitute the vast majority of the hemocytes (95%) and act as scavenger cells in homeostatic as well as challenged conditions. In the embryo, they are necessary for the elimination of apoptotic bodies and for the clearance of cellular debris. Prior to the formation of the BBB, they also contribute to this function within the developing CNS (Franc et al., 1996; Kurant et al., 2008; Manaka et al., 2004). Once the glial cells form the BBB, however, plasmatocytes no longer have access to the CNS and by late embryogenesis the scavenging function within the nervous system is taken up by glia (Kurant et al., 2008). Thus, plasmatocytes and glia can be considered as the macrophages acting outside and inside the nervous system, respectively. The question then arises about the similarity of their transcriptional landscapes. Do glia express the same immune pathways as the hemocytes, in addition to the neural pathways, or do they represent a distinct class of macrophages?

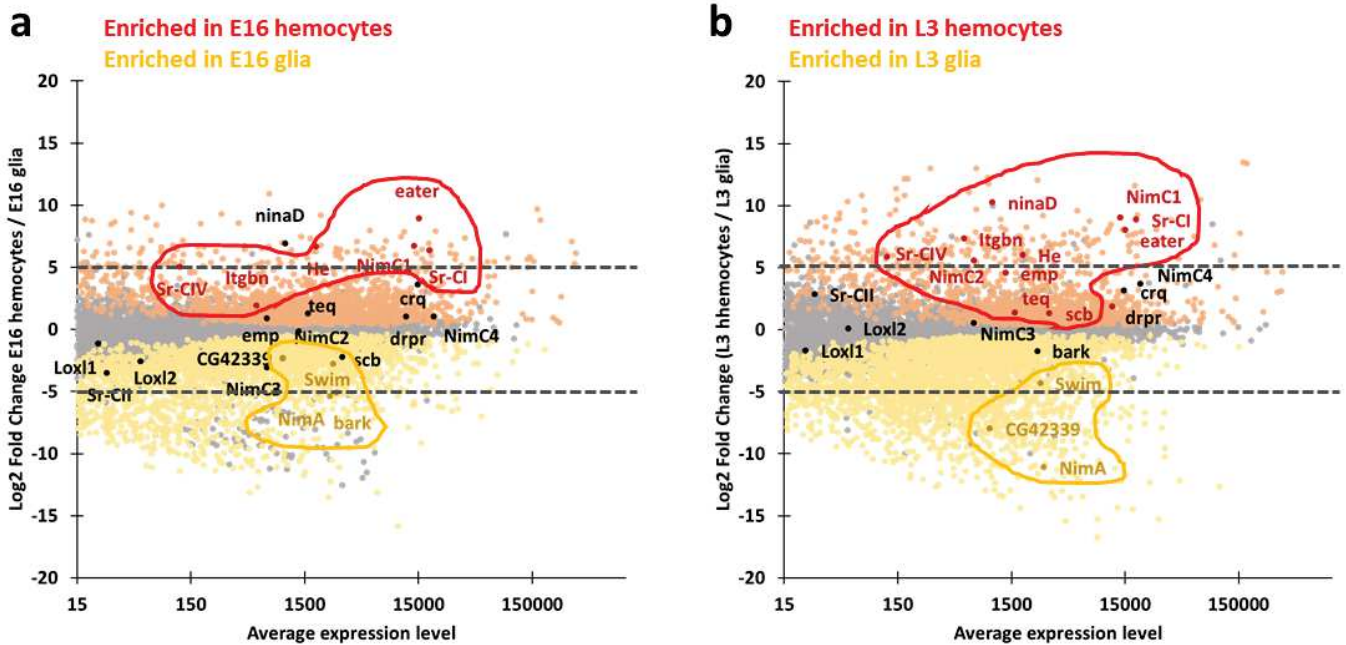
A key molecular signature of macrophages is the expression of scavenger receptors (Kocks et al., 2005; Kurucz et al., 2007; Lebestky et al., 2000; Lemaitre et al., 1996; Rämét et al., 2002; Tepass et al., 1994) which I therefore used to compare hemocytes and glia in embryos and larvae. The scatter plots comparing E16 (**Figure 17-a**) or L3 hemocytes and glia (**Figure 17-b**) allowed us to identify three groups of receptors based on their relative levels of expression. Group 1 (in red) is upregulated in hemocytes, group 2 (black) is expressed in both cell types at comparable levels and group 3 (in yellow) is upregulated in glia. The group I receptors have been shown to be involved in the phagocytosis of gram-positive and gram-negative bacteria (Pearson et al., 1995; Rämét et al., 2001), nicely matching the hemocyte-specific function of pathogen phagocytosis. Group 1 includes very well studied receptors such as *eater*, *Nimrod C1* (*NimC1*), *Nimrod C2* (*NimC2*), *Scavenger receptor class C, type I* (*Sr-CI*) and *Scavenger receptor class C, type IV* (*Sr-CIV*). *Eater*, *NimC1* and *NimC2* belong to the Nimrod sub-family Nimrod C-type, transmembrane proteins containing multiple epidermal growth factor (EGF) like repeats called Nimrod (NIM) repeats (Kurucz et al., 2007). SR-CI and SR-CIV belong to the *Drosophila* class C scavenger receptors, SR-CI contains several known motifs, including two complement control protein (CCP) domains, a somatomedin B domain and a Meprin, A-5 protein, and receptor protein-tyrosine phosphatase Mu (MAM) domain (Rämét et al., 2001). As for the receptors that are commonly expressed in hemocytes and glia (Group 2, in black), three known receptors are found in this group, *Nimrod C4* (*NimC4*), *draper* (*drpr*) and *croquemort* (*crq*) a member of the class B scavenger receptors homologous to the mammalian CD36, which was believed to be only expressed in hemocytes (Franc et al., 1996). A common trait of these receptors is that they are involved in the phagocytosis of apoptotic bodies. In fact, *Drpr* and *NimC4*, members of the Nimrod family, act both in hemocytes and in glia to promote the clearance of apoptotic cells (Kurant et al., 2008;

Manaka et al., 2004; Melcarne et al., 2019; Roddie et al., 2019). The three genes that are specifically upregulated in glia (Group 3, in yellow) have not been investigated for their role in phagocytosis. CG42339, the ortholog of the human SBSPON (somatomedin B and thrombospondin type 1 domain containing), and Secreted Wg-interacting molecule (Swim) both contain a somatomedin B domain similarly to Sr-CI. Nimrod A (NimA) is a member of the Nimrod family and is a Drpr-like receptor containing the same domains: one NIM domain followed by EGF repeats and one Emilin (EMI) domain (Callebaut et al., 2003). Its *C. elegans* orthologs are involved in the phagocytosis of apoptotic cells (Kurucz et al., 2007; Mangahas and Zhou, 2005). Finally, the number of scavenger receptors specifically upregulated in hemocytes increases in the larva (6 at E16, 11 at L3, shown in red, **Figure 17-a,b**). In parallel, the number of receptors commonly expressed between hemocytes and glia decreases in the larva (13 at E16, 8 at L3, shown in black).

To validate the expression of these genes, I performed quantitative reverse transcription PCR (RT-qPCR) assays on glia and hemocytes specifically sorted from L3. I tested and confirmed the purity of the sorted populations by quantifying the levels of expression of *Hemolectin (Hml)* and *reverse polarity (repo)*, known genes specifically expressed in hemocytes and glia, respectively (**Sup. Figure 4**). The qPCR data showed that *NimC1* and *eater* are significantly upregulated in L3 hemocytes compared to glia (**Figure 17-c,d**), *NimC4* and *drpr* show comparable expression levels (**Figure 17-e,f**); *crq* seems to be more expressed in hemocytes however the difference is not statistically significant (**Figure 17-g**), similarly to what was found in the transcriptome. *NimA* was only found in glia (**Figure 17-h**). The finding that some receptors are commonly expressed between glia and hemocytes, whereas others are cell-specific recapitulates their common function of phagocytosing apoptotic cells and at the same time emphasizes the role

of hemocytes in host defense against pathogens. In addition, the higher levels of the scavenger receptors in L3 than in E16 hemocytes prove an increased functional specificity of hemocytes in the larva.

To conclude, hemocytes have a strong commitment to the immune function in L3 and they express different phagocytic receptors than glia.



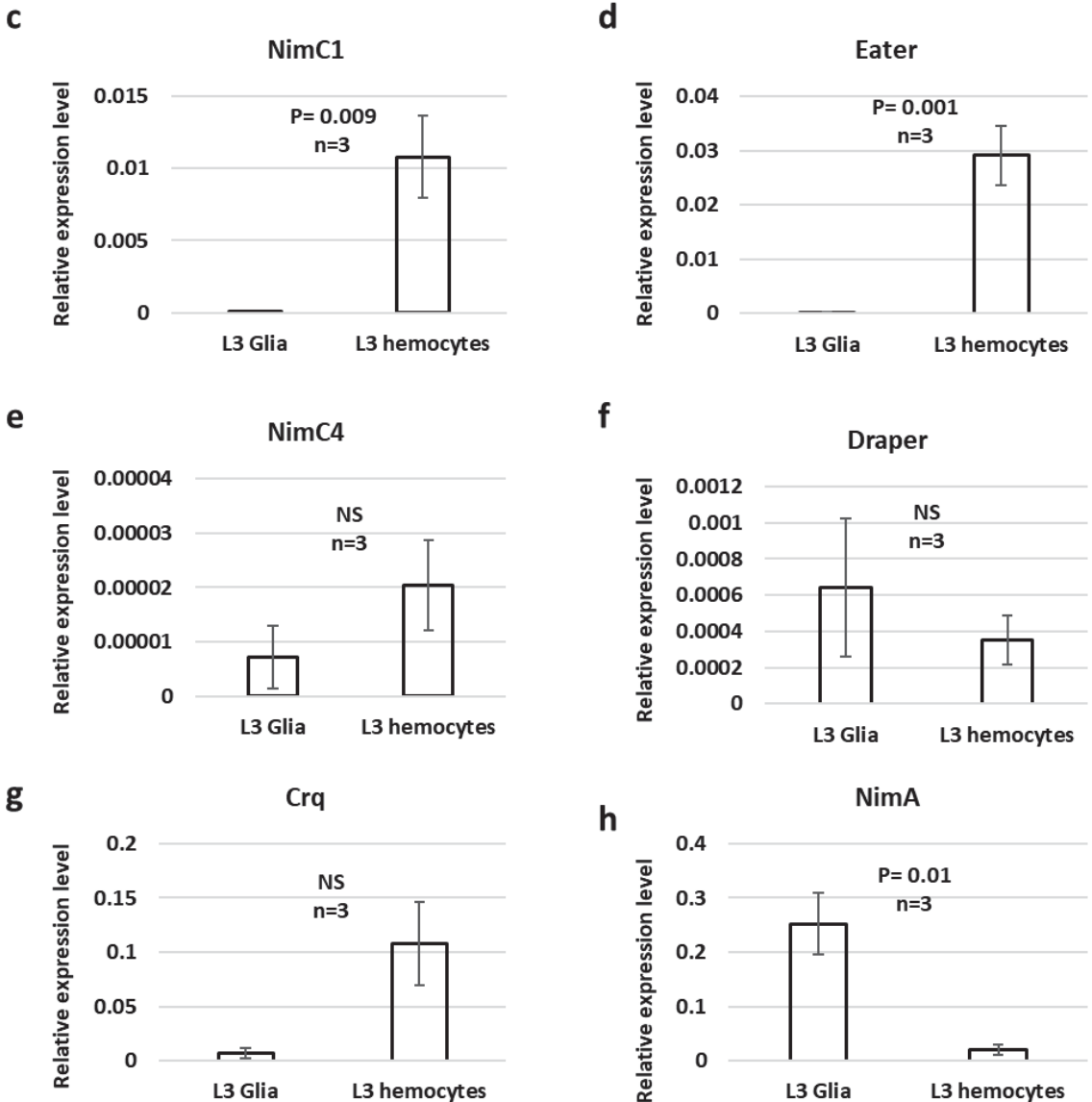


Figure 17: comparing scavenger receptors expression in glia and hemocytes

(a,b) Scatter plots as in figure 2-c and figure 3-c showing subsets of scavenger receptors. In black, scavenger receptors non-differentially expressed between E16 glia and hemocytes (a) or L3 hemocytes and glia (b). In red, scavenger receptors that are upregulated in E16 hemocytes (a) or L3 hemocytes (b) and in yellow, scavenger receptors that are upregulated in E16 glia (b) or L3 glia (b). (c-f) Quantitative reverse transcription PCR (RT-qPCR) on hemocytes and glia sorted from third instar wandering larvae. The levels of NimC1 (c), eater (d), NimC4 (e) and Drpr (f) were quantified and relative expression levels were calculated based on the expression of Actin5C and Ribosomal Protein 49. N=3, p-value calculated using Student’s t-test and indicated on graphs, error bars represent the standard deviation.

5. Stage-specific transcriptional landscapes

Since the transition from the embryo to the larva involves massive physiological changes that accompany the switch from a closed and isolated environment to an open environment, I suspected that such changes have a strong impact on the organism as a whole. To this purpose, I assessed whether there is a stage-specific molecular signature that is common to all cell types. To focus the analysis on stage-specific differences, I combined all cells deriving from E16 embryos and compared their transcriptomes to those of (combined) cells deriving from L3. The dot plot comparing E16 to L3 (**Figure 18-a**) reveals that fewer genes are upregulated in L3 (1746 genes in purple) compared to E16 (3386 genes in green). This could be explained by the fact that at E16 the different cell types cells are overall closer to each other than at L3, as seen in **Figure 14-a**. In addition, the different cell types express fewer genes at L3 than at E16 and become more specialized. Interestingly, I do find genes that are upregulated in one stage compared to the other regardless of the above mentioned quantitative differences. This strongly suggests a stage-specific signature common to all cells.

The GO term analysis shows that the genes upregulated in L3 are mainly involved in tRNA aminoacylation, telomere maintenance and DNA repair (**Figure 18-b**). Aminoacyl-tRNA synthetases attach aminoacids to tRNAs and as such are key regulators of protein synthesis. Their levels increase upon starvation to compensate for the limiting amounts of aminoacids, in line with the fact that wandering larvae no longer feed (Putzer and Laalami, 2013). I was puzzled by the upregulation of genes involved in different DNA repair pathways such as nucleotide-excision repair, DNA mismatch repair and double strand break (DSB) repair. I hence assessed the expression levels of the genes belonging to the DNA repair term (**Figure 18-d**) on the scatter plot: Rad50, NBS and meiotic recombination 11 (Mre11), which form the MRN complex, as well as

Telomere fusion (*Tefu*), the ortholog of Ataxia Telangiectasia Mutated (*ATM*). These genes have been implicated in Double Strand Break (DSB) repair (Wang et al., 2014b), in DNA replication restart after replication stress (Gatei et al., 2014) and in preventing telomere fusion as well as chromosome breakage in *Drosophila* (Ciapponi et al., 2004). To first validate the stage-specific enrichment of these transcripts, I performed qPCR quantification on RNA extracted from whole embryos and whole larvae: *Rad50* (**Figure 19-a**) and *Mre11* (**Figure 19-b**) are upregulated in L3 and *NBS* tends to be upregulated (**Figure 19-c**). I then estimated the relative levels of DSB using the phosphorylated H2AV (γ H2AV) *Drosophila* marker, the equivalent of γ H2AX in mammals. When DNA damage forms double stranded breaks, it triggers the phosphorylation of the histone H2AX, a variant of the H2A protein family (Redon et al., 2002). H2AX is phosphorylated by kinases such as ATM and the phosphorylated protein, γ H2AX, is the first step in recruiting and localizing DNA repair proteins. Western blot assays on histone extracts from E16 embryos and L3 larvae show that the levels of γ H2AV decrease in L3 compared to E16 (**Figure 19-d**), contrary to my initial hypothesis. I tested the potential differences in proliferation and apoptosis levels, since previous studies in *Drosophila* have shown that null mutants of *Rad50*, *Mre11* and *NBS* cause increased apoptosis and cell division defects (Ciapponi et al., 2004; Gao et al., 2009; Gorski et al., 2004) and knock-down of *Rad50*, *Mre11*, *NBS* and *tefu* in enterocyte in the adult fly caused hyperproliferation and increased apoptosis (Park et al., 2018). I assessed whether these genes are needed to control proliferation and apoptosis at L3 stage and used tested RNAi lines (Park et al., 2018) to knock-down of *Rad50*, *Mre11* and *NBS*. No effect on the number of proliferating nor apoptotic cells at L3 stage was found (data not shown). These findings suggest that the DSB repair complex is required for additional processes that are stage-specific.

The genes upregulated in E16 are involved in cytoplasmic translation and cuticle development (**Figure 18-c**). Genes included in the GO term cytoplasmic translation code for ribosomal proteins and translation machinery, such as translation initiation factors. Ribosomal protein transcripts and ribosomes are known to be maternally deposited and highly abundant in early embryos (Qin et al., 2007) but new ribosomes are built by the end of embryogenesis, hence in line with the transcriptomic data. The ‘cuticle development genes’, plotted in (**Figure 18-e**), include 42 genes coding for Cuticular proteins, 21 genes coding for members of the Tweedle family and 7 genes coding for CPR cuticle protein family. All these genes are predicted to be present in the extracellular matrix (ECM) (Cornman, 2009; Karouzou et al., 2007; Naba et al., 2016), which implies a role of all three cell types in the deposition of ECM molecules and the cuticle at embryonic stage, a feature that has already been described for the hemocytes (Brown, 2011; Martinek et al., 2008).

My data reveal that different developmental stages are associated with specific molecular signatures, regardless of the cell type.

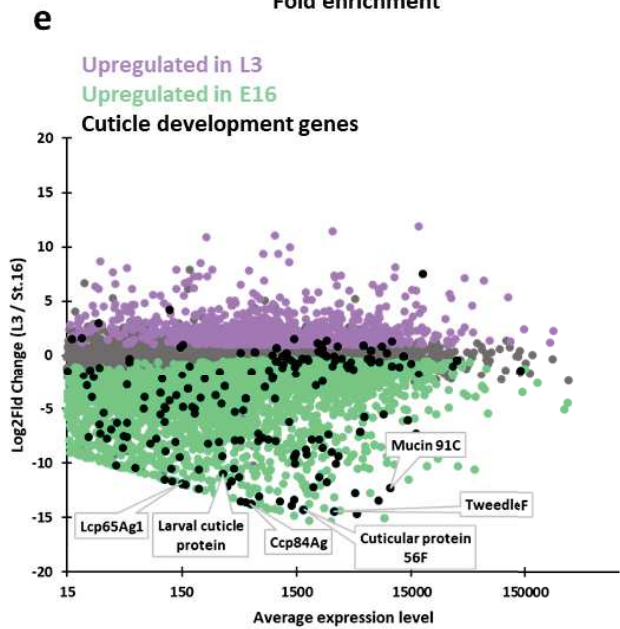
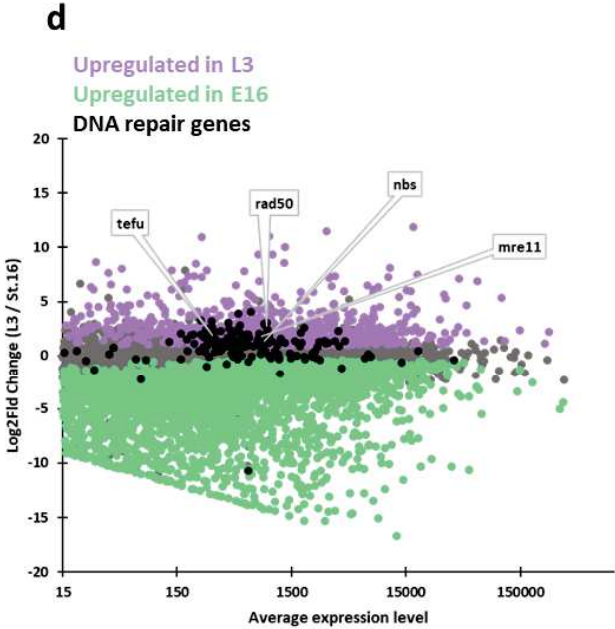
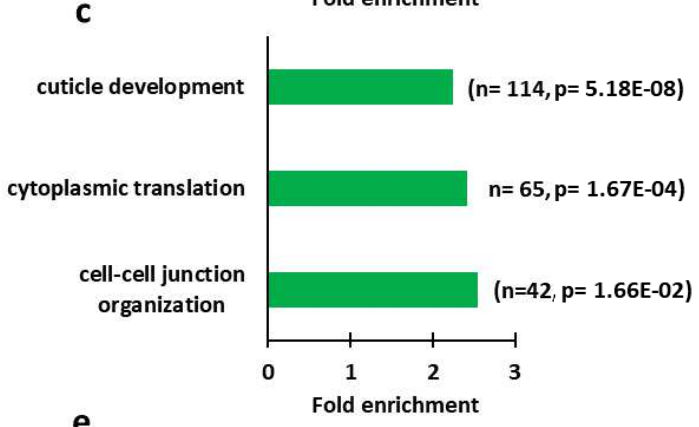
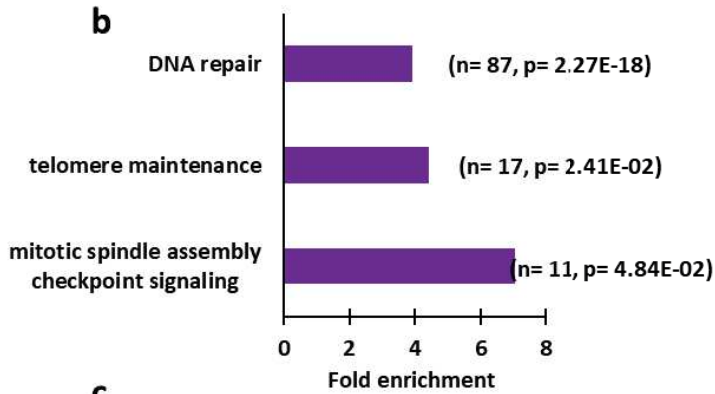
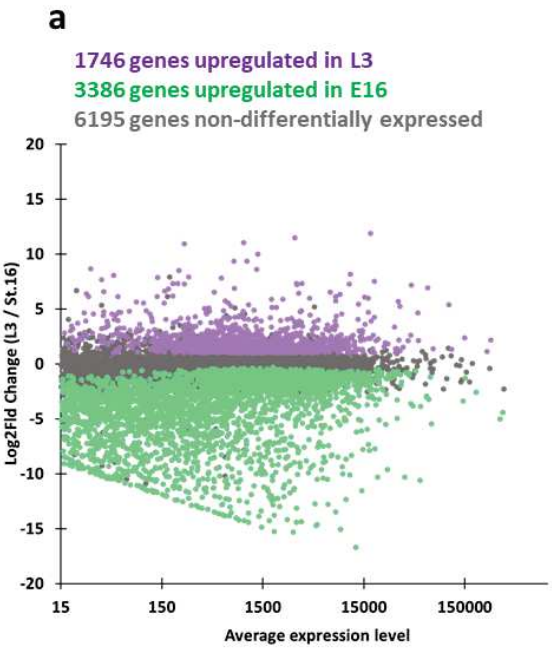


Figure 18: Comparing the expression profile of cells from E16 and L3.

(a) The x-axis is the average gene expression levels ($n = 3$), and the y-axis is the \log_2 fold change L3 cells/E16 cells. Genes significantly enriched in E16 are shown in green, genes significantly enriched in L3 are shown in purple and non-differentially expressed genes are shown in gray. (b) Gene Ontology (GO) term enrichment analysis in L3 (purple) and E16 (green). The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets. (c-d) Scatter plots as in (a) highlighting genes involved in cuticle development and DNA repair. In (c) in black, all genes belonging to the GO term cuticle development. In (d) in black, all genes belonging to the GO term DNA repair.

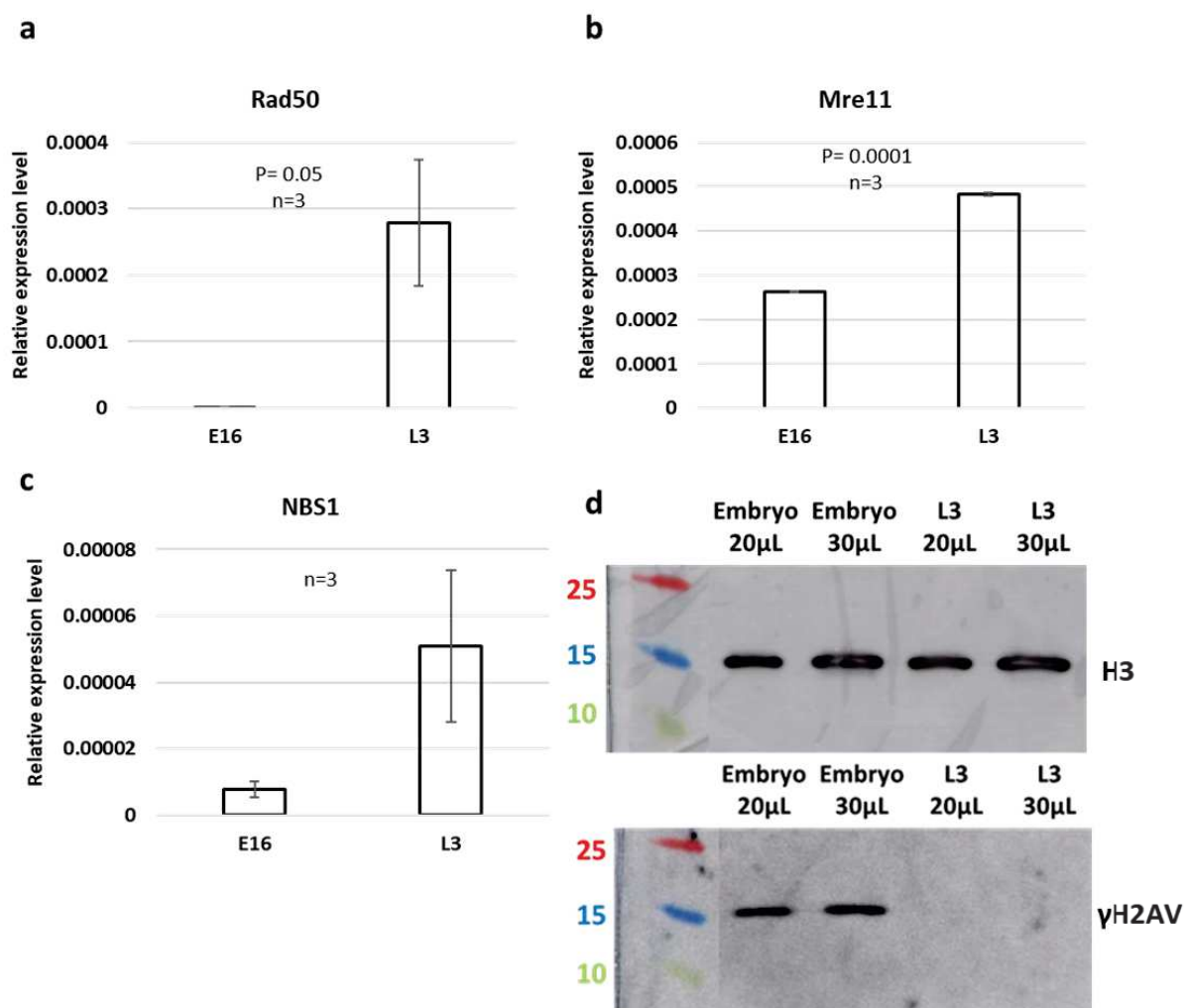


Figure 19: Quantification of components of DNA repair and γ H2AV.

(a,c) Quantitative reverse transcription PCR (RT-qPCR) on cells from stage 16 embryos and third instar wandering larvae. The levels of Rad50 (a), Mre11 (b) and NBS (c) were quantified and relative expression levels were calculated based on the expression of Actin5C and Ribosomal Protein 49. N=3, p-value calculated using Student's t-test and indicated on graphs, error bars represent the standard deviation. (d) Western blot quantification of γ H2AV. Histones extracted from cells from stage 16 embryos (E16) and third instar larvae (L3) were blotted on nitrocellulose membranes and probed with anti- γ H2AV antibody. An antibody that recognizes all histones H3, anti-H3, was used to confirm equal loading of lysates.

IV- Discussion:

Understanding the bases of cellular identity makes the object of intense investigation and genome-wide techniques now allow unbiased analyses at unprecedented resolution. I identified the molecular signature that characterizes a cell type upon RNAseq assays on sorted *Drosophila* neurons, glia and hemocytes, representative examples of related and unrelated cell types. I here showed that the molecular identity of a cell results from the combined action of intrinsic and extrinsic cues and I established the relative impact of origin, function and time on the three cell types. I identify cell-specific as well as stage-specific signatures and I show that cells maintain their identity during development but adopt different states to accommodate to the evolving needs of the organism.

1. Cell identity is defined by a relatively low number of transcripts and is heavily impacted by temporal cues

The identification of cell subtypes within the same cell subpopulation requires the analysis of the most expressed genes, on which the current single cell approaches are based. In this way, even relatively similar cells can be distinguished. By comparison, bulk transcriptomic assays allow deeper analyses and better resolution when different cell populations are compared. The bulk transcriptomic data allow us to compare related and unrelated cell types and to take into account the progression occurring between two developmental stages.

The majority of genes are not differentially expressed between neurons, glia and hemocytes, both in the embryo and in the larva, so levels of expression per se are not indicative of cell identity. The three cell types share a very high number of transcripts (around 25% of all expressed genes). Such high number of similar genes cannot be due to the common origin/function

because neurons and hemocytes share neither. This goes along with another study that analyzed genes expression at different stages of development and found that 48.8% of all Flybase genes are commonly expressed between all stages (Daines et al., 2011). No more than 2000 genes are upregulated in each cell type (**Figure 14-b,c**), hence, a relatively low number of genes is sufficient to provide cell identity. In average, the transcripts enriched in a specific cell type are expressed at high levels, suggesting that the cell-specific transcripts are under the control of signaling pathway rather than being constitutively expressed at basal levels. Thus, cell identity is defined by a combination of cell-specific transcript enrichment and high levels of expression. Of note, even amongst the common transcripts, many are expressed at moderate levels and seem to have a house keeping function while more cell-specific functions are associated to those that are expressed at higher levels. For instance, common genes expression at relatively low levels in neurons and glia (# of reads < 300) are involved in DNA replication while among genes expressed at high levels (> 3000) I find genes involved in axons.

Cell identity involves the stable expression of specific morphological and functional features, even though cells may lose or acquire some traits and assume different states over time. For this reason, I expected each cell type to group together at different stages. Much to my surprise, however, the comparative analysis of the three cell types in the embryo and in the larva shows that the developmental stage is the first factor of heterogeneity, more than the cell type. Embryonic hemocytes are closer to embryonic glia and neurons than to larval hemocytes, matching the observation that the embryonic and the larval hemocytes share many fewer transcription factors (n=22) than those that are common to the three embryonic cell types (n=122). Whether the impact of the developmental stage is due to the fact that the epigenome of “young”, embryonic, cells, is not yet well established, awaits further investigations.

Amongst the stage-specific genes, ECM related GO terms emerged as being upregulated in E16 compared to L3 cells. In line with the fact that the ECM is a key player in morphogenesis and organogenesis, different types of embryonic cells hence express large quantities of ECM transcripts to prepare for the larval life and the immense growth happening during this stage. L3 cells show an enrichment of transcripts involved in DNA damage and repair. In particular, the three components of the MRN complex (Rad50, Mre-11 and NBS) and Tefu, that respond to DNA double-strand break and are involved in replication fork dynamics, telomere maintenance and even response to viral infection (Ciapponi et al., 2004; Syed and Tainer, 2018). The increase of the expression of these genes is seemingly not due to DSB since there is more DBSs markers in L3 than in larvae. A recent study showed that the yeast equivalent of the MRN complex (MRX) is involved in controlling expression level through modulating chromatin interaction and repressing the loci near its binding site (Forey et al., 2021). Together with the finding that fewer genes are expressed in general in L3 than in E16 cells, the observation that DNA repair genes are more expressed in L3 cells may open novel perspectives to understand the developmental mechanisms regulating open vs. closed chromatin states.

In sum, a relatively low number of highly expressed transcripts characterizes each cell type and different cell types group together according to the developmental stage.

2. Neurons, hemocytes and glia, differences and commonalities

Fly neurons and glia differentiate in the ectoderm from the multipotent stem cells called neuroblasts. These two cell types constantly interact to ensure the homeostasis of the nervous system, like their vertebrate counterparts. Fly neurons sense, process and transmit information (dendrite and axonal growth, synaptogenesis); fly glia ensure neuronal development, function and survival (regulation of axon guidance and ensheathment, modulation of synaptic activity and nervous system insulation) (Yildirim et al., 2019). In addition, glia also act as the microglia, vertebrate immune cells that invade the nervous system during development and prune synapses during circuit establishment (Neumann et al., 2009; Wu et al., 2015). Fly glia represent the major phagocytic cell type in the CNS and eliminate nearly all neuronal debris generated during development (Kurant, 2011; Tasdemir-Yilmaz and Freeman, 2014). They are needed for the clearance of apoptotic bodies and debris within the nervous system at embryonic and larval stages (Abrams et al., 1993; Rogulja-Ortmann et al., 2007; Sonnenfeld and Jacobs, 1995). Hemocytes differentiate in the mesoderm, provide a myeloid-like function and contribute to humoral immunity (Imler and Bulet, 2005; Kragol et al., 2001; Tepass et al., 1994). These motile cells patrol the organism and constitute the main phagocytic cell population outside the nervous system (Lemaitre et al., 1996; Tepass et al., 1994). Hence, fly glia and neurons share the same origin while glia and hemocytes the immune functions.

Origin has a significant impact on the transcriptional landscapes: glia and neurons share a large number of molecular signatures, even though they become slightly more diversified by L3 (31,3% of commonly upregulated genes in the embryo, 27,1% in the larva, **Figure 14-b,c**). The transcripts commonly upregulated in embryonic glia and neurons belong to 47 GO terms for biological processes (≥ 2 -fold enrichment) and to 28 GO terms in the larva. Importantly, a

substantial fraction of these shared GO terms concerns processes related to axons, synapses, and glia, i.e., cell-specific processes. In contrast, neurons and hemocytes share many fewer transcripts (5,9% in the embryo, 3,8% in the larva), with no GO term significant enrichment in the embryo and 24 shared GO terms in the larva (2_≥fold enrichment), none of whose relates to a cell-specific process. Finally, embryonic glia and hemocytes share the same very low degree of similarity as that observed between embryonic neurons and hemocytes (5,7%), with a relatively low number of GO terms (15), none of whose relates to cell-specific processes. The lack of hemocyte-specific functions amongst these GO terms is likely due to the fact that the major function of the embryonic hemocytes is to secrete ECM components and to phagocytose apoptotic cells, both of which are also done by glia. The situation changes in the larva, a stage at which hemocytes and glia share 18,3% of upregulated transcripts and a very high number of GO terms (197), a significant number of whose is related to immunity, including phagocytosis. The presence of hemocyte-specific GO in the larva but not in the embryo calls for a real developmental switch and not for a bias in the definition of the GO terms. This combined with the reduced number of expressed genes at L3 compared to E16 suggests a higher specialization, with more refined groups of genes being expressed in the larva.

Plotting the scavenger receptors in E16 and L3 cells emphasizes the acquisition of specific phagocytic potentials of L3 hemocytes, enriched in transcripts associated with pathogen clearance. The comparison also shows that glia are not just macrophages that are also able to perform neural functions. Rather, fly glia constitute a specific type of macrophages with their own identity, much like vertebrate microglia, which also express very specific immune features compared to macrophages and monocytes. Since NimA remains expressed in glia in the adult brain, based on the single cell data visualized on SCoPe (<http://scope.aertslab.org/>, Davie et al., 2018), pursuing

the analysis of this and other glial-specific genes will help deciphering the role and mode of action of glia in the immune response, which so far has relied only on the study of Draper and NimC4.

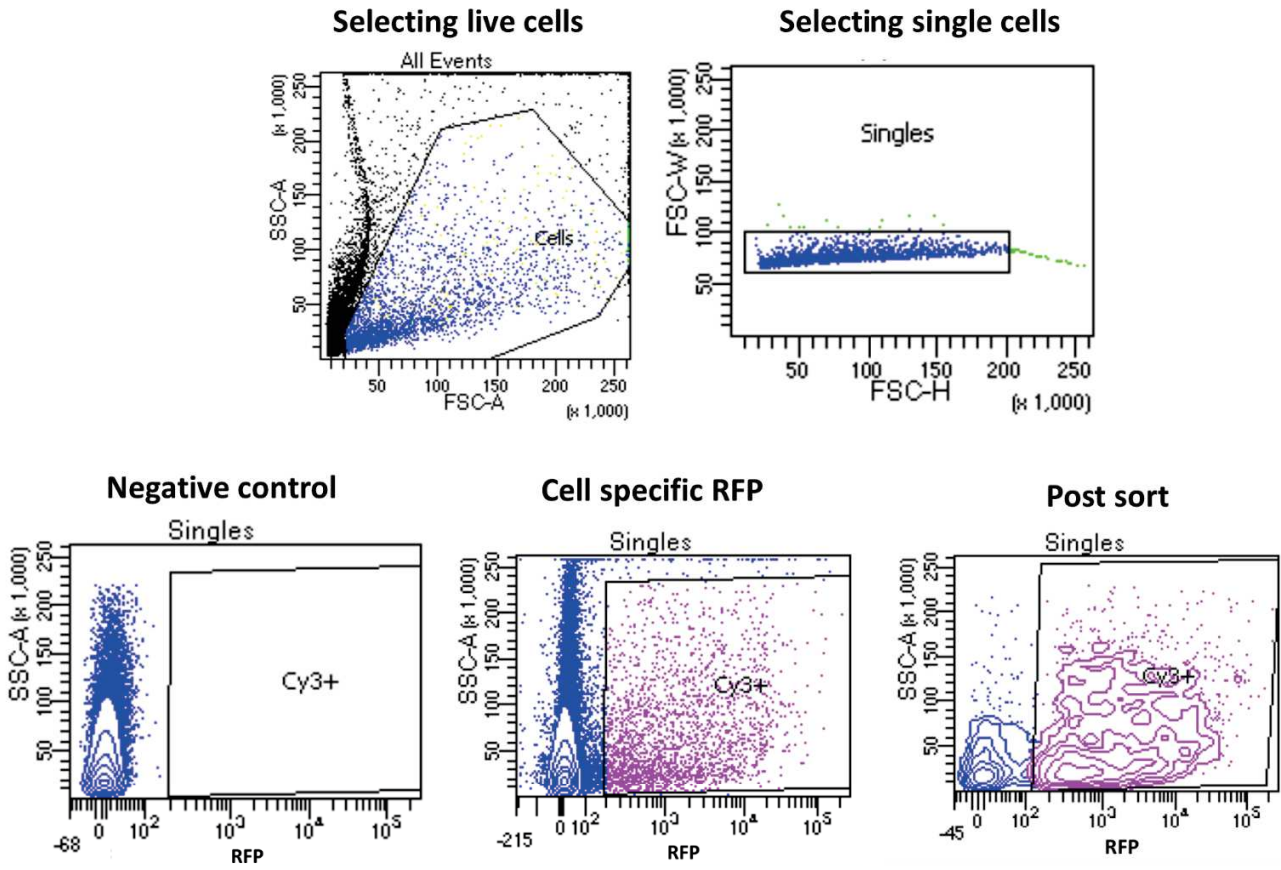
In sum, cell identity is defined by the origin and by the function(s). The relative impact of these parameters depends on the cell type, with neurons remaining stable over development, while hemocytes become more specialized immune cells as the animal hatches, and glia show an intermediate neural and immune phenotype. Indeed, glial cells are at the interface between the nervous and the immune system: the neural origin shared with neurons is already apparent in the embryo, while immune function shared with hemocytes becomes apparent in the larva, suggesting that hemocytes and glia tend to diversify after embryogenesis. Finally, since transcription factors control the coordinated expression of several downstream, effector genes that implement cell-specific functions, I also analyzed the transcription factors upregulated in the different cell types and at the two stages. This analysis reveals that the identity as well as the number of cell-specific transcription factors differ between embryos and larvae. There are many more cell-specific factors in embryonic neurons than in embryonic hemocytes or glia. In addition, the number of the transcription factors specific to hemocytes or glia (but not that of neurons) increases significantly in the larva, calling for a diversification of glial and hemocyte functions and/or subtypes during development.

The most parsimonious interpretation of these data is that the functions of the CNS neurons remain relatively stable throughout development, whereas glia, and even more so hemocytes, are more plastic and change as the animal gets exposed to the outside world.

The unbiased identification of common and specific transcriptional pathways has allowed the better understanding of the mechanisms underlying the acquisition of cell identity and to define

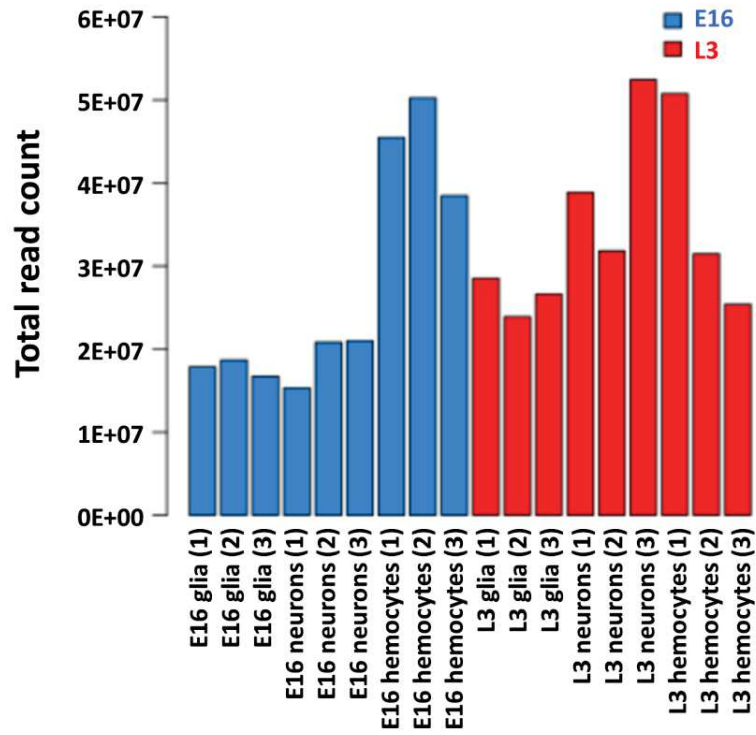
the relative impact of intrinsic and extrinsic cues in the process. The RNAseq data also highlight an intermediate neuro-immune phenotype of the fly glial cells. Thus, simple organisms rely on relatively few cell types with multiple potentials whereas the evolution of more complex organisms entails a more refined division of labor accompanied by the appearance of a dedicated immune cell type, the microglia. Given the importance of the immune resident cells of the CNS in development and homeostasis, fly glia represent an interesting tool to unravel the role of myeloid-like cells in diseases as severe and diverse as brain tumor, neurodegeneration and autoimmune diseases.

V- Supplementary data:



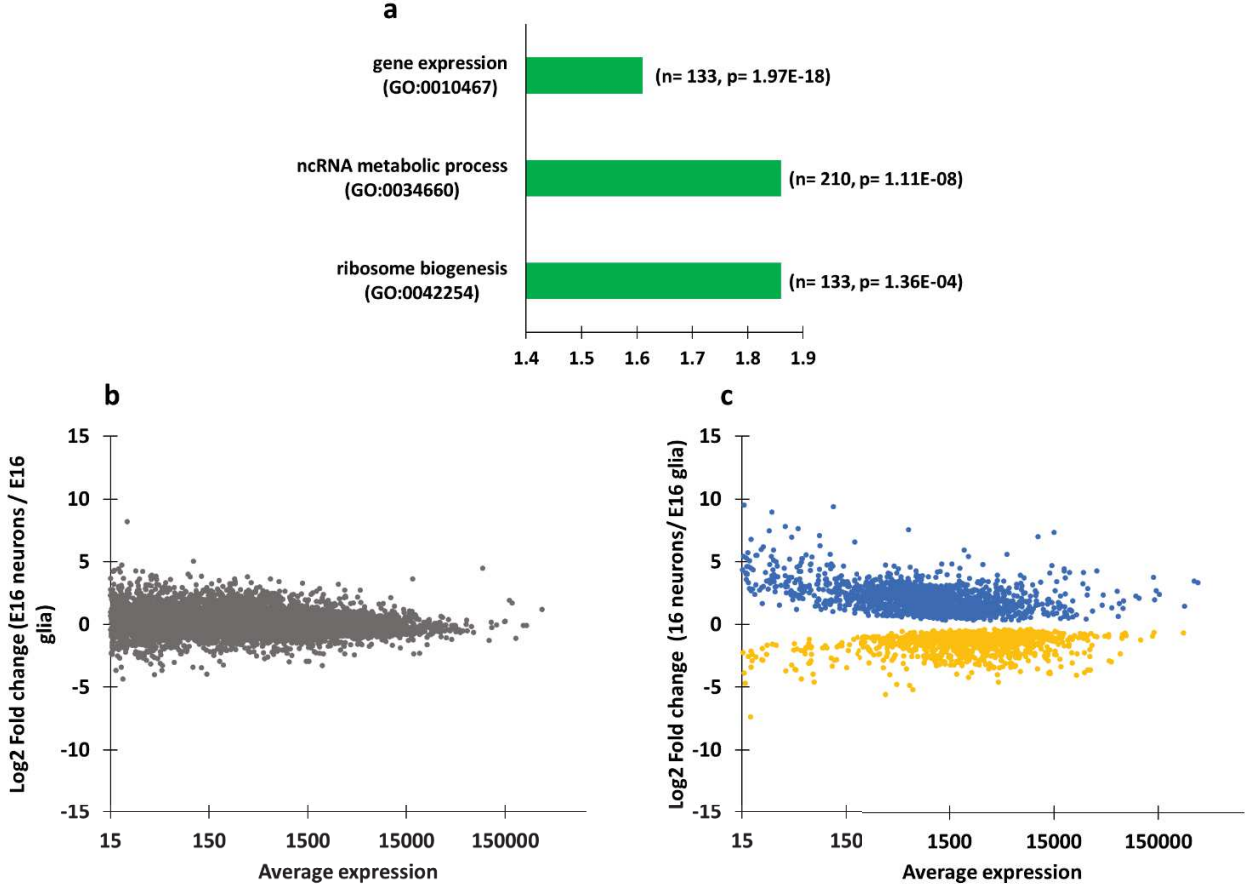
Sup. Figure 1: An example of the FACS sorting method used for all samples.

Cell are first selected based on the forward scatter (FSC-A) and the side scatter (SSC-A) to eliminate dead cells and debris. Doublets are removed and only single cells are taken into account. The gate for sorting Cy3/RFP positive cells is based on the negative control. Only Cy3/RFP positive are sorted. A post sort step is done at the end to test the purity of the sorted population.



Sup. Figure 2: Number of mapped reads per sample.

The y-axis represents the total read count. The x axis-represents the different sample: E16 glia, neurons and hemocytes (blue) and L3 glia, neurons and hemocytes (red).



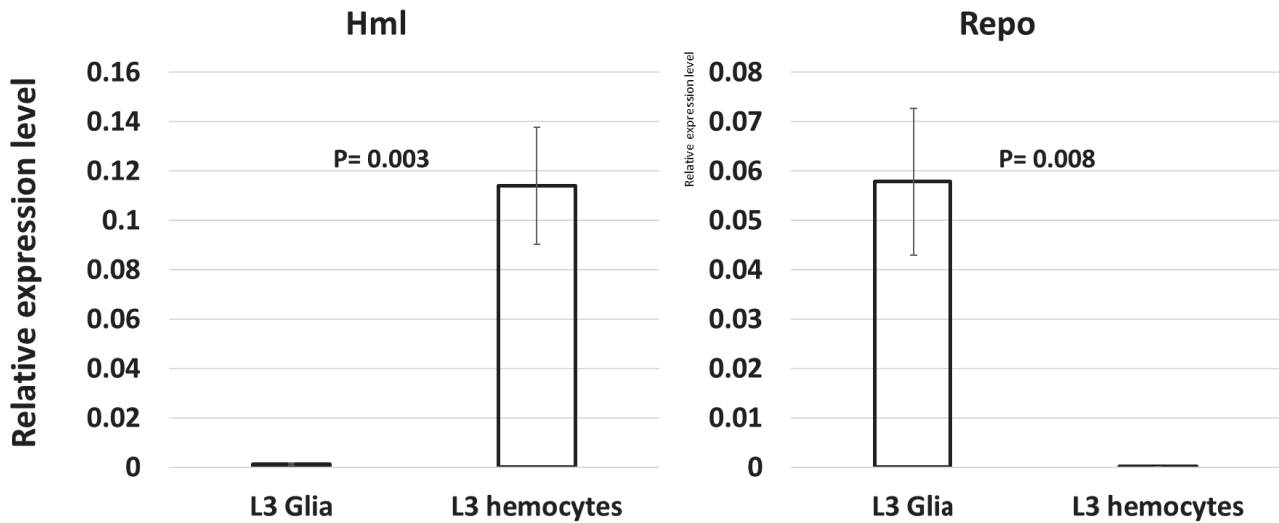
Sup. Figure 3: Comparison between commonly expressed genes and upregulated genes.

(a) Gene Ontology (GO) term enrichment analysis on genes commonly expressed between E16 glia and neurons. The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets. (b,c) Dot plots comparing the expression levels of non-differentially genes versus differentially expressed genes. y-axis is the log2 fold change E16 neurons/ E16 glia. Non-differentially expressed genes are shown in gray (b). The x-axis is the average gene expression levels (n = 3), Genes significantly enriched in E16 neurons are shown in blue, genes significantly enriched in E16 glia are shown in yellow (c).

	Forward	Reverse
Act5C	TGCTGCACTCCAAACTTCCA	GCAGCAACTTCTTCGTCACA
RP49	GACGCTTCAAGGGACAGTATCTG	AAACGCGGTTCTGCATGAG
Repo	CTCCGCCAAGTAGTTCCTCC	AGGCAGTAAAGGTGGTTCTCG
Hml	CGAGGCAAATCACGATGCTG	ACGGGCACTTGACGTTGTAT
Eater	ACGATCCATCTAACCGATGTGT	CGCAGTTATCCTTGACGTT
NimC1	TTCGCCATTTTACGGCATGG	GTCCTGTAGGCAGTCTCATCTT
NimC4	CTCGGGCTGAACGAAGCTAT	CCAAGGGATGAACCTGACCC
Drpr	GTGGCAGGGTGGGTAGC	TGATTCATGCCGTAATGTGTGC
Crq	CGATCATCGAAGCGGGAAGT	TGCAGCACATGGTGGATACG

Sup. Table 1: Primer used for RT-qPCR

Act5C: Actin 5C and RP49: Ribosomal Protein 49 were used for normalization. Repo: Reverse polarity, Hml: Hemolectin, NimC1: Nimrod C1, NimC4: Nimrod C4, Drpr: Draper, Crq: croquemort



Sup. Figure 4: RT-qPCR on *repo* and *hml*

To test the purity of the sorted glia and hemocytes used for the RT-qPCR experiment, I assessed the expression of *repo* and *hml* in glia and hemocytes respectively. N=3. Error bars represent the standard deviation, the p-value was calculated using Student t-test.

CHAPTER II

Characterization of cell-specific epigenetic marks in the *Drosophila* neurons and glia

In chapter I established the cell-specific expression profile of neurons and glia and how it shifts through development. In order to achieve a deeper understanding of cellular identity I decided to go beyond gene expression and investigate epigenetic modifications. In collaboration with Thomas Boutet, a PhD student in our lab, who contributed to the optimization and the performance of the CUT&RUN technique, I studied three well known histone modifications: H3K4me3, H3K27me3 and H3K9ac in neurons and in glia at embryonic stage 16 and at wandering third instar larva. This allowed us to first identify the distribution of these marks in the two cell types. In addition, these results show that while H3K4me3 alone is indicative of active transcription, the combination of H3K4me3 and H3K9ac leads to higher levels of expression. Moreover, H3K4me3 is found on all expressed genes including cell-specific and general genes. Finally, H3K9ac mostly marks genes that are involved in the cell-specific functions, hence defining cell identity.

I- Introduction:

Gene expression profiling (Chapter II) had allowed me to identify the cell-specific molecular features of neurons and glia as well as the developmental transcriptional changes. Since gene expression is tightly associated to a combination of epigenetic processes, I decided to complement my analysis on cell identity by studying the impact on gene expression of these modifications, in particular post-translational histone modifications. I therefore investigated the distribution of three well established marks: H3K9ac, H3K4me3 and H3K27me3.

H3K4me3 is the hallmark of active transcription. It is a broad mark found at the transcription start site (TSS) of active genes. Even though the exact mechanisms by which H3K4me3 activates transcription is still unknown, H3K4me3 was shown to promote H3K9ac on the neighboring residue within the same histone (Foulds et al., 2013). Correspondingly, H3K4me3 and H3K9ac have been reported to coexist on many active promoters in different organisms (Brusslan et al., 2015; Heintzman et al., 2007; Wang et al., 2008; Yin et al., 2011). H3K27me3, on the other hand, is a mark associated to Polycomb (Pc) and is found spread widely on repressed genes (Cai et al., 2021; Kundu et al., 2017; Nakagawa and Kitabayashi, 2018). Therefore, together these three marks represent the ideal candidates to study the correlation between chromatin states and expression profiles. I also used these marks to investigate the cell-specific epigenetic signature and its evolution through development in neurons and glia. The choice of neurons and glia is based on the fact that they originate from the same precursor then differentiate into functionally and morphologically different cell types. Thus, studying these two cell types constitute a good model to help reveal the cell-specific epigenetic modifications related to cell identity.

II- Material and methods:

1. Fly strains and genetics

All flies were raised on standard media at 25°C. For the Chromatin Immuno-Precipitation and sequencing (ChIP-seq) *Oregon R (OR)* flies were used. For the CUT&RUN the transgenic line *repo-nRFP* 43.1 on the III chromosome was used for the isolation of glia. The *repo-nRFP* line recapitulates the full *repo* expression pattern (Laneve et al., 2013). The *elav-nRFP* 28.2 insertion on the III chromosome was used to purify embryonic neurons. As for larval neurons *nSyb-Gal4* was crossed with *UAS-RFP* flies (RRID:BDSC_8547) to obtain *nSyb-Gal4/+; UAS-RFP/+*. The *Oregon-R* strain was used as the control for the cell sorting. For immunolabeling, the driver *Repo-Gal4 (II); UAS-mCD8GFP* was used for the experiments on embryos and *Repo-Gal4 (II)* alone was used for the experiment on wandering third instar larvae (L3). *UAS-Gcn5-RNAi* (Bloomington B#33981) was used to knock-down *Gcn5*.

2. FACS sorting and CUT&RUN

Embryonic cells were isolated as follows. Staged egg laying was carried out to produce embryonic stage 16 (E16) embryos. *repo-nRFP*, *elav-nRFP* and *OR* strains were amplified, several hundreds of flies of the different genotypes were then transferred into cages and raised on a yeast apple juice agar at 25°C. After a pre-lay period of 30 minutes, the agar plates and yeast were replaced with fresh plates and flies were left to lay for 3 hours at 25°C. Agar plates were then removed and kept at 25°C, embryos were collected 11 hours and 40 minutes AEL, when they reached stage 16. Embryos were then isolated from the medium and washed on a 100 µm mesh. The collected embryos were transferred into a cold solution of phosphate-buffered saline (PBS) in

a Dounce homogenizer on ice. Embryos were disassociated using the large clearance pestle then filtered through a 70 μm filter and live cells were FACS sorted directly after.

L3 neurons and glia were isolated from *nSyb/+; UAS-RFP/+* and *repo-nRFP* larvae respectively. Staged lays of 3 h were carried out at 25°C and wandering larvae were collected 108–117 h AEL. Larval brains were dissected in cold PBS on ice and transferred to tubes containing 0.5 μg of collagenase IV in 220 μL of PBS 1x. Brains were incubated at 37°C on a Thermomixer with shaking at 500 rpm for 20 minutes. Then the brains were dissociated by pipetting up and down with 10-gauge needles and syringes and filtered through a 70 μm filter. The cells were sorted using FACS Aria II at 4°C in three independent biological replicates for each genotype. Live cells were first selected based on the forward scatter and side scatter and only single cells were sorted according to the RFP signal. *OR* cells were used as a negative control to set the gate and sort the RFP positive cells. Around 10,000 cells were sorted for each replicate. The purity of the sorted populations was assessed by carrying out a post-sort step. The FACS sorter was set up to produce cell pools displaying at least 80% of purity on the post-sort analysis. Around 10,000 cells were collected in 1x PBS for each replicate. The purity of the sorted populations was assessed by carrying out a post-sort step. The FACS sorter was set up to produce cell pools displaying at least 80% of purity on the post-sort analysis.

The CUT&RUN protocol was adapted from the published method (Skene et al., 2018).

All buffers were supplemented with cOmplete™, EDTA-free Protease Inhibitor Cocktail (11873580001). All incubation and washing steps, if not specified, were performed on a balancelle at 4°C.

Nuclei from sorted cells were extracted on ice for 10 minutes in extraction buffer (20mM HEPES-KOH pH 7,5, 10 mM KCl, 0,5mM spermidine tetrahydrochloride, 0,1% triton X-100, 20% glycerol) and fixed 2 minutes in 1% PFA. Formaldehyde crosslink reaction is quenched by addition of 1.25 M Glycine. Concanavalin A-coated magnetic beads (BioMag® Plus Concanavalin A, cat#86057-3) (5 μ L for n samples) were washed 3 times in binding buffer (20 mM HEPES-KOH, 10 mM KCl, 1 mM CaCl₂, 1 mM MnCl₂) and incubated with the isolated nuclei for 10 minutes. After removing the supernatant, bead-bound nuclei were incubated 10 minutes in blocking buffer (20mM HEPES-KOH pH 7,5, 150mM NaCl, 0,5mM spermidine tetrahydrochloride, 0,1% BSA (mpBio™, CAT NO. 160069), 2mM EDTA), washed 3 times in wash buffer (20mM HEPES-KOH pH 7,5, 150mM NaCl, 0,5mM spermidine tetrahydrochloride, 0,1% BSA) and resuspend in 50 μ L wash buffer. Bead-bound-nuclei were split into n tubes and incubated overnight with primary antibody 1:50 (rabbit anti-H3K27me3 Cell Signaling C36B11, rabbit anti-H3K4me3 abcam8580, rabbit anti-H3K9ac active motif 39585). Nuclei were washed 3 times in wash buffer, resuspend in 100 μ L and incubated one hour with the secondary antibody 1:100 (guinea-pig anti-rabbit, ABIN101961). Bead-bound nuclei were washed 3 times and incubated one hour in 200 μ L wash buffer containing 0,7 ng/ μ L pA-MNase. Beads were washed 3 times, resuspend in 150 μ L wash buffer and incubated 10 minutes in ice-water. Cleavage by the MNase is induced by adding 2 mM CaCl₂. Digestion was stopped after 30 minutes by addition of 1 volume of stop-buffer (340mM NaCl, 20mM EDTA, 4mM EGTA, 100 μ g/mL RNase A, 50 μ g/mL glycogen, 100pg/mL yeast spike-in) followed by 20 minutes incubation at 37°C. SDS (0,1%) and Proteinase K (150 μ g/mL) were added to the samples to digest proteins and to reverse crosslinking and incubated for one hour at 50°C. DNA was then extracted using Phenol/Chloroform extraction and MaXtract™ High Density tubes (QIAGEN, 129046). Library

preparation was performed following the protocol published by Nan Liu on protocol.io ([dx.doi.org/10.17504/protocols.io.wvgfe3w](https://doi.org/10.17504/protocols.io.wvgfe3w)) and using the library preparation kit NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® #E7645L and the Multiplex Oligos for Illumina® #E6442S. Adapter concentration used for the preparation was downsized to 0,15 µM (**Figure 20**).

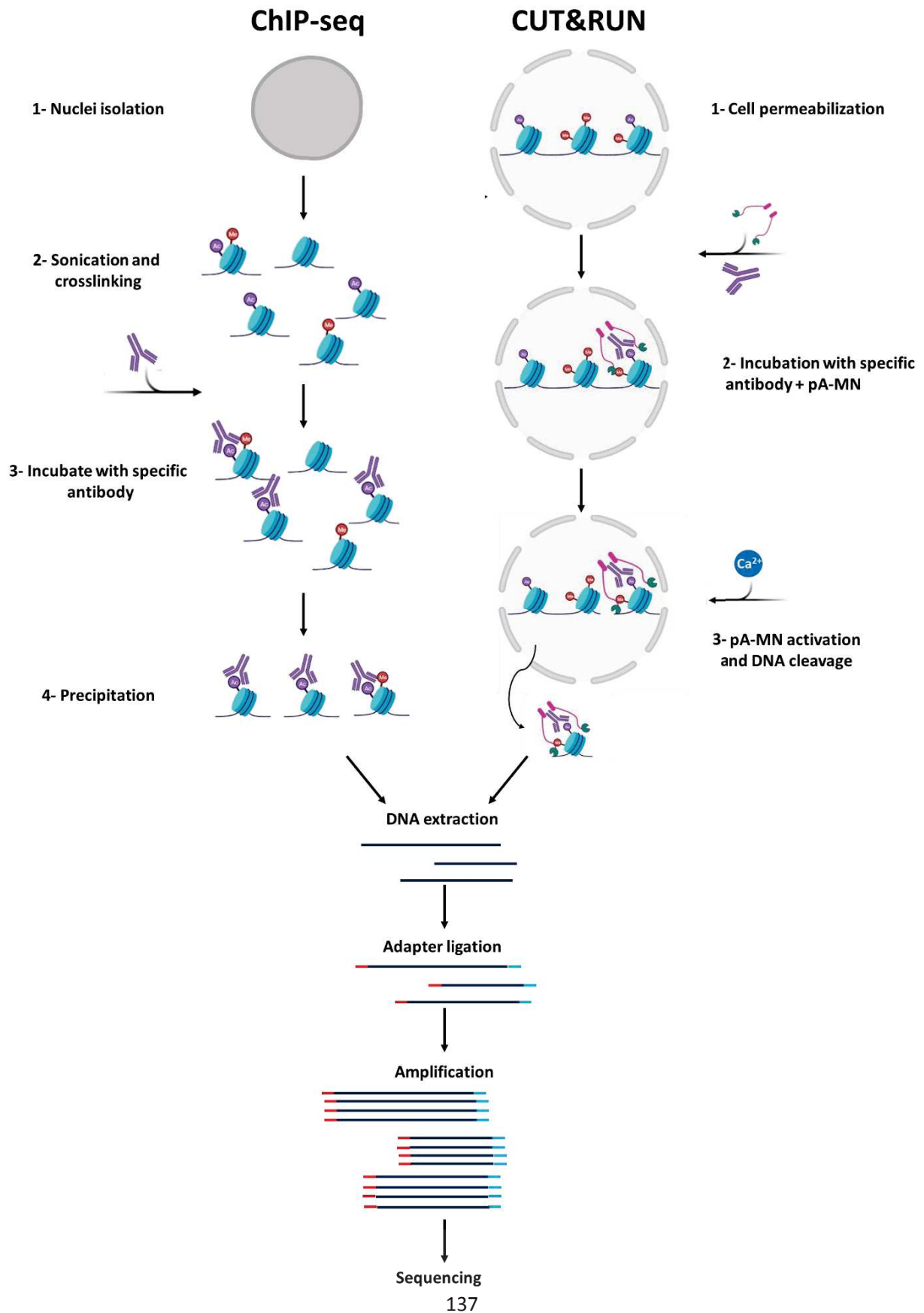


Figure 20: Comparing ChIP-seq and CUT&RUN techniques.

ChIP-seq: Nuclei are isolated, DNA crosslinked then fragmented using sonication. DNA is incubated with specific antibodies then precipitated. Crosslinking is reversed and DNA is extracted followed by library preparation and sequencing.

CUT&RUN: Cells are permeabilized to allow the introduction of the specific antibody and the pA-MNase. pA-MNase binds the antibody. Ca^{2+} is added to activate the pA-MNase that cuts the DNA around the binding site of the antibody. Cut DNA sequences are eluted from the cell. DNA is extracted followed by library preparation and sequencing.

3. FACS sorting and ChIP-seq

The protocol used for the Chromatin immunoprecipitation and sequencing (ChIP-seq) follows two published methods: (Bonn et al., 2012; Sandmann et al., 2006). Collected *OR* embryos dechorionated in 25% bleach for 5 minutes then washed thoroughly. Embryos were incubated in crosslinking solution (1 mM EDTA, 0.5 mM EGTA, 100 mM NaCl, 1.8% formaldehyde (v/v), 50 mM HEPES, pH 8.0, 1 vol. of Heptane) for 15 minutes on a shaker. Crosslinking was stopped using 125 mM glycine, 0.1% Triton X-100 (v/v) in PBS 1x. Embryos were washed 3 x 10 minutes in PBS containing 0.1% Triton-X (PTX) then dried by removing all supernatant and snap frozen in liquid nitrogen.

Embryos were transferred into a Dounce homogenizer containing 10 mL of PTX supplemented with protease inhibitor (PI) and disassociated with the loose pestle 30 times. The lysate was centrifuged at 400 g for a minute at 4°C to pellet debris. The supernatant was centrifuged again at 2000 g for 10 min to pellet the cells. Cells were resuspended in 20 mL HB buffer (15 mM Tris-HCl (pH 7.4), 0.34 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA and 0.2 mM EGTA) supplemented with (PI), filtered through a 70 µm filter and nuclei were isolated using the tight Pestle. The lysate was centrifuged 3300 g for 5 minutes at 4°C to pellet the nuclei. Nuclei were washed again in HB then resuspended in 3 mL of blocking buffer (5% (wt/vol) BSA in PTX 1X)

supplemented with PI. Nuclei were syringed 10x through an 18-G needle then 10x through a 22-G needle to individualize them and avoid clumping. Nuclei were separated into 2 lots. One lot was labeled with mouse anti-Elav (DSHB) to label neuronal nuclei and the other was labeled with mouse anti-Repo (DSHB) to label glial nuclei and incubated for 1 hour at room temperature (RT). Nuclei were washed then incubated with secondary antibody Cy3 anti-mouse (Jackson) for 30 minutes at RT then washed and filtered again. Nuclei were sorted according to the RFP signal using FACS Aria II (BD Biosciences) at 4°C in two independent biological replicates. Non-labeled nuclei were used as a negative control to set the gate and sort the RFP positive cells. Around 1,000,000 nuclei per replicate were sorted.

The ChIP-seq was done in collaboration with the labs of Dr. Giacomo Cavalli (IGH, Montpellier) and Dr. Nicola Iovino (Max Planck Institute, Freiburg).

Sorted nuclei were resuspended in 300 µl of RIPA buffer (140 mM NaCl, 1 mM EDTA, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) SDS, 0.1% (wt/vol) sodium deoxycholate and 10 mM Tris-HCl (pH 8.0)) + PI and incubated on ice for 10 minutes. Samples were sonicated for 15 cycles 30 seconds ON/OFF then centrifuged at 12000 g for 10 minutes. Protein A–Sepharose CL4B (Seph A) beads were added to the chromatin and the mix was incubated on a rotating wheel at 4°C for 1 hour then beads were pelleted by centrifugation at 1600 g at 4°C for 2 minutes. 30 µL per sample were removed to be used as input. For the remaining, for each sample 3 µg of Rabbit anti-H3K9ac (Active motif 39585) were added and samples were incubated overnight on a rotating wheel. Immunocomplexes were purified and beads were washed for 10 minutes in the following buffers: 1x with RIPA, 4x with RIPA500 (500 mM NaCl, 1 mM EDTA, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) SDS, 0.1% (wt/vol) sodium deoxycholate and 10 mM Tris-HCl (pH 8.0)), 1x with LiCl

(Combine 250 mM LiCl, 1 mM EDTA, 0.5% (vol/vol) IGEPAL CA-630, 0.5% (wt/vol) sodium deoxycholate and 10 mM Tris-HCl (pH 8.0)) and 2x with TE (1 mM EDTA and 10 mM Tris-HCl (pH 8.)). DNA was eluted twice using elution buffer (1% SDS, 0.1M NaHCO₃). 20 µl of NaCl 5 M were added to samples and input and incubates overnight at 65°C. 10 µg of RNase A then 20 µg of Proteinase were added. DNA was extracted using Phenol/Chloroform followed by library preparation and sequencing.

4. CUT&RUN pipeline

Conda environments (Conda, v4.8.2 (Anaconda Software Distribution, Anaconda Inc)). were used to install and run bioinformatic tools used for Cleavage Under Targets & Release Using Nuclease (CUT&RUN) dataset analysis. Raw sequences were trimmed using trim-galore v0.6.6 -paired --length 25 --trim1 --retain_unpaired. Quality control before and after trimming was performed using fastQC v0.11.9 and multiqc v1.9. Trimmed paired-end sequences were aligned to the *Drosophila melanogaster* August 2014 Dm6 (BDGF release 6 + ISO1 MT/Dm6) reference genome using Bowtie2 v2.4.2 --very-sensitive-local --dovetail --no-unal -I 25 -X 700 --fr -1. Resulting alignment files were converted to bam, sorted and indexed using samtools v1.3.1. Multiple mapping reads (quality score < 10) were sorted out. Peak calling was performed on merged replicates using MACS2 callpeak v 2.2.7.1 --format BAMPE --gsize 120000000 --keep-dup=all --qvalue x. Pvalue used for peak calling was determine for each sample separately based on MACS2 cutoff analysis. The pvalues were chosen in the proximity of the inflexion points observed when plotting the pscore against the average read length. Peak annotation was performed using HOMER annotatePeaks. Extensive analysis and comparison of the resulting files was performed using R studio as well as multiple packages: xlsx v0.6.5, readxl v1.3.1, dplyr v1.0.6,

ggplot2 v2016, gdata v2.18.0, tidyr v1.1.3, ggpubr v0.4.0, htmltools v0.5.1.1, corrplot v0.90, ggrepel v0.9.1, magrittr v2.0.1, DESeq2 v1.32.0. R Core Team (2021) (Anders and Huber, 2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

5. ChIP-seq analysis pipeline

ChIP-seq pipeline: The data analysis was done using the GalaxEast platform, the Galaxy instance of east of France (<http://www.galaxeast.fr/>). First, summary statistics were computed on the raw FastQ Illumina files of the dataset using the quality control tool for high throughput sequence data FastQC (Babraham Bioinformatics). The raw files were then converted to FastQ Sanger using FastQ Groomer for downstream analysis. Data were then mapped to the *Drosophila melanogaster* August 2014 Dm6 (BDGF release 6 + ISO1 MT/Dm6) reference genome using Bowtie2 (Langmead and Salzberg, 2012). Peak calling was then carried out using MACS2 (Zhang et al., 2008), the input was used as control. Peaks were then annotated using Homer (Heinz et al., 2010).

6. Data analysis and visualization

Peak profiles were generated using Deeptools2 (Ramírez et al., 2016). Gene ontology studies were done using the Panther classification system version 16 (<http://pantherdb.org/>) (Mi et al., 2010; Thomas et al., 2003) for the identification of biological processes, Bonferroni correction was used for all analyses. and data representation was done using Microsoft Excel.

7. Immunolabeling and BBB permeability assay

Embryos were dechorionated in bleach for 5 minutes and rinsed in water. They were then fixed in 50% heptane/50% PFA for 25 minutes on a shaker at RT and devitellinized in methanol/heptane for 1 minute. embryos were then rinsed in methanol followed by PTX (0.3% Triton X-100 in PBS 1x) and incubated in blocking reagent (Roche) for 1 hour at RT. The embryos were incubated with primary antibodies (mouse anti-Repo (DSHB), rabbit anti-Neurexin IV (Gift from M. Bhat. Banerjee et al., 2010) in blocking reagent overnight at 4 °C, washed in PTX 3 × 10 minutes at RT, incubated with secondary antibodies (FITC anti-mouse and Cy3 anti-rabbit (Jackson)) for 1 h at RT and washed 3 × 10 minutes at PTX. The embryos were mounted in Vectashield (Vector #H-1000) and analyzed by confocal microscopy (Leica spinning disk). For L3 larvae, CNSs were dissected in cold PBS 1x and transferred into wells containing 4% PFA and fixed for 2 hours at RT. They were then washed in PTX for 1 hour and incubated in blocking reagent for 1 hour at RT. CNSs were then incubated with primary antibodies (mouse anti-Repo, rabbit anti-PH3 (Millipore)) ON at 4 °C, then washed in PTX 3 x 10 minutes and incubated in secondary antibodies (FITC anti-mouse and Cy3 anti-rabbit (Jackson)) for 1 hour at RT. After washing in PTX CNSs were mounted in Vectashield and analyzed using confocal microscopy.

The Blood-Brain barrier (BBB) permeability assay was done as follow: Dissected L3 larvae with exposed CNS were for 30 minutes in 50 mM of 4 kDa Dextran (Invitrogen). Then were fixed in 4% PFA for 2 hours at RT and mounted.

III- Result:

1. Correlation between H3K4me3, H3K9ac and H3K27me3

H3K4me3 and H3K9ac are usually linked to transcriptional activation. H3K27me3, on the other hand, marks silenced genes. Therefore, I hypothesize that these marks are directly involved in the acquisition of the specific transcriptional signature displayed by glia and neurons. I used a new technology to investigate the distribution of the three marks in FACS sorted E16 and L3 glia and neurons. Cleavage Under Targets & Release Using Nuclease (CUT&RUN) (Skene and Henikoff, 2017) is an antibody-targeted nuclease based strategy that allows mapping of transcription factors and histone marks using a limited number of cells (10,000 cells per replicate). This provides a great advantage compared to Chromatin immune precipitation and sequencing (ChIP-seq) that requires around 1,000,000 fixed nuclei per replicate. CUT&RUN uses a protein-A-tagged-MNase (micrococcal nuclease) to direct cleavage of chromatin at specific sites, is characterized by a low background and requires less sequencing depth, in comparison to ChIP-seq, which uses sonication to fragment chromatin. CUT&RUN allowed multiplexed sequencing of large numbers of libraries and made it possible to determine the positions of the three marks in the genome of neurons and glia.

To analyze the data, I first compared the genome-wide distribution of H3K4me3, H3K9ac and H3K27me3 in all the different samples (E16 glia, E16 neurons, L3 glia and L3 neurons). I computed the average scores for every genomic region for every sample and generated a heatmap representing the Spearman correlation coefficients (r) (**Figure 21**). The correlation coefficient between the replicates is above 0.8, highlighting the replicability of the method and the robustness of the histone mark profiles across cell types.

Since H3K4me3 and H3K9ac are both indicative of active transcription, I expect them to correlate together to a certain extent. The heatmap shows a correlation coefficient of 0.65 on average between the different samples, thus indicating some overlap but an overall differential distribution between the two marks. As expected, no correlation was found between H3K4me3 and H3K27me3 or between H3K9ac and H3K27me3, representing their different roles in gene expression regulation.

Interestingly, H3K9ac does not cluster based on the developmental stage but on the cell type: all E16 and L3 neurons cluster together while E16 and L3 glia cluster together. The clear separation between the two cell types indicates a cell-specific signature. The same trend is seen with the repression mark H3K27me3. However, in the case of H3K4me3, E16 neurons (2) groups with E16 glia (2) and E16 neurons (1) shows higher correlation with E16 glia (2) ($r = 0.9$) than with L3 neurons ($r = 0.89$ and 0.88). In addition, the differences between the cell types are more striking in the case of H3K9ac and H3K27me3 while H3K4me3 seems to share more commonalities in the different cell types. Thus, all together this comparison indicates that H3K9ac and H3K27me3 are more cell type specific than H3K4me3.

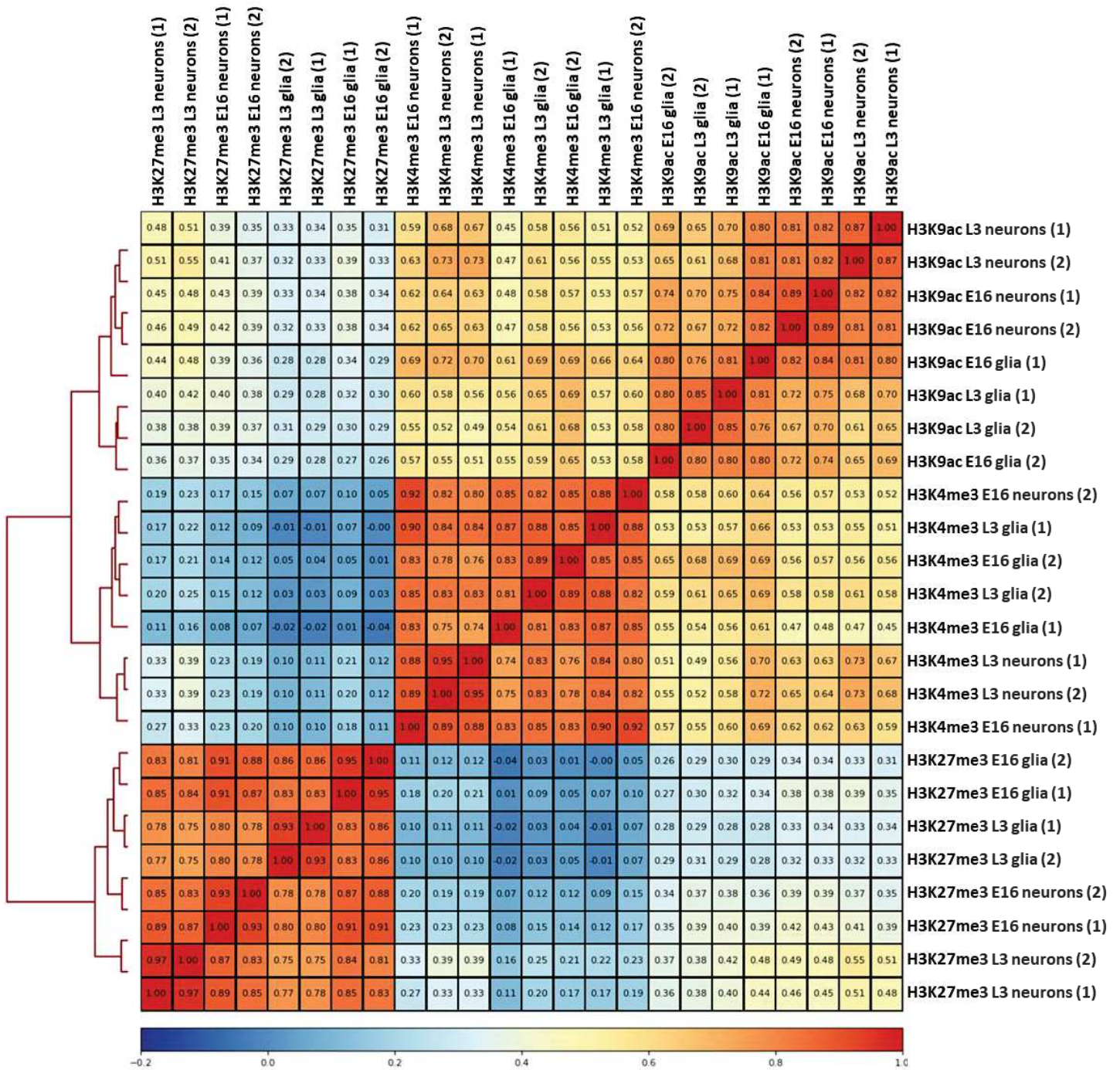


Figure 21: Heatmap showing the correlation between all samples.

Comparison of the genome-wide distribution of H3K4me3, H3K9ac and H3K27me3 in E16 glia, E16 neurons, L3 glia and L3 neurons. The average scores for every genomic region for every sample is computed and a heatmap representing the Spearman correlation coefficients (r) is generated. R values are indicated inside each square. Color code indicated the r value going from -0.2 (blue) to 1 (red).

2. H3K9ac, H3K4me3 and H3K27me3 peak at specific genes

To validate the specificity of the data in a tissue-specific manner, I looked at the distribution of the three marks in neurons and glia on genes that are known to be repressed or expressed in a specific cell type according to the literature.

abdominal A (abd-A) encodes a homeobox-containing TF that is a component of the Bithorax complex and one of the Hox genes that contributes to the developmental fate of embryonic segments (Jijakli and Ghysen, 1992). Wide peaks of H3K27me3 spanning the entire *abd-A* locus are observed in the glia and neuron data, while H3K4me3 and H3K9ac are very low at this locus (**Figure 22-a,b**). Furthermore, I hypothesized that the hemocyte fate determinant *serpent (srp)*, involved in the activation of hemocyte-specific genes (Lebestky et al., 2000), must also be repressed in neurons and glia. There is no record of the mode of epigenetic repression of this gene in neurons and glia. Interestingly, in both neurons and glia I see a wide distribution of H3K27me3 throughout the *srp* gene accompanied by very low H3K4me3 and H3K9ac (**Figure 22-c,d**). I also assessed in neurons the state of methylation of the glial fate determinant *glide/gcm*. Similarly than *srp*, *gcm* displays methylation of H3K27 in neurons (**Figure 22-e**). This highlights the role of H3K27me3 in the repression of TFs necessary for the induction of other fates.

I then assessed the distribution of the activation marks H3K4me3 and H3K9ac on known loci, such as the *tramtrack (tkk)* gene, a repressor of neuronal fate known to be expressed in glia (Giesen et al., 1997). Enrichments of H3K9ac and H3K4me3 as well as an absence of H3K27me3 are observed in the promoter of *tkk* in E16 and L3 glia (**Figure 22-f**). For neurons, *embryonic lethal abnormal vision (elav)* is a pan-neuronal RNA binding protein often used as a neuron-specific marker (Robinow and White, 1991). H3K4me3 and H3K9ac are strongly enriched at the *elav*

promoter region accompanied by a lack of H3K27me3, highlighting the active state of this gene. Interestingly, the presence of the active marks is only detected at E16 and not at L3, in concordance with the fact that *elav* is expressed almost 5 times less at L3 than at E16 (**Figure 22-g**).

Overall, these results highlight a strong correlation between active histone marks and tissue-specific genes as well as repressive marks around unrelated genes. This suggests that the apposition of histone marks may be directly involved in the processes of differentiation of the cell.

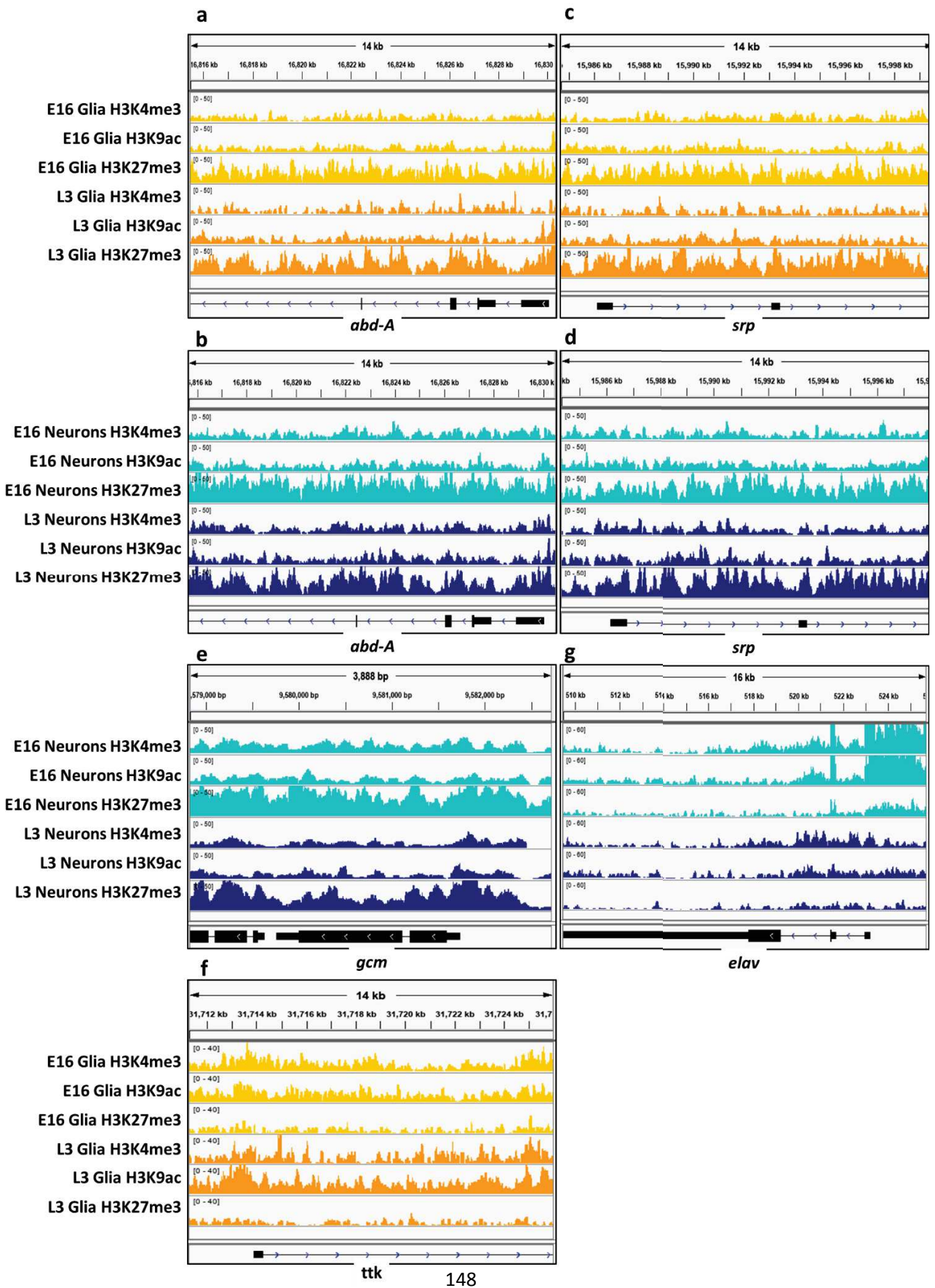


Figure 22: Visualization of peaks in Integrative Genomics Viewer (IGV)

Visualization of H3K4me3, H3K9ac and H3K27me3 peaks over the *abdominal A (abd-A)* and *serpent (srp)* genes in E16 and L3 glia (a-c), as well as in E16 and L3 neurons (b-d). In neurons, are also shown *glial cell missing (gcm)* (e) and *embryonic lethal abnormal vision (elav)* (f). Also, *tramtrack (ttk)* is shown in glia (f). E16 glia are shown in yellow, L3 glia in orange, E 16 neurons in light blue and L3 neurons in navy. The samples are indicated on the left. The same data range was set for all samples over the same gene.

3. H3K4me3 and H3K9ac peaks are found around the TSS

While H3K27me3 is found spanning the entire gene locus, H3K9ac and H3K4me3 are hallmark of nucleosomes spanning the TSS of active genes. Therefore, I focused the analysis on the distribution of H3K4me3 and H3K9ac relatively to the TSS. I generated the profile of peaks and heatmap for the two marks in E16 glia and neurons.

The peak profile represents the read count frequency (y-axis) over a genomic region spanning 3000 bp before and after the TSS with the TSS at the center (x-axis). The profiles of H3K4me3 and H3K9ac show a first small peak starting before the TSS followed by a sharper and bigger peak after the TSS (**Figure 23-a,b**). This is the case for the two marks and the two peaks represent the two nucleosomes around the TSS. The corresponding heatmap also emphasizes the enrichment of H3K4me3 and H3K9ac before and after the TSS with a lack of the marks right at the TSS. This highlights the expected profile for H3K4me3 and H3K9ac (Barth and Imhof, 2010).

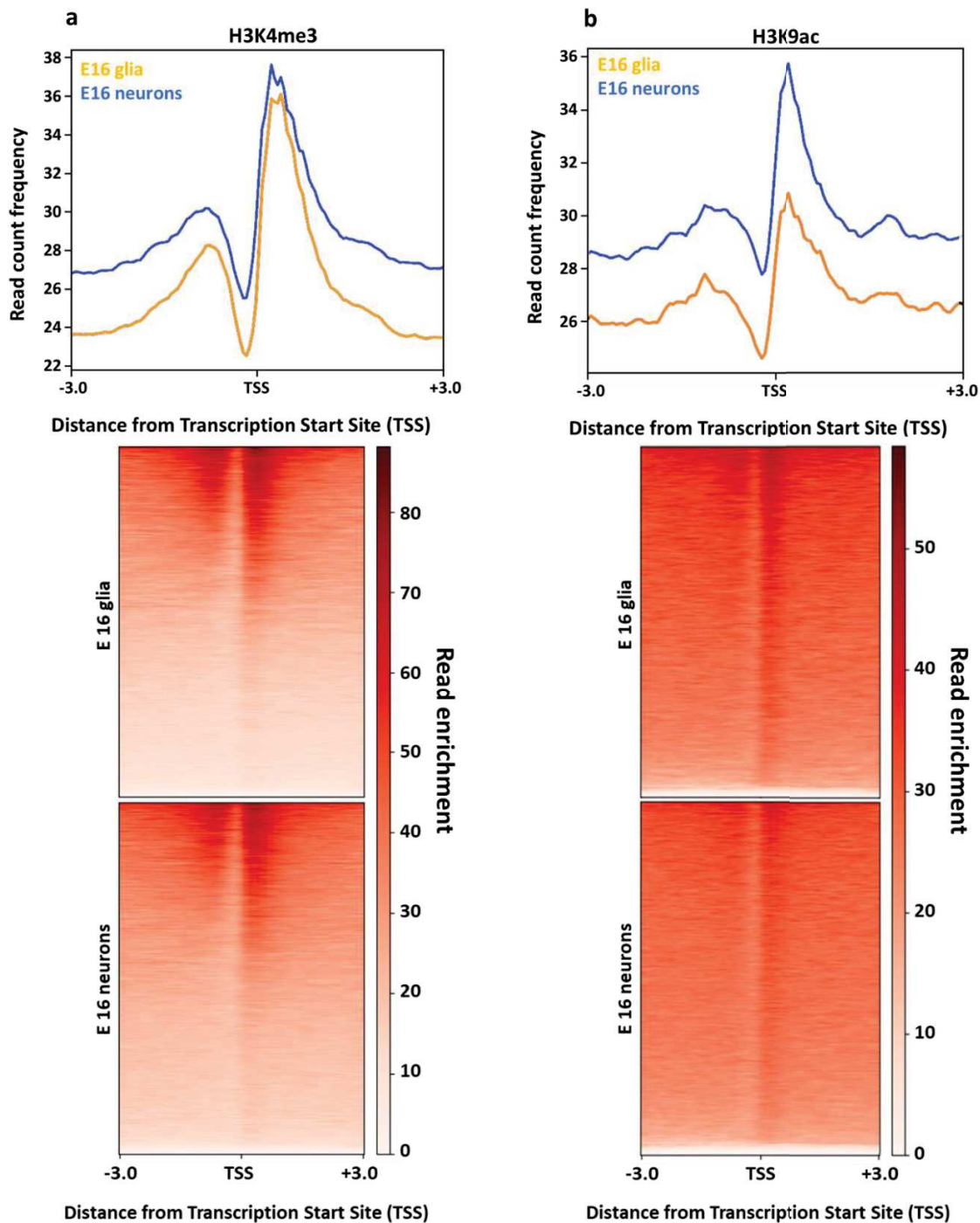


Figure 23: Peak profile and heatmap for H3K4me3 and H3K9ac.

Profile and heatmap of H3K4me3 (a) and H3K9ac (b) in E16 glia (yellow) and L3 (neurons). The x-axis presents the genomic region from 3 Kb before to 3 Kb after the transcription start site (TSS). The y-axis presents the read count frequency. Color code of heatmap goes from 0 (white) to 90 (red) for H3K4me3 and from 0 (white) to 60 (red) for H3K9ac.

4. Colocalization of H3K4me3 and H3K9ac identifies genes expressed at high levels.

I had previously established the expression profile of neurons and glia at E16 and L3 and since gene expression has been tightly linked to histone modifications, I compared the presence of histone marks on gene loci with their expression levels.

For H3K4me3 and H3K9ac peaks, I took into account genes presenting the marks at ± 1000 bp around the TSS, which encompass the two nucleosomes around the TSS. The genes were then clustered according to their histone mark profiles: Group 1 genes present no mark, Group 2 genes display H3K27-methylated, group 3 genes are only H3K9-acetylated but not H3K4-methylated, Group 4 genes have H3K4me3 but not H3K9ac and group 5 genes show colocalization of both marks. The same distribution of Group size was found in the different samples. 10% of all genes belonged to groups 2 and around 8% to group 3 compared to around 20% belonging to group 4. Group 5 included 12% of genes and the rest presented none of the marks (**Sup. Figure 5**).

I then compared the expression levels of the different groups of genes using the mRNA-seq data. As expected, the genes that have H3K27me3 are expressed at very low levels, or even not at all. To my surprise, genes that have no mark are expressed at very low levels, highlighting the importance of active marks on gene expression. Interestingly, I found that genes presenting both H3K4me3 and H3K9ac (group 5) are expressed at significantly higher levels than those presenting only one of the two marks or none (**Figure 24**). This analysis was done on E16 as well as L3 neurons and glia combined, however analyzing each sample separately leads to similar observations (**Sup. Figure 6**).

In addition, there is a clear difference between the expression levels of group 3 and group 4 genes. H3K4me3 alone, unlike H3K9ac, correlates with an overall active transcription, which goes along with it being the hallmark of active genes. However, the fact that genes presenting both are expressed at an even higher level suggests a cooperative interaction between the two marks and thus a specific role of H3K9ac in gene expression. Of note, not all genes that show H3K4me3 are highly expressed since around 10% of H3K4-methylated genes are not expressed at all.

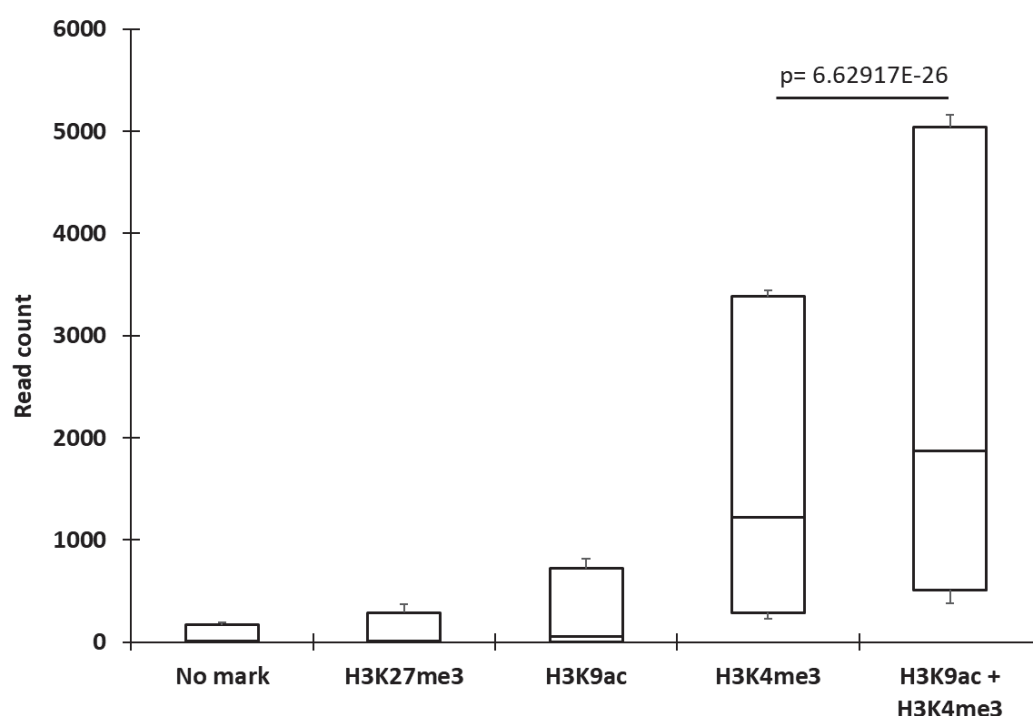


Figure 24: Correlation between histone marks and gene expression.

The x-axis shows genes that have none of the three marks, the genes that have H3K27me3, the genes that have H3K9ac but no H3K4me3 around the TSS, the genes that have H3K4me3 but not H3K9ac around the TSS and the genes showing both H3K4me3 and H3K9ac around the TSS. The y-axis represents the normalized reads count. For each group of genes, the graph was plotted based on the 1st quartile, median and 3rd quartile of gene expression. Error bars show the standard error. p shows the p-value that was computed using Student's t-test.

5. H3K9ac is linked to cell-specific functions

The comparison of the CUT&RUN data with the transcriptomes indicates that colocalization of H3K4me3 and H3K9ac is associated with stronger gene activation than that observed for genes that are labeled only by H3K4me3. This implies a role of H3K9ac other than broad transcriptional activation. To further investigate this, I looked at the function of genes in each group using GO term analysis on each cell type at E16.

I analyzed the genes carrying the repressive mark H3K27me3. GO term analysis on these genes in both neurons and glia identified biological processes related to the development of other tissues. For instance, I find in this group the GO terms posterior head segmentation, cell fate specification, leg disc development and digestive system development (**Figure 25-a,b**). In addition, in glia only I find the terms neuron differentiation (**Figure 25-a**). This indicates the role of H3K27me3 in the repression of other fates, in line with previous reports.

The genes of group 3 presenting H3K9ac alone do not show any enrichment of biological processes involved in glial or neuronal function. In fact, the GO term analysis yielded no results which means it was not possible to group these genes in specific biological processes. This combined with the low expression of these genes implies that H3K9ac alone is not indicative of the state of the cell and that its function is dependent on its combination with other marks such as H3K4me3.

The presence of H3K4me3 as well as the colocalization of H3K4me3 and H3K9ac correlate with high expression. I decided to take into account genes expressed at high levels in the case of H3K4me3 and the combination of H3K4me3 and H3K9ac to limit the number of genes assessed and thus increase the efficiency of the GO term analysis.

I assessed the genes highly expressed (# reads > 3000) and presenting H3K4me3 but not H3K9ac in E16 glia. GO term analysis identified many significantly enriched biological processes (p-value < 0.05, fold enrichment >3), such as terms related to cytoplasmic translation, protein localization to endoplasmic reticulum, regulation of cytoskeleton organization ATP and peptide metabolic process. Therefore, mainly broad terms, found in many cells were identified, these do not reflect glia identity (**Figure 25-c**). Next, I ran the same analysis on highly expressed genes presenting both H3K4me3 and H3K9ac. This revealed a striking difference since most biological processes identified are related to glial functions. For instance, we find the terms establishment of glial blood-brain barrier, glial cell differentiation, regulation of axonogenesis, regulation of synapse structure or activity and neuron projection guidance, all of which are glia-specific functions (**Figure 25-d**).

To find out whether this specificity is only present in glia, I investigated the function of genes presenting H3Kme3 or H3K4me3 and H3K9ac in E16 neurons. Two main neuronal functions (axo-dendritic transport and regulation of synapse organization) are identified in the group of genes presenting only H3K4me3 (**Figure 25-e**). However, the majority of neuron-specific functions were found in the genes that show both H3K4me3 and H3K9ac (**Figure 25-f**), implying a link between cell-specific genes and the presence of H3K9ac.

These results show that while H3K4me3 is needed for gene expression in general, by itself it does not highlight cell identity as it is found on gene expressed in most cell types such as those involved in translation and peptide synthesis. On the other hand, H3K9ac alone does not define active transcription, however its combination with H3K4me3 marks cell identity through the expression of cell-specific genes.

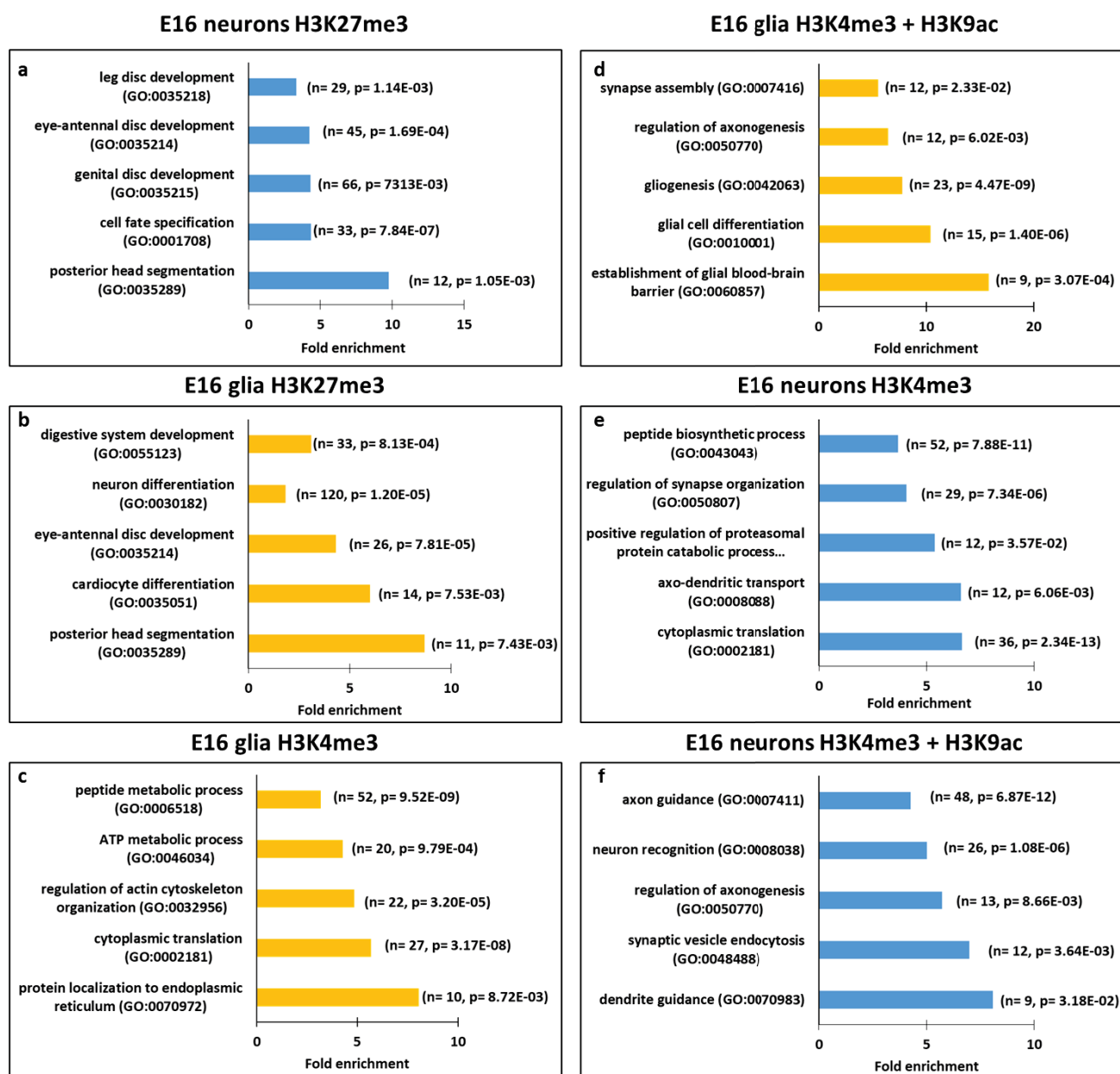


Figure 25: Gene ontology analysis at E16.

Gene Ontology (GO) term enrichment analysis on genes labeled by different marks in neurons or glia: H3K27me3 in E16 neurons (a) and glia (b); H3K4me3 (c) and H3K4me3 as well as H3K9ac (d) in E16 glia; H3K4me3 (e) and H3K4me3 and H3K9ac (f) in E16 neurons. Only genes presenting the marks at ± 1000 bases around the TSS are taken into account. The fold enrichments of GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

6. The evolution of the three marks through development

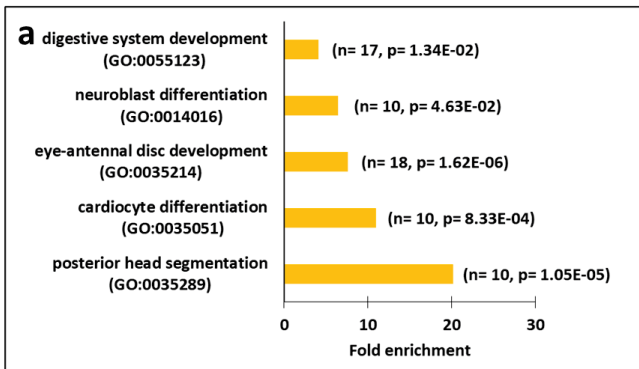
To identify the impact of the developmental stage on glia and neurons, I compared the profile of H3K4me3, H3K9ac and H3K27me3 at E16 and L3.

I first identified the genes that show H3K27me3 in both stages in glia. GO term analyses on these genes showed a continuous repression of the genes identified previously. In both stages, H3K27me is found on genes involved in head segmentation, cardiocyte differentiation, neuroblast differentiation and antennal development (**Figure 26-a**). However, the GO term analysis did not allow the identification of any stage-specific repression since genes H3K27-methylated only in one stage or the other did not group in any biological process. The same result was found when comparing E16 and L3 neurons (**Figure 26-b**). This indicates that cell fate related processes are repressed in early development and remain repressed during the maturation of the organism. This means that no new process is repressed nor de-repressed in neurons and glia at L3.

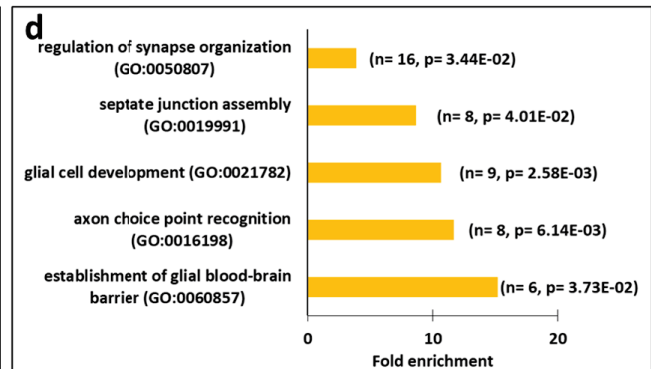
In addition, I assessed the distribution of the activation marks at the different stages. For that, I used genes presenting both H3K4me3 and H3K9ac and expressed at high levels (> 3000). The genes showing H3K9ac and H3K4me3 in the two stages display nervous system related GO term such as photoreceptor cell fate commitment, synapse assembly and learning or memory (**Figure 26-c**). These functions relate to both neuron and glia. Most of the glia-specific functions are found only in the genes presenting H3K9ac and H3K4me3 in E16 glia. Among those, are: establishment of glial blood-brain barrier, axon choice point recognition and glial cell development (**Figure 26-d**). Strikingly, no glia-specific biological processes were found in the genes presenting H3K4me3 and H3K9ac in L3 glia.

In neurons, we find GO terms such as axon extension, synaptic vesicle endocytosis and neuron recognition enriched in the genes showing H3K4me3 and H3K9ac in E16 as well as L3 neurons (**Figure 26-e**), while other neuronal functions such as neuron projection guidance, central nervous system development and axon guidance are only enriched in E16 (**Figure 26-f**). Interestingly, the GO terms axo-dendritic transport and regulation of synapse organization are only found in the genes presenting H3K4me3 and H3K9ac in L3 (**Figure 26-g**). Both terms were found in the E16 genes showing only H3K4me3 and no H3K9ac which means that these genes get H3K9-acetylated later in development. Even though these two GO terms are found in L3 neurons, these comprise only 20 genes, meaning that the majority of the neuronal genes show H3K4me3 and H3K9ac in E16 neurons or at E16 and L3 neurons similarly to the trend seen in glia.

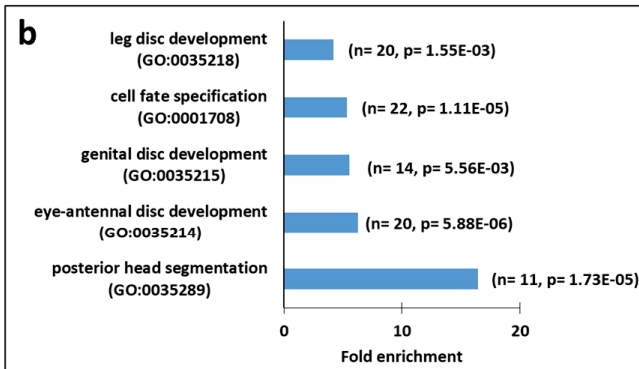
E16 + L3 glia H3K27me3



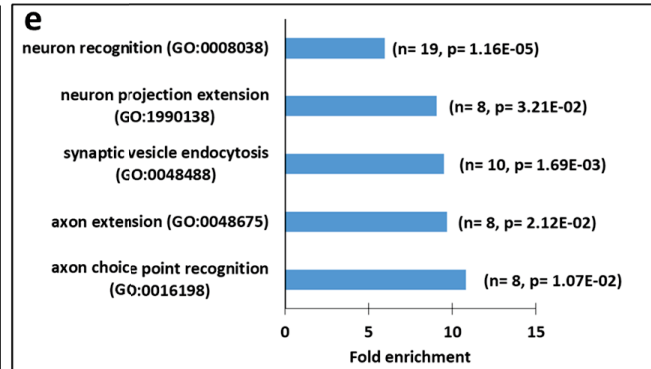
E16 glia H3K4me3 + H3K9ac



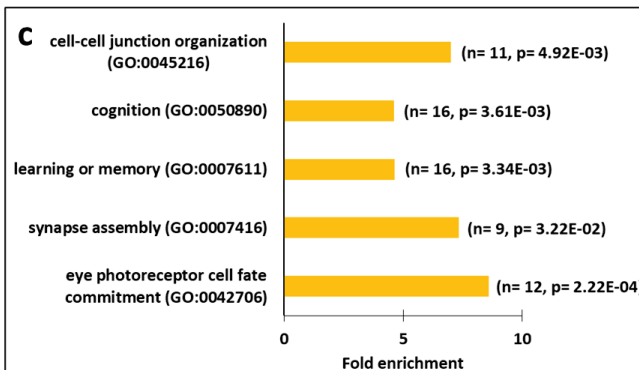
E16 + L3 neurons H3K27me3



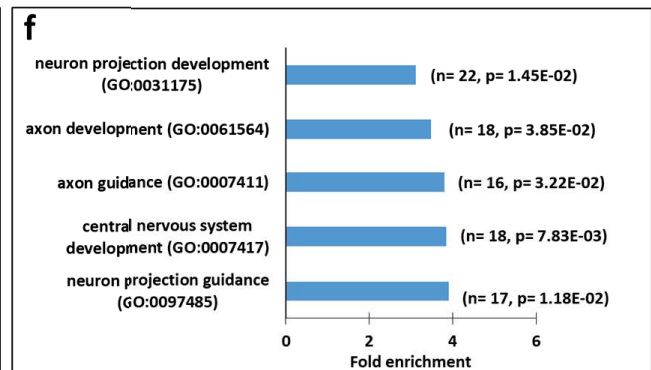
E16 + L3 neurons H3K4me3 + H3K9ac



E16 + L3 glia H3K4me3 + H3K9ac



E16 neurons H3K4me3 + H3K9ac



L3 neurons H3K4me3 + H3K9ac

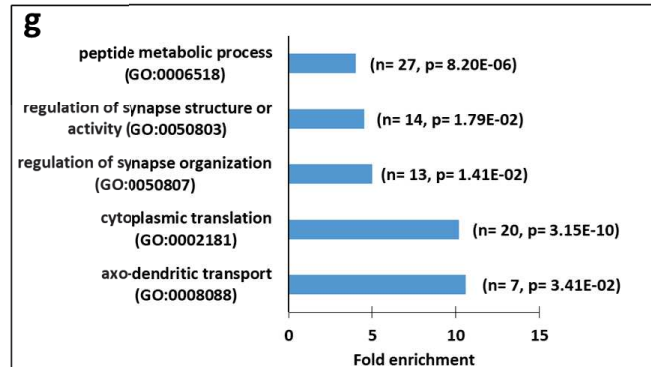


Figure 26: Gene ontology analysis at E16 and L3.

GO term enrichment analysis on genes showing H3K27me3 in E16 and L3 glia (a) and neurons (b). GO term enrichment analysis on genes showing H3K4me3 and H3K9ac in E16 and L3 glia (c), in E16 glia only (d), in E16 and L3 neurons (e), in E16 neurons only (f) and in L3 neurons only (g). Only genes showing the marks at ± 1000 bases around the TSS are taken into account. The fold enrichments of GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

To link these differences in activation marks to expression, I compared the transcriptomes of E16 and L3 glia or neurons and identified genes that are upregulated at one stage compared to the other (**Figure 27-a,b**). In concordance with the results found in comparing H3K4me3 and H3K9ac, genes upregulated in E16 glia are involved in the blood brain barrier, glia development and synapses (**Figure 26-c**). In addition, genes upregulated in L3 glia show no glial function (**Figure 27-d**). A similar trend is observed in the case of neurons. E16 neurons express genes involved in axons, synapse and the development of neurons at higher levels than L3 neurons (**Figure 27-e**). In addition, no neuron-specific function is enriched in genes upregulated in L3 compared to E16 neurons (**Figure 27-f**).

To investigate whether the loss of specific signatures at L3 is due to a loss of H3K9ac or H3K4me3, I compared the presence of the two marks at L3 on the cell-specific genes that present both marks at E16. Results show that on average 70% of genes that show the combination of H3K4me3 and H3K9ac at E16 but not at L3 lose H3K9ac. On the other hand, only around 15% of these genes lose H3K4me3 at L3. Thus, the genes involved in cell-specific functions show H3K4me3 and H3K9ac early in development, then lose H3K9ac later on. This shows again that H3K9ac is needed in combination with H3K4me3 to increase expression level since these genes are upregulated in E16 glia compared to L3 glia.

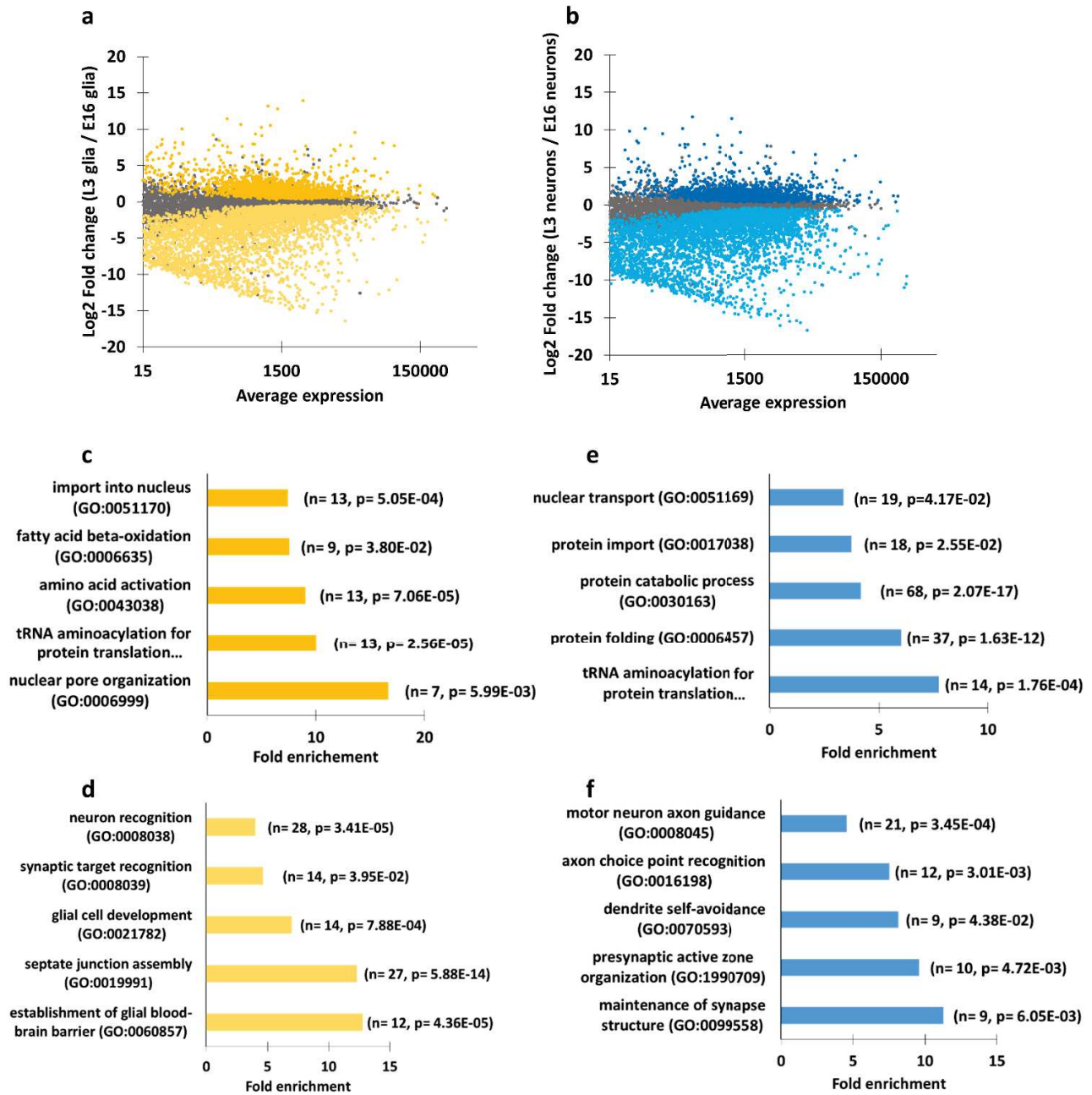


Figure 27: Transcriptome analysis.

(a,d) Transcriptome comparison of E16 and L3 glia (a) or E16 and L3 neurons (d). The x-axis is the average gene expression levels ($n = 3$), and the y-axis is the \log_2 fold change E16 glia/ L3 glia. Genes significantly enriched (adjusted p-value < 0.05 , Fold change > 1.5 , # reads > 15) in E16 glia are shown in yellow, genes significantly enriched in L3 glia are shown in orange, genes significantly enriched in E16 neurons are shown in light blue, genes significantly enriched in L3 neurons are shown in navy and non-differentially expressed genes are shown in gray. (b,c,e,f) GO term enrichment analysis on genes upregulated in E16 glia (b), L3 glia (c), E16 neurons (e) and L3 neurons (f). The fold enrichments for a subset of representative GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

To summarize, these results show that H3K27me3 remains relatively unchanged throughout the life of the cell. This is probably due to the fact that effector genes such as fate determinants must be repressed at all times to maintain cell identity. Similarly, the activation mark H3K4me3 seems to be relatively stable. In contrast, H3K9ac shows an important shift between E16 and L3 following the same trend as expression. Most cell-specific functions that give the cell its identity have H3K4me3 and H3K9ac at E16 only. Consequently, these genes are expressed at higher levels in E16 than at L3. This again shows the tight relation between the two marks and expression levels but also relates to the state of cells at a certain time point. For instance, the repression of genes modulating other fates alone is not enough to consolidate cell fate, it must be accompanied by the activation of the correct genes.

7. Comparing CUT&RUN and ChIP-seq data

Because of the novelty of the CUT&RUN technique, I complemented and confirmed my data by performing a ChIP-seq assay on H3K9ac for sorted E16 glia and neurons.

I first compared the profile of the H3K9ac peaks obtained with ChIP-seq and CUT&RUN. The distribution of the peak over the genomic region is similar, with a small peak starting before the TSS followed by a bigger peak after the TSS. This is especially the case in glia. The main difference is the height of the peak representing the read count frequency (**Figure 28-a,b, Figure 24-b**). As reported before, the read count is clearly lower in the ChIP-seq, which is likely due to the lower signal to noise ratio in CUT&RUN assays (Skene and Henikoff, 2017).

I also tested the correlation between H3K9ac and gene expression. Even though H3K4me3 was not included in the analysis and we know that the combination of both is needed to indicate active expression, we still see a significant increase in expression for genes presenting H3K9ac in glia and in neurons (**Figure 28-c,d**).

I then proceeded with GO term analyses on genes that are highly expressed and presenting H3K9ac around the TSS. As expected, these genes are indeed involved in cell-specific functions. In glia, I find terms related to glia function such as establishment of glial blood-brain barrier, septate junction assembly and glial cell differentiation. In neurons, I find the terms dendrite guidance and synapse assembly (**Figure 28-e,f**). These results are in concordance with the data found using CUT&RUN and thus further validate my findings.

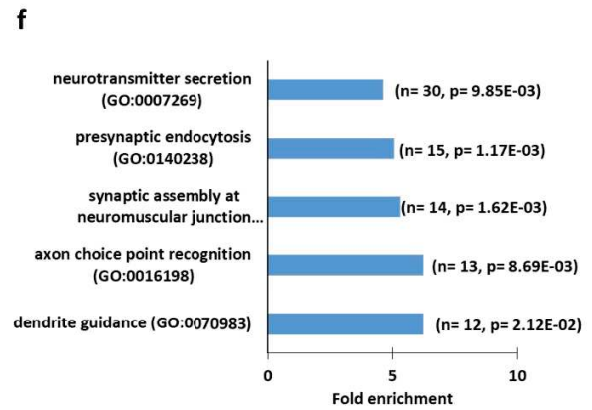
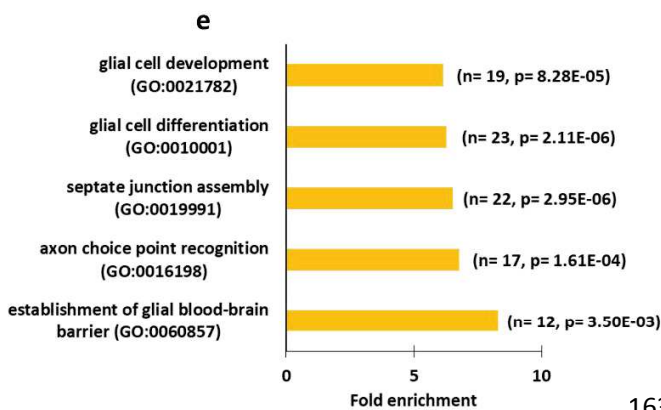
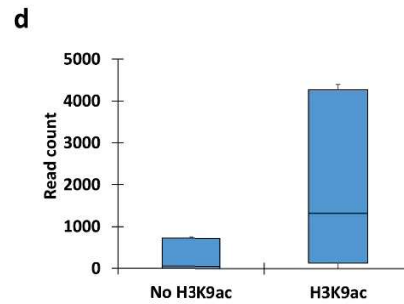
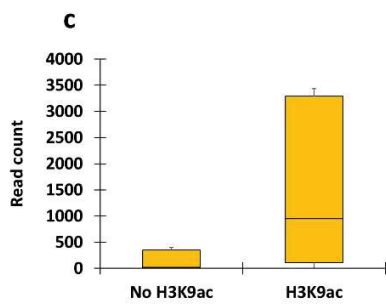
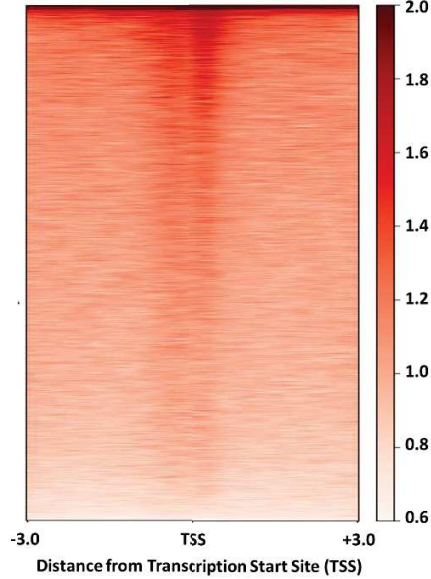
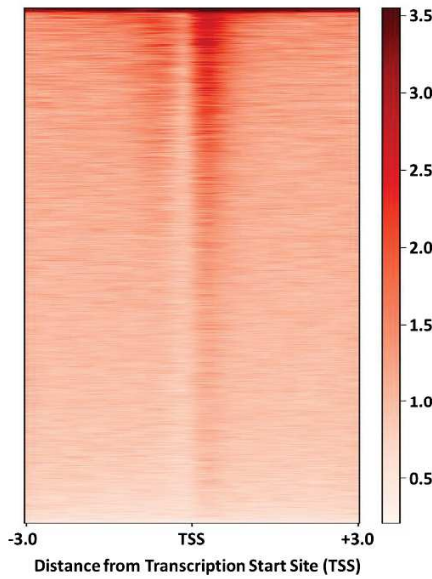
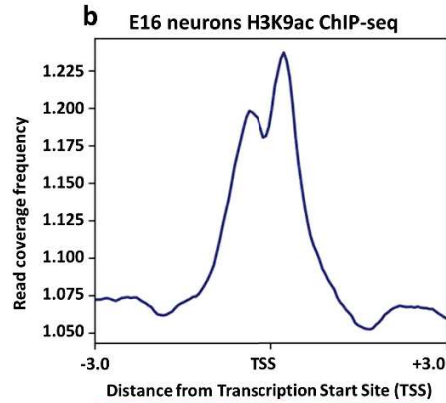
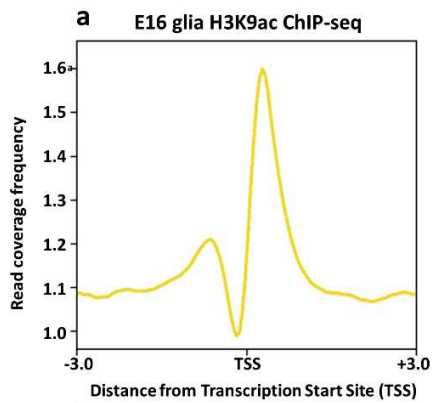


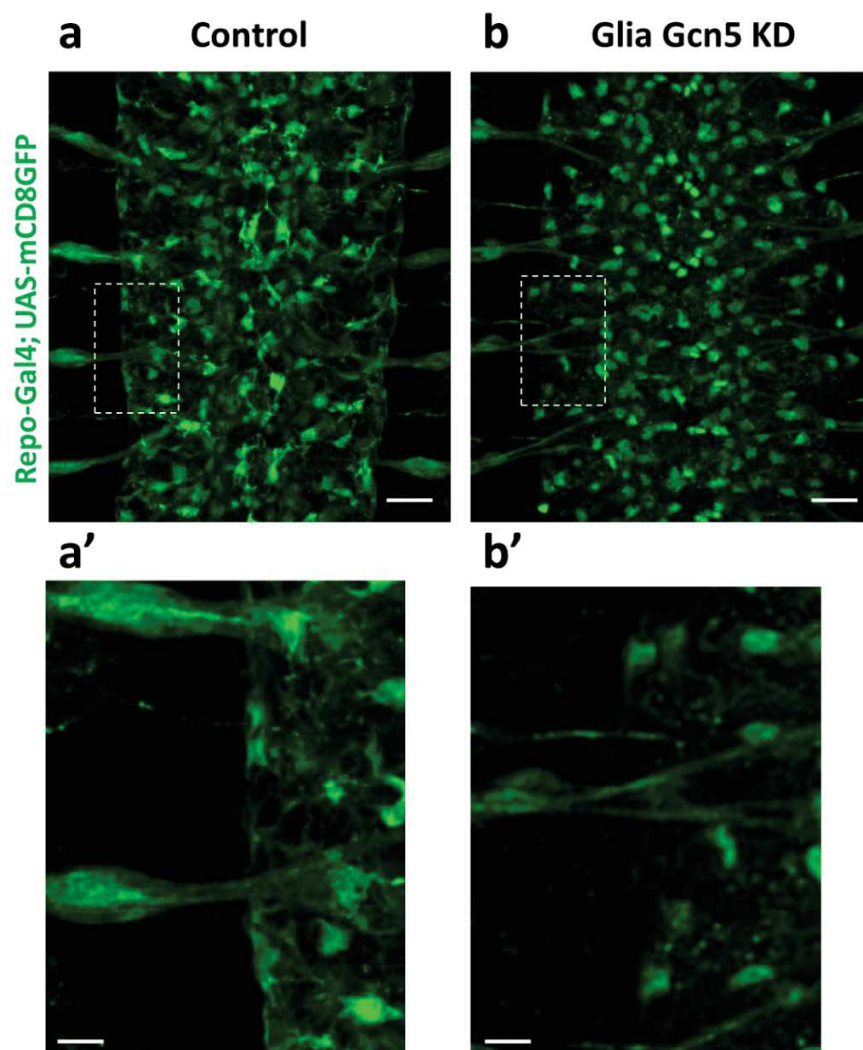
Figure 28: ChIP-seq data analysis.

(a,b) Profile and heatmap of H3K9ac in E16 glia (a) and E16 neurons (b). The x-axis presents the genomic region from 3 Kb before to 3 Kb after the transcription start site (TSS). The y-axis presents the read count frequency. (c,d) Correlation between H3K9ac and expression levels. The x-axis shows the genes that have no H3K9ac and those that have H3K9ac around the TSS. The y-axis represents the normalized reads count. For each group of genes, the graph was plotted based on the 1st quartile, median and 3rd quartile of gene expression. Error bars show the standard error. p shows the p-value that was computed using Student's t-test. (e,f) GO term enrichment analysis on genes having H3K9ac in E16 glia (e) and in E16 neurons (f). The fold enrichments of GO terms are displayed on the x-axis, the number of genes and the P-value of the GO term enrichment are indicated in brackets.

8. Preliminary data on Gcn5 knock-down in glia

To understand the functional relevance of H3K9ac in glia and neurons I used RNAi to knock-down Gcn5, the main histone acetyltransferase (HAT) targeting H3K9 (Grant et al., 1999; Helmlinger et al., 2021). Since one of the most prominent GO terms found in glial genes presenting H3K9ac is the establishment of the Blood-Brain Barrier (BBB) I decided to first assess this structure. The BBB in *Drosophila* is formed by two glial subtypes, the Subperineural glia (SPG) and the Perineural glia (PNG). I used the pan-glial *repo-Gal4* driver combined with the *UAS-mCD8GFP* reporter that localizes the GFP to the membrane of glial cells allowing the tracking of cell morphology. For the knock-down, the *repo-Gal4; UAS-mCD8GFP* line was crossed with the *UAS-Gcn5-RNAi* line. Confocal microscopy imaging on late embryos (stage 16-17) deriving from this cross (KD) and from the control driver alone (control) highlighted differences in the ventral nerve cord (VNC). The barrier around the VNC that is formed by glial cells is disrupted in the KD compared to the control animals. In addition, the VNC of the KD is slightly less dense than in the control (**Figure 29-a,b**).

To explore the role of H3K9ac on the BBB further, I labeled glia-specific KD embryos and control embryos with anti-Repo, to label glia nuclei, and anti-Neurexin IV (Nrx-IV), since Nrx-IV is a transmembrane protein that is critical for septate junction formation between the SPGs. My analysis showed that Nrx-IV is highly expressed and shows H3K4me3 and H3K9ac around the TSS in glia. Control animals present a continuous Nrx-IV signal around the VNC colocalizing with flat Repo positive cells representing the surface glia (**Figure 29-c,d**). Gcn5 KD animals display a disrupted Nrx-IV signal and the shape and position of the glia cells are altered. These results indicate that Gcn5 affects SPGs and the integrity of the BBB.



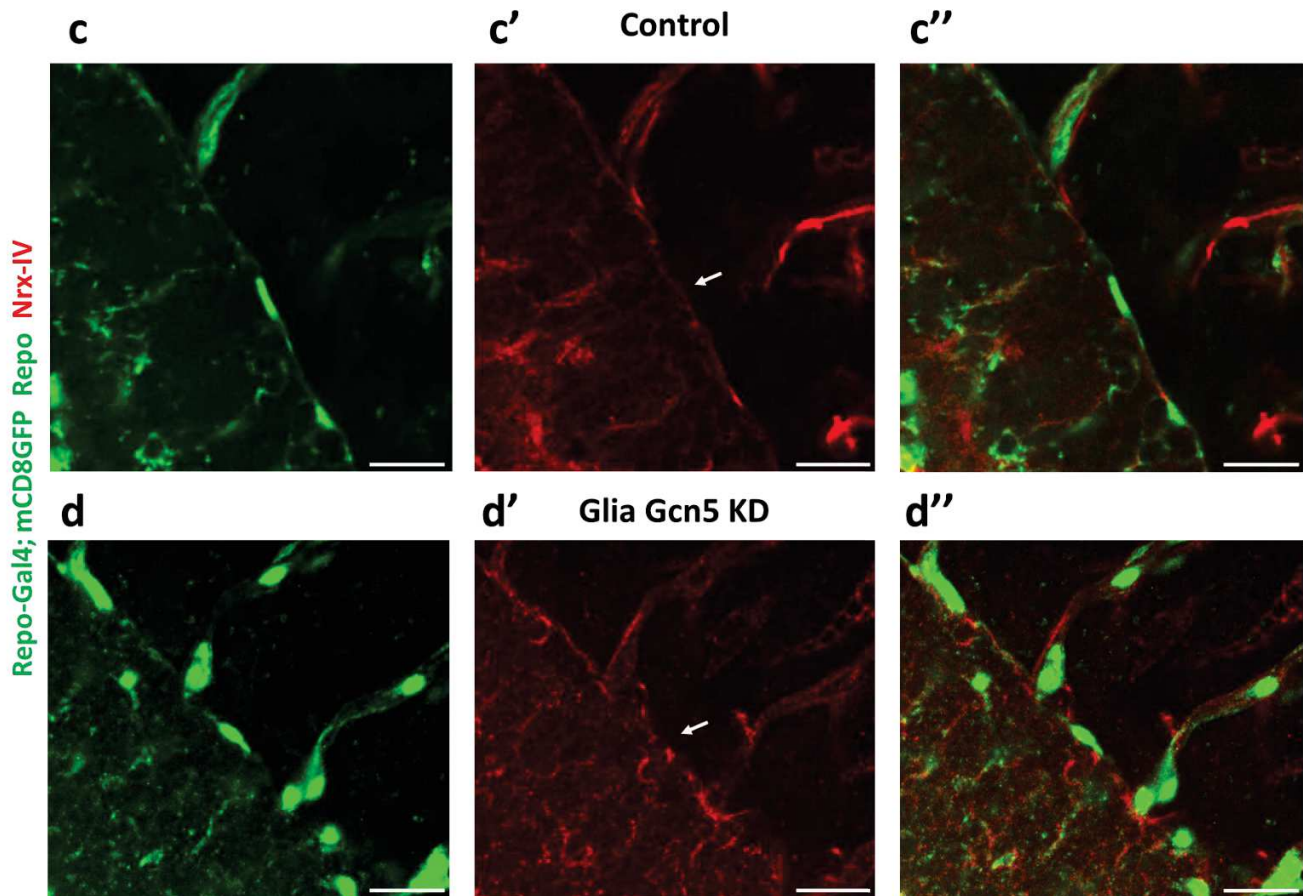


Figure 29: Glia-specific Gcn5 downregulation phenotypes in the embryo.

(a,b) Projection of confocal images on E16 embryos from *repo-Gal4;UAS-mCD8GFP/+* (a) and *repo-Gal4;UAS-mCD8GFP;UAS-Gcn5-RNAi* (b) showing glia cells (green). (a',b') zoomed image representing the dashed square in (a,b). (c-d'') Projection of confocal images on E16 embryos from *repo-Gal4; UAS-mCD8GFP/+* (c-c'') and *repo-Gal4; UAS-mCD8GFP;UAS-Gcn5-RNAi* (d-d'') immunolabeled with mouse anti-Repo (green) and rabbit anti-Neurexin IV (red).

I then tracked this phenotype through development and studied the CNS of wandering larvae. The size of the CNS in glia-specific *Gcn5 KD* decreases significantly compared to that of the control CNS (**Figure 30-a**). During larval stages, the two subtypes of surface glia behave differently to keep up with the growing CNS. PG glia are known to proliferate while SPG do not proliferate but grow in size. To test if the smaller CNS size is due to altered proliferation of PNG, I assessed the proliferation of glia cells in the L3 CNS using an anti-phospho-histone H3 (PH3) antibody, a marker for mitotic cells (Hendzel et al., 1997). The analysis indicates a lower number of PH3 positive PNG in *Gcn5 KD* animals (**Figure 30-b,c**). This indicates that the disruption of the BBB might be due to *Gcn5 KD* affecting both SPG and PNG through the disruption of septate junctions and proliferation, respectively.

The main function of the BBB is to isolate the CNS and control the diffusion of molecules. For this reason, I performed permeability assay using fluorescent dextran. In normal conditions where the BBB is intact, the 4 KDa dextran cannot penetrate inside the CNS. This analysis showed variable results between the different replicates that led to no conclusive finding so far. However, as shown in the figure, in some cases the difference between the control and glia specific *Gcn5 KD* is striking (**Figure 30-d,e**). I am currently optimizing the permeability assay to reduce the technical variability.

In sum, the data suggest a direct link between *Gcn5 KD* and the function of genes carrying H3K9ac in glia. There is clear impact of *Gcn5 KD* on the formation of the BBB even though more experiments are needed to reach a conclusive result. I am also assessing the impact of *Gcn5 KD* specifically in neurons. I have found that *Gcn5 KD* in neurons causes developmental lethality at

late pupal stages. This suggests that H3K9ac might be needed again during metamorphosis when most of the fly's tissues need to be reestablished.

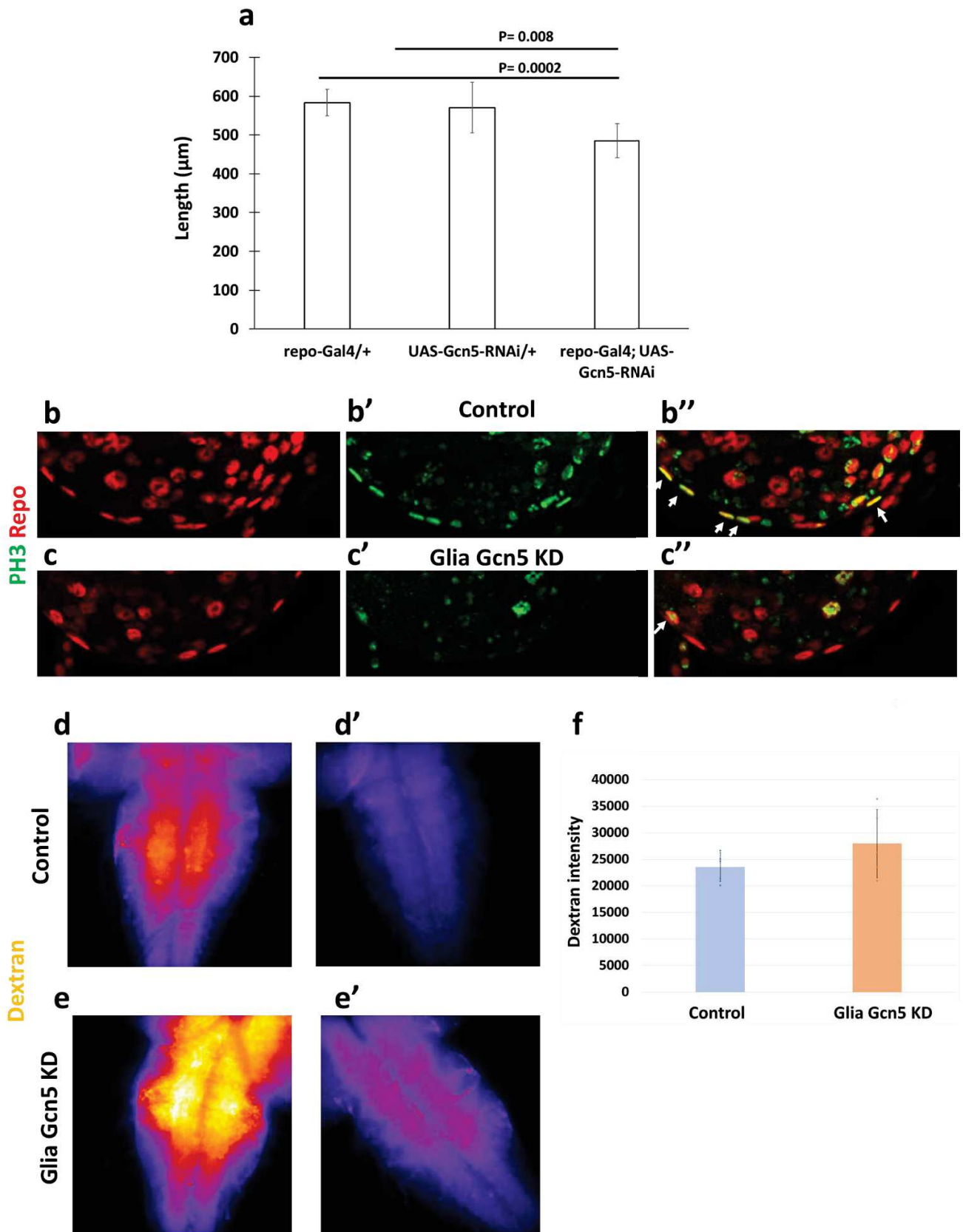


Figure 30: Glia-specific Gcn5 downregulation phenotypes in the larva.

(a) Graph comparing the size of the CNS in L3 from *repo-Gal4/+*, *UAS-Gcn5-RNAi/+* and *UAS-Gcn5-RNAi* animals. The y-axis represents the length of the CNS. Error bars represent the standard deviation and p indicates the p-value computed using Student's t-test. (b-c'') Confocal image of the L3 brain lobe from *repo-Gal4/+* (b-b'') and *repo-Gal4; UAS-Gcn5-RNAi* (c-c'') animals. CNSs are labeled with mouse anti-Repo (red) and rabbit anti-PH3 (green). (d-e') BBB permeability assay: confocal images showing the ventral nerve cord (VNC) of L3 from *repo-Gal4/+* (d,d') and *repo-Gal4; UAS-Gcn5-RNAi* (e,e'). The fluorescent dextran is shown in fire colors. VNCs with the most penetration in both genotypes (d,e) and VNCs with the least penetration (d',e') are shown. (f) quantification of the dextran intensity n = 6. Bars represent the standard deviation.

IV- Discussion:

While the role of epigenetic modifications in stem cells maintenance and in cell differentiation has become evident (Atlasi and Stunnenberg, 2017; Srinageshwar et al., 2016; Wu and Sun, 2006), the impact of these epigenetic marks on differentiated cells and on the definition of cellular identity remains less understood. In chapter I, I identified the cell-specific gene expression profile of different cell types. Since the molecular signature of a cell is defined by its transcriptome and its epigenome, I investigated the correlation between the cell-specific expression profile and three histone marks: H3K4me3, H3K9ac and H3K27me3. Using the cutting-edge CUT&RUN technique I established the distribution of these marks in neurons and glia. These two cell types emerge from the same precursor to take on highly specialized functions and morphologies. The acquisition of different identities calls for epigenetic modifications reflecting this transition. My analyses link histone marks to the expression profile of cells and highlight the presence of cell-specific epigenetic signatures. In addition, I identified the evolution of the epigenetic signature through development and showed that it is a dynamic feature. Finally, I showed that modulating histone modifications in a specific cell type impinges on the function of the cell.

1. Correlation between expression levels and histone marks

The relationship between histone marks and transcriptional activity has been intensively studied for the past two decades and histone marks are efficient predictors of gene expression (González-Ramírez et al., 2021; Karlić et al., 2010; Read et al., 2019). In my study, I tested the relationship between H3K4me3, H3K9ac, H3K27me3 and expression levels. As expected, H3K27me3 shows negative correlation with expression, consistent with its role as a transcriptional

repressor (Wiles and Selker, 2017). While H3K9ac is usually linked to activation, its presence without H3K4me3 does not correlate with expression. This highlights that H3K9ac is not sufficient to activate transcription, while H3K4me3 shows high overall correlation with expression.

The importance of the combinatorial effect of histone marks is shown by the finding that genes presenting both H3K4me3 and H3K9ac are expressed at higher levels than those presenting only H3K4me3 or H3K9ac. Interestingly, it was demonstrated that H3K4me3 modulates the acetylation of neighboring H3K9 within the same nucleosome and that the two marks seem to cooperate to activate gene expression (Foulds et al., 2013). H3K4me3 has been shown to promote transcription initiation, and H3K9ac induces RNA polymerase II pause release by directly recruiting the super elongation complex (SEC) to chromatin (Gates et al., 2017). This suggests a cooperative relationship between the two marks possibly inducing the increased expression.

2. H3K9ac reflects cell identity

Beyond the correlation between expression and histone marks, the purpose of my study is to link cell identity to histone signature. Indeed, I found that H3K9ac marks cell-specific genes involved in the development and function of neurons and glia.

The first indicator of H3K9ac being a cell-specific modification is the correlation heatmap. The heatmap split the data by cell type for H3K9ac and H3K27me3 but not for H3K4me3. This indicates differential distribution of H3K9ac and H3K27me3. In addition, GO term analyses identified cell-specific biological processes in the group of genes having both H3K9ac and H3K4me3 in neurons and in glia. This was not the case for genes showing H3K4me3 alone as these genes are involved in functions shared by different cell types. A concrete example of this

observation is the case of the cell-specific genes *elav* and *ttk* that are enriched for H3K4me3 and H3K9ac around the TSS in neurons and glia respectively. Another recent study showed differential distribution of H3K9ac between HeLa and lymphoblastoid and suggested that this could be linked to cell-type specific transcription (Halsall et al., 2021). However, this was done on immortalized cells *in vitro* and was not shown on differentiated *in vivo* until now.

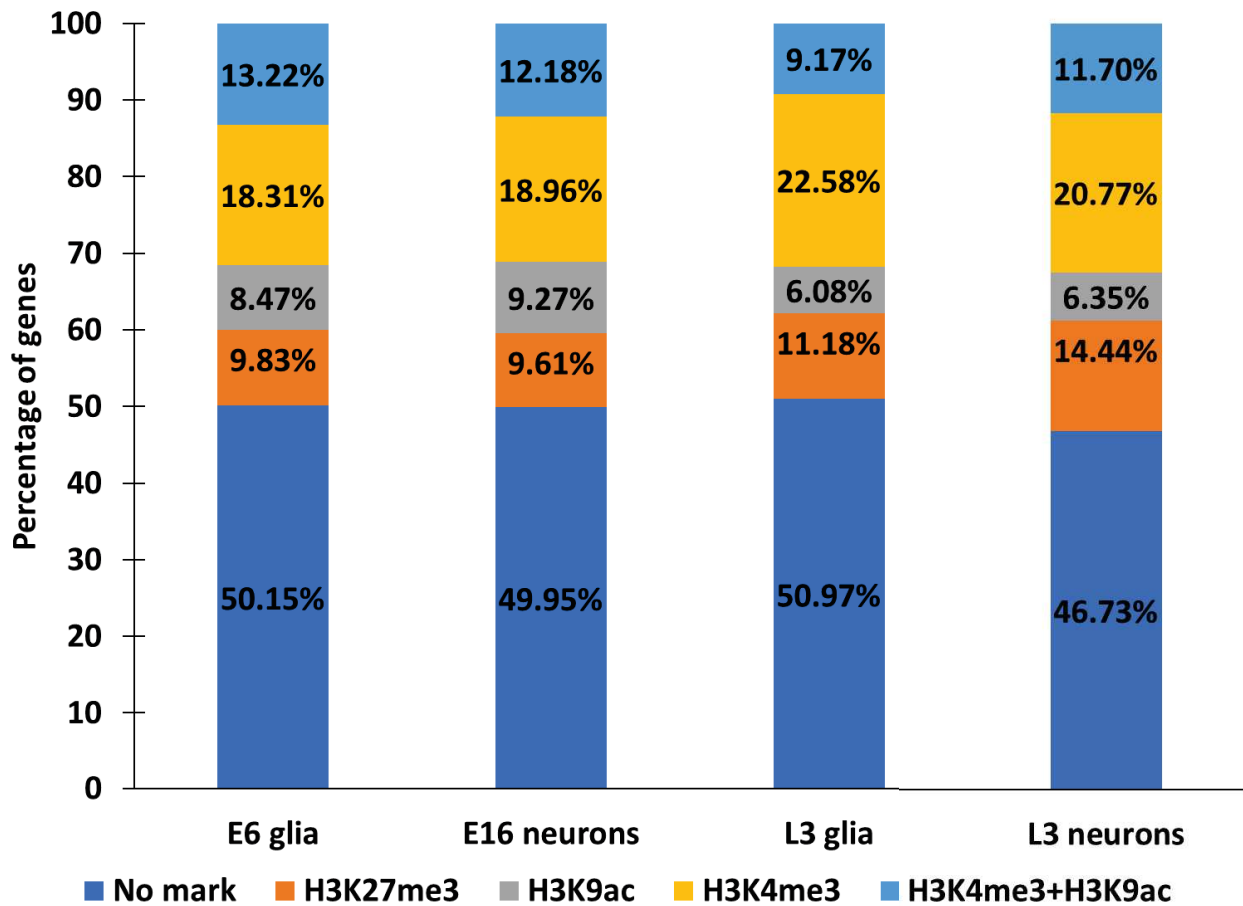
3. Modulation of H3K4me3, H3K9ac and H3K27me3 profiles through development

To investigate how the histone profile of cells behaves at different developmental stages, I compared the distribution of the three marks in E16 and L3 glia and neurons. This analysis indicates that H3K27me3 is a stable mark: the biological processes that are repressed at E16 remain repressed at L3 and no specific processes are found in a single stage in both neurons and glia. This goes along with the idea that H3K27me3 is the physical support of the PcG mediated epigenetic memory of cell identity (Hugues et al., 2020). Studies indicate that PcG, in particular PRC2, is not required to initiate repression but is necessary to maintain gene repression (Reinberg and Vales, 2018; Schuettengruber et al., 2007). Therefore, PRC2 and subsequently H3K27me3, preserve the gene expression profile that reflects cell identity through the constant repression of other genes.

I found that genes involved in cell-specific processes present the H3K4me3 and H3K9ac combination either at E16 only or at E16 and L3. However, few cell-specific genes show H3K9ac and H3K4me3 at L3 only. This is due to most genes involved in the cell-specific processes losing H3K9ac at L3 but maintaining H3K4me3. Thus, unlike H3K4me3, H3K9ac is not a stable mark. This is concordant with the fact that most of these genes are less expressed at L3 than at E16, even though they remain expressed at high levels. The stability of H3K4me3 has been suggested in the context of mitosis where the mark is found in the daughter cell. A study in *Drosophila* showed that

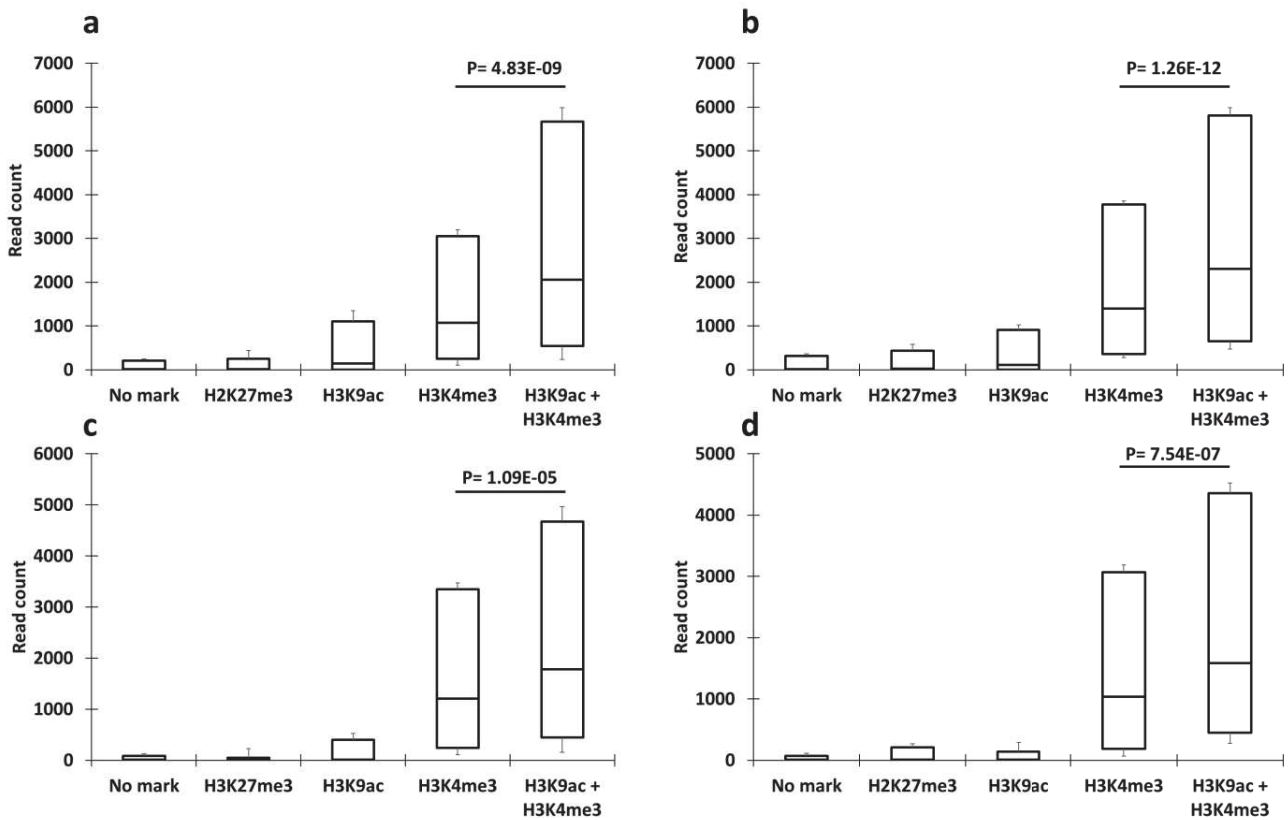
the mark itself is not inherited by the daughter cell but it is the H3K4 methyltransferase Trithorax that remains associated to the replicating DNA (Petruk et al., 2012). On the other hand, the peculiar finding that H3K9ac is lost on cell-specific genes in L3 raises questions regarding the role of this mark. One hypothesis could be that H3K9ac is needed early in the cell to boost the expression of cell-specific genes. The H3K9ac mediated increase in cell-specific genes might contribute to establishing the epigenome of the newly differentiated cell. Once the epigenome is established H3K27me3 continuously represses unwanted genes and H3K4me3 sustains the expression of specific genes at relatively high levels.

V- Supplementary data:



Sup. Figure 5: Distribution of H3K27me, H3K9ac and H3K4me3 in the different samples.

The x-axis shows the percentage of genes. The y-axis represents the different samples. The percentage of each group of genes is shown inside the bars.



Sup. Figure 6: Correlation between histone marks and gene expression.

(a) E16 glia, (b) E16 neurons, (c) L3 glia and (d) L3 neurons. The x-axis shows genes that have none of the three marks, the genes that have H3K27me3, the genes that have H3K9ac but no H3K4me3 around the TSS, the genes that have H3K4me3 but not H3K9ac around the TSS and the genes showing both H3K4me3 and H3K9ac around the TSS. The y-axis represents the normalized reads count. For each group of genes, the graph was plotted based on the 1st quartile, median and 3rd quartile of gene expression. Error bars show the standard error. P shows the p-value that was computed using Student's t-test.

CHAPTER III

Investigating cell identity at the single cell level:

In chapters I and II I explored the mechanisms underlying cell identity by investigating expression profiles and linking them to cell-specific epigenetic signatures. This work was done on population of cells belonging to the same type, which allowed me to establish the genome-wide signatures at high depth, including all annotated genes. However, the cells were sorted according to single markers, which does not allow the distinction of subtype-specific features.

The different subtypes of glia and neurons and their proper characteristic have been extensively studied (Freeman, 2015; Li et al., 2017; Sancer and Wernet, 2021; Yildirim et al., 2019). On the other hand, hemocytes are only subdivided into two types: plasmatocytes that make around 95% of hemocytes and the remaining are crystal cells. However, the diversity of functions of the hemocytes going from the secretion of extra cellular matrix (ECM) to phagocytosis make it unlikely that all these cells are able to perform all the different functions (Brown, 2011; Kurant et al., 2008; Martinek et al., 2008). This suggests the presence of subtypes with different potentials.

To evaluate the heterogeneity of hemocytes, we submitted them to a single cell RNA sequencing (scRNA-seq) assay. The identification of expression profiles of single cells allows the direct assessment of the fundamental unit of life, the cell. The main steps in scRNA-seq comprise: cell isolation, cell lysis, reverse transcription, amplification and transcript coverage. This technology uses unique molecular identifiers (UMI), barcodes that can be applied to detect and quantify the unique transcripts thus each read can be assigned to its original cell (Hwang et al., 2018). We also compared the hemocyte bulk transcriptome in the embryo and in the larva. The mRNA-seq data established that hemocytes acquire their immune potential by L3. We validated

this finding functionally by showing that L3 hemocytes are more potent phagocytes than E16 hemocytes. The scRNA-seq on L3 hemocytes identified 14 different subtypes characterized by different markers and enrichment of different biological processes, likely reflecting specific identities. While in the previous chapters we explored the stability of cell identity through development, here we explored this upon immune challenge. We identified two different subtypes of lamellocytes that emerge upon challenge, presenting different markers. We also investigated the impact of immune challenge on the subtypes identified previously. Interestingly, we found that the number of cells within the different clusters changes upon challenge. However, the main markers identifying the clusters remain constant, which indicates that the subtypes reflect identities that remain present after immune challenge.

To summarize, this work allowed us to establish the first Atlas of *Drosophila* hemocytes and to highlight the existence of different cell identities, with different potentials within the same cell population.

**Manuscript: Temporal specificity and heterogeneity of *Drosophila*
immune cells**

Pierre B Cattenoz^{1,2,3,4,*}, Rosy Sakr^{1,2,3,4,†}, Alexia Pavlidaki^{1,2,3,4,†}, Claude Delaporte^{1,2,3,4}, Andrea Riba^{1,2,3,4}, Nacho Molina^{1,2,3,4}, Nivedita Hariharan^{5,6}, Tina Mukherjee⁵ & Angela Giangrande^{1,2,3,4,**}

The EMBO Journal, March 2020

DISCUSSION AND PERSPECTIVES

1. Cell identity at the transcriptional and epigenetic levels

mRNA sequencing on glia, neurons and hemocytes allowed us to demonstrate how the cell-specific gene expression profile reflects the origin and the function of a cell. Neurons and glia showed the highest similarity, due to their common origin and micro-environment. Glia and hemocytes share the function of phagocytosis, which was shown in the expression of scavenger receptors in both cell types. The genome-wide analysis also helped demonstrating that, while some receptors are expressed at similar levels, others are more specific to either cell type, indicating different phagocytic properties. The hemocytes express scavenger receptors involved in the phagocytosis of pathogens exclusively (Kocks et al., 2005; Kurucz et al., 2007; Rämetsch et al., 2002). We found and validated that the glial receptor *NimA* is not expressed in hemocytes nor in neurons. Deeper investigation of its expression profile and phenotype will help us better understand its role. Typically, is *NimA* expressed in all glia, in a specific subtype? Based on the data from the published single cell transcriptome of adult fly CNS (Davie et al., 2018), *NimA* is found in ensheathing glia, astrocyte-like glia and a few cortex glia cells. It does not show any expression in subperineurial and perineurial glia. Interestingly, the three cell types expressing the scavenger receptor *NimA* have been linked to phagocytosis in the adult brain (Hakim-Mishnaevski et al., 2019).

We have attempted to reveal the expression profile of *NimA* in the embryo glia using *in situ hybridization* assays, we have so far not succeeded, nor did our attempts to generate a Crispr/Cas9 *NimA* mutant (while other genes could be targeted in the laboratory). We decided to generate *NimA* mutants ourselves because only one line existed and there was little information

about the mutation. Luckily, another mutant has surfaced recently. In addition, we have also acquired another *NimA* mutant that was produced by WellGenetics (Taiwan). I will soon be testing these lines and assessing the loss of function phenotype on *NimA* null animals. We are also currently establishing better tools and protocols for the *in-situ* assays. Needless to say, *NimA* might be expressed in other cells with phagocytic potential such as ovarian follicle, epithelial cells (Serizier and McCall, 2017), which will be interesting to analyze for comparison with glia.

The genome-wide analyses highlighted a clear link between histone modifications and the cell-specific expression profile. H3K4me3 displays strong correlation with expression but not with cell specificity. This goes along with H3K4me3 being a general mark of active transcription (Heintzman et al., 2007; Ruthenburg et al., 2007). Specificity is highlighted in genes that present a combination of H3K4me3 and H3K9ac. These genes are expressed at high levels and are involved in cell-specific functions in neurons and glia. In addition, the glia-specific *Gcn5* knockdown leads to disruption in glia-specific processes. Altogether, these findings indicate that H3K9ac is associated to cell identity. The mechanism behind the high expression of genes presenting both H3K9ac and H3K4me3 can be explained by several observations: SAGA (Spt-Ada-Gcn5 acetyltransferase), the main modifier of H3K9ac, is required for RNA polymerase II recruitment at the promoter (Baptista et al., 2017) and H3K9ac is needed for transcriptional elongation (Gates et al., 2017).

As per the specificity of H3K9ac, this could be due to the recruitment of SAGA to specific loci by sequence-specific transcription factors. While some studies indicate that SAGA acts at specific genes, others indicate that it functions as a general transcription factor, thus this aspect of

SAGA is still controversial (Baptista et al., 2017; Bonnet et al., 2014; Krebs et al., 2011; Nagy et al., 2009).

The acetylation of H3K9 on genes that are specific to glial identity could be mediated by glial fate determinants. Gcm, the glial fate determinant is transiently expressed early in glia and is needed to activate the expression of subsequent glial genes. This combined with the finding that H3K9ac is only present on specific genes at early stages alludes to a possible role of Gcm in the cell specific deposition of H3K9ac. Previous studies in our lab established the targets of Gcm through a DNA adenine methyltransferase identification (DamID) screen (Cattenoz et al., 2016). One of these targets is Gcn5, the HAT subunit of SAGA. In addition, other works in our lab aimed to identify the physical interaction between Gcm and different histone modifiers. Immunoprecipitation assays showed that there is no protein-protein interaction between Gcm and Gcn5 nor with Ada2b, however Gcm was found to interact with Tra1, another subunit of SAGA and with CBP which in turn interacts with Gcn5 (Xu et al., 1998). All these results combined suggest that Gcm might be involved in the activation of Gcn5 on the transcriptional level and could recruit SAGA to specific genes via its interaction with Tra1 or/and CBP as well. Since acetylation is a very dynamic mechanism and we are investigating late embryonic stages, at which Gcm is no longer expressed, we can also hypothesize that while Gcm might be involved in H3K9ac early on, other factors are involved in the maintenance of the marks later in the differentiated cell. Repo is a glial transcription factor induced by Gcm and known to be necessary for execution of the glial differentiation program (Trébuchet et al., 2019; Yuasa et al., 2003). Studies in our lab revealed that Repo also interacts with CBP (Flici et al., 2014), however we still do not know whether Repo interacts with components of SAGA. Therefore, assessing whether we find Repo-SAGA

interaction and/or whether components of SAGA are transcriptionally regulated by Repo can help us understand the role of Repo in H3K9ac in differentiated glia.

In addition, SAGA has been shown to function independently from acetylation. Therefore, even if SAGA functions as a general transcription factor, H3K9ac could still be deposited specifically (Fischer et al., 2021).

On the other hand, we cannot exclude the role of HDACs in this mechanism. If H3K9ac is indeed deposited on all active genes, HDACs can be recruited to deacetylate specific loci. This is especially relevant since HDACs are known to be needed for differentiation and their inhibition affects cell fate specification, in particular, in the nervous system (Balasubramaniyan et al., 2006; Lee et al., 2004; Shukla and Tekwani, 2020). In *Drosophila*, Ttk, a glia specific gene is able to recruit the chromatin modifying complex containing an HDAC subunit NuRD and was shown to block the expression of neuronal genes. Based on this we can speculate that Ttk represses neuronal genes and other non-glia-specific genes through NuRD (Giesen et al., 1997; Murawsky et al., 2001; Reddy et al., 2010). To test this, we can assess the genome wide distribution of H3K9ac in glia from mutant *ttk* embryos which is possible because even though we have reduction in glia cell numbers we still find Repo positive glia in these mutants (Giesen et al., 1997).

A number of studies have identified a plethora of histone modifications and the link between these modifications and gene expression has become evident. Many of them have been implicated in different diseases and histone modifiers have become targets for the development of therapeutic approaches. In contrast, we still do not know the relevance of this variety of marks. We here show for the first time *in vivo* and in differentiated cells that a specific histone mark,

H3K9ac, plays a role outside of simple transcriptional activation since its presence does not reflect the expression status of the gene but its function and the processes it is involved in.

2. The evolution of cell identity through development

Amongst the three cell types investigated, the most striking shift in expression profile was seen in hemocytes. These cells show very few specific transcripts enriched at E16 but this changes drastically at L3, where they express transcripts associated to immune related biological processes. This shift in the expression profile goes along with hemocytes acquiring new functions at L3, which can be seen clearly in their increased phagocytic ability. This drastic change is most likely due to the fact that the role of hemocytes in the embryo is different than in the larva. The embryo is completely isolated from the surrounding by the chorion, thus during this stage the role of hemocytes is mainly to secrete ECM and to phagocytose apoptotic bodies. At larval stages, however, the animal interacts with its environment and is susceptible to immune challenges so hemocytes take on an immune role.

In the case of neurons and glia, we found that most cell-specific biological processes are enriched in E16 or commonly expressed at E16 and L3 but very few are specific to L3. However, the enrichment of these genes at E16 does not mean they are not expressed L3. Genes involved in cell-specific processes are still expressed at high levels at L3, just lower than at E16. This correlates with the fact that cell-specific genes lose H3K9ac at L3, implying that H3K9ac is needed at early stages to induce high expression of cell-specific genes but not to maintain gene expression. To test this hypothesis, we first need to show correlation between H3K9ac and the expression levels of the specific genes. This will be done through modulating H3K9ac levels in a specific cell type and assessing the expression of the cell-specific genes identified in our study. Thus, modulating the

levels of H3K9ac at a specific stage would help pinpoint the exact role of H3K9ac. For instance, we could hypothesize that modulating H3K9ac during embryonic stages would have more impact on cell identity than modulating it at L3.

Due to the dynamic nature of acetylation, the depletion of H3K9ac in specific genes at L3 may be either due to lower HAT activity or increased HDAC activity. Interestingly, assessing the levels of the SAGA components *Gcn5* and *Ada2b* showed no differences between E16 and L3 in glia and neurons. On the other hand, *HDAC1* is expressed at 4 folds higher in L3 than at E16 in glia. Needless to say, the expression levels of these factors do not necessarily indicate the activity of the complexes they are associated with, however this does open the doors to the possibility that HDACs overactivation could be responsible for the lower levels of H3K9ac. If this indeed is the case, then it raises the question of what could be the relevance of active deacetylation of specific genes? This can be assessed by studying the impact of inhibiting the deacetylation of H3K9ac at later stages, through HDAC1 inhibitors. This will allow us to understand whether this deacetylation plays a role in the development and/or function of the cells.

If the loss of H3K9ac is due to the fact that most neuronal and glial functions remain unchanged between E16 and L3, then assessing the profile of H3K9ac in hemocytes will help us validate this hypothesis. Unlike neurons and glia, hemocytes acquire new functions after embryonic stages so we would expect new genes to show H3K9ac at L3 compared to E16.

3. Cell identity on the single cell level

The role of hemocytes goes beyond immunity. They perform many other roles crucial for the homeostasis of the organism. Our data identify 14 different clusters of hemocytes, each having

its own molecular signature and active pathways reflecting a specific function and for some a distinct identity. For instance, the PL-AMP cluster express high levels of antimicrobial peptides which indicates their dedication to the humoral response. In addition, the expression profile of PL-vir1 indicates their involvement in anti-viral response. Beyond immunity, we also find clusters dedicated to homeostasis such as PL-Lsp, which represents the amino acids reservoir. This suggests that different cells within the same population diverge and acquire specific features. This analysis raised the question of whether what we identified is real distinct identities within the hemocyte population or whether it is transient states of the cells captured at a certain time point. In this case it seems to be a combination of both. For instance, some subtypes show clear molecular signatures indicating their functions as mentioned above. In addition, the fact that these signatures remain even after immune challenge, and we do not see strong transcriptomic modifications induced by the challenge, further emphasizes the idea of true identity. Other subtypes however, including PL-0 and PL-1 for example, do not show a specific molecular signature. These cells might be in a transient state as they do not seem to be specialized but might represent the reservoir of functions needed in case of immune or other challenges. The immune challenge did cause shifting in the number of cells in a few clusters, indicating that specific hemocyte clusters may preferentially survive or proliferate upon challenge. Moreover, we still do not know whether the different subtypes identified also present specific epigenetic marks differentiating them from each other and potentially influencing their expression profile. With the current advancement in technology, it seems that we will soon be able to perform single cell RNA-seq and CHIP-seq on the same cell, thus opening the door to a whole new understanding of cell identity. In sum, this analysis suggests that single cells do not go through major changes upon challenge, but the shift can be clearly seen when looking at the population as a whole. This highlights the importance of

bulk transcriptome analysis and the fact that even though single cell RNA-seq can help us answer questions regarding individual cell identity it might overlook the big picture.

Understanding the many aspects of cell identity is not only important from a fundamental point of view since accurate identification of cell identity is also crucial to understanding diseases and targeting therapies. The cell being the smallest functional unit of life, it is the starting point of deregulations that accumulate leading to alteration in cell identity and eventually impacting the entire organism. One example is cancer, in which the loss of cell identity is a critical early step in neoplastic transformation. It is impossible to identify discrete deregulations if we do not understand the biology of cell identity in homeostatic conditions. Here we explore cellular identity at the molecular level and highlight a clear link between expression profile, epigenetic signature, origin and function of the cell. We were able to show *in vivo* that H3K9ac is a cell-specific epigenetic mark, thus opening the doors to explore the functional role of this mark and its implication in cell development, function and identity. We also indicate that cell identity is constantly maintained but it is able to shift and evolve allowing the cells to accommodate to the needs of the organism. Given the evolutionary conservation of the main biological processes, we expect the impact of these *in vivo* findings to go beyond fly glia, neurons and hemocytes and concern the bases of cell identity in other metazoans as well.

REFERENCES

- Abbott, N.J. (2005). Dynamics of CNS barriers: evolution, differentiation, and modulation. *Cell. Mol. Neurobiol.* *25*, 5–23.
- Abdolhosseini, F., Azarkhalili, B., Maazallahi, A., Kamal, A., Motahari, S.A., Sharifi-Zarchi, A., and Chitsaz, H. (2019). Cell Identity Codes: Understanding Cell Identity from Gene Expression Profiles using Deep Neural Networks. *Sci. Rep.* *9*, 2342.
- Abrams, J.M., White, K., Fessler, L.I., and Steller, H. (1993). Programmed cell death during *Drosophila* embryogenesis. *Development* *117*, 29–43.
- Allahverdi, A., Yang, R., Korolev, N., Fan, Y., Davey, C.A., Liu, C.-F., and Nordenskiöld, L. (2011). The effects of histone H4 tail acetylations on cation-induced chromatin folding and self-association. *Nucleic Acids Res.* *39*, 1680–1691.
- ALLFREY, V.G., FAULKNER, R., and MIRSKY, A.E. (1964). ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc. Natl. Acad. Sci. U. S. A.* *51*, 786–794.
- Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhhattar, R., et al. (2007). New nomenclature for chromatin-modifying enzymes. *Cell* *131*, 633–636.
- Amoutzias, G.D., Robertson, D.L., Van de Peer, Y., and Oliver, S.G. (2008). Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* *33*, 220–229.
- An, P.N.T., Shimaji, K., Tanaka, R., Yoshida, H., Kimura, H., Fukusaki, E., and Yamaguchi, M. (2017). Epigenetic regulation of starvation-induced autophagy in *Drosophila* by histone methyltransferase G9a. *Sci. Rep.* *7*.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Andres, M.E., Burger, C., Peral-Rubio, M.J., Battaglioli, E., Anderson, M.E., Grimes, J., Dallman, J., Ballas, N., and Mandel, G. (1999). CoREST: a functional corepressor required for regulation of neural-specific gene expression. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 9873–9878.
- Antignano, F., Burrows, K., Hughes, M.R., Han, J.M., Kron, K.J., Penrod, N.M., Oudhoff, M.J., Wang, S.K.H., Min, P.H., Gold, M.J., et al. (2014). Methyltransferase G9A regulates T cell differentiation during murine intestinal inflammation. *J. Clin. Invest.* *124*, 1945–1955.
- Arents, G., Burlingame, R.W., Wang, B.C., Love, W.E., and Moudrianakis, E.N. (1991). The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 10148–10152.
- Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* *18*, 643–658.
- Avila, A.M., Burnett, B.G., Taye, A.A., Gabanella, F., Knight, M.A., Hartenstein, P., Cizman, Z., Di Prospero, N.A., Pellizzoni, L., Fischbeck, K.H., et al. (2007). Trichostatin A increases SMN expression and survival in a mouse model of spinal muscular atrophy. *J. Clin. Invest.* *117*, 659–671.
- Avvakumov, N., and Côté, J. (2007). The MYST family of histone acetyltransferases and their intimate

- links to cancer. *Oncogene* 26, 5395–5407.
- Awasaki, T., and Ito, K. (2004). Engulfing action of glial cells is required for programmed axon pruning during *Drosophila* metamorphosis. *Curr. Biol.* 14, 668–677.
- Awasaki, T., Lai, S.-L., Ito, K., and Lee, T. (2008). Organization and postembryonic development of glial cells in the adult central brain of *Drosophila*. *J. Neurosci.* 28, 13742–13753.
- Ayyaz, A., Li, H., and Jasper, H. (2015). Haemocytes control stem cell activity in the *Drosophila* intestine. *Nat. Cell Biol.* 17, 736–748.
- B, A., PB, C., and A, G. (2016). The early life of a fly glial cell. *Wiley Interdiscip. Rev. Dev. Biol.* 5, 67–84.
- Baker, W.K. (1968). Position-effect variegation. *Adv. Genet.* 14, 133–169.
- Balasubramanian, V., Boddeke, E., Bakels, R., Küst, B., Kooistra, S., Veneman, A., and Copray, S. (2006). Effects of histone deacetylation inhibition on neuronal differentiation of embryonic mouse neural stem cells. *Neuroscience* 143, 939–951.
- Bandarra, D., Biddlestone, J., Mudie, S., Muller, H.A., and Rocha, S. (2014). Hypoxia activates IKK-NF- κ B and the immune response in *Drosophila melanogaster*. *Biosci. Rep.* 34.
- Banerjee, S., Bainton, R.J., Mayer, N., Beckstead, R., and Bhat, M.A. (2008). Septate junctions are required for ommatidial integrity and blood-eye barrier function in *Drosophila*. *Dev. Biol.* 317, 585–599.
- Banerjee, S., Blauth, K., Peters, K., Rogers, S.L., Fanning, A.S., and Bhat, M.A. (2010). *Drosophila* neurexin IV interacts with Roundabout and is required for repulsive midline axon guidance. *J. Neurosci.* 30, 5653–5667.
- Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase. *Nature* 384, 641–643.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124.
- Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J. Biol. Chem.* 280, 17732–17736.
- Baptista, T., Grünberg, S., Minoungou, N., Koster, M.J.E., Timmers, H.T.M., Hahn, S., Devys, D., and Tora, L. (2017). SAGA Is a General Cofactor for RNA Polymerase II Transcription. *Mol. Cell* 68, 130–143.e5.
- Barnes, V.L., Laity, K.A., Pilecki, M., and Pile, L.A. (2018). Systematic Analysis of SIN3 Histone Modifying Complex Components During Development. *Sci. Rep.* 8, 17048.
- Barres, B.A. (2008). The mystery and magic of glia: a perspective on their roles in health and disease. *Neuron* 60, 430–440.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837.

- Barth, T.K., and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. *Trends Biochem. Sci.* *35*, 618–626.
- Baymaz, H.I., Karemaker, I.D., and Vermeulen, M. (2015). Perspective on unraveling the versatility of “co-repressor” complexes. *Biochim. Biophys. Acta* *1849*, 1051–1056.
- Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.* *5*.
- Bedford, M.T., and Clarke, S.G. (2009). Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* *33*, 1–13.
- Benito, E., Urbanke, H., Ramachandran, B., Barth, J., Halder, R., Awasthi, A., Jain, G., Capece, V., Burkhardt, S., Navarro-Sala, M., et al. (2015). HDAC inhibitor-dependent transcriptome and memory reinstatement in cognitive decline models. *J. Clin. Invest.* *125*, 3572–3584.
- Bennett, S.A., Tanaz, R., Cobos, S.N., and Torrente, M.P. (2019). Epigenetics in amyotrophic lateral sclerosis: a role for histone post-translational modifications in neurodegenerative disease. *Transl. Res.* *204*, 19–30.
- Bernardoni, R., Miller, A.A., and Giangrande, A. (1998). Glial differentiation does not require a neural ground state. *Development* *125*, 3189–3200.
- Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 8695–8700.
- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J. 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* *120*, 169–181.
- Bian, C., Xu, C., Ruan, J., Lee, K.K., Burke, T.L., Tempel, W., Barsyte, D., Li, J., Wu, M., Zhou, B.O., et al. (2011). Sgf29 binds histone H3K4me2/3 and is required for SAGA complex recruitment and histone H3 acetylation. *EMBO J.* *30*, 2829–2842.
- Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M.C., Tassani, S., Piva, F., et al. (2013). An estimation of the number of cells in the human body. *Ann. Hum. Biol.* *40*, 463–471.
- Binggeli, O., Neyen, C., Poidevin, M., and Lemaitre, B. (2014). Prophenoloxidase activation is required for survival to microbial infections in *Drosophila*. *PLoS Pathog.* *10*, e1004067.
- Bittern, J., Pogodalla, N., Ohm, H., Brüser, L., Kottmeier, R., Schirmeier, S., and Klämbt, C. (2021). Neuron-glia interaction in the *Drosophila* nervous system. *Dev. Neurobiol.* *81*, 438–452.
- Bjerling, P., Silverstein, R.A., Thon, G., Caudy, A., Grewal, S., and Ekwall, K. (2002). Functional divergence between histone deacetylases in fission yeast by distinct cellular localization and in vivo specificity. *Mol. Cell Biol.* *22*, 2170–2181.
- Blanc, R.S., and Richard, S. (2017). Arginine Methylation: The Coming of Age. *Mol. Cell* *65*, 8–24.
- Boeger, H., Griesenbeck, J., Strattan, J.S., and Kornberg, R.D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Mol. Cell* *14*, 667–673.
- Boland, M.J., Nazor, K.L., and Loring, J.F. (2014). Epigenetic regulation of pluripotency and differentiation. *Circ. Res.* *115*, 311–324.

- Bonn, S., Zinzen, R.P., Perez-Gonzalez, A., Riddell, A., Gavin, A.-C., and Furlong, E.E.M. (2012). Cell type-specific chromatin immunoprecipitation from multicellular complex samples using BiTS-ChIP. *Nat. Protoc.* *7*, 978–994.
- Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev.* *28*, 1999–2012.
- Van De Bor, V., and Giangrande, A. (2002). *glide/gcm*: at the crossroads between neurons and glia. *Curr. Opin. Genet. Dev.* *12*, 465–472.
- Van De Bor, V., Zimniak, G., Papone, L., Cerezo, D., Malbouyres, M., Juan, T., Ruggiero, F., and Noselli, S. (2015). Companion Blood Cells Control Ovarian Stem Cell Niche Microenvironment and Homeostasis. *Cell Rep.* *13*, 546–560.
- Bossing, T., and Technau, G.M. (1994). The fate of the CNS midline progenitors in *Drosophila* as revealed by a new method for single cell labelling. *Development* *120*, 1895–1906.
- Bossing, T., Udolph, G., Doe, C.Q., and Technau, G.M. (1996). The embryonic central nervous system lineages of *Drosophila melanogaster*. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev. Biol.* *179*, 41–64.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* *118*, 401–415.
- Brankatschk, M., Dunst, S., Nemetschke, L., and Eaton, S. (2014). Delivery of circulating lipoproteins to specific neurons in the *Drosophila* brain regulates systemic insulin signaling. *Elife* *3*.
- Brennan, C.A., and Anderson, K. V (2004). *Drosophila*: the genetics of innate immune recognition and response. *Annu. Rev. Immunol.* *22*, 457–483.
- Briggs, S.D., Bryk, M., Strahl, B.D., Cheung, W.L., Davie, J.K., Dent, S.Y., Winston, F., and Allis, C.D. (2001). Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* *15*, 3286–3295.
- Brink, D.L., Gilbert, M., Xie, X., Petley-Ragan, L., and Auld, V.J. (2012). Glial processes at the *Drosophila* larval neuromuscular junction match synaptic growth. *PLoS One* *7*, e37876.
- Brower-Toland, B., Riddle, N.C., Jiang, H., Huisinga, K.L., and Elgin, S.C.R. (2009). Multiple SET methyltransferases are required to maintain normal heterochromatin domains in the genome of *Drosophila melanogaster*. *Genetics* *181*, 1303–1319.
- Brown, N.H. (2011). Extracellular matrix in development: insights from mechanisms conserved between invertebrates and vertebrates. *Cold Spring Harb. Perspect. Biol.* *3*.
- Brown, S.W. (1966). Heterochromatin. *Science* *151*, 417–425.
- Brownell, J.E., and Allis, C.D. (1996). Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation. *Curr. Opin. Genet. Dev.* *6*, 176–184.
- Brownell, J.E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* *84*, 843–851.
- Brückner, K., Kockel, L., Duchek, P., Luque, C.M., Rørth, P., and Perrimon, N. (2004). The PDGF/VEGF receptor controls blood cell survival in *Drosophila*. *Dev. Cell* *7*, 73–84.

- de Bruijn, M.F.T.R., and Speck, N.A. (2004). Core-binding factors in hematopoiesis and immune function. *Oncogene* 23, 4238–4248.
- Brusslan, J.A., Bonora, G., Rus-Canterbury, A.M., Tariq, F., Jaroszewicz, A., and Pellegrini, M. (2015). A Genome-Wide Chronological Study of Gene Expression and Two Histone Modifications, H3K4me3 and H3K9ac, during Developmental Leaf Senescence. *Plant Physiol.* 168, 1246–1261.
- Bu, P., Evrard, Y.A., Lozano, G., and Dent, S.Y.R. (2007). Loss of Gcn5 acetyltransferase activity leads to neural tube closure defects and exencephaly in mouse embryos. *Mol. Cell. Biol.* 27, 3405–3416.
- Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Lieberman Aiden, E., Li, S., Manikandan, L., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. *Nat. Commun.* 12, 719.
- Callebaut, I., Mignotte, V., Souchet, M., and Mornon, J.P. (2003). EMI domains are widespread and reveal the probable orthologs of the *Caenorhabditis elegans* CED-1 protein. *Biochem. Biophys. Res. Commun.* 300, 619–623.
- Campuzano, S., Carramolino, L., Cabrera, C. V, Ruíz-Gómez, M., Villares, R., Boronat, A., and Modolell, J. (1985). Molecular genetics of the achaete-scute gene complex of *D. melanogaster*. *Cell* 40, 327–338.
- Candau, R., Moore, P.A., Wang, L., Barlev, N., Ying, C.Y., Rosen, C.A., and Berger, S.L. (1996). Identification of human proteins functionally conserved with the yeast putative adaptors ADA2 and GCN5. *Mol. Cell. Biol.* 16, 593–602.
- Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., et al. (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* 496, 377–381.
- Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr. Opin. Genet. Dev.* 14, 155–164.
- Carré, C., Szymczak, D., Pidoux, J., and Antoniewski, C. (2005). The histone H3 acetylase dGcn5 is a key player in *Drosophila melanogaster* metamorphosis. *Mol. Cell. Biol.* 25, 8228–8238.
- Carrozza, M.J., Utley, R.T., Workman, J.L., and Côté, J. (2003). The diverse functions of histone acetyltransferase complexes. *Trends Genet.* 19, 321–329.
- Cattenoz, P.B., Popkova, A., Southall, T.D., Aiello, G., Brand, A.H., and Giangrande, A. (2016). Functional Conservation of the Glide/Gcm Regulatory Network Controlling Glia, Hemocyte, and Tendon Cell Differentiation in *Drosophila*. *Genetics* 202, 191–219.
- Cattenoz, P.B., Sakr, R., Pavlidaki, A., Delaporte, C., Riba, A., Molina, N., Hariharan, N., Mukherjee, T., and Giangrande, A. (2020). Temporal specificity and heterogeneity of *Drosophila* immune cells. *EMBO J.* e104486.
- Di Cerbo, V., and Schneider, R. (2013). Cancers with wrong HATs: the impact of acetylation. *Brief. Funct. Genomics* 12, 231–243.
- Chatterjee, S., Angelakos, C.C., Bahl, E., Hawk, J.D., Gaine, M.E., Poplawski, S.G., Schneider-Anthony, A., Yadav, M., Porcari, G.S., Cassel, J.-C., et al. (2020). The CBP KIX domain regulates long-term memory and circadian activity. *BMC Biol.* 18, 155.
- Chia, W., Somers, W.G., and Wang, H. (2008). *Drosophila* neuroblast asymmetric divisions: cell cycle regulators, asymmetric protein localization, and tumorigenesis. *J. Cell Biol.* 180, 267–272.

- Choe, K.-M., Werner, T., Stöven, S., Hultmark, D., and Anderson, K. V (2002). Requirement for a peptidoglycan recognition protein (PGRP) in Relish activation and antibacterial immune responses in *Drosophila*. *Science* 296, 359–362.
- Christensen, J., Agger, K., Cloos, P.A.C., Pasini, D., Rose, S., Sennels, L., Rappsilber, J., Hansen, K.H., Salcini, A.E., and Helin, K. (2007). RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3. *Cell* 128, 1063–1076.
- Church, R.B., and Robertson, F.W. (1966). A biochemical study of the growth of *Drosophila melanogaster*. *J. Exp. Zool.* 162, 337–351.
- Ciapponi, L., Cenci, G., Ducau, J., Flores, C., Johnson-Schlitz, D., Gorski, M.M., Engels, W.R., and Gatti, M. (2004). The *Drosophila* Mre11/Rad50 complex is required to prevent both telomeric fusion and chromosome breakage. *Curr. Biol.* 14, 1360–1366.
- Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304.
- Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* 18, 407–422.
- Cloos, P.A.C., Christensen, J., Agger, K., Maiolica, A., Rappsilber, J., Antal, T., Hansen, K.H., and Helin, K. (2006). The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* 442, 307–311.
- Clough, E., Moon, W., Wang, S., Smith, K., and Hazelrigg, T. (2007). Histone methylation is required for oogenesis in *Drosophila*. *Development* 134, 157–165.
- Clough, E., Tedeschi, T., and Hazelrigg, T. (2014). Epigenetic regulation of oogenesis and germ stem cell maintenance by the *Drosophila* histone methyltransferase Eggless/dSetDB1. *Dev. Biol.* 388, 181–191.
- Collepardo-Guevara, R., Portella, G., Vendruscolo, M., Frenkel, D., Schlick, T., and Orozco, M. (2015). Chromatin Unfolding by Epigenetic Modifications Explained by Dramatic Impairment of Internucleosome Interactions: A Multiscale Computational Study. *J. Am. Chem. Soc.* 137, 10205–10215.
- Colonques, J., Ceron, J., and Tejedor, F.J. (2007). Segregation of postembryonic neuronal and glial lineages inferred from a mosaic analysis of the *Drosophila* larval brain. *Mech. Dev.* 124, 327–340.
- Conaway, R.C., and Conaway, J.W. (1993). GENERAL INITIATION FACTORS FOR RNA POLYMERASE II. *Annu. Rev. Biochem.* 62, 161–190.
- Cooper, G.M. (2000). Regulation of Transcription in Eukaryotes.
- Corbo, J.C., and Levine, M. (1996). Characterization of an immunodeficiency mutant in *Drosophila*. *Mech. Dev.* 55, 211–220.
- Cornman, R.S. (2009). Molecular evolution of *Drosophila* cuticular protein genes. *PLoS One* 4, e8345.
- Corona, D.F., Längst, G., Clapier, C.R., Bonte, E.J., Ferrari, S., Tamkun, J.W., and Becker, P.B. (1999). ISWI is an ATP-dependent nucleosome remodeling factor. *Mol. Cell* 3, 239–245.
- Costa, F.F. (2008). Non-coding RNAs, epigenetics and complexity. *Gene* 410, 9–17.
- Crozatier, M., Ubeda, J.-M., Vincent, A., and Meister, M. (2004). Cellular immune response to parasitization in *Drosophila* requires the EBF orthologue *collier*. *PLoS Biol.* 2, E196.
- Ct, W., and Morris, J.R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* 293, 1103–

1105.

- Cubas, P., de Celis, J.F., Campuzano, S., and Modolell, J. (1991). Proneural clusters of achaete-scute expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes Dev.* *5*, 996–1008.
- Curtis, B.J., Zraly, C.B., Marendza, D.R., and Dingwall, A.K. (2011). Histone lysine demethylases function as co-repressors of SWI/SNF remodeling activities during *Drosophila* wing development. *Dev. Biol.* *350*, 534–547.
- Czermin, B., Schotta, G., Hülsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional association of SU(VAR)3-9 and HDAC1 in *Drosophila*. *EMBO Rep.* *2*, 915–919.
- Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* *111*, 185–196.
- Daban, J.R. (2000). Physical constraints in the condensation of eukaryotic chromosomes. Local concentration of DNA versus linear packing ratio in higher order chromatin structures. *Biochemistry* *39*, 3861–3866.
- Daines, B., Wang, H., Wang, L., Li, Y., Han, Y., Emmert, D., Gelbart, W., Wang, X., Li, W., Gibbs, R., et al. (2011). The *Drosophila melanogaster* transcriptome by paired-end RNA sequencing. *Genome Res.* *21*, 315–324.
- Dallman, J.E., Allopenna, J., Bassett, A., Travers, A., and Mandel, G. (2004). A conserved role but different partners for the transcriptional corepressor CoREST in fly and mammalian nervous system formation. *J. Neurosci.* *24*, 7186–7193.
- Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, L., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., et al. (2018). A Single-Cell Transcriptome Atlas of the Aging *Drosophila* Brain. *Cell* *174*, 982-998.e20.
- Dehé, P.-M., Pamblanco, M., Luciano, P., Lebrun, R., Moinier, D., Sendra, R., Verreault, A., Tordera, V., and Géli, V. (2005). Histone H3 lysine 4 mono-methylation does not require ubiquitination of histone H2B. *J. Mol. Biol.* *353*, 477–484.
- Deshmukh, S., Ponnaluri, V.C., Dai, N., Pradhan, S., and Deobagkar, D. (2018). Levels of DNA cytosine methylation in the *Drosophila* genome. *PeerJ* *6*, e5119.
- Deubzer, H.E., Schier, M.C., Oehme, I., Lodrini, M., Haendler, B., Sommer, A., and Witt, O. (2013). HDAC11 is a novel drug target in carcinomas. *Int. J. Cancer* *132*, 2200–2208.
- Dinger, M.E., Baillie, G.J., and Musgrave, D.R. (2000). Growth phase-dependent expression and degradation of histones in the thermophilic archaeon *Thermococcus zilligii*. *Mol. Microbiol.* *36*, 876–885.
- Doe, C.Q. (1992). Molecular markers for identified neuroblasts and ganglion mother cells in the *Drosophila* central nervous system. *Development* *116*, 855–863.
- Doherty, J., Logan, M.A., Taşdemir, O.E., and Freeman, M.R. (2009). Ensheathing glia function as phagocytes in the adult *Drosophila* brain. *J. Neurosci.* *29*, 4768–4781.
- Dorigo, B., Schalch, T., Bystricky, K., and Richmond, T.J. (2003). Chromatin fiber folding: requirement for the histone H4 N-terminal tail. *J. Mol. Biol.* *327*, 85–96.
- Dover, J., Schneider, J., Tawiah-Boateng, M.A., Wood, A., Dean, K., Johnston, M., and Shilatifard, A.

- (2002). Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J. Biol. Chem.* *277*, 28368–28371.
- Duan, Z., Zarebski, A., Montoya-Durango, D., Grimes, H.L., and Horwitz, M. (2005). Gfi1 coordinates epigenetic repression of p21Cip/WAF1 by recruitment of histone lysine methyltransferase G9a and histone deacetylase 1. *Mol. Cell. Biol.* *25*, 10338–10351.
- Dumstrei, K., Wang, F., Nassif, C., and Hartenstein, V. (2003). Early development of the *Drosophila* brain: V. Pattern of postembryonic neuronal lineages expressing DE-cadherin. *J. Comp. Neurol.* *455*, 451–462.
- Ebens, A.J., Garren, H., Cheyette, B.N., and Zipursky, S.L. (1993). The *Drosophila* anachronism locus: a glycoprotein secreted by glia inhibits neuroblast proliferation. *Cell* *74*, 15–27.
- Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes Dev.* *18*, 2973–2983.
- Egger, B., Chell, J.M., and Brand, A.H. (2008). Insights into neural stem cell biology from flies. *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* *363*, 39–56.
- Eichner, J., Chen, H.T., Warfield, L., and Hahn, S. (2010). Position of the general transcription factor TFIIF within the RNA polymerase II transcription preinitiation complex. *EMBO J.* *29*, 706–716.
- Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev. Biol.* *339*, 240–249.
- Eleftherianos, I., and Revenis, C. (2011). Role and importance of phenoloxidase in insect hemostasis. *J. Innate Immun.* *3*, 28–33.
- Elfring, L.K., Deuring, R., McCallum, C.M., Peterson, C.L., and Tamkun, J.W. (1994). Identification and characterization of *Drosophila* relatives of the yeast transcriptional activator SNF2/SWI2. *Mol. Cell. Biol.* *14*, 2225–2234.
- Endoh, M., Endo, T.A., Endoh, T., Isono, K. ichi, Sharif, J., Ohara, O., Toyoda, T., Ito, T., Eskeland, R., Bickmore, W.A., et al. (2012). Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. *PLoS Genet.* *8*.
- Evans, C.J., and Banerjee, U. (2003). Transcriptional regulation of hematopoiesis in *Drosophila*. *Blood Cells. Mol. Dis.* *30*, 223–228.
- Fei, J., Torigoe, S.E., Brown, C.R., Khuong, M.T., Kassavetis, G.A., Boeger, H., and Kadonaga, J.T. (2015). The prenucleosome, a stable conformational isomer of the nucleosome. *Genes Dev.* *29*, 2563–2575.
- Feller, C., Forné, I., Imhof, A., and Becker, P.B. (2015). Global and specific responses of the histone acetylome to systematic perturbation. *Mol. Cell* *57*, 559–571.
- Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. *Nature* *421*, 448–453.
- Ferreira, R., Ohneda, K., Yamamoto, M., and Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell. Biol.* *25*, 1215–1227.
- Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* *73*, 1897–1901.
- Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in

Drosophila. *Nature* 332, 853–856.

Fischer, V., Plassard, D., Ye, T., Reina-San-Martin, B., Stierle, M., Tora, L., and Devys, D. (2021). The related coactivator complexes SAGA and ATAC control embryonic stem cell self-renewal through acetyltransferase-independent mechanisms. *Cell Rep.* 36, 109598.

Flici, H., Cattenoz, P.B., Komonyi, O., Laneve, P., Erkosar, B., Karatas, O.F., Reichert, H., Berzsenyi, S., and Giangrande, A. (2014). Interlocked loops trigger lineage specification and stable fates in the *Drosophila* nervous system. *Nat. Commun.* 2014 51 5, 1–15.

Forey, R., Barthe, A., Tittel-Elmer, M., Wery, M., Barrault, M.-B., Ducrot, C., Seeber, A., Krietenstein, N., Szachnowski, U., Skrzypczak, M., et al. (2021). A Role for the Mre11-Rad50-Xrs2 Complex in Gene Expression and Chromosome Organization. *Mol. Cell* 81, 183-197.e6.

Fossett, N., and Schulz, R.A. (2001). Functional conservation of hematopoietic factors in *Drosophila* and vertebrates. *Differentiation.* 69, 83–90.

Foulds, C.E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T.L., Malovannaya, A., Hamilton, R.A., Gates, L.A., Zhang, Z., Li, C., et al. (2013). Proteomic analysis of coregulators bound to ER α on DNA and nucleosomes reveals coregulator dynamics. *Mol. Cell* 51, 185–199.

Franc, N.C., Dimarcq, J.L., Lagueux, M., Hoffmann, J., and Ezekowitz, R.A. (1996). Croquemort, a novel *Drosophila* hemocyte/macrophage receptor that recognizes apoptotic cells. *Immunity* 4, 431–443.

Freeman, M.R. (2015). *Drosophila* Central Nervous System Glia. *Cold Spring Harb. Perspect. Biol.* 7.

Freeman, M.R., Delrow, J., Kim, J., Johnson, E., and Doe, C.Q. (2003). Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function. *Neuron* 38, 567–580.

Furdas, S.D., Kannan, S., Sippl, W., and Jung, M. (2012). Small molecule inhibitors of histone acetyltransferases as epigenetic tools and drug candidates. *Arch. Pharm. (Weinheim).* 345, 7–21.

Gao, G., Bi, X., Chen, J., Srikanta, D., and Rong, Y.S. (2009). Mre11-Rad50-Nbs complex is required to cap telomeres during *Drosophila* embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 10728–10733.

Gao, L., Cueto, M.A., Asselbergs, F., and Atadja, P. (2002). Cloning and functional characterization of HDAC11, a novel member of the human histone deacetylase family. *J. Biol. Chem.* 277, 25748–25755.

Garcea, R.L., and Alberts, B.M. (1980). Comparative studies of histone acetylation in nucleosomes, nuclei, and intact cells. Evidence for special factors which modify acetylase action. *J. Biol. Chem.* 255, 11454–11463.

Garces, A., and Thor, S. (2006). Specification of *Drosophila* aCC motoneuron identity by a genetic cascade involving even-skipped, grain and zfh1. *Development* 133, 1445–1455.

Gatei, M., Kijas, A.W., Biard, D., Dörk, T., and Lavin, M.F. (2014). RAD50 phosphorylation promotes ATR downstream signaling and DNA restart following replication stress. *Hum. Mol. Genet.* 23, 4232–4248.

Gates, L.A., Shi, J., Rohira, A.D., Feng, Q., Zhu, B., Bedford, M.T., Sagum, C.A., Jung, S.Y., Qin, J., Tsai, M.-J., et al. (2017). Acetylation on histone H3 lysine 9 mediates a switch from transcription initiation to elongation. *J. Biol. Chem.* 292, 14456–14472.

Gaudet, P., Livstone, M.S., Lewis, S.E., and Thomas, P.D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief. Bioinform.* 12, 449–462.

Di Gennaro, E., Bruzzese, F., Caraglia, M., Abruzzese, A., and Budillon, A. (2004). Acetylation of

- proteins as novel target for antitumor therapy: review article. *Amino Acids* 26, 435–441.
- Gershenzon, N.I., Trifonov, E.N., and Ioshikhes, I.P. (2006). The features of *Drosophila* core promoters revealed by statistical analysis. *BMC Genomics* 7, 161.
- Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.
- Ghysen, A., and Dambly-Chaudiere, C. (1989). Genesis of the *Drosophila* peripheral nervous system. *Trends Genet.* 5, 251–255.
- Gibney, E.R., and Nolan, C.M. (2010). Epigenetics and gene expression. *Heredity (Edinb.)* 105, 4–13.
- Giesen, K., Hummel, T., Stollewerk, A., Harrison, S., Travers, A., and Klämbt, C. (1997). Glial development in the *Drosophila* CNS requires concomitant activation of glial and repression of neuronal differentiation genes. *Development* 124, 2307–2316.
- Gobert, V., Gottar, M., Matskevich, A.A., Rutschmann, S., Royet, J., Belvin, M., Hoffmann, J.A., and Ferrandon, D. (2003). Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. *Science* 302, 2126–2130.
- González-Ramírez, M., Ballaré, C., Mugianesi, F., Beringer, M., Santanach, A., Blanco, E., and Di Croce, L. (2021). Differential contribution to gene expression prediction of histone modifications at enhancers or promoters. *PLoS Comput. Biol.* 17, e1009368.
- Gorski, M.M., Romeijn, R.J., Eeken, J.C.J., de Jong, A.W.M., van Veen, B.L., Szuhai, K., Mullenders, L.H., Ferro, W., and Pastink, A. (2004). Disruption of *Drosophila* Rad50 causes pupal lethality, the accumulation of DNA double-strand breaks and the induction of apoptosis in third instar larvae. *DNA Repair (Amst.)* 3, 603–615.
- Gottschling, D.E., Aparicio, O.M., Billington, B.L., and Zakian, V.A. (1990). Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 63, 751–762.
- Grant, P.A., and Berger, S.L. (1999). Histone acetyltransferase complexes. *Semin. Cell Dev. Biol.* 10, 169–177.
- Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., et al. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev.* 11, 1640–1650.
- Grant, P.A., Eberharter, A., John, S., Cook, R.G., Turner, B.M., and Workman, J.L. (1999). Expanded lysine acetylation specificity of Gcn5 in native complexes. *J. Biol. Chem.* 274, 5895–5900.
- Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. *Curr. Opin. Genet. Dev.* 20, 134–141.
- Grigorian, M., Mandal, L., and Hartenstein, V. (2011). Hematopoiesis at the onset of metamorphosis: terminal differentiation and dissociation of the *Drosophila* lymph gland. *Dev. Genes Evol.* 221, 121–131.
- Guan, D.-L., Ding, R.-R., Hu, X.-Y., Yang, X.-R., Xu, S.-Q., Gu, W., and Zhang, M. (2019). Cadmium-induced genome-wide DNA methylation changes in growth and oxidative metabolism in *Drosophila melanogaster*. *BMC Genomics* 20, 356.
- Guccione, E., Bassi, C., Casadio, F., Martinato, F., Cesaroni, M., Schuchlantz, H., Lüscher, B., and Amati, B. (2007). Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* 449, 933–937.

- Guelman, S., Kozuka, K., Mao, Y., Pham, V., Solloway, M.J., Wang, J., Wu, J., Lill, J.R., and Zha, J. (2009). The double-histone-acetyltransferase complex ATAC is essential for mammalian development. *Mol. Cell. Biol.* *29*, 1176–1188.
- Gupta-Agarwal, S., Franklin, A. V., Deramus, T., Wheelock, M., Davis, R.L., McMahon, L.L., and Lubin, F.D. (2012). G9a/GLP histone lysine dimethyltransferase complex activity in the hippocampus and the entorhinal cortex is required for gene activation and silencing during memory consolidation. *J. Neurosci.* *32*, 5440–5453.
- Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N., and Müller, J. (2012). The role of the histone H2A ubiquitinase Sce in Polycomb repression. *Development* *139*, 117–127.
- Hakim-Mishnaevski, K., Flint-Brodsky, N., Shklyar, B., Levy-Adam, F., and Kurant, E. (2019). Glial Phagocytic Receptors Promote Neuronal Loss in Adult *Drosophila* Brain. *Cell Rep.* *29*, 1438-1448.e3.
- Halsall, J.A., Andrews, S., Krueger, F., Rutledge, C.E., Ficz, G., Reik, W., and Turner, B.M. (2021). Histone modifications form a cell-type-specific chromosomal bar code that persists through the cell cycle. *Sci. Rep.* *11*, 3009.
- Halter, D.A., Urban, J., Rickert, C., Ner, S.S., Ito, K., Travers, A.A., and Technau, G.M. (1995). The homeobox gene repo is required for the differentiation and maintenance of glia function in the embryonic nervous system of *Drosophila melanogaster*. *Development* *121*, 317–332.
- Han, M., and Grunstein, M. (1988). Nucleosome loss activates yeast downstream promoters in vivo. *Cell* *55*, 1137–1145.
- Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* *21*, 396–420.
- Harrison, D.A., Binari, R., Nahreini, T.S., Gilman, M., and Perrimon, N. (1995). Activation of a *Drosophila* Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. *EMBO J.* *14*, 2857–2865.
- Hartenstein, V. (1993). *Atlas of Drosophila Development* (Cold Spring Harbor Laboratory Press).
- Hartenstein, V., and Campos-Ortega, J.A. (1984). Early neurogenesis in wild-type *Drosophila melanogaster*. *Wilhelm Roux's Arch. Dev. Biol.* *193*, 308–325.
- Hartenstein, V., Rudloff, E., and Campos-Ortega, J.A. (1987). The pattern of proliferation of the neuroblasts in the wild-type embryo of *Drosophila melanogaster*. *Roux's Arch. Dev. Biol. Off. Organ EDBO* *196*, 473–485.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* *39*, 311–318.
- Helmlinger, D., Hardy, S., Eberlin, A., Devys, D., and Tora, L. (2006). Both normal and polyglutamine-expanded ataxin-7 are components of TFTC-type GCN5 histone acetyltransferase-containing complexes. *Biochem. Soc. Symp.* 155–163.
- Helmlinger, D., Papai, G., Devys, D., and Tora, L. (2021). What do the structures of GCN5-containing complexes teach us about their function? *Biochim. Biophys. Acta. Gene Regul. Mech.* *1864*.
- Hendzel, M.J., Wei, Y., Mancini, M.A., Van Hooser, A., Ranalli, T., Brinkley, B.R., Bazett-Jones, D.P., and Allis, C.D. (1997). Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic

- chromosome condensation. *Chromosoma* 106, 348–360.
- Herz, H.-M., Madden, L.D., Chen, Z., Bolduc, C., Buff, E., Gupta, R., Davuluri, R., Shilatifard, A., Hariharan, I.K., and Bergmann, A. (2010). The H3K27me3 demethylase dUTX is a suppressor of Notch- and Rb-dependent tumors in *Drosophila*. *Mol. Cell. Biol.* 30, 2485–2497.
- Herz, H.-M., Morgan, M., Gao, X., Jackson, J., Rickels, R., Swanson, S.K., Florens, L., Washburn, M.P., Eissenberg, J.C., and Shilatifard, A. (2014). Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. *Science* 345, 1065–1070.
- von Hilchen, C.M., Bustos, A.E., Giangrande, A., Technau, G.M., and Altenhein, B. (2013). Predetermined embryonic glial cells form the distinct glial sheaths of the *Drosophila* peripheral nervous system. *Development* 140, 3657–3668.
- Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velázquez, I.A., Zenn, H.M., Kost, N., Pohl, W., et al. (2016). Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin. *Nat. Commun.* 7, 11310.
- Hoffmann, J.A. (2003). The immune response of *Drosophila*. *Nature* 426, 33–38.
- Van Holde, K.E., Sahasrabudhe, C.G., and Shaw, B.R. (1974). A model for particulate structure in chromatin. *Nucleic Acids Res.* 1, 1579–1586.
- Holstege, F.C.P., Tantin, D., Carey, M., Van Der Vliet, P.C., and Timmers, H.T.M. (1995). The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.* 14, 810–819.
- Holt, M.T., David, Y., Pollock, S., Tang, Z., Jeon, J., Kim, J., Roeder, R.G., and Muir, T.W. (2015). Identification of a functional hotspot on ubiquitin required for stimulation of methyltransferase activity on chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10365–10370.
- Holz, A., Bossinger, B., Strasser, T., Janning, W., and Klapper, R. (2003). The two origins of hemocytes in *Drosophila*. *Development* 130, 4955–4962.
- Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y., and Cheng, X. (2010). Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat. Struct. Mol. Biol.* 17, 38–43.
- Hosoya, T., Takizawa, K., Nitta, K., and Hotta, Y. (1995). glial cells missing: a binary switch between neuronal and glial determination in *Drosophila*. *Cell* 82, 1025–1036.
- Hou, S.X., Zheng, Z., Chen, X., and Perrimon, N. (2002). The Jak/STAT pathway in model organisms: emerging roles in cell movement. *Dev. Cell* 3, 765–778.
- Hugues, A., Jacobs, C.S., and Roudier, F. (2020). Mitotic Inheritance of PRC2-Mediated Silencing: Mechanistic Insights and Developmental Perspectives. *Front. Plant Sci.* 11, 262.
- Hultmark, D. (2003). *Drosophila* immunity: paths and patterns. *Curr. Opin. Immunol.* 15, 12–19.
- Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* 2018 508 50, 1–14.
- Imler, J.-L., and Bulet, P. (2005). Antimicrobial peptides in *Drosophila*: structures, activities and gene regulation. *Chem. Immunol. Allergy* 86, 1–21.
- Ito, K., and Hotta, Y. (1992). Proliferation pattern of postembryonic neuroblasts in the brain of *Drosophila melanogaster*. *Dev. Biol.* 149, 134–148.

- Ito, A., Kawaguchi, Y., Lai, C.-H., Kovacs, J.J., Higashimoto, Y., Appella, E., and Yao, T.-P. (2002). MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation. *EMBO J.* *21*, 6236–6245.
- Ito, K., Urban, J., and Technau, G.M. (1995). Distribution, classification, and development of *Drosophila* glial cells in the late embryonic and early larval ventral nerve cord. *Roux's Arch. Dev. Biol. Off. Organ EDBO* *204*, 284–307.
- Iwase, S., Lan, F., Bayliss, P., de la Torre-Ubieta, L., Huarte, M., Qi, H.H., Whetstine, J.R., Bonni, A., Roberts, T.M., and Shi, Y. (2007). The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* *128*, 1077–1088.
- Jan, Y.N., and Jan, L.Y. (1998). Asymmetric cell division. *Nature* *392*, 775–778.
- Jang, I.-H., Chosa, N., Kim, S.-H., Nam, H.-J., Lemaitre, B., Ochiai, M., Kambris, Z., Brun, S., Hashimoto, C., Ashida, M., et al. (2006). A Spätzle-processing enzyme required for toll signaling activation in *Drosophila* innate immunity. *Dev. Cell* *10*, 45–55.
- Jenett, A., Rubin, G.M., Ngo, T.-T.B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B.D., Cavallaro, A., Hall, D., Jeter, J., et al. (2012). A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* *2*, 991–1001.
- Jensen, L.R., Amende, M., Gurok, U., Moser, B., Gimmel, V., Tzschach, A., Janecke, A.R., Tariverdian, G., Chelly, J., Fryns, J.-P., et al. (2005). Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. *Am. J. Hum. Genet.* *76*, 227–236.
- Jijakli, H., and Ghysen, A. (1992). Segmental determination in *Drosophila* central nervous system: analysis of the abdominal-A region of the bithorax complex. *Int. J. Dev. Biol.* *36*, 93–99.
- Jiravanichpaisal, P., Lee, B.L., and Söderhäll, K. (2006). Cell-mediated immunity in arthropods: hematopoiesis, coagulation, melanization and opsonization. *Immunobiology* *211*, 213–236.
- Jung, S.-H., Evans, C.J., Uemura, C., and Banerjee, U. (2005). The *Drosophila* lymph gland as a developmental model of hematopoiesis. *Development* *132*, 2521–2533.
- Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* *339*, 225–229.
- Kadamb, R., Mittal, S., Bansal, N., Batra, H., and Saluja, D. (2013). Sin3: insight into its transcription regulatory functions. *Eur. J. Cell Biol.* *92*, 237–246.
- Karlič, R., Chung, H.-R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proc. Natl. Acad. Sci. U. S. A.* *107*, 2926–2931.
- Karouzou, M. V., Spyropoulos, Y., Iconomidou, V.A., Cornman, R.S., Hamodrakas, S.J., and Willis, J.H. (2007). *Drosophila* cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. *Insect Biochem. Mol. Biol.* *37*, 754–760.
- Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* *85*, 8136–8140.
- Khochbin, S., Verdel, A., Lemercier, C., and Seigneurin-Berny, D. (2001). Functional significance of histone deacetylase diversity. *Curr. Opin. Genet. Dev.* *11*, 162–166.
- Kim, T., and Buratowski, S. (2009). Dimethylation of H3K4 by Set1 recruits the Set3 histone deacetylase complex to 5' transcribed regions. *Cell* *137*, 259–272.

- Kim, J., Kim, J.-A., McGinty, R.K., Nguyen, U.T.T., Muir, T.W., Allis, C.D., and Roeder, R.G. (2013). The n-SET domain of Set1 regulates H2B ubiquitylation-dependent H3K4 methylation. *Mol. Cell* *49*, 1121–1133.
- Kimura, A., Matsubara, K., and Horikoshi, M. (2005). A decade of histone acetylation: marking eukaryotic chromosomes with specific codes. *J. Biochem.* *138*, 647–662.
- Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P., and Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell* *133*, 813–828.
- King, I.F.G., Francis, N.J., and Kingston, R.E. (2002). Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol. Cell Biol.* *22*, 7919–7928.
- Kingston, R.E., and Narlikar, G.J. (1999). ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* *13*, 2339–2352.
- Klose, R.J., Yamane, K., Bae, Y., Zhang, D., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature* *442*, 312–316.
- Knoepfler, P.S., and Eisenman, R.N. (1999). Sin meets NuRD and other tails of repression. *Cell* *99*, 447–450.
- Kocks, C., Cho, J.H., Nehme, N., Ulvila, J., Pearson, A.M., Meister, M., Strom, C., Conto, S.L., Hetru, C., Stuart, L.M., et al. (2005). Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in *Drosophila*. *Cell* *123*, 335–346.
- Konsoula, Z., and Barile, F.A. (2012). Epigenetic histone acetylation and deacetylation mechanisms in experimental models of neurodegenerative disorders. *J. Pharmacol. Toxicol. Methods* *66*, 215–220.
- Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science* *184*, 868–871.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* *128*, 693–705.
- Kragol, G., Lovas, S., Varadi, G., Condie, B.A., Hoffmann, R., and Otvos, L.J. (2001). The antibacterial peptide pyrrolicin inhibits the ATPase actions of DnaK and prevents chaperone-assisted protein folding. *Biochemistry* *40*, 3016–3026.
- Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Mol. Cell* *44*, 410–423.
- Kremer, M.C., Jung, C., Batelli, S., Rubin, G.M., and Gaul, U. (2017). The glia of the adult *Drosophila* nervous system. *Glia* *65*, 606–638.
- Krzemien, J., Crozatier, M., and Vincent, A. (2010). Ontogeny of the *Drosophila* larval hematopoietic organ, hemocyte homeostasis and the dedicated cellular immune response to parasitism. *Int. J. Dev. Biol.* *54*, 1117–1125.
- Kumar, S., and Mohapatra, T. (2021). Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression. *Front. Cell Dev. Biol.* *9*, 628415.
- Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017).

- Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol. Cell* 65, 432–446.e5.
- Kuo, M.H., Brownell, J.E., Sobel, R.E., Ranalli, T.A., Cook, R.G., Edmondson, D.G., Roth, S.Y., and Allis, C.D. (1996). Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* 383, 269–272.
- Kurant, E. (2011). Keeping the CNS clear: glial phagocytic functions in *Drosophila*. *Glia* 59, 1304–1311.
- Kurant, E., Axelrod, S., Leaman, D., and Gaul, U. (2008). Six-microns-under acts upstream of Draper in the glial phagocytosis of apoptotic neurons. *Cell* 133, 498–509.
- Kurucz, E., Márkus, R., Zsámboki, J., Folkl-Medzihradzky, K., Darula, Z., Vilmos, P., Udvardy, A., Krausz, I., Lukacsovich, T., Gateff, E., et al. (2007). Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Curr. Biol.* 17, 649–654.
- Lachner, M., O’Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116–120.
- Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* 12, 34–44.
- Lake, C.M., Holsclaw, J.K., Bellendir, S.P., Sekelsky, J., and Hawley, R.S. (2013). The development of a monoclonal antibody recognizing the *Drosophila melanogaster* phosphorylated histone H2A variant (γ -H2AV). *G3 (Bethesda)*. 3, 1539–1543.
- Lakowski, B., Roelens, I., and Jacob, S. (2006). CoREST-like complexes regulate chromatin modification and neuronal gene expression. *J. Mol. Neurosci.* 29, 227–239.
- Landgraf, M., and Thor, S. (2006). Development of *Drosophila* motoneurons: specification and morphology. *Semin. Cell Dev. Biol.* 17, 3–11.
- Laneve, P., Delaporte, C., Trebuchet, G., Komonyi, O., Flici, H., Popkova, A., D’Agostino, G., Taglini, F., Kerekes, I., and Giangrande, A. (2013). The Gcm/Glide molecular and cellular pathway: new actors and new lineages. *Dev. Biol.* 375, 65–78.
- Lanot, R., Zachary, D., Holder, F., and Meister, M. (2001). Postembryonic hematopoiesis in *Drosophila*. *Dev. Biol.* 230, 243–257.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036.
- Laurent, B.C., Treitel, M.A., and Carlson, M. (1991). Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. *Proc. Natl. Acad. Sci. U. S. A.* 88, 2687–2691.
- Laurent, B.C., Treich, I., and Carlson, M. (1993). The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. *Genes Dev.* 7, 583–591.
- Lazo-Gómez, R., Ramírez-Jarquín, U.N., Tovar-Y-Romo, L.B., and Tapia, R. (2013). Histone deacetylases and their role in motor neuron degeneration. *Front. Cell. Neurosci.* 7, 243.
- Lebestky, T., Chang, T., Hartenstein, V., and Banerjee, U. (2000). Specification of *Drosophila* hematopoietic lineage by conserved transcription factors. *Science* 288, 146–149.
- Lee, J.H., Hart, S.R.L., and Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic

stem cell differentiation. *Genesis* 38, 32–38.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhatar, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432–435.

Lee, M.G., Villa, R., Trojer, P., Norman, J., Yan, K.-P., Reinberg, D., Di Croce, L., and Shiekhatar, R. (2007). Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* 318, 447–450.

Lemaitre, B., Kromer-Metzger, E., Michaut, L., Nicolas, E., Meister, M., Georgel, P., Reichhart, J.M., and Hoffmann, J.A. (1995). A recessive mutation, immune deficiency (*imd*), defines two distinct control pathways in the *Drosophila* host defense. *Proc. Natl. Acad. Sci. U. S. A.* 92, 9465–9469.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J.M., and Hoffmann, J.A. (1996). The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* 86, 973–983.

LERNER, A.B., and FITZPATRICK, T.B. (1950). Biochemistry of melanin formation. *Physiol. Rev.* 30, 91–126.

LEWIS, E.B. (1950). The phenomenon of position effect. *Adv. Genet.* 3, 73–115.

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D.J., Quake, S.R., and Luo, L. (2017). Classifying *Drosophila* Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. *Cell* 171, 1206–1220.e22.

Li, M.-Y., Zhu, M., Linghu, E.-Q., Feng, F., Zhu, B., Wu, C., and Guo, M.-Z. (2016). Interleukin-13 suppresses interleukin-10 via inhibiting A20 in peripheral B cells of patients with food allergy. *Oncotarget* 7, 79914–79924.

Lian, T., Gaur, U., Wu, Q.I., Tu, J., Sun, B., Yang, D., Fan, X., Mao, X., and Yang, M. (2018). DNA methylation is not involved in dietary restriction induced lifespan extension in adult *Drosophila*. *Genet. Res. (Camb)*. 100, e1.

Limmer, S., Weiler, A., Volkenhoff, A., Babatz, F., and Klambt, C. (2014). The *Drosophila* blood-brain barrier: development and function of a glial endothelium. *Front. Neurosci.* 8, 365.

Lin, C.-H., Li, B., Swanson, S., Zhang, Y., Florens, L., Washburn, M.P., Abmayr, S.M., and Workman, J.L. (2008). Heterochromatin protein 1a stimulates histone H3 lysine 36 demethylation by the *Drosophila* KDM4A demethylase. *Mol. Cell* 32, 696–706.

Lin, R.J., Nagy, L., Inoue, S., Shao, W., Miller, W.H.J., and Evans, R.M. (1998). Role of the histone deacetylase complex in acute promyelocytic leukaemia. *Nature* 391, 811–814.

Liu, N., Zhang, Z., Wu, H., Jiang, Y., Meng, L., Xiong, J., Zhao, Z., Zhou, X., Li, J., Li, H., et al. (2015). Recognition of H3K9 methylation by GLP is required for efficient establishment of H3K9 methylation, rapid target gene repression, and mouse viability. *Genes Dev.* 29, 379–393.

Loenarz, C., Ge, W., Coleman, M.L., Rose, N.R., Cooper, C.D.O., Klose, R.J., Ratcliffe, P.J., and Schofield, C.J. (2010). PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an N-epsilon-dimethyl lysine demethylase. *Hum. Mol. Genet.* 19, 217–222.

Luger, K., and Richmond, T.J. (1998). The histone tails of the nucleosome. *Curr. Opin. Genet. Dev.* 8, 140–146.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.

- Lusser, A., Urwin, D.L., and Kadonaga, J.T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* *12*, 160–166.
- MacDonald, J.M., Beach, M.G., Porpiglia, E., Sheehan, A.E., Watts, R.J., and Freeman, M.R. (2006). The *Drosophila* cell corpse engulfment receptor Draper mediates glial clearance of severed axons. *Neuron* *50*, 869–881.
- Maconochie, M., Nonchev, S., Morrison, A., and Krumlauf, R. (1996). Paralogous Hox genes: function and regulation. *Annu. Rev. Genet.* *30*, 529–556.
- Majid, T., Griffin, D., Criss, Z. 2nd, Jarpe, M., and Pautler, R.G. (2015). Pharmacologic treatment with histone deacetylase 6 inhibitor (ACY-738) recovers Alzheimer’s disease phenotype in amyloid precursor protein/presenilin 1 (APP/PS1) mice. *Alzheimer’s Dement. (New York, N. Y.)* *1*, 170–181.
- Makhijani, K., Alexander, B., Tanaka, T., Rulifson, E., and Brückner, K. (2011). The peripheral nervous system supports blood cell homing and survival in the *Drosophila* larva. *Development* *138*, 5379–5391.
- Manaka, J., Kuraishi, T., Shiratsuchi, A., Nakai, Y., Higashida, H., Henson, P., and Nakanishi, Y. (2004). Draper-mediated and phosphatidylserine-independent phagocytosis of apoptotic cells by *Drosophila* hemocytes/macrophages. *J. Biol. Chem.* *279*, 48466–48476.
- Mangahas, P.M., and Zhou, Z. (2005). Clearance of apoptotic cells in *Caenorhabditis elegans*. *Semin. Cell Dev. Biol.* *16*, 295–306.
- Marks, P., Rifkind, R.A., Richon, V.M., Breslow, R., Miller, T., and Kelly, W.K. (2001). Histone deacetylases and cancer: causes and therapies. *Nat. Rev. Cancer* *1*, 194–202.
- Marmorstein, R., and Zhou, M.-M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb. Perspect. Biol.* *6*, a018762.
- Martín-Bermudo, M.D., Martínez, C., Rodríguez, A., and Jiménez, F. (1991). Distribution and function of the lethal of scute gene product during early neurogenesis in *Drosophila*. *Development* *113*, 445–454.
- Martinek, N., Shahab, J., Saathoff, M., and Ringuette, M. (2008). Haemocyte-derived SPARC is required for collagen-IV-dependent stability of basal laminae in *Drosophila* embryos. *J. Cell Sci.* *121*, 1671–1680.
- Martínez-Balbás, M.A., Bauer, U.M., Nielsen, S.J., Brehm, A., and Kouzarides, T. (2000). Regulation of E2F1 activity by acetylation. *EMBO J.* *19*, 662–671.
- McGinty, R.K., Kim, J., Chatterjee, C., Roeder, R.G., and Muir, T.W. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* *453*, 812–816.
- Melcarne, C., Lemaitre, B., and Kurant, E. (2019). Phagocytosis in *Drosophila*: From molecules and cellular machinery to physiology. *Insect Biochem. Mol. Biol.* *109*, 1–12.
- Menne, T. V. and Klämbt, C. (1994). The formation of commissures in the *Drosophila* CNS depends on the midline cells and on the Notch gene. *Development* *120*, 123–133.
- Mersfelder, E.L., and Parthun, M.R. (2008). Involvement of Hat1p (Kat1p) catalytic activity and subcellular localization in telomeric silencing. *J. Biol. Chem.* *283*, 29060–29068.
- Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., and Thomas, P.D. (2010). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* *38*.
- Mi, W., Guan, H., Lyu, J., Zhao, D., Xi, Y., Jiang, S., Andrews, F.H., Wang, X., Gagea, M., Wen, H., et al. (2017). YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nat.*

Commun. 8, 1088.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.

Miller, A.A., Bernardoni, R., and Giangrande, A. (1998). Positive autoregulation of the glial promoting factor glide/gcm. *EMBO J.* 17, 6316–6326.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12902–12907.

Minucci, S., Nervi, C., Lo Coco, F., and Pelicci, P.G. (2001). Histone deacetylases: a common molecular target for differentiation treatment of acute myeloid leukemias? *Oncogene* 20, 3110–3115.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303, 343–348.

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature* 457, 413–420.

Moss, T.J., and Wallrath, L.L. (2007). Connections between epigenetic gene silencing and human disease. *Mutat. Res.* 618, 163–174.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* 111, 197–208.

Muratoglu, S., Georgieva, S., Pápai, G., Scheer, E., Enünlü, I., Komonyi, O., Cserpán, I., Lebedeva, L., Nabirochkina, E., Udvardy, A., et al. (2003). Two different Drosophila ADA2 homologues are present in distinct GCN5 histone acetyltransferase-containing complexes. *Mol. Cell. Biol.* 23, 306–321.

Murawsky, C.M., Brehm, A., Badenhorst, P., Lowe, N., Becker, P.B., and Travers, A.A. (2001). Tramtrack69 interacts with the dMi-2 subunit of the Drosophila NuRD chromatin remodelling complex. *EMBO Rep.* 2, 1089–1094.

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

Mustachio, L.M., Roszik, J., Farria, A., and Dent, S.Y.R. (2020). Targeting the SAGA and ATAC Transcriptional Coactivator Complexes in MYC-Driven Cancers. *Cancer Res.* 80, 1905–1911.

Naba, A., Clauser, K.R., Ding, H., Whittaker, C.A., Carr, S.A., and Hynes, R.O. (2016). The extracellular matrix: Tools and insights for the “omics” era. *Matrix Biol.* 49, 10–24.

Nagy, Z., and Tora, L. (2007). Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene* 26, 5341–5357.

Nagy, Z., Riss, A., Romier, C., le Guezennec, X., Dongre, A.R., Orpinell, M., Han, J., Stunnenberg, H., and Tora, L. (2009). The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. *Mol. Cell. Biol.* 29, 1649–1660.

Nakagawa, M., and Kitabayashi, I. (2018). Oncogenic roles of enhancer of zeste homolog 1/2 in hematological malignancies. *Cancer Sci.* 109, 2342–2348.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9

- methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110–113.
- Nambu, J.R., Lewis, J.O., Wharton, K.A.J., and Crews, S.T. (1991). The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* 67, 1157–1167.
- Narayanan, A., Ruyechan, W.T., and Kristie, T.M. (2007). The coactivator host cell factor-1 mediates Set1 and MLL1 H3K4 trimethylation at herpesvirus immediate early promoters for initiation of infection. *Proc. Natl. Acad. Sci. U. S. A.* 104, 10835–10840.
- Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 154, 490–503.
- Neault, M., Mallette, F.A., Vogel, G., Michaud-Levesque, J., and Richard, S. (2012). Ablation of PRMT6 reveals a role as a negative transcriptional regulator of the p53 tumor suppressor. *Nucleic Acids Res.* 40, 9513–9521.
- Neigeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* 108, 845–858.
- Nekrasov, M., Wild, B., and Müller, J. (2005). Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* 6, 348–353.
- Neumann, H., Kotter, M.R., and Franklin, R.J.M. (2009). Debris clearance by microglia: an essential link between degeneration and regeneration. *Brain* 132, 288–295.
- Neuwald, A.F., and Landsman, D. (1997). GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends Biochem. Sci.* 22, 154–155.
- Nishida, M., Kato, M., Kato, Y., Sasai, N., Ueda, J., Tachibana, M., Shinkai, Y., and Yamaguchi, M. (2007). Identification of ZNF200 as a novel binding partner of histone H3 methyltransferase G9a. *Genes Cells* 12, 877–888.
- Nitsch, S., Zorro Shahidian, L., and Schneider, R. (2021). Histone acylations and chromatin dynamics: concepts, challenges, and links to metabolism. *EMBO Rep.* 22, e52774.
- Ogawa, H., Ishiguro, K.-I., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* 296, 1132–1136.
- Okada, Y., Scott, G., Ray, M.K., Mishina, Y., and Zhang, Y. (2007). Histone demethylase JHDM2A is critical for Tnp1 and Prm1 transcription and spermatogenesis. *Nature* 450, 119–123.
- Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). *Science* 183, 330–332.
- Olsson, I., Bergh, G., Ehinger, M., and Gullberg, U. (1996). Cell differentiation in acute myeloid leukemia. *Eur. J. Haematol.* 57, 1–16.
- Omoto, J.J., Lovick, J.K., and Hartenstein, V. (2016). Origins of glial cell populations in the insect nervous system. *Curr. Opin. Insect Sci.* 18, 96–104.
- Ong, C.-T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* 12, 283–293.
- Ou, J., He, Y., Xiao, X., Yu, T.-M., Chen, C., Gao, Z., and Ho, M.S. (2014). Glial cells in neuronal development: recent advances and insights from *Drosophila melanogaster*. *Neurosci. Bull.* 30, 584–594.
- Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence

that chromatin structure is a repeating unit. *Cell* 4, 281–300.

Ouzounis, C.A., and Kyrpides, N.C. (1996). Parallel origins of the nucleosome core and eukaryotic transcription from Archaea. *J. Mol. Evol.* 42, 234–239.

P, S., and AH, B. (2018). Systemic and local cues drive neural stem cell niche remodeling during neurogenesis in *Drosophila*. *Elife* 7.

Palhan, V.B., Chen, S., Peng, G.-H., Tjernberg, A., Gamper, A.M., Fan, Y., Chait, B.T., La Spada, A.R., and Roeder, R.G. (2005). Polyglutamine-expanded ataxin-7 inhibits STAGA histone acetyltransferase activity to produce retinal degeneration. *Proc. Natl. Acad. Sci. U. S. A.* 102, 8472–8477.

Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S.M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell* 75, 543–555.

Pankotai, T., Komonyi, O., Bodai, L., Ujfaludi, Z., Muratoglu, S., Ciurciu, A., Tora, L., Szabad, J., and Boros, I. (2005). The homologous *Drosophila* transcriptional adaptors ADA2a and ADA2b are both required for normal development but have different functions. *Mol. Cell. Biol.* 25, 8215–8227.

Park, J.-S., Jeon, H.-J., Pyo, J.-H., Kim, Y.-S., and Yoo, M.-A. (2018). Deficiency in DNA damage response of enterocytes accelerates intestinal stem cell aging in *Drosophila*. *Aging (Albany, NY)*. 10, 322–338.

Pearson, A., Lux, A., and Krieger, M. (1995). Expression cloning of dSR-CI, a class C macrophage-specific scavenger receptor from *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 4056–4060.

Peco, E., Davla, S., Camp, D., Stacey, S.M., Landgraf, M., and van Meyel, D.J. (2016). *Drosophila* astrocytes cover specific territories of the CNS neuropil and are instructed to differentiate by Prospero, a key effector of Notch. *Development* 143, 1170–1181.

Pereanu, W., Shy, D., and Hartenstein, V. (2005). Morphogenesis and proliferation of the larval brain glia in *Drosophila*. *Dev. Biol.* 283, 191–203.

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* 12, 1577–1589.

Peterson, C.L., and Herskowitz, I. (1992). Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell* 68, 573–583.

Petruk, S., Sedkov, Y., Johnston, D.M., Hodgson, J.W., Black, K.L., Kovermann, S.K., Beck, S., Canaani, E., Brock, H.W., and Mazo, A. (2012). TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. *Cell* 150, 922–933.

Pile, L.A., and Wassarman, D.A. (2000). Chromosomal localization links the SIN3-RPD3 complex to the regulation of chromatin condensation, histone acetylation and gene expression. *EMBO J.* 19, 6131–6140.

Pillus, L. (2008). MYSTs mark chromatin for chromosomal functions. *Curr. Opin. Cell Biol.* 20, 326–333.

Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol. Cell* 55, 5–14.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* 300, 131–135.

- Popkova, A., Bernardoni, R., Diebold, C., Van de Bor, V., Schuettengruber, B., González, I., Busturia, A., Cavalli, G., and Giangrande, A. (2012). Polycomb controls gliogenesis by regulating the transient expression of the Gcm/Glide fate determinant. *PLoS Genet.* *8*, e1003159.
- Popova, E.Y., Barnstable, C.J., and Zhang, S.S.-M. (2014). Cell type-specific epigenetic signatures accompany late stages of mouse retina development. *Adv. Exp. Med. Biol.* *801*, 3–8.
- Prokop, A., and Technau, G.M. (1991). The origin of postembryonic neuroblasts in the ventral nerve cord of *Drosophila melanogaster*. *Development* *111*, 79–88.
- Putzer, H., and Laalami, S. (2013). Regulation of the Expression of Aminoacyl-tRNA Synthetases and Translation Factors. (Austin (TX): andes Bioscience).
- Qi, D., Larsson, J., and Mannervik, M. (2004). *Drosophila* Ada2b is required for viability and normal histone H3 acetylation. *Mol. Cell. Biol.* *24*, 8080–8089.
- Qin, X., Ahn, S., Speed, T.P., and Rubin, G.M. (2007). Global analyses of mRNA translational control during early *Drosophila* embryogenesis. *Genome Biol.* *8*, 1–18.
- Rahnamaeian, M., Cytryńska, M., Zdybicka-Barabas, A., Dobszlaff, K., Wiesner, J., Twyman, R.M., Zuchner, T., Sadd, B.M., Regoes, R.R., Schmid-Hempel, P., et al. (2015). Insect antimicrobial peptides show potentiating functional interactions against Gram-negative bacteria. *Proceedings. Biol. Sci.* *282*, 20150293.
- Rämet, M., Pearson, A., Manfruelli, P., Li, X., Koziel, H., Göbel, V., Chung, E., Krieger, M., and Ezekowitz, R.A. (2001). *Drosophila* scavenger receptor CI is a pattern recognition receptor for bacteria. *Immunity* *15*, 1027–1038.
- Rämet, M., Manfruelli, P., Pearson, A., Mathey-Prevo, B., and Ezekowitz, R.A.B. (2002). Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. *Nature* *416*, 644–648.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44*, W160-5.
- Rando, O.J. (2012). Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Curr. Opin. Genet. Dev.* *22*, 148–155.
- Rao, R.C., and Dou, Y. (2015). Hijacked in cancer: the MLL/KMT2 family of methyltransferases. *Nat. Rev. Cancer* *15*, 334.
- Rea, S., Eisenhaber, F., O’Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., et al. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* *406*, 593–599.
- Read, D.F., Cook, K., Lu, Y.Y., Le Roch, K.G., and Noble, W.S. (2019). Predicting gene expression in the human malaria parasite *Plasmodium falciparum* using histone modification, nucleosome positioning, and 3D localization features. *PLoS Comput. Biol.* *15*, e1007329.
- Reddy, B.A., Bajpe, P.K., Bassett, A., Moshkin, Y.M., Kozhevnikova, E., Bezstarosti, K., Demmers, J.A.A., Travers, A.A., and Verrijzer, C.P. (2010). *Drosophila* Transcription Factor Tramtrack69 Binds MEP1 To Recruit the Chromatin Remodeler NuRD. *Mol. Cell. Biol.* *30*, 5234.
- Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., and Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.* *12*, 162–169.

- Rehorn, K.P., Thelen, H., Michelson, A.M., and Reuter, R. (1996). A molecular aspect of hematopoiesis and endoderm development common to vertebrates and *Drosophila*. *Development* *122*, 4023–4031.
- Reinberg, D., and Vales, L.D. (2018). Chromatin domains rich in inheritance. *Science* *361*, 33–34.
- Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* *43*, 73–81.
- Ren, C., Morohashi, K., Plotnikov, A.N., Jakoncic, J., Smith, S.G., Li, J., Zeng, L., Rodriguez, Y., Stojanoff, V., Walsh, M., et al. (2015). Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* *22*, 161–168.
- Rice, J.C., Briggs, S.D., Ueberheide, B., Barber, C.M., Shabanowitz, J., Hunt, D.F., Shinkai, Y., and Allis, C.D. (2003). Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol. Cell* *12*, 1591–1598.
- Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* *38*, 413–443.
- Riss, A., Scheer, E., Joint, M., Trowitzsch, S., Berger, I., and Tora, L. (2015). Subunits of ADA-two-A-containing (ATAC) or Spt-Ada-Gcn5-acetyltransferase (SAGA) Coactivator Complexes Enhance the Acetyltransferase Activity of GCN5. *J. Biol. Chem.* *290*, 28997–29009.
- Rizki, T.M., and Rizki, R.M. (1978). Larval adipose tissue of homoecotic bithorax mutants of *Drosophila*. *Dev. Biol.* *65*, 476–482.
- Rizki, T.M., Rizki, R.M., and Bellotti, R.A. (1985). Genetics of a *Drosophila* phenoloxidase. *Mol. Gen. Genet.* *201*, 7–13.
- Robinow, S., and White, K. (1991). Characterization and spatial distribution of the ELAV protein during *Drosophila melanogaster* development. *J. Neurobiol.* *22*, 443–461.
- Roddie, H.G., Armitage, E.L., Coates, J.A., Johnston, S.A., and Evans, I.R. (2019). Simu-dependent clearance of dying cells regulates macrophage function and inflammation resolution. *PLoS Biol.* *17*, e2006741.
- Roeder, R.G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* *21*, 327–335.
- Rogulja-Ortmann, A., Lürer, K., Seibert, J., Rickert, C., and Technau, G.M. (2007). Programmed cell death in the embryonic central nervous system of *Drosophila melanogaster*. *Development* *134*, 105–116.
- de Rooij, J.D.E., Hollink, I.H.I.M., Arentsen-Peters, S.T.C.J.M., van Galen, J.F., Berna Beverloo, H., Baruchel, A., Trka, J., Reinhardt, D., Sonneveld, E., Zimmermann, M., et al. (2013). NUP98/JARID1A is a novel recurrent abnormality in pediatric acute megakaryoblastic leukemia with a distinct HOX gene expression pattern. *Leukemia* *27*, 2280–2288.
- Rossaert, E., Pollari, E., Jaspers, T., Van Helleputte, L., Jarpe, M., Van Damme, P., De Bock, K., Moisse, M., and Van Den Bosch, L. (2019). Restoration of histone acetylation ameliorates disease and metabolic abnormalities in a FUS mouse model. *Acta Neuropathol. Commun.* *7*, 107.
- Rössl, A., Denoncourt, A., Lin, M.-S., and Downey, M. (2019). A synthetic non-histone substrate to study substrate targeting by the Gcn5 HAT and sirtuin HDACs. *J. Biol. Chem.* *294*, 6227–6239.
- Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* *70*, 81–120.

- Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schäfer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., et al. (2007). Heterochromatin formation in *Drosophila* is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol. Cell* 26, 103–115.
- Ruhl, D.D., Jin, J., Cai, Y., Swanson, S., Florens, L., Washburn, M.P., Conaway, R.C., Conaway, J.W., and Chrivia, J.C. (2006). Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes. *Biochemistry* 45, 5671–5677.
- Ruiz-Gómez, M., and Ghysen, A. (1993). The expression and role of a proneural gene, *achaete*, in the development of the larval nervous system of *Drosophila*. *EMBO J.* 12, 1121–1130.
- Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* 25, 15–30.
- Rutschmann, S., Jung, A.C., Hetru, C., Reichhart, J.M., Hoffmann, J.A., and Ferrandon, D. (2000). The Rel protein DIF mediates the antifungal but not the antibacterial host defense in *Drosophila*. *Immunity* 12, 569–580.
- Ryu, H., Lee, J., Hagerty, S.W., Soh, B.Y., McAlpin, S.E., Cormier, K.A., Smith, K.M., and Ferrante, R.J. (2006). ESET/SETDB1 gene expression and histone H3 (K9) trimethylation in Huntington's disease. *Proc. Natl. Acad. Sci. U. S. A.* 103, 19176–19181.
- Saha, R.N., and Pahan, K. (2006). HATs and HDACs in neurodegeneration: a tale of disconcerted acetylation homeostasis. *Cell Death Differ.* 13, 539–550.
- Sahakian, E., Powers, J.J., Chen, J., Deng, S.L., Cheng, F., Distler, A., Woods, D.M., Rock-Klotz, J., Sodre, A.L., Youn, J.-I., et al. (2015). Histone deacetylase 11: A novel epigenetic regulator of myeloid derived suppressor cell expansion and function. *Mol. Immunol.* 63, 579–585.
- Sancer, G., and Wernet, M.F. (2021). The development and function of neuronal subtypes processing color and skylight polarization in the optic lobes of *Drosophila melanogaster*. *Arthropod Struct. Dev.* 61, 101012.
- Sandman, K., Pereira, S.L., and Reeve, J.N. (1998). Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. *Cell. Mol. Life Sci.* 54, 1350–1364.
- Sandmann, T., Jakobsen, J.S., and Furlong, E.E.M. (2006). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nat. Protoc.* 1, 2839–2855.
- Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411.
- Sapountzi, V., Logan, I.R., and Robson, C.N. (2006). Cellular functions of TIP60. *Int. J. Biochem. Cell Biol.* 38, 1496–1509.
- Schirmeier, S., Matzat, T., and Klämbt, C. (2016). Axon ensheathment and metabolic supply by glial cells in *Drosophila*. *Brain Res.* 1641, 122–129.
- Schmid, K.J., and Tautz, D. (1997). A screen for fast evolving genes from *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 9746–9750.
- Schmid, A., Chiba, A., and Doe, C.Q. (1999). Clonal analysis of *Drosophila* embryonic neuroblasts: neural cell types, axon projections and muscle targets. *Development* 126, 4653–4689.
- Schotta, G., Ebert, A., and Reuter, G. (2003). SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. *Genetica* 117, 149–158.

- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735–745.
- Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* 8, 9–22.
- Schwartz, Y.B., and Pirrotta, V. (2013). A new world of Polycombs: unexpected partnerships and emerging functions. *Nat. Rev. Genet.* 14, 853–864.
- Sedkov, Y., Cho, E., Petruk, S., Cherbas, L., Smith, S.T., Jones, R.S., Cherbas, P., Canaani, E., Jaynes, J.B., and Mazo, A. (2003). Methylation at lysine 4 of histone H3 in ecdysone-dependent development of *Drosophila*. *Nature* 426, 78–83.
- Serizier, S.B., and McCall, K. (2017). Scrambled Eggs: Apoptotic Cell Clearance by Non-Professional Phagocytes in the *Drosophila* Ovary. *Front. Immunol.* 8, 1642.
- Seto, E., and Yoshida, M. (2014). Erasers of histone acetylation: the histone deacetylase enzymes. *Cold Spring Harb. Perspect. Biol.* 6, a018713.
- Seum, C., Bontron, S., Reo, E., Delattre, M., and Spierer, P. (2007). *Drosophila* G9a is a nonessential gene. *Genetics* 177, 1955–1957.
- Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98, 37–46.
- Shepherd, D. (2000). Glial dependent survival of neurons in *Drosophila*. *Bioessays* 22, 407–409.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953.
- Shikama, N., Lutz, W., Kretzschmar, R., Sauter, N., Roth, J.-F., Marino, S., Wittwer, J., Scheidweiler, A., and Eckner, R. (2003). Essential function of p300 acetyltransferase activity in heart, lung and small intestine formation. *EMBO J.* 22, 5175–5185.
- Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* 81, 65–95.
- Shimaji, K., Tanaka, R., Maeda, T., Ozaki, M., Yoshida, H., Ohkawa, Y., Sato, T., Suyama, M., and Yamaguchi, M. (2017). Histone methyltransferase G9a is a key regulator of the starvation-induced behaviors in *Drosophila melanogaster*. *Sci. Rep.* 7.
- Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* 311, 844–847.
- Shukla, S., and Tekwani, B.L. (2020). Histone Deacetylases Inhibitors in Neurodegenerative Diseases, Neuroprotection and Neuronal Differentiation. *Front. Pharmacol.* 11, 537.
- Simon, R.P., Robaa, D., Alhalabi, Z., Sippl, W., and Jung, M. (2016). KATching-Up on Small Molecule Modulators of Lysine Acetyltransferases. *J. Med. Chem.* 59, 1249–1270.
- Simpson, P. (1997). Notch signalling in development: on equivalence groups and asymmetric developmental potential. *Curr. Opin. Genet. Dev.* 7, 537–542.
- Sims, R.J. 3rd, Chen, C.-F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* 280, 41789–41792.

- Skeath, J.B., and Carroll, S.B. (1992). Regulation of proneural gene expression and cell fate during neuroblast segregation in the *Drosophila* embryo. *Development* *114*, 939–946.
- Skeath, J.B., and Carroll, S.B. (1994). The achaete-scute complex: generation of cellular pattern and fate within the *Drosophila* nervous system. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* *8*, 714–721.
- Skene, P.J., and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* *6*.
- Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* *13*, 1006–1019.
- Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu. Rev. Biochem.* *72*, 449–479.
- Sonnenfeld, M.J., and Jacobs, J.R. (1995). Macrophages and glia participate in the removal of apoptotic neurons from the *Drosophila* embryonic nervous system. *J. Comp. Neurol.* *359*, 644–652.
- Sorrentino, R.P., Carton, Y., and Govind, S. (2002). Cellular immune response to parasite infection in the *Drosophila* lymph gland is developmentally regulated. *Dev. Biol.* *243*, 65–80.
- Sousa-Nunes, R., Yee, L.L., and Gould, A.P. (2011). Fat cells reactivate quiescent neuroblasts via TOR and glial insulin relays in *Drosophila*. *Nature* *471*, 508–513.
- Sparmann, A., and van Lohuizen, M. (2006). Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* *6*, 846–856.
- Spedale, G., Timmers, H.T.M., and Pijnappel, W.W.M.P. (2012). ATAC-king the complexity of SAGA during evolution. *Genes Dev.* *26*, 527–541.
- Srinageshwar, B., Maiti, P., Dunbar, G.L., and Rossignol, J. (2016). Role of Epigenetics in Stem Cell Proliferation and Differentiation: Implications for Treating Neurodegenerative Diseases. *Int. J. Mol. Sci.* *17*.
- Di Stefano, L., Walker, J.A., Burgio, G., Corona, D.F. V, Mulligan, P., Näär, A.M., and Dyson, N.J. (2011). Functional antagonism between histone H3K4 demethylases in vivo. *Genes Dev.* *25*, 17–28.
- Stegmaier, P., Kel, A.E., and Wingender, E. (2004). Systematic DNA-binding domain classification of transcription factors. *Genome Inform.* *15*, 276–286.
- Stofanko, M., Kwon, S.Y., and Badenhorst, P. (2010). Lineage Tracing of Lamellocytes Demonstrates *Drosophila* Macrophage Plasticity. *PLoS One* *5*, e14051.
- Stork, T., Engelen, D., Krudewig, A., Silies, M., Bainton, R.J., and Klambt, C. (2008). Organization and function of the blood-brain barrier in *Drosophila*. *J. Neurosci.* *28*, 587–597.
- Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. *Nature* *403*, 41–45.
- Sugai, F., Yamamoto, Y., Miyaguchi, K., Zhou, Z., Sumi, H., Hamasaki, T., Goto, M., and Sakoda, S. (2004). Benefit of valproic acid in suppressing disease progression of ALS model mice. *Eur. J. Neurosci.* *20*, 3179–3183.
- Suganuma, T., and Workman, J.L. (2008). Crosstalk among Histone Modifications. *Cell* *135*, 604–607.
- Sun, Z.-W., and Allis, C.D. (2002). Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* *418*, 104–108.
- Sures, I., and Gallwitz, D. (1980). Histone-specific acetyltransferases from calf thymus. Isolation,

- properties, and substrate specificity of three different enzymes. *Biochemistry* 19, 943–951.
- Swaminathan, A., and Pile, L.A. (2010). Regulation of cell proliferation and wing development by *Drosophila* SIN3 and String. *Mech. Dev.* 127, 96–106.
- Syed, A., and Tainer, J.A. (2018). The MRE11-RAD50-NBS1 Complex Conducts the Orchestration of Damage Signaling and Outcomes to Stress in DNA Replication and Repair. *Annu. Rev. Biochem.* 87, 263–294.
- Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., et al. (2002). G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* 16, 1779–1791.
- Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S.-J., Pachter, L., Martin, D.I.K., and Boffelli, D. (2014). Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res.* 24, 821–830.
- Tamaru, H. (2010). Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev.* 24, 1465–1478.
- Tamkun, J.W. (1995). The role of brahma and related proteins in transcription and development. *Curr. Opin. Genet. Dev.* 5, 473–477.
- Tanaka, Y., Katagiri, Z.-I., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. *Gene* 397, 161–168.
- Tang, Z., Chen, W.-Y., Shimada, M., Nguyen, U.T.T., Kim, J., Sun, X.-J., Sengoku, T., McGinty, R.K., Fernandez, J.P., Muir, T.W., et al. (2013). SET1 and p300 act synergistically, through coupled histone modifications, in transcriptional activation by p53. *Cell* 154, 297–310.
- Tasdemir-Yilmaz, O.E., and Freeman, M.R. (2014). Astrocytes engage unique molecular programs to engulf pruned neuronal debris from distinct subsets of neurons. *Genes Dev.* 28, 20–33.
- Taunton, J., Hassig, C.A., and Schreiber, S.L. (1996). A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science* 272, 408–411.
- Tepass, U., Fessler, L.I., Aziz, A., and Hartenstein, V. (1994). Embryonic origin of hemocytes and their relationship to cell death in *Drosophila*. *Development* 120, 1829–1837.
- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* 13, 2129–2141.
- Tirode, F., Busso, D., Coin, F., and Egly, J.M. (1999). Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol. Cell* 3, 87–95.
- Tissenbaum, H.A., and Guarente, L. (2001). Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature* 410, 227–230.
- Torigoe, S.E., Urwin, D.L., Ishii, H., Smith, D.E., and Kadonaga, J.T. (2011). Identification of a rapidly formed nonnucleosomal histone-DNA intermediate that is converted into chromatin by ACF. *Mol. Cell* 43, 638–648.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Trébuchet, G., Cattenoz, P.B., Zsámboki, J., Mazaud, D., Siekhaus, D.E., Fanto, M., and Giangrande, A.

- (2019). The Repo Homeodomain Transcription Factor Suppresses Hematopoiesis in *Drosophila* and Preserves the Glial Fate. *J. Neurosci.* *39*, 238.
- Tremblay, M., Sanchez-Ferras, O., and Bouchard, M. (2018). GATA transcription factors in development and disease. *Development* *145*.
- Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* *128*, 651–654.
- Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? *Mol. Cell* *28*, 1–13.
- Truman, J.W. (1990). Metamorphosis of the central nervous system of *Drosophila*. *J. Neurobiol.* *21*, 1072–1084.
- Truman, J.W., and Bate, M. (1988). Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila melanogaster*. *Dev. Biol.* *125*, 145–157.
- Ueda, J., Tachibana, M., Ikura, T., and Shinkai, Y. (2006). Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP. *J. Biol. Chem.* *281*, 20120–20128.
- Ugur, B., Chen, K., and Bellen, H.J. (2016). *Drosophila* tools and assays for the study of human diseases. *Dis. Model. Mech.* *9*, 235–244.
- Vahid, F., Zand, H., Nosrat-Mirshekarlou, E., Najafi, R., and Hekmatdoost, A. (2015). The role dietary of bioactive compounds on the regulation of histone acetylases and deacetylases: a review. *Gene* *562*, 8–15.
- Valle, C., Salvatori, I., Gerbino, V., Rossi, S., Palamiuc, L., René, F., and Carri, M.T. (2014). Tissue-specific deregulation of selected HDACs characterizes ALS progression in mouse models: pharmacological characterization of SIRT1 and SIRT2 pathways. *Cell Death Dis.* *5*, e1296.
- Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* *10*, 252–263.
- Varga-Weisz, P.D., Wilm, M., Bonte, E., Dumas, K., Mann, M., and Becker, P.B. (1997). Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* *388*, 598–602.
- Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* *24*, 374–386.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* *131*, 58–69.
- Verreault, A. (2000). De novo nucleosome assembly: new pieces in an old puzzle. *Genes Dev.* *14*, 1430–1438.
- Villares, R., and Cabrera, C. V (1987). The achaete-scute gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* *50*, 415–424.
- Vincent, S., Vonesch, J.L., and Giangrande, A. (1996). Glide directs glial fate commitment and cell fate switch between neurones and glia. *Development* *122*, 131–139.
- Vinson, C.R., Sigler, P.B., and McKnight, S.L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* *246*, 911–916.
- Volkenhoff, A., Weiler, A., Letzel, M., Stehling, M., Klämbt, C., and Schirmeier, S. (2015). Glial

- Glycolysis Is Essential for Neuronal Survival in *Drosophila*. *Cell Metab.* 22, 437–447.
- Walker, M.P., LaFerla, F.M., Oddo, S.S., and Brewer, G.J. (2013). Reversible epigenetic histone modifications and Bdnf expression in neurons with aging and from a mouse model of Alzheimer's disease. *Age (Dordr.)* 35, 519–531.
- Waltzer, L., Ferjoux, G., Bataillé, L., and Haenlin, M. (2003). Cooperation between the GATA and RUNX factors Serpent and Lozenge during *Drosophila* hematopoiesis. *EMBO J.* 22, 6516–6525.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.
- Wang, H., Ouyang, Y., Somers, W.G., Chia, W., and Lu, B. (2007). Polo inhibits progenitor self-renewal and regulates Numb asymmetry by phosphorylating Pon. *Nature* 449, 96–100.
- Wang, L., Kounatidis, I., and Ligoxygakis, P. (2014a). *Drosophila* as a model to study the role of blood cells in inflammation, innate immunity and cancer. *Front. Cell. Infect. Microbiol.* 3, 113.
- Wang, Q., Goldstein, M., Alexander, P., Wakeman, T.P., Sun, T., Feng, J., Lou, Z., Kastan, M.B., and Wang, X.-F. (2014b). Rad17 recruits the MRE11-RAD50-NBS1 complex to regulate the cellular response to DNA double-strand breaks. *EMBO J.* 33, 862–877.
- Wang, X., Pan, L., Wang, S., Zhou, J., McDowell, W., Park, J., Haug, J., Staehling, K., Tang, H., and Xie, T. (2011). Histone H3K9 trimethylase Eggless controls germline stem cell maintenance and differentiation. *PLoS Genet.* 7, e1002426.
- Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 40, 897–903.
- Waterland, R.A. (2006). Epigenetic mechanisms and gastrointestinal development. *J. Pediatr.* 149, S137–42.
- Watts, R.J., Schuldiner, O., Perrino, J., Larsen, C., and Luo, L. (2004). Glia engulf degenerating axons during developmental axon pruning. *Curr. Biol.* 14, 678–684.
- Weber, A.N.R., Tauszig-Delamasure, S., Hoffmann, J.A., Lelièvre, E., Gascan, H., Ray, K.P., Morse, M.A., Imler, J.-L., and Gay, N.J. (2003). Binding of the *Drosophila* cytokine Spätzle to Toll is direct and establishes signaling. *Nat. Immunol.* 4, 794–800.
- Wiegand, R.C., and Brutlag, D.L. (1981). Histone acetylase from *Drosophila melanogaster* specific for H4. *J. Biol. Chem.* 256, 4578–4583.
- Wiles, E.T., and Selker, E.U. (2017). H3K27 methylation: a promiscuous repressive chromatin mark. *Curr. Opin. Genet. Dev.* 43, 31–37.
- Winston, F., and Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends Genet.* 8, 387–391.
- Wittschieben, B., Otero, G., De Bizemont, T., Fellows, J., Erdjument-Bromage, H., Ohba, R., Li, Y., Allis, C.D., Tempst, P., and Svejstrup, J.Q. (1999). A novel histone acetyltransferase is an integral subunit of elongating RNA polymerase II holoenzyme. *Mol. Cell* 4, 123–128.
- Wolfe, S.A., Nekludova, L., and Pabo, C.O. (2000). DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* 29, 183–212.
- Wolffe, A.P., and Matzke, M.A. (1999). Epigenetics: regulation through repression. *Science* 286, 481–

486.

- Wood, W., and Martin, P. (2017). Macrophage Functions in Tissue Patterning and Disease: New Insights from the Fly. *Dev. Cell* *40*, 221–233.
- Woodcock, C.L., and Dimitrov, S. (2001). Higher-order structure of chromatin and chromosomes. *Curr. Opin. Genet. Dev.* *11*, 130–135.
- Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* *99*, 42–52.
- Woodcock, K.J., Kierdorf, K., Pouchelon, C.A., Vivancos, V., Dionne, M.S., and Geissmann, F. (2015). Macrophage-derived upd3 cytokine causes impaired glucose homeostasis and reduced lifespan in *Drosophila* fed a lipid-rich diet. *Immunity* *42*, 133–144.
- Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* *67*, 545–579.
- Wozniak, R.J., Klimecki, W.T., Lau, S.S., Feinstein, Y., and Futscher, B.W. (2007). 5-Aza-2'-deoxycytidine-mediated reductions in G9A histone methyltransferase and histone H3 K9 di-methylation levels are linked to tumor suppressor gene reactivation. *Oncogene* *26*, 77–90.
- Wu, H., and Sun, Y.E. (2006). Epigenetic Regulation of Stem Cell Differentiation. *Pediatr. Res.* *59*, 21–25.
- Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol. Cell. Biol.* *28*, 7337–7344.
- Wu, Y., Dissing-Olesen, L., MacVicar, B.A., and Stevens, B. (2015). Microglia: Dynamic Mediators of Synapse Development and Plasticity. *Trends Immunol.* *36*, 605–613.
- Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., et al. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* *442*, 86–90.
- Xu, W., Edmondson, D.G., and Roth, S.Y. (1998). Mammalian GCN5 and P/CAF acetyltransferases have homologous amino-terminal domains important for recognition of nucleosomal substrates. *Mol. Cell. Biol.* *18*, 5659–5669.
- Xu, W., Edmondson, D.G., Evrard, Y.A., Wakamiya, M., Behringer, R.R., and Roth, S.Y. (2000). Loss of Gcn5l2 leads to increased apoptosis and mesodermal defects during mouse development. *Nat. Genet.* *26*, 229–232.
- Yamane, K., Toumazou, C., Tsukada, Y., Erdjument-Bromage, H., Tempst, P., Wong, J., and Zhang, Y. (2006). JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* *125*, 483–495.
- Yamane, K., Tateishi, K., Klose, R.J., Fang, J., Fabrizio, L.A., Erdjument-Bromage, H., Taylor-Papadimitriou, J., Tempst, P., and Zhang, Y. (2007). PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol. Cell* *25*, 801–812.
- Yamauchi, T., Yamauchi, J., Kuwata, T., Tamura, T., Yamashita, T., Bae, N., Westphal, H., Ozato, K., and Nakatani, Y. (2000). Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 11303–11306.
- Yang, X.-J., and Grégoire, S. (2005). Class II histone deacetylases: from sequence to function, regulation,

- and clinical implication. *Mol. Cell. Biol.* *25*, 2873–2884.
- Yang, X.-J., and Seto, E. (2003). Collaborative spirit of histone deacetylases in regulating chromatin structure and gene expression. *Curr. Opin. Genet. Dev.* *13*, 143–153.
- Yang, X.-J., and Seto, E. (2008). Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol. Cell* *31*, 449–461.
- Yang, X.-J., and Ullah, M. (2007). MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene* *26*, 5408–5419.
- Yang, Y., and Bedford, M.T. (2013). Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* *13*, 37–50.
- Yang, X.J., Ogryzko, V. V., Nishikawa, J., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* *382*, 319–324.
- Yao, Y.L., Yang, W.M., and Seto, E. (2001). Regulation of transcription factor YY1 by acetylation and deacetylation. *Mol. Cell. Biol.* *21*, 5979–5991.
- Yildirim, K., Petri, J., Kottmeier, R., and Klämbt, C. (2019). *Drosophila* glia: Few cell types and many conserved functions. *Glia* *67*, 5–26.
- Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2011). A high-resolution whole-genome map of key chromatin modifications in the adult *Drosophila melanogaster*. *PLoS Genet.* *7*, e1002380.
- Yokoyama, A., Wang, Z., Wysocka, J., Sanyal, M., Aufiero, D.J., Kitabayashi, I., Herr, W., and Cleary, M.L. (2004). Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with menin to regulate Hox gene expression. *Mol. Cell. Biol.* *24*, 5639–5649.
- Yokoyama, R., Pannuti, A., Ling, H., Smith, E.R., and Lucchesi, J.C. (2007). A plasmid model system shows that *Drosophila* dosage compensation depends on the global acetylation of histone H4 at lysine 16 and is not affected by depletion of common transcription elongation chromatin marks. *Mol. Cell. Biol.* *27*, 7865–7870.
- Yoo, S.Y., Pennesi, M.E., Weeber, E.J., Xu, B., Atkinson, R., Chen, S., Armstrong, D.L., Wu, S.M., Sweatt, J.D., and Zoghbi, H.Y. (2003). SCA7 knockin mice model human SCA7 and reveal gradual accumulation of mutant ataxin-7 in neurons and abnormalities in short-term plasticity. *Neuron* *37*, 383–401.
- You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. (2001). CoREST is an integral component of the CoREST-human histone deacetylase complex. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 1454–1458.
- Younossi-Hartenstein, A., Nguyen, B., Shy, D., and Hartenstein, V. (2006). Embryonic origin of the *Drosophila* brain neuropile. *J. Comp. Neurol.* *497*, 981–998.
- Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* *286*, 7983–7989.
- Yuasa, Y., Okabe, M., Yoshikawa, S., Tabuchi, K., Xiong, W.C., Hiromi, Y., and Okano, H. (2003). *Drosophila* homeodomain protein REPO controls glial differentiation by cooperating with ETS and BTB transcription factors. *Development* *130*, 2419–2428.
- Ziegenfuss, J.S., Biswas, R., Avery, M.A., Hong, K., Sheehan, A.E., Yeung, Y.-G., Stanley, E.R., and Freeman, M.R. (2008). Draper-dependent glial phagocytic activity is mediated by Src and Syk family kinase signalling. *Nature* *453*, 935–939.

