

Technische Universität München  
Department of Electrical Engineering and Information Technology  
Bio-Inspired Information Processing

Diplomarbeit

Biological and computational framework for  
the identification of neuronal circuits in  
*Drosophila* brains

Thomas Templier

*Date of Submission:*

August 31, 2011

*Supervisors:*

Dr. Hiromu Tanimoto

Prof. Dr.-Ing. Werner Hemmert



## Abstract

One of the powerful approaches to understand the organization of neural circuits is to activate or inactivate a specific neuronal component in isolation, and then to measure the consequence of this manipulation. The fruit fly *Drosophila melanogaster* exhibits a wide variety of behaviours, and at the same time, targeted gene expression systems allow specific functional manipulation in intact freely behaving flies. The Gal4/UAS system is one of the transgene expression system in *Drosophila*, and it consists of a "Gal4 driver" defining the cells to express a transgene and UAS-transgene. By transgenically expressing a protein that inhibits synaptic vesicle release in various sets of cells using different Gal4 drivers, a previous study has isolated approximately one hundred Gal4 driver lines that impaired responses to appetitive or aversive sensory stimuli. However, responsible neurons in these Gal4 lines remain to be identified.

The goal of this study is to develop an experimental system to systematically visualize the expression pattern of Gal4 and to establish a computational framework that quantitatively compares the expression pattern in the brain, in order to identify the cells repeatedly labelled in a group of Gal4 lines showing a same behavioural phenotype. To this end, I designed a large scale biological and computational production pipeline to image immunolabelled fly brains using confocal microscopy and to detect repeatedly labelled neurons. For quantitative comparison, each three-dimensional confocal image was non-linearly transformed to fit into a standard brain by employing state-of-the-art image registration tools. I developed a critical labelling method that is optimized for a transformation reference. Furthermore, I implemented a series of machine vision algorithms into the pipeline system in order to address numerous sources of biological noise that would degrade the detection quality. The production pipeline I established serves as a computational hypothesis generator for neurons repeatedly labelled in different Gal4 lines and represents the first attempt for a semi-automated image processing system. This pipeline is versatile and has possibilities of further applications to other high-noise biological materials.



# Acknowledgments

I would like to thank my supervisor Dr. Hiromu Tanimoto who taught me numerous of essential biological concepts, and for his attention throughout my thesis. I want to thank Prof. Dr.-Ing. Werner Hemmert for the academic supervision. Many thanks belong to Vladimiro Thoma and Anja Friedrich for the introduction to fly biology and laboratory practice.

I want to thank Pavel Tomancak and his group, Johannes Schindelin and Benjamin Schmidt for my visit to the Max Planck Institute in Dresden, and the Fiji community who provided me substantial help for the design of the plugins. I'm grateful to Hanchuan Peng for providing the software BrainAligner and answering to my requests for some functionalities.

The line pJFRC66 was kindly provided by Barret Pfeiffer, Rubin Lab, JFRC/HHMI.

Finally I thank my mother, my father and my brothers for their support.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction to behavioral genetics . . . . .	1
1.1.1	Behavioral genetics . . . . .	1
1.1.2	Overall project . . . . .	2
1.1.3	Genetic tools and existing database . . . . .	2
1.2	Behavioral assay . . . . .	4
1.2.1	Sugar response . . . . .	4
1.2.2	Behavioural assay for electric shock responses. . . . .	5
1.3	Immunohistochemistry . . . . .	6
1.3.1	Principle . . . . .	6
1.3.2	Neuropil markers . . . . .	7
1.3.3	Anti-Green-Fluorescence-Protein antibody . . . . .	9
1.3.4	Nuclei markers . . . . .	10
1.3.5	Antibody combination for the large scale screening . . . . .	12
1.4	Fluorescent confocal microscopy . . . . .	14
1.4.1	Fluorescent microscopy . . . . .	14
1.4.2	Confocal microscope . . . . .	16
1.5	Image Registration . . . . .	16
1.5.1	General principle . . . . .	16
1.5.2	Intensity-based and landmarks-based registration . . . . .	17
1.5.3	Two-steps algorithm . . . . .	17
1.6	Goal of the Diploma Thesis . . . . .	18
1.6.1	Central nervous system . . . . .	18
1.6.2	Overlapping neurons . . . . .	19
1.6.3	Existing techniques . . . . .	20
1.6.4	Biological and computational pipeline for a density map of nuclei . . . . .	20
<b>2</b>	<b>Materials and methods</b>	<b>22</b>
2.1	Fly strains . . . . .	22
2.1.1	Use of the Kyoto fly database . . . . .	22

## Contents

2.1.2	Double-reporter line : nls-lacZ and myrGFP . . . . .	22
2.2	Fly preparation . . . . .	22
2.2.1	Brain dissection . . . . .	22
2.2.2	Mounting . . . . .	23
2.2.3	Immunohistochemistry for the large scale screening . . . . .	24
2.2.4	Scanning with the fluorescent confocal microscope . . . . .	25
2.3	Image processing with Fiji . . . . .	26
2.4	Brain registration . . . . .	26
2.4.1	Contrast enhancement . . . . .	26
2.4.2	Creation of the average reference brain . . . . .	26
2.4.3	Large scale registration . . . . .	30
2.5	Nuclei detection . . . . .	30
2.6	Glia detection and removal . . . . .	31
2.7	Nuclei removal based on GFP expression . . . . .	33
<b>3</b>	<b>Results</b>	<b>34</b>
3.1	Quadruple staining . . . . .	34
3.1.1	Absence of cross reactions between antibodies . . . . .	34
3.1.2	Optimization of dilution factors for the third channel . . . . .	34
3.2	Registration . . . . .	37
3.2.1	Overview . . . . .	37
3.2.2	Nuclei registration . . . . .	37
3.2.3	Brains partially damaged . . . . .	38
3.3	Removal of optic lobes . . . . .	40
3.3.1	Technique employed . . . . .	40
3.3.2	Reasons for not using this plugin . . . . .	42
3.4	Nuclei detection . . . . .	43
3.4.1	Quality . . . . .	43
3.4.2	Statistics . . . . .	44
3.5	Nuclei variability . . . . .	44
3.5.1	Preliminary results . . . . .	44
3.5.2	Results . . . . .	46
3.6	Detection and removal of glial cells . . . . .	48
3.6.1	Goal . . . . .	48
3.6.2	Intensity profiles of glial cells in the registration channel . . . . .	48
3.6.3	Methods . . . . .	50
3.6.4	Threshold rough estimation . . . . .	52
3.6.5	Threshold determination . . . . .	52
3.6.6	Final validation . . . . .	53
3.6.7	Application on the large scale screening . . . . .	57
3.7	Removal of nuclei based on GFP signal . . . . .	58



3.7.1	Preliminary observations . . . . .	58
3.7.2	Definition of the neighborhood . . . . .	65
3.7.3	Score of maximum intensity . . . . .	65
3.7.4	Tests on weak and strong lines . . . . .	66
3.7.5	Statistics for the large scale screening . . . . .	67
3.8	Density map . . . . .	67
3.8.1	Plugin . . . . .	67
3.8.2	Radius of nuclei . . . . .	67
3.8.3	3D display . . . . .	67
<b>4</b>	<b>Discussion</b>	<b>73</b>
4.1	Nuclei removal based on GFP channel and registration channel	73
4.1.1	Positive nls-lacZ labeling and negative GFP labeling .	73
4.1.2	Lines with strong glial cells expression . . . . .	74
4.1.3	Effect of false positive glia removal . . . . .	74
4.1.4	Comparison of the two removal techniques . . . . .	75
4.2	Density map . . . . .	76
4.2.1	Number of nuclei visualized simultaneously . . . . .	76
4.3	Further applications . . . . .	78
	<b>List of Figures</b>	<b>83</b>
	<b>List of Tables</b>	<b>84</b>
	<b>References</b>	<b>88</b>

*Contents*

# Chapter 1

## Introduction

### 1.1 Introduction to behavioral genetics

#### 1.1.1 Behavioral genetics

Behavioral genetics is a field of study that aims to fill the immense gap between two extremes. On one side, the genome (structure), raw data containing all information necessary to the development of an animal in a proper environment. On the other side, the behaviour (function), the very end of organisms interacting with their environment. Behavioral genetics addresses the question of the inheritance of behavioural traits and seeks to understand both the genetic and environmental contributions to individual variations in behavior.

*Drosophila* is an ideal model organism for addressing questions regarding the structure and function of neuronal circuits in combination to behavioural measurements. On the one hand, the fruit fly's nervous system is much simpler than in vertebrates: its brain consists of about  $10^5$  neurons, whereas the human brain is estimated to contain  $10^{11}$  neurons. Being one of the most studied organisms on earth, we have access to very advanced tools in genetics that allow us to investigate the dissection of complicated circuits. On the other hand, the easy handling of flies in a laboratory environment is a main reason why this animal has been thoroughly studied for decades. *Drosophila* flies reproduce massively and quickly (thousands of progenies 20 days after having put 5 males and 5 females in a vial), are small organisms, and can be tested in behavioral assays almost without restriction on animal protection.

### **1.1.2 Overall project**

Animals rely on sensory inputs to perceive their environment. Such stimuli are detected by sensory systems and are further processed by higher-order neurons. Eventually, motor neurons drive suitable behavioral responses.

The work presented here is part of a joint project of research groups addressing the issue of comprehensively characterizing functional neuronal circuits that drive responses to sugar and electric shock in *Drosophila*. Flies have been previously genetically modified, so that various random subsets of neurons are inhibited in the fly brain. Hundred of these lines have been selected based on the exhibition of an abnormal behaviour in response to sugar and electric shock. We aim to procede to a large scale screening of the brains of flies belonging to these hundred selected lines, in order to identify neurons that have been repeatedly inhibited in several lines. These common neurons, also called overlapping neurons, could be involved in the circuits driving response to sugar and electric shock.

### **1.1.3 Genetic tools and existing database**

#### **Gal4/UAS system**

The GAL4/UAS system is a very powerful genetic tool used to characterize functional neuronal circuits in *Drosophila*. It consists of two components: a gene named GAL4 that encodes the protein Gal4 responsible for the activation of the transcription of specific genes in yeast, and a DNA sequence called Upstream Activation Sequence (UAS) that plays the role of a promoter region. In cells where both Gal4 proteins and the UAS are present, the Gal4 protein binds to the UAS, which activates the transcription of the gene in the DNA sequence following the position of UAS. This gene can be selected by the geneticist among a large set of genes well characterized.

All cells of an animal have the same genome, but all cells do not activate the transcription of all their genes. Only some sets of cells activate the transcription of a specific gene, this specificity being based on the neighboring genome that contains many regulatory elements of gene expression. The GAL4 gene is typically inserted on a place of the DNA sequence, so that it is activated only in some subsets of cells. We say that the GAL4 gene has been inserted under the control of an endogenous gene. The UAS being present in all cells, only cells that activate the transcription of GAL4 will express the gene chosen by the geneticist that follows the Upstream Activation Sequence in the DNA sequence. The general principle is shown in Figure 1.1.

1.1 Introduction to behavioral genetics

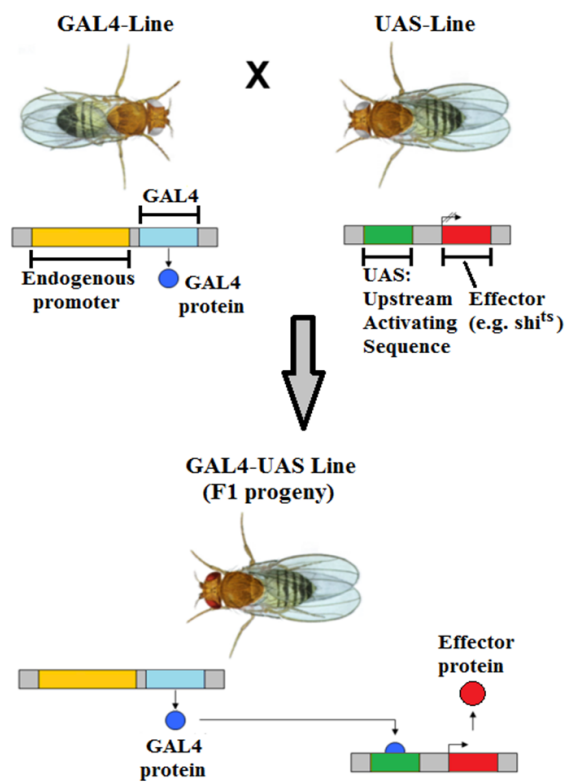


Figure 1.1: General principle of the UAS/Gal4 system (Perrimon et al., 1991)

## **Effector and reporter genes**

The gene following the UAS can be an effector or a reporter gene. Effector genes are used for the manipulation of cellular functions, and reporter genes are used for visualization of specific cellular structures.

Effector genes, like dTrpA1, Shibire or TNT, code for proteins that have an effect on the cell. For example, the protein coded by the *ts1* allele of the SHIBIRE gene (Van Der Blik and Meyerowitz, 1991), is serving to inactivate the neuronal output. Shibire<sup>*ts1*</sup> inhibits vesicle release at presynaptic terminals, and thus impairs neurotransmission of chemical synapses. Over 30° Celsius, the protein becomes dominant negative.

The strategy used is to raise flies in normal temperature at 25° Celsius to allow a normal development. Before testing, the flies are exposed to a high temperature, typically over 33° Celsius, for 20 minutes, which activates the protein coded by SHIBIRE<sup>*ts1*</sup>. All neurons have a spontaneous activity, which insures us that the neuron has used its amount of vesicles and is then in an inhibited state. Flies are then subjected for one minute to following assays described in next section 1.2.

Reporter genes, like *mcd8::GFP* or *nls-lacZ* code for proteins having none or little effect on the cell. Typically, these proteins can be detected by immunofluorescence by the mean of fluorescent microscopy, "reporting" the location of the expressing cell. Candidate components of neuronal circuits corresponding to abnormal behaviour can thus be visualized.

## **1.2 Behavioral assay**

Around one thousand Gal4 lines have been crossed to shibire<sup>*ts1*</sup> lines, namely, lines of flies that have the gene SHIBIRE in their genome. The progeny have been subjected to two different behavioural assays and screened for a defective response. Around 300 lines have been already isolated.

The progeny of the crosses of these 300 lines is tested in the following two behavioral assays for defective stimulus responses. In our case, neuronal circuits involved in the perceptions and response memories to sugar and electric shocks are investigated.

### **1.2.1 Sugar response**

The behavioural assay for sugar responses enables the experimenter to test the behavior of the flies towards perception and response to sugar. Flies are inserted in the sugar arena, that is shown in Figure 1.2. The flies are given a binary choice between two pieces of filter papers containing sucrose solution

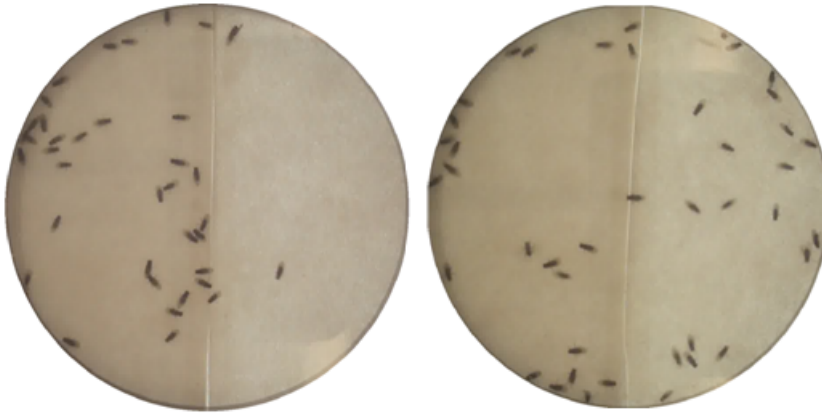


Figure 1.2: Two sugar arenas. On the left part of each arena: sucrose solution filter paper. On the right of each arena: water on filter paper. On the left arena: normal behaviour. On the right arena: abnormal behaviour.

or water. The whole setup is backlit and flies are confined into a small space like a turned petri dish. Starved wild-type flies gather on the sugar side, like in Figure 1.2, left. In the arena right, the flies show an abnormal behaviour, not showing any preference for the sugar. To assess the behaviours, the following components are observed : sugar preference, grooming, courtship, walking, proboscis extension, activity during experiment and activity before experiment.

So far, 40 lines with an abnormal behaviour towards perception and response to sugar have been found.

### 1.2.2 Behavioural assay for electric shock responses.

An electric grid is present on the floor of the petri dish arena. Similarly to the sugar arena, only the right half delivers electric shocks. The arena is shown in Figure 1.3 with an normal behaviour on the left arena, and an abnormal behaviour on the right arena. The following behaviours are assessed : avoidance, jumping, flying, number of supine flies, time for rolling over, grooming, time to re-disperse, activity during experiment, activity before experiment.

So far, 33 lines with an abnormal behaviour have been found.



Figure 1.3: Two electric shock arenas. On the left of each arena: electric shock. On the right of each arena: no electric shock. On the left arena: normal behaviour. On the right arena: abnormal behaviour.

## **1.3 Immunohistochemistry**

### **1.3.1 Principle**

#### **Antibodies**

Immunohistochemistry provides a set of powerful tools for the investigation of neuronal circuits. It makes use of the general biological system of antigens and antibodies and is used to visualize the localisation of specific structures.

Antibodies are complex proteins produced by animals to label or neutralise several foreign elements present in the organism. Antibodies are characterized by two identical binding domains at the extremity of a Y-shaped protein, that are specific to a class of antibodies. They bind to specific structures presenting the complementary domain called the antigen.

#### **Principle**

Assume we are interested in a protein  $P$  in a species  $S$ , and this protein is also present in a second species  $S'$ . The immunohistochemistry is done in two steps. In the first step, we let an animal of the species  $S$  produce antibodies specific to the protein  $P$ . Then, these antibodies, called primary antibodies, are collected and inserted in an animal of the species  $S'$ . The antibodies bind to the proteins  $P$ . In the second step, antibodies are artificially produced with two characteristics, and are called secondary antibodies. They bind to antibodies from the first species  $S$ , and they present a label on their base. The label in our case is a fluorescent structure. These secondary antibodies are



injected in the animal of the species S'. They will bind to the already present primary antibodies bound themselves to the proteins P. Finally, fluorescent structures are therefore bound to the proteins P in the species S', which gives the localisations of the proteins P.

### Fluorescent labels

The label of the secondary antibodies is in our case a fluorophore. Fluorophores have the determinant characteristic of emitting light with a higher wavelength in response to a light stimulus with a lower wavelength. Fluorophores have a specific absorption wavelength and a specific emission wavelength. In this work, we used secondary antibodies labeled with fluorophores having three different ranges of emission wavelength : 488 *nm*, 568 *nm*, and 633 *nm*.

It is important to note that all primary antibodies raised in the same species will be targeted by secondary antibodies having the same emission wavelength. That is, it is for instance impossible to distinguish two types of structures labeled by two primary antibodies raised in mouse. This principle is of crucial importance for the choice of the antibody combination.

### 1.3.2 Neuropil markers

The neuropil refers to the regions of the brain where synapses are present. Several antibodies enable the labeling of these regions.

#### Antibody anti-synapsin

The antibody anti-synapsin binds to a protein called synapsin. This protein binds to synaptic vesicles. Thus, the signal is primarily found in presynaptic regions, as shown in Figure 1.4. It labels the major compartments in the CNS. Some of them are coloured in Figure 1.14. Most of the counter stainings done in my laboratory are done with this antibody, that has shown a strong reliability. The quality of the staining is high, leading to few evenness and reproducible features.

An appreciable characteristic of this staining is the clear intensity gradients inside the neuropil. These contrasts are not so high in some other neuropil stainings, like for exemple anti-discs-large antibody. These gradients are of high importance for the registration process, as we will see later.

## *Introduction*

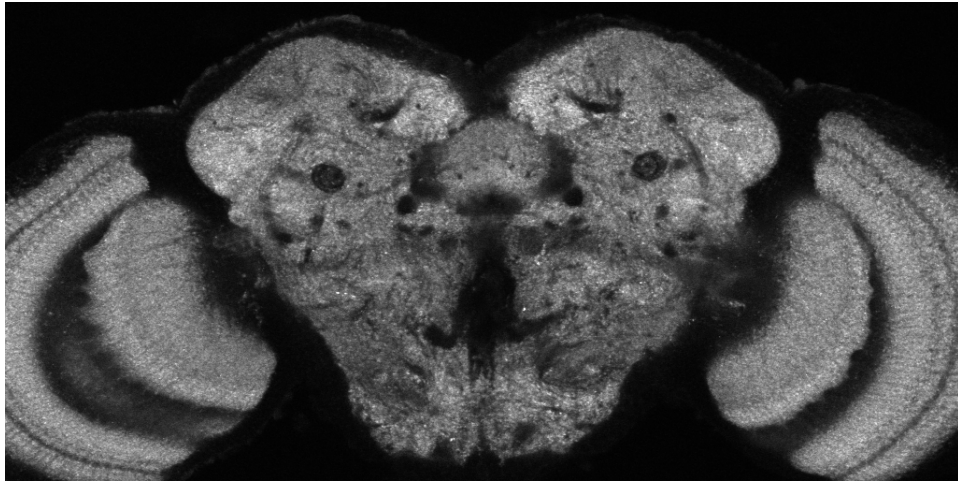


Figure 1.4: Slice of a brain stained with the antibody anti-synapsin

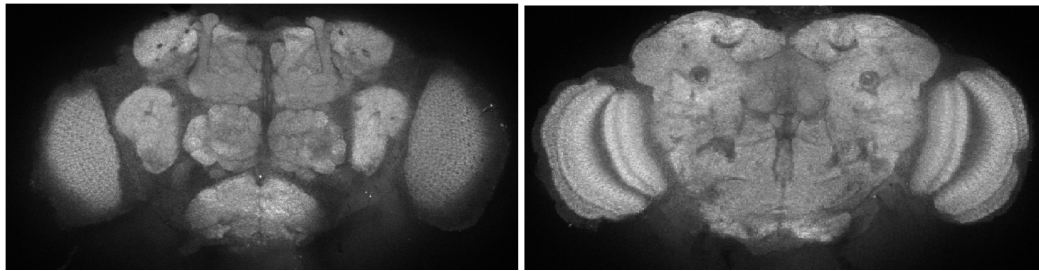


Figure 1.5: Slices of a brain stained with the antibody anti-NCadherine

### **Anti-NCadherine**

The antibody anti-NCadherine is rather weak and difficult to stain. It produces less contrast than anti-synapsin. This antibody could however be useful because it is raised in the rat. This specie does not overlap with the most commonly used host animals used to raise antibodies like mouse and rabbit. Figure 1.5 shows such a staining.

### **Anti-Discs-Large**

The antibody anti-discs-large is a rather strong neuropile marker and is quite reliable. However, the contrasts observed with this staining are not as clear as with synapsin. Being also raised in mouse, this antibody does not offer better conditions than anti-synapsin, thus we preferred the latter one. Figure 1.6 shows a brain stained with anti-discs-large.

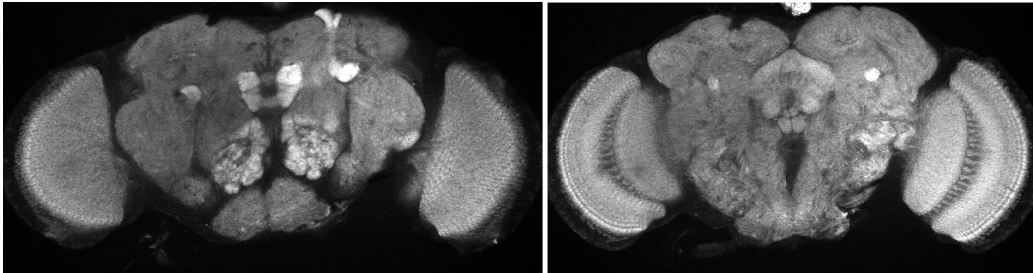


Figure 1.6: Slices of a brain stained with the antibody anti-discs-large

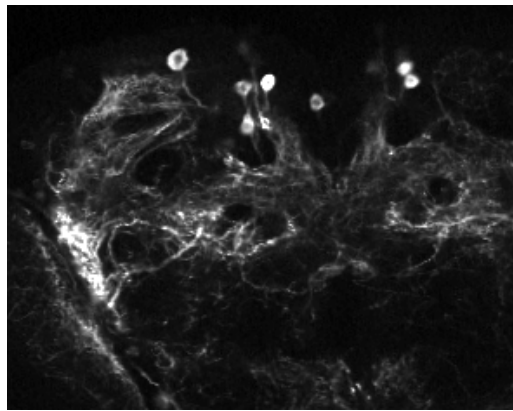


Figure 1.7: Anti-GFP staining. We see that membranes of somas, dendrites and axons are labeled.

### 1.3.3 Anti-Green-Fluorescence-Protein antibody

Visualization of the morphology of the GAL4-expressing neurons was done with UAS-myrfGFP. This protein is a modified version of the green fluorescence protein (GFP). It consists of the GFP with a fatty acid tail. The fatty acid tail presents a hydrophobic extremity, similar to the fatty acids that form plasma membranes. Thus, myrfGFP is localized in plasma membranes.

The GFP protein exhibits a bright fluorescence when exposed to blue light. However, the fluorescence of the GFP protein is rather weak. That is why we use an anti-GFP antibody to enhance the fluorescence of the GFP. The secondary antibody chosen must have the same wavelength as the one of the fluorescence emitted by the GFP protein, namely,  $488nm$ . Thus, the signal emitted by the GFP itself is strengthened by the signal of the secondary antibody. Figure 1.7 gives an example of an anti-GFP staining in a Gal4-UAS-myrfGFP line. Plasma membranes in the nuclei are labelled, as well as plasma membranes in axons and dendrites.

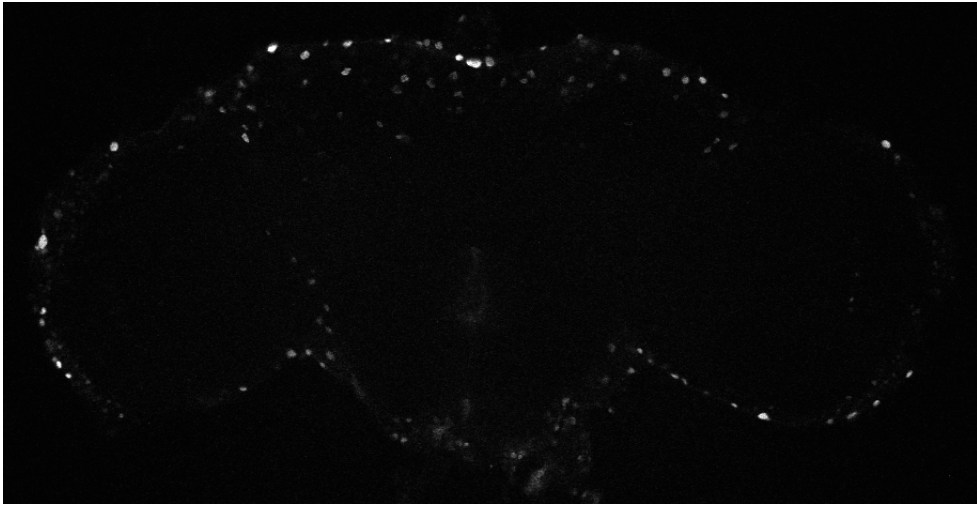


Figure 1.8: Anti-beta galactosidase staining. Only nuclei of cells expressing nls-lacZ are labeled.

### 1.3.4 Nuclei markers

#### Anti-beta-galactosidase

At the same time of detecting cellular morphology with myrGFP, nuclei of GAL4-expressing cells were labelled by UAS-nls-lacZ. NLS stands for the nuclear localization signal, which is an amino acid sequence stemming from the gene *myc*, that is serving the specific transport of structures. Elements containing a nuclear localization signal will be transported to the nuclei. Thus, UAS-nls-lacZ is only present in the nuclei of the cells expressing Gal4. lacZ is beta-galactosidase from bacteria and serves as a marker. It can be used as a marker since it does not have a particular biological function in the fly. Using the antibody against beta-galactosidase, nls-lacZ results in clear dots visible at the confocal microscope. The Figure 1.8 shows an example.

#### Glial cells and anti-elaV

**Glial cells** Glial cells (or glia) are cells present in the nervous system that provide support to the neurons. The antibody anti-repo labels specifically the nuclei of glia. It has been used to stain the brain shown on Figure 1.9, where we see the nuclei of glia. Using a nuclei detector program described later, we counted on a single brain around 4000 glia.

**Reporter gene expression in neurons and glia** myrGFP and nls-lacZ label neurons as well glia, and thus it is very difficult for machine vision

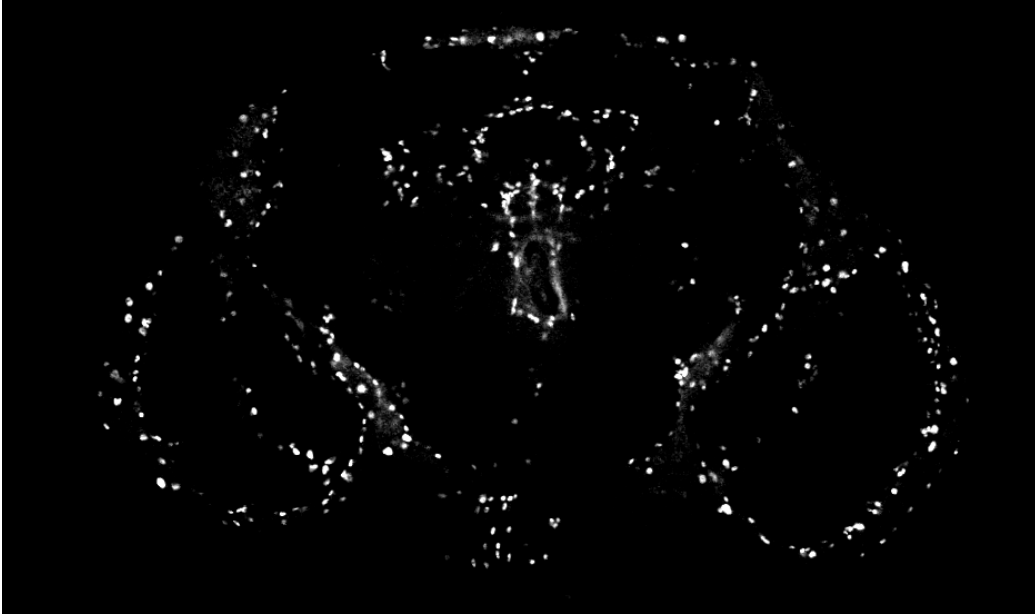


Figure 1.9: Staining with the antibody anti-repo showing glial cells in the fly brain

algorithms to distinguish neurons and glia with these transgenes. Trained human experts may be able to distinguish them based on their morphology visualized with myrGFP only in particular cases.

**Need for glia detection** As we are interested in characterizing neuronal circuits, and it is less likely that the acute responses to electric shock or sugar are directly controlled by glia, we would like to exclude glial cells from our analyses. Considering the molecular nature of Shi dynamin, the effect is well documented on neurons (Van Der Blik and Meyerowitz, 1991), but less is known on the consequences for glia.

By driving expression of GAL80 in glial cells, that is a suppressor of GAL4, it is possible to transgenically 'remove' the GAL4 expression in glia. However, it would take months of work, because one has to make a new transgenic line. I have therefore to deal with Gal4 lines containing a lot of glia, which may impair our goal that focuses on only neurons.

The antibody against Repo protein is available from rabbit and mouse. However, rabbit anti-Repo is polyclonal and very limited in amount, and therefore it is not optimal for a large-scale immunohistochemical use. Thus, using anti-repo from mouse is the only solution to label glial cells. It will be later explained why this antibody has not been used for the large scale

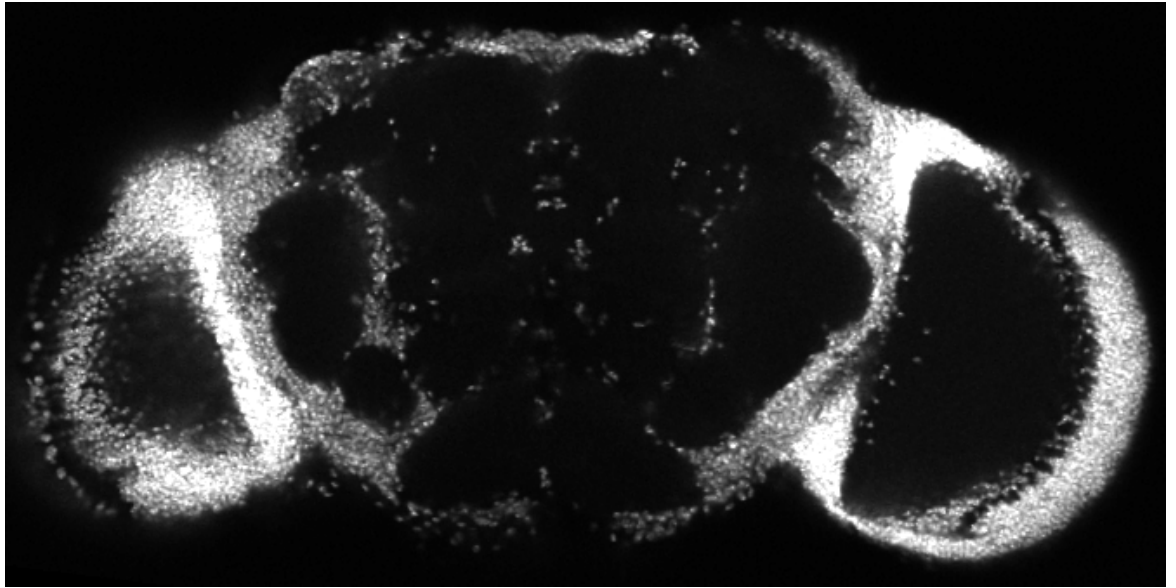


Figure 1.10: Staining with the anti-body elavV

screening.

**Anti-elav antibody** elav is a transcription factor expressed exclusively in neurons, and thus, not in glial cells (Robinow and White, 1991). An idea we retained is that this difference could be used to make a distinction between neurons and glia. Being able to detect the protein elav would provide a way to distinguish them because glial nuclei are negatively stained. The anti-elav antibody is available from the rat and from the mouse. The latter one is widely used and my laboratory has experienced it for a long time with a well established protocol.

An example of this staining is given in Figure 1.10.

### 1.3.5 Antibody combination for the large scale screening

The table 1.1 summarizes the major characteristics of each antibody, that led us to make our final choice for the large scale screening. Our needs are multiple and can be outlined in the following points.

- Labelling of plasma membranes of Gal4-expressing cells
- Labelling of the nuclei of GAL4-expressing cells

### 1.3 Immunohistochemistry

	GFP	Synapsin	NCadherine	Discs large	elav	repo	$\beta$ galactosidase
label	plasma membrane	neuropile	neuropile	neuropile	neuron nuclei	glial cells nuclei	lacZ protein in nuclei
animal	rabbit, rat, mouse	mouse	rat	mouse	rat, mouse	mouse	chicken, rabbit
intensity	average	average	low	high	high	average	average
reliability	average	high	low	average	high	high	high

Table 1.1: Principal characteristics of available antibodies

## *Introduction*

- Reliable neuropile marker for a registration reference
- Labeling of all nuclei in the cortex for better registration of nls-lacZ signals
- Detection and exclusion of glial cells

Our microscope has a limited number of channels, so I should meet all the criteria above with three channels. Particular attention has to be put on the species where the antibodies have been raised.

Fullfilling the condition of the first item, leads us to use anti-GFP on the  $488nm$  channel. This channel should exclusively contain this staining.

For the second condition, I need an anti- $\beta$ galactosidase staining on another channel. To enable an automated detection of nuclei, this staining can't be combined to another staining on the same channel.

The two first conditions requiring two different channels, it means that the three last conditions have to be fullfilled on only one channel. Firstly, it is sure that this channel must contain a neuropile marker. Using then the antibody anti-repo would have been of great help to reliably detect glia. Repo is rather reliable and straightforward to detect with the microscope. It would mean having a neuropile marker and anti-repo on the same channel, as shown in Figure 1.11, with synapsin as neuropil marker. As visible in the picture, the repo signal is rather strong and present throughout the brain. The locations of these dots being variable among different brain, it seems that it would impair the registration process. Additionnally, the quality of the two stainings repo and synapsin have to be high, in order to enable us to systematically distinguish the repo blobs from the neuropile background. That is why we decided not to use anti-repo. The remaining solution is to use the antibody anti-elav and a neuropil marker on the same channel. Anti-elav labels all neuron nuclei in cortex regions, which will be useful for the accurate registration of these regions. Moreover, it does not label glial cells, which will be used to distinguish neurons from glia.

The final staining we chose for the large scale screening is given in Table 1.2.

## **1.4 Fluorescent confocal microscopy**

### **1.4.1 Fluorescent microscopy**

Fluorescent microscopy is an imaging technique that enables the study of biological tissues using the fluorescence properties of molecules. Fluorophores,



#### 1.4 Fluorescent confocal microscopy

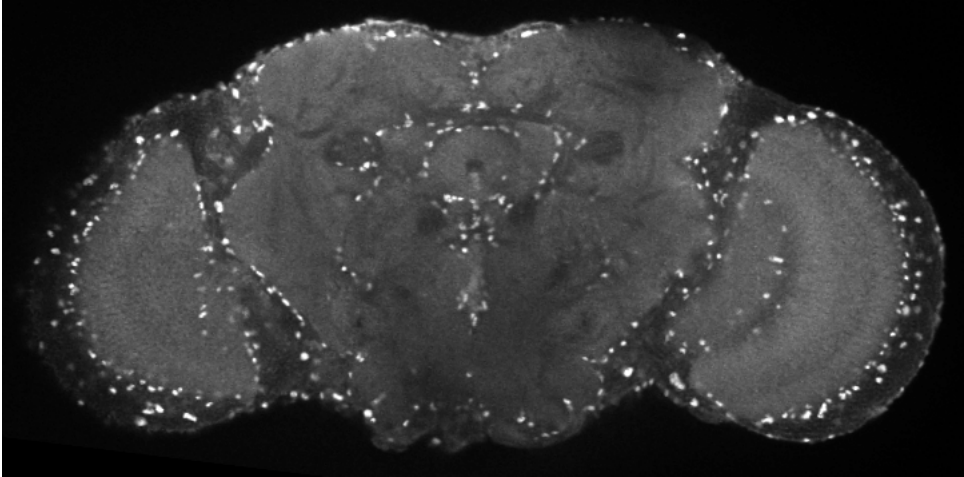


Figure 1.11: Anti-synapsin and anti-repo on the same channel

Channel	First channel	Second channel	Third channel
Antibody	anti-GFP	anti-synapsin; anti-elav	anti- $\beta$ galactosidase
Animal	Rabbit	Mouse; Rat	Chicken
Wavelength	488nm	563	633nm

Table 1.2: Staining retained for the large scale screening

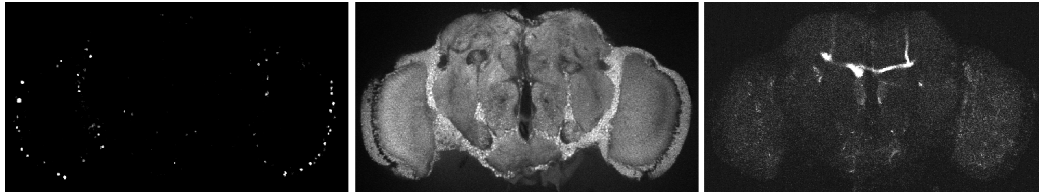


Figure 1.12: Three images of three different channels acquired from the same biological sample. Left: 488 nm; Middle 560 nm; Right 630 nm

as mentioned in 1.3.1, can emit light with a higher wavelength than the one of the light they can absorb. In fluorescent microscopes, quasi-monochromatic light produced by a laser illuminates a biological sample containing fluorophores. They absorb the light and emit then light with a higher and specific wavelength, that is then received by a detector. The specificity of the emitted light is a crucial feature of fluorophores. It makes possible to distinguish the lights emitted by fluorophores with different emission wavelengths. The microscope I use has a detector able to receive three different ranges of wavelength around  $490nm$ ,  $560nm$  and  $630nm$ . Therefore it is possible to acquire three different images corresponding to these three ranges of wavelengths. These ranges are referred as channels. Figure 1.12 shows the images of the three channels of a brain stained with antibodies emitting in the three ranges.

## 1.4.2 Confocal microscope

Confocal microscopy is an optical imaging technique that enables the acquisition of signals being exclusively located in the focal plan. It is used to acquire images of thick biological tissues. The whole sample is imaged step by step, producing slice images of each focal plan of the tissue. Such slices are widely used in three dimensional reconstructions of biological samples. In our case, I use a step of  $1.5\mu m$  to acquire brain images. The thickness of a *Drosophila* brain being roughly  $150nm$ , I acquire an amount of approximately 100 slices per brain.

## 1.5 Image Registration

### 1.5.1 General principle

Comparing neuronal structures between hundreds of different brains is only conceivable if we are able to observe them in a common reference coordinate



Figure 1.13: General registration process. Left: subject (template) brain; Middle: transformed (warped) brain; Right : reference (target) brain

system (Peng et al., 2011). Image registration consists in transforming a subject image (also called template image) to a target image (also called reference image). The subject image and the target image must present common features and can stem from the same object at different timepoints, from different view points, or from different objects. In our case, our objects are *drosophila* brains from different individuals that present characteristic features common to all of them. Figure 1.13 shows a sample brain (left) being registered to a target brain (right). The result brain (also called transformed brain or warped brain) is shown in the middle.

### 1.5.2 Intensity-based and landmarks-based registration

Registration algorithms can be classified in landmarks-based (Jenett et al., 2006; Peng et al., 2011) and intensity-based techniques (Peng et al., 2011; Klein et al., 2010; Rohlfing and Maurer, 2003) for the evaluation of similarity of two different images. Intensity-based approaches compare features such as correlation or mutual information (Rohlfing and Maurer, 2001). Landmarks-based approaches evaluate the similarity of two images based on the pairing of points or regions of interest called landmarks. These landmarks can be automatically detected (Peng et al., 2011) or manually defined by the user (Jenett et al., 2006).

Intensity-based approaches tend to be more used in fields requiring real time registration (Klein et al., 2010). For our purpose, the approach leading to the best results is a landmarks-based method (Peng et al., 2011).

### 1.5.3 Two-steps algorithm

Most of registration algorithms are performed in two distinct steps (Klein et al., 2010; Peng et al., 2011; Rohlfing and Maurer, 2001; Rueckert et al.,

1999). The first one consists in a rigid or affine transformation. A rigid transformation  $\mathbf{R}$  is defined by an orthogonal matrix  $A$  and a vector  $\vec{t}$  such that for every vector  $\vec{v}$  of an euclidean space, we have  $\mathbf{R}(\vec{v}) = A\vec{v} + \vec{t}$ . A rigid transformation preserves distances of every pair of points. This transformation can be seen as a composition of rotations and translations. Rigid transformations belong to the affine transformations, that allow also scaling. The first step of the registration process allows a rough alignment of the sample to the target image. The sample image is rotated, scaled and translated to maximize its similarity with the reference image.

The second step consists in locally maximizing similarities between target and reference images. It is also called warping. The target image is distorted to match local similarities.

After these two steps, we say that the subject image has been registered to the target image.

## 1.6 Goal of the Diploma Thesis

### 1.6.1 Central nervous system

The brain of *Drosophila* is estimated to have around 100.000 neurons (Pfeiffer et al., 2008). The neuronal cell bodies (also called somas) are located on the surface of the brain, which is called cortex. Axons and dendrites spread from the somas and connect to each other through a bridges called synapses. The place where synapses are present is called the neuropile. Figure 1.14 shows in an optical slice the neuropile of a *Drosophila* brain. It can be segregated in clear neuropil structures that are shown with different colors in 1.14. The brain structure is symmetrical and presents approximately the same shape among different organisms. The structures labeled in Figure 1.14 are reproducibly present among all *Drosophila* organisms. The central part of the brain is called central brain, and the two parts aside are the optic lobes. Optic lobes are directly connected to the photoreceptors of the compound eyes, and visual processing takes place in these.

From a developmental point of view, neurons are produced by a unique type of stem cell called neuroblasts. The number of neuroblasts is conserved in different insect species and is estimated to be around 100 (Younossi-Hartenstein et al., 1996) in *Drosophila*. Each neuroblast forms a uniquely identifiable cell that appears at the same time and position in every individual. All neurons stemming from the same neuroblasts typically form one bundle, called the primary axon tract (Spindler and Hartenstein, 2010). The pattern of these tracts is highly invariant among individuals. Thus, looking

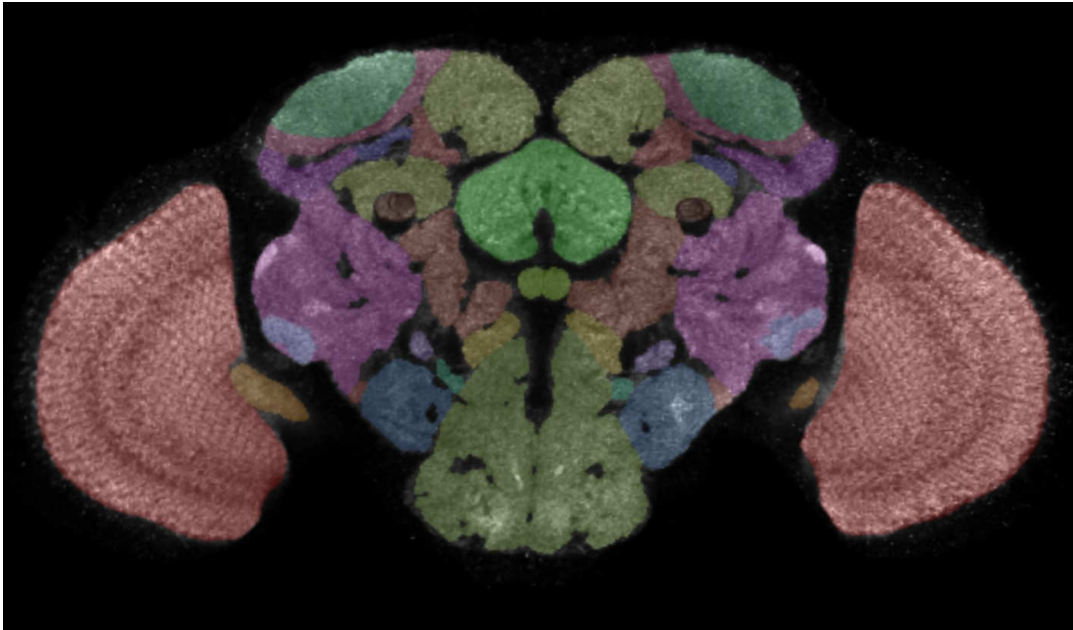


Figure 1.14: One slice of the brain of *Drosophila* (from Strausfeld (2007))

at the primary tract of a neuron should let an expert viewer determine from which neuroblast this cell stems. Whereas the location of primary tract is clearly reproducible among neurons of the same neuroblast and also among different individuals, the location of the nuclei is slightly more variable. Given this variability, it is hard to assess the identity of a neuron based on the location of its nuclei. However, the variability of the locations of the nuclei, as far as I know, has never been quantified.

### 1.6.2 Overlapping neurons

In order to achieve comprehensive functional characterization of entire neuronal circuits, specific behavioural defects have to be attributed to single neurons. To find neuronal candidates that participate in the circuit, it is essential to analyze and compare expression patterns of several lines. The goal of the project is to analyze the expression patterns of the lines that show the same behavioural defects in greater detail, because neurons responsible for the behavioural defects should be labeled in these Gal4 driver lines. However, expression patterns in single Gal4 lines are unlikely to unambiguously suggest candidate neurons, as the number of Gal4-expressing neurons vary from severals to thousands, and thus it also labels cells irrelevant for the behavioural defects. To tackle this problem, I aim to find neurons that have

## *Introduction*

been repeatedly inhibited in different Gal4 lines. These analyses should highlight individual neurons, or at least groups of neurons, that are important for a given behaviour, as they will be repeatedly labeled in lines exhibiting the same behavioral defect. In this report, I name such neurons overlapping neurons.

### **1.6.3 Existing techniques**

#### **Manual inspection**

The basic way to find overlapping neurons is a manual approach. The anatomy expert has to analyze images of brains of genetically modified flies. The number of cells expressing Gal4 in each line, varies from a few dozens to several thousands. Reminding that in addition a *drosophila* brain has around 100.000 neurons, the task of systematically find so-called overlapping neurons seems to be barely conceivable. We are convinced that we can search for these overlapping neurons only with the help of an automated approach.

#### **Automated identification of lineages**

Cardona et al. (2010) succeeded in identifying neurons based on the characteristic shape of the secondary axon tract. The same approach could be used with the primary axon tracts in adult *Drosophila* brains. However, this method relies on an automatic segmentation of the neuron axon tract. It has been possible because the authors used a genetic technique that visualizes almost only one or two neurons labeled in the whole larva brain. In our case, at least several dozens of cells are labeled, which makes the segmentation of the primary tract a very difficult image processing task. That is why I did not try to adapt this approach for our purpose.

### **1.6.4 Biological and computational pipeline for a density map of nuclei**

To find overlapping neurons among the expression patterns of different Gal4 lines, I plan to systematically investigate the expression pattern of approximately 100 different lines.

Less is known about the variability of the nuclei position among individuals. However, it seems that, after inspection of expression patterns of brains from the same line based on the nuclei position, this variability should be low enough to be able to detect a same neuron labeled in different lines, based on its location.

## 1.6 Goal of the Diploma Thesis

The first requirement is to manipulate brains that have been transformed to a generic reference brain in order to consider the nuclei locations in a unique common coordinate system. I aim to adopt a strategy based on image registration and the location of the nuclei of Gal4 expressing cells. The use of image registration of fly brains is challenging and represents the first goal. Cortex regions containing most of the nuclei of the brain being of high importance for my purpose, I will additionally try to reach a high accuracy of registration in these regions. So far, I don't have knowledge of any attempt to do it.

The second requirement is to produce an hypothesis generator that highlights regions in the reference brain, where there is a high density of nuclei. We speak of hypothesis, because a manual inspection is absolutely necessary. Considering the nuclei location variability, if two nuclei from different lines are close to each other, it does not imply that they belong to the same neuron. Inspection of the first channel (anti-GFP channel) has to be manually performed to confirm if overlapping neurons are the same or not. The hypothesis generator will be a three dimensional density map giving the probability to find nuclei of Gal4 expressing cells.

The work accomplished in this thesis will provide new informatic tools for the search of overlapping neurons among hundreds of Gal4 *Drosophila* lines.

The aims can be outlined as in the following.

- Establishment of a biological pipeline for a large scale screening of Gal4 lines
- Establishment of a computational pipeline to process image data of the screening
- Hypothesis generator: 3D density map of nuclei distribution

# Chapter 2

## Materials and methods

### 2.1 Fly strains

#### 2.1.1 Use of the Kyoto fly database

There is a collection of 4000 different GAL4 lines at the Kyoto Drosophila Resource Center. This collection is called the Kyoto NP stock collection (S et al., 2002). Each line has a different Gal4 insertion pattern. 1081 lines of this stock have been previously selected and crossed by Vladimiro Thomas with the lines presented in the next paragraph.

#### 2.1.2 Double-reporter line : nls-lacZ and myrGFP

The line pJFRC66-10XUAS-DSCP-IVS-myr::GFP has been kindly provided by Barret Pfeiffer, Rubin Lab, JFRC/HHMI (vectors described in Pfeiffer et al. (2010)). Both myrGFP and nls-lacZ are present in the genome of this line. It is therefore possible to visualize in the same brain a labeling of neurons with anti-GFP and anti- $\beta$ galactosidase. We crossed this line to the Gal4 lines from the Kyoto NP stock collection and selected the progeny for our large scale screening.

### 2.2 Fly preparation

#### 2.2.1 Brain dissection

Fly preparation is a critical step that has to be done very carefully. Brains are removed with forceps from the heads of the flies under the light microscope. Having nicely prepared brains is of major importance for the registration



step. In our case, the quadruple staining I apply requires even more attention, because the nuclei located in the cortex at the surface of the brain are labeled and of high interest. A very small tear done with a forceps may damage considerably nuclei regions at the surface of the brain, and the brain is more likely to be unsuitable for further analysis. The following protocol has been used for dissection, storage and staining.

- Dissect fly brains under Phosphate Buffered Saline (PBS) as in Wu and Luo (2006)
- Store fresh brains in ice-cold Phosphate Buffered Triton (PBT) (0.1% Triton X-100 in PBS), 1% HCHO (PCR tube)
- Fix brains in 200  $\mu$ l PBT (0.1% in PBS), 4% HCHO, room temperature (RT), agitation, 45'
- Rinse 1x and wash 2x with PBT 0.1 % for 20', RT, agitation
- Block 30', 3% Normal Goat Serum (NGS) in 0.1%PBT (blocking solution), RT, agitation.
- Prepare primary antibodies solution in blocking solution (200  $\mu$ l/ tube)
- Remove blocking solution, add antibodies solution.
- Incubate over night (ON), 4 ° C, agitation.
- Rinse brains shortly with 0.1% PBT, remove after a short mix and wash 3x with 0.1% PBT 20', 4 ° C, agitation.
- Prepare secondary antibodies solution in blocking solution (200  $\mu$ l/ tube)
- Incubate ON (at least 6 h), 4 ° C, agitation
- Rinse brains shortly with 0.1% PBT, wash 3x 10 min + 1x 1h, 4 ° C, agitation.
- Mount on microscope slide with a cover slip

### 2.2.2 Mounting

Thirty to seventy brains are mounted on a same microscope slide. They are placed on a grid I designed, as shown in Figure 2.1, in order to ease the systematic scanning of the brains. The grid is used to mount all cover glasses. The glasses have to be aligned very carefully, as it is a major cause of misalignment in the scanning step at the confocal microscope. The mounting process follows the protocol in use in the laboratory, as detailed in the following.

## Materials and methods

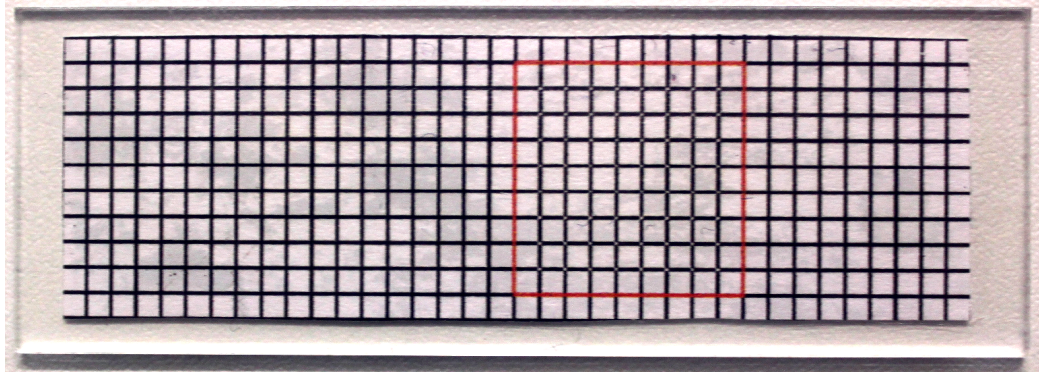


Figure 2.1: Mounting grid. Brains are placed at the lines intersections in the red square. Therefore, the position of the brains is the same for all glasses

- Prepare cover glass: label with date, genotypes
- Stick up the microscope slide to the reference mounting grid
- Prepare spacer between cover glasses with specific double-sided tape in use in the laboratory
- Put brains in PBS or PBT in a small glass bowl
- Put coverslip on glass slide
- Put a drop of Vectashield on glass slide
- Prepare a sufficient number of small droplets on coverslip at the intersections of the grid
- Transfer brains with pipette tip or loop to Vectashield droplets. Correct position and push to coverslip.
- Put second coverslip over first (with small mismatch) and mount brains with Vectashield using drops from pipette.

20 to 70 brains are mounted on each glass.

### 2.2.3 Immunohistochemistry for the large scale screening

Tables 2.2.3 and 2.2 give the antibody concentrations I used to realize the quadruple staining for the following antibodies: neuropile marker anti-synapsin, anti-GFP, anti-elav and anti- $\beta$ -galactosidase.

Antibody	Concentration
anti-GFP	1:1000
anti-synapsin	1:100
anti-elav	1:200
anti- $\beta$ galactosidase	1:25000

Table 2.1: Primary antibody concentrations added to blocking solution.

Antibody	Concentration
Alexa anti-rabbit 488	1:1000
Cy3 anti-rat	1:200
Cy3 anti-mouse	1:250
Alexa anti-chicken	1:1250

Table 2.2: Secondary antibody concentrations added to blocking solution.

### 2.2.4 Scanning with the fluorescent confocal microscope

Brains are scanned using three channels :  $488nm$ ,  $563nm$  and  $633nm$ . The step size for the acquisition is  $1.5 \mu m$  and leads to an amount of roughly 100 slices/brain.

Brain locations on the mounting glasses are the same for each glass, which enables to save once the locations of the brains. Images are acquired overnight with a speed of roughly 15 min/brain.

The same laser scanning intensity has been used to scan all brains of the large scale screening. The laser intensity of the registration channel has been manually set and did not require optimization.

The laser intensity of the GFP channel has been more carefully chosen: a Gal4 line has been selected based on a very strong GFP expression . We set the laser intensity in order to reach the limit of saturation for this brain. Therefore, saturation leading to blurry images has been avoided for the majority of the brains.

The laser intensity of the nuclei channel has been chosen similarly to the intensity of the GFP channel.

In order to compensate the intensity attenuation in deeper focal plans, we increased the intensity of roughly 10% between the top and the bottom of the sample.

## **2.3 Image processing with Fiji**

Fiji (Abramoff et al., 2004) is an open source software that aims to assist research in life sciences imaging. Its architecture provides a high extensibility with the possibility for one to write his own plugins. Fiji is written in java, and supports plugins in Java, Jython and more programming languages. It sets a powerful framework enabling the handling of stack images, consisting of slices of a biological samples.

All the plugins I implemented in this work have been written in Java and Jython, using the Fiji built-in editor. They are aimed to be used in Fiji. The major ones are listed in Table 2.4. The source codes of the plugins are freely available under GPL License in a subversion repository at <https://flylacz.svn.sourceforge.net/svnroot/flylacz>.

## **2.4 Brain registration**

Brain registration has been performed with the program BrainAligner Peng et al. (2011).

### **2.4.1 Contrast enhancement**

The staining used for the registration channel is different from the conventional one and presents less contrasts.

Images are preprocessed before the registration step. I used a local contrast enhancer (Zuiderveld, 1994) based on limited adaptative histogram equalization. It has been implemented by Stephan Saalfeld and released as a freely available plugin of Fiji. I designed a plugin that makes use of this implementation and applies the contrast enhancement to entire stacks and processes all images of our database. Table 2.3 gives the parameters I used. I used the option "Fast" to have convenient processing times. It takes approximately 5 seconds to process each brain. Figure 2.2 shows an example of an original brain and the resulting contrasted image for two slices.

### **2.4.2 Creation of the average reference brain**

#### **Reference single brain**

In a first time, a single reference brain has been selected among two hundred brains. This reference brain has to be perfectly prepared and have a nice staining. Particular attention has to be put on the quality of the nuclei labeling by anti-elav. It should be chosen as an average-looking brain, with

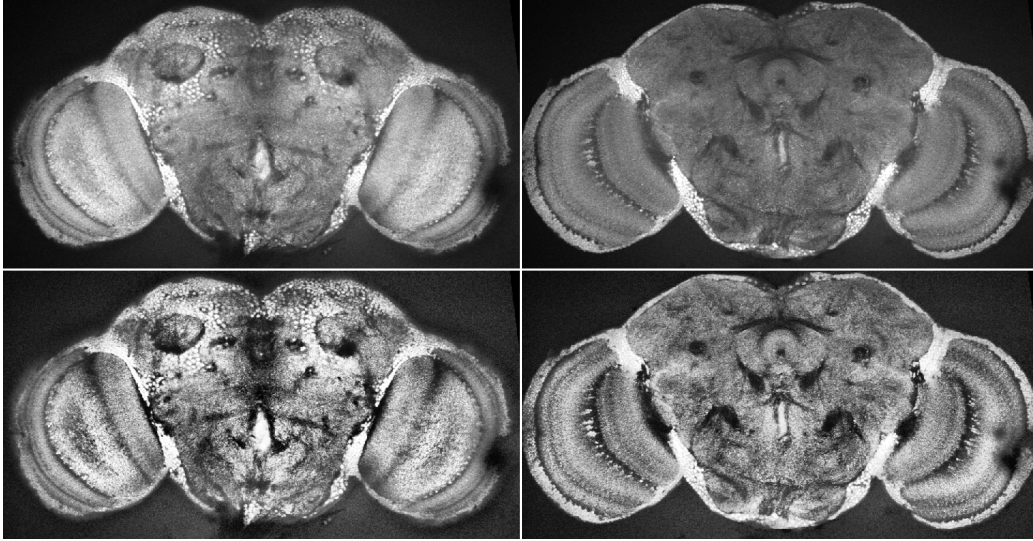


Figure 2.2: Effect of the contrast enhancement on two slices of a brain stained with anti-synapsin and anti-elav

Parameter	Value
Block size	127
Histogram bins	256
Maximum slope	3
Version	Fast

Table 2.3: Parameters for the contrast enhancement

## Materials and methods

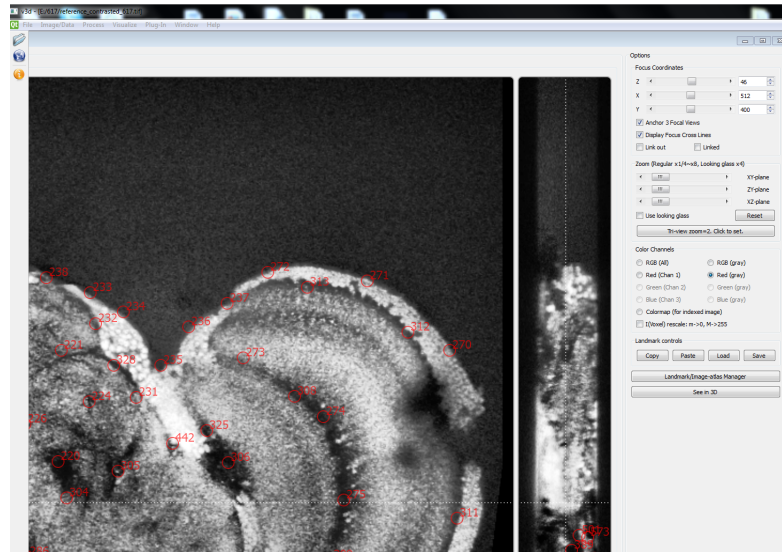


Figure 2.3: Screenshot of the V3D tool. The red circles indicate the locations of the landmarks (represented as circles, but landmarks are points)

average positions of optic lobes. A mistake would be to take a too "nice" brain that would differ too much from the majority of the brains.

### Landmarks definition

The landmarks have been defined with the image visualization and analysis system V3D (Peng et al., 2010). I defined 617 landmarks that are points with high curvature and high intensity gradient. They represent also reproducible features common to all brains. Figure 2.3 shows a screenshot of V3D during the manual selection of landmarks.

### Registration and averaging

200 brains have been registered to the single reference brain. The quality of the registration has been manually checked, and I selected 64 well registered brains. The criteria of the quality I used are : general shape, small distortions, accuracy of the neuropile structures alignment, accuracy of nuclei alignment.

The average of the 64 brains was then computed with a plugin I implemented.

Then, I applied the local contrast enhancer to this average brain, and obtained the final average reference brain that has been used in our large

2.4 Brain registration

Plugin	Type	Language	Input	Output	Processing time	Use
acquire	Fiji Plugin	Jython	.oib images, names of lines, manual evaluation of images (rotation, suppression)	8 bit images; one .tif per channel; files sorted in folders	Immediate	first step of acquisition
mergeAll	Fiji Plugin	Jython	folder of sorted monochannel images (lines-index-channel.tif)	folder of sorted RGB images (line-index.tif)	few sec/image	before registration (input of brain aligner)
splitAll	Fiji Plugin	Jython	folder of sorted RGB images (line-index.tif)	folder of sorted monochannel images (lines-index-channel.tif)	few sec/image	after registration
register	Linux script	bash	hard drive containing folder of sorted RGB images (line-index.tif), folder containing reference brain and landmarks	hard drive containing registered images and retained landmarks	4 sec/landmark	registration
evaluate	Fiji Plugin	Jython	reference; folder of sorted monochannels images (line-index.tif), manual evaluation of quality	folders of good and bad registered brains	Immediate	after registration
removeNuclei	Fiji Plugin	Jython	nuclei file; images directory	removed nuclei file (based on Glia or GFP signal)	few seconds/brains	after nuclei detection
densityMap	Fiji Plugin	Jython	nuclei file; images directory	interactive density map	2sec/line to draw	final visualization

Table 2.4: Plugins and bashes implemented and available at the subversion repository.

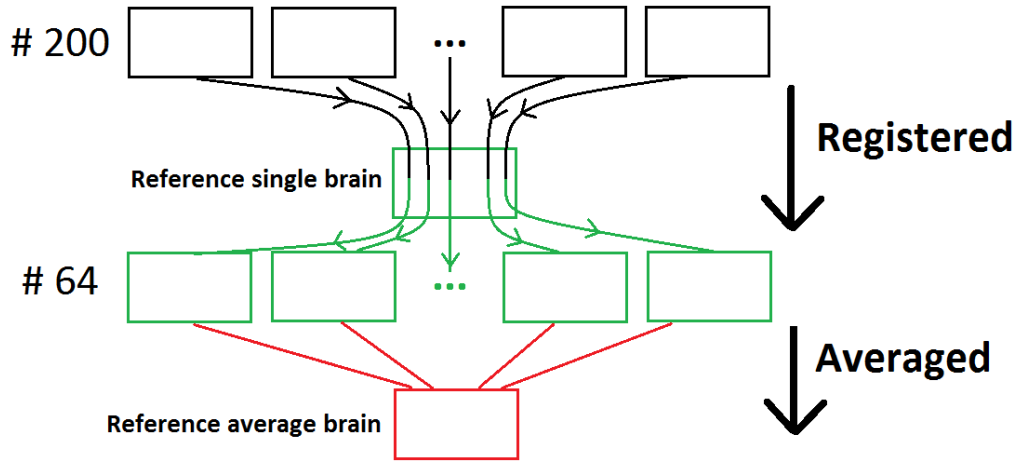


Figure 2.4: Schema of the process of creation of the average reference brain

scale screening. The process is summarized in Figure 2.4 and the average reference brain is shown on Figure 2.5.

### 2.4.3 Large scale registration

For the large scale registration, the average reference brain has been used as target brain. We defined 628 landmarks on this image. I wrote a script that takes advantage of multi-core computers and optimizes the memory consumption. On a 2x2.33 GHz, 8GB RAM, we registered approximately 3 brains/hour. The script is user-friendly and does not require any programming skills.

## 2.5 Nuclei detection

I used an automated nuclei detector to distinguish nuclei in confocal images of brains stained with anti- $\beta$ galactosidase and brains stained with anti-repo. The algorithm has been initially developed (Santella et al., 2010) to detect blobs in confocal images of C-elegans. The software package is written in Matlab MATLAB (2010) and is freely available under the GNU General Public License (GPL) at <http://sourceforge.net/projects/starrynite/files/blob-slice%20cell%20detection>. I changed several parts of the code to suit our purpose. The modified version is also available under the GNU General Public License (GPL) at <https://flylacz.svn.sourceforge.net/svnroot/flylacz>.



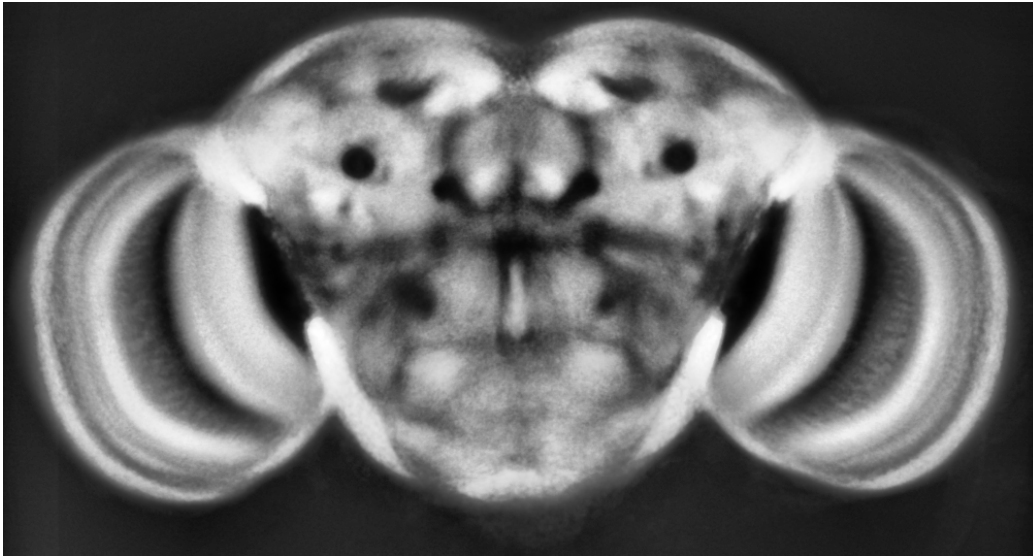


Figure 2.5: Average reference brain used for the large scale screening.

The most important parameter that has been changed is the intensity threshold for noise detection. The Table 2.5 gives the list of parameters and corresponding values I used and changed for the detection of nls-lacZ expressing cells and glial cells.

## 2.6 Glia detection and removal

I developed a method that makes use of the specificity of the elav antibody that labels exclusively neurons. The method relies on the characteristic intensity decrease of the staining at locations of glial cells in dense labeled nuclei regions. The technique uses local tests based on concentric 3D rings splitted in quadrants. Two major parameters have to be set by the user. The first one is the intensity threshold between the inside and the outside of nuclei. The second is the score of ring quadrants satisfying the threshold criteria. The parameters I used to ensure a glia removal with a very low probability of false positives are given in Table 2.6. As it will be discussed in 3.5.2, the removal rate is rather low with these safe parameters. The implemented plugin takes as inputs the nuclei data file, and the registration channel. The output is the nuclei data file containing only neurons. The plugin has been implemented in Fiji.

Parameter	Value for nls-lacz nuclei	Value for glial cell
xy resolution	0.69	0.69
z resolution	1.5	1.5
nucleus diameter	7	9
parameters staging	[200,500,800,1500]	[200,500,800,1500]
noise intensity threshold	[2,3,4,5,6,7]	[15,19,19,21,23,25]
filter size	1	1
distance to merge overlapping nuclei	[4,3,6,3,4,3,2,3,3]	[2,1,8,1,7,1,6,1,5,1,5]
mergescore	[-300,-200,-100,-35,-15,-20]	[-400,-300,-200,-95,-75,-60]
split score	[100,100,20,19,17,5]	[200,200,80,78,50,20]
merged aspect ratio	[2,2.5,2,1,8,1,4,1]	[2,2.5,2,1,8,1,4,1]

Table 2.5: Main parameters used for the detection of nls-lacz expressing cells and glial cells

## 2.7 Nuclei removal based on GFP expression

Parameter	Value
Threshold	30
Score	7

Table 2.6: Parameters used to remove glia based on the registration channel

Case	Parameter	Value	Case	Value
Safe	Threshold	8	Less safe	9
	Number of voxels below threshold	3		30

Table 2.7: Parameters used to remove lacz nuclei based on the GFP expression

## 2.7 Nuclei removal based on GFP expression

Similarly to the precedent paragraph, I aimed to reduce the number of nuclei detected in the nuclei channel, based on the GFP expression. Detected nuclei with showing a lacZ signal but no GFP signal on the first channel are removed. For this purpose I implemented a plugin that mainly takes two parameters as input. The first is the intensity threshold  $t$  in the GFP channel that a voxel located in a detected nuclei should have. The second parameter is the number of voxels  $n$  of a nuclei that should have an intensity greater than the intensity threshold. That is, nuclei that possess more than  $n$  voxels with an intensity greater than  $t$  are kept, and the others removed. I used at least two sets of parameters that are given in Table 2.7. The first set corresponds to a very safe removal, putting priority on the avoidance of false positive removals. The second set is less safe and can be used to dramatically decrease the number of nuclei visualized in the density map.

# Chapter 3

## Results

### 3.1 Quadruple staining

#### 3.1.1 Absence of cross reactions between antibodies

We realized an immunohistochemical quadruple staining that makes use of the three available channels of the microscope in order to obtain as much information as possible. The antibody combination we retained is described in 2.2.3. We did not notice any cross reaction between the antibodies. Figure 3.1 shows the three channels acquired with the confocal microscope, merged on the same image with different colors. Figure 3.2 shows the channels on three separate images.

This staining quality has been reached reproducibly for the majority of the brains we prepared for the large scale screening.

#### 3.1.2 Optimization of dilution factors for the third channel

The staining conditions and the protocol for the staining with the antibodies of the two first channels were well established in our laboratory and provided already good results (Rabbit anti-GFP, Rat anti-elav, Mouse anti-synapsin). For the last channel with the anti- $\beta$ galactosidase staining, I had to optimize the dilution factors for the primary and secondary antibodies. The common method to establish a protocol for antibody staining is to look at publications where the antibody is used, to have a rough idea of the dilution factors needed (Simone and DiNardo, 2010; Mukherjee et al., 2011; Kolahgar et al., 2011), and then to adjust the parameters starting from these conditions.

Using the conditions used in the literature lead us to unsatisfactory results that could have impaired the automated nuclei detection. Figure 3.3

### 3.1 Quadruple staining

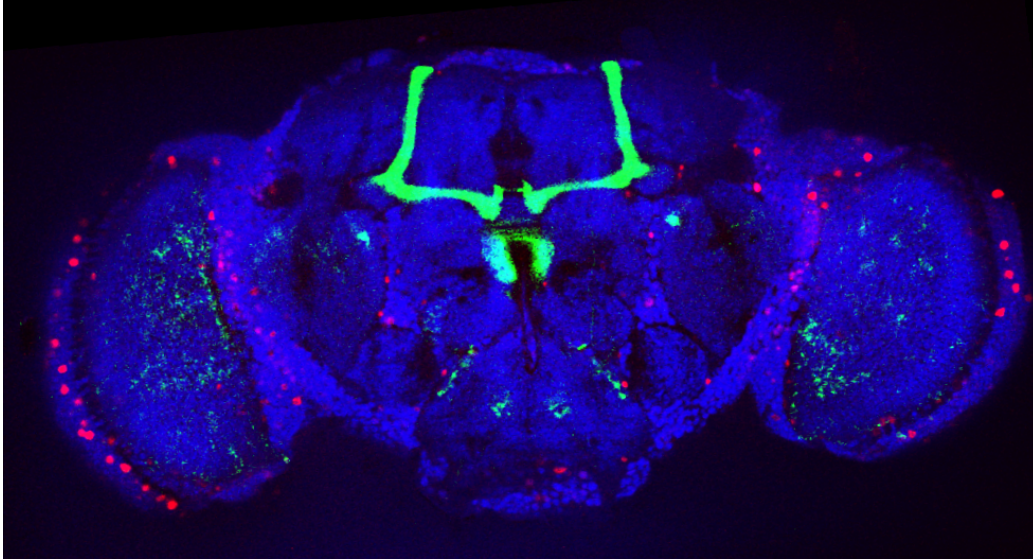


Figure 3.1: Three channels of the quadruple staining merged on a same image. Green : GFP channel (named also first or green channel); Red : registration channel (named also second or reference channel); Blue : anti- $\beta$ galactosidase channel (named also third or lacz or nuclei channel)



Figure 3.2: Three channels of the quadruple staining. Left : GFP channel; Middle : registration channel; Right : anti- $\beta$ galactosidase channel

## Results

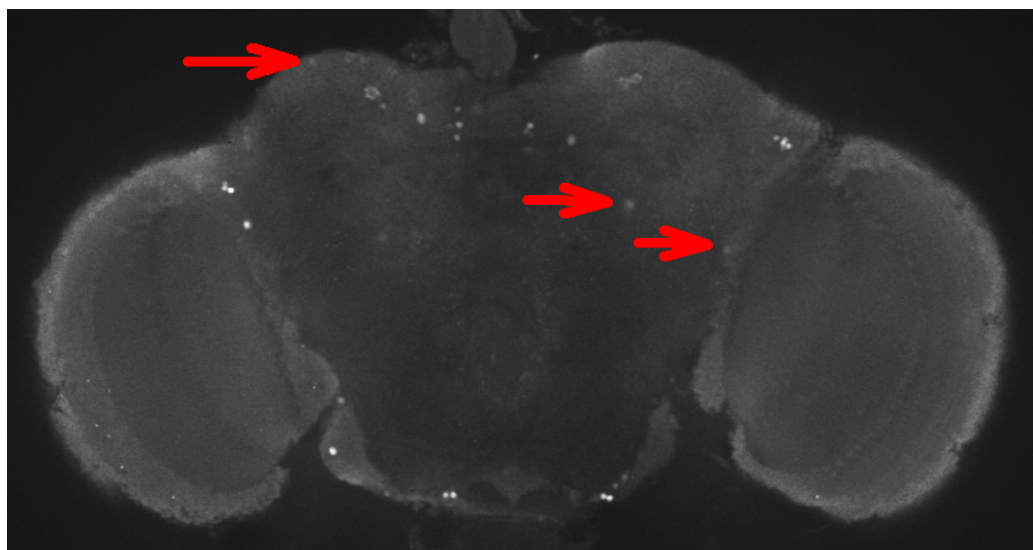


Figure 3.3: Nuclei channel with concentrations of 1:250 for primary and secondary antibody

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
Primary antibody	1:1000	1:1000	1:5000	1:25000	1:25000	1:100000
Secondary antibody	1:250	1:1250	1:1250	1:1250	1:250	1:250

Table 3.1: Antibody concentrations tested for the optimization of the anti- $\beta$ galactosidase staining

shows a slice of a brain stained with concentrations of 1:250 for primary and secondary antibody. The background seemed to be far too high, and we may have lost some lacz signal, as shown by the arrows.

To find the optimal conditions, we tested several dilution factors combinations on two different lines. One line has been selected with a high expression intensity and the second one with a low expression intensity. Table 3.1 gives the dilution factors we used.

At this time, we did not know what technique we would use for the automated detection of nuclei, and thus, what would be the requirements for a good staining. We simply seeked to have a high signal to noise ratio, that is, having a low background with a good signal. The combination number 4 gave the better results and can be seen on Figure 3.4 on the bottom right. The background is low, and the signal is high.

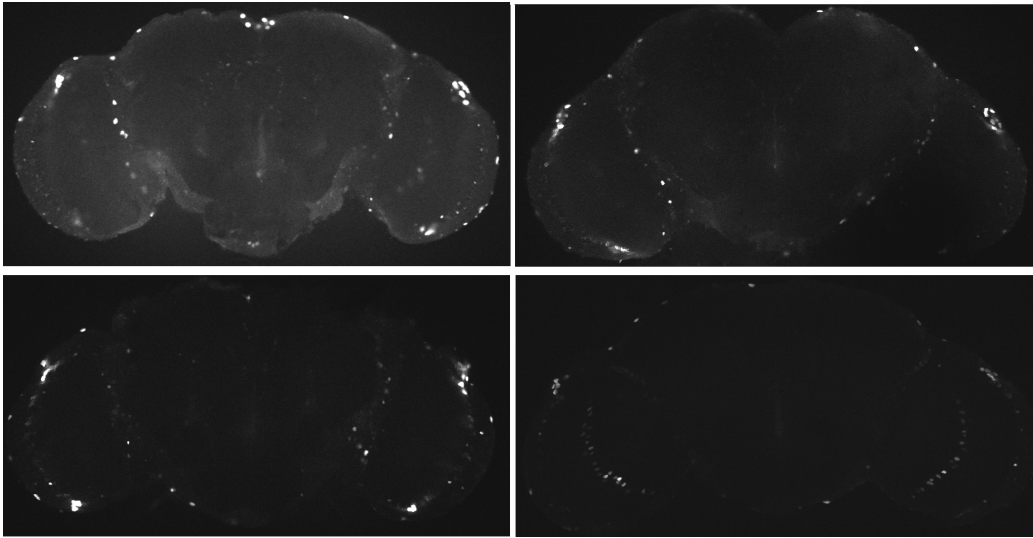


Figure 3.4: Slices of brains stained with different antibody concentrations from Table 3.1. Top left: Test 1; Top right: Test 2; Bottom left: Test 3; Bottom right: Test 4

## 3.2 Registration

### 3.2.1 Overview

So far, 197 brains have been registered to the average reference brain. A high acceptance rate has been reached for the pairs of landmarks retained with  $67\% \pm 19\%$ . Figure 3.5 shows the histogram of the landmarks acceptance rates. In Peng et al. (2011), the authors claim that transformations with an acceptance rate above 75% are good. In our case, 67% of the registered brains have a score above 70%. I manually checked the quality registration of the 197 registered brains. It led me to discard 75 brains and to keep 122 brains. It corresponds to a success rate of 62%, which is similar to the value given by the acceptance rate.

### 3.2.2 Nuclei registration

A reason why I decided to use the antibody anti-elav, is that this labeling is useful for the registration of nuclei regions. The classical staining used for fly brain registration is a neuropile staining, which does not label cortex regions at the surface of the brain. We were able to define landmarks in characteristic nuclei regions, and thus we took into account location information that would not have been possible with a simple neuropile staining. The quality of the

## Results

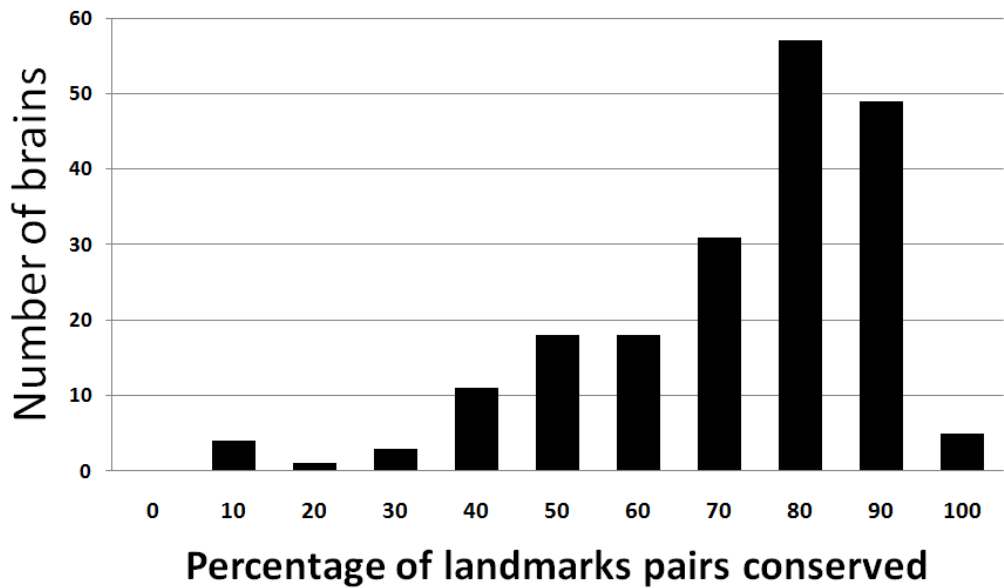


Figure 3.5: Histogram of landmarks pairs conserved

nuclei registration can only be assessed by manual evaluation and I estimate having maximal errors around  $3\mu\text{m}$  among the 122 successfully registered brains. This value is far below the nuclei location variability. Figure 3.6 shows examples taken from several successfully registered brains with a high accuracy in nuclei regions.

### 3.2.3 Brains partially damaged

As claimed in Peng et al. (2011), BrainAligner succeeds in aligning brains that have been partly damaged. The case I encountered most is that an optic lobe has been cut during preparation. Figure 3.7 shows an example of registration that succeeded with an optic lobe that has been cut. In cases where almost no brain remains in a given line at the end of the selection process (after preparation and after registration), we can still use images with brains partly damaged. It gives reliable information on only one side, that can nevertheless be used in the density map. It can avoid us having to recross and reprepare brains of a specific line, which takes more than three weeks.



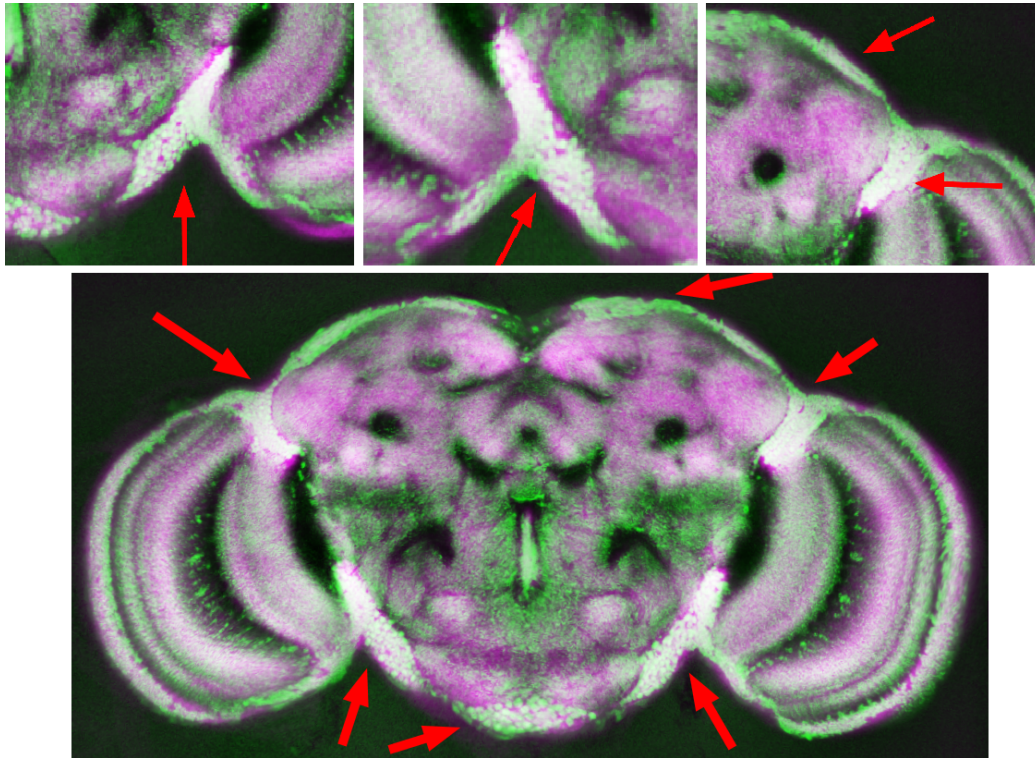


Figure 3.6: Some examples of the accuracy of the registration in nuclei regions. Green: transformed brain; Magenta: reference brain

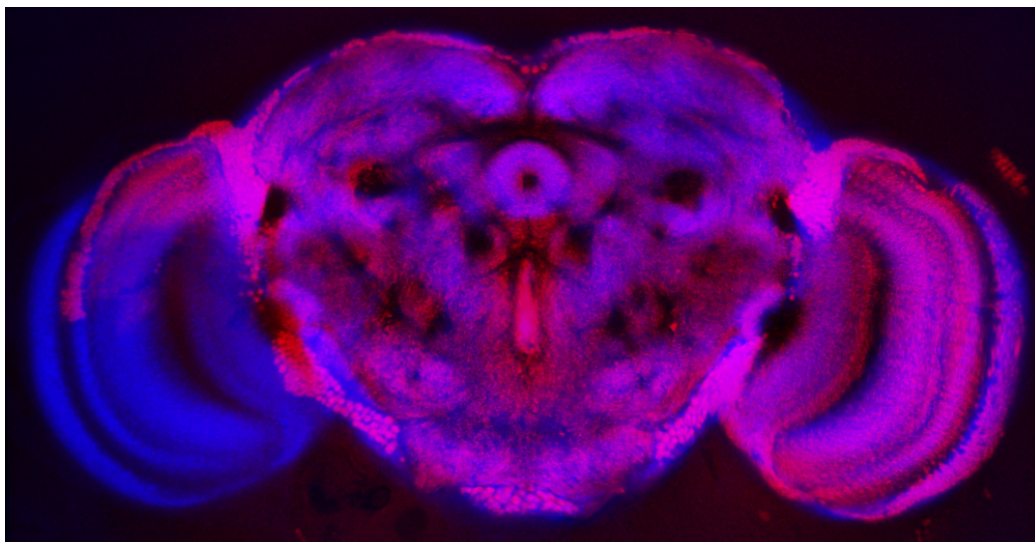


Figure 3.7: Successful registration of a brain with only one optic lobe to the reference average brain.



Figure 3.8: Brain with with clearly rotated optic lobes

### **3.3 Removal of optic lobes**

#### **3.3.1 Technique employed**

As mentioned in 1.5.3, all registration algorithms are done in two steps. First, an affine nonrigid transformation is computed, then a nonrigid transformation (warping) is applied. Images of whole adult brains have a high variability in the orientation of the optical lobes. They are weakly bound to the central brain. That is why we often have to deal with images where optic lobes are twisted, like in Figure 3.8 . In this project, we may focus our investigations first on the central brain, as candidate neurons in the optic lobes would principally show that the tested flies had an abnormal behavior because of an impairment in the optic perception circuits. However, bad oriented optic lobes can have a substantial effect on the first step of the registration process, namely the rigid transformation. Such a problem is shown in Figure 3.9.

For the calculation of the best rigid transformation, hundreds of different sets of parameters (scale, translation, rotation) are compared, and the set giving a brain that overlaps the reference brain with a high score (based on mutual information, or on correlation) is retained. That is why in the case of Figure 3.9, the transformed brain has been too much translated to the top.

Removing the optical lobes physically is barely conceivable. The brain preparation is a crucial step, and preparing registerable whole brains is al-

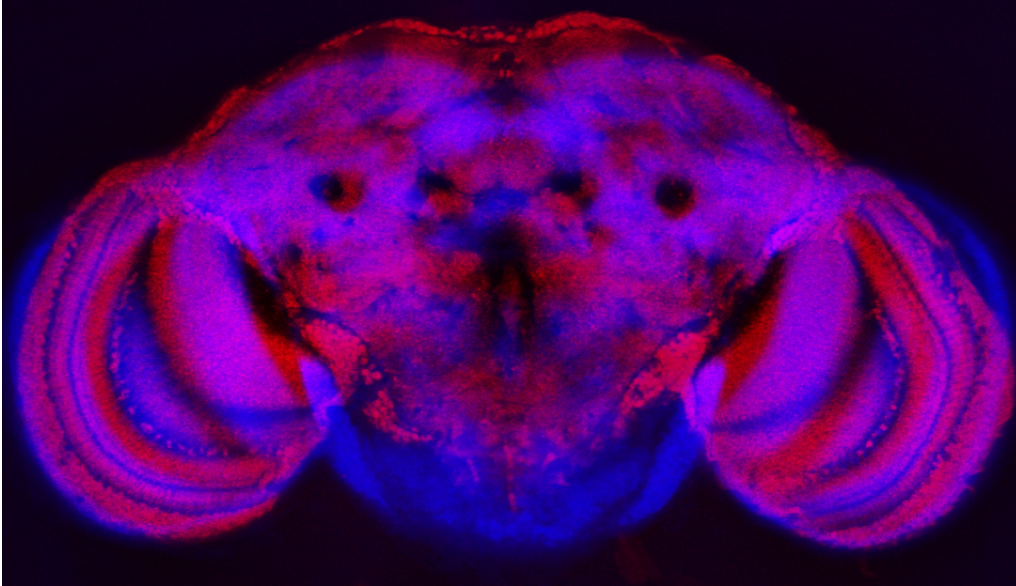


Figure 3.9: Result of affine registration. Blue: target brain; Red: result of the transformation of the subject brain

ready a complicated task. A way that has been investigated, is to computationally cut the optic lobes in the confocal images. The only attempt I am aware of has been published in Lam et al. (2010). The authors presented two different approaches.

The first approach is based on minimisation of a three dimensional energy function created from segmentation of 2D slices. They show that this method is more reliable, but its implementation is not freely available.

The second approach is more intuitive and is based on the concept of shortest path detector. The idea is to define four points on an image brain that are the inputs of the shortest path detector. The algorithm finds the path having the smallest intensity possible between two pairs of these points, that corresponds to the border of the central brain with the optic lobes.

I have used a shortest path detector implemented by Benjamin Schmidt. It is an implementation of the Dijkstra algorithm (Dijkstra, 1959) in java, and can be used in Fiji. The only important parameter is the weight of the distance compared to the intensity. With a high value of this parameter, the algorithm tries to minimize in priority the distance, and the path may go through high intensity regions. With a low value, the algorithm tries to minimize the intensity, and the path may be rather long. We fixed manually the best value after evaluation of the process applied to different brains. The best results are obtained with a value of 0.1. Figure 3.10 shows how the four

## Results

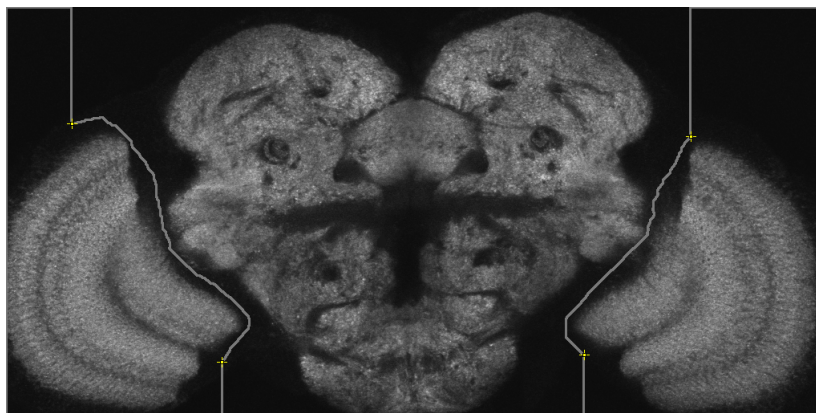


Figure 3.10: Definition of the four points needed for the optic lobes removal. The line in gray is the shortest path found.

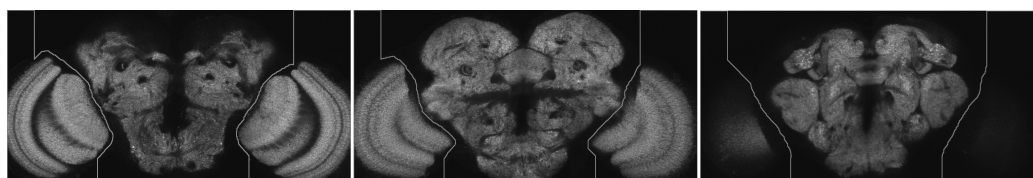


Figure 3.11: Exemples of optic lobes removal in three slices of a same brain.

points are defined, and the shortest path found by the algorithm. Figure 3.11 shows the results for three slices of the same brain. We had similar results for the majority of the tested brains, which was quite satisfactory.

### 3.3.2 Reasons for not using this plugin

The plugin implemented in this section has not been used in the large scale screening, for several reasons. Firstly, the release of the BrainAligner program in May 2011 enabled me to overcome in most of the cases the problems presented here, that we could not have avoided with other brain registration programs we tested. Secondly, this technique requires to have a dark separation between the central brain and the optic lobes, which is not the case with the staining we retained, as it can be seen on Figure 3.2. Thirdly, since our approach for finding overlapping neurons is based on nuclei locations, we have to pay a particular attention to the nuclei present on the pathway determined by the algorithm. The pathway may cut at the wrong place, removing nuclei belonging to the central brain. Figure 3.12 shows the computed separation that has been calculated on a channel stained with synapsin (in red). The image has been merged with an elav staining that labels nuclei (green). We

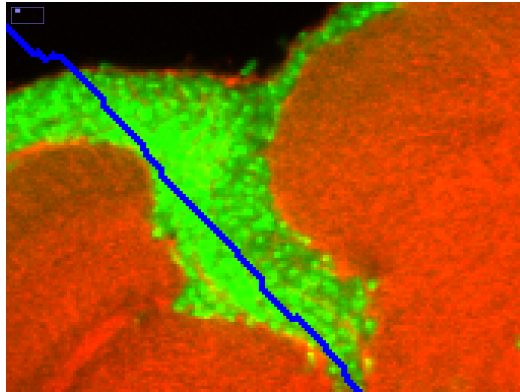


Figure 3.12: Shortest path computed by the Dijkstra algorithm on a synapsin channel. The elav channel (green) has been merged to the image for visualization of nuclei. The path may remove nuclei belonging to the central brain.

see that it is hard to determine a clear separation between nuclei belonging to the central brain and nuclei belonging to the optic lobes.

## 3.4 Nuclei detection

### 3.4.1 Quality

The nuclei detection has been performed on the third channel (nuclei channel). The parameters I used have been chosen based on manual evaluation. After manual inspection, I did not find any false negative detection. Figure 3.13 shows an example of detected nuclei. The detected nuclei are represented in magenta and the original image is in green. Nuclei have been drawn with a greater diameter than the one detected to be well visible on the picture. The white overlay of green and magenta shows that the nuclei have been successfully detected.

After manual inspection, I did not find any false negative detection, which means that all nuclei are detected. Nevertheless, I've observed some false positive detections, that more precisely are double detections. It happens for nuclei elongated in the z-axis, which leads to a double detection in two distinct z slices. I estimate the happening of false positive detections around 1/500 after evaluation of several brains, which is satisfactory. Moreover, the effect on our final goal of display of density map is insignificant. One nuclei out of 500 displayed twice instead of once, does not change the information given by the density map.

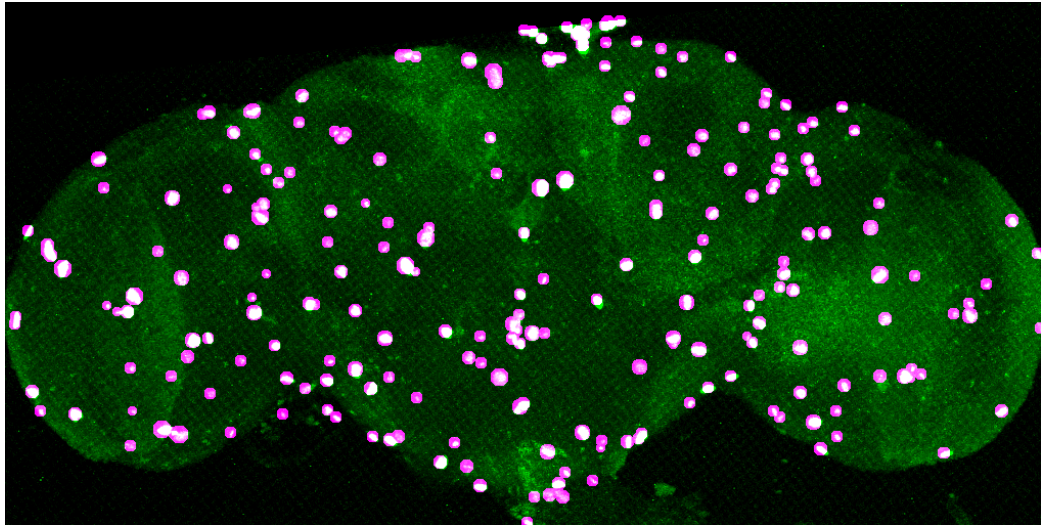


Figure 3.13: Maximum intensity Z projection. Merge of detected nuclei on the third channel (magenta), and third channel (green). The white overlay of green and magenta shows that the nuclei have been successfully detected.

### 3.4.2 Statistics

Figure 3.14 shows the average number of nuclei detected per line with the standard deviation. The lines are referenced by numbers on the x-axis. The standard deviation is drawn on the corresponding bars (if more than one brain is available per line). The small values of the standard deviations confirms us that the nuclei detection is coherent and satisfactory.

The average number of detect nuclei in the third channel is 1600. The maximum and minimum average numbers of detected nuclei are respectively 9713 and 79. Figure 3.15 shows the nuclei and GFP channels in a 3D reconstruction of a brain with 9895 detected nuclei. It represents approximately one tenth of all nuclei of the fly brain. Figure 3.16 shows the nuclei and GFP channels in a 3D reconstruction of a brain with 1159 detected nuclei.

## 3.5 Nuclei variability

### 3.5.1 Preliminary results

After three months of investigation, I obtained the first results that showed that the nuclei variability was low enough to have a chance to succeed with the nuclei-based approach for finding overlapping neurons among different Gal4 lines. The preparation conditions were not optimal thus the quality of

### 3.5 Nuclei variability

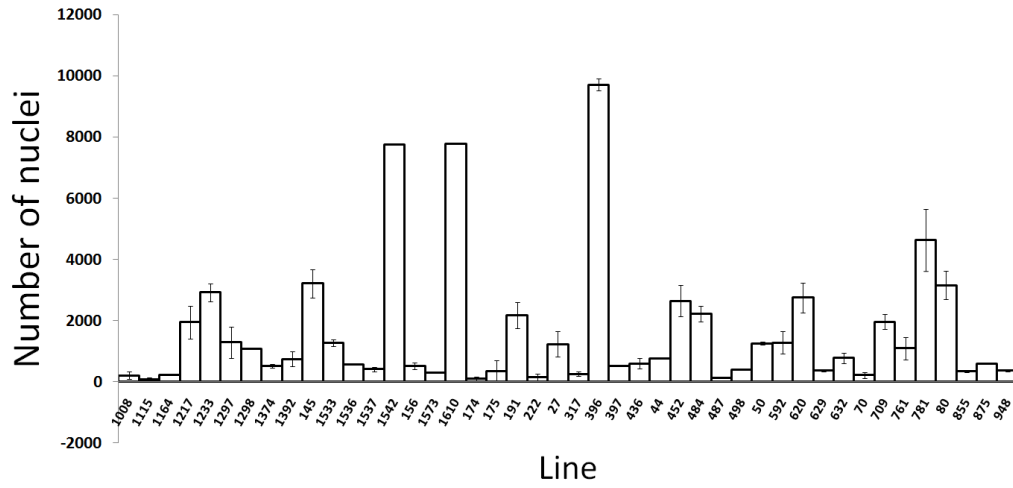


Figure 3.14: Number of nuclei detected in the third channel of several brains of different lines. Lines are referenced with numbers on the x-axis.

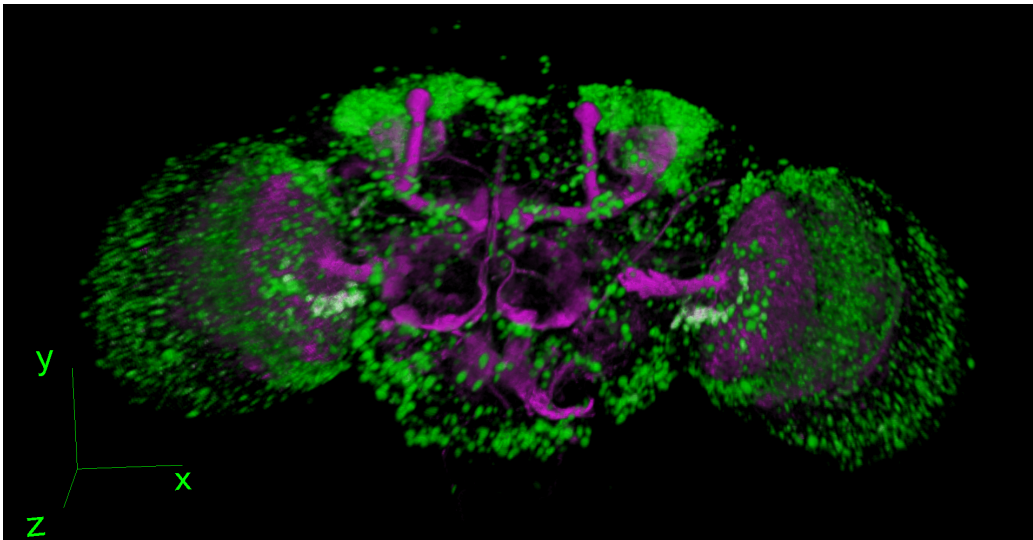


Figure 3.15: Nuclei (green) and GFP (magenta) channels in a 3D reconstruction of a brain with 9895 detected nuclei.

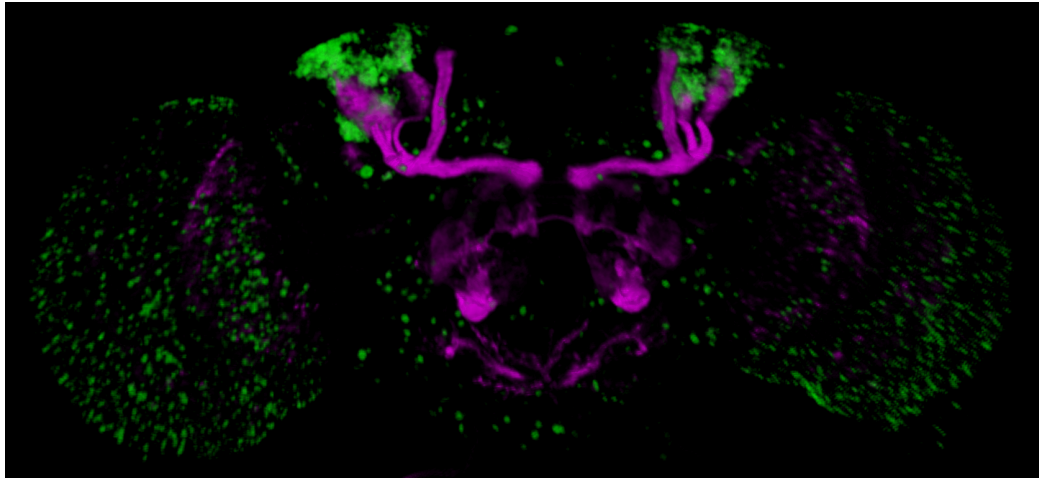


Figure 3.16: Nuclei (green) and GFP (magenta) channels in a 3D reconstruction of a brain with 1165 detected nuclei.

the registration was also not optimal because of high distortions. Moreover, the registration had been performed with the first registration algorithm I tried (Elastix (Klein et al., 2010)) and I did not optimize the registration parameters. Nevertheless, the accuracy was good enough to assess on the chance of the approach to succeed. Figure 3.17 shows this result.

### **3.5.2 Results**

Figure 3.18 shows the third channel of three brains of the same line. It appears clearly that the neurons shown with the arrows are reproducibly in the same region in three different brains of the same Gal4 line. However, it is hard to evaluate the variability of nuclei in most of the lines.

Another way to assess the nuclei variability is to look at lines having clusters in their expression patterns. Figure 3.19 shows the nuclei channel of three different brains of the same line. The cluster structures (mainly composed of Kenyon cells) appear in each brain and clearly overlap. In the rest of the brain, this image shows how difficult it is to assess the nuclei variability in dense nuclei regions.



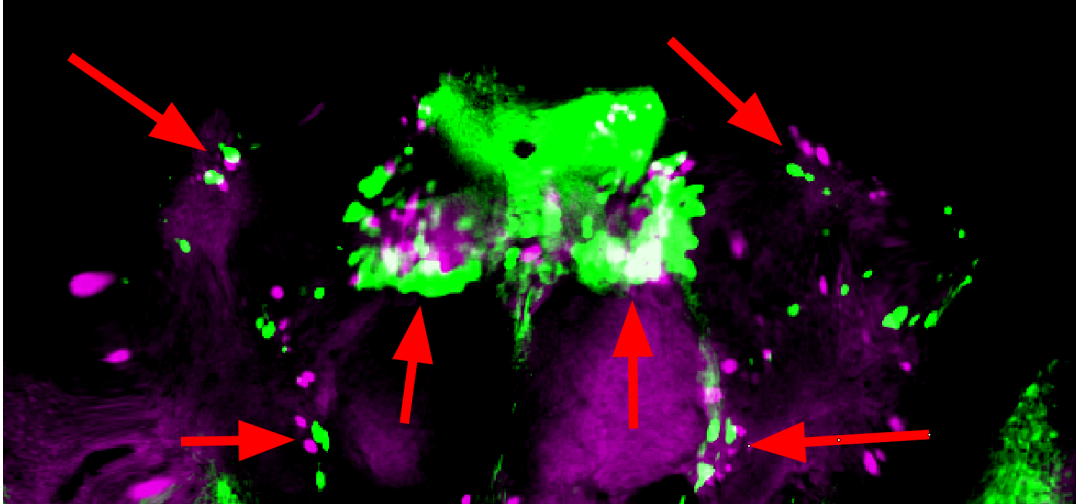


Figure 3.17: Anti- $\beta$ galactosidase channel of one brain registered to another brain of the same Gal4 line (DDC). The arrows show locations where nuclei are thought to correspond to each other and are close to each other.

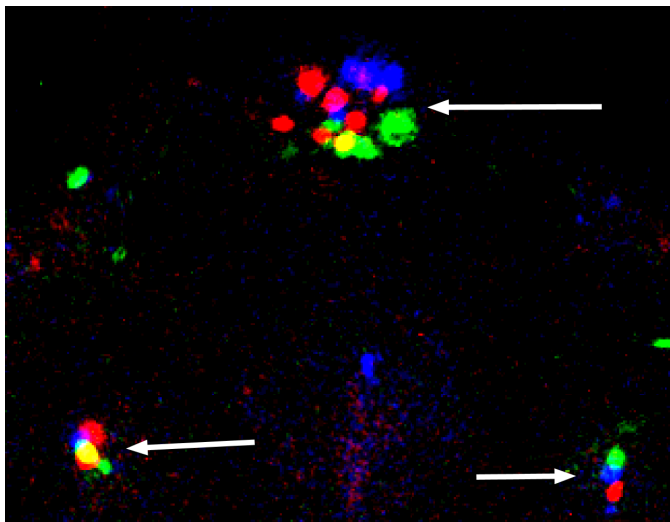


Figure 3.18: Anti- $\beta$ galactosidase channel of three brains from the same line registered to the average reference brain. The arrows show locations where nuclei are thought to correspond to each other and are close to each other.

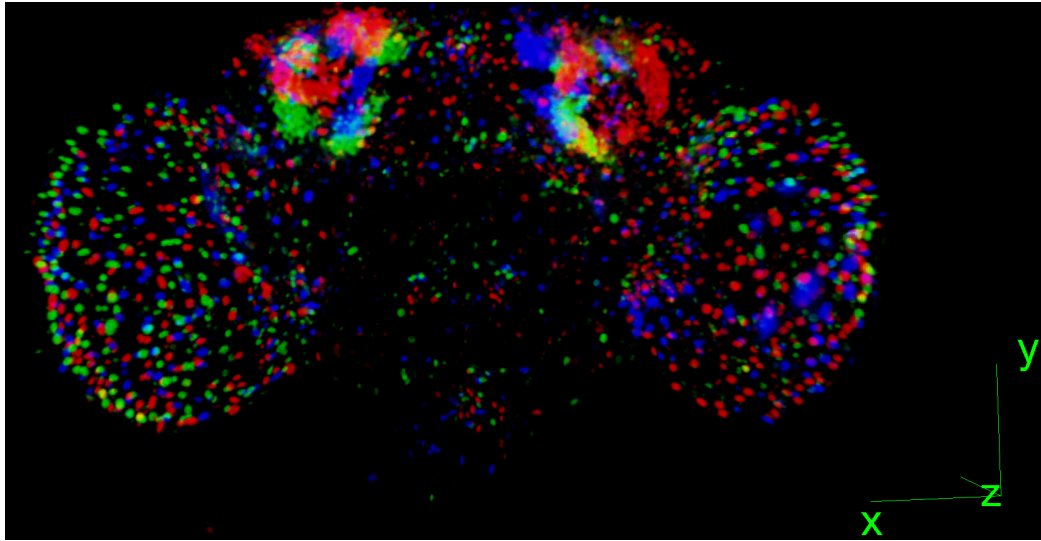


Figure 3.19: 3D reconstruction of the Anti- $\beta$ galactosidase channel of three brains from the same line registered to the average reference brain. The clear clusters enable to judge the nuclei location variability.

## 3.6 Detection and removal of glial cells

### 3.6.1 Goal

### 3.6.2 Intensity profiles of glial cells in the registration channel

#### Staining used to test the technique

In order to test our glia removal technique, we did not use the same staining as the one we used for the large scale screening. This staining and the following ones are reminded in Table 3.2. The ideal staining we should have used to test our glia removal technique is anti-repo on the first channel, to reliably see the glia, and anti-synapsin and anti-elaV on the second channel (Staining 1 in Table 3.2), namely, the same as for the large scale screening. However, this antibody combination is not possible as both anti-repo and anti-synapsin have been raised in mouse. After intensive search, it seems that anti-repo is commercially only available from the mouse. Thus, in order to use a neuropil staining, I should have found a neuropil marker raised in another animal than mouse. The only antibody satisfying this condition, as far as I know, is NCadherine (Iwai et al., 2002), raised in rat (Staining 2 in Table 3.2). The problem is that this antibody is far weaker than synapsin, as

### 3.6 Detection and removal of glial cells

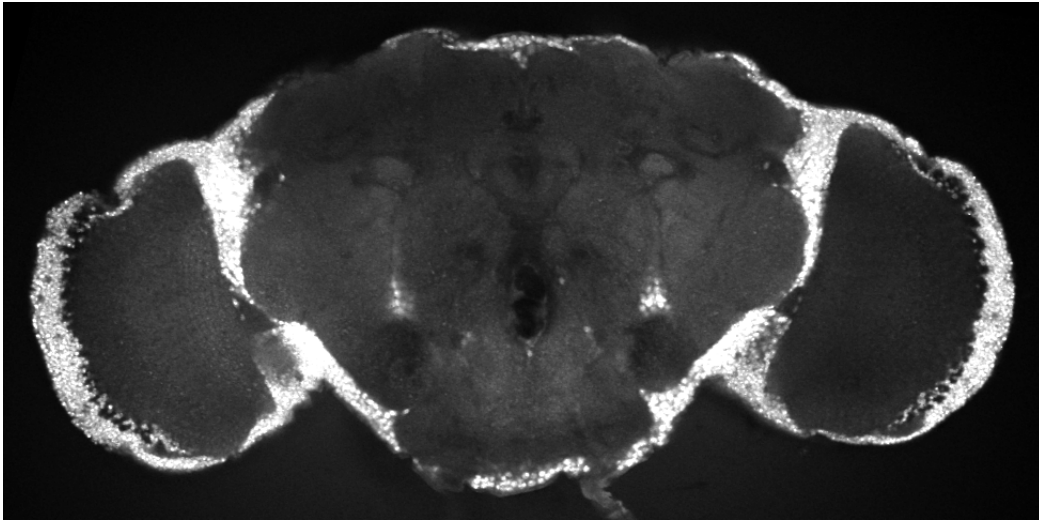


Figure 3.20: Brain stained with antibodies anti-NCadherine and anti-elav. The signal of anti-NCadherine is very weak compared to the signal of anti-elav.

	Channel 1	Channel 2	Channel 3
Large scale staining	anti-GFP	anti-synapsin; anti-elav	anti- $\beta$ galactosidase
Staining 1		anti-repo (m)	anti-synapsin (m)
Staining 2		anti-repo (m)	anti-NCadherine (r)
Staining 3		anti-repo(m)	anti-synapsin(m)

Table 3.2: Large scale staining and three possible stainings for the investigation of glias.

explained in 1.3.2, thus setting a high laser intensity is necessary to obtain a good signal. Doing this leads elav to saturate, elav being also rather strong. Such a combination is shown in Figure 3.20. A solution would be to adjust the secondary antibody dilutions, but both NCad and elav have been raised in rat, and thus are labeled with the same secondary antibody. Finally, elav is also available from the mouse (Overstreet et al., 2003), but it does not help since anti-repo is raised in mouse(Staining 3 in Table 3.2).

That is why we used the combination anti-repo on one channel and anti-elav only on the other channel. As we will see later, we will be able to remove glia located in the cortex only, that is, outside the neuropil staining. Therefore the non presence of neuropil staining do not prevent our investigation to be valid with the staining retained for the large scale screening.

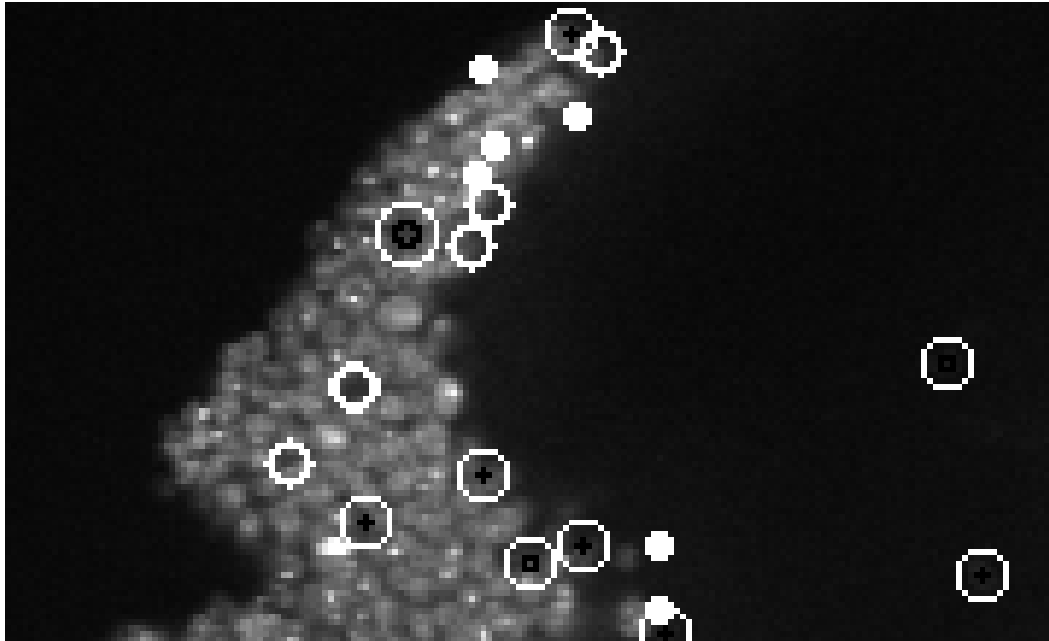


Figure 3.21: Concentric rings used to compute the average intensity. White:  $r(7)$ , black:  $r(4)$

### 3.6.3 Methods

#### Average intensity of concentric rings

The goal is to detect intensity decreases in blobs of different sizes detected by the nuclei detection algorithm. We implemented a new technique suited to our purpose. As observed in the *elaV* channel in the cortex regions, glia are characterised by an intensity decrease in the cell body because anti-*elaV* labels specifically neurons and not glia. The approach consists in comparing the average intensity outside the dot with the average intensity inside the dot. To do this, we use concentric rings centered on the location of the center of mass given by the nuclei detector, and compute the average intensity of these rings. The fifth ring gives for example the average intensity of all pixels located at a distance of 5 voxels to the center of mass. On Figure 3.21, the sixth and third rings are drawn respectively in white and black around detected nuclei.

#### Intensity profiles of cortex glia on the registration channel

With the information given by the *repo* channel, we manually selected different locations of glia. Then, based on the *elaV* channel, we selected a rep-

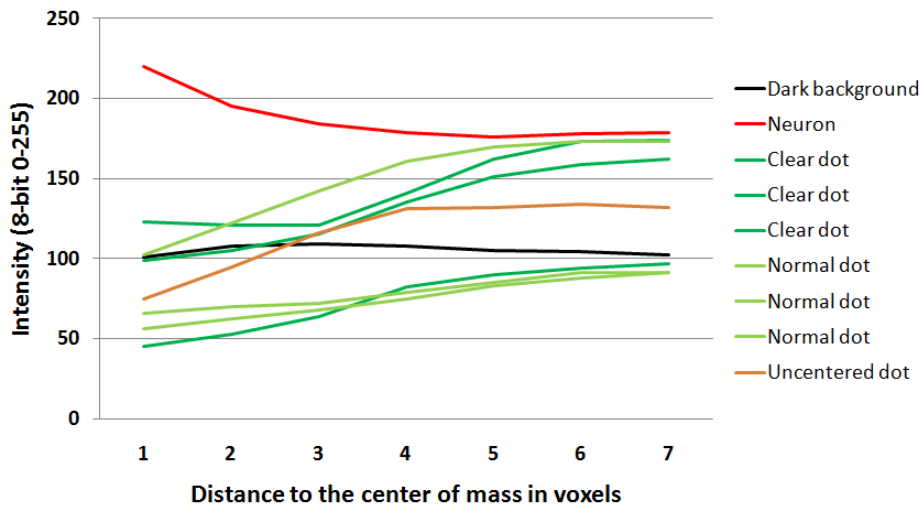


Figure 3.22: Intensity profiles of representative locations based on average intensity of concentric rings

representative sample of dots having different intensity profiles, that are shown in Figure 3.22. The x-axis is the radius of the ring (in voxels) centered on the center of mass of the nuclei, and the y-axis is the average intensity value of the corresponding ring.

The dots have, as expected, profiles with a clear intensity increase from the center of mass to the outside. We named "clear dots", dots where we visually observed a sharp intensity decrease. "Normal dots" presented an average intensity decrease.

A representative intensity profile of a visible neuron can be seen at the top of the graph. It shows a clear intensity increase in the center of mass. In orange, the intensity profile of an uncentered dot is drawn. We manually chose a translated center of mass, to simulate an inaccuracy in the automatic determination of the center of mass, which might occur in the real lacZ nuclei detection. There is a sharp intensity increase between the third and fourth ring due to the translation.

Random places in the background have approximately constant profiles, as represented in dark.

We can conclude from this graph that glia intensity holes have a characteristic feature that can be revealed by the method of concentric rings. The position of the glia is complementary to the neuronal nuclei and thus computationally detectable. The next steps define a threshold and then validates the choice of the best value to avoid false positive detections, namely, detecting real neurons as glia.

### **3.6.4 Threshold rough estimation**

Places with glial cells may appear more or less bright on the registration channel, depending on the staining and the location. We need therefore a local relative criteria to discriminate the dots, instead of a simple absolute threshold.

The basic idea is to compare the intensity in the neighborhood outside the dot and the intensity inside the dot. A critical parameter to set is the definition of the neighborhood and the inside of the dot. Given the computed diameter  $d$  by the nuclei detector, we have several possibilities to define the neighborhood. The neighborhood could be defined as the voxels at a distance of  $(d + \alpha)$  or  $(\alpha * d)$  voxels from the center of mass,  $\alpha$  being a positive integer in the first case, and a real greater than 1 in the second case. We chose the first possibility after manual inspection, because the distance between the surface of a blob and the close neighborhood does not depend on the diameter of the blob. This neighborhood has been manually estimated to be at a distance of 2 voxels of the surface of the blob. This value is confirmed in the next paragraph.

Knowing the radius of each dot, Figure 3.22 gives a good estimation of how to define the neighborhood of a dot. It appears that choosing  $\alpha = 2$  voxels reflects best the variations of intensity. The absolute differences between the ring 10 and the ring 4 vary from more than 100 to around 30 for the last dot on the graph.

We have selected dots in the elav channel based on visual observation, that clearly look like what we expect from a glia. Thus the value 30 gives a rough estimate for a possible threshold, that would very likely avoid false positive glia detection. We precise the value of the threshold in the following part.

### **3.6.5 Threshold determination**

The glia detection is only a help that will simplify the analysis of the density map. The first priority is then to detect as many glia as possible, while avoiding false positives. It is more acceptable to have a lot of false negatives, that is, glia that are not removed with our criterion. The idea is thus to set a very secure value for the threshold.

What would be false positive cases? One example is, an intensity decrease in the registration channel at the location of a detected nuclei in the beta-galactosidase channel, whereas this nucleus is not a glia. A simple way to avoid such cases is to set a high threshold for the comparison between inside and outside the dots, in order to be sure to detect only very clear dots. After

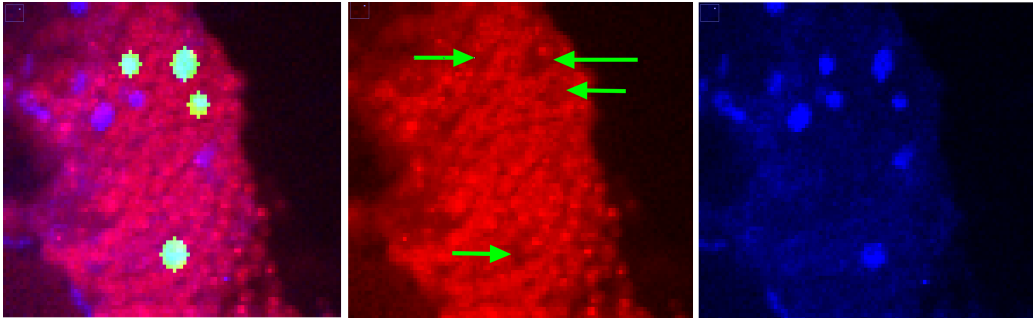


Figure 3.23: Glia detection with a threshold value of 30. Left: elav channel (red), repo channel (blue) and detected dots (green); Middle: elav channel; Right: repo channel

manual check we confirmed that the value 30 for a threshold between the inside and the outside leads to the detection of only clear dots, as it can be seen for example in Figure 3.23.

A crucial point to mention, is that we observe that with this method, we only remove glia that are in the cortex region, as it can be seen in Figure 3.24. It means, since we checked how glia in dense elav staining look like, that we remove only dots that have the desired profile. It insures us also that the staining around the nuclei is strong, which means that we can consider that elav worked well in this region, that is, neurons have been labeled. Knowing the typical intensity profile of a neuron shown in Figure 3.22, that exhibits a clear intensity decrease from the inside to the outside, we can claim that, in addition to the successful manual checks, we highly likely do not remove false positives.

### 3.6.6 Final validation

#### Test on retained staining

In order to validate the developed technique and the threshold value, we have to manually check the removed blobs on the real staining we use for the large scale screening, that is, with an anti-synapsin and anti-elav staining. After examination of several images, it appears that almost all nuclei evaluated as glia are in dense elav labeled regions. An example is given in Figure 3.25. The same reasoning as above can be applied to ensure us that they are not false positive, because the staining in these regions is the same as in the images tested precedently.

However, some nuclei are removed in regions with weak or no elav labeling in the middle of the brain. After manual inspection, we categorized these

*Results*

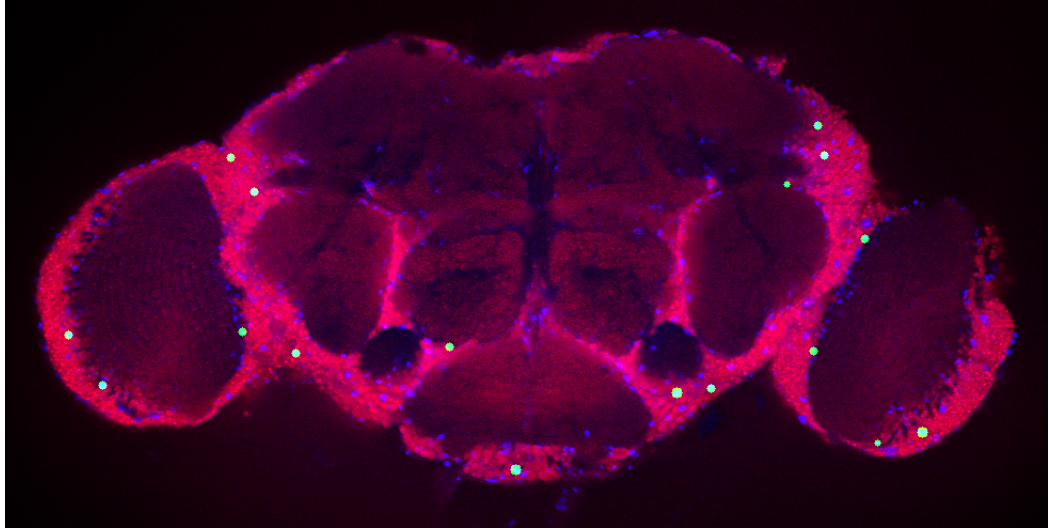


Figure 3.24: Repo (blue), elav (red) and Repo nuclei detected as glia (green). The nuclei detected as glia are only in regions with dense elav staining.

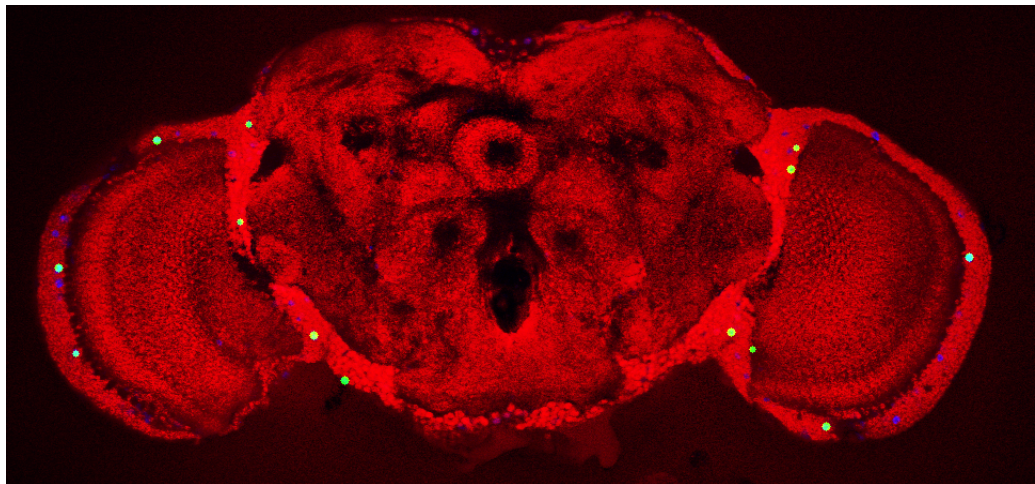


Figure 3.25: Repo (blue), synapsin and elav (red) and dots detected as glia (green). The nuclei detected as glia are only in regions with dense elav staining.



### 3.6 Detection and removal of glial cells

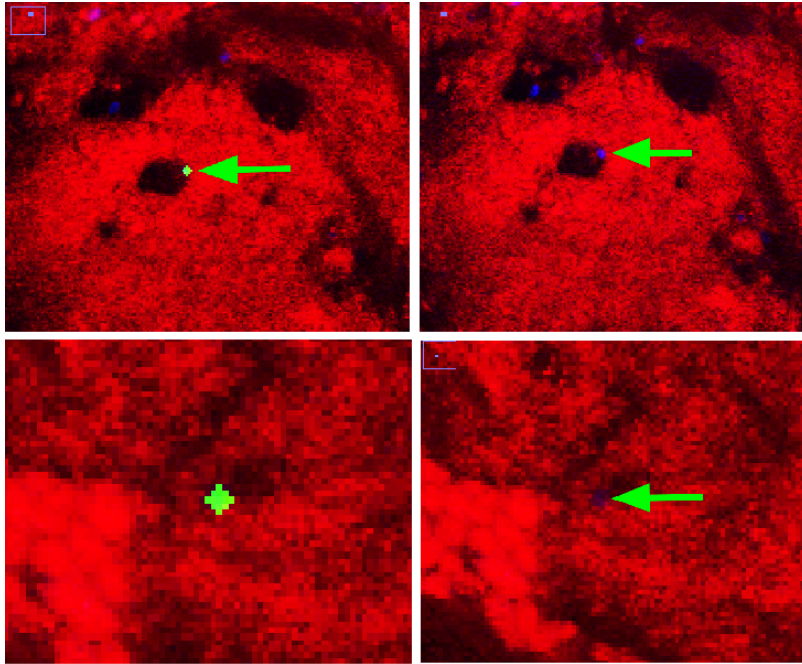


Figure 3.26: Two kinds of nuclei detected as glia in the neuropile region. Top: first sort; Bottom: second sort

cases in two sorts, as represented in Figure 3.26.

The first sort (top on Figure 3.26) is characterized by the unlikely simultaneous happening of the following items :

- clear intensity decrease in the neuropile staining
- nls-lacZ detected blob located in the intensity decrease of the neuropile staining
- the intensity decrease in the neuropile staining has approximately the same size as the nls-lacZ detected dot

Even if the fulfilling of these three conditions seems to be very unlikely, we would like to have an estimation of what would be the parameters for a very safe removal. If we consider that the elav staining is perfectly reliable, such cases are not a problem, as a neuron would have been stained and would exhibit a clear decreasing characteristic as in 3.22.

The second sort (down on 3.26) appears more often. It is characterized by the two items :

- partial intensity decrease at the border between a clear non-neuropile region and a neuropile region

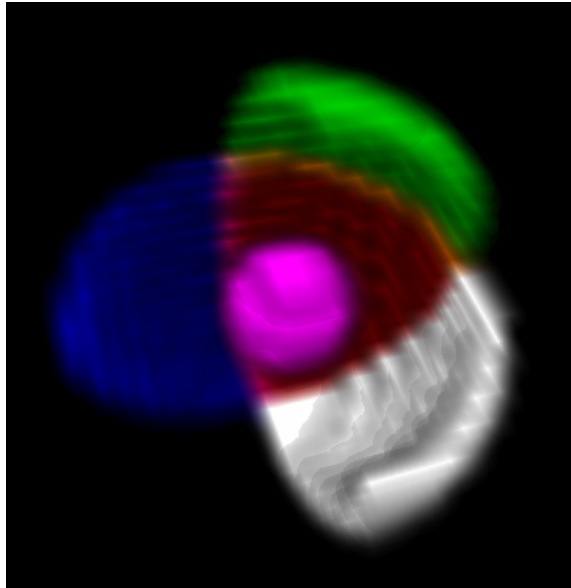


Figure 3.27: 3D ring splitted in 8 different quadrants. Four quadrants are represented with different colors. The center is represented in magenta

- nls-lacZ detected blob located in this border

The dot is removed because the average intensity on the outside is clearly greater than the one inside, because of the close neuropile region. In this case, we lack the information telling that the removed dot is in a dense *elav* labeled region, which means that the *elav* staining worked well in this region. Assuming that the *elav* staining is not so good (although it has been observed in dozens of brains that it is quite reliable), it could lead to a false positive removal.

### **Eight quadrants rings**

In order to exclude these two cases, we precised the criterion of dot removal. Instead of computing the average intensity on whole rings, we splitted the rings in eight quadrants. Four rings are represented with different colors on Figure 3.27. The center of mass is represented as a sphere in magenta.

We apply the threshold criteria to each of these eighths of rings, which gives a score between 0 and 8. 8-scoring dots have an intensity decrease in all eight tested directions. To avoid cases of the second sort mentioned above, a score threshold of 5 or 6 should be sufficient for a systematic removal. To also avoid cases of the first sort, a threshold of 5 or 6 is also enough. With a score threshold of 5, a false positive should correspond to a neuropil region having

a clear intensity decrease in almost all directions. This decrease should have a dot shape and have roughly the same size as the nls-lacZ signal detected at this location. We consider this as highly unlikely, and did not find any of them in our manual inspections.

We tested different combinations of intensity difference thresholds and score thresholds. After manual evaluation of numerous brains of different lines, we chose an intensity difference threshold of 30, and a score threshold of 5. With these values, we did not find any dot detected as a glia in neuropil regions. All nuclei belong to dense elav labeled regions.

To conclude this paragraph, we resume the reasons why we claim that we do not remove false positive nuclei with a very high probability with the parameters defined above.

- anti-elav staining is reliable
- neurons show a clear intensity decrease whereas we select only clear intensity increases
- threshold values are set far from critical values
- nuclei in neuropile regions are not excluded with the method of the splitted rings because only dots with uniform intensity decrease are removed
- only nuclei in dense elav labeled regions are removed, that insures that the elav staining worked well in this region

#### 3.6.7 Application on the large scale screening

With the set of safe parameters (30;5), we obtained a removal rate of  $8\% \pm 8\%$  on 14 lines. After manual inspection, it appears that this set of parameters is roughly satisfactory. Removed nuclei almost exclusively located in dense elav labelled regions. Figure 3.28 gives an example of successful glial cell detection in a dense elav region.

I tested other parameters, trying to increase the removal rate. A good set of parameters I found is (20;5). After manual evaluation of several brains of different lines, I did not find any clear false positive removal. With values below 20 for the intensity threshold, I started to clearly increase the number of false positives. I think that this set of parameters is a good compromise between a high removing rate and the insurance of avoiding false positive removals. Figure 3.29 shows the average percentages of removed nuclei for each line, with the two sets of parameters. As expected, the removal rate is

## Results

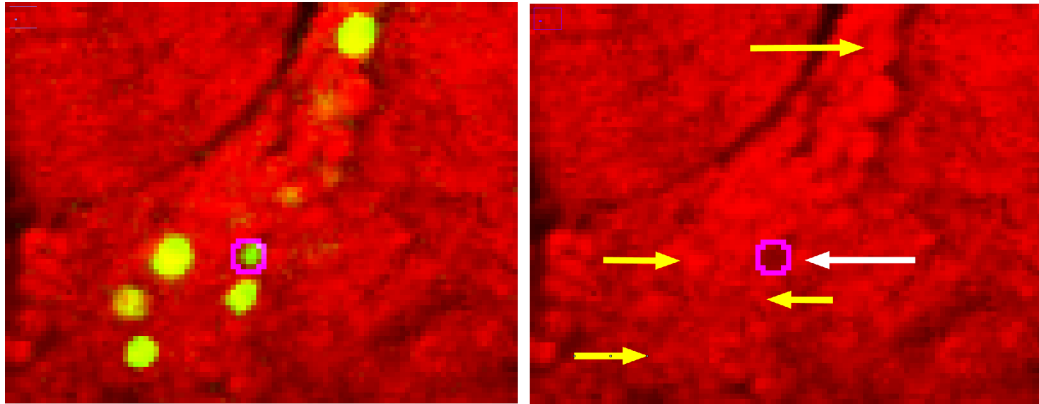


Figure 3.28: Result of the glia detection in a dense elav labeled region. Red: registration channel (synapsin, elav); Green: third channel; Magenta: nuclei detected as glia

higher with (20,5) than with (30,5). The rather low standard deviations show that the technique is quite reliable and the removal is reproducible among brains of the same line.

Finally, Figure 3.30 shows the the better glial detection of (20,5) towards (30,5). Three additional glial cells are detected with (20,5) and are shown with the arrows.

## 3.7 Removal of nuclei based on GFP signal

### 3.7.1 Preliminary observations

Examining the first stainings with the double reporter line expressing both myrGFP and nls-lacZ, it appears that a lot of nuclei revealed by the anti-beta-galactosidase staining are not visible at all in the GFP channel. Figure 3.31 shows such examples. Reasons for such a discrepancy in the labelling are discussed in 4.1.

The density map as hypothesis generator should highlight the overlapping neurons in a small region of nuclei of different lines. If the density map reveals a region with high concentration of nuclei, then the user has to check manually the GFP channels of the corresponding images. The user principally observes the GFP channel and the registration channel. The reference channel gives an insight into the location of the candidate region and enables a possible manual evaluation for the glia. The user observes the fine structures of the neuron with the GFP channel and can determine if the neuron are the same or not.

### 3.7 Removal of nuclei based on GFP signal

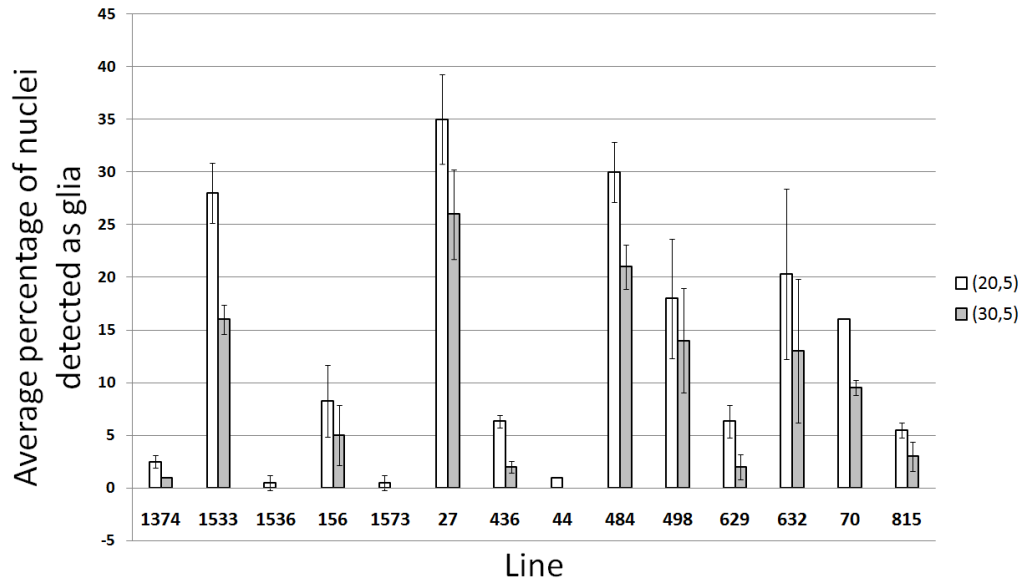


Figure 3.29: Average percentages of nuclei detected as glia for two sets of parameters

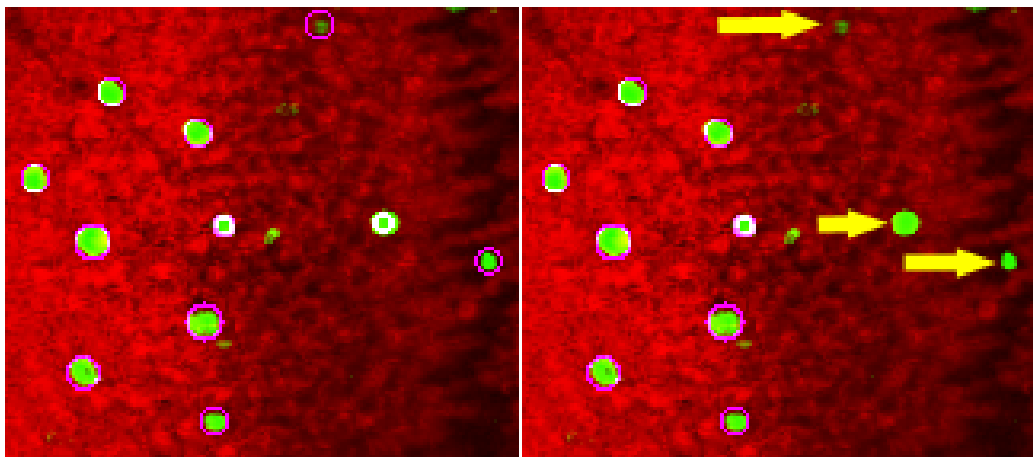


Figure 3.30: With the parameters (20,5), three additional glial cells are detected. Red: registration channel; Magenta: detected glia; Green: nuclei channel

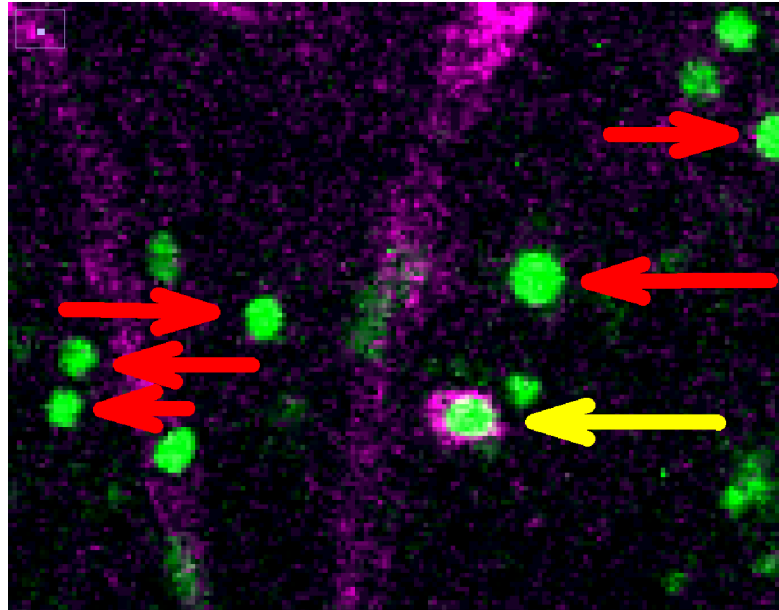


Figure 3.31: GFP channel (magenta) and nuclei channel (green) merged in an image. Red arrows indicate nuclei without GFP expression, and the yellow arrow shows a nuclei with GFP expression

If the density map observation leads the user to look more precisely at supposed overlapping nuclei that do not have any GFP expression, then the manual determination is impossible. It means that I could discard such nuclei from the density map analysis, based on the presence of GFP signal in a small area around the nuclei.

Similarly to the glial cells removal, I have to consider what would be the worst case of a false removal ? It would be a candidate neuron labeled with nls-lacZ, with no signal in the nuclei on the GFP channel, but with a clear signal in other parts of the neuron. The non detection of GFP signal would lead us to remove this nuclei, that would not appear in the density map. In the worst case, the absence of GFP signal in the nuclei should be reproducible among brains of the same line and across the different lines, which is very unlikely to happen.

### Histogram of the GFP channel

Similarly to the strategy used for the glia removal, the goal is to set very safe thresholds, to avoid false positive detections. Therefore, as some lines exhibit strong GFP expressions and others rather weak expressions, I first focused the tests on a line with a weak GFP expression.

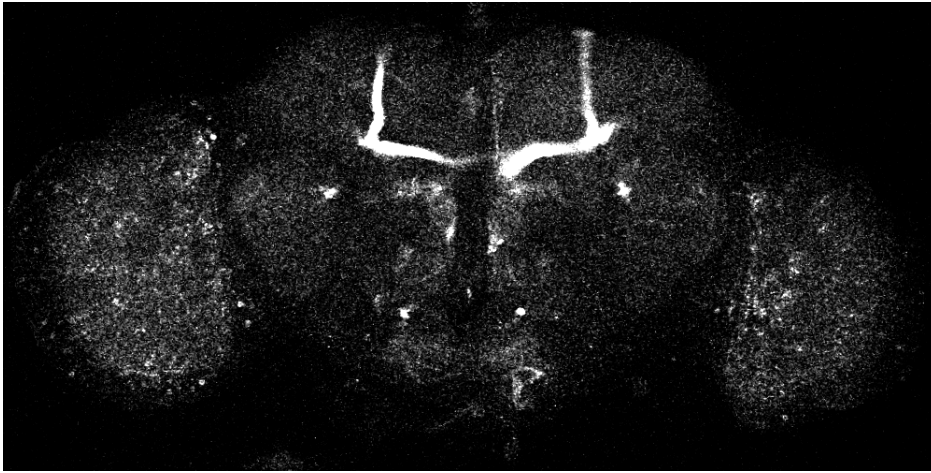


Figure 3.32: GFP channel of a weak line. The range of the intensity has to be set very narrow, leading to a lot of noise

As explained in section 2.2.4, I use the same laser settings for all the brains scanned. It leads to very weak signals for weak lines. Figure 3.32 shows a slice of the weak expression line I used in this section.

The range of the intensity for the display has been set very narrow in order to visualize the signal on screen. The brightest regions of the picture have an intensity of around 60, on a scale of 255 (8-bit images). Figure 3.33 is the intensity histogram of this image and is typical for weak lines.

The majority of the pixels has an intensity lower than 4 and corresponds to the background containing no GFP signal. It gives a first rough estimation of possible thresholds intensity between 5 and 20.

### **Average intensity of spheres on the GFP channel**

To decide if there is some GFP signal in a detected lacZ dot, a first idea is to observe the average intensity in and around the dot. Figure 3.34 shows the intensity histogram of the dots with a radius corresponding to the the radius of the dot (left) and twice the radius of the dot (right).

In both cases, there is a clear border at the intensity of five. We could define a removal criteria: if the average intensity in the GFP channel of a nuclei is smaller than the threshold 5, then the nuclei is discarded. Setting a threshold at five would be satisfactory because it enables us to remove a large majority of dots. However, the average intensity of the nuclei might not be the best value to test. Figure 3.35 shows a lacZ blob with an extremely weak GFP signal in it. We think that it must represent the weakest signal we can manually detect. Such a dot is likely to be removed with an average

Results

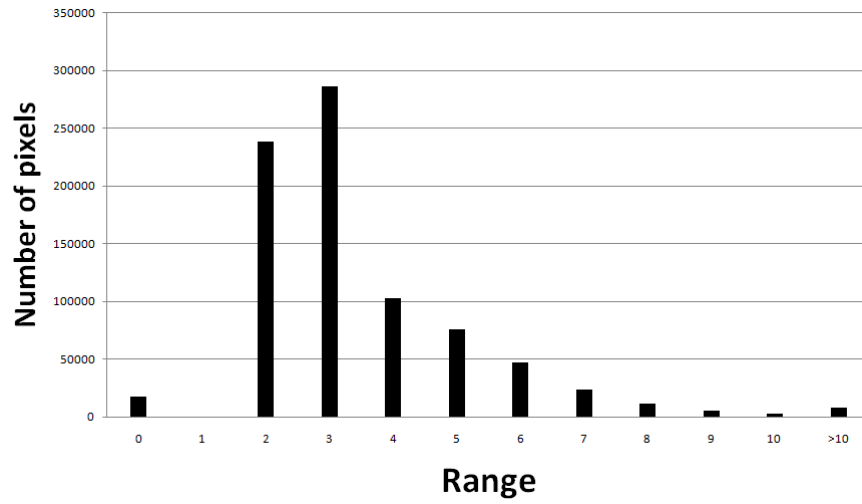


Figure 3.33: Intensity histogram of Figure 3.32

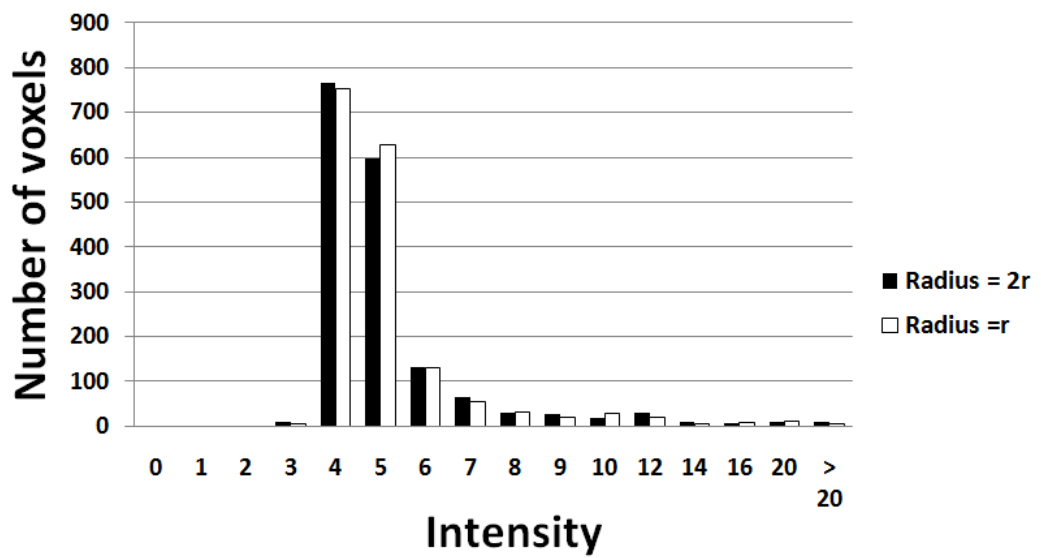


Figure 3.34: Histograms of average intensities of spheres centered on the center of mass of detected nuclei with two different radius



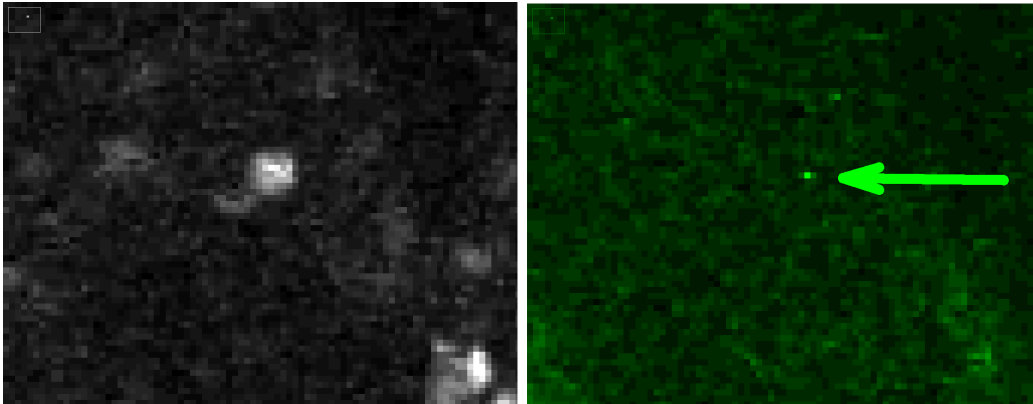


Figure 3.35: A lacZ blob with an extremely weak GFP signal in it. The highest pixel intensity of the blob is fifteen.

intensity criterion. That is why instead of using an average intensity criterion, we should focus on a maximum intensity criterion.

### Maximum intensity of spheres on the GFP channel

If the GFP signal is very weak in a small part of the nuclei with the rest being almost zero, only the maximum value can tell us that there may be a GFP signal. The corresponding criterion would be: if a nuclei does not have any voxel in the GFP channel with an intensity greater than a threshold, then the nuclei is discarded.

Figure 3.36 gives the intensity histogram of the maximum values of each detected nuclei.

A clear separation around the intensity 7-8 is to be observed. Setting a threshold around these values would be satisfactory concerning the number of nuclei removed. Moreover, these values correspond to the finest structures I were able to detect by eye, trying to follow the finest possible dendrites or axons. Figure 3.37 shows such a weak structure. The weakest detectable pixels have an intensity of 9-11.

The value 10 appears thus to be a good threshold. It would enable a removal of 70% of the nuclei. However, taking the image noise into account, it may not be appropriate to decide the removal of a nuclei based on only one voxel. In the next paragraph, I enhance the criterion and increase the removal rate without raising the risk of false positive removals.

Results

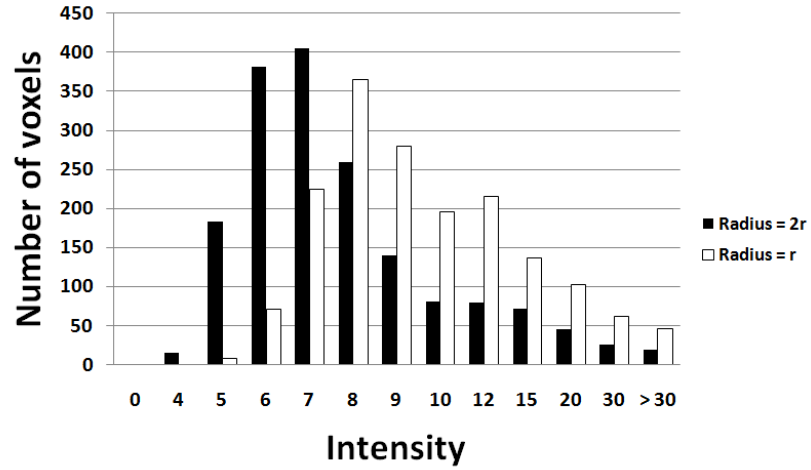


Figure 3.36: Histograms of spheres maximum intensity centered on the center of mass of detected nuclei. Left: same radius as the nuclei; Right: twice the nuclei radius

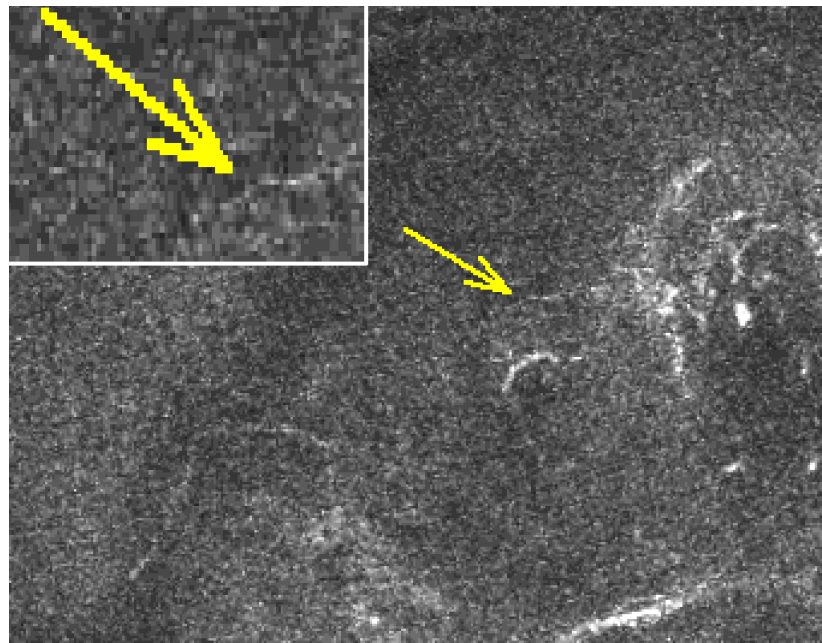


Figure 3.37: A dendrite or axon of a GFP expressing cell labeled with anti-GFP. It represents one of the weakest signals we are able to detect manually.

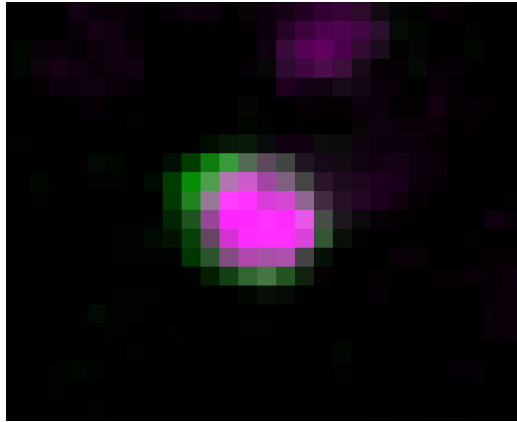


Figure 3.38: Typical profile of a lacZ nuclei with GFP expression. The GFP expression is located in the plasma membrane and the lacZ expression is located in the nuclei.

### 3.7.2 Definition of the neighborhood

Given the radius of detected nuclei in the nuclei channel, I had to define the neighborhood of the nuclei where it is checked if there is a GFP signal. The lacZ protein is only located in the nuclei, whereas the GFP protein is located in the plasma membranes. It leads to a GFP signal that has a ring form, with lacZ signal inside the ring. This typical feature is shown in Figure 3.38.

Manual inspections of numerous such lacZ nuclei with GFP expression, we estimated that the majority of the GFP signal is located in a sphere centered on the center of mass of the nuclei with a radius equals to the radius of the nuclei plus two. This definition of the neighborhood is used for the following paragraphs.

### 3.7.3 Score of maximum intensity

Not removing a nuclei because only at least one voxel has an intensity greater than a threshold may be a bit too much selective. A nuclei with only one bright voxel does not really help for the manual inspection. The criterion that consists of at least a few voxels having an intensity greater than a threshold seems to be more appropriate. Figure 3.39 gives for example the histogram of the number of pixels in a nuclei having an intensity greater than the threshold value 10.

If I require that a nuclei of interest should have at least 3 pixels with an intensity greater than 10, then the removing rate grows from 70% to 86%.

To summarize, all detected nuclei having less than 4 pixels with an inten-

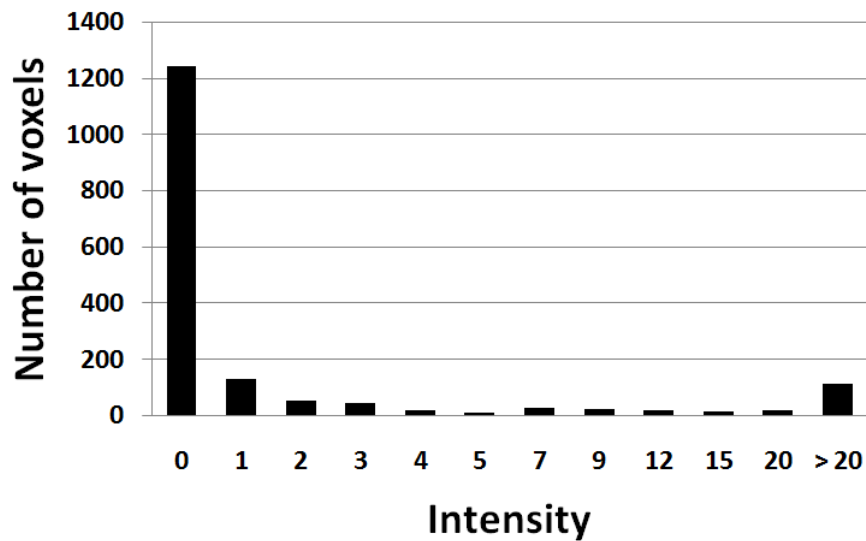


Figure 3.39: Histogramm of the number of pixels having an intensity greater than 10.

sity greater than 10 are removed. It corresponds to a removing rate of 86%. For other brains of the same line, I had very similar results with removing rates of 81 and 83 %.

### 3.7.4 Tests on weak and strong lines

To conclude this section, I present some results of the technique applied to several brains of several lines, in order to definitely fix the two parameters : the intensity threshold and the voxels score that gives the number of voxels that must have an intensity greater than the intensity threshold.

The value of 10 for the intensity threshold seemed to be satisfactory and has been determined by the weakest manually detectable GFP structures. A higher value would lead to false positive removals, and a lower value would lead to false negative removals. Therefore we focused only on the score parameter.

After manual evaluation of several brains, it appeared that a value of around 20-40 for the score parameter seemed to be satisfactory. I then tested five different parameters with 14 different lines. The results of the average percentages of remaining nuclei are given in the Figure 3.40. As expected, if the value of the score parameter increases, then the percentage of remaining voxels decreases.

A very important information given by this graph is the standard devia-

tion of the remaining nuclei. It shows that the nuclei remaining percentages are reproducible in brains of the same line. It means that the number of nuclei having a GFP expression is approximately the same in all brains of a same line.

We observe a very low removal rate in the line 484. This is the line with a lot of glia that have already been noticed in section 3.4.2 and shown in 3.15.

The final parameters we chose is 10 for the intensity threshold, and 30 for the score threshold.

#### 3.7.5 Statistics for the large scale screening

We applied the nuclei removal based on GFP expression to the 48 lines that have so far been acquired. Figure 3.41 gives the histogram of the removal rates in 48 lines (122 brains). The average of the removal rate is 11 % and is satisfactory. Figure 3.42 gives an example of the nuclei removal on a whole brain and Figure 3.43 shows the removal in a zoomed region.

## 3.8 Density map

### 3.8.1 Plugin

A plugin has been developed to allow a manual inspection of the density maps. The user can set several parameters for the display of the density map. The major functionality consist of the possibility to chose which lines are displayed in the density map. It is done by simple click in a table listing the available lines. Figure 3.44 shows a screenshot of the plugin.

Several parameters can be fixed for the display of the density map. The most important one is the radius.

### 3.8.2 Radius of nuclei

Figures 3.45 shows a density map of nuclei of the same line with two different radius (20 voxels and 40 voxels).

### 3.8.3 3D display

Using embedded Fiji functionalities, the user can chose to look at the density maps in a 3D reconstruction. It can reveal informations that could have been missed with the classical view. Figure 3.46 shows a density map visualized in three dimensions. The clusters of Kenyon cells appear clearly on the image.

Results

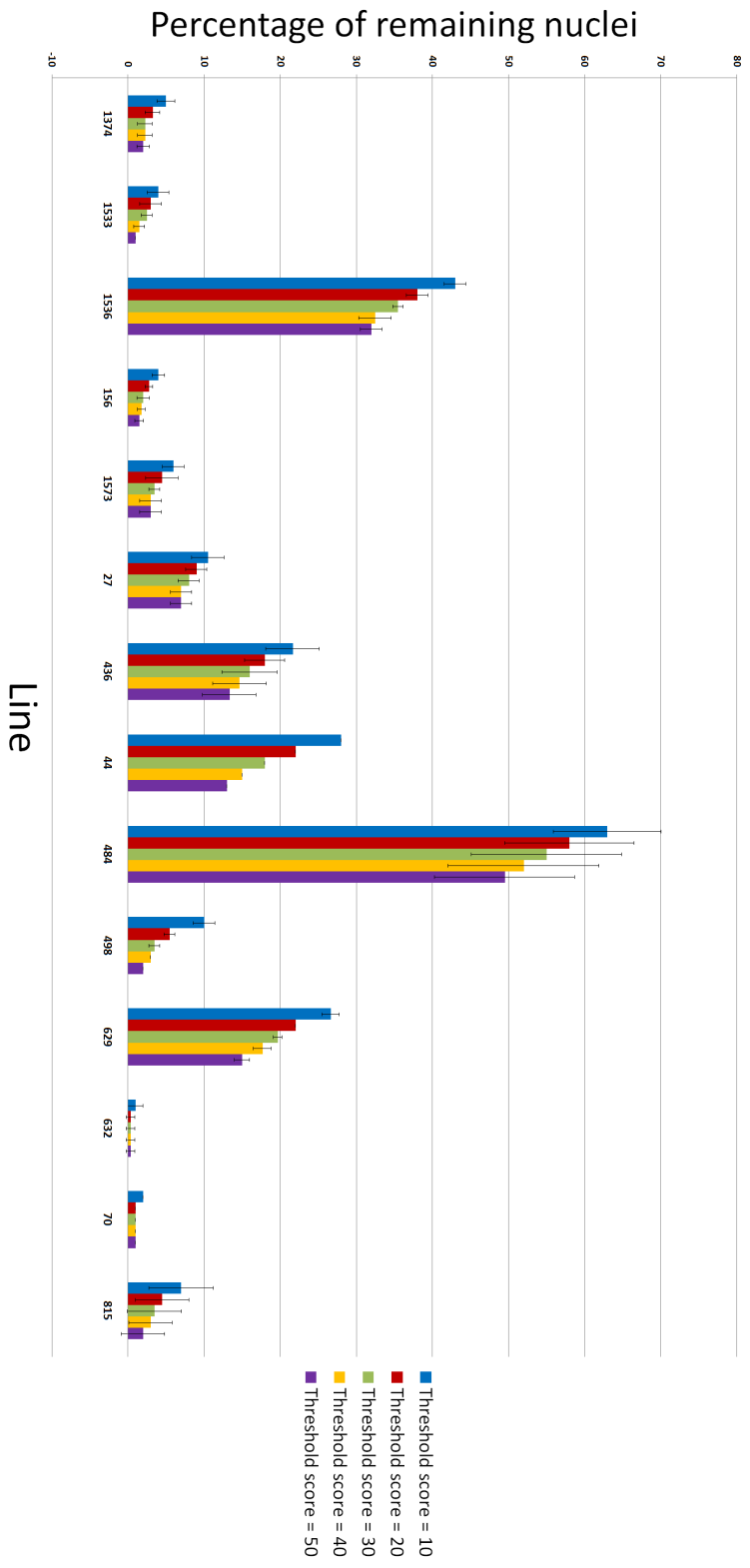


Figure 3.40: Average percentage of remaining nuclei after nuclei removal based on GFP expression in fourteen lines with five different values for the score parameter.

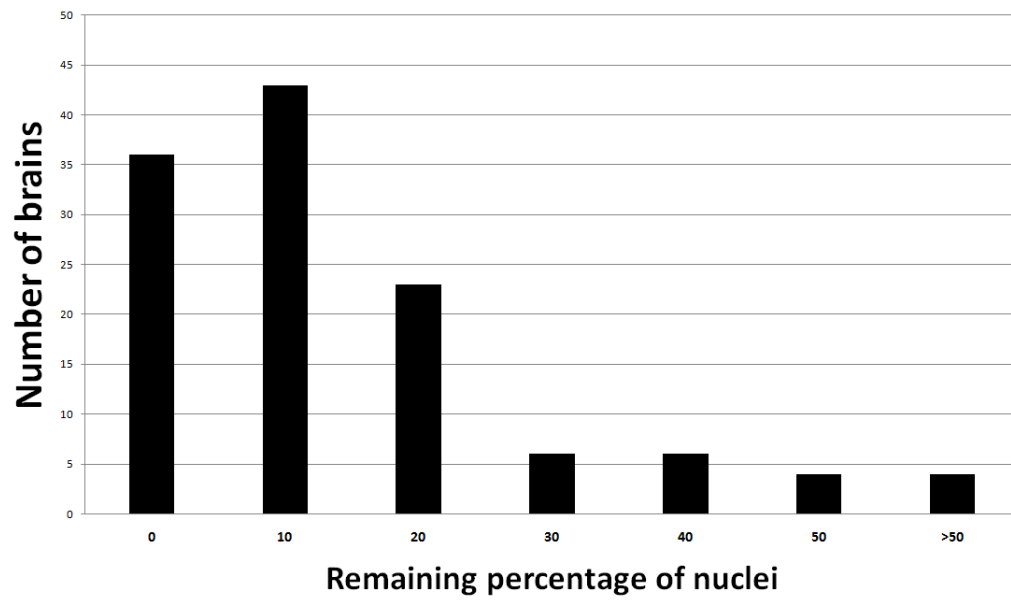


Figure 3.41: Histogramm of the removal rates of 48 different lines (122 images)

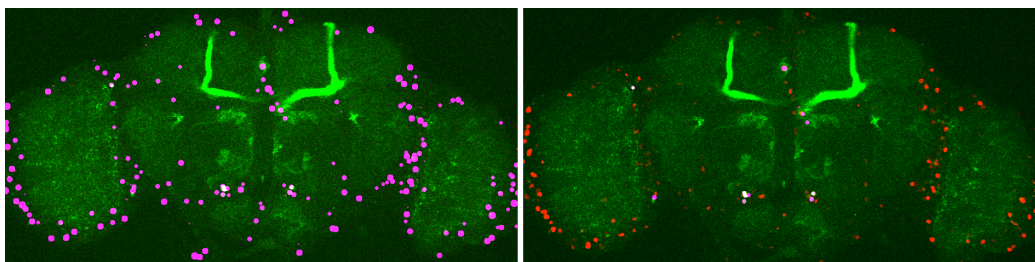


Figure 3.42: Example of nuclei removal based on GFP expression. Green : GFP; Magenta: detected nuclei (left) and remaining nuclei (right); Red: nuclei channel

*Results*

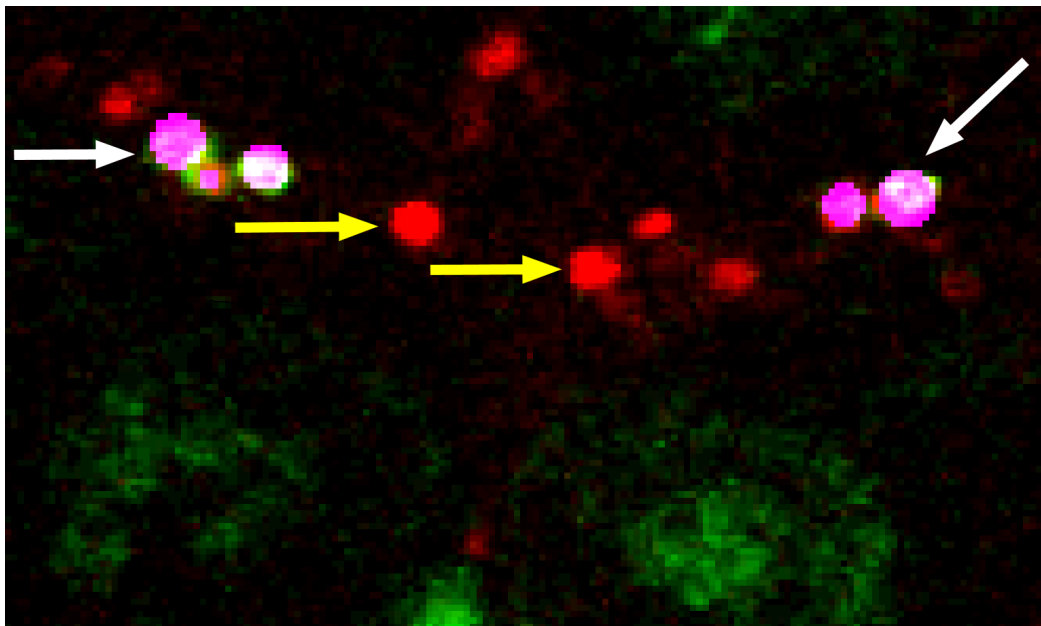


Figure 3.43: Nuclei removal based on the GFP expression. Green: GFP channel; Red: Nuclei channel; Magenta: remaining nuclei presenting GFP signal



### 3.8 Density map

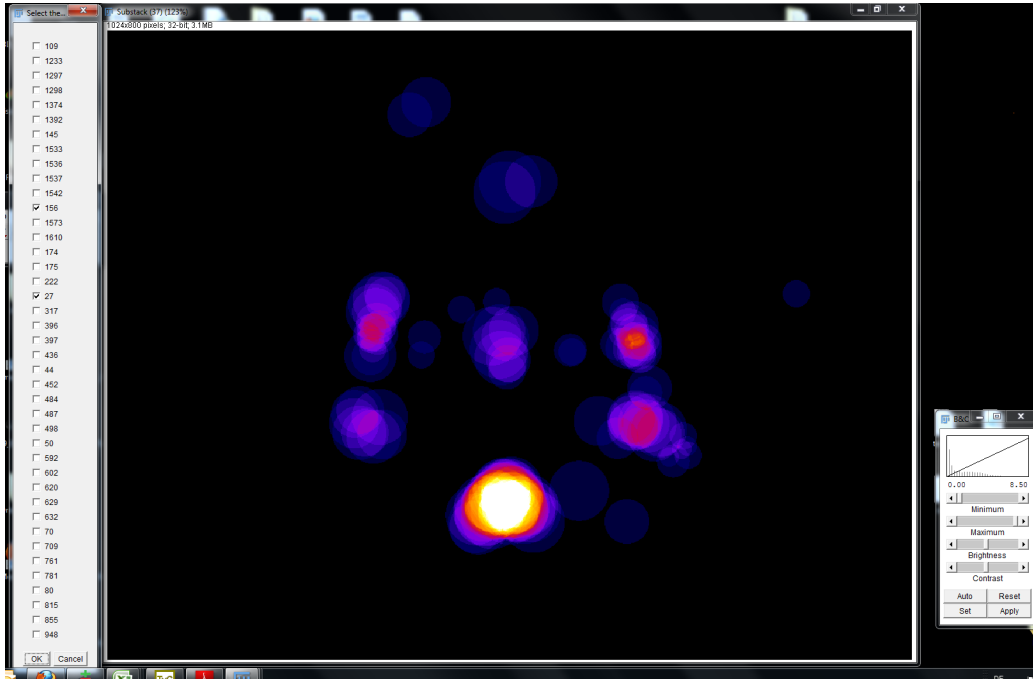


Figure 3.44: Screenshot of the visualization plugin for the density map developed for the software Fiji. The user can select on the table on the left the lines to be displayed.

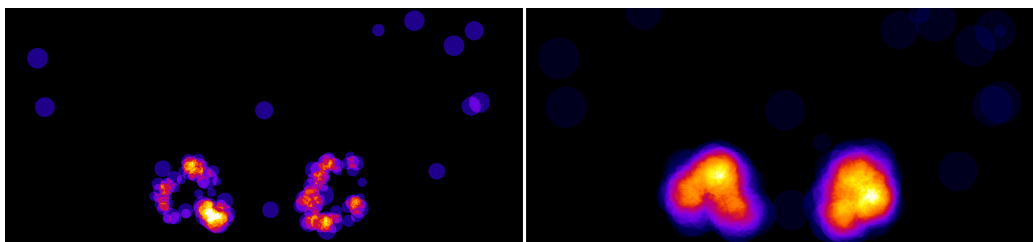


Figure 3.45: Two density maps displayed with different radius for the nuclei. Left: radius 20 voxels; Right: radius 40 voxels

*Results*

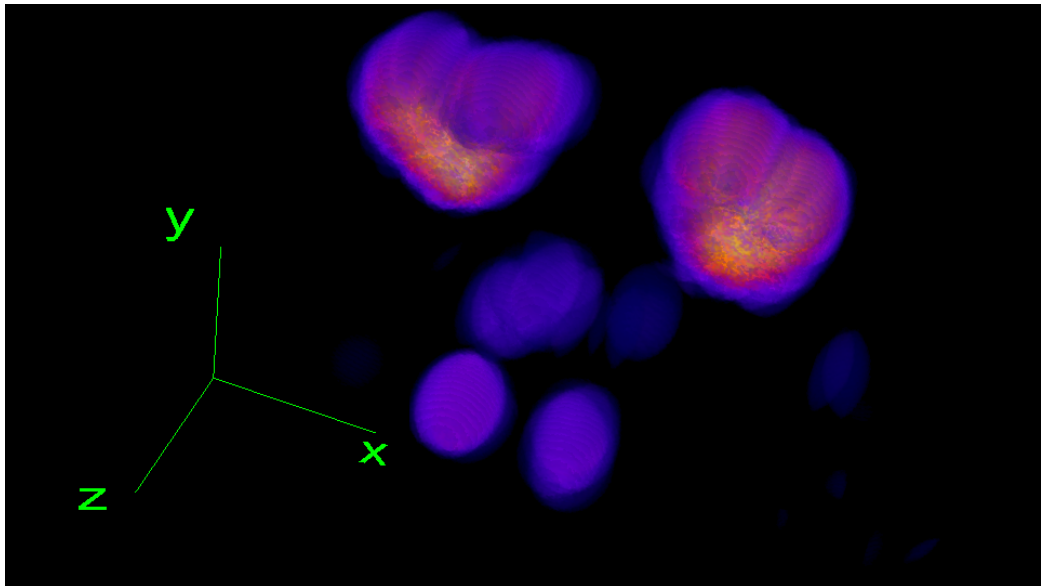


Figure 3.46: Three-dimensional view of a density map.

# Chapter 4

## Discussion

### 4.1 Nuclei removal based on GFP channel and registration channel

#### 4.1.1 Positive nls-lacZ labeling and negative GFP labeling

The first images we acquired stemming from the double reporter line that enables a simultaneous staining of GFP proteins and nls-lacZ proteins were quite surprising. The majority of the cells having their nuclei labeled in the nuclei channel did not have any GFP signal. We first thought that it would be a major impairment for our goal of identifying overlapping neurons based on the nuclei locations.

Several reasons could contribute to the simultaneous positive anti- $\beta$ -galactosidase staining and GFP negative staining. 1. myrGFP binds on membranes, whereas the lacZ proteins are concentrated in the nuclei. Therefore very weak signals can be detected in the lacZ channel while nothing is detected in the GFP channel. 2. myrGFP and nls-lacZ may have been both expressed during the embryonal stage of the development and then inactivated in the adult stage. nls-lacZ might be more persistent than myrGFP proteins. 3. The anti-GFP antibody is not as efficient as anti-beta-galactosidase. 4. Because of microscope characteristics, the channel used for GFP does not enable to see very weak signal.

For our purpose, we found a way to overcome this discrepancy by automatically discarding nuclei that do not show any GFP expression. The question remains to exactly know the reasons of this staining difference, that could have a substantial impact on our approach.

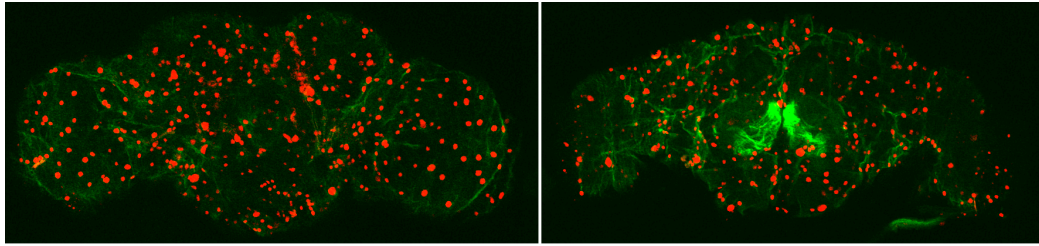


Figure 4.1: Line with a high amount of glia. It has been manually checked on the registration channel that most of the nuclei of this picture are glia. Green: GFP channel; Red: nuclei channel

### 4.1.2 Lines with strong glial cells expression

In cases of Gal4 lines having a strong glial expression, the removal of nuclei based on the GFP expression can be problematic. Such lines have a typical expression pattern as can be seen in Figure 4.1. Manual inspection of the registration channel containing the elav labeling shows us that almost all glia located at the ventral and dorsal surfaces have a Gal4 expression. Dendrites and axons of these glia spread out in cortex regions to provide support to nuclei of neurons located in the cortex. Therefore, lacz signals of nuclei present also a GFP signal that is not due to the GFP expression of the nuclei, but to the glia innervation. Figure 4.2 shows that with our technique, we are unable to discard these nuclei. We see in green the GFP channel, in red the lacz channel, and in magenta the remaining nuclei that have a GFP expression. Almost all nuclei are kept on this cortex layer because of the GFP signal that may come from only a few glial cells. The distinction is even not very clear with manual inspection. It is hard to determine if the remaining nuclei really have a GFP expression, or if they are bound to a glia with a GFP expression.

So far, 49 lines have been analysed, and we found 3 such lines. The removal rates of these lines are 50%, 58% and 38%. The number of remaining detected nuclei varies between approximately 1000 and 3000, which is far higher than average. As a consequence, we may exclude these lines in a first time in the exploration of the density map. As we discuss later, it is not practical to display lines with more than one thousand nuclei.

### 4.1.3 Effect of false positive glia removal

The results of section 3.7 show that the glial removal rate is very low when using the safest parameters we determined. Compared to the high removal rates provided by the criterion based on the coexpression of GFP and lacz,

#### 4.1 Nuclei removal based on GFP channel and registration channel

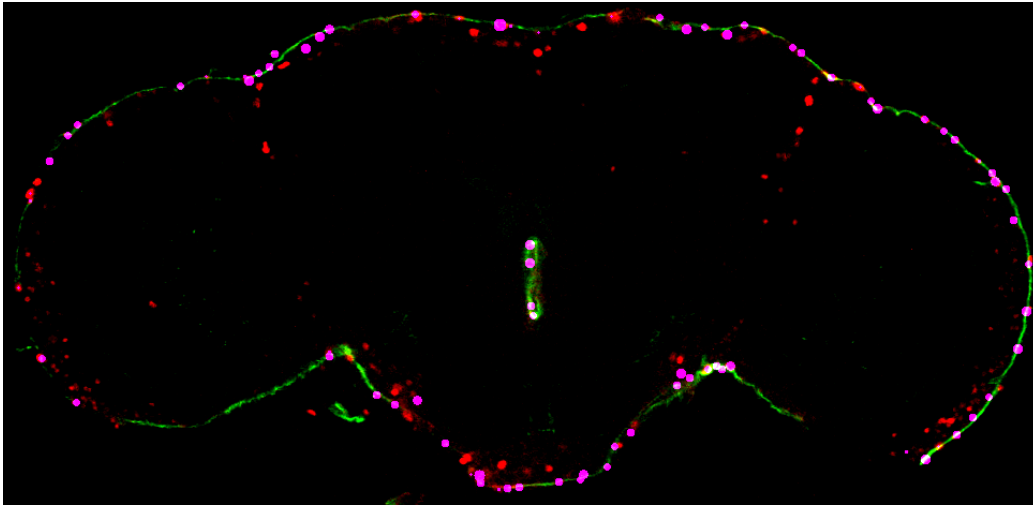


Figure 4.2: Line with a high amount of glia. Green: GFP channel; Red: nuclei channel; Magenta: remaining nuclei presenting a GFP signal

this removal has little or no effect on the density map, when used with the safest parameters. However, it appears after manual inspection, that choosing less safe parameters leads to false positive only in the middle of the brain and not in cortex regions. They have been characterised by the unlikely simultaneously fulfilling of several conditions : the detected nuclei are present in the central brain, at a location where there is an intensity decrease in the registration channel due to the neuropile staining. Nonetheless, given the nuclei location variability, these low probability cases are not reproducible. Moreover, we create the density map using several brains per line. It leads to a minimization of the effect of rare glia false positive removals present in single brains. The user can therefore try different parameters for the glia removal and select a set of parameters that seems to produce only rare false positive removals, without impairing the search for overlapping neurons in the density map. We estimate a such set of parameters to have a threshold of 20 and a eighth-ring score above 4.

#### 4.1.4 Comparison of the two removal techniques

We used two different ways to discard nuclei. The first is based on the presence or not of GFP signal in nuclei locations, and the second on the intensity profile of the detected nuclei in the registration channel, determining if the nuclei is a neuron nuclei or a glial nuclei. The results show that if we choose the safest sets of parameters to avoid false positive removals, the

## *Discussion*

GFP-based removal rate is far greater than the glia-based removal. With the other set of parameter (20,5), the removal rate is still far higher with the GFP expression criterion.

## **4.2 Density map**

### **4.2.1 Number of nuclei visualized simultaneously**

#### **Number of lines visualized**

A major impairment that users will seem to face with the analysis of the density map is the high number of remaining nuclei in numerous lines, even after use of our two nuclei removal techniques. Figure 4.3 shows the density map of 20 lines in twelve slices.

With so many lines displayed at the same time, the density map tends to display the nuclei regions uniformly. The Kenyon cells cluster seems to be the region being the most repeatedly labeled. A brief outline with handling of the density map suggests that the visualization should occur with a limited amount of lines. Given that the 100 isolated Gal4 lines have been tested with precise criteria, the visualization should be based on these groups. For example, one could start to visualize only lines of flies exhibiting an impairment in the sugar response without locomotive defects, or any group that can be formed based on fine behavioral assessments.

#### **Radius**

The choice of the radius has to be done by the user and should be changed in accordance with specific needs. I estimate the optimal radius to be around 40 voxels to reflect the nuclei variability. A functionality that has not been implemented and that could enhance the visualization, would be to draw the nuclei with a decreasing intensity starting from the center of mass. It would still take the nuclei location variability but would reduce the tendency of nuclei regions to be uniformly labeled.

#### **Customizing removal thresholds**

A more direct way to reduce the number of nuclei would be to adjust the nuclei removal techniques for each image. Threshold could be set depending on the number of nuclei detected, on the average intensity of the GFP channel or could be manually set by the user. It would probably lead to a high

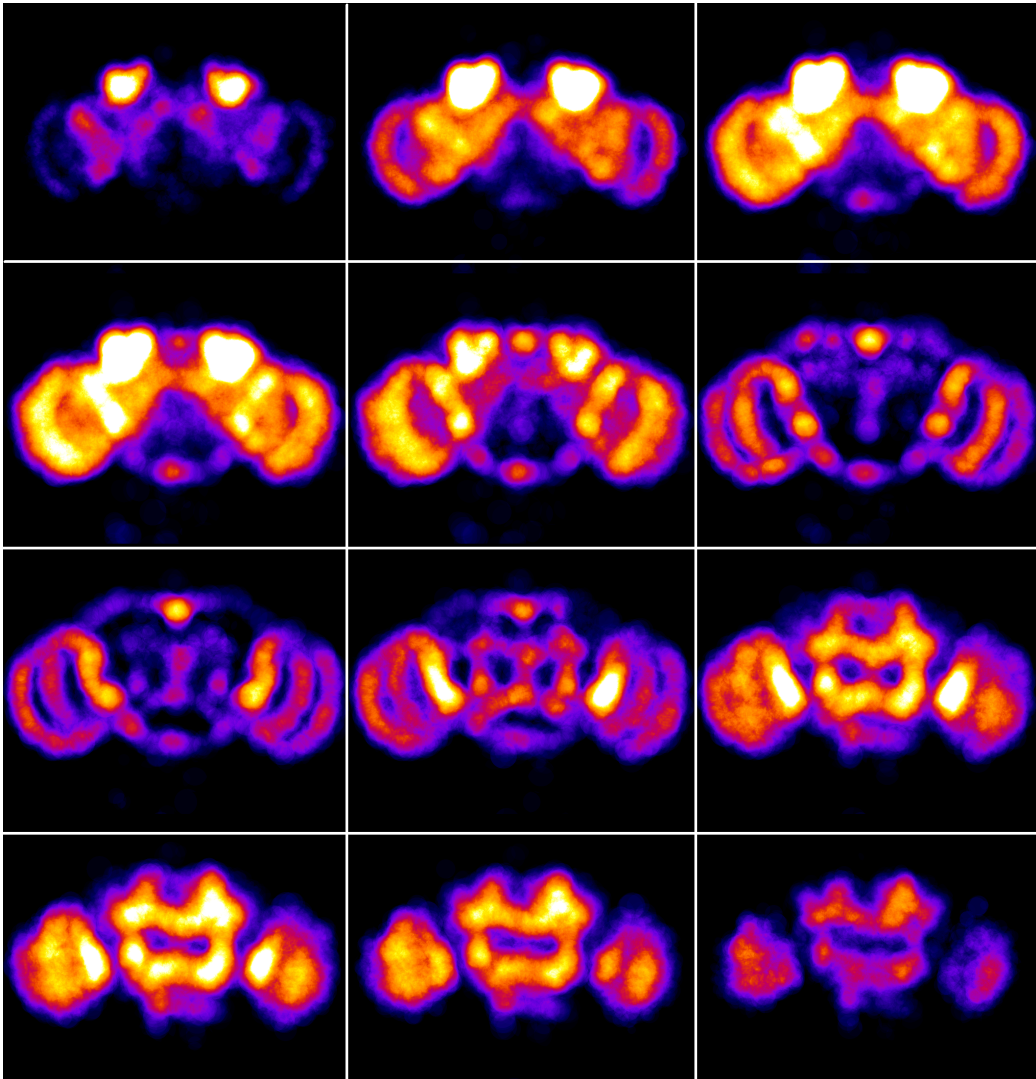


Figure 4.3: Nuclei density map of 20 different lines

## *Discussion*

increase of the removal rate. I think that this way should provide a major improvement.

### **4.3 Further applications**

The biological and computational pipeline established in this work are suited for further applications.

The information obtained on nuclei could be used to produce statistics towards the expression pattern of each Gal4 lines. The sparseness of expression patterns could be automatically assessed, which is an useful information to evaluate the interest in a line. A manual segmentation could be performed on the reference brain to identify the principal structures of the neuropile. It would then possible to have statistical information about the location of the nuclei towards these structures.

The search for overlapping neurons repeatedly labeled in different Gal4 lines in Ventral Nerval Cords in *Drosophila* could take advantage of our technique. Ventral Nerval Cords (VNC) consist mainly of the same type of cells as the brain. Neuron nuclei, synapses and plasma membranes can be labeled similarly to the brain. Registration of VNC has already been reported in Jefferis et al. (2007). It shows that registration with the our quadruple staining may have a chance to succeed.

The general project including this work aims to characterize neuronal circuits involved in the response to sugar and electric shock stimuli. The approach could be used for any behaviour that can be easily evaluated by observation.



# List of Figures

1.1	General principle of the UAS/Gal4 system (Perrimon et al., 1991) . . . . .	3
1.2	Two sugar arenas. On the left part of each arena: sucrose solution filter paper. On the right of each arena: water on filter paper. On the left arena: normal behaviour. On the right arena: abnormal behaviour. . . . .	5
1.3	Two electric shock arenas. On the left of each arena: electric shock. On the right of each arena: no electric shock. On the left arena: normal behaviour. On the right arena: abnormal behaviour. . . . .	6
1.4	Slice of a brain stained with the antibody anti-synapsin . . . .	8
1.5	Slices of a brain stained with the antibody anti-NCadherine . .	8
1.6	Slices of a brain stained with the antibody anti-discs-large . .	9
1.7	Anti-GFP staining. We see that membranes of somas, dendrites and axons are labeled. . . . .	9
1.8	Anti-beta galactosidase staining. Only nuclei of cells expressing nls-lacZ are labeled. . . . .	10
1.9	Staining with the antibody anti-repo showing glial cells in the fly brain . . . . .	11
1.10	Staining with the anti-body elav . . . . .	12
1.11	Anti-synapsin and anti-repo on the same channel . . . . .	15
1.12	Three images of three different channels acquired from the same biological sample. Left: 488 nm; Middle 560 nm; Right 630 nm . . . . .	16
1.13	General registration process. Left: subject (template) brain; Middle: transformed (warped) brain; Right : reference (target) brain . . . . .	17
1.14	One slice of the brain of <i>Drosophila</i> (from Strausfeld (2007)) .	19

*List of Figures*

2.1	Mounting grid. Brains are placed at the lines intersections in the red square. Therefore, the position of the brains is the same for all glasses . . . . .	24
2.2	Effect of the contrast enhancement on two slices of a brain stained with anti-synapsin and anti-elav . . . . .	27
2.3	Screenshot of the V3D tool. The red circles indicate the locations of the landmarks (represented as circles, but landmarks are points) . . . . .	28
2.4	Schema of the process of creation of the average reference brain	30
2.5	Average reference brain used for the large scale screening. . . . .	31
3.1	Three channels of the quadruple staining merged on a same image. Green : GFP channel (named also first or green channel); Red : registration channel (named also second or reference channel); Blue : anti- $\beta$ galactosidase channel (named also third or lacz or nuclei channel) . . . . .	35
3.2	Three channels of the quadruple staining. Left : GFP channel; Middle : registration channel; Right : anti- $\beta$ galactosidase channel . . . . .	35
3.3	Nuclei channel with concentrations of 1:250 for primary and secondary antibody . . . . .	36
3.4	Slices of brains stained with different antibody concentrations from Table 3.1. Top left: Test 1; Top right: Test 2; Bottom left: Test 3; Bottom right: Test 4 . . . . .	37
3.5	Histogram of landmarks pairs conserved . . . . .	38
3.6	Some examples of the accuracy of the registration in nuclei regions. Green: transformed brain; Magenta: reference brain . . . . .	39
3.7	Successful registration of a brain with only one optic lobe to the reference average brain. . . . .	39
3.8	Brain with with clearly rotated optic lobes . . . . .	40
3.9	Result of affine registration. Blue: target brain; Red: result of the transformation of the subject brain . . . . .	41
3.10	Definition of the four points needed for the optic lobes removal. The line in gray is the shortest path found. . . . .	42
3.11	Exemples of optic lobes removal in three slices of a same brain.	42
3.12	Shortest path computed by the Djikstra algorithm on a synapsin channel. The elav channel (green) has been merged to the image for visualization of nuclei. The path may remove nuclei belonging to the central brain. . . . .	43

3.13 Maximum intensity Z projection. Merge of detected nuclei on the third channel (magenta), and third channel (green). The white overlay of green and magenta shows that the nuclei have been successfully detected. . . . . 44

3.14 Number of nuclei detected in the third channel of several brains of different lines. Lines are referenced with numbers on the x-axis. . . . . 45

3.15 Nuclei (green) and GFP (magenta) channels in a 3D reconstruction of a brain with 9895 detected nuclei. . . . . 45

3.16 Nuclei (green) and GFP (magenta) channels in a 3D reconstruction of a brain with 1165 detected nuclei. . . . . 46

3.17 Anti- $\beta$ galactosidase channel of one brain registered to another brain of the same Gal4 line (DDC). The arrows show locations where nuclei are thought to correspond to each other and are close to each other. . . . . 47

3.18 Anti- $\beta$ galactosidase channel of three brains from the same line registered to the average reference brain. The arrows show locations where nuclei are thought to correspond to each other and are close to each other. . . . . 47

3.19 3D reconstruction of the Anti- $\beta$ galactosidase channel of three brains from the same line registered to the average reference brain. The clear clusters enable to judge the nuclei location variability. . . . . 48

3.20 Brain stained with antibodies anti-NCadherine and anti-elav. The signal of anti-NCadherine is very weak compared to the signal of anti-elav. . . . . 49

3.21 Concentric rings used to compute the average intensity. White:  $r(7)$ , black:  $r(4)$  . . . . . 50

3.22 Intensity profiles of representative locations based on average intensity of concentric rings . . . . . 51

3.23 Glia detection with a threshold value of 30. Left: elav channel(red), repo channel(blue) and detected dots(green); Middle: elav channel; Right: repo channel . . . . . 53

3.24 Repo (blue), elav (red) and repo nuclei detected as glia (green). The nuclei detected as glia are only in regions with dense elav staining. . . . . 54

3.25 Repo (blue), synapsin and elav (red) and dots detected as glia (green). The nuclei detected as glia are only in regions with dense elav staining. . . . . 54

3.26 Two kinds of nuclei detected as glia in the neuropile region. Top: first sort; Bottom: second sort . . . . . 55

*List of Figures*

3.27	3D ring splitted in 8 different quadrants. Four quadrants are represented with different colors. The center is represented in magenta . . . . .	56
3.28	Result of the glia detection in a dense elav labeled region. Red: registration channel (synapsin, elav); Green: third channel; Magenta: nuclei detected as glia . . . . .	58
3.29	Average percentages of nuclei detected as glia for two sets of parameters . . . . .	59
3.30	With the parameters (20,5), three additional glial cells are detected. Red: registration channel; Magenta: detected glia; Green: nuclei channel . . . . .	59
3.31	GFP channel (magenta) and nuclei channel (green) merged in an image. Red arrows indicate nuclei without GFP expression, and the yellow arrow shows a nuclei with GFP expression . . .	60
3.32	GFP channel of a weak line. The range of the intensity has to be set very narrow, leading to a lot of noise . . . . .	61
3.33	Intensity histogram of Figure 3.32 . . . . .	62
3.34	Histograms of average intensities of spheres centered on the center of mass of detected nuclei with two different radius . . .	62
3.35	A lacZ blob with an extremely weak GFP signal in it. The highest pixel intensity of the blob is fifteen. . . . .	63
3.36	Histograms of spheres maximum intensity centered on the center of mass of detected nuclei. Left: same radius as the nuclei; Right: twice the nuclei radius . . . . .	64
3.37	A dendrite or axon of a GFP expressing cell labeled with anti-GFP. It represents one of the weakest signals we are able to detect manually. . . . .	64
3.38	Typical profile of a lacZ nuclei with GFP expression. The GFP expression is located in the plasma membrane and the lacZ expression is located in the nuclei. . . . .	65
3.39	Histogramm of the number of pixels having an intensity greater than 10. . . . .	66
3.40	Average percentage of remaining nuclei after nuclei removal based on GFP expression in fourteen lines with five different values for the score parameter. . . . .	68
3.41	Histogramm of the removal rates of 48 different lines (122 images) . . . . .	69
3.42	Example of nuclei removal based on GFP expression. Green : GFP; Magenta: detected nuclei (left) and remaining nuclei (right); Red: nuclei channel . . . . .	69

3.43	Nuclei removal based on the GFP expression. Green: GFP channel; Red: Nuclei channel; Magenta: remaining nuclei presenting GFP signal . . . . .	70
3.44	Screenshot of the visualization plugin for the density map developed for the software Fiji. The user can select on the table on the left the lines to be displayed. . . . .	71
3.45	Two density maps displayed with different radius for the nuclei. Left: radius 20 voxels; Right: radius 40 voxels . . . . .	71
3.46	Three-dimensional view of a density map. . . . .	72
4.1	Line with a high amount of glia. It has been manually checked on the registration channel that most of the nuclei of this picture are glia. Green: GFP channel; Red: nuclei channel . .	74
4.2	Line with a high amount of glia. Green: GFP channel; Red: nuclei channel; Magenta: remaining nuclei presenting a GFP signal . . . . .	75
4.3	Nuclei density map of 20 different lines . . . . .	77

# List of Tables

1.1	Principal characteristics of available antibodies . . . . .	13
1.2	Staining retained for the large scale screening . . . . .	15
2.1	Primary antibody concentrations added to blocking solution. .	25
2.2	Secondary antibody concentrations added to blocking solution.	25
2.3	Parameters for the contrast enhancement . . . . .	27
2.4	Plugins and bashes implemented and available at the subversion repository. . . . .	29
2.5	Main parameters used for the detection of nls-lacz expressing cells and glial cells . . . . .	32
2.6	Parameters used to remove glia based on the registration channel	33
2.7	Parameters used to remove lacz nuclei based on the GFP expression . . . . .	33
3.1	Antibody concentrations tested for the optimization of the anti- $\beta$ galactosidase staining . . . . .	36
3.2	Large scale staining and three possible stainings for the investigation of glias. . . . .	49

# Bibliography

- Abramoff, M. D., Magelhaes, P. J., and Ram, S. J. (2004). Image processing with imagej. *Biophotonics International*, 11(7):36–42.
- Cardona, A., Saalfeld, S., Arganda, I., Pereanu, W., Schindelin, J., and Hartenstein, V. (2010). Identifying neuronal lineages of drosophila by sequence analysis of axon tracts. *The Journal of Neuroscience*, 30(22):7538–7553.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271. 10.1007/BF01386390.
- Iwai, Y., Hirota, Y., Ozaki, K., Okano, H., Takeichi, M., and Uemura, T. (2002). Dn-cadherin is required for spatial arrangement of nerve terminals and ultrastructural organization of synapses. *Molecular and Cellular Neuroscience*, 19(3):375 – 388.
- Jefferis, G. S., Potter, C. J., Chan, A. M., Marin, E. C., Rohlfsing, T., Jr., C. R. M., and Luo, L. (2007). Comprehensive maps of drosophila higher olfactory centers: Spatially segregated fruit and pheromone representation. *Cell*, 128(6):1187 – 1203.
- Jenett, A., Schindelin, J., and Heisenberg, M. (2006). The virtual insect brain protocol: creating and comparing standardized neuroanatomy. *BMC Bioinformatics*, 7(1):544.
- Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. (2010). elastix: A toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*, 29(1):196 –205.
- Kolahgar, G., Bardet, P.-L., Langton, P. F., Alexandre, C., and Vincent, J.-P. (2011). Apical deficiency triggers jnk-dependent apoptosis in the embryonic epidermis of drosophila. *Development*, 138(14):3021–3031.

## Bibliography

- Lam, S. C. B., Ruan, Z., Zhao, T., Long, F., Jenett, A., Simpson, J., Myers, E. W., and Peng, H. (2010). Segmentation of center brains and optic lobes in 3d confocal images of adult fruit fly brains. *Methods*, 50(2):63 – 69. Spatial gene imaging in the brain.
- MATLAB (2010). *version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.
- Mukherjee, P., Gildor, B., Shilo, B.-Z., VijayRaghavan, K., and Schejter, E. D. (2011). The actin nucleator wasp is required for myoblast fusion during adult drosophila myogenesis. *Development*, 138(11):2347–2357.
- Overstreet, E., Chen, X., Wendland, B., and Fischer, J. A. (2003). Either part of a drosophila epsin protein, divided after the enth domain, functions in endocytosis of delta in the developing eye. *Current Biology*, 13(10):854 – 860.
- Peng, H., Chung, P., Long, F., Qu, L., Jenett, A., Seeds, A. M., Myers, E. W., and Simpson, J. H. (2011). Brainaligner: 3d registration atlases of drosophila brains. *Nat Meth*, 8(6):493–498.
- Peng, H., Ruan, Z., Long, F., Simpson, J. H., and Myers, E. W. (2010). V3d enables real-time 3d visualization and quantitative analysis of large-scale biological image data sets. *Nat Biotech*, 28(4):348–353.
- Perrimon, N., Noll, E., McCall, K., and Brand, A. (1991). Generating lineage-specific markers to study drosophila development. *Developmental Genetics*, 12:238–252.
- Pfeiffer, B., Ngo, T., Hibbard, K., Murphy, C., Jenett, A., Truman, J., and Rubin, G. (2010). Refinement of tools for targeted gene expression in drosophila. *Genetics*, 186(2):735–755.
- Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., Scully, A., Carlson, J. W., Wan, K. H., Laverty, T. R., Mungall, C., Svirskas, R., Kadonaga, J. T., Doe, C. Q., Eisen, M. B., Celniker, S. E., and Rubin, G. M. (2008). Tools for neuroanatomy and neurogenetics in drosophila. *Proceedings of the National Academy of Sciences*, 105(28):9715–9720.
- Robinow, S. and White, K. (1991). Characterization and spatial distribution of the elav protein during drosophila melanogaster development. *Journal of Neurobiology*, 22(5):443–461.



- Rohlfing, T. and Maurer, C. (2001). Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint. In Niessen, W. and Viergever, M., editors, *Medical Image Computing and Computer-Assisted Intervention at MICCAI 2001*, volume 2208 of *Lecture Notes in Computer Science*, pages 111–119. Springer Berlin / Heidelberg. 10.1007/3-540-45468-3-14.
- Rohlfing, T. and Maurer, C. R. (2003). Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. *IEEE Transactions on Information Technology in Biomedicine*, 7(1):16–25.
- Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721.
- S, H., K, I., Y, S., M, T., A, A., H, T., T, A., F, M., H, N., T, T., R, U., T, U., M, Y., and S, G. (2002). Getdb, a database compiling expression patterns and molecular locations of a collection of gal4 enhancer traps. *Genesis*, 34(1-2):58–61.
- Santella, A., Du, Z., Nowotschin, S., Hadjantonakis, A.-K., and Bao, Z. (2010). A hybrid blob-slice model for accurate and efficient detection of fluorescence labeled nuclei in 3d. *BMC Bioinformatics*, 11(1):580.
- Simone, R. P. and DiNardo, S. (2010). Actomyosin contractility and discs large contribute to junctional conversion in guiding cell alignment within the drosophila embryonic epithelium. *Development*, 137(8):1385–1394.
- Spindler, S. and Hartenstein, V. (2010). The drosophila neural lineages: a model system to study brain development and circuitry. *Development Genes and Evolution*, 220(1):1–10.
- Strausfeld, Fischbach, I. J. (2007). Nomenclature of the fly brain. An attempt to merge the nomenclatures of Strausfeld, Fischbach, Ito and Jenett.
- Van Der Blik, A. M. and Meyerowitz, E. M. (1991). Dynamin-like protein encoded by the drosophila shibire gene associated with vesicular traffic. *Nature*, 351(6325):411–414.
- Wu, J. S. and Luo, L. (2006). A protocol for dissecting drosophila melanogaster brains for live imaging or immunostaining. *Nat. Protocols*, 1(4):2110–2115.

*Bibliography*

- Younossi-Hartenstein, A., Nassif, C., Green, P., and Hartenstein, V. (1996). Early neurogenesis of the drosophila brain. *Journal of Comparative Neurology*, 370(3):313–329.
- Zuiderveld, K. (1994). *Contrast limited adaptive histogram equalization*, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA.

# Erklärung der Selbstständigkeit

Hiermit versichere ich, die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die Zitate deutlich kenntlich gemacht zu haben.

.....  
Ort, Datum

.....  
Thomas Templier