Linköping University Medical Dissertations No 1458

**Mechanisms of Olfactory sensory neuron class
maintenance in *Drosophila* :**

It is all about design and equilibrium

Shadi Jafari

Linköping University

Department of Clinical and Experimental Medicine

Linköping University, Sweden

Linköping 2015

© Shadi Jafari 2015

Cover picture by Shadi Jafari

Published articles have been reprinted with the permission of the copyright holder.

Printed in Sweden by LiU-Tryck, Linköping, Sweden, 2015

ISBN: 978-91-7519-085-3

ISSN 0345-0082

حکمت وزیدن باد رقصیدن شاخه ها نیست بلکه امتحان ریشه هاست.

امام علی (ع)

"The wind does not blow to make leaves dance, but to test their roots."

Imam Ali (AS)

I was not born with knowledge, but, being fond of it, I was eager to seek it through diligence.

- Confucius

CONTENTS

<u>Abstract</u>	1
<u>List of papers</u>	3
<u>Abbreviations</u>	4
<u>Introduction</u>	5
<u>Cell, genome and gene regulation</u>	5
<u>Gene expression regulation</u>	6
<u>Transcription factors</u>	8
<u>DNA binding motifs</u>	10
<u>Combinatorial interactions of TFs</u>	11
<u>Cis-regulatory modules</u>	12
<u>Promoters, Enhancers/Silencers and Insulators</u>	13
<u>Binary vs rheostat regulation of gene expression</u>	14
<u>Limits of Combinatorial regulation</u>	15
<u>Epigenetic mechanisms</u>	16
<u>Effects and functions of histone modifications and epigenetic codes</u>	18
<u>Whole genome regulation and nucleus localization of expressing genes</u>	19
<u>Neuronal specification</u>	21
<u>Generating neuronal diversity</u>	22
<u>Olfactory system</u>	24
<u>Drosophila olfactory system</u>	24

<u>Odorant receptors</u>	28
<u>Drosophila Odorant receptors</u>	31
<u>Terminal selector genes</u>	33
<u>Gene Expression Robustness to environmental changes</u>	33
<u>Aims</u>	37
<u>Materials and methods</u>	39
<u>The Gal4/UAS/TARGET system</u>	39
<u>OR markers</u>	39
<u>RNAi methodology and Reverse genetic screening</u>	40
<u>Electrophoretic Mobility Shift Assay</u>	41
<u>Results</u>	45
<u>Paper I</u>	45
<u>Principal findings</u>	45
<u>Paper II</u>	47
<u>Principal findings</u>	47
<u>Paper III</u>	49
<u>Principal findings</u>	49
<u>Discussion</u>	51
<u>OSN class specification and maintenance</u>	51
<u>Repression modulates OR expression</u>	53
<u>Interacting subnetworks specify OSN classes</u>	54
<u>OR genes promoters are enhansosomes</u>	56
<u>Low-affinity motifs can be advantageous for gene regulation</u>	57

<u>Change in the TF composition of a cell can reprogram an adult cell to a new fate</u>	58
<u>Modern environments and Human diseases</u>	61
<u>Popular science abstract</u>	62
<u>How our neurons are doing what they should do?</u>	62
<u>Acknowledgements</u>	64
<u>References</u>	67

Abstract

How the cellular diversity of our body is generated and maintained is still a great mystery regardless of the wealth of research that has been done on this issue. The greatest complexity is found in the nervous system that contains a vast number of neurons and displays a great diversity in cell types and classes. For example the *Drosophila* olfactory system is a complex but defined set of neurons with extremely high specificity and sensitivity. The 34 OSN classes are each defined by their expression of a specific odorant receptor (OR). During development each OSN chooses one OR from 60 different OR genes in the genome to express. Furthermore, a cell is subject to immense challenges during its life cycle. Confronting each challenge the cell needs to perform its function and maintain its fate. OSNs continue to express the same OR during their whole life regardless of fluctuations in the environment.

Although the olfactory system is remarkably conserved across the phyla, it is still unclear how an OSN chooses to express a particular OR from a large genomic repertoire. In this thesis the final steps of the specification and diversification for establishing an OSN identity is addressed. We find seven transcription factors that are continuously required in different combinations for the expression of the *Drosophila* ORs. The TFs can in different background context both activate and repress OR expression, making the regulation more economical. We also imply that repression is crucial for correct OR gene expression. We further show that short DNA sequences from OR gene promoters are sufficient to drive OSN class specific expression. These regions contain clusters of TF binding motifs, which we show to be sensitive to any change in their composition or to changes of the internal or external environment. We demonstrate that the chromatin state is responsible for the clusters response to environmental challenges. We reveal that Su(var)3-9 controls the OSN response to environmental stresses. We address the epigenetic mechanisms that initiate and pertain the single OR expression to a single OSN class. Our results show that OSNs have an epigenetic switch marking the end of development and the transition to mature OSNs. This switch supplies the expression of a single OR gene.

List of papers

Paper I

Combinatorial activation and repression by seven transcription factors specify *Drosophila* odorant receptor expression

Shadi Jafari, Liza Alkhori, Alexander Schleiffer, Anna Brochtrup, Thomas Hummel and Mattias Alenius.

PLoS Biology (Vol. 10) March 2012

Paper II

Cis-regulatory mechanisms for robust olfactory sensory neuron class-restricted odorant receptor gene expression in *Drosophila*.

Shadi Jafari and Mattias Alenius

PLoS Genetics March 2015

Paper III

***Drosophila* olfactory sensory neurons have two phases of gene expression regulation**

Shadi Jafari and Mattias Alenius

Manuscript

Abbreviations

AL, Antennal lobe

OR, Odorant receptor

OSN, Olfactory sensory neuron

OE, Olfactory epithelium

OB, Olfactory bulb

TF, Transcription factor

TSS, Transcription start-site

GTFs, General transcription factors

RNAPII, RNA polymerase II

TBP, TATA-box binding protein

TAFs, TBP associated factors

TFBS, Transcription factor binding sites

PWM, Positional weight matrices

GRNs, Gene regulatory networks

CRMs, Cis-regulatory modules

LCRs, Locus control regions

NE, Nuclear envelope

NM, Nuclear membrane

NPC, Nuclear pore complex

CNS, Central nervous system

GPCRs, G protein-coupled receptors

IRs, Ionotropic receptors

SOPs, Sense organ precursors

Introduction

Cell, genome and gene regulation

In 1925, the pioneering cell biologists, E. B. Wilson, emphasized: “the key to every biological problem must finally be sought in the cell; for every living organism is, or at some time has been, a cell”. Cells are the ultimate elements of life. With a deeper understanding of cells, we can start to challenge the substantial problem of life: despite their apparent diversity, living things are fundamentally similar inside. This is mostly as a result of the common molecular code, the genome, in which the information about a whole organism is written.

A genome includes the complete set of Deoxyribonucleic acid (DNA) that carries the hereditary information of a eukaryotic organism and is divided into structures called chromosomes. Almost all cells of an organism hold a copy of the DNA that was in the fertilized egg from which the whole organism developed. Each cell, however, can have very distinct appearance, function and respond differently to extracellular stimuli. These differences are because cells use the genes differently. Genes are the working subunits of DNA. Each gene contains a particular set of instructions, usually coding for a particular protein or for a particular function. Genes are indeed used as templates to build functional cellular products in a process called gene expression. The processes behind gene expression are of complex nature, with many different proteins involved in recognizing, binding to, reading, and ultimately translating the messages within the genes, the DNA code, into fully functional proteins that are required in different processes of the cell.

The enormous complexity of cellular fate and the challenge the response to the environment makes is perhaps best appreciated in the mammalian nervous system. The mammalian nervous system is thought to contain thousands of neuronal cell types that are distinguished on the basis of unique location and morphology (Masland, 2004). The human CNS contains an immense number of cells ($> 10^{11}$) that are from different types. Estimates hold that there are at least 10,000 different classes of neurons and glia in the human nervous system, although the definition of neuronal class is debated (Muotri and Gage, 2006). Diversity exists between individual neurons of a type to produce neuronal classes. Moreover, neurons from a

class can differ from other classes in important molecular details, expressing different combinations of ion channels, for instance, providing cells with various excitation thresholds and distinctive firing patterns. The neurons are quite robust in nature and preserve their class for a lifetime. However, neurons are still greatly plastic and can change dynamically in response to both internal and external environmental fluctuations. The present flexibility is consequently another layer of challenge for the gene regulatory system to maintain a balance between flexibility and robustness.

The main goal of this thesis is to understand the relationships among as many as possible of different components from cell fate specifiers within a cell. It also aims to address how these components respond to different challenges. This requires all cell components to be measured at the same time. Unfortunately, despite all the recent innovation in molecular biology, it is excessively confounded to address all the mechanisms that are involved in the cell fate decisions and maintenance at once. Nevertheless a concise comprehensive overview will help us to better understand the living organisms and consequently help us to address the associated health problems more in depth and hopefully find new treatments for these problems.

Gene expression regulation

The abundance of different proteins in a cell at any given point is controlled by several regulatory systems that influence each other to varying degrees. Among these systems are the initiation complex, the cis-regulatory modules and the epigenetic mechanisms. These systems allow cells to respond to environmental changes and maintain their cell type specific expression patterns. The principal regulatory system is the regulation of the timing and rate of transcription initiation and elongation. The process of gene expression begins with transcription in the cell nucleus.

Regulated eukaryotic gene transcription involves the assembly of an initiation complex at the core promoter region and regulatory complexes at promoter-enhancer/operator regions (Tjian and Maniatis, 1994). In eukaryotes, the term ‘core promoter’ is often used to focus on the DNA region in the immediate upstream area of the transcription start-site (TSS), which is

Fig 1. Regulation of transcription. A summary of promoter elements and regulatory motifs.

assumed to bind the initiation complex, general transcription factors (GTFs), and RNA polymerase II (RNAPII). GTFs include the TATA-box-binding protein (TBP) and its associated factors (TAFs). The initiation complex transcribes genes that produce mRNAs, long ncRNAs, and a number of small regulatory ncRNAs which, by a combination of base pairing and interaction with proteins, regulate other RNAs. In the standard view of RNAPII promoter function (Fig 1), the core promoter consists of several binding motif elements around the TSS, which bind core components of the initiation complex (Deaton and Bird, 2011; Kadonaga, 2012; Ohler and Wassarman, 2010; Riethoven, 2010; Sandelin et al., 2007). The core elements are presented briefly in Fig 1. In the classical model, the regulatory input to the core promoter consists of transcription factors binding to sites, either in the promoter region within several hundred base pairs of the TSS (at proximal elements) or further away (at distal elements) (Fig 1) (Lenhard et al., 2012).

For the RNA polymerases to bind and start transcribing, several other facilitating proteins are needed. These include the so called GTFs which bind the promoter region of every gene. The binding of the general TFs on their own produces only low levels of transcriptional activity. This activity is increased or decreased by other sequence-specific TFs, estimated that are to

be around 1400 in humans (Vaquerizas et al., 2009), which bind to regions of the DNA called enhancers and silencers that can exist inside and outside the gene region.

Transcription factors

Approximately 10% of the protein-coding genes of most organisms are devoted to transcription factors, making them one of the largest classes of proteins in the cell. In most cases, a specified transcription factor identifies its cis-regulatory sequence/motif, which is different from those recognized by all the other regulators in the cell. Although there are occasions where different TFs from the same family can bind similar binding motifs.

Transcription factors are specialized proteins competent of interacting with DNA and either up- or down-regulating the transcription of genes. The TFs are highly regulated with respect to cellular location (cytoplasmic vs. nuclear), stabilization, and post-translational modification. TFs are divided into six major super families (helix–loop–helix proteins, zinc finger proteins, β sheet DNA recognition proteins, homeodomain proteins, leucine zipper proteins and helix–turn–helix proteins) based on their similar properties of recognition and binding to DNA binding motifs. In general, DNA binding domains of different members of a given TF super family are extremely well-conserved, but the rest of the protein readily diverges between homologs (Cheatle Jarvela and Hinman, 2015).

The surfaces of the TFs are extensively complementary to the special surface features of the double helix that display binding motifs' sequences. Because the major groove of DNA is wider and displays more molecular features than does the minor groove, nearly all transcription regulators make the majority of their contacts with the major groove.

Each transcription factor contacts with the DNA making a series of hydrogen bonds, ionic bonds, and hydrophobic interactions. Generally each individual contact is weak, but together the 20 or so contacts that are typically formed add together to ensure that the protein-DNA bond is both highly specific and very strong. This binding puts into action a series of responses that ultimately specify which genes are to be transcribed and at what rate.

Many DNA binding motifs do not contain sufficient information to be specifically picked out from their DNA background. Random appearance of a given binding sequence all over the

genome would be expected to be high. Although for complex TFs, like zinc fingers, the DNA binding motifs appear to be very specific. One way to increase fidelity is to form dimers (hetero- or homodimers) or build other complexes to produce cooperative binding at the DNA (Funnell and Crossley, 2012). For dimerization the two monomers make nearly identical contacts with DNA. This arrangement doubles the length of the cis-regulatory sequence recognized by the transcription factors and greatly increases both the affinity and the specificity of transcription regulator binding. Transcription regulators may form heterodimers with more than one partner protein; in this way, the same transcription regulator can be “reused” to create several distinct DNA-binding specificities. (Alberts, Johnson et al. 2014)

Most sequence specific TFs and the initiation complex interact via a general mediator complex and a number of proteins called co-factors that do not bind the DNA themselves. Co-factors are a broad class of multi-subunit proteins that can act as co-activators or co-repressors. The mode of action for co-regulators varies, some can interact directly with Pol II and general TFs, and others can interact with chromatin modifying factors such as histone modifiers or nucleosome remodelers, but also directly bind to nucleosomes with histone modifications. Often the protein–protein interactions between transcription factors and co-activators are too weak for them to assemble on DNA; however, the appropriate combination of cis-regulatory sequences can facilitate the assembly of these complexes on DNA.

An individual transcription factor can often participate in more than one type of regulatory complex. They can be a part of activator or repressor complexes. While the general TFs and the mediator complexes are common to the transcriptional machinery of every gene, the order of specific TFs and co-factors can vary between genes.

Each eukaryotic gene is therefore regulated by a combination of proteins, all of which must be present to a specific level to express the gene at its proper level and instability on the concentration of TFs and co-factors influences the timing and rate of transcription of genes, providing a mechanism of gene expression regulation.

DNA binding sites (motifs)

The basic elements of binding code are usually short (5–20 bp) DNA sequences recognized by TFs. The set of regulatory motifs of a gene determines its expression during development, tissue-specification and respond to various external signals (Matys et al., 2006). Mutations in the transcription factor binding sites (TFBS) are known to underlie several human diseases and cause a substantial component of the phenotypic variability within and across species (Wray, 2007). A comprehensive knowledge of TFBS is thus critical for understanding the mechanism of transcriptional regulation, disease etiology and phenotypic variability.

Identification of TFBS usually involves experimentally identifying binding sites, constructing a model or a motif to represent the set of binding sites for a TF and searching for novel instances of binding sites using the model. Additionally, binding sites for an unknown TF can be identified computationally through de novo motif discovery.

The first method for searching TFBS relied on a number of experimental techniques that determine DNase I hypersensitive regions. Hypersensitivity to the DNase I enzyme represents open chromatin regions that are likely to harbor functional TFBS. Subsequently several follow-up experiments can be done to define the precise boundaries of TFBS, such as deletion/mutation experiments referred to ‘promoter bashing’. Nevertheless, these experiments do not identify the associated TFs. To this end, several other techniques like the Electro-Mobility Shift Assay (EMSA), Systematic Evolution of Ligands by EXponential enrichment (SELEX) and protein-binding DNA microarrays have been developed to utilize the TFBS/TF relations (Elnitski et al., 2006). The most common high-throughput technique for in vivo identification of binding sites for a specific TF is chromatin immunoprecipitation of bound DNA followed either by hybridization (ChIP-chip) or sequencing (ChIP-seq) (Elnitski et al., 2006).

In the next step individual binding sites are assigned a quality score corresponding to the strength of experimental evidence. The experimentally determined binding sites for a TF provide a collection by which a concise representation, or a model, is designed that is referred to as a motif. This collection of sites can be represented as the consensus motif. The consensus model is defined by the degree of base pair match from a seed set of strings. In

contrast positional weight matrices (PWM) model provides a probabilistic representation of binding sites. Unlike the consensus representation, a PWM captures relative preference for all four bases at each position (Stormo, 2000). However, the PWM representation assumes independence among positions within a binding site which may not always be biologically reasonable (Bulyk et al., 2002; Man and Stormo, 2001).

Knowing the motif sequence for a TFBS it can be searched in the whole genome. Regardless, the characteristic problem with binding site search is that the binding motifs are short and degenerate, which leads to a high-error rate in a genome-wide scan both in sensitivity as well as specificity. A variety of strategies have been used (most notably, evolutionary conservation) to reduce the false-positive rate. A fundamental concept in biology is the conservation of cis-regulatory elements that drive developmental gene expression. Non-coding regions of the genome that are conserved across distantly related species are likely to be selected and are thus expected to harbor functional elements. Transcription factor binding sequences can diverge extensively between closely related species (Borneman et al., 2007; Schmidt et al., 2010), implying high turnover of cis-regulatory elements (Odom et al., 2007). Interestingly, regulatory mechanisms that specify certain cell types and organs can be preserved through the evolution (Davidson and Erwin, 2006). Furthermore, enhancers and other elements involved in development are often complex, and it is unlikely that they frequently arise *de novo* from nonfunctional sequence by random mutations.

Combinatorial interactions of TFs

Part of the complexity of eukaryotic gene regulatory programs is achieved through combinatorial interactions among TFs. The requirement for TF interactions is reflected in the genomic clustering of binding sites. By searching *Drosophila* promoter sequences for dense clusters of putative binding sites novel developmental genes have been discovered (Berman et al., 2002). Although searching for clusters of binding sites for a given collection of TFs is useful, it can result in high error rates without the knowledge of functionally interacting TFs. Thus, independent approaches are needed to determine the potentially interacting TFs. For instance, a greater than expected frequency of co-localized binding sites for two distinct TFs is indicative of their interaction (Hannenhalli and Levy, 2002; Pilpel et al., 2001). Also, if the genes that are putative targets of a pair of TFs have similar expression patterns across

multiple spatio-temporal contexts (e.g. cell types, developmental times or environmental perturbations), then the two TFs are likely to interact functionally (Banerjee and Zhang, 2003; Pilpel et al., 2001).

Cis-regulatory modules

To understand the transcriptional regulation it is important to identify the network of cis-regulatory modules (CRMs), which can be regarded as the regulatory code installed within the genome. CRMs encode the precise instructions for transcriptional activation and repression. Control of gene regulation involves the CRM architecture, the factors which modulate CRM activity and their ability to regulate gene transcription.

The definition of a CRM is the advanced form of the original definition of an enhancer as a cis-acting DNA sequence that increases the rate of transcription from a related promoter independent of its location or orientation relative to the RNA start site (Blackwood and Kadonaga, 1998). A CRM is therefore defined as a DNA sequence containing TFBSs clustered into modular structures (typically 50 bp to 1.5 kbp in size) that act to regulate various aspects of the transcription process and generate cell specific gene expressions. In this classification CRMs include, but are not limited to, enhancers, silencers, promoters, locus control regions (LCRs), and insulators (Bushey et al., 2008; Sandelin et al., 2007; Wallace and Felsenfeld, 2007; Wray et al., 2003).

A CRM affects the regulation by orchestrating the input from active transcription factors and associated co-factors expressed in a cell at a given time, and integrate them into one united transcriptional output. A single gene may contain multiple CRMs and the ability of these CRMs to function in a combinatorial manner is necessary to generate precise four dimensional expression patterns. Furthermore, the regulation of chromatin structure and nuclear organization plays an important role in the control of CRM function (Fedorova and Zink, 2008; Fraser and Bickmore, 2007; Jiang and Pugh, 2009). CRMs are further characterized by the ways in which they affect the overall probability, proportion and rate of gene transcription.

CRMs do not function in isolation but are components of large gene regulatory networks. Gene regulatory networks (GRNs) explain the gene expression states that direct a cell to

establish a particular fate (Levine and Davidson, 2005). GRNs are predominantly composed of intercellular signaling molecules, transcription factor proteins, and CRMs. Because these networks instruct the specification of a particular cell type or structure, changes to these networks result in the evolution of animal morphology. A single gene within these networks can contain many CRMs, each of which contributes to the final transcription rate. Interactions between these CRMs modulate the expression of a gene through time and space and provide a unique code for when, where, and at which level a gene is transcribed (Ben-Tabou de-Leon and Davidson, 2007).

Recent studies have revealed novel functional aspects at the level of CRMs. A new subtype of CRM has been very recently described. A variety of recently developed whole-genome assays suggested that many of the crucial developmental patterning genes in *Drosophila* are regulated by multiple enhancers that direct extensively overlapping patterns of gene expression and employ a similar regulatory ‘logic’ (Zeitlinger et al., 2007). The newly identified CRMs are sometimes termed ‘shadow enhancers’ because they map to more remote locations than the ‘classical’ or primary enhancers situated close to the gene (Barolo, 2012; Hong et al., 2008). Shadow enhancers ensure reliable expression of their related genes when cells are subject to genetic and environmental variation (Frankel et al., 2010).

Promoters, Enhancers/Silencers and Insulators

The differences and similarities of various classes of CRMs should always be considered when we refer to CRMs. Although they all are involved in transcription regulation each has a specified role in this process. Promoter sequences are typically located directly upstream or at the 5' end of the transcription initiation site. Promoter sequences define the direction of transcription and indicate which DNA strand will be transcribed; this strand is known as the sense strand.

Like promoters enhancer sequences are regulatory DNA sequences that enhance the transcription of an associated gene. Transcription factors can bind to enhancer sequences located upstream or downstream from an associated gene, resulting in stimulation or enhancement of transcription of the related gene. Enhancer sequences act upon genes on the same DNA molecule; however, enhancer sequences can be located thousands of base pairs

away from the transcription start site of the gene being regulated. Because DNA is folded and coiled in the nucleus, the enhancer may actually be located near the transcription start site in the folded state. Additionally, enhancer sequences can be positioned in both forward or reversed sequence orientations and still affect gene transcription. A silencer is another type of CRMs that inhibits transcription of the gene independent of position and direction.

“Insulator” is the name given to a class of CRMs that protect an expressing gene from regulatory noise of the surrounding and does so in two principal ways. The first way is by blocking the action of a distal enhancer on a promoter (Geyer and Corces, 1992; Kellum and Schedl, 1992). Enhancer blocking only occurs if the insulator is situated between the enhancer and the promoter, not if it is placed elsewhere. Therefore the insulator generates direction to an enhancer to only stimulate expression of genes located on its unblocked side. The second way in which insulators protect genes is by acting as “barriers” (Sun and Elgin, 1999) that prevent the progress of nearby heterochromatin that might otherwise silence expression. Some insulators are able to act both as enhancer blockers and barriers.

Binary vs rheostat regulation of gene expression

Classically CRMs can be divided into enhanceosomes, highly cooperative and coordinated units, or billboards that are more flexible functional units (Arnosti and Kulkarni, 2005). In the enhanceosome model the exact/highly cooperative arrangement of TFBSs and the structure of the complex are critical. Any minor change of a single binding site in the enhanceosome complex can nullify functionality. In contrast, the billboard elements have a flexible organization of TFBSs and their transcriptional output is measured as the sum effect of the bound transcription factors.

In another category of classification a CRM could be the binary response model in which the CRM will act as part of an “on/off” switch for transcription. In this all or nothing model the CRM would increase or decrease the proportion of cells transcribing a gene, but not the rate of transcription initiation (the binary model) (Walters et al., 1995). By contrast, the rheostatic response model points to CRMs as regulators of transcription rates quantitatively. In this model a CRM would increase /decrease the rate of initiation of transcription of its associated gene. The rheostatic model is traditionally the most widely used description of CRMs

(Fiering et al., 2000). However, none of the models is a sharp description of what could be found in the world of CRMs. Overall, the necessity for sharply regulated gene expression in space and time, as well as the continuous need for certain genes to be expressed, implies alternate forms of control.

A study by Rossi et al. showed that transcription from a reporter construct was rheostatic when either activator or repressor proteins alone were present in the cell. However, when both activator and repressor proteins were present and competing for the same regulatory element the transcriptional response was converted to a binary on/off switch (Rossi et al., 2000). In other words CRMs' function is dictated by the transcription factors bound to them, and the presence or absence of these transcription factors in the cells is dictated by the internal and external environments. In some contexts a CRM might function as any of the described models. Therefore the dual functionality of CRMs could be an important mechanism through which the same gene participates in multiple pathways and developmental contexts.

Limits of Combinatorial regulation

TFs specify gene expression pattern necessary for a unique cell identity and function. Systematically, TFs recruit epigenetic modifiers to specific DNA sequence motifs present in promoters and enhancers to impact chromatin state (Rosenfeld et al., 2006). A TF may contribute to enforcing a particular cell fate by simultaneously activating genes required for maintaining the function and identity of that cell while antagonizing lineage inappropriate genes (Qi et al., 2013).

One TF may exhibit completely different genome-wide binding patterns and regulate non-overlapping sets of target genes in different cell types (Pimkin et al., 2014). This control mechanism stresses the very detailed genetic control of nervous system development. In the past decades numerous cellular complexities has been shown in systems like olfactory system that show a monogenic expression of over 1000 odor receptors (Buck, 2000; Masland, 2004). This level of complexity raises the question about whether the spatial and temporal patterning is enough for an identification code of TFs and co-factors combination for individual neuronal subtypes. Indeed, very few TFs have been found to be restricted to small neuronal

populations (Gray et al., 2004) and distinct neuron classes might share the same TF (and co-factor) codes (Shykind, 2005).

Epigenetic mechanisms

The properties and characteristics of cells depend on the genetic factors, as well as on environmental factors that are linked by the epigenetic machinery. Epigenetic literally means “above genetic” and is defined as changes and patterns in gene expression, which can be inherited through mitosis and sometimes also meiosis, but which are not caused by any changes in the actual DNA sequence (Kouzarides, 2007). Epigenetic control enables modification of transcriptional regulation of gene expression in neurons. Epigenetic control in the mature nervous system guides plasticity processes and long-lasting cellular neuronal responses (Borrelli et al., 2008).

Table 1. Different Classes of Modifications Identified on Histones

Chromatin modifications	Residues modified	Functions regulated
Ascetilation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 K-me2 K-me3	Transcription, Repair
Methylation (arginines)	R-me1 R-me2a R me2s	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription

Overview of different classes of modification identified on histones. The functions that have been associated with each modification are shown.

Chromatin is the state of how DNA is packed within the nucleus. As a first step of condensation, 147 base pairs of DNA are wrapped around a specific protein core complex, an octamer of the four core histones (H3, H4, H2A, H2B), forming a DNA-protein complex

Fig 2. Chromatin is condensed DNA to histone and non-histone nuclear proteins and compresses to form a chromosome. (Tonna et al., 2010)

called a nucleosome (Fig 2), which is found in all eukaryote genomes giving the DNA a typical “beads on a string” like structure (Khorasanizadeh, 2004). Positioning of nucleosomes throughout a genome has a significant regulatory function by modifying the *in vivo* availability of binding sites to transcription factors (TFs) and the general transcription machinery and thus affecting DNA-dependent processes such as transcription, DNA repair, replication and recombination (Radman-Livaja and Rando, 2010).

Histones, and particularly their N-terminal tails, retain large number and types of modifications. There are at least eight distinct types of modifications found on histones (Table 1.). The most available information is regarding the small covalent modifications: acetylation, methylation, and phosphorylation. Modifications on histones are dynamic and rapidly changing. Acetylation, methylation, phosphorylation, and other modifications can appear and disappear on chromatin within minutes of stimulus arriving at the cell surface (Kouzarides, 2007). The N-terminal tails can be modified on several sites simultaneously. Typically one acetyl group binds each site while the methyl groups can occupy the sites in mono-, di-, or trimethyl form for lysines and mono-, or demethyl form for arginines to add more complexity to the epigenetic modifications (Kouzarides, 2007).

Effects and functions of histone modifications and epigenetic codes

Heterochromatin was originally defined as highly condensed chromatin by Heitz (Passarge, 1979). Heterochromatin is thought to consist of regular nucleosomal arrays, which inhibit access by nucleases and contain a high proportion of transcriptionally inactive repetitive sequences interspersed by relatively few genes (Elgin and Grewal, 2003; Grewal and Moazed, 2003). Heterochromatin is also subdivided into constitutive heterochromatin, that is always condensed, and facultative heterochromatin, which may decondense in some circumstances. Chromosomal regions around the centromeres and telomeres are examples of constitutive heterochromatin. Facultative heterochromatin interspersed along chromosome arms contain genes that should be silenced from a certain point in development onwards. In organisms with large genomes, constitutively heterochromatic regions are also found along chromosome arms. Euchromatin is considered to be open because of irregular nucleosome spacing. This open chromatin is relatively gene rich and is potentially transcriptionally active (Elgin and Grewal, 2003). However, these differences are not always clear-cut.

Two models have been proposed to explain how the histone modifications act by. The first model suggests that modifications mainly result in a change in chromatin structure by a change in net charge of histones (Zheng and Hayes, 2003). Acetylation is associated with a reduced positive charge of lysines on histones, leading to less interaction with the negatively charged DNA (Shahbazian and Grunstein, 2007). The second model suggests a “histone code”. In this model different histone modifications act sequentially or in combination to

achieve contexts that regulates biological outcomes (Jenuwein and Allis, 2001). The effects of a modification on transcription are contextual dependent. Also, combinations of repressive and activating modifications co-exist in certain chromatin environments.

Among biologists acetylation and phosphorylation are generally associated with active transcription. Sumoylation, deamination and proline isomerization usually are found in transcriptionally silent regions whereas methylation and ubiquitylation are linked with both activation and repression.

Epigenetic marks of silent chromatin in higher eukaryotes are histone hypoacetylation, di- or trimethylation of lysine 9 at histone H3 (H3K9 di- or trimethylation) as well as cytosine methylation (Fischle et al., 2003a; Fischle et al., 2003b; Grewal and Moazed, 2003). Euchromatin is characterized by histone hyperacetylation and dimethylation of lysine 4 at histone H3 (H3K4 dimethylation) (Fischle et al., 2003b).

Recent studies have described chromatin regions further into five chromatin states (Filion et al., 2010). Heterochromatin can be subdivided into three chromatin types: GREEN (HP1 and H3K9me enriched), BLUE (PcG protein and H3K27me3 enriched) and BLACK (very low transcriptional activity enriched in noncoding elements). Also the deacetylase Rpd3 occurs in the GREEN and BLUE chromatin. Euchromatic regions contain RED and YELLOW chromatin with high levels of mRNA and RNA polymerase. H3K9me2 and H3K27me3 marks are low while H3K4me2 and H3K79me3 levels are characteristically high. Only RED is enriched in H3K36me3 and the chromodomain-containing protein MRG15.

Whole genome regulation and nucleus localization of expressing genes

In differentiated cells, the open chromatin and associated expressing genes are likely to be dispersed within the nucleus whereas the closed chromatin is more peripheral as resulting from the association of the heterochromatin protein HP1 with lamina at the nuclear envelop.

During cell differentiation, individual genes, alongside larger chromosome regions are repositioned within the nuclear space. This repositioning correlates with tissue specific gene expression profiles (Bickmore and van Steensel, 2013; Schneider and Grosschedl, 2007).

Fig 3. Models for changes in nuclear envelope-genome interactions during specification.

(A) Repositioning of cell type specific genes: Going through the path to a fully developed cell genes required for pluripotency or an alternate differentiation pathway are repositioned to the transcriptionally repressive nuclear periphery. Entailed genes for differentiation or cell type maintenance are kept in the nuclear interior. (B) Expression of cell type specific regulators during cell differentiation repositions chromosomes or nuclear territories to the nuclear periphery, influencing their transcriptional activity. (C) Genes necessary for cell type specification can be repositioned to the NPC for transcriptional activation (black arrow) or other regulation (white), such as establishment of chromatin boundaries or non-expressed genes (Talamas and Capelson, 2015)

Gene repositioning in part depends on losing or gaining interactions with major nuclear compartments such as the nuclear envelope (NE). Accumulating evidence has demonstrated tissue specific presence and functions of various NE components. Components of the NE, including the nuclear lamina, the nuclear membrane (NM) and the nuclear pore complex (NPC), come in close contact with the underlying genome to work together and execute tissue specific gene expression programs (Akhtar and Gasser, 2007; Amendola and van Steensel, 2014; Arib and Akhtar, 2011; Van de Vosse et al., 2011).

Existing knowledge supports a model in which cell type specific interactions between the NE and the genome, carry out regulatory functions which contribute to the correct establishment of tissue specific gene expression (Fig 3). For tethering heterochromatin to the nuclear periphery and for repositioning critical developmental genes to a silencing nuclear compartment tissue specific expression of Lamin isotypes are necessary (Fig 3A). In this process inner nuclear membrane proteins, have the ability to interact with chromatin bound regulators and histone modifying complexes. Another highly cell type specific component of this process are the nuclear envelope transmembrane proteins that likely drive the reorganization of chromosomes and large genomic regions needed for certain paths of differentiation (Fig 3B). Finally, components of the NPC are functionally implicated in regulation of developmentally induced active genes and in setting up boundaries between chromatin domains (Fig 3C) (Talamas and Capelson, 2015).

Together the inter-connected roles of nuclear compartments are essential for cell fate determination. That is why Talamas and Capelson call the NE composition another “cellular code” for specifying tissue specific gene expression programs through its contacts with the underlying chromatin.

Neuronal specification

The neuronal heterogeneity and the interconnections in nervous system of each living organism are different. Such complexity is the result of millions of years of evolution. The mouse brain contains approximately 75 million neurons (Williams, 2000). The human brain is composed of approximately 100 billion neurons. The total number of synapses in the human neocortex is estimated to be 0.15 quadrillion (10^{15}), and a typical neuron has

approximately 5,000–200,000 synapses (Pakkenberg et al., 2003). Thus, understanding cell-type variation in the nervous system is an important question regarding the existence of a large number of specialized types with unique molecular and functional assets, morphologies and axonal projections.

The central nervous system (CNS) neurons are generated from a limited population of multipotent neural progenitors. A neural progenitor goes through repeated asymmetric cell divisions during development. In each division, the progenitor self-renews itself, while producing a post mitotic neuron, or an intermediate precursor that can undergo limited rounds of self-renewal divisions to produce a series of neurons. Several studies have implied that early neuronal fate determination and cell type specification are dependent on the function of Neurogenin family proteins in many neural tissues, including the cortex, cranial sensory ganglia, OSNs and the spinal cord (Yuan and Hassan, 2014).

A central challenge in the determination of neuronal cell fate is the correlation between external and internal environmental information. As cellular context change over time, the cellular response to external cues will vary between progenitor cells. As a result, the birth of a neuron at a specific time and position determines its identity. The diversity of single neurons provides the underpinning for how neuronal circuits operate; therefore, it is essential to identify the participating neuronal subpopulations, and to determine the functions of the neurons at the cellular and molecular levels.

Generating neuronal diversity

Previous studies have suggested three major strategies for generating neuronal diversity. First, the CNS neural progenitors are heterogeneous and produce discrete neuronal lineages. The fate mapping experiments in mice, for instance, show that fluorescent-dye-labeled progenitors in medial and lateral ganglionic eminences, when transplanted into non-labeled host brains, give rise to cortical interneurons and striatum plus olfactory bulb neurons, respectively (Wichterle et al., 2001). Importantly, it is the original site of the donor tissue but not the site of transplantation that predicts the fate of the transplanted tissue, suggesting that progenitors in these regions of the brain are cell-autonomously different. This is further supported by in vitro cell culture experiments in which, cultured cortical and retina neural

progenitors still generate neuronal lineages that are indistinguishable from those they produced in vivo (Cayouette et al., 2003; Cayouette and Raff, 2003; Shen et al., 2006).

Fig 4. A combination of several systems acting in different substrates (DNA, RNA or proteins) might be involved in neuronal differentiation. (Muotri and Gage, 2006)

Second, a given multipotent neural progenitor often produces different types of neurons in an invariant order; thus, neurons acquire distinct temporal identities according to their birth-order within a lineage. Such birth-order dependent temporal identities are likely to be predetermined before the neurons were born (Baumgardt et al., 2007; Thor, 2013). The cortical neurons shortly after their final mitotic division migrate to the layer typical of their birthday, even when they are transplanted into an environment in which host neurons are destined for other layers (Desai and McConnell, 2000). Moreover, cultured cortical neural progenitors still produce layer-specific neurons in appropriate order and timings (Shen et al., 2006). This further suggests that the specification of the temporal identity largely relies on cell-intrinsic cues, although environmental cues have also been suggested to modulate the temporal identity of cortical neurons (Frantz and McConnell, 1996).

Third, at its final division, a neural progenitor or intermediate precursor can produce two sibling postmitotic cells with distinct cell fates, a process called binary cell fate specification (Huang et al., 2007; Schneider and Bowerman, 2007). The core of binary cell fate specification lies in the Notch signaling pathway and specific TFs so called terminal selector genes (Endo et al., 2007). The lineage-dependent functions and outcomes of terminal selector genes are likely due to the intrinsic property each neuroblast acquires. The intrinsic property

can be some lineage specific factors that interact with the terminal selector genes, or lineage-specific epigenetic modifications that make the genome respond to the terminal selector genes differently. Although there is no evidence for the lineage-specific epigenetic modifications in the fly, it remains an interesting possibility since several chromatin-remodeling proteins have been shown to be important for neurogenesis in both mouse and *Drosophila* (Yoo and Crabtree, 2009).

Olfactory system

During the endless process of evolution, animals have specialized sophisticated sensory modalities to interact with the external world. Among all these senses, the chemo sensation is the oldest one. Olfaction plays a key role for the life as all animals are embedded in a world of odors that function as signals able to trigger vital behaviors such as to eat, find mates, and avoid dangers. The olfactory system appears to be much more complex than visual or auditory system, which discriminate only between two simple parameters such as wavelength or frequency (Hallem and Carlson, 2006). The olfactory system performs the complex task of discriminating the quality and assessing the concentration of thousands of different odorants that differ in shape, size and electric charge (Persaud, 2013).

This complex identification is achieved through the interaction of volatile molecules with a large number of specialized Odorant Receptors (ORs), which are expressed in the Olfactory Sensory Neurons (OSNs) (Jiang and Matsunami, 2015). Researchers have revealed a system with a precise organization that translates a chemical structure to a topographic fingerprint of activated neurons. This generates maps of chemical properties of an odor that are converted into meaningful neural information, bringing forth proper behavioral response (Laissue and Vosshall, 2008; Mori et al., 2006).

***Drosophila* olfactory system**

In *D. melanogaster* adults, as well as in most insects, peripheral olfactory system has two pairs of organs, the antennae and the maxillary palps (Fig 5). Each antenna contains ~1200 OSNs, whereas each maxillary palp contains ~120 OSNs (Hallem and Carlson, 2004). Both organs are covered by sensory hairs, named sensilla, which house and protect the dendrites of up to four OSNs (Fig 5). A total of about 410 olfactory sensilla cover the antenna, while the

maxillary palp has about 60 olfactory sensilla (Martin et al., 2013). Although these organs respond to overlapping sets of odors, maxillary palp lies close to the labellum that is involved in the taste sense, and seems that the olfactory input via maxillary palp enhances taste-mediated behaviors (Shiraiwa, 2008). *Drosophila* maxillary palp is a structure that protrudes from the mouth parts and is covered by basiconic sensilla. Such as for all insects, the major olfactory organ of *D. melanogaster* is the antenna. The antenna is covered by three distinct morphological and functional classes of sensilla: basiconic, trichoid and coeloconic sensilla (Fig 5). Different sensilla types are distributed in a highly stereotyped fashion over the surface of the antenna.

OSNs send their axons to one of about 50 glomeruli located in the antennal lobe (AL) (Couto et al., 2005). Each glomerulus then sends the olfactory information to the higher centers of the brain. This functional organization is remarkably similar to that of the olfactory bulb in vertebrates where the signal is transmitted from olfactory epithelium neurons to the glomeruli in the olfactory bulb and from there to the cortex in the brain (Hildebrand and Shepherd, 1997).

The OSNs are thought to arise in a temporal process (Ray and Rodrigues, 1995). This process begins with the appearance of founding cells (FCs; Reddy et al., 1997) that arise in stereotyped locations and temporal sequence in the early pupal stage. An FC recruits adjacent cells into nascent sensillum clusters. A final cell division precedes the differentiation of sensillum cell types.

Fig 5. The *Drosophila* olfactory organs (A) The adult head with antennae and maxillary palps (B) four morphological types of olfactory sensilla on the antennal surface (C) *Drosophila* OSN morphology (Benton et al., 2006)

The adult antenna arises from the eye-antennal imaginal disk whose growth and patterning in the larva sets the stage for OSN differentiation in the pupa. Imaginal discs are small epithelial islands that are put aside during embryonic development. They give rise to adult structures during pupal metamorphosis. Homothorax is the primary homeotic factor that controls the specification of each disc into its respective structure (Morata, 2001). Most imaginal discs, including the leg and antennal discs are patterned by the expression of several genes and signaling pathways. For example, gain of function mutations of the homeotic gene *antennapedia* converts the antennae into legs (Schneuwly et al., 1987). *engrailed* (*en*) is expressed early in disc development and activates *hedgehog* (*hh*), which then activates both *wingless* (*wg*) and *decapentaplegic* (*dpp*). These genes are expressed in wedge-like and specify the ventral and dorsal fate, respectively, thereby, layering a new axis on top of the anterior–posterior compartments (Ling et al., 2014). At the center of the disc, where the expression of *wg* and *dpp* meet, the combination of these signaling pathways establish the proximal–distal axis of the antenna. Epidermal growth factor receptor (EGFR) signaling is critical for delineating the proximal–distal axis of the antenna.

Sense organs on the antenna derive from sense organ precursors (SOPs) in the eye-antennal disc. Disc development begins when the combinatorial action of the homeodomain transcription factors Homothorax, Extradenticle and Distalless, as well as the basic helix-loop-helix (bHLH) protein Spineless give the initial antennal identity to SOPs (Haynie and Bryant, 1986). Notch signaling controls the selection of a single progenitor from the undifferentiated field of epidermal cells in the antennal disc (Rhyu et al., 1994). *atonal* (*ato*), a bHLH transcription factor-encoding gene, specifies the progenitors that give rise to coeloconic sensilla, while *amos*, a second bHLH transcription factor-encoding gene specifies the basiconic and trichoid SOPs (Goulding et al., 2000; zur Lage et al., 2003). Basiconic and trichoid sensillar choice during development appears to be controlled by the dosage of the Runx family transcription factor Lozenge (Lz), which regulates *amos* expression; high levels of Lz produce basiconic sensilla, while lower levels produce trichoidea (Gupta et al., 1998). Diversification decisions regarding sensillar subtype identity are regulated by additional factors; Rotund (Rn), Dachshund (Dac) and Engrailed (En) (Li et al., 2013).

Once the precursor potentials are determined by the early patterning factors, each multipotent precursor undergoes several rounds of asymmetric divisions to generate 1–4 terminally differentiated OSNs in the same sensillum (Endo et al., 2007). Each consecutive asymmetric division from a single multipotent precursor is associated with binary segregation of possible cell fates. Notch signaling is a common pathway that is utilized for such binary segregation, and also contributes to proper segregation of OSN fates within each sensillum through lateral inhibition.

Because Notch signaling is used broadly in all sensilla, the intermediate precursor cells must retain a cellular memory for Notch signaling to act on during each asymmetric division. These mechanisms govern chromatin regulation. Recent works have connected the mechanism of Notch signaling in OSN specification to the chromatin modifiers Hamlet (Ham, a homolog of Prdm16) and C-terminal-binding protein (CtBP) (Endo et al., 2012). Ham functions as a repressor of Notch signaling and complexes with CtBP to reduce the amount of activating H3K4 methylation and increase the amount of repressive H3K27 methylation around Notch target genes. Further, the transcriptional co-repressor Atrophin (Atro) has been shown to regulate OR expression in Notch-on OSNs (Alkhorri et al., 2014). These data suggest that modulation of chromatin and epigenetic states are critical for proper segregation of alternate OSN fates.

In order to achieve the class-specific connectivity of 60 different OSN classes to non-overlapping glomeruli, OSN axons of the same type must converge onto the same glomerulus and synapse with the proper PNs. One simple solution to this problem is the use of olfactory receptors as instructive cues to govern glomerular targeting in order to compartmentalize connectivity in such a diverse system. Indeed, this strategy has been adopted in mammalian olfactory system (Chou et al., 2010; Feinstein et al., 2004; Imai et al., 2006; Nakashima et al., 2013). Despite the organizational similarities of olfactory system in mammals and *Drosophila*, *Drosophila* OSNs do not require receptor function for glomerular targeting. Indeed some of the early patterning factors, such as Hh, also pattern the axonal projections and glomerular connectivity of OSNs (Chou et al., 2010). In addition, molecular mechanisms that regulate OR expression, such as Notch signaling, and the transcription factors Acj6 and Pdm3, also play significant roles in controlling glomerular targeting (Komiyama et al., 2004).

It is thought that combinations of cell surface and guidance molecules expressed by OSN classes ensure proper connectivity in stepwise fashion that includes OSN axon trajectory selection, interglomerular and intraglomerular interactions that establish glomerular boundaries, and OSN–PN matching that ensures OSN wiring specificity (Barish and Volkan, 2015).

In *Drosophila* OR gene expression starts in the second half of OSN development and is considered as the final step of OSN differentiation. There are some known transcription factors that are expressed in later stages of OSN development and regulate OR expression in postmitotic OSNs. The more extensively studied of these factors is *Acj6*, which is required for OR expression in a subset of OSN classes (Ayer and Carlson, 1992). Recent studies identified 16 different *acj6* splice isoforms, each of which differentially regulates subsets of ORs in the maxillary palp (Bai and Carlson, 2010). *Pdm3* is also POU domain transcription factor that has primarily been investigated in the palps, where it is required for activation of some ORs (Tichy et al., 2008). Although *pdm3* is broadly expressed in the antenna and maxillary palp, its function in each OSN class is not clear.

Odorant receptors

Vertebrates and insects use similar strategies to recognize and discriminate odors. Large families of receptors detect odors in both vertebrates and insects, although the insects OR repertoire is considerably smaller than that of vertebrates. Moreover, the general tuning of ORs are also shared by mammals and insects.

Odorant receptors are part of the large, diverse superfamily of G protein-coupled receptors (GPCRs) with seven membrane-spanning domains. ORs were first explained in rats (Buck and Axel, 1991), as GPCRs based in part on physiological and biochemical evidence indicating that odorant binding activates G proteins. The vertebrate odorant receptor genes constitute large families, e.g. mouse has approximately 1200 odorant receptor genes (Johnston and Desplan, 2008).

Insect olfaction is mediated by receptors of multiple Types. The fly OR repertoire is considerably smaller than that of mammals, consisting of 62 ORs (Robertson et al., 2003). A few members of the Gustatory receptor (Gr) family are also expressed in olfactory organs,

where some have been found to mediate response to CO₂ (Jones and Reekie, 2007; Kwon et al., 2007). Recently, another family of ~60 receptors called Ionotropic receptors (IRs) has been identified, of which several are expressed in OSNs of coeloconic sensilla (Benton et al., 2009). GRs are predicted to contain seven transmembrane domains like ORs, whereas IRs are related to ionotropic glutamate receptors and are predicted to contain three transmembrane domains and a pore loop.

In vertebrates, each neuron expresses only a single receptor gene. This phenomenon is known as the “one receptor, one neuron” rule (Fuss and Ray, 2009). This rule forms the logical basis of the combinatorial strategy of odorant recognition. Furthermore, in vertebrates, a single OR allele is expressed in each OSN class (Chess et al., 1994; Malnic et al., 1999). The expression of a functional vertebrate OR turns on a negative feedback loop (Lewcock and Reed, 2004). This negative feedback downregulates the expression of the H3K9 demethylase Lsd1 (Dalton et al., 2013) and locks the expressed OR allele into a stable, robust expression state while suppressing the expression of other ORs (Lyons et al., 2013). Without monoallelic expression, vertebrate OR expression is not stable and robust within a single OSN class (Vassalli et al., 2011).

Early after the discovery of insect OR genes it was clear that insect and vertebrate ORs are distinct from one another. Insect receptors do not show similarity with other GPCRs. Indeed, bioinformatics approach unveiled receptors with seven-transmembrane regions that were specifically expressed in olfactory organs but the insect receptors adopt a membrane topology that is the reverse of GPCRs (Smart et al., 2008). Moreover, most fly olfactory neurons express two distinct receptors: a universal co-receptor, Or83b or Orco, and one of the common ORs (Larsson et al., 2004).

If we consider the Orco-ORs assembly a single receptor complex, then the one receptor–one neuron rule also applies to insects, although there are exceptions to this rule (Goldman et al., 2005; Hallem et al., 2006).

From each neuron, a single axon projects to the primary olfactory center in the brain; Olfactory Bulb (OB) in vertebrates and Antennal Lobe (AL) in insects; a specialized area of the forebrain that serves as the first station of the odorant information. In this center, these

A.

B.

Fig 6. A. The vertebrate nasal cavity (left) contains several olfactory subsystems (Kaupp, 2010)
B. In the olfactory epithelium (small box on the left side), the olfactory sensory neurons (OSNs) expressing a given odorant receptor, project their axons to the same glomerulus. In turn, each glomerulus is connected to a second-order olfactory neuron that transmits the signal to the cortex. The olfactory neurons that express the same receptor are represented by the same color. (Menini et al., 2004)

axons synapse with the dendrites of projection neurons within the glomeruli (Fig 6).

The other known characteristic of the olfactory system is that although neurons that express a given OR are randomly distributed throughout the olfactory epithelium, they converge their axons into specific glomeruli in the primary olfactory center (Buck, 2005; Mombaerts et al., 1996a; Mombaerts et al., 1996b), in a spatially invariant pattern (Mombaerts et al., 1996b; Mori et al., 1999; Ressler et al., 1994; Vassar et al., 1994) and create a functional unit called the OSN class (Fuss and Ray, 2009).

In such model, neuronal activity in a given glomerulus reflects the stimulation of one specific type of OR in the nose or the antenna. As an odor is typically a blend of molecules that can be recognized by different receptors, it is the combination of activated glomeruli that defines the unique neuronal representation of an odor (Ache and Young, 2005).

OSNs respond to odors with a sequence of action potentials that reflect the quality, intensity, and temporal structure of the odor stimulus (Pellegrino et al., 2010). The signals generated by OSNs are transmitted from the peripheral olfactory system to the higher centers of the brain, where processing takes place. To date, many efforts have been made to understand anatomy, molecular processes and behavioral responses that underlie olfactory perception with particular attention to mammals and insects. In particular, *Drosophila melanogaster* olfaction has been a field of deep interest. The outstanding reason for *Drosophila* to be an excellent model system for the study of odor coding is that its olfactory system is similar in organization to that of other insects and vertebrates yet is small in size and easily flexible to molecular, genetic, and electrophysiological analysis (Hallem and Carlson, 2004).

***Drosophila* Odorant receptors**

In 1999, a large gene family encoding candidate ORs was identified in *D. melanogaster* (Clyne et al., 1999; Gao and Chess, 1999; Vosshall et al., 1999). The *D. melanogaster* OR

gene family contains 62 genes and encodes a novel family of seven transmembrane-domain proteins. These genes express selectively in the olfactory sensory neurons in the antennae and maxillary palps. The OR genes are widely distributed throughout the genome, and some members exist in small collections (Robertson et al., 2003). The presence of these small clusters could suggest a mechanism of gene duplication that would determine the mechanism of expansion of this family (Ramdya and Benton, 2010). In some of these collections the genes share a higher degree of sequence similarity with each other than with the rest of the other OR genes like *Or22a/b*, *Or59b/c* or *Or85b/d*.

D. melanogaster ORs are highly different in sequence both within and between species. *Drosophila* clade ORs share less than 20% amino acid identity and do not show any primary sequence similarity to either mammalian ORs or any other known GPCR. Two major evidences suggest that *Drosophila* ORs are not divergent G-protein-coupled receptors (GPCRs), which is the class of proteins that mammalian ORs belong to (Keller and Vosshall, 2008; Thomas and Robertson, 2008). First of all, insect ORs possess a transmembrane topology that is the reverse of typical GPCRs, with the N-terminal located intracellular and the C-terminus located extracellular (Benton et al., 2006; Lundin et al., 2007; Smart et al., 2008; Wistrand et al., 2006). Second, although insect olfactory transduction mechanisms are still controversial, in contrast to mammals, the evidence for the involvement of G protein-mediated second messengers remains unclear. A further key difference in OR biology reflects the existence of a highly conserved OR, Or83b or Orco, that is co-expressed with other conventional ORs in most, if not all, olfactory neurons (Krieger et al., 2003; Larsson et al., 2004). Many studies suggest that Orco is indeed a chaperone; it forms a heteromeric complex with conventional ORs and helps the receptor localization in the OSN membrane where endures in this complex, suggesting that Orco could act as a co-receptor in olfactory signaling (Larsson et al., 2004; Nakagawa et al., 2005; Neuhaus et al., 2005).

By genetically introducing individual ORs into the mutant neuron and recording the electrophysiological response against a panel of 110 odorants (Dobritsa et al., 2003; Hallem and Carlson, 2004, 2006) and by OR reporter transgenes (Couto et al., 2005) it has been shown that most, if not all, antennal OSNs express only one functional OR and stress the existence of combinatorial receptor codes for odorants, similar to those in mammals. Hallem

EA et al. showed that many ORs respond to common ligands, reason for which one odor typically can activate multiple receptors. Nevertheless, rather than to be narrowly and broadly tuned, the insect ORs present a continuum of tuning breadths. Thus, each OR determines multiple aspects of odor coding in *Drosophila* (Hallem et al., 2006).

Terminal selector genes

To form a functionally cohesive mature nervous system the terminal differentiation of the post-mitotic neurons is a fundamental step. Each neuron gains its unique final properties in this step through the activation of distinct strings of terminal differentiation genes.

Antonio Garcia-Bellido first formulated the selector gene theory in 1975 (Garcia-Bellido, 1975) to describe a class of genes that directs the fates of groups of cells. After discovery of unique combinations of transcription factors to be required to maintain the expression profile of a particular mature neuron class in *C. elegans*, the term is now used more broadly to describe the function of genes that specify cell type and classes (Hobert, 2008). In the later concept terminal selector genes define assets of neuron classes without touching overall neuronal identity and act relatively early in developmental pathways. Terminal selector genes should be expressed throughout the life of a neuron and are constantly needed to regulate themselves and their downstream genes through terminal selector motifs (Transcription factor binding sites). The simplicity of terminal selector motifs offers a powerful space for evolution.

Gene Expression Robustness to environmental changes

Robustness supports the system to maintain its function despite external and internal perturbations. Robustness has been observed from the level of gene transcription to the level of systemic homeostasis (Kitano, 2004; Little et al., 1999). A developmental process can be considered robust if variation in this process is uncorrelated with variation in genetic, environmental or physiological conditions (Nijhout, 2002).

One of the key problems is to coordinate different steps when the environment dictates the speed of development like during body patterning and growth of the embryo. These developmental milestones have to be synchronized with environmental challenges for proper

development. One hypothesis is synchronized tissue patterning and whole-body development progress (Fig 7). The second model of robustness claims a relative progression of patterning to a developmental endpoint therefore the patterning would be independent of the duration of development (Fig 7). The third hypothesis states that tissue patterning may only be synchronized with key developmental events referred to as developmental milestones (Fig 7).

Fig 7. Hypotheses to explain how organ and whole-body development are coordinated. (a-b) Hypothesis 1: Whole-body development and individual tissue patterning are tightly coordinated throughout development. (c, d) Hypothesis 2: Whole-body development and individual tissue patterning are coordinated only at key physiological transitions. Note that the curve of the two lines between the developmental events is illustrative showing two ways the curves could differ under altered developmental conditions. Tissue patterning of a reference/condition is represented in black dashed lines; fast developers are represented in orange and slow developers are shown in blue. Stars and circles symbolize developmental events 1 and 2. (Oliveira et al., 2014)

Developmental milestones are behaviors or physical signs seen in animal infants and children at different stages of life as they grow and develop. Thus if the duration of development varies for example if the flies are kept in a low temperature, at each milestone a fundamental patterning would take place in a robust manner while showing greater variability between the milestones. Consequently, if patterning was to drift in rate with respect to whole body

development, developmental milestones balance the progression of patterning to achieve the correct stage by the onset of the milestone (Oliveira et al., 2014).

Several studies in *Drosophila* have identified the developmental milestones and their related gene expression and their correlation to different life stages (Devaud et al., 2003). Also several groups have explored how the genetic background contributes to the robustness of development (Matsuda et al., 2013; Nien et al., 2011).

Although the genetic background is, more or less, robust within a species the environmental conditions are tremendously variable not only over different generations but along one generations' lifespan. However, how environmental conditions affect the development within a species is less studied. There is even less information about mechanisms and principals involved in buffering environmental conditions post developmentally.

Aims

The aim of this project was to understand:

How various olfactory sensory neuron classes are specified in *Drosophila*? What factors, DNA binding motifs and regulatory processes are involved in this specification? (Paper I).

How the promoter organization generates specified expression? What role does the composition of the promoter play? How to generate an extensive combinatorial code? How much information is required for specified expression? What role does TF motif strength, combinatorial order and direction play for robustness of enhancer expression? How is stable expression encoded in the OR gene regulation? (Paper II)

What are the temporal aspects of OSN class specification? How do TFs and CRMs respond to different internal and external challenges over the life time of the flies? How do OR gene regulation change upon ageing? (Paper III)

In general, we tried to provide a broader understanding of the mechanisms of neuronal formation and specification, as well as robustness.

Materials and methods

The Gal4/UAS/TARGET system

The GAL4/UAS system is a standard technique for studying biological questions in *Drosophila*. Gal4 is a yeast transcription factor that binds selectively to the ‘upstream activation sequence’ (UAS). Depending of the choice of promoter that express the Gal4, any given gene downstream the UAS can be mis-expressed in a wide variety of spatial and temporal compartments. The transgene expression can then only be targeted in either the desired time or space, but not both. The TARGET system offers a method to direct transgene expression both temporally and spatially at the same time. Here, the conventional GAL4-upstream activator, UAS, is regulated via a temperature sensitive Gal4 repressor, GAL80^{ts} (McGuire et al., 2004). In the permissive temperature (18°C) the GAL4 transcriptional activity is repressed while at the restrictive temperature (29°C) the GAL80^{ts} is inactivated and repression is relieved.

OR markers

The OR markers are composed of OR predicted promoter regions with the open reading frame encoding a fusion protein consisting of the mCD8 extracellular and transmembrane domains and four tandem copies of GFP. The constructs are then inserted randomly in the fly genome and despite that the predicted promoters vary in sequence length they faithfully express in the right neurons (Couto et al., 2005). Experimentally proven, promoter fusions may not give a complete representation of the real expression pattern of a gene, both spatially and temporally. For assurance of the OR markers to be faithful presenters of endogenous ORs, the expression of the OR reporters were tested versus ORs’ endogenous mRNA by in situ hybridization. Only OR genes for which the reporter line and in situ hybridization labeled identical subsets of OSNs were chosen for future experiments.

The scattered expression of the neurons expressing one specific receptor in the olfactory tissue is detectable by fluorescent microscopy. We can also follow how the neurons’ afferents all converge upon the same target in the olfactory center, forming so called glomeruli. By this means an extensive anatomical, molecular and functional map of the *Drosophila* olfactory

system are formed, This methodology allows for direct visualization of changes in i) establishment of neural identity during neurogenesis (loss or gain of OR expression), ii) axon guidance and iii) target and synapse formation.

RNAi methodology and Reverse genetic screening

Traditionally forward genetic screening has been the method which genetics uses to find genes involved in biological processes. The method involves random mutagenesis, scoring of the desired phenotype and the identification of mutated gene by sequencing. This approach is time-consuming and genes with function early in development and other tissue could be missed. With the availability of complete genome sequences and the ability to inactivate genes' function a systemic exploring by reverse genetic have arisen in the last decade. Using reverse genetic strategies, a gene is targeted for disruption, and subsequently the researcher goes on to characterize the phenotypic consequences. In the fly, genes can be disrupted by deletions, RNAi or by site-specific knock-out and knock-ins (Okray and Hassan, 2013). A gene that is predicted to be involved in a certain process can be inhibited and the phenotype outcome investigated.

The inducible RNAi method is based on the principle of endogenous gene silencing mediated by the siRNA system upon expression of a double stranded RNA or by microRNA (Lee et al., 1993; Pillai, 2005). The approach was first developed in cultured cells and in vivo in plants and *C. elegans* (Fire et al., 1998). In these in vivo models the RNAi was systemic and could not be restricted to single cell types. In *Drosophila*, RNAi is cell autonomous (Roignant et al., 2003), and combined with the *Gal4-UAS* system for controlled gene expression; it allows the silencing of the gene of interest in almost any tissue. There are major advantages of this technique compared with the classical random mutagenesis screen. First, it circumvents lethality and positional cloning, thus preventing two major obstacles in classical random mutagenesis screens and reduces the work after the screen substantially. Second, the RNAi libraries make it possible to choose groups of genes to screen, addressing specific questions. We have taken use of the multiple RNAi libraries around the world the first established in the Dickson laboratory (Dietzl et al., 2007) the Vienna *Drosophila* RNAi Centre (VDRC), and the DGRC Kyoto Japan and the TRiP lines in the Bloomington *Drosophila* Stock Center (BDSC). The libraries make the system flexible with possibilities to include interesting genes

at any time point also during the phenotypic analysis, thus facilitating addressing any questions along the way.

However, no techniques are perfect and the main draw backs of RNAi is the risk of off targeting effects and that the knock down are not always complete, thus making the function of some genes difficult to reveal. To circumvent these draw backs we use different RNAi lines and when possible assess the knock down with RNA in situ hybridization or antibody staining. Furthermore, to increase the RNAi knock down efficiency, we have been using a sensitized background i.e. over expressing Dicer2 which is the effector molecule of the RNAi machinery (Bernstein et al., 2001). Over expression of Dicer2 has shown a 6% increase in false-positive results (Dietzl et al., 2007). To evaluate the reliability of Dicer2 overexpression we tested carefully for each driver and assay in all our experiments in our control samples.

Electrophoretic Mobility Shift Assay

The electrophoretic mobility shift assay (EMSA), also known as gel retardation or band shift assay, is a rapid and sensitive means for detecting sequence-specific DNA-binding proteins (Fried and Crothers, 1981; Garner and Revzin, 1981). This assay can be used to determine, in both a qualitative and quantitative manner, if a particular transcription factor is present within the nuclei of the cells or tissue of interest or to identify an unknown DNA binding protein which may control the expression of your gene of interest. The assay is based upon the ability of a transcription factor to bind in a sequence specific manner to a radio-labeled oligonucleotide probe and retard its migration through a non-denaturing polyacrylamide gel. Either crude nuclear extracts or purified factors can be used as a source of the DNA-binding protein.

Since its first publication, in 1981, several improvements and variant techniques of EMSA were reported. Originally described as a method to qualitatively detect protein-DNA interactions, gel retardation assays became one of the methods to map interaction sequences and domains not only in DNA but in RNA-protein interactions as well. EMSA was also adapted in order to allow the determination of quantitative parameters including complex stoichiometry, binding kinetics and affinity.

Several features made EMSA one of the most popular methods among researchers that study protein-nucleic acid interactions. Probably, the main advantages of EMSA when compared to other methods may be considered as follows: (1) EMSA is a basic, easy to perform, and robust method able to accommodate a wide range of conditions; (2) EMSA is a sensitive method, using radioisotopes to label nucleic acids and autoradiography, it is possible to use very low concentrations and small sample volumes (Hellman and Fried, 2007). Although less sensitive than radioisotopes, the wide variety of labels that can be used makes EMSA a very versatile method; (3) EMSA can also be used with a wide range of nucleic acid sizes and structures as well as a wide range of proteins, from small oligonucleotides to heavy transcription complexes; (4) Under the right conditions a gel retardation assay can separate the distribution of proteins between several nucleic acids within a single sample (Fried and Daugherty, 1998) or distinguish between complexes with different protein stoichiometry and/or binding site distribution (Fried and Crothers, 1981); (5) Finally, but not less important, it is possible to use both crude protein extracts and purified recombinant proteins enabling the identification of new nucleic acid interacting proteins or characterization of specific proteins and its targets.

Despite its sensitivity, versatility and usually easy to perform protocols, EMSA is often considered to bear a number of limitations. Dissociation can occur during electrophoresis since samples are not at equilibrium during the run, thus preventing detection. Additionally, complexes that are not stable in solution may be stable in the gel requiring very short runs so that the observed pattern relates to what happens in solution. EMSA does not provide a straightforward measure of the weights or entities of the proteins as mobility in gels is influenced by several other factors. Also, EMSA does not directly provide information on the nucleic acid sequence the proteins are bound to. However, this problem may usually be overcome using Footprinting approaches as described further ahead. Kinetic studies using EMSA are limited since the time resolution for a regular EMSA protocol consists of the time required to mix the binding reaction and for the electrophoretic migration to occur before the mix enters the gel. Only processes that have relaxation times larger than the interval required for solution handling are suitable for kinetic studies.

There are several alternatives to EMSA used in the analysis of nucleic acid-protein interactions with its own advantages and disadvantages when compared to EMSA. Footprinting, Nitrocellulose filter binding, Microfluidic mobility shift assay (MMSA), The yeast one-, two- and three-hybrid systems, Chromatin immunoprecipitation (ChiP) assays, The Systematic Evolution of Ligands by Exponential Enrichment (SELEX), Protein microarray, Nucleic acid microarrays, Ribonucleoprotein Immunoprecipitation and chip hybridization – Microarray (RIP-chip), Crosslinking and Immunoprecipitation (CLIP) and Photoactivable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) and High-Throughput Sequencing – Fluorescent Ligand Interaction Profiling (HiT-FLIP) are just a few of these alternative methods that each has its advantages and limitations compared to EMSA. Among mentioned EMSA alternatives the yeast one-, two- and three-hybrid systems and ChiP capture complexes in vivo and the binding reactions can be studied under different cellular conditions and at different time points. However they also have important limitations. Yeast hybrid systems are performed in yeast (*Saccharomyces cerevisiae*), which may not be the true endogenous context, and is limited to DNA-protein interactions. ChiP requires high-quality antibodies that are available only for a limited number of proteins. To circumvent this, epitope-tagged proteins could be used although it usually implies the introduction of modified genes into the endogenous locus in order to obtain expression at physiological levels. This method does not distinguish between proteins that bind directly to the genomic DNA and those that only interact with other proteins that do bind.

Development of EMSA alternative approaches was made possible, at least in part, by the availability of more sensitive and specific equipment and tools. Although EMSA cannot achieve a high throughput level it remains a valuable tool to confirm the detected interactions.

Results

Paper I

Several studies had been performed on the required transcription factors (TFs) for specification of each OSN class. These studies identified some regulators, principally *Acj6* and *Pdm3*, important for specification of some OSN classes foremost in the maxillary palp of *Drosophila* (Bai et al., 2009; Ray et al., 2008; Tichy et al., 2008). Having a clear picture of the *Drosophila* receptor-neuron-glomeruli map, these studies left the most interesting question pertaining to a systematic approach to the specification of the olfactory system. The completion of the *Drosophila* genome sequencing helped us to run a systematic RNAi screen on post mitotic OSNs with all available putative non-lethal TFs in *Drosophila*, 611 in total (www.flytf.org). In this screen we focused on the direct regulators of *Drosophila* OR expression. We identified a set of seven TFs that regulate the complete OR collection of the adult *Drosophila* olfactory system.

Principal findings

We identified 7 TFs that lead to a strong and penetrant loss of OR expression. The factors belong to different TF super families: *acj6* (POU-homeobox; Hox), *Eip93F* (Psq likehelix-turn-helix; HTH), *Xbp1* (bZIP), *onecut* (cut-Hox), *Sim* (PAS-bHLH), *Fer1* (basic helix loop helix; bHLH) and *zf30C* (C2H2; zinc finger). The last two genes, *Fer1* and *zf30C*, were not previously characterized (Fig 2, Paper I). All seven TFs were expressed in the adult antenna in various patterns. The patterns showed little restriction to domains or sensilla groups and none of the TFs were expressed in only one or only a few OSN classes. *acj6* and *Xbp1* were ubiquitously expressed in the whole antenna. The broad expression in the adult antenna suggested that the factors are part of a regulator network that control OR gene expression and likely not sensilla or OSN lineage specification. TARGET system experiments revealed that the factors are continuously required in the adult fly for expression of the ORs (Fig 3, Paper I). Furthermore, the broad expression of the TFs indicated that all ORs in the antenna are regulated by the 7 TFs. An expansion of the RNAi analyses to the majority of ORs expressed in the antenna revealed that the 32 ORs tested required at least one of the seven TFs for expression and for most of them, indeed, multiple TFs are required in different combinations

for the expression of the 32 ORs tested (Fig 4, Paper I). 17 of the 32 ORs required TF that generated unique combinations. The TF requirement demonstrated no correlation to OSN class location or sensilla group.

Next we asked if the identified TFs directly regulate OR gene expression by binding to upstream DNA sequences. Acj6, Xbp1 and Onecut have known DNA binding motifs in vertebrates. Bioinformatics search for the motifs upstream of OR genes revealed that most OR promoters included at least one binding motif for the required TF. It became also obvious that the region most proximally upstream the OR gene contained highest motif density and the highest sequence conservation in most ORs. Transgenes with only the proximal region of the cis regulatory region expressed in the right OSN class however half of them were also ectopically expressed in other OSN classes, indicating that the distal region acts to repress expression in some OSN classes (Fig 5D and 5E, Paper I). It was later shown that one co-repressor, Atrophin, segregated OSN classes through interaction with distal region of the promoter (Alkhorri et al., 2014).

Having identified the binding motifs for the TFs we constructed transgenes with different mutations of the OR regulatory region. Mutating one motif (*acj6* or *onecut*) in one OR promoter (*Or19a*) resulted in loss of expression for that particular OR in vivo, demonstrating that the motifs are necessary for OR expression (Fig 5B, Paper I). We further mutated *Xbp1* motif in the *Or98a* core promoter and our results suggest that the distinct activities of factors can be defined by the position of their binding motif in the regulatory regions of the OR genes (Fig 6, Paper I).

Paper II

In Paper I we demonstrated that there is not a direct regulatory “code” but, rather, an enrichment of motifs upstream of regulated OR genes (Jafari et al., 2012). Similarly, the identified TFs in *Drosophila* show little of the spatial and temporal specificity expected for combinatorial regulation, implying that it is neither the presence of a TF in an OSN nor the motif upstream of the OR that restricts expression to one OSN class.

Class-specific OR expression is generated in *Drosophila* in part by input from upstream repressive regions (Jafari et al., 2012; Miller and Carlson, 2010). Nevertheless other mechanisms involved in the stability of the OR expression in the adult flies were unknown.

In this paper, we addressed the cis-regulatory mechanisms that result in the precise and robust expression of *Drosophila* ORs.

Principal findings

We showed that high affinity binding motifs are not sufficient to drive expression. A pair of *acj6* Hox/Pou motif dimers placed upstream of a synthetic minimal promoter fused to *CD8:GFP* did not induce expression (Fig 1B, Paper II) indicating that motifs bound by other TFs or spatial arrangements are required to support *acj6* function. However, *Acj6* likely bound the low affinity Hox/Pou dimer cluster upstream *Or85a* gene and regulated the expression of *Or85a cluster*. We also showed that the *onecut* Hox/Cut dimers produced expression specific to Ab3a OSN class neurons, which express *Or22a* that is regulated by *Onecut* (Fig 1E, Paper II). The potential of *Onecut* motif dimers to drive the expression shows that motif dimers specific to one TF present sufficient regulatory information to specifically drive gene expression in a single OSN class.

In our next experiments we showed that a 36bp cluster from the *Or59b* upstream region with overlapping TF motifs (*acj6*, *Fer1* and *pdm3*) produced expression into two OSN classes: Ab2a, which expresses *Or59b* and innervates the DM4 glomerulus, and Ab7b, which expresses *Or67c* and innervates the VC4 glomerulus (Fig 2, Paper II). The expression by the cluster was regulated by these three TFs but this TF combination could not segregate the Ab2a and Ab7a classes. An interesting finding was that the expression of the cluster was sensitive to the loss of one copy of *acj6* but not of *pdm3*. The loss of expression was

attenuated by lowering the copy number of both factors (Fig 3, Paper II) revealing that the ratio of Acj6 to Pdm3 creates a restricted expression of *Or59b* cluster to two OSN classes. Our results also demonstrated that the composition and relative positions of motifs within the cluster define and restrict the expression of *Or59b* cluster (Fig 4, Paper II).

qPCR experiments results showed a slight increase in the mRNA levels of almost all the ORs upon exposure to short environmental stressful conditions. We attained that *Or59b* cis-regulatory region maintains a robust class specific expression during environmental strains. On the contrary of *Or59b* cis-regulatory region, the 36bp cluster lacks the regulatory information required to maintain fine-tuned class-specific expression during external stresses (Fig 5D, Paper II). The lack of fine-tuned regulatory information by the cluster was fully rescued by the addition of an E-box to the cluster. This rescue emphasizes the role of cooperative function of single motifs around the cluster in the stabilization of the expression produced by the cluster (Fig 6C, Paper II).

Last, we showed that Su(var)3-9, the enzyme that trimethylates H3K9 and is required for stable and robust class-specific OR expression in mice (Lyons et al., 2013; Magklara et al., 2011) modulates gene expression driven by the cluster (Fig 7A, Paper II). The sum of our experiments suggested that the Hox- and Pou-binding proteins open chromatin and let bHLH proteins to bind the E-box that induce expression. The bHLH binding competes with the Hox/Pou proteins and cooperation between additional E-boxes stabilizes bHLH binding and buffers *Or59b* cluster expression from variation in epigenetic and environmental states.

Paper III

In paper III we address the epigenetic mechanisms that initiate and pertain the single OR expression to a single OSN class over the life time of flies. Our results show that OSNs have an epigenetic switch marking the end of development and the transition to mature OSNs. This switch supplies the expression of a single OR gene. Our results further show that similar to vertebrates, H3K9 methylation status set by *Su(var)3-9* and *Su(var)3-3* (*dLsd1*) is a key factor in the commencement and maintenance of single OR expression in *Drosophila* OSNs.

Principal findings

Temperature shift experiments at different periods of the flies' life cycle generated different phenotypes. A switch at 80-90h after pupa formation (APF) resulted in ectopic expression of *Or59b* cluster in 70% of the flies. These flies showed no loss of expression phenotype. The loss of expression phenotype increased after the first days of the flies' life. These results indicate that there is a regulatory switch during the first days after hatching from a permissive to a more repressive state.

We also observed that OSNs have a rapid reaction to temperature shift and only 12 hours exposure to low temperature resulted in phenotypic variations (Fig 1, Paper III). Despite the rapid reaction to temperature stress the recovery phase was more extended and needed almost 72 hours. The 72 hours of ambient temperature did not recover the ectopic expression phenotype when the shift was done 80-90h APF. These results demonstrate that recovery of the gene regulatory specificity after environmental changes is a slow process and the changes at the end of OSN development tend to be stable or very hard to reverse.

Our results showed that the *Su(var)3-9* phenotype (Fig 3, Paper III) was dominant in the flies that were moved to 14°C when they were 80-90 hours APF. However, the *Or59b* cluster displayed a significant reduction in the number of the flies that lost the expression and an increase in the number of flies with the OSN restricted expression phenotype (Fig 3, Paper III). These results indicate that other epigenetic factors buffer the *Su(var)3-9* phenotype in the OSNs, when the olfactory system is stressed in the pupa stage.

Next we showed that in *Drosophila* like in vertebrates *Su(var)3-3* is required for initiation of *Or59b* cluster expression. *Or59b* cluster showed a loss-of-expression phenotype in *Su(var)3-3^{+/-}* flies indicating that demethylation of H3K9 is essential for transcription initiation. A

switch to low temperature during late pupa stage augmented *Su(var)3-3* phenotype (Fig 4, Paper III) that signifies the role of *Su(var)3-3* as a key epigenetic factor in the initiation of *Or59b* cluster expression.

Discussion

The regulatory mechanisms that control the genetic program of an organism comprise many factors. These include transcription factors, co-regulators, regulatory sequences in the DNA and chromatin structure and modification. The same mechanisms buffer cellular environment against surrounding challenges and make it possible for the cell to not only survive but also keep its identity and responsibility in the collection of all the organisms' cells. We tried to address just a small set of these mechanisms in this thesis.

OSN class specification and maintenance

There are few studies about the post mitotic factors that are required to maintain a neuron class and how they do so. In our first paper we showed that expression of each *Drosophila* OR gene is controlled by specific transcription factors. Given the large number of OR genes, and the singular expression of an OR gene in each OSN, the most cost-conscious model indicates a combinatorial code of transcription factors that control expression of different receptors in *Drosophila*. Taking the advantage of RNAi technology in *Drosophila* we identified seven transcription factors, required in various combinations for the class specific expression as well as maintenance of 32 odorant receptors in post mitotic olfactory sensory neurons. Bioinformatics analyses revealed binding motifs of the necessary transcription factors upstream odorant receptor genes expressed by them. These findings suggest that there is a direct regulatory function for these factors. These TFs can act as both activators and repressors of OR gene expression. Our set of TFs act only in late stages of pupal development around the onset of OR expression which leaves us with an open question about the primary regulators of OR genes to turn on the expression in each OSN class.

It would be interesting to know if the fundamental signalling pathways that are involved in the cellular fate specification are the keys to turning on not only the OR gene expression in the first place but also the final regulatory TFs.

There is evidence that suggests that these TFs alone cannot entirely explain how OSNs select a particular OR to express. Presence of binding motifs for specific TFs upstream of many OR genes does not correspond completely to the number of ORs that are affected in the TF knockdowns. Similarly, all seven selector genes in this study are expressed across the

antenna but still are required for the expression of some few ORs. This specificity could be as a result of combinatorial or competitive interactions at specific motifs, or the neuron-specific patterns of TF expression. Another scenario is that the TF activity is limited, such that only a small fraction of target genes can be activated in a given neuron. TF activity can be limited by low expression levels, or silencing by inhibitors (Kaern et al., 2005). In addition, many TFs can function as either a transcription activator or repressor. The presence of either co-activators or co-repressors, can also limit a TF to deterministically activate a specific OR gene in a given cell. More over different TF splice forms and the expression levels of each TF in different OSN classes might play a role. Finally, neuronal cell type specification is also subject to epigenetic control, through histone modifications, DNA methylation and the remodelling of chromatin architectures (Hsieh and Gage, 2005). While some TFs are capable of recruiting enzymatic co-factors to modify chromatin structures, other TFs only bind target promoters at open chromatin states (Maier et al., 2004). Therefore, cell diversity can emerge if a group of neurons inherit distinct chromatin states, even if they share the same TF codes. Recent years have also started to recognize the important roles of non-coding RNA, including microRNA and various types of small interfering RNAs (siRNA), in controlling gene expression at both chromatin and post-transcription levels. These new regulatory molecules will most probably contribute to the generation of OSN classes as well (Mattick and Makunin, 2005).

Deep sequencing of the antenna transcriptome including small RNAs might identify regulatory RNAs that can stabilize expression.

TFs from distinct families with divergent binding motifs make it possible to replace one TF from one family with another one from the same family. It might even be a competition among TFs as we have shown that low affinity binding motifs also are functional. We show such an example in our second paper where Pdm3 and Acj6 compete for the same Pou motif in *Or59b* cluster.

There are likely additional TFs, as about half the RNAi lines have limited function. It is likely that another round of screening could give a few additional TFs. Another perhaps more fruitful path would be to focus on the TFs in the two published antenna transcriptomes and go through a larger set of ORs for those.

Our results show that the OR regulatory TFs can be considered as terminal selector genes as they are required post-developmentally. They are expressed throughout the life of mature OSNs and are constantly needed to regulate ORs. They are regulating OR genes through their DNA binding motifs that are found upstream OR genes. We also show that these motifs are necessary for OR gene expression. It is not still clear if these TFs auto regulate their own genes expression or not. Auto regulation of terminal selector genes can maintain the cell fate.

We anticipate that a thorough lineage map of TF expression patterns in the context of developing OSNs as well as chromatin landscape of different OSNs and precursors will help reconcile our knowledge about OR regulation in Drosophila in the future.

Repression Modulates OR Expression

Systematic large scale expression analyses in mouse (Gray et al., 2004) and in *Drosophila* (Adryan and Teichmann, 2010) demonstrate that very few transcription factors have expression patterns limited to sets of specific neuron types or classes, implying that restricted transcription factor expression is not sufficient to dictate each neuron class in the nervous system. The expression analyses of the transcription factors in mouse and *Drosophila* did not include various isoforms. One of the selector genes in the OSNs, *acj6* has 16 different isoforms that provide a vast selection of combinations. However, most of the selector genes in the OSNs produce a single isoform and none of the terminal selector genes in the OSNs showed restricted expression to the OSN classes they produced. One simple solution to the enigma is that repression restricts OR expression to each class.

We found strong evidence that repression restricts OR expression. Our OR gene promoter analysis revealed that the identified TF motifs are enriched within a short region immediate upstream the transcription start site. These short regions are moderately conserved and sufficient for OSN expression. Upstream mouse ORs similar regions with motifs enrichments have been reported. These short cis regulatory regions were able to reproduce a normal expression in the OSN (Plessy et al., 2012; Vassalli et al., 2002). Interestingly, the proximal promoters with high motif concentration are sufficient for expression but in half of the cases exhibit ectopic expression in other OSN classes, which indicates that restricted expression is due to repression from distal parts of the cis regulatory regions. The nature of this long range

interaction remains to be identified, but one repressive co-factor complex, the dream complex, has been shown to control the expression of the carbon dioxide receptor *Gr21a* to the correct OSN class (Sim et al., 2012). Another example is the Atro complex that segregates ORs expressed in the Nba OSN classes and helps establish the Nab fate during OSN development (Alkhori et al., 2014)

Most of the motifs for the selector genes that we identified are found in the proximal region, implying a positive function. However, part of the repression is a direct function of the selector genes that switch to repress OR expression dependent on OR and OSN class. Some selector gene motifs are located outside the proximal region as in the core promoter next to the transcription start site and found to repress OR expression, which suggest that motif location can define the regulatory function of the factor. Yet, most selector gene motifs were enriched to the proximal region irrespective if they activate or repress expression, implying that each selector gene gets instructed if to activate or repress expression in that OSN class. One way to instruct the role of a factor is to build a decisive cluster that needs to be assembled in a certain way. The composition of these clusters and the distance or order of motifs defies the role of the factors as activators or repressors (paper II, Jafari and Alenius, 2015). The common cis regulatory structure with repressive and activator regions upstream the OR, indicates that a set of co-factors define terminal selector gene function and that terminal selector complexes rather than single genes define the OR expression.

Interacting subnetworks specify OSN classes

Although the neuron is extremely rigid its function is still plastic and features like the number of dendrites and synapses needs to be regulated. This plasticity is likely moderated by a separate set of TFs or different co-factors without the risk of the cell fate change. This way the cell fate is locked after these genes are turned on post mitotically. A locked fate lets the cell to use the signalling pathways that turned the terminal selector genes on to be reused for other purposes without affecting the cell fate as well.

Another issue is that all OSN classes express genes that define the type like neurotransmitters, ion channels and other determinants for function. One possibility is that there are type selector genes or OSN master regulatory factors which would regulate

expression of OSN type specific genes in all classes. These determinant genes for the OSN type express cilia proteins and ORs. None of the identified TFs in our screen regulated all ORs expression, demonstrating that none of the seven factors is a selector gene for the OSN neuron type. Of course the OSN type selector gene could be overlooked in the systematic screen. However, another tractable explanation would be that the neuron type is a sum of properties regulated by separate networks. Such networks could be found regulating OSN ciliary proteins which have more than 500 genes implicated for their maintenance. The cilia are lost in mutants of the winged helix transcription factor Rfx. Rfx motifs can be found upstream genes required for cilia formation. These results suggest Rfx to be a terminal selector gene for cilia function. However, Rfx was not found to regulate OR expression and suggest that cilia homeostasis is a separate unit, a subnetwork in the OSNs specification.

OSN type subnetworks can even be variable in different OSN classes. OSNs are cholinergic neurons. In some OSN classes Choline acetyltransferase, requires *Acj6*, one of the terminal selector genes. In other OSN classes Choline acetyltransferase is expressed independent of *Acj6* (Lee and Salvaterra, 2002). In other words, although expression of Choline acetyltransferase is a common feature of all the OSNs and therefore an OSN type property, its expression is regulated by different subnetworks in various OSN classes. This variation in subnetworks indicates that there is not a terminal selector subnetwork for the OSN type.

Xbp1 is a terminal selector gene for OR expression with an evolutionary conserved function as a selector gene for the Unfolded Protein Response (UPR) and is activated upon the increase of unfolded proteins. In this way *Xbp1* is part of several subnetworks, ORs production and the UPR. ORs are expressed in large quantities and *Xbp1* coupled folding and OR expression properties might be beneficial to the OSN. Even if the *Xbp1* related regulation of folding and transport in the OSNs needs to be further tested, the collected evidence suggests that subnetworks build the OSN type and that selector genes can be part of different subnetworks.

Subnetworks seem also to define each OSN class. All OSNs from one class first project their axons to one glomerulus and thereafter initiate expression of the OR. In our screen, knocking down of *Dll* produced loss of one OSN class axon outgrowth but did not affect OR expression for that class (data not shown). This phenotype proposes that axon out growth and

OR expression can be segregated and be two different subnetworks that have separate regulation. Segregation of the selector genes that regulate OR expression and the ones that regulate axon convergence to one glomerulus create a problem in how new classes are produced by evolution. Both a novel glomerulus and a novel expression pattern for that class have to happen independently. On the other hand, subnetworks could provide an elegant solution to the dichotomy between the plasticity required for memory formation and the stability of the neuron class. Stable subnetworks that possibly are self-maintained and have limited flexibility could maintain central properties such as OR expression without being affected by the continuous regulated flexible subnetworks that produce the homeostasis of axon and synaptic growth required for memory formation.

OR genes promoters are enhanceosomes

Enhancers have been shown to be the main agents of morphological diversity during evolution (Wittkopp et al., 2008). Any changes that affect TF binding properties to enhancers have the potential to modify expression of different genes in a tissue-specific manner without compromising the survival of the organism. Sequence alterations such as deletions, insertions and nucleotide substitutions in enhancers have been shown to be responsible for morphological diversity (Wittkopp and Kalay, 2012). The arrangement of TF binding motifs at enhancers is very complex. This complexity is the cause of high sensitivity to even tiny changes in regulatory sequences. Any changes can have significant effects in the transcriptional output by modifying binding affinity, binding site number or changing the spacing between TFs, among many other possible scenarios (Levine, 2010).

CRMs are divided into enhanceosomes or billboards (Arnosti and Kulkarni, 2005). In the enhanceosome model the position of binding motifs and the structure of the complex are critical. In this model all the TFs play equally important role and any change in the TFs or TFBSs can change the outcome expression. Our observations from different manipulations of *Or59b* cluster support the enhanceosome model. In contrast, the billboard elements perform as an information display. They have a flexible organization of binding motifs and their transcriptional output is measured as the sum effect of the bound transcription factors (Kulkarni and Arnosti, 2003). However, a billboard CRM could be a combination of one enhanceosome surrounded with spread single motifs. In our experiment a DNA sequence as

short as 36 bps can display a CRM role. We also show that surrounding E-boxes can increase the strength of expression while their presence is not crucial for the restricted expression of the cluster and gene at an ambient environment. While any change in the composition of the cluster interfered with its operation, the single E-boxes could be moved or removed without any considerable affect in the favourable environment. We therefore think that a more cautious look on the proposed billboard CRMs is necessary. This may reveal the presence of very short enhanceosomes within the billboard CRMs.

Low-affinity motifs can be advantageous for gene regulation

Our results show that the low affinity motifs can produce motif clusters that are recognized and bound by TFs. We show that the structure of the motif cluster can specify and also limit the function of broadly expressed TFs to regulate OR gene expression to a single OSN class. More complex clusters, like the *Or59b* cluster, combine the information input from several TFs that compete in binding to DNA. In such a context the low affinity of a motif becomes more critical as stable binding of one TF interferes with the high plasticity that is required. In our studies low affinity motifs exchanged for high affinity motifs did not induce expression (data not shown). Some have proposed that low-affinity genome-binding interactions can be unconditionally ignored (Fisher et al., 2012; Stamatoyannopoulos, 2012). Our findings suggest a risk of discarding functional sequences if we overlook the low affinity motifs.

Experimentally, binding site predictions have been supported by taking evolutionary sequence conservation into account (Hallikas et al., 2006). Short motifs are evolutionarily unstable and selection might regulate the birth and death of weak motifs. Even a complex cluster such as the *Or59b* cluster can only be found as a unit in the *melanogaster* clade and shows a large number of changes among species in the clade. This observation shows that ignorance of non-conserved low affinity motifs risks filtering out true regulators as lower-affinity binding sites seem to be less constrained with respect to sequence variation. Even in cases when the presence of the cluster is highly conserved this type of degenerate motif conservation can be easily missed. This is presumably because, for each non-consensus binding motif, there are multiple alternative sequences with similar affinity and thus equivalent functionality.

Moreover non-conserved Or59b cluster could be a result of altered TF or OR expression patterns in other species. It would be interesting and is now possible to investigate how similar (or different) the TF expression patterns and requirements for OR regulations are across all Drosophila species.

Although many predicted low-affinity sites are poorly conserved; overall TF occupancy on an enhancer may be maintained despite significant sequence turnover. This may occur either through the rapid gain and loss of individual sites, or through the maintenance of relatively weak binding affinity at a site that is unstable at the level of DNA sequence (Swanson et al., 2011; Tuch et al., 2008). Indeed, there are many more ways to apply a weak binding site than a strong site. Weaker binding sites, and the enhancers containing them, have a far greater flexibility without strongly impacting transcriptional output. A thermodynamics-based simulation of enhancer evolution has shown that there is a greater number of plasticity solutions using weak TF sites than using high-affinity sites for a given gene expression problem (He et al., 2012).

Change in the TF composition of a cell can reprogram an adult cell to a new fate

Oliveira et al. (Oliveria et al., 2014) showed greater variability in response to changes in temperature and alterations in physiology between developmental milestones. Exaggerated responses are observed regardless of the fact that tissue patterning remains robust against environmental and physiological perturbations. In our study cold exposure caused ectopic expression of the cluster in developing flies. In fully developed OSNs yet we observed stereotyped expression after low temperature encounter. The *Or59b* cluster insertion showed sensitive to H3K9 methylation status in both stages.

Briggs, King, and Gurdon in the 1950s demonstrated that irreversible genomic changes are not the reason of the stability of the differentiated state that occurs during differentiation (Briggs and King, 1952; Gurdon et al., 1958). It is now well established that TFs can induce distantly related cell fates (Vierbuchen and Wernig, 2011). Indeed the differentiated state of a cell needs to be constantly instructed by the intrinsic regulators of a cell (Blau and Baltimore, 1991). In other words, a change in the TF composition of a cell can reprogram an adult cell to

a new fate. In mice it has been shown that in vivo transition from exocrine pancreatic cells into insulin-producing beta cells by mis-expression of pancreatic developmental TFs does not require transition through a proliferative progenitor state (Zhou et al., 2008). It is therefore important to address if there is a possibility that environmental changes could change a cell TF composition and consequently the cell fate. Different aspects of OR regulation in *Drosophila* olfactory system points to such possibility. First, all known TFs that are required for regulation of ORs in *Drosophila* are expressed across the antenna (Jafari et al., 2012) and make it possible to use a new TF in unique situations. Second, the identified TFs can activate one OR in one class of OSNs while repress it in another class (Jafari et al., 2012) and they have the potential to change their role if the necessary elements are provided. Last, there are TF binding motifs in the promoter of ORs that are not required for the regulation of their downstream OR genes in the normal environment (Jafari and Alenius, 2015) but could be employed upon changes in the environment. Together these potentials make *Drosophila* fully differentiated OSNs prone to cell fate change by applying a combination of recruiting new TF binding motifs, shifting a TF's role from repressor to activator or applying a new TF program by altering the ratio of the genes expression. Our results together with other studies point to epigenetic machinery as the main coordinator of the mentioned element. Indeed with a modified heterochromatin status new motifs and consequently new TFs could be recruited.

It is generally believed that cells sequester unneeded genes into densely packed heterochromatin, which is thought to inhibit binding of transcription factors, thus preventing activation of transcription at that locus (Beisel and Paro, 2011). Experiments by Taylor and Jones provided initial evidence that somatic cells could also exhibit cell fate plasticity (Taylor and Jones, 1979). They found that treatment with 5-azacytidine, an inhibitor of DNA methylation, initiated spontaneous differentiation of fibroblasts into muscle and fat cells. This suggested that DNA methylation is important for preventing expression of genes that regulate differentiation into alternative lineages.

Our results show that a switch to low temperature direct after development (80-90h AFP) results in ectopic expression and with no loss of expression phenotypes. The ectopic phenotypes during the plastic phase indicates that the number of OSNs that can express *Or59b* exceed the ones in the *Or59b* class. Hence, it is possible that several OSN classes

compete to express different ORs and chromatin status is the final determinant of the cells' specificity.

Which epigenetic factors are the last determiners of OSNs identity needs to be addressed in the future.

As the cluster produces expression to the *Or59b* class despite the change in the ambient temperature during development we suppose the expression of *Or59b* is secured regardless of any environmental change in the early adult post developmental stage. Interestingly, after day 4, loss of expression phenotypes appears, indicating that there is a switch from a permissive heterochromatin status to a more restricted chromatin state which is indication of a more stable identity for the adult OSNs. We later observed that low temperature exposure of *Su(var)3-9^{+/-}* flies during OSN development buffered against environmentally induced loss of expression phenotypes. This buffering suggests a different mechanism from *Su(var)3-9* for the initial opening of the chromatin in developing OSNs. Our results support a model in which *Su(var)3-3* opens the chromatin to initiate the OR expression and lets the necessary TFs to bind the promoter. This binding of necessary TFs could be the trigger for the switch in the epigenetic regulation from *Su(var)3-3* in developing OSNs to *Su(var)3-9* in mature OSNs.

We also found that cold exposure of *Su(var)3-3^{+/-}* flies augmented the loss of expression phenotype. The OSN restricted phenotype in *Su(var)3-9^{+/-}* flies confronted with low temperature ponders in the possibility of *Su(var)3-9* acting as a defence mechanism in developing olfactory system against environmental fluctuations. Our model suggests the ectopic expression phenotype that we see in cold exposed developing OSNs is because cold exposure disturbs *Su(var)3-3* and consequently there is a downregulation in cellular *Su(var)3-9* levels.

*It would be interesting to continue studying about this epigenetic factors and their protecting role in the future. It would also be noteworthy to test if low or high temperature exposures shortly during the larva stage would put a mark on our neurons or not. This could be tested by our cluster that has been shown sensitive to temperature fluctuations. In addition, it is necessary to further evaluate the role of epigenetic factors in general and *Su(var)* family especially in the cell specification and robustness in the future.*

Modern environments and Human diseases

It is hypothesized that modern environments increase the genetic contribution to human disease risk. The genome of any living organism has a unique genetic variation that normally has little or no effect on phenotypic variation but that under atypical conditions, rare in the history of a population, generates heritable phenotypic variation (Paaby and Rockman, 2014). This genetic variation provides a pool of standing genetic alternatives poised to facilitate adaptation when the rare condition becomes common. Open chromatin during OSN development provides a flexible OR expression for OSNs and the general internal and external environment can then dictate the OR that is expressed. Flexible OR expression in a new environment provides more open options to the *Drosophila* population to survive new environmental challenges as the variability of the existing phenotypes increases in the newly hatched flies. After a bottle neck period the population can adapt to the new situation and only the best new phenotype for the new environment would survive. If the environment changes back the population will have a novel bottle neck. As starvation induces similar changes in *Drosophila* it is likely that even changes of food source or availability can broaden the phenotypes in the population and secure that novel food sources will be detected.

Popular science abstract

How our neurons are doing what they should do?

To be healthy we need each cell to be located in the right place and do the right thing. The most complicated organ in our body is our nervous system with billions of neurons. Each of these nerve cells has a specific function and location in a circuit and together they form our brain. Each nerve cell should get its specific instruction on what to do and where to go. To study and understand this process we looked at some well-defined characterised neurons. The olfactory system and olfactory neurons are known and distinguished by their specific odorant receptors. Mankind has thousands of different odorant receptor genes, but each olfactory neuron in the nose houses only one receptor.

Since the olfactory systems of all organisms basically are similar, we chose to study one of the smallest olfactory systems; the one of the fruit fly (*Drosophila melanogaster*). Unlike vertebrates' olfactory system which is very complicated, fruit fly has only 60 odorant receptors. The details of the construction of its olfactory neurons are well studied as well. The purpose of our studies is to understand how different components in a cell interact and together form the olfactory system. In the olfactory system we looked at how a neuron gets its properties. We also studied how it is determined which olfactory receptor one neuron should choose. We later looked at the neurons in different environments and studied how they protected their identity and correctly accomplished their role in an undesired environment.

For establishing this study we first studied almost all the possible proteins that could have a role in this process. The big challenge was to understand how they interact in the cell. We found that the election of olfactory receptor was determined by different combination of proteins that together make a complex code. Moreover, we found that part of the code is opening of the genome in the right place and the correct amount. Changes in opening of the genome can cause development of cancer, which makes us believe that they are important for both the emergence and growth of cancer. Now we can further compare the differences / similarities between cells in how they use the genome. We used different receptors and looked at the proteins that control other cell-specific processes such as the ability to maintain

their function and also to adapt to the changed environment. Disorders in such processes can cause uncontrolled cell growth as in cancer. Such disorders can also cause degeneration of cells, such as in Parkinson's. We hope that our research provides a better understanding of how the future can support the growth and regeneration of neurons both in body and in vitro. But also contribute to the development of new drugs for the treatment of cancer.

Acknowledgement

I came to Sweden in 2007, when I decided to change my path in science. I happened to meet so many incredible people in my path. I am so glad that they helped me and provided this opportunity for me. I want to express my sincere gratitude to everybody who contributed to my work.

First, I am forever grateful to my advisor, Mattias Alenius. Words cannot express my deep gratitude for his guidance, support and understanding during my graduate study. And his knowledge has always benefited me over the years.

I would also like to thank my co-advisors, **Stefan Thor** and **David Engblom** for their assistance and guidance with their broad knowledge and expertise in science in general and cell biology, genetics and *Drosophila* in particular.

The members of the Alenius lab will be my precious memory: **Olivia Forsberg**, for her great help with my research with her technical support in the lab and as a good friend. **Liza Alkhori**, I enjoyed a lot working with you and discussing science, politics and life. You helped me a lot in the science and life and Swedish! Wish you and your family all the best. **Johanna Karlsson**, I should claim that I admire your discipline and try to develop mine to that level. Thank you for very clean and in order lab and good luck with parenthood. **Anujaianthi Kuzhandaivel**, we had a lot in common while being very different. It was always great to have a person with such close cultural background in the lab. It was nice to talk about the things that sounded very strange to others! **Anita Öst**, I always get extra energy when you are around. Somebody told me that you were a perfect example of a scientist and I agree with that person. Thank you for being you! And good luck with your incredibly interesting research! **Sebastian W Schultz**, although it was a short time I learned critical thinking from you. It is always nice to hear that you are successful. **Eduardo Hatano**, you gave me very practical information about work and scientific world. Not everybody is that much generous. Wish you, Vivien and little Thomas a safe journey to US and a happy life always and anywhere you are.

Many people through different ways have provided help during my graduate study. In particular, I am grateful to **Jordi Altimiras**, a very smart person, for helpful discussions about science and introducing me to Mattias. Many thanks also go to his family.

All my fantastic colleagues in the past: **Annika Starkenberg, Anna Angel, Helén Ekman, Carolin Jönsson, Daniel Karlsson, Caroline Bivik, Erika Gunnar, Johannes Stratmann**, and my dear Iranian friends; **Behzad Yaghmaeian** and **Shahrzad Bahrampour**. You will be always in my heart. I have had the best possible “chats” with you!

I would like also thank all the inhabitants at Lab1 floor 13 especially **Mikael Sigvardsson, Marie Larsson** and **Jan-Ingvar Jönsson** for enlightening conversations.

All my wonderful present colleagues at cell biology department and especially floor 12: **Peter Strålfors, Åsa Jufvas, Cecilia Jönsson, Meenu Rajan, Siri Fagerholm, and Elin Nyman**. I could never wish for a better working environment. It has always been refreshing to talk to you in any situation, from hard and frustrating to exciting and happy publication occasions. I feel home while I am with you. Thank you all!

It was also nice to know **Mats Söderström, Annika Thorsell, Lovisa Holm, Abdul Maruf Asif Aziz, Daniel Nätt**, who thought me all I know about qPCR, and the rest of Thorsell group for nice and refreshing “fika” times.

All my dear friends, Iranians in particular, I would like to thank you all. You are so many that I cannot name you one by one but you know one by one that I love you for all you have done for me and wish you all the best in your lives.

My beloved family, my parents who dedicated all they had to raise me. Because of them I am who I am now. **My father** taught me to be strong and take risks if I wished to develop. I owe you all I have. Thank you for being supportive any time that I needed your support. **My mother** taught me to be kind and forgiving if I wanted to have peace in my life. Thank you for helping me any time that I needed help. My husband, **Farshid**, who has gone through all the up and downs with me and have supported me all the way. My little love, **Ermia**, who is the reason of all I do. I would never be in this position if I did not have you, Thank you! I wish you a bright, peaceful, successful and healthy future.

Alireza, Abdolreza, and Amirreza for being the best brothers I could have wished for. You never forget your sister even though she is far from you. You never forget to call me even when I am really busy to call you and that means always! I could not be as successful if I did not know that you were taking good care of our parents and I did not have to worry for them.

Papers

The articles associated with this thesis have been removed for copyright reasons. For more details about these see:

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-120992>