TRANSCRIPTOME-WIDE ANALYSIS OF IROC TARGET GENES IN THE *DROSOPHILA* OLFACTORY SYSTEM

by

İbrahim İhsan Taşkıran

B.S., Electrical and Electronics Engineering, TOBB University of Economy and Technology, 2015

B.S., Computer Engineering, TOBB University of Economy and Technology, 2015

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2018

For those who can ride their bikes with no handlebars…

ACKNOWLEDGEMENTS

This project was supported by the funds of TUBITAK project number 215Z369.

İbrahim İhsan Taşkıran was supported by a TUBITAK-BIDEP fellowship for his MSc. study.

ABSTRACT

TRANSCRIPTOME-WIDE ANALYSIS OF IROC TARGET GENES IN THE *DROSOPHILA* OLFACTORY SYSTEM

Sensing the environment is vital for an organism and it is provided by divergent sensory neurons. These neurons are generated from sensory organ precursor (SOP) cells that differentiate into neurons with diverse spectral sensitivity. The molecular mechanisms underlying neuron diversification in the olfactory system are largely unknown. *Drosophila melanogaster* olfactory system represents a stunning example of neuron diversification and it is an amenable system to study these mechanisms because of its reduced quantity and explicit definition. The *Drosophila* olfactory system comprises two olfactory organs, the antenna and maxillary palp, and 50 different olfactory receptor neuron (ORN) classes. ORNs are housed in sensilla and each sensillum contains one to four neurons.

In the framework of this study we focus on homeodomain transcription factors from the TALE class, *Iroquois Complex (IroC)*. The analysis of enhancer trap lines shows that *IroC* is expressed in both olfactory organs. In order to identify IroC target genes in the olfactory system, I used a genome-wide transcriptome approach and performed RNA Sequencing on antenna and maxillary palp tissue isolated from *IroC* triple mutant flies. I identified 11 differentially expressed OR genes and 23 transcription factors. Differentially expressed ORs were validated by qRT-PCR.

ÖZET

DROSOPHİLA KOKU SİSTEMİNDE IROC HEDEF GENLERİNİN TRANSKRİPTOM GENELİ ANALİZİ

Bir organizma için dış çevreyi algılamak çok hayatidir ve bu algılama çeşitli duyu sinirleri tarafından sağlanmaktadır. Bu sinir hücreleri, daha sonra çeşitli spektral hassasiyeti olan sinir hücrelerine dönüşen öncül duyu organı hücreleri (SOP) tarafından üretilirler. Koku sistemindeki bu sinir hücresi çeşitlenmesinin moleküler mekanizması halen tamamen bilinmemektedir. *Drosophila melanogaster* koku sistemi, sinir hücresi çeşitlenmesine çok iyi bir örnektir. Bu sistemin elemanlarının az sayıda olması ve şimdiye kadar iyi tanımlanmış olması bu sistemi çalışmaya uyumlu hale getirmiştir. *Drosophila* koku sistemi anten ve maksiller palp olarak adlandırılan iki koku organı ve 50 farklı koku sinir hücresi (ORN) sınıfından oluşur. Koku sinir hücreleri sensilaların içerisinde bulunurlar ve her sensila birden dörde kadar sayıda sinir hücresi barındırabilir.

Bu çalışma çerçevesinde biz bir TALE sınıfına ait gen düzenleyici ailesinden olan iroC (Iroquois Complex) gen düzenleyicilerine odaklandık. Enhancer trap analizleri *IroC* genlerinin iki koku organında da ifade edildiğini gösterdi. IroC hedef genlerinin hangileri olduğunu tespit etmek için genom geneli transkriptom yaklaşımı kullandım ve *IroC* üçlü mutant dokularından izole edilmiş anten ve maksiller palp dokularına RNA sekanslama deneyini yaptım. Sonuç olarak 11 tane diferansiyel bir şekilde ifade edilmiş koku reseptör geni ve 23 tane diferansiyel bir şekilde ifade edilmiş gen düzenliyici buldum. Difarensiyel bir biçimde ifade edilmiş koku reseptör genlerini qRT-PCR yöntemi ile doğruladım.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

bp	Base pair
g	Gram
kb	Kilobase
L	Liter
ml	Milliliter
mm	Millimeter
M	Molar
ng	Nanogram
nm	Nanometer
rpm	Revolutions per minute
v	Volume
w	Weight
µg	Microgram
µm	Micrometer
µl	Microliter

LIST OF ACRONYMS/ABBREVIATIONS

Acj6	Abnormal chemosensory jump 6
AD	Activating domain
AL	Antennal lobe
BD	Binding domain
cDNA	Complementary DNA
CNS	Central nervous system
DNA	Deoxyribonucleic acid
FC	Fold change
FRT	FLP recombination target
GFP	Green fluorescent protein
GO	Gene Ontology
GPCR	G-protein coupled receptor
IroC	Iroquois complex
LOF	Loss of function
nt	Nucleotide
NTC	No template control
OBP	Odor binding protein
OR	Olfactory receptor
ORN	Olfactory receptor neuron
pb	Palp basiconic
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pdm3	POU domain motif 3
PFA	Paraformaldehyde
pH	Power of hydrogen
PN	Projection neuron

pval	p value
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNA-Seq	RNA Sequencing
RT	Room temperature
RT-PCR	Reverse transcription polymerase chain reaction
SOP	Sensory organ precursor
TF	Transcription factor
UAS	Upstream activating sequence

1. INTRODUCTION

For an organism, sensing the environment is necessary for several reasons. It has much importance from not being hunted by a predator to finding a mating partner. In order to understand the outside world and act according to it we, as human beings, have developed several sensory systems like vision, hearing, taste, touch, and olfaction. Sensory systems are basically composed of two main parts; detection and interpretation. In both humans and fruit flies, detection takes place in sensory organs with the help of sensory receptors and interpretation takes place in higher brain structures. The connection between detection and interpretation is provided by sensory neurons.

1.1. Olfaction

The most ancient sense is the sense of smell or olfaction (Strausfeld and Hildebrand, 1999). Smell is different from other senses because it resides deep in the limbic system and it performs intimately with the hippocampus and amygdala (Smitka *et al.*, 2012). This is the reason why smell is so closely connected to emotion and memory.

For the detection of odors, the responsible cells are called olfactory receptor neurons (ORNs), which harbor olfactory receptors (ORs). Binding of odor molecules to the ORs excites ORNs and these neurons project their axons to olfactory processing centers (olfactory bulb in humans, antennal lobe in *Drosophila*) and from there, they synapse to secondary projection neurons, which transmit signals to higher brain centers (Hildebrand and Shepherd, 1997).

For an ORN, there are two main choices during development; which receptor to express and which place in the olfactory processing center to target. The ORNs which are expressing the same OR target their axons to the same place in the olfactory processing center (Hallem and Carlson, 2004).

The similarity of olfactory systems in different animals makes studying the system in model organisms suitable. To give an example, while an adult human has around 50.000.000 ORNs and 1000 ORs, *Drosophila melanogaster* has around 1200 olfactory neurons in one antenna and 60 ORs in total (Vosshall *et al.*, 1999).

1.2. Organization of the *Drosophila melanogaster* Adult Olfactory System

An adult *Drosophila melanogaster* has two pairs of main olfactory sensory organs; the antenna and the maxillary palp. The antenna originates from the antenna primordium and maxillary palp originates from the palpus primordium of the eye-antennal imaginal disc of *Drosophila* larva (Figure 1.1) (Kenyon *et al.*, 2003).

Each antenna harbors around 1200 ORNs, while each maxillary palp harbors around 120 ORNs. On the surface of both organs, there are sensilla, which are hair-like structures that contain olfactory neurons and supporting cells. Sensilla can be found in three different morphological types, which are basiconic, coeloconic and trichoid. While the antenna harbors all types of sensilla, maxillary palp houses only basiconic sensilla. In the antenna the number of ORNs in a sensillum can be between 1 and 4, while in the maxillary palp the number of the ORNs in each sensillum is 2 (Figure 1.2) (Hallem and Carlson, 2004).

Neurons and supporting cells develop from sensory organ precursors (SOPs). First, the transcriptional factors Atonal and Amos initiate differentiation. The fate of coeloconic sensilla and basiconic sensilla is determined by Atonal and the fate of basiconic and trichoid sensilla is determined by Amos. After initial specification, the transcription factor Lozenge helps to distinguish basiconic sensilla from trichoid sensilla. After that with the help of Notch signaling, supporting cells and different types of neurons are formed (Figure 1.3) (Endo *et al.*, 2007).

The *Drosophila* developmental stage of each step is also shown in Figure 1.3B

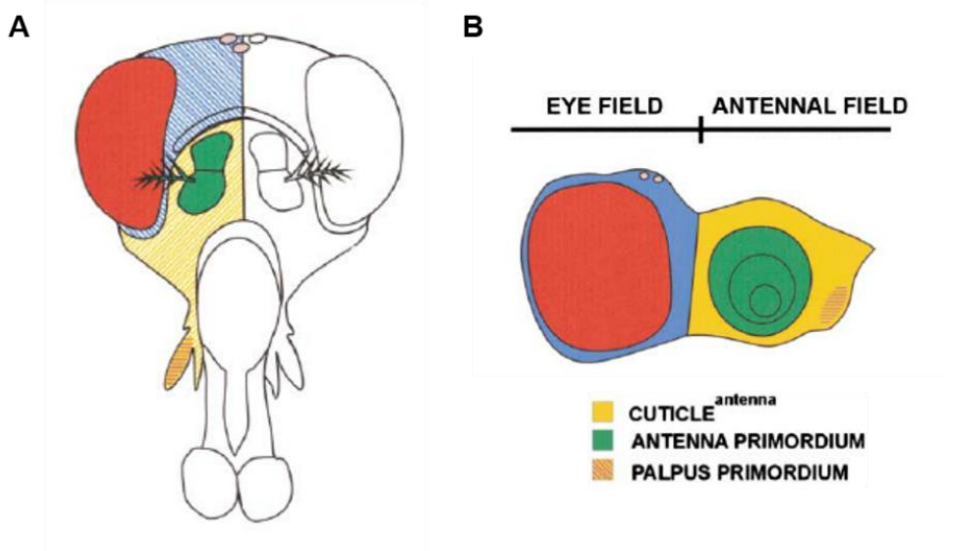

Figure 1.1. The *Drosophila* olfactory organs and their origins. (A) Adult head of *Drosophila*. Antenna and maxillary palp are shown in green and orange, respectively. (B) Eye-antennal imaginal disc of *Drosophila* larva. Antenna primordium and palpus primordium are shown in green and orange, respectively (adapted from Kenyon *et al.*, 2003).

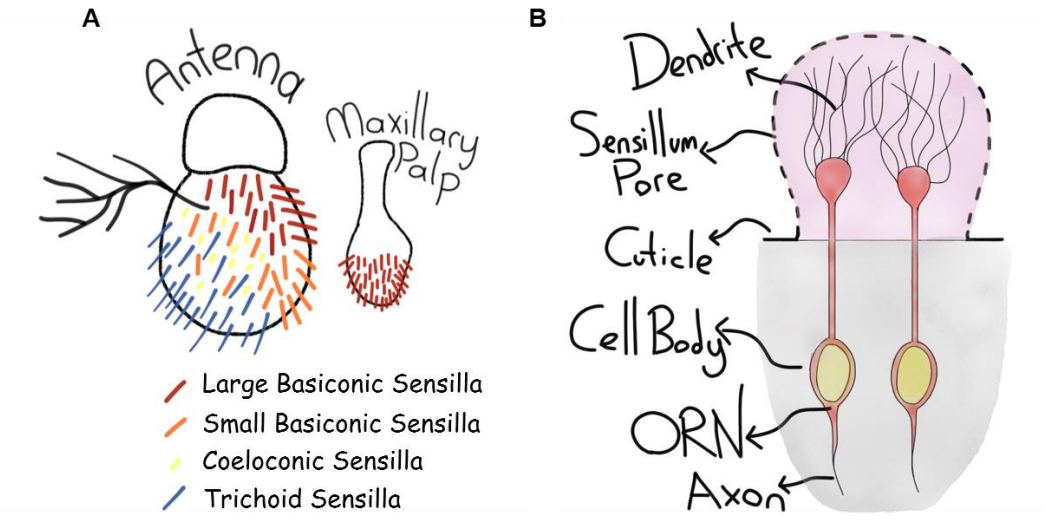

Figure 1.2. Sensilla and their morphological types. (A) Types of sensilla in each organ. There are three types of sensilla; basiconic, coeloconic, and trichoid. (B) Closer look at one sensillum harboring two ORNs.

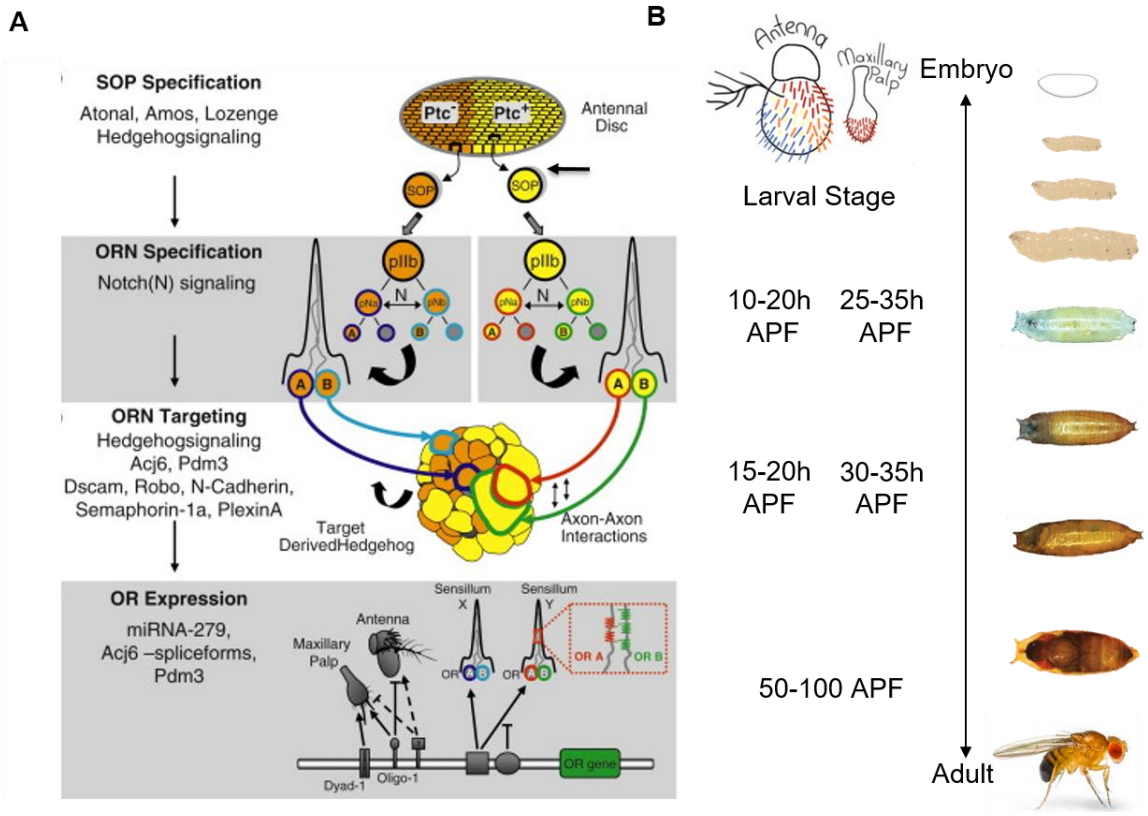

Figure 1.3. Development of ORNs. First differentiation starts in the antennal disc at the larval stage. Atonal, Amos, and then Lozenge determine the fate of sensilla. After initial fate determination, Notch signaling initiates ORN diversification (adapted from Brochtrup and Hummel, 2011).

ORN differentiation is finalized by projection of axons to their first relay station in the brain, the antennal lobe (AL). In the antennal lobe, there are 43 spherical units called glomeruli that are formed by ORN terminal and dendrites of secondary projection neurons. ORNs expressing the same type of OR project to the same glomerulus in the antennal lobe in a stereotyped fashion. Secondary projection neurons transmit the signal to higher brain centers such as the mushroom body and lateral horn. Besides these neurons, there are other types of neurons in the antennal lobe called local interneurons which are responsible for information transfer between glomeruli (Jefferis *et al.*, 2005).

1.3. Olfactory Receptor Genes

OR genes encode receptor proteins. These proteins have seven trans-membrane domains. The type of these receptors is still in controversial. While one group claims that these neurons are Odor-Gated Ion Channels (Sato *et al.*, 2008), the other claims that they represent both Cyclic Nucleotide-Gated Ion Channels and G Protein-Coupled Receptors (Wicher *et al.*, 2008) (Figure 1.4).

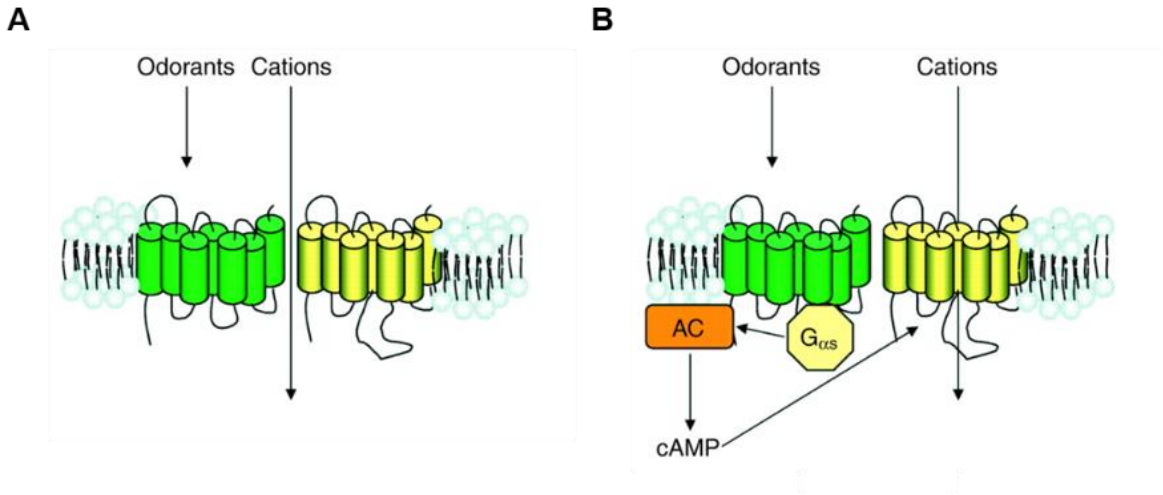

Figure 1.4. Two argued models of the membrane topology of olfactory receptor proteins. (A) Odor-Gated Ion Channels. (B) Cyclic Nucleotide-Gated Ion Channels and G Protein-Coupled Receptors (adapted from Sato *et al.*, 2008).

ORs make a complex with a co-receptor called Or83b (Orco), which is an integral part of a functional OR. The list of the ORs can be seen in Table 1.1 The first column represents the general sensilla type, which can be trichoid, basiconic, or coeloconic. The second column represents the name of each ORN. In the third and fourth column the name of the receptors and their co-receptor is given. Lastly, in the fifth column, the name of the glomerulus targeted by each ORN is given.

Table 1.1. Overview of ORs (taken from Münch and Galizia, 2016)

Sensillum	OSN	Receptor(s)	Co-receptor	Glomerulus	Sensillum	OSN	Receptor(s)	Co-receptor	Glomerulus
TRICHODEA (ANTENNA)	at1A	Or67d	Orco	DA1	BASICONICA (ANTENNA)	ab1A	Or42b	Orco	DM1
	at2A	Or83c	Orco	DC3		ab1B	Or92a	Orco	VA2
	at2B	Or23a	Orco	DA3		ab1C	Gr21a, Gr63a		V
	at3A	Or19a, Or19b	Orco	DC1		ab1D	Or10a, Gr10a	Orco	DL1
	at3	Or2a	Orco	DA4m		ab2A	Or59b	Orco	DM4
	at3	Or43a	Orco	DA4l		ab2B	Or33b, Or85a	Orco	DM5
	at4A	Or47b	Orco	VA1v		ab2B	Or33b	Orco	DM5
	at4B	Or65a, b, c	Orco	DL3		ab2B	Or85a	Orco	DM5
	at4C	Or88a	Orco	VA1d		ab3A	Or22a, Or22b	Orco	DM2
COELOCONICA (ANTENNA)	ac1	Ir31a	Ir8a	VL2p		ab3B	Or85b	Orco	VM5d
	ac1	Ir75d	Ir25a	VL1		ab4A	Or7a	Orco	DL5
	ac1	Ir92a	Ir25a, Ir76b	VM1		ab4B	Or33a, Or56a	Orco	DA2
	ac2	Ir75a	Ir8a	DP1l		ab4B	Or33a	Orco	DA2
	ac2	Ir75d	Ir25a	VL1		ab4B	Or56a	Orco	DA2
	ac2	Ir41a	Ir25a, Ir76b	VC5		ab5A	Or82a	Orco	VA6
	ac3A	Ir75a, Ir75b, Ir75c	Ir8a	DL2d/v		ab5B	Or33b, Or47a	Orco	DM3
	ac3B	Or35a	Orco, Ir76b	VC3		ab5B	Or33b	Orco	DM3
	ac3B	Or35a	Orco, Ir76b	VC3		ab5B	Or47a	Orco	DM3
	ac4	Ir84a	Ir8a	VL2a		ab6A	Or13a	Orco	DC2
	ac4	Ir75d	Ir25a	VL1		ab6B	Or49b	Orco	VA5
	ac4	Ir76a	Ir25a, Ir76b	VM4		ab7A	Or98a	Orco	VM5v
BASICONICA (PALP)	pb1A	Or42a	Orco	VM7d		ab7B	Or67c	Orco	VC4
	pb1B	Or71a	Orco	VC2		ab8A	Or43b	Orco	VM2
	pb2A	Or33c, Or85e	Orco	VC1		ab8B	Or9a	Orco	VM3
	pb2A	Or33c	Orco	VC1		ab9	Or67b	Orco	VA3
	pb2A	Or85e	Orco	VC1		ab9	Or69aA, Or69aB	Orco	D
	pb2B	Or46a	Orco	VA7l		ab10A	Or67a	Orco	DM6
	pb3A	Or59c	Orco	VM7v		ab10B	Or49a, Or85f	Orco	DL4
	pb3B	Or85d	Orco	VA4		ab10B	Or49a	Orco	DL4

1.4. Olfactory Receptor Gene Choice

In general, each ORN expresses only one OR. Unlike mammalian ORNs, *Drosophila* ORNs have been suggested to use a deterministic model (Figure 1.5) for OR gene choice by an ORN cell in order to express a particular receptor, while vertebrates display stochastic expression of single receptor (Fuss and Ray, 2009). In the deterministic model, unlike the stochastic model, receptor expression is accomplished by a combinatorial code of transcription factors. Supporting this model, it has been shown that a region 500 bp upstream of the transcription start site of a receptor gene is sufficient for expression of that gene (Ray *et al.*,

2007). Thus, these observations suggest that a combination of regulatory elements (cis-elements) and transcription factors work together to determine the expression of receptor genes.

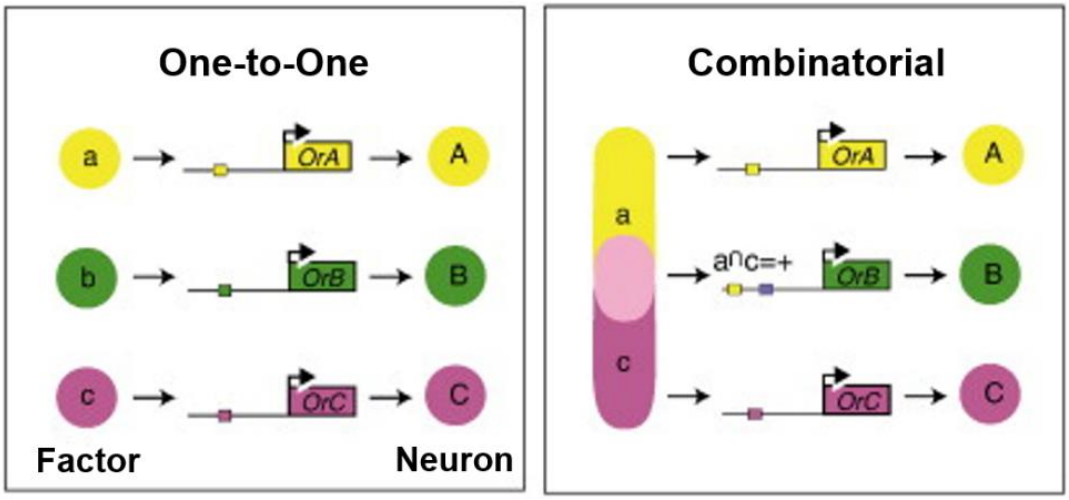

Figure 1.5. Combinatorial manner of OR gene choice (adapted from Fuss and Ray, 2009)

Several transcription factors and motifs have been identified as an element in this combinatorial manner. For example, Dyad-1, and Oligo-1 are identified motifs that drive expression of OR genes in the maxillary palp and repress their expression in the antenna (Ray *et al.*, 2007).

Pdm3 (POU domain motif 3) and Acj6 (Abnormal chemosensory jump 6) transcription factors have a role in OR gene choice (Ray *et al.*, 2008; Ray *et al.*, 2007; Tichy *et al.*, 2008). They are required for expression of certain OR genes. Known regulatory elements and transcription factors are summarized in Figure 1.6. Pdm3 transcription factor regulates the *Or42a* expression in the maxillary palp and has an important function in axonal targeting. The transcription factor Acj6 also known with its receptor gene expression and axonal targeting regulation. Transcription factors atonal, amos, and lozange play a role in initial diversification of sensilla types. Dyad-1 and Oligo-1 motifs play a role in the receptor gene expression in the maxillary palp. The transcription factor Hamlet regulates the neuron diversity by regulating

notch signaling. The transcription factor Rotund regulates sensilla subtype specification. The transcription factor Engrailed controls neuron/glia fate decision.

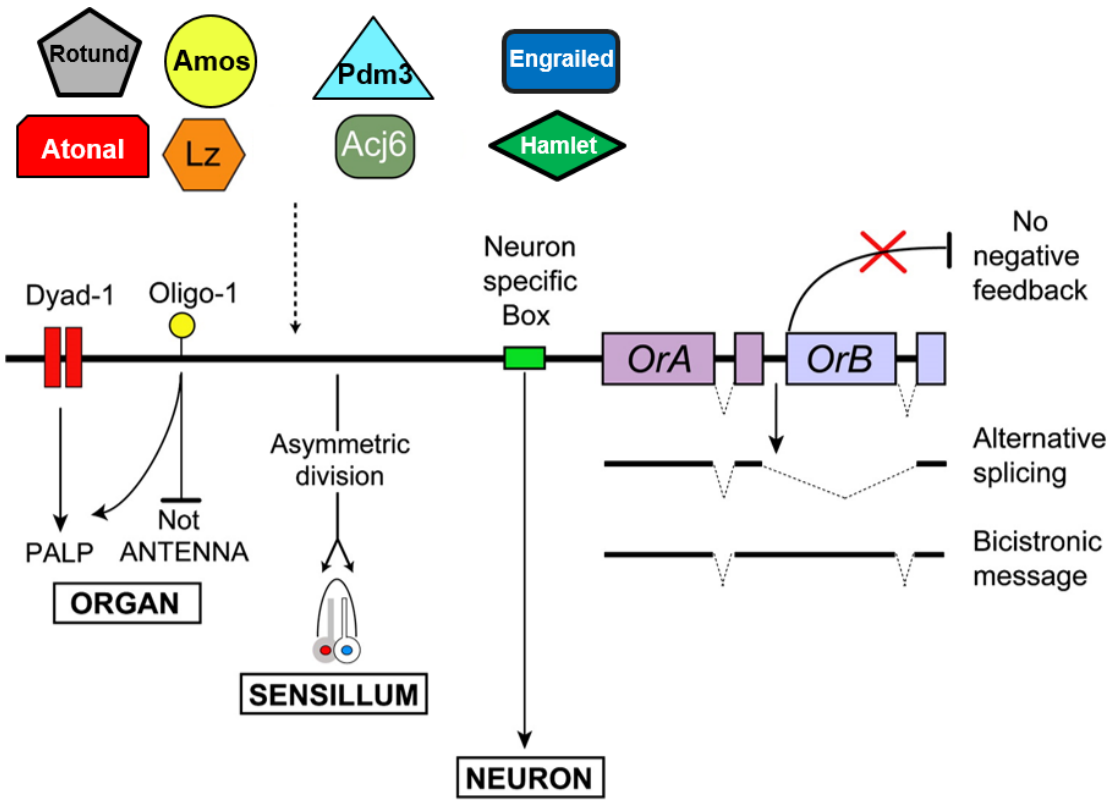

Figure 1.6. Summary of known receptor gene expression mechanisms (adapted from Ray *et al.*, 2007)

1.5. Iroquois Complex

Araucan (*ara*), *caupolican* (*caup*), and *mirror* (*mirr*) are members of the Iro family and all together, they constitute the *Iroquois Complex* (*IroC*). These genes encode TALE class homeodomain transcription factors and share a homeodomain and a Iro box. These three genes are conserved as a cluster from flies to mammals (Figure 1.7). They have very vital roles in development. *ara* and *caup* are closely related to each other than *mirr* in both genomic position and expression pattern. *ara* and *caup* also share the same common cis-regulatory regions. (J.-L. Gómez-Skarmeta *et al*, 1996) (McNeill *et al.*, 1997).

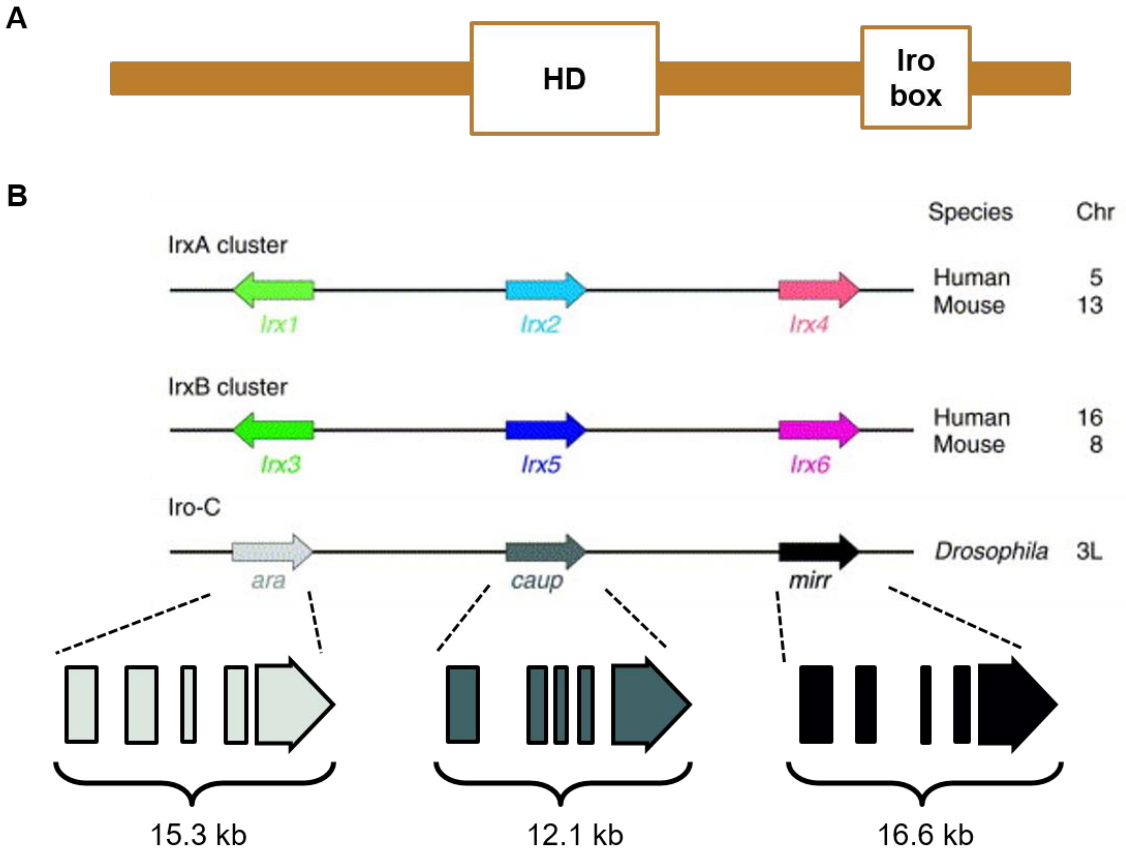

Figure 1.7. Genomic and organization of *IroC* genes and structure of proteins. (A) All Iro proteins have a homeodomain of the TALE class (HD) and the Iro box. (B) All IroC transcription factors are clustered together and it is conserved from vertebrates to *Drosophila* (adapted from J. L. Gómez-Skarmeta and Modolell, 2002)

Iro proteins have diverse role in development. These are heart ventricular identity specification (Bao *et al.*, 1999), neural plate differentiation (Bellefroid *et al.*, 1998), organizer formation during gastrulation (Kudoh and Dawid, 2001), planar cell polarity organization in the eye (McNeill *et al.*, 1997), sensory organ compartmentalization (J.-L. Gómez-Skarmeta *et al.*, 1996), and receptor expression (Mazzoni *et al.*, 2008).

Iro transcription factors can form heterodimers and homodimers. It is known that mirr binds to DNA as a homodimer (Bilioni *et al.*, 2005). There are also putative binding sites for Iro transcription factors. Ara binds to an ASC enhancer TTAATTAA (J.-L. Gómez-Skarmeta *et al.*,

1996), AAAACACGTGTTAA is mirr binding site and ACANNTGT is a minimal recognition sequence for mirr (Bilioni *et al.*, 2005), binding sites for Ara/Caup is ACAN(2–8)TGT (Carrasco-Rando *et al.*, 2011), and ACACGTGT is a mirr binding site (Andreu *et al.*, 2012).

1.6. Genetic Tools for *Drosophila*

Drosophila melanogaster is well-established model organism, which provides many sophisticated tools for gene manipulation. Here, I will describe two basic genetic tools that have been used in this study, the Gal4/UAS and FLP/FRT System.

1.6.1. Gal4/UAS System

The Gal4/UAS system is used for the targeted expression of a gene of interest in a cell or tissue of interest. It is composed of two main elements; Gal4 protein and Upstream Activating Sequence (UAS). Gal4 protein is encoded by *gal4* gene identified in the yeast *Saccharomyces cerevisiae*. It has two domains known as binding domain (BD) and activating domain (AD). Gal4 protein binds to UAS through its binding domain and activates the downstream gene through its activating domain. Specific promoter or enhancer regions cloned to the upstream region of the *gal4* gene can be used as a driver to induce a gene under the control of UAS sequence (Figure 1.8) (Duffy, 2002).

1.6.2. FLP/FRT System

The FLP/FRT System is used in this study in order to create homozygous mutant tissues. The system is composed of two main elements; FLP recombinase and FLP recombination target (FRT) sequence. FLP recombinase is encoded by the *flp* gene derived from yeast and recognizes FRT sequences and triggers site-specific recombination between FRT sites on two different alleles. Like in the Gal4/UAS System, a specific promoter or enhancer region cloned to the upstream region of the *flp* gene is used as a driver to induce recombination between FRT sites in specific cells or tissues (Theodosiou and Xu, 1998) (Figure 1.9).

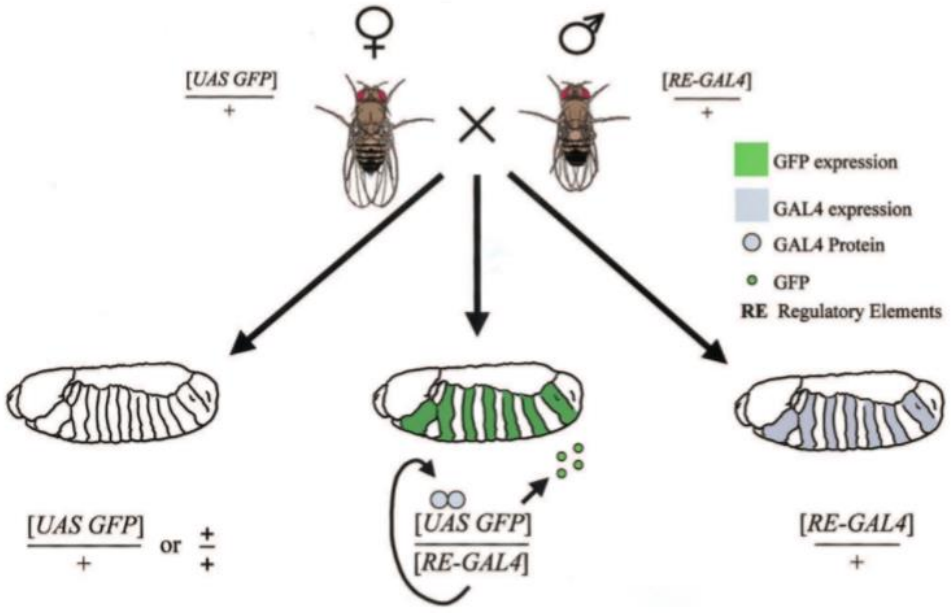

Figure 1.8. The Gal4/UAS System (adapted from Duffy, 2002).

1.7. RNA Sequencing

With the invention of next-generation technologies, sequencing techniques are now improved in terms of taking short times, reduced amount of cost and sequencing massive scale. RNA Sequencing (RNA-Seq) is one of the next-generation technologies. The Illumina platform was used in this study. To use this platform RNA is extracted from cells, cDNA libraries are prepared, fragmented and adaptors are ligated to the end of cDNA fragments. During sequencing, fragments are attached randomly to the surface of Illumina flow cell with the help of adapter sequences. Then, single stranded bridge formation of the fragments occurs. After that bridge amplification and denaturation happens. Finally, fluorescent-labeled nucleotides are added, and the signals from each base are detected with laser excitation during amplification. As a result, sequence data is obtained (Wilhelm and Landry, 2009).

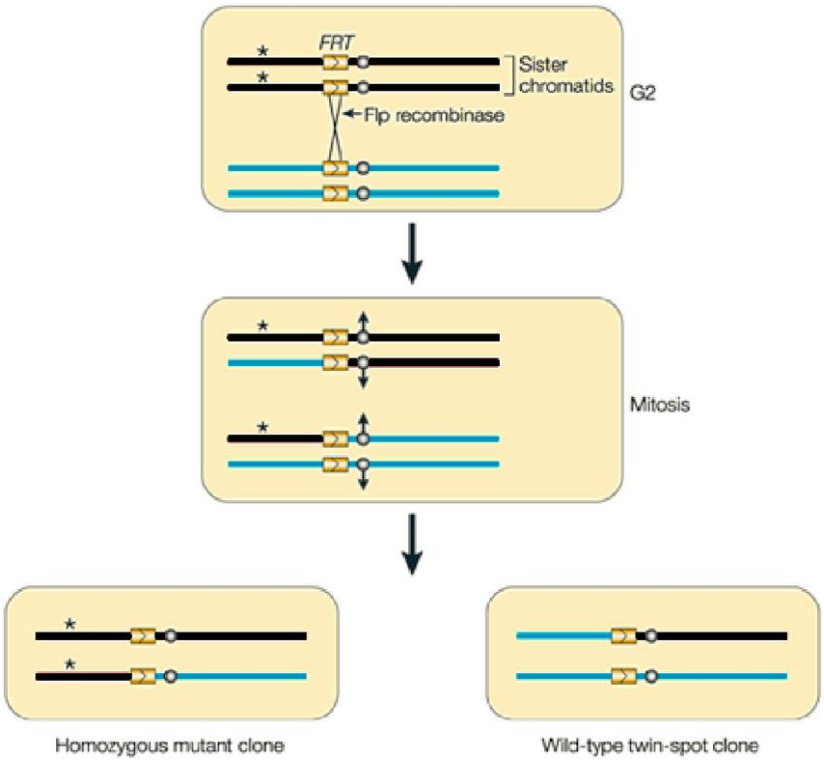

Figure 1.9. FLP/FRT System (adapted from Beckingham *et al.*, 2007)

2. AIM OF THE STUDY

In order to understand and solve the olfactory receptor gene choice mechanism and illuminate the required transcription factors, the main purpose of this study is to identify the role of *IroC* transcription factor family in transcriptome-wide analysis and identify the downstream genes regulated by IroC transcription factors.

IroC genes are expressed in the ORNs in both antenna and maxillary palp and we hypothesize that *IroC* may have a role in olfactory receptor gene choice mechanism. In order to test this hypothesis, we aimed to combine loss-of-function experiments with transcriptome-wide analysis and to get all of the olfactory receptor genes and transcription factors regulated by IroC proteins. Loss-of-function experiments were done in antenna and maxillary palp by using FLP/FRT system and by using *IroC* triple mutant flies. Transcriptome-wide analysis was done by RNA Sequencing experiments.

Our second aim was to find IroC transcription factors binding sites *de novo* from differentially expressed OR genes resulted from our transcriptome-wide analysis. *De novo* motif discovery was done by using RSAT bioinformatics tool.

3. MATERIALS AND METHODS

3.1. Biological Material

Drosophila melanogaster lines were stored in air-permeable transparent tubes including fly food at 25^{o}C and 70% humidity with a 12h:12h light and dark cycle. Commercially available fly food (Genesee Scientific Nutri-Fly^{TM} Bloomington Formulas) was prepared and used according to company's instruction for the *Drosophila*. Flies were transferred to new tubes with freshly prepared fly food once a month. Fly lines used in this study are listed in Table 3.1.

Table 3.1. *Drosophila melanogaster* lines used in this study.

Transgene	Inserted Chr. No.	Description
Gal4 Drivers		
ey-Gal4	1,2	Expresses Gal4 under the control of *eyeless* gene enhancer.
IroC-Gal4	3	Expresses Gal4 under the control of *IroC*
mirr(DE)-Gal4	3	Gal4 expressed in the dorsal part of eye
UAS Constructs		
UAS-mCD8::GFP	2,3	Encodes GFP fused cell membrane bound protein under the control of UAS.
UAS-GFPnls	2,3	Encodes GFP with nuclear localization signal under the control of UAS.
UAS-mCD8::chRFP		Encodes RFP fused cell membrane bound protein under the control of UAS.
UAS-FLP	1,2,3	Encodes flippase under the control of UAS.
Deficiency Lines		
iro^{DFM3}	3	Chromosomal deficiency spanning *araucan*, *caupolican* and the promoter of *mirror*

Table 3.1. *Drosophila melanogaster* lines used in this study (cont.)

Transgene	Inserted Chr. No.	Description
General Stocks		
w^{1118}	1	Mutant allele of eye pigment gene causes white eye phenotype.
CL	3	Putative cell lethal mutation
GMR-hid	3	Recombined construct of pro-apoptotic gene hid under the control of GMR promoter
FRT80B	3	Allow FLP-mediated site-specific recombination on the chromosome arm 3L
His2Av-mRFP	3	Expresses mRFP-tagged Histone 2Av constitutively
yw ; QB		Flies carrying balancer chromosomes Sp / CyO ; TM2 / TM6B

3.2. Chemicals and Supplies

All chemical material used in this study are listed under the corresponding titles.

3.2.1. Chemical Supplies

Chemicals used in this study are listed in Table 3.2.

3.2.2. Buffers and Solutions

Solutions and mixtures used in this study are listed in Table 3.3.

3.2.3. Oligonucleotide Primers

Oligonucleotides were synthesized commercially in Macrogen company (South Korea), and produced lyophilized oligonucleotides were dissolved in dH_2O as final concentration of 100 pmol/µL. Dissolved oligonucleotides were stored at -20^oC as stock. For PCR, 1:10 diluted (10 pmol/µL) stock oligonucleotides were used. Oligonucleotides used in this study are listed in Table 3.4.

Table 3.2. Chemical list used in this study

Chemical	**Producing Company**
1 kb Marker	NEB, USA (N3232L)
100 bp Marker	NEB, USA
Ethidium Bromide Solution	Sigma Life Sciences, USA (E1510)
$MgCl_2$	Riedel-de Haen, Germany (13152)
Paraformaldehyde	Sigma-Aldrich, USA (P6148)
Phenol:Chloroform:Isoamyl alcohol	Sigma-Aldrich, USA (P2069)
Triton X-100	AppliChem, USA (A4975)
Trizol	Invitrogen (15596-026)
Tween-20	Roche, USA (11332465001)

Table 3.3. Solution and Mixture list used in this study

Solution / Mixture	**Ingredient**
Formaldehyde Solution (4%)	160 g/l PFA, pH 7.4
PAXD	50 g BSA 3 g Sodium Dexoycholate 0.3% Triton X-100 In PBS
PBS (1x)	137 mM NaCl 2.7 mM KCl 10 mM Na_2HPO_4 1.8 mM KH_2PO_4
PBST	PBS (1x) 0.1% Tween-20
PBSTX	PBS (1x) 0.05% Triton X-100
PBX3	PBS (1x) 0.3% Triton X-100

Table 3.4. Oligonucleotide list used in this study

Primer Name	Primer Sequence (5'-3')	T_m(°C)
Or33c-F	TGGTGCGATTCCACAACCTG	60.5
Or33c-R	AGACCAGCGAGATTGTGTCG	60.5
Or83b-F	TGTTTGCTTCGGTTCTGGGT	58.4
Or83b-R	GGTAGGTGGCTTCTTGGCAT	60.5
Or47a-qPCR-F	AGAAATGTGGAAACGCCCCT	58.4
Or47a-qPCR-R	CTGTTCCGCCTCATTCGACA	60.5
Or43b-qPCR-F	CCACCGATTCCTACCCTGTG	62.5
Or43b-qPCR-R	TGGCATCACAAAAATTTAGCATGG	60.1
Or19a-qPCR-F	ACGCTGCCAGATTCCAATGA	58.4
Or19a-qPCR-R	CGCTCCCTTAATCATCACCGT	61.3
Or47b-qPCR-F	ATCCCACGCCTGGCTTTTTA	58.4
Or47b-qPCR-R	GTGCCATCCCCGAAATCCAA	60.5
Or9a-qPCR-F	AAAGAGCCCTCCTCATTGCC	60.5
Or9a-qPCR-R	GCGAACAATCGTCGAGAAGG	60.5
Or85a-qPCR-F	TGCCATTGAATCGGTCACCA	58.4
Or85a-qPCR-R	CGCAGAGTGTTTCCATATTCGAC	62.9
Or71a-qPCR-F	ACAATTCGGATTGGCCGGAT	58.4
Or71a-qPCR-R	TCATGGTCTTGGTAAACAGAGGT	61.1
Gapdh1-qPCR-F	GACTCGACTCACGGTCGTTT	60.5
Gapdh1-qPCR-R	CACCACATACTCGGCTCCAG	62.5

3.2.4. Antibodies

Antibodies used in this study are listed in Table 3.5.

3.2.5. Embedding Media

In immunohistochemistry experiments, stained tissues were embedded in Vectashield embedding medium (Vector Laboratories, Inc.), Fluoromount-G embedding medium (Southern Biotech) or 80% glycerol. Embedded tissues were kept at 4^{o}C until visualization.

Table 3.5. Antibody list used in this study

Name	Antigen	Species	Dilution	Source
Primary Antibodies				
Elav	Elav	Rat	1:20	Hybridoma
Elav	Elav	Mouse	1:20	Hybridoma
GFP	GFP	Rabbit	1:500	Invitrogen
GFP	GFP	Mouse	1:1000	Promega
DsRed	RFP	Rabbit	1:1000	Clontech
Secondary Antibodies				
Alexa 488	Rabbit	Goat	1:800	Invitrogen
Alexa 488	Rat	Goat	1:800	Invitrogen
Alexa 488	Mouse	Goat	1:800	Invitrogen
Alexa 555	Mouse	Goat	1:800	Invitrogen
Alexa 555	Rabbit	Goat	1:800	Invitrogen
Alexa 637	Rat	Goat	1:800	Invitrogen
Alexa 647	Rabbit	Goat	1:800	Invitrogen
Alexa 647	Mouse	Goat	1:800	Invitrogen

3.2.6. Disposable Labware

Disposable labware used in this study are listed in Table 3.6.

Table 3.6. Disposable labware used in this study

Material	**Manufacturer**
Micropipette Tips	Greiner Bio-One, Belgium
Microscope cover glass	Fisher Scientific, UK
Microscope slides	Fisher Scientific, UK
PCR tubes (200 µl)	Bio-Rad, USA

Table 3.6. Disposable labware used in this study (cont.)

Material	**Manufacturer**
Pipette Tips (10 - 200 - 1000 µl)	VWR, USA
Plastic Pasteur pipettes	TPP Techno Plastic Products AG, Switzerland
Syringe (1cc)	Becton, Dickinson and Company, USA
Test Tubes, (0.5 - 1 - 1,5 - 2 ml)	Citotest Labware Manufacturing, China
Test Tubes, (15 - 50) ml	Becton, Dickinson and Company, USA

3.2.7. Equipment

Equipment used in this study is listed in Table 3.7.

Table 3.7. Equipment used in this study

Equipment	**Manufacturer**
Autoclave	Astell Scientific Ltd., UK
Centrifuges	Eppendorf, Germany (Centrifuge 5424, 5417R)
Cold Room	Birikim Elektrik Soğutma
Confocal Microscope	Leica Microsystems, USA (TCS SP5)
Electrophoresis Equipment	Bio-Rad Labs, USA
Freezers	Arçelik, Turkey
Gel Documentation System	Bio-Rad Labs, USA (Gel Doc XR)
Heating Block	Fisher Scientific, France

Tablo 3.7 Equipment used in this study (cont.)

Equipment	Manufacturer
Heating Magnetic Stirrer	IKA, China (RCT Basic)
Incubator	Weiss Gallenkamp, USA (Incubator Plus Series)
Laboratory Bottles	Isolab, Germany
Micropipettes	Eppendorf, Germany
Microwave oven	Vestel, Turkey
pH meter	WTW, Germany (Ph330i)
Refrigerators	Arçelik, Turkey
Stereo Microscope	Olympus, USA (SZ61)
Thermal Cycler	Bio-Rad Labs, USA (C1000 Thermal Cycler)
Vortes Mixer	Scientific Industries, USA (Vortex Genie2)

3.3. Histological Techniques

3.3.1. Immunohistochemistry

Tissues were dissected in PBS on dissection pad and collected in a test tube containing PBS on ice. Dissected tissues were fixed in 4% PFA for 30 minutes at RT. The fixed tissues were washed 3 times for 15 minutes in PBX3 at RT then blocked in PAXD for 30 minutes at RT. The blocked tissues were then incubated with primary antibodies diluted in PAXD overnight at 4°C. On the second day, the tissues were washed 3 times with PBX3 for 15 minutes at RT. Then, washed tissues were incubated with secondary antibody for 2 hours at RT. After that, the tissues were washed 3 times in PBX3 for 15 minutes. Then, the tissues were mounted on a microscope

slide with embedding medium. Finally, mounted tissues were visualized under confocal microscope. All the steps were performed on a shaker unless otherwise stated.

3.4. Molecular Biological Techniques

3.4.1. Isolation of RNA

All the antenna and maxillary palp were dissected on CO_2 fly pads and directly collected in Trizol reagent. For 300 antennae and 300 maxillary palp, 400 µl Trizol was used. Collected tissues were homogenized with hand held homogenizer then incubated 5 minutes at RT. 80 µl chloroform was added to the tube and shaken vigorously by hand for 15 seconds. After that, tissues were incubated for 3 minutes at RT and centrifuged for 15 minutes at 12000 g at 4°C. The aqueous phase was transferred into new tube and 10 µg of RNase free glycogen was added to the tube in order to visualize small amount of RNA pellet. 200 µl isopropanol was added to the tube and incubated for 10 minutes at RT. The mixture was centrifuged for 10 minutes at 12000 g at 4°C and supernatant was discarded. Remaining pellet was washed by adding 400 µl 75% EtOH and centrifuging for 5 minutes at 7500 g at 4°C. The EtOH was discarded and the pellet was left to air-dry for 5 minutes. Finally, the pellet was dissolved in 50 µl of DEPC-treated dH_2O and stored in -70°C freezer.

3.4.2. cDNA Synthesis

To synthesize cDNA from obtained RNA, a cDNA synthesis kit (Invitrogen) was using according to manufacturer's suggestions. Briefly, 8 µg of isolated RNA was incubated with 1 µl 10mM dNTP mix and 1 µl 500 µg/ml oligo(dT) primer in a total volume of 10 µl for 5 minutes at 65°C then, chilled on ice at least 1 minute. The reaction mix was prepared in incubation time by mixing 2 µl 10x reverse transcriptase buffer, 4 µl 25mM $MgCl_2$, 2 µl 0.1M DTT, and 1 µl 40units/µl RNaseOUT. The reaction mixture was mixed gently with the sample mixture and centrifuged briefly. The final mixture was incubated for 2 minutes at 42°C and 1 µl of reverse transcriptase was added to the mixture. The reaction was incubated for 50 minutes at 42°C, for 15 minutes at 70°C, and chilled on ice. The reaction was briefly centrifuged and 1 µl

of RNaseH was added to the reaction. Finally, the reaction was incubated for 20 minutes at 37°C and stored at -20°C.

3.4.3. Reverse Transcription Polymerase Chain Reaction (RT-PCR)

The cDNA of differentially expressed OR genes found by RNA-Sequencing analysis were amplified by polymerase chain reaction. Reactions were performed in a total volume of 25 µl containing 1 µl cDNA, 0.5 µ dNTP (10mM), 0.5 µl forward primer (10 pmol/µl), 0.5 µl reverse primer (10 pmol/µl), 2.5 µl Taq polymerase buffer (10x), 2 µl $MgCl_2$ (25mM), 0.125 µl Taq polymerase, and 17.875 µl dH_2O. The standard PCR program comprised a cycle of 2' at 95°C, followed by 30'' at 95°C, 30'' at 56°C, and 30'' at 68°C for 30 cycles and finally a cycle of 10' at 68°C. PCR products were run and compared on the same 2% agarose gel.

3.4.4. Quantitative Reverse Transcription Polymerase Chain Reaction (qRT-PCR)

The cDNA of differentially expressed OR genes found by RNA-Sequencing analysis were amplified by quantitative reverse transcription polymerase chain reaction. Reactions were performed in a total volume of 10 µl containing 2 µl cDNA, 0.25 µl forward primer (10 pmol/µl), 0.25 µl reverse primer (10 pmol/µl), 5 µl Master Mix containing SYBR Green, and 2.5 µl dH_2O. The qPCR program comprised a cycle of 10' at 95°C, followed by 15'' at 95°C, 45'' at 57°C, and SCAN for 30 cycles. For melting curve, temperature was started from 60°C and increased 1°C in every second until 95°C. PCR products were run on the 2% agarose gel. 3 biological replicate was done. For each biological replicate, 3 technical replicates and no template control (NTC) were set for each gene. ΔΔ Ct values were used for fold change calculation.

3.4.5. Agarose Gel Electrophoresis

2% agarose gel (w/v) was prepared with 1X TAE buffer and 30 ng/ml ethidium bromide solution. By finalizing the concentration of loading dye to 1x, samples were prepared and loaded

on agarose gel with 1kb or 100bp markers. The gel was run at 90 V for 30-60 min, and visualized under UV (Gel-DOC, Bio-Rad, USA).

3.5. Bioinformatics Analysis Techniques

For RNA-Sequencing experiment, RNA was isolated from iro^{DFM3} clonal and control flies. 300 whole antennae and 300 maxillary palps of 1-10 days old 150 flies (75 males and 75 females) were dissected for each biological replicate. 3 biological replicates were done for each genotype. Isolated RNAs were sent to Genomics Core, UZ Leuven, Belgium on dry ice. cDNA synthesis, library preparation, and RIN value measurement were done by Genomics Core. Illumina HiSeq 4000 System was used for 50 bp single-end sequencing. Resulting raw reads were sent to us by Genomics Core and RNA-Sequencing analysis was done starting from raw reads.

3.5.1. Quality Control and Removing Adaptors

Starting from raw reads (fastq file), firstly, quality control analysis was done by using FastQC sequencing quality control tool which is effective on Illumina reads. According to results of the software, adaptors were removed and in order to improve mappability, reads were trimmed from 3' end due to read quality decrease.

3.5.2. Aligning Raw Reads to Genome

After quality control, resulting reads were mapped to the genome by using STAR tool which is a gapped mapper tool and effective on *Drosophila* genome. In order to align of raw reads to genome, at first, genome indexes files were generated with STAR tool by using *Drosophila melanogaster* genome file in fasta file format and annotation file in gtf file format. Latest release of *Drosophila melanogaster* genome file (dm6) and annotation file (BDGP6.90) were downloaded from Ensembl. After that, quality controlled raw reads were mapped to the genome by using STAR tool and generated genome indexes file. As a result of this step, required file for differential gene expression analysis (.count file) was obtained.

3.5.3. Differential Gene Expression Analysis

After aligning reads to genome, gene count file was obtained. DeSeq2 tool was used with count file as an input and differential expression analysis was done. List of all differentially expressed genes were obtained. From that file, OR genes and transcription factors were separated as a new list.

3.5.4. *De Novo* Motif Discovery

RSAT Dyad Analysis web browser (Medina-Rivera *et al*., 2015) was used for *de novo* motif discovery. 1000 bp upstream region of significantly differentially expressed OR genes and transcription factors were used. As a result, enriched motifs on these sequences were obtained.

3.5.5. GO Term Analysis

GOrilla (Eden *et al*, 2009) and GO:TermFinder (Boyle *et al*., 2004) web browsers were used for GO Term Analysis. All of the significantly differentially expressed genes resulted from RNA Sequencing experiment was used as an input. As a result, enriched GO Terms on this set of genes were obtained.

3.6. Experiments for Functional Analysis

Functional Analysis of *IroC* genes was performed as loss-of-function experiments.

3.6.1. Loss-of-Function Experiments: Clonal Tissue Generation

In order to discover the OR genes and transcription factors regulated by *IroC* genes, $IroC^{DFM3}$ deficient fly lines were used. The deletion of these genes results is homozygous lethality. In order to generate homozygous mutant antenna and maxillary palp tissues, FLP/FRT system is used to overcome this problem. This method provides making homozygous mutant

cells in specific tissues by using FLP recombinase enzyme and this enzyme causes recombination on flippase recognition target sites on the chromosome while the other cells in other tissues remain heterozygous for the deficiency. Eyeless promoter which is expressed in eye, antenna, and maxillary palp was used to drive FLP recombinase.

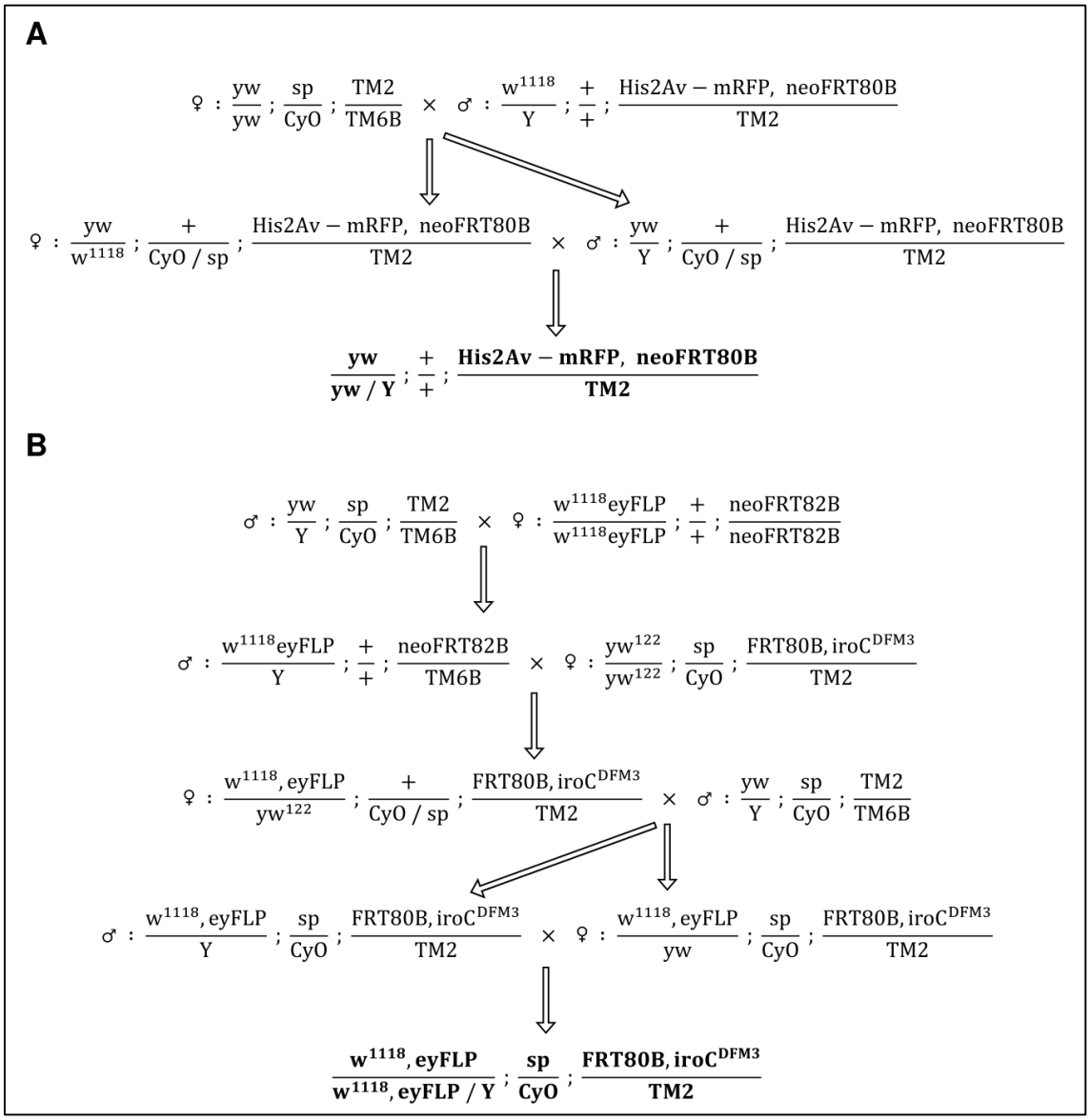

Figure 3.1. Generation of FRT His2Av and FRT iroC^{DFM3} lines.

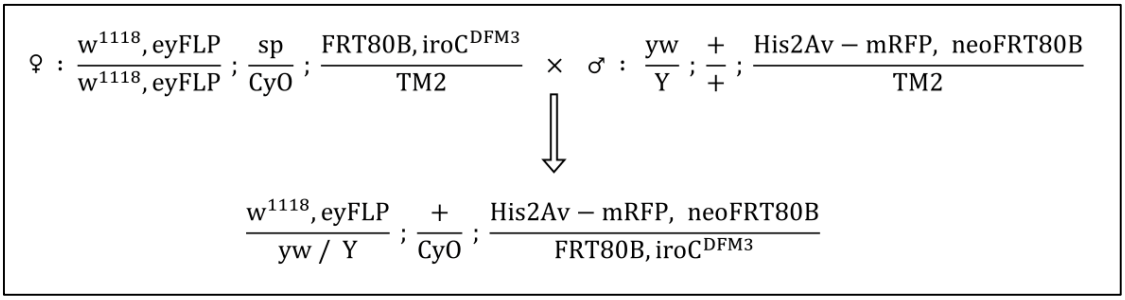

Figure 3.2. FLP / FRT crosses for creating clonal tissues with His2Av-mRFP flies.

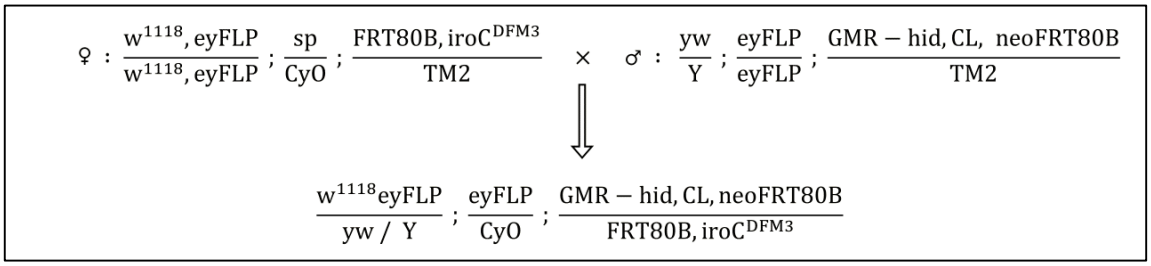

Figure 3.3. FLP / FRT crosses for creating clonal tissues with GMR-hid, CL flies

4. RESULTS

4.1. IroC is Expressed in Olfactory Receptor Neurons in the Antenna and the Maxillary Palp

In order to analyze *IroC* expression profile in the antenna and maxillary palp, Mirr(DE)-Gal4 and IroC-Gal4 lines were used. Mirr-Gal4 line was prepared by LacZ-expressing $mirr^{DE}$ with P[Gal4] element by P-element exchange procedure (Morrison and Halder, 2010). The LacZ-expressing iro^{rF209} line has been shown to reflect the expression pattern of *ara* and *caup* (J.-L. Gómez-Skarmeta *et al*., 1996). Here, I used an iroC-Gal4 line that was prepared by replacing the P[lacZ] element with a P[Gal4] element (Wernet *et al.*, 2006). The physical map of the *IroC* locus is shown in Figure 4.1. *ara* and *caup* are close to each other than *mirr*. They are conserved as cluster as in vertebrates. In iro^{DFM3}, *ara* and *caup* are deleted and regulatory region of *mirr* is deleted. P element exchange location of iro^{rF209} and $mirr^{DE}$ is shown with triangles.

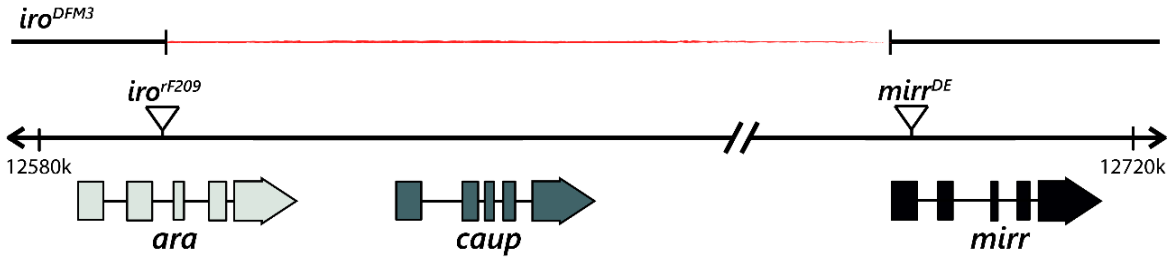

Figure 4.1. Physical Map of the *IroC* Locus. Triangles represent the insertion sites of P-elements and the red line represents the extent of deletion in the iro^{DFM3} mutant. *ara* and *caup* are close to each other and *mirr* is more divergent in terms of genomic location. In iro^{DFM3} allele, *ara* and *caup* is totally deleted and also regulatory region of *mirr* is deleted.

Expression analysis of Mirr-Gal4 line crossed with UAS-nGFP showed that *mirr* is expressed in a subset of neurons in both adult antenna and adult maxillary palp assessed by staining with the neuron-specific antibody elav. GFP is also expressed in the dendrites, which harbor ORs. (Figure 4.2).

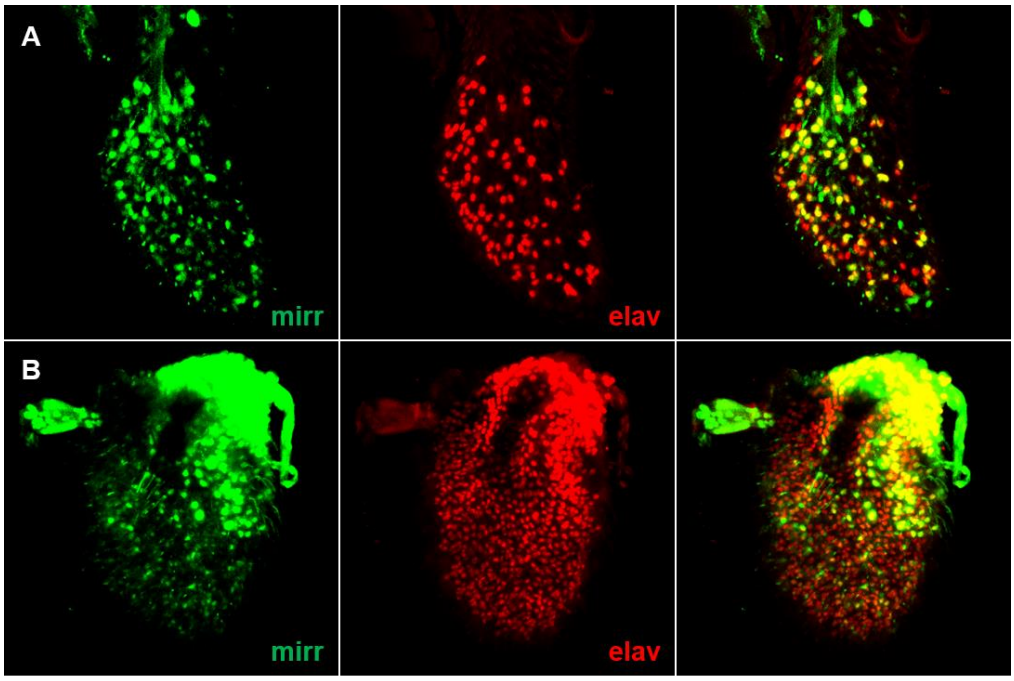

Figure 4.2. *mirr* is expressed in neurons. Antibody staining of maxillary palps (A) and antennae (B) of Mirr-Gal4 UAS-nGFP expressing flies using antibodies against GFP (green) and the neuronal marker elav (red). Co-localization can be observed in yellow.

Expression analysis of IroC-Gal4 line crossed with UAS-mCD8::GFP showed that *ara* and *caup* are expressed in a subset of neurons in both adult antenna and adult maxillary palp assessed by staining with the neuron-specific antibody elav. GFP is also expressed in the dendrites, which harbor ORs. (Figure 4.3).

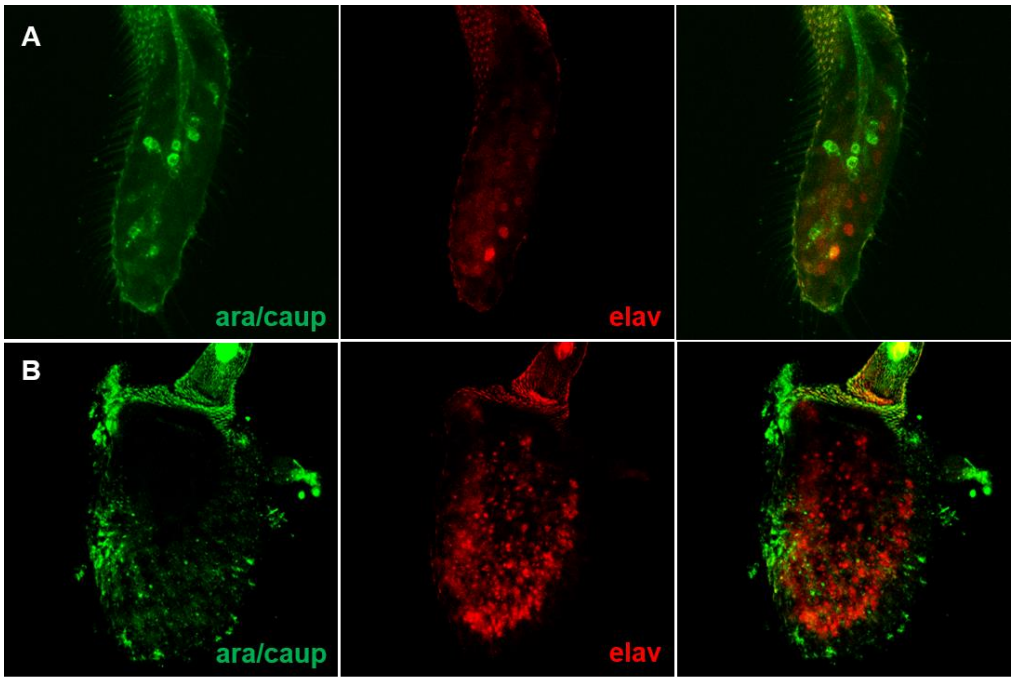

Figure 4.3. *ara* and *caup* are expressed in neurons. Antibody staining of maxillary palps (A) and antennae (B) of IroC-Gal4 UAS UAS-mCD8::GFP expressing flies using antibodies against GFP (green) and the neuronal marker elav (red).

4.2. Generation of FRT *iroC*^{DFM3} Lines for Use in RNA Sequencing Experiments

In order to get insight into which genes might be regulated by iroC, I wanted to make use of the *iroC* triple mutant *iroC*^{DFM3}. However, this line is homozygous lethal at embryonic stage and thus olfactory tissues cannot be analyzed in the adult. Thus, I used the FLP/FRT system to generate mutant olfactory tissue in an otherwise heterozygous animal. The *eyeless*-Gal4 driver, which is expressed in eye, antenna and maxillary palp was used to drive the expression of the flippase enzyme. In order to observe the homozygous tissue clones the FRT80B His2Av-mRFP allele was used with FRT80B *iroC*^{DFM3} allele at the same chromosome. To visualize the generated homozygous clones, eyes of the mutant flies were stained with a DsRed antibody (Figure 4.4). The *iroC*^{DFM3} allele on its own was used as a negative control and His2Av-mRFP alone was used as a positive control. In clonal regions, the His2Av-mRFP allele should not be present, so those regions would not give a DsRed signal. As expected those regions

did not give DsRed signal. However, the sizes of the generated clones were small and were considered to be too small to reflect significant changes in gene expression when analyzed by RNA Sequencing.

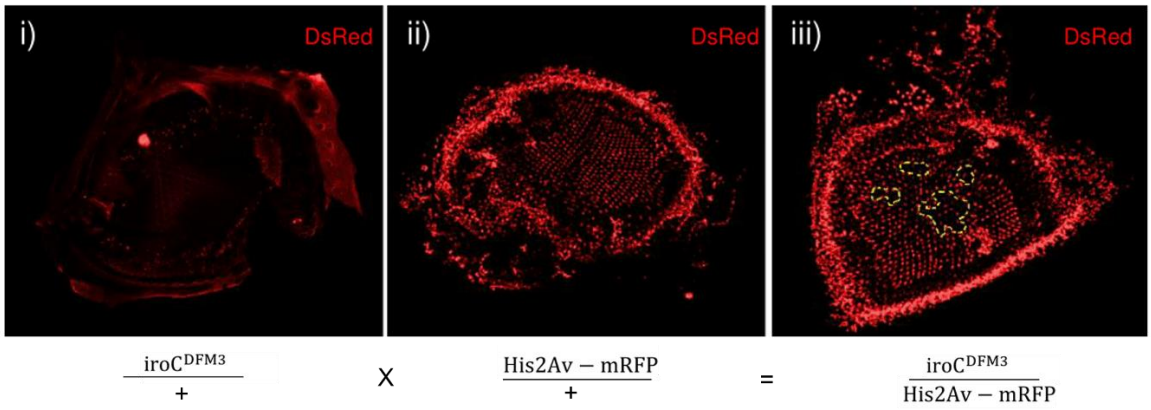

Figure 4.4. Staining of eyes of adult mutant flies with DsRed antibody to visualize clones. (i) Negative control. (ii) Positive control. (iii) Yellow dashed regions show homozygous *IroC*^{DFM3} regions.

One way of increasing the size of the clones is to use a cell-lethal (CL) allele. If a CL allele is present in a cell in a homozygous manner, this cell will undergo apoptosis. With the help of this allele, the cells that do not contain the *IroC*^{DFM3} allele will undergo apoptosis, so the proportion of homozygous clonal region will increase. Thus, to increase the size of the mutant clones, FRT80B His2Av-mRFP allele was changed to FRT80B GMR-Hid CL allele.

This allele contains an additional construct, GMR-Hid, which expresses the apoptotic gene *hid* under the control of an eye-specific promoter. While iroC deficient flies do not develop eyes this construct will allow the generation of *IroC*^{DFM3}–deficient whole-eye mutant flies allowing an easy selection of flies of the correct genetic background (Figure 4.5). The *IroC*^{DFM3} allele on its own was used as a negative control and GMR-Hid CL alone was used as a positive control. The flies were phenotyped by following specific markers and observing the development of the eyes in clonal mutant flies and sorted for RNA isolation.

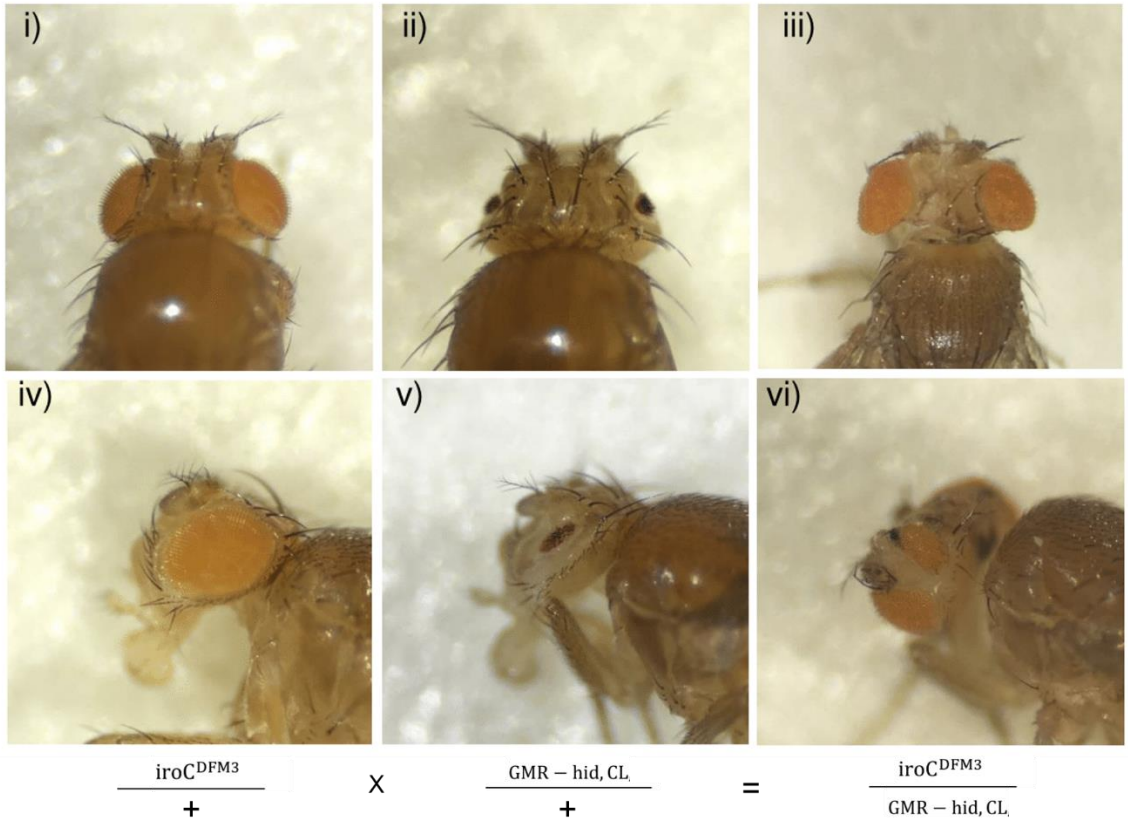

Figure 4.5. Eye morphology of the mutant flies. (i, ii, iii) Dorsal view of the eyes. (iv, v, vi) Side view of the eyes. (i, iv) Positive control. (ii, v) Negative control. (iii, vi) Abnormal eye formation is observed in clonal eyes of the flies.

4.3. Isolation of RNA from Clonal Olfactory Tissues

RNAs were isolated from two different genotypes. Homozygous clonal mutant flies carrying the FRT80B *iro*^{DFM3} allele and the FRT80B GMR-Hid CL allele on the same chromosome and eyFLP on second chromosome were generated as described above. Flies carrying only a FRT80B *iro*^{DFM3} allele were chosen as control. While the mutant tissues contain larger regions of homozygous mutant tissue they still contain some heterozygous mutant tissue for *iro*^{DFM3}. 150 flies were used for each genotype which corresponds to 300 whole antennae and 300 maxillary palps. In order to minimize any kind of variation due to differences in gender and or age of the used flies, 1-10 day old flies were used and 75 males and 75 females were

selected for dissection. For each genotype, 3 biological replicates were prepared. The concentration and purity of isolated RNAs are shown in Table 4.1. The concentrations of the three control samples were 41.7, 58.4 and 74.9 ng/ul and for the homozygous mutant clones 61.2, 80.1 and 72.5. The ratio of absorbance at 260 nm and 280 nm should be between 1.8 and 2.2. The ratio of pure RNA is 2.0. The ratio of absorbance at 260 nm and 280 nm of our result were very close to 2.0 and they were in desired level. For sequencing, requested amount of total RNA should be more than 500 ng. Total RNA amount of our samples was higher than 2000 ng. RIN values of our samples were measured by the Genomics Core and they were at a desired level (RIN>6.0). cDNA library preparation and RNA sequencing experiments were performed by the Genomics Core, UZ Leuven, Belgium.

Table 4.1. Isolated RNA values. Concentration and absorbance values can be seen.

	Concentration(ng/µl)	**260/280**	**260/230**	**Total (ng)**
Heterozgous *iro*^{DFM3} - 1	41.7	1.94	1.63	2085
Heterozgous *iro*^{DFM3} - 2	58.4	1.92	1.77	2920
Heterozgous *iro*^{DFM3} - 3	74.9	1.93	1.76	3745
Clonal *iro*^{DFM3} - 1	61.2	1.92	2.03	3060
Clonal *iro*^{DFM3} - 2	80.1	2.04	1.67	4005
Clonal *iro*^{DFM3} - 3	72.5	1.95	1.92	3625

4.4. Bioinformatics Analysis of RNA Sequencing Results

Raw reads were obtained from the Illumina HiSeq 4000 System. 50 bp single-end option was selected and one lane was used for sequencing. The total number of raw reads that were obtained for each sample are summarized in Table 4.2. The required number of reads for 3 biological replicates was around 10-15 million for differential expression analysis. Sequence depth of our samples were at least 22 million, which is considered sufficient for differential expression analysis.

Table 4.2. Total number of raw reads. Number of reads were varying from 22 million to 32 million. These numbers were enough for differential gene expression for *Drosophila*.

Sample Name	**Read Counts**	**Sample Name**	**Read Counts**
Heterozgous iro^{DFM3} - 1	29.349.145	Clonal iro^{DFM3} - 1	32.403.132
Heterozgous iro^{DFM3} - 2	27.885.795	Clonal iro^{DFM3} - 2	22.298.500
Heterozgous iro^{DFM3} - 3	26.472.190	Clonal iro^{DFM3} - 3	22.676.951

A pipeline for differential expression analysis of raw reads was prepared and is summarized in Figure 4.6. After the raw reads were obtained from the sequencing machine a quality control analysis was done by FastQC and FastQC-mcf tool. FastQC is a quality control tool for high throughput sequence data and it does quality check on raw reads. FastQC-mcf is tool for trimming sequence from raw reads. This analysis yielded the so-called QC plots. In parallel, genome and annotation files were obtained from the Ensembl Genome Browser. The STAR tool was used for both generating genome index files and mapping raw read to the *Drosophila melanogaster* genome using genome indexes. After successful mapping, downstream analysis was done on gene count files by using DeSeq2 tool. This step resulted in a list of genes that were differentially expressed. Using those lists, De Novo motif finding and GO-Term analysis was done by using RSAT, GOrilla, and GO:TermFinder. RSAT is a web-based tool for analysis of spaced dyads in a set of DNA Sequence. GOrilla and GO:TermFinder are web-based tools for GO term analysis from a set of gene name.

4.4.1. Quality Control, Adaptor Removal and Trimming

Quality control analysis was done on raw reads using the FastQC sequencing quality control tool (Andrews *et al.*, 2010), which has previously been shown to be effective on Illumina reads. However, the quality of the 3’ ends of raw reads was not at the desired level. Thus, in

order to increase mappability and get better results, 10 bp from the 3' end was trimmed from each read using the FastQC-mcf software. The reads with a poor quality were defecated.

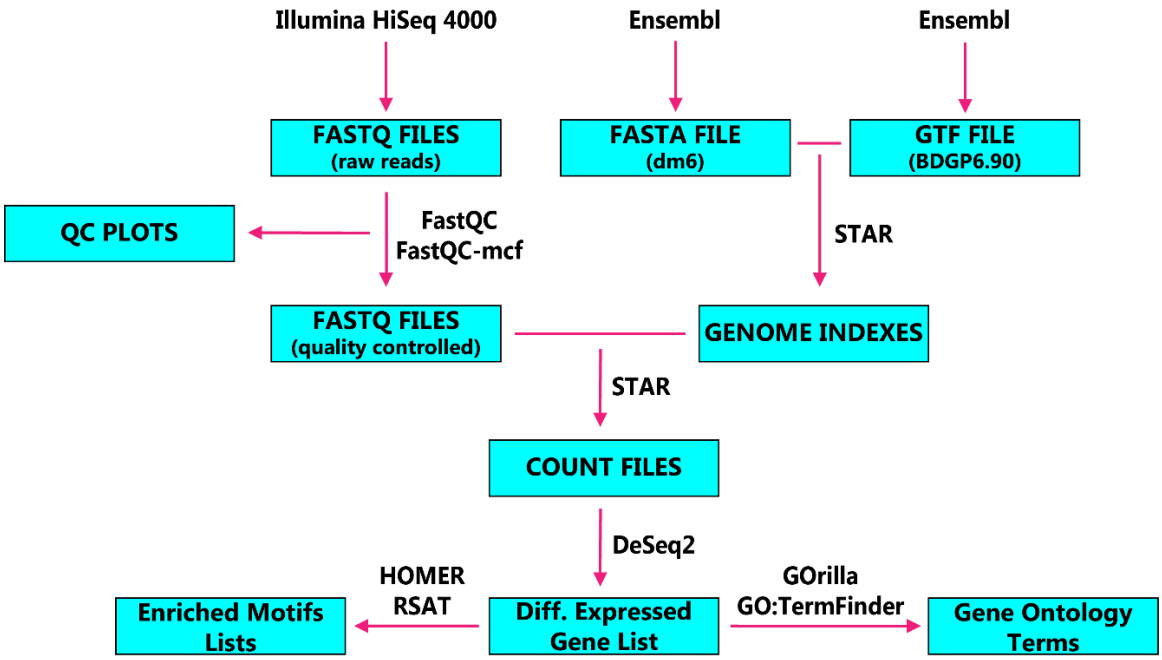

Figure 4.6. RNA-Sequencing pipeline. Name of the files and figures are shown in blue boxes. Used tools and browsers are shown with purple arrows.

Moreover, we detected overrepresented reads. Running a BLAST analysis on these overrepresented reads revealed that these reads are belong to mitochondrial RNA (Boratyn *et al.*, 2012). The method used for mapping raw reads was toward to the genome. Because of that, they wouldn't align to the genome and those reads were posing a danger.

4.4.2. Generating Genome Indexes and Count Files

In the next step, after quality control the trimmed raw reads were mapped to the *Drosophila* genome using the STAR tool (Dobin *et al.*, 2013). STAR is a gapped mapper tool and it is working quite well on *Drosophila* genome. I used the latest release of the *Drosophila melanogaster* genome file (dm6) and annotation file (BDGP6.90). These files were downloaded from the Ensembl Genome Browser (Zerbino *et al.*, 2017). The genome file was in fasta file

format and the annotation file was in gtf file format. Using these two files, genome index files were created before mapping the raw reads to the genome. The genome index files that were generated and fastq files were then used to map the quality controlled raw reads to the genome. As a result, so-called gene count files were obtained that were then used for differential expression analysis.

4.4.3. Finding Differentially Expressed Genes

To identify differentially expressed genes in mutant versus control samples, gene count files were used as an input to the DeSeq2 tool (Love *et al.*, 2014). By using negative binomial generalized linear models, DeSeq2 tool serves methods to test differential expression.

Principle Component Analysis (PCA) is used to assess whether each biological replicate is in the same orientation. It is expected that clonal and control samples should group together separately. Results of the PCA first component shows that control and mutant samples differ from each other (Figure 4.7). Sample-to-sample map distances of control and mutant samples is obtained by using OR gene counts (Figure 4.8). As expected, the three controls and three mutant replicates clustered together but clearly separate from each other.

As an output of DeSeq2 analysis, a list of differentially expressed genes was obtained. MA-plot (log ratio-mean average) is used to show expression amount of each gene and fold change of those genes compared to control samples. The MA-plot for *IroC* is shown in Figure 4.9. Each gene is represented with black dots and significantly differentially expressed genes are shown with red dots. The ORs and transcription factors we are interested in are among red dots.

A list of significantly differentially expressed OR genes and transcription factors are shown in Table 4.3 and Table 4.4., respectively. In first column, names of the OR genes are given. Mean count represents the expression amount of a gene and it correlates with the number of neurons, which express that gene. Fold change shows ratio of expression amount of a gene

in control samples and expression amount of that gene in clonal samples in log_2. P value represents how significant the result is.

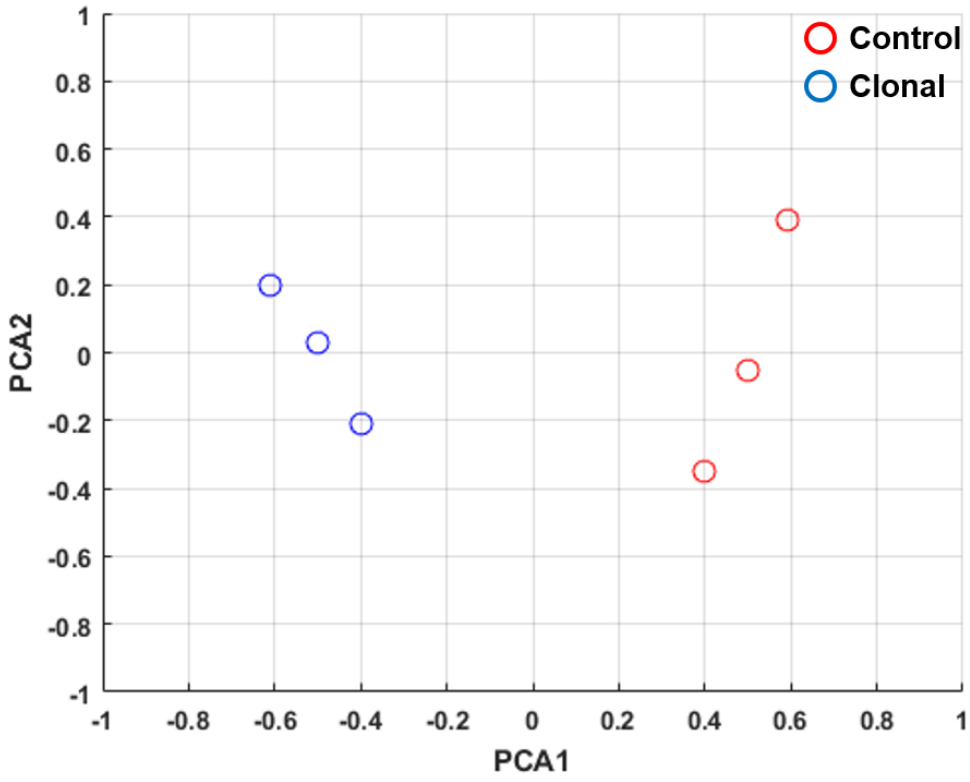

Figure 4.7. Principle Component Analysis (PCA) result. It can be inferred that control and clonal samples were different from each other when the OR gene expression is compared.

Table 4.3. Differentially expressed OR genes. (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001)

OLFACTORY RECEPTOR	MEAN COUNT	FOLD CHANGE (LOG2)	P VALUE
OR47A	1236.71	1.05	****
OR43B	1474.46	0.81	****
OR33C	251.53	0.77	****
OR47B	2087.87	0.66	***

Tablo 4.4. Differentially expressed OR genes. (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001) (cont.)

OLFACTORY RECEPTOR	MEAN COUNT	FOLD CHANGE (LOG2)	P VALUE
OR85A	787.40	0.61	***
ORCO	54059.51	0.51	**
OR71A	352.57	0.71	**
OR9A	1556.74	0.48	**
OR59B	4225.04	0.49	*
OR67A	905.89	0.51	*
OR82A	399.01	0.56	*

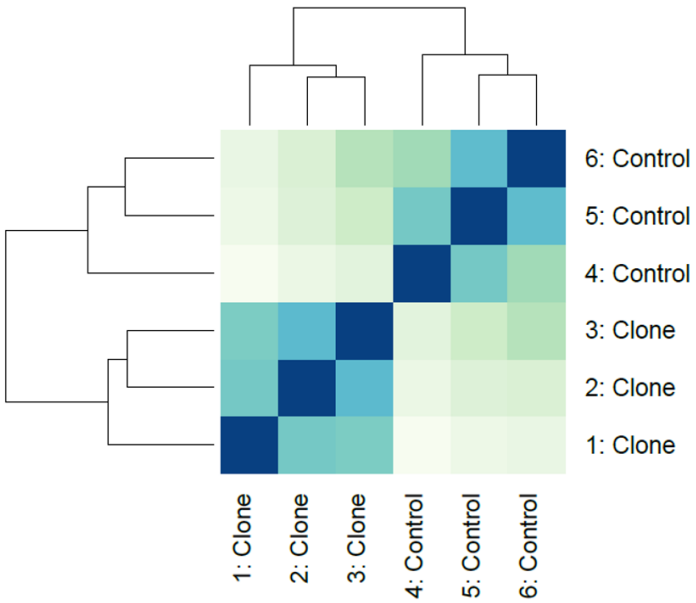

Figure 4.8 Sample-to-sample map distances of control and clonal samples. It can be inferred that control samples were close to each other and had distance from clonal samples and vice versa.

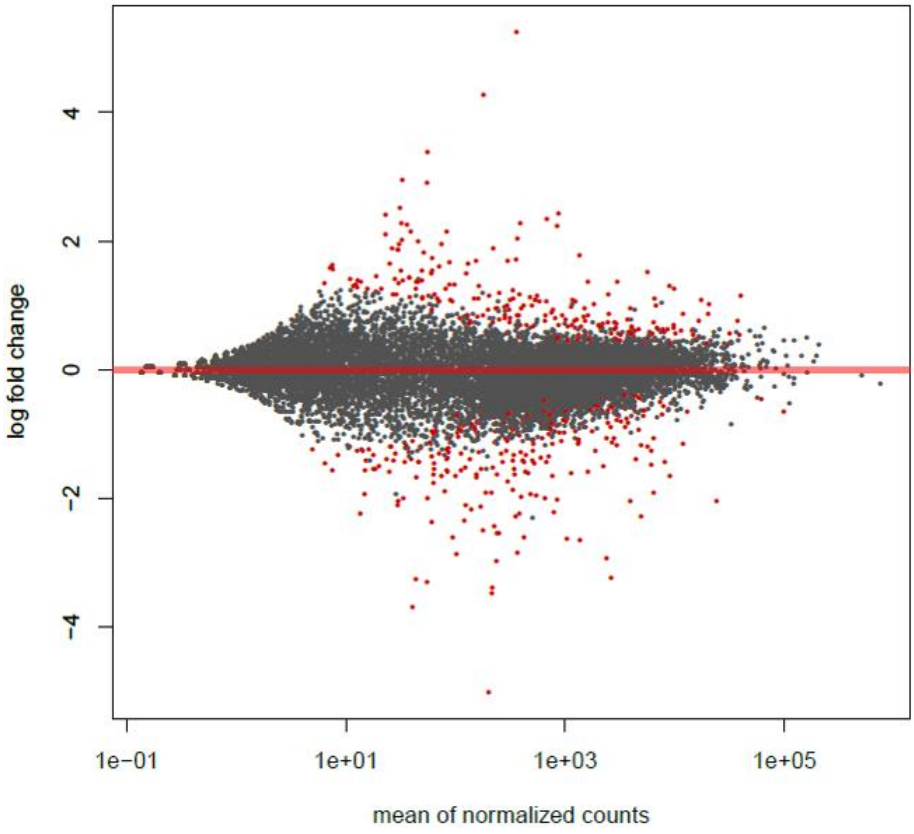

Figure 4.9. MA-plot for *IroC*, control vs clonal. Each dot represents a gene. Red dots represent significantly differentially expressed genes.

4.4.4. GO Term Analysis

Gene Ontology (GO) Term analysis was done on all differentially expressed genes that were obtained by RNA Sequencing analysis in order to perform enrichment analysis on our gene set. Two different servers were used; GOrilla (Eden *et al.*, 2009) and GO::TermFinder (Boyle *et al.*, 2004).

The predominant gene ontology term with the lowest P value was odorant binding. The GO-Term number of odorant binding is GO:0005549. 26 genes were annotated to the term. All of them are significantly differentially expressed genes including OR genes and odor binding

proteins (OBPs). OBPs are secreted in the perilymph in the sensilla by support cells (Swarup, Williams, and Anholt, 2011). They function in the delivery of hydrophobic odorants to ORs (Graham and Davies, 2002).

Results of this analysis are summarized in Table 4.5. The other three enriched gene ontology terms are oxidoreductase activity, tetrapyrole binding, and heme binding. In all of these terms cytochrome proteins are enriched.

Table 4.5. Differentially expressed transcription factors. (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001)

TRANSCRIPTION FACTOR	MEAN COUNT	FOLD CHANGE (LOG2)	P VALUE
NERFIN-1	236.33	-2.97	****
UBX	31.92	2.27	****
REL	5191.23	-0.62	****
SP1	9741.38	0.7	***
HNF4	5683.45	0.46	***
BI	1307.42	0.51	***
PDM3	2908.72	0.47	**
CG32006	3154.86	0.45	**
HEY	34.41	0.97	**
FD64A	5.57	-1.11	**
SS	8590.65	0.43	**
ETS21C	756.41	-0.49	**
BRK	363.21	-0.59	**
PDP1	14561.25	0.45	**
ACJ6	1578.32	0.47	**
NK7.1	1429.88	0.38	*
BTD	369.38	0.52	*
KEN	1251.01	-0.39	*
KAY	4952.65	-0.33	*
TRL	761.09	0.34	*
GATAE	17.03	-0.9	*
FER3	33.53	0.88	*
CT	1272.19	0.33	*

Table 4.6. GO - Genetic Functions term analysis. The term which has the lowest P value is odorant binding. (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001)

Gene Ontology term	P-value	Genes annotated to the term
odorant binding	***	Obp83b, Or67a, Or43b, Or9a, Obp69a, Obp8a, Or85a, Obp99b, Obp56a, Obp56d, Obp56c, Or33c, Obp57c, Or71a, Or59b, Orco, Obp83cd, Obp56h, Obp58c, Or47b, Obp19b, Or82a, Obp57a, Obp19c, Obp56b, Or47a
oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	**	Cyp4p1, Cyp6a14, Desat2, Fad2, Plod, Cyp9b1, Cyp4p2, Cyp6d4, Cyp9b2, Cyp6t3, CG8630, Cyp6a22, Cyp6t1, Skeletor, Cyp4d21, Cyp6a8, Cyp6a23, Cyp4p3, Cyp6g2, Pdi, fid, Cyp313a1, Cyp12d1-p, Cyp6a17, Cyp4e3, Cyp4e1, Cyp6a21, Cyp6w1, Cyp6a2, Cyp313b1
tetrapyrrole binding	**	CG16957, Cyp4p1, Cyp6a14, Cyp9b1, CG3556, Cyp4p2, Cyp6d4, Cyp9b2, Cyp6t3, Cyp6a22, Cyp6t1, Cyt-b5, Cyp4d21, Cyp6a8, Cyp6a23, Cyp4p3, Irc, Cyp6g2, Cyp313a1, Cyp6a17, Cyp12d1-p, Cyp4e3, Cyp4e1, Cyp6a21, Cyp6w1, Cyp6a2, Cyp313b1
heme binding	**	CG16957, Cyp4p1, Cyp6a14, Cyp9b1, Cyp4p2, Cyp6d4, Cyp9b2, Cyp6t3, Cyp6a22, Cyp6t1, Cyt-b5, Cyp4d21, Cyp6a8, Cyp6a23, Cyp4p3, Irc, Cyp6g2, Cyp313a1, Cyp6a17, Cyp12d1-p, Cyp4e3, Cyp4e1, Cyp6a21, Cyp6w1, Cyp6a2, Cyp313b1

The analysis of biological process terms revealed the terms ‘cell development’, ‘cell differentiation’, and ‘organismal development’ as the most significant terms as shown by the lowest P values, which is in line with our expectations and previously identified functions for iroC. GO-Term numbers are GO:0060284, GO:0045595, GO:0010720, GO:0045597, GO:0048522, and GO:2000026. Results can be seen in Table 4.6. When we compared the results with the differentially expressed OR genes, all of the ORs were downregulated. Thus, we can say that iroC transcription factors have a positive regulation on cell development and differentiation and it is consistent with GO Term-Biological Process results. However, it is

known that iroC transcription factors have a negative regulation role on growth (Barrios *et al.*, 2015).

Table 4.7. GO - Biological Process term analysis. (*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001)

GO term	Description	P-value
GO:0060284	regulation of cell development	****
GO:0045595	regulation of cell differentiation	****
GO:0010720	positive regulation of cell development	***
GO:0045597	positive regulation of cell differentiation	***
GO:0048522	positive regulation of cellular process	***
GO:2000026	regulation of multicellular organismal development	***

4.4.5. Search for Known Motifs and De Novo Motif Discovery

In order to identify known IroC binding sites within the region 1000 bps upstream of the transcription start site of differentially expressed olfactory receptor genes, a Python script that I generated was used. 3 different motifs were used. Several binding sites were previously described in the literature. These were $ACAN_{(2\text{-}8)}TGT$ for the *ara* and *caup* binding site found by EMSA (Carrasco-Rando *et al.*, 2011), TTAATTAA for the *ara* binding site found by a DNase I protection assay (J.-L. Gómez-Skarmeta *et al.*, 1996), and ACANNTGT for the *mirr* binding site found by MEME bioinformatics software and P-labeled oligonucleotides with point mutations in EMSA (Bilioni *et al.*, 2005). 5, 3, 3, 2, 1, 1, and 1 binding site were found in the analyzed region of *Or43b*, *Or85a*, *Or47b*, *Or83b*, *Or33c*, *Or71a*, and *Or19a*, respectively (Figure 4.10). On *Or43b* there are 1 ara binding site and 4 ara/caup binding sites. On *Or85a* there are 3 ara/caup binding sites. On *Or47b* there are 3 ara/caup binding sites. On *Or83b* there are 2 ara/caup binding sites. On *Or33c* there is only 1 ara binding site. On *Or71a* and *Or19a* there is only 1 ara/caup binding site.

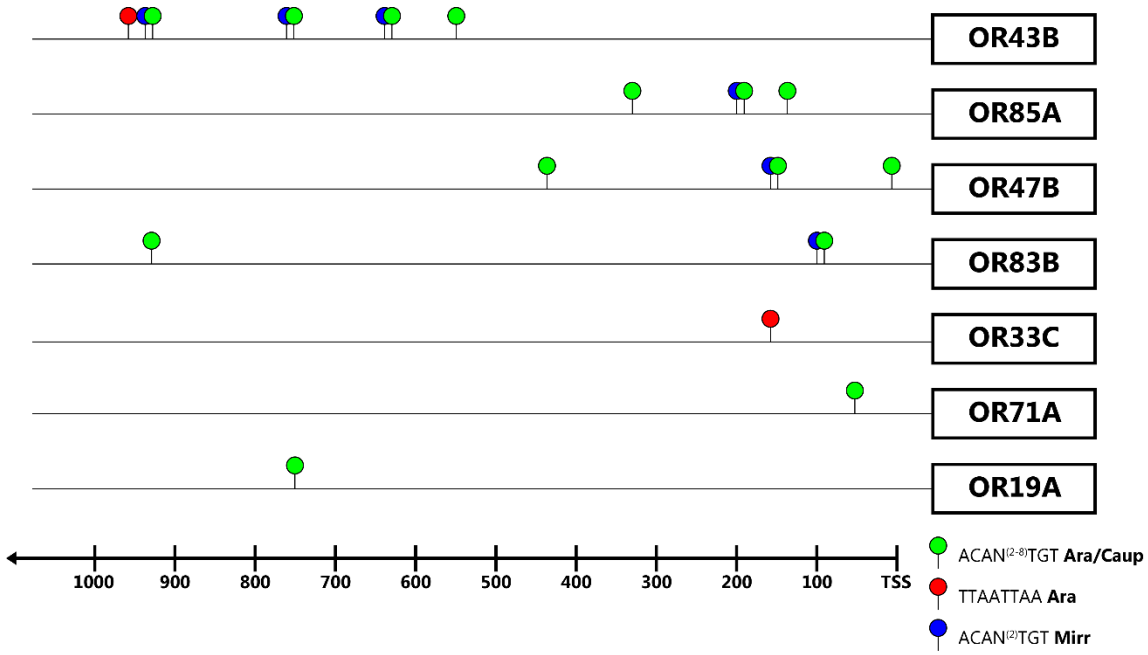

Figure 4.10. Known motifs enriched on differentially expressed OR genes. 3 different known motifs were used; $ACAN_{(2\text{-}8)}TGT$ represented with green circle, TTAATTAA represented with red circle, and ACANNTGT represented with blue circle.

In order to identify known IroC binding sites within the region 1000 bps upstream of the transcription start site of differentially expressed transcription factors, a Python script that I generated was used. Same motifs mentioned before were used. 2, 2, 1, 1, 2, 3, 2, 1, 1, 3, 1, and 1 binding site were found in the analyzed region of *nerfin-1*, *sp1*, *hnf4*, *bi*, *pdm3*, *cg32006*, *fd64a*, *ss*, *ets21c*, *brk*, *pdp1*, and *acj6*, respectively. (Figure 4.11). On *nerfin-1*, *sp1*, and *fd64a* there are 4 ara/caup binding sites. On *hnf4*, *bi*, *ets21c*, *pdp1*, and *acj6* there is only 1 ara/caup binding site. On *pdm3* there are 1 ara and 1 ara/caup binding site. On *cg32006* and *brk* there are 3 ara/caup binding sites. On *ss* there is only 1 ara binding site.

For *De Novo* motif analysis, again 1000 bps upstream region of transcription start site of differentially expressed OR genes were used. The analyzed genes were *Or47a*, *Or43b*, *Or33c*, *Or47b*, *Or85a*, *Orco*, *Or71a*, *Or9a*, and *Or19a*. RSAT Dyad analysis (Medina-Rivera *et al.*, 2015) was used. Two enriched motifs were found, which are ACGCCC on 10 sites and AAGGAA on 20 sites (Figure 4.12)

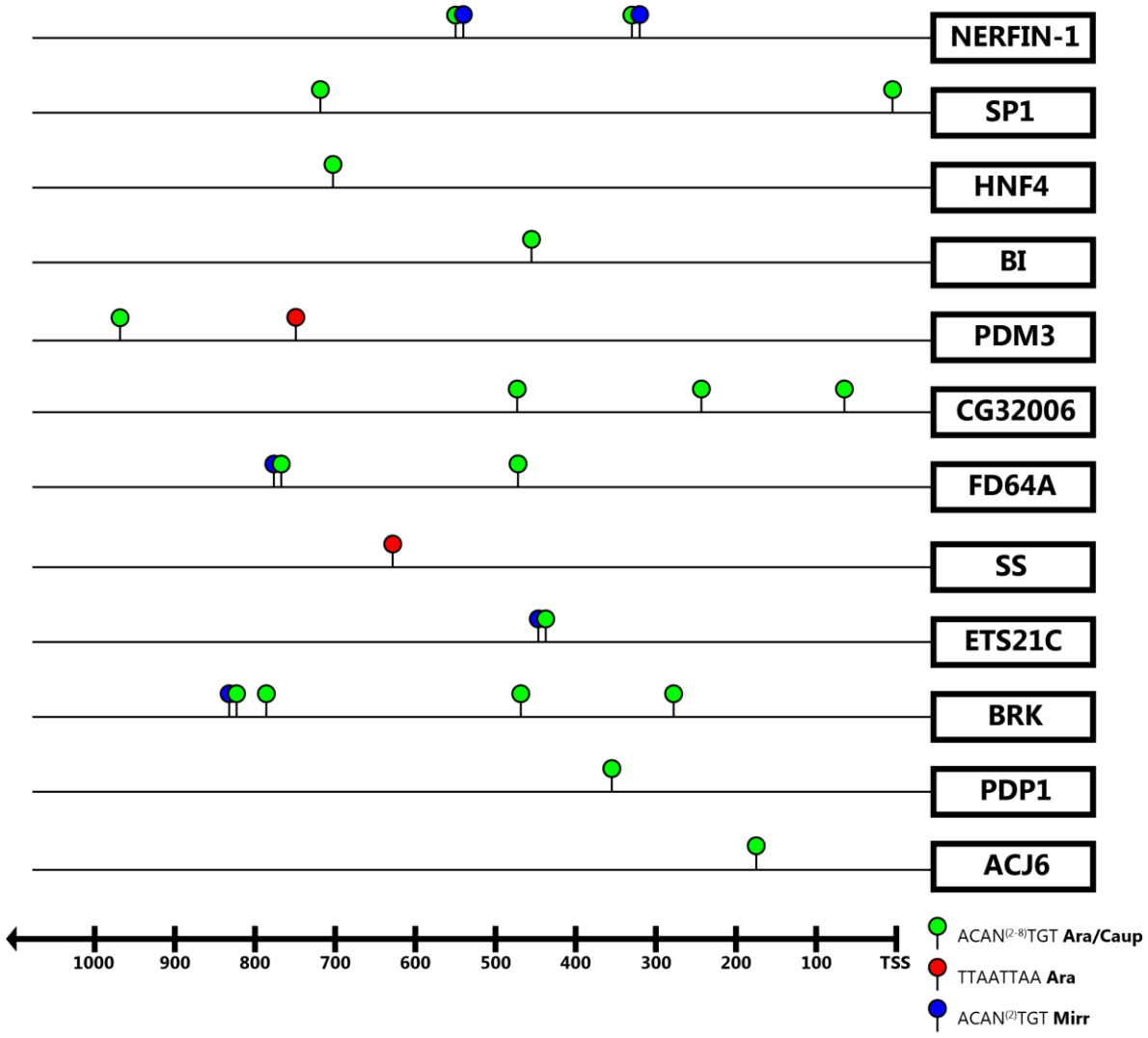

Figure 4.11. Known motifs enriched on differentially expressed transcription factors. 3 different known motifs were used; $ACAN_{(2\text{-}8)}TGT$ represented with green circle, TTAATTAA represented with red circle, and ACANNTGT represented with blue circle.

Enriched motifs found *de novo* were used to find in which site they are present in the upstream region of OR genes. Resulted figure can be seen in Figure 4.13.

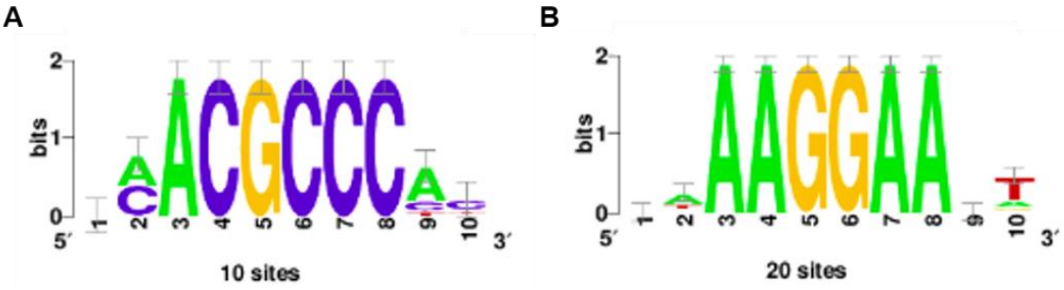

Figure 4.12. Two enriched motifs in the upstream region of differentially expressed OR genes. (A) ACGCCC was enriched in 10 sites. (B) AAGGAA was enriched in 20 sites.

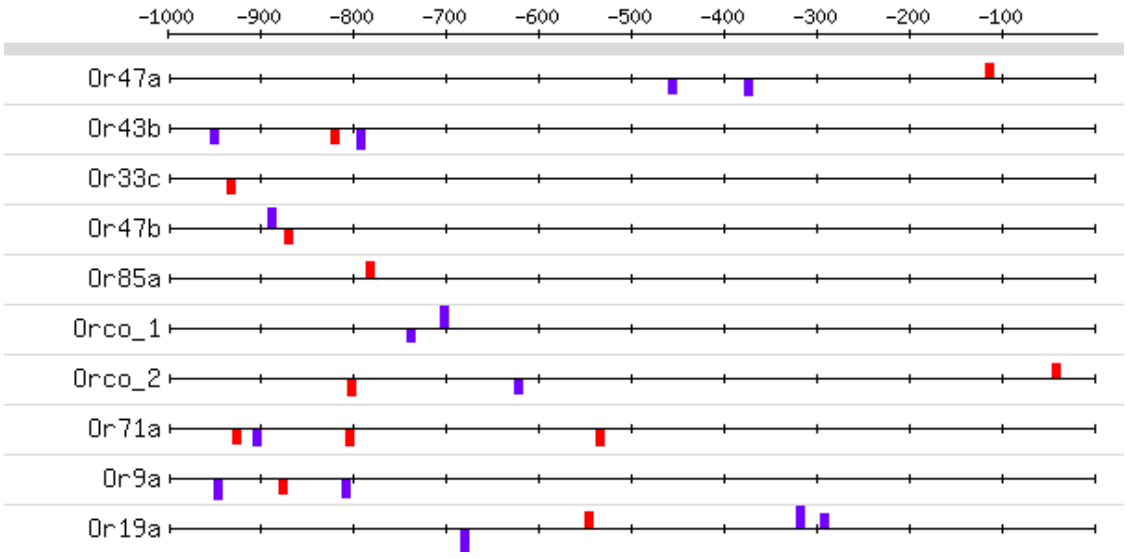

Figure 4.13. Two enriched motifs found *de novo* on differentially expressed OR genes. Red boxes represent AAGGAA and blue boxes represent ACGCCC.

Enriched motifs found *de novo* were used to find in which site they are present in the upstream region of transcription factors. Results are shown in Figure 4.14. While the ACGCCC motif is highly represented, there is only 1 AAGGAA motif on the upstream region of *hey* gene.

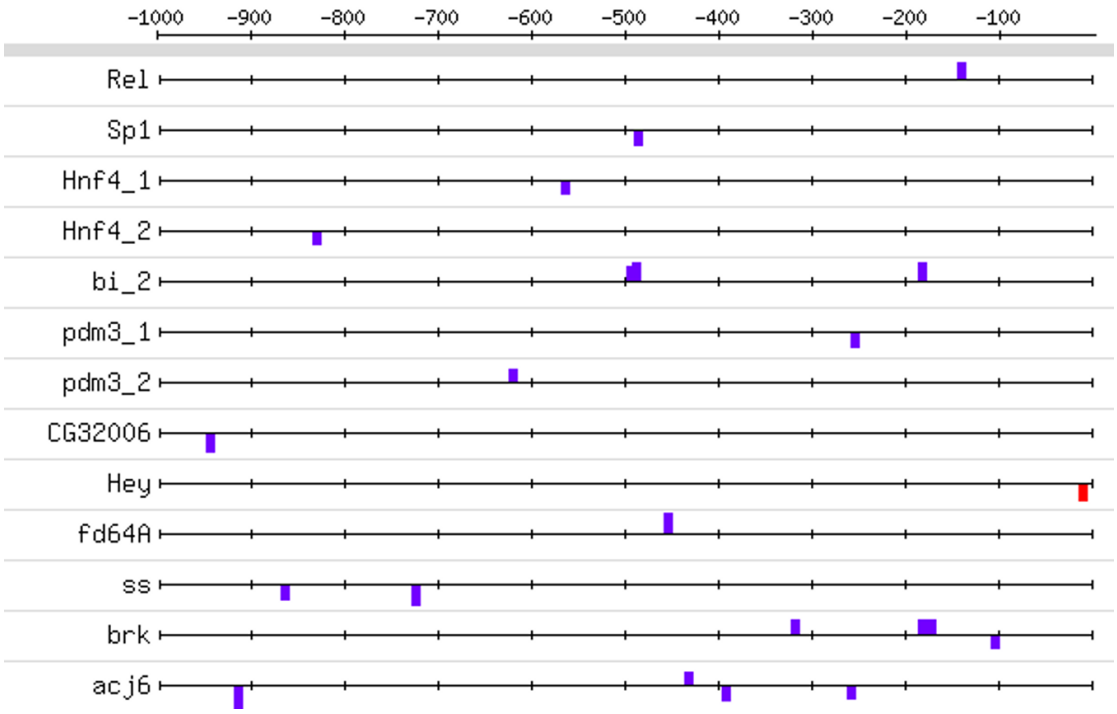

Figure 4.14. Two enriched motifs found *de novo* on differentially expressed transcription factors. Red boxes represent AAGGAA and blue boxes represent ACGCCC.

4.5. Validation of RNA Sequencing Results by qRT-PCR

RNA Sequencing results of differentially expressed OR genes were validated by qRT-PCR experiments. The cDNA of control and clonal samples isolated from antenna and maxillary palp were amplified with specific primers. Primers were designed by using Primer-BLAST web browser (Ye *et al.*, 2012). While designing primers for qRT-PCR, the size of the PCR product was selected to be lower than 250 bps. Also, all primers were designed in a way to ensure that they span exon-exon junction in order to eliminate genomic DNA contamination. *Gapdh1* gene was used as a control to normalize results. 3 biological replicates were done. For each replicate 3 technical replicates and 2 no-template control reactions were prepared. Result were consistent with RNA Sequencing results (Figure 4.15). The fold changes of the genes are 0.45, 0.48, 0.33, 0.30, 0.38, 0.54, 0.70, 0.59, and 0.53 in *Or9a*, *Or19a*, *Or33c*, *Or43b*, *Or47a*, *Or47b*, *Or71a*, *Or85a*, and *Orco*, respectively, while the fold changes of the genes in RNA Sequencing results are 0.71, 0.57, 0.58, 0.56, 0.48, 0.63, 0.61, 0.65, and 0.70, respectively.

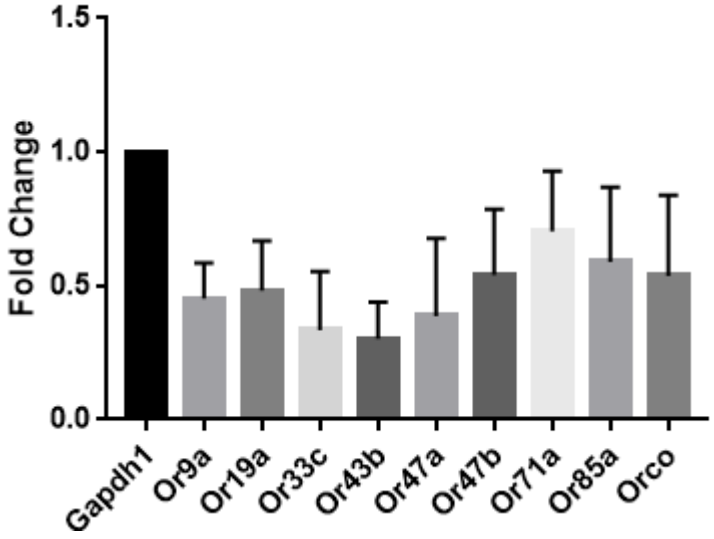

Figure 4.15. qRT-PCR analysis result. *Gapdh1* gene was used as a control. According to qRT-PCR results, all of the genes were downregulated, which is consistent with RNA Sequencing results.

4.6. Improvement of FLY Organ Development Network (FlyOde)

FlyOde is an online platform which is constructed in order to gather all the information about sensory organ development (Koestler *et al.,* 2015). It was initially constructed for eye development. It helps to see gene and protein relationships in a well visualized manner. In this study, we aimed to improve the network by adding genes/proteins relevant for olfactory organ development to the network. Initially, we created FlyOde olfactory system ontology in order to be able to input data to the network. Ontology for antenna and maxillary palp are shown in Figure 4.16 and Figure 4.17, respectively.

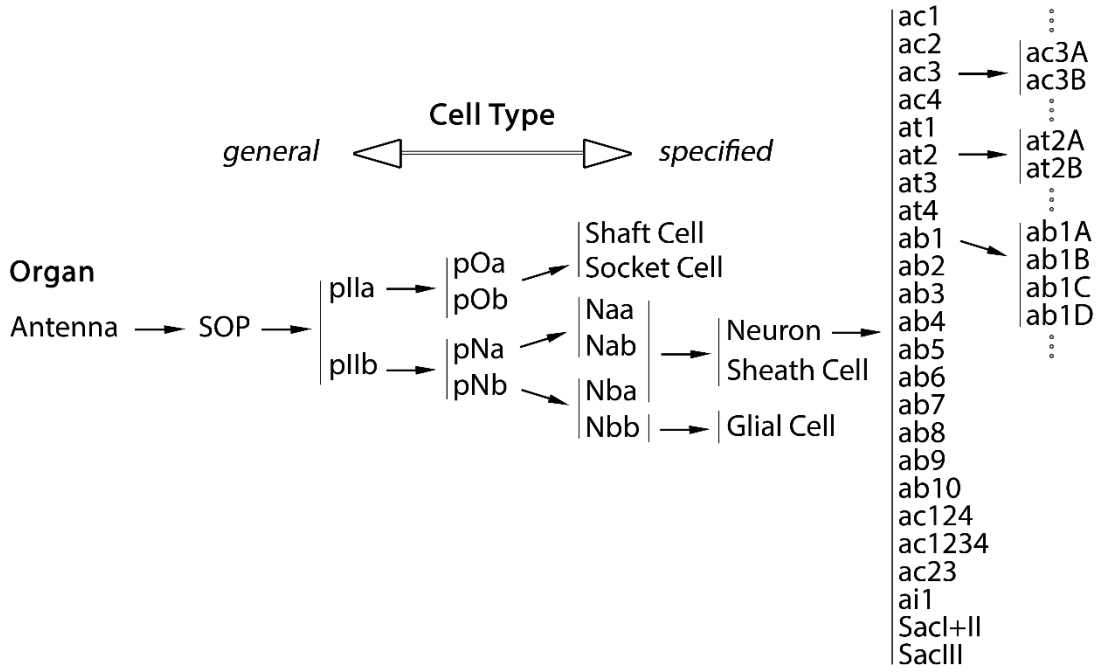

Figure 4.16. Ontology of antenna development for FlyOde

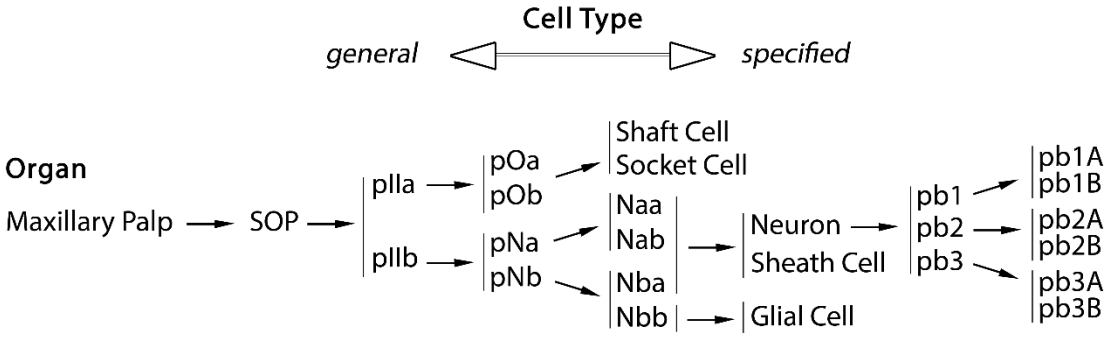

Figure 4.17. Ontology of maxillary palp development for FlyOde

5. DISCUSSION

Sensing the environment for an organism is like holding a torch in the utter darkness. It is crucial from cellular level to organismal level. To be able to sense the environment, we have developed several sensory organs throughout evolution. The most ancient sense among all senses is the sense of smell, olfaction. Olfaction also differs from other senses in terms of its close connection to emotion and memory.

In the olfactory system, as in all sensory systems, detection and interpretation are two key words in order to understand the underlying mechanism of the sensory system. Detection takes place in the sensory organs with the help of sensory neurons. In the olfactory system, in order to be able to sense different odors, a large number of olfactory neurons expressing one receptor out of a large repertoire of olfactory receptors, is used.

For an ORN, there are two main choices in development; which OR gene from this large repertoire to express, and which glomeruli to target in the antennal lobe. The underlying mechanisms that enable these two choices to be made precisely is largely unknown, however experimental evidence indicates that this occurs in a combinatorial manner.

In the framework of this thesis I performed a transcriptome-wide analysis of target genes of IroC proteins, a group of transcription factors expressed in the two olfactory organs of *Drosophila*. We hypothesized that *IroC* may have a role in the OR gene choice mechanism. Identifying target OR genes and transcription factors following by further analysis may illuminate the role of IroC transcription factors in the regulation of OR gene choice.

Despite identified transcription factors and sequence motifs, OR gene choice mechanism is still something of a puzzle. Here, we tried to solve a part of this puzzle by identifying target genes of IroC transcription factors.

Firstly, we have shown the expression of *IroC* genes in both olfactory organs of *Drosophila*, antenna and maxillary palp in a subset of ORNs and hypothesize that they are involved in OR regulation.

In order to test our hypothesis, we performed a loss-of-function experiment followed by RNA Sequencing. To this end, we needed to create *IroC* homozygous deficient flies. Homozygous *IroC* deficiency was lethal. To be able to overcome this lethality we created clonal tissues in the antenna and maxillary palp by using the FLP/FRT system. For tissue-specific *flippase* expression we chose the *eyeless* driver, which is widely used for this type of experiments due to its early expression in the eye, antenna, and maxillary palp.

5.1. Loss of *IroC* Genes Results in the Decrease of OR Gene Expression

RNA Sequencing analysis identified a large number of differentially expressed genes. From that list, firstly, we sorted the OR genes. In order to identify genes, whose expression changed significantly, I set the P value threshold as 0.01 and this resulted in 9 significantly downregulated OR genes, which are *Or47a*, *Or43b*, *Or33c*, *Or47b*, *Or85a*, *Orco* (*Or83b*), *Or71a*, *Or9a*, and *Or19a* in significance order.

Orco is a co-receptor that almost all OR proteins are in a complex with. We can infer two different results from the downregulation of *Orco*. Firstly, *IroC might* regulate the expression of Orco directly or alternatively loss of iroC might cause the mis-specification of cells and loss of neurons and thus be observed as downregulation in the expression of this co-receptor. To distinguish between these possibilities, we analyzed the upstream region of the *orco* gene and identified two putative binding sites for *ara* and *caup*. The analysis of the second possibility is currently underway in our lab.

The other OR genes that appear to be downregulated are located in different sensilla types. *Or19a* is expressed in the at3A ORNs together with *Or19b* in the antennal trichoid sensilla. *Or33c* is expressed in the pb2A ORNs together with *Or85e* in the maxillary palp

basiconic sensilla. *Or85a* is expressed in the ab2B ORNs together with *Or33b* in the antennal basiconic sensilla. *Or47a* is expressed in the ab5B ORNs together with *Or33b* in the antennal basiconic sensilla. It appears interesting that these 4 downregulated OR genes are all co-expressed with other ORs. This result correlates with the study of IroC transcription factors in the regulation of *rhodopsin* genes in *Drosophila* (Mazzoni *et al.*, 2008) where *IroC* genes were shown to induce co-expression of *rhodopsin* genes by breaking "one receptor-one neuron" rule. The same mechanism might be valid in the regulation of OR genes as well. Analysis of their upstream regions revealed IroC transcription factor binding sites in *Or19a* and *Or33c*. However, no known binding motifs could be identified in the upstream region of *Or85a* and *Or47a*, which both co-express *Or33b*. I suggest that these genes might be regulated indirectly by a different mechanism that regulates the *Or33b* gene or there is still an unidentified binding motif of IroC transcription factors.

The other four OR genes that were downregulated in the *IroC* deficient flies are expressed in following sensilla types: *Or47b* is expressed in the at4A ORN in the antennal trichoid sensilla. *Or71a* is expressed in the pb1B ORN in the maxillary palp basiconic sensilla. *Or43b* and *Or9a* are expressed in the antennal basiconic ab8A and ab8B sensilla, respectively. Analysis of IroC transcription factor binding sites revealed their presence in the upstream region of *Or47b*, *Or71a*, and *Or43b*

5.2. Loss of *IroC* Genes Results in the Differential Expression of TFs

In addition to the OR genes the regulation of other transcription factor appears as a plausible way of controlling ORN diversity. Thus, I analyzed the differential expression of transcription factors. In order to get the genes, whose expression changed significantly, I set the P value threshold to 0.01 and identified 15 transcription factors. Five of them, *nerfin-1*, *rel*, *fd64a*, *ets21c*, and *brk* were upregulated, while the others, *ubx*, *sp1*, *hnf4*, *bi*, *pdm3*, *cg32006*, *hey*, *ss*, *pdp1*, and *acj6* were downregulated in the iroC loss of function background.

nerfin-1 had the lowest P value and its expression was increased almost 8-fold in *IroC* deficient tissues compared to controls. It has a role in neuronal development, neuron differentiation, dendrite morphogenesis, and axon guidance. It is a transcription factor and it is known that Nerfin-1 protein represses Notch activity in medulla neurons (Xu *et al.*, 2017). Moreover, the transcription factor hey is known to have a role in modifying Notch activity. Notch is one of the key signaling pathways that plays a role in neuron diversification. iroC is known to modify Notch activity during early eye development. It is interesting that two other Notch modulators appear to be regulated by iroC. May be IroC regulates OR expression indirectly by regulating *hey* and *nerfin-1*. *Nerfin-1* has also been shown to function in neuron development by regulating dendrite morphogenesis similar to the *ets21c* gene, which is also up-regulated in our RNASeq experiments (Parrish *et al.*, 2006). Again, IroC might be regulating OR expression indirectly by regulating *nerfin-1* and *ets21c*.

pdm3 is one of the downregulated genes in the *IroC* deficient tissues and it encodes a POU domain transcription factor. Pdm3 has previously been shown to regulate OR gene expression and axon targeting of olfactory neurons. In *pdm3* mutant maxillary palp tissue, *Or42a* expression is lost (Tichy *et al.*, 2008). In our RNA Sequencing results, however, expression of *Or42a* is not changed. We can infer that IroC transcription factors do not regulate the expression of *pdm3,* as this would result in a change in *Or42a* expression, but rather change the number of cells that express *pdm3* normally. Also we can say that minor amount of pdm3 transcription factor is enough for regulation of *Or42a* expression. Therefore, we can say that IroC transcription factors are probably upstream effectors of the transcription factor pdm3.

acj6 encodes a POU domain transcription factor and similar to *pdm3*, it has a function in OR gene expression and axon targeting of olfactory neurons (Bai *et al*., 2009). It is known to regulate the expression of the OR genes *Or42a*, *Or33c*, *Or85e*, *Or46a*, and *Or59c* in the maxillary palp and *Or42b*, *Or92a*, *Or85a*, and *Or85b* in the antenna. Except for *Or33c* and *Or85a*, the expression levels of other *acj6*-regulated OR genes are not changed in our results. We conclude that, like *pdm3*, *acj6* is not regulated by IroC transcription factors but IroC transcription factors are probably upstream effectors of the transcription factor pdm3.

5.3. *De Novo* Motif Analysis Resulted Two Unique Motifs

By using the 1000 bp upstream region of significantly differentially expressed OR genes, we found two motifs enriched in those sequences. These motifs are ACGCCC in 10 sites and AAGGAA in 20 sites. When we checked whether these motifs are also present in the upstream region of differentially expressed transcription factors, we found that ACGCCC was highly enriched. In order to check whether these motifs were identified before or there are similar motifs to our motifs, we did similarity analysis on motif database by using ACGCCC and AAGGAA. The analysis is resulted that there are no identical or similar motifs in the database. We can conclude that we found two unique motifs in the upstream region of OR genes. To confirm this result, further analysis should be done by generating transgenic lines via cloning minimal promoter constructs of these motifs.

5.4. Future Directions

In order to take the study forward and be able to see the whole picture more experiments need to be done. We established that iroC genes clearly play a role in OR gene expression, although it is unclear if all three genes are involved. Thus, this study could be extended to other mutant lines on which RNA Sequencing experiments could be performed. These lines are $mirr^{E48}$, iro^{EGP1}, iro^{EGP6}, and iro^{EGP7} (Figure 5.1). The $mirr^{E48}$ line is only null for *mirr* gene and it is homozygous lethal. iro^{EGP1} line is only null for *caup* gene and the mutant is homozygous viable. iro^{EGP6} line is null for both *ara* and *caup* gene and it is homozygous lethal. iro^{EGP7} line is null for all *IroC* genes and it is also homozygous lethal. These lines have been recently acquired and are in the process to be prepared for RNA Sequencing. Analysis and comparison of target genes would yield invaluable information about the way these three transcription factors work.

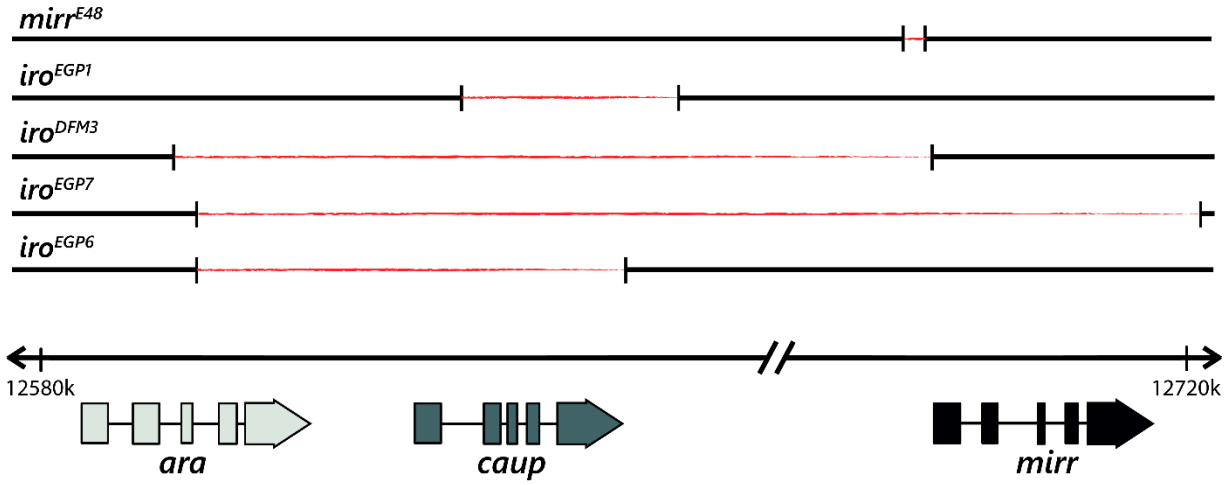

Figure 5.1. Physical map of all IroC mutant lines. The *mirr*^{E48} line is only null for *mirr* gene. *iro*^{EGP1} line is only null for *caup* gene. *iro*^{EGP6} line is null for both *ara* and *caup* gene. *iro*^{EGP7} line is null for all *IroC* genes.

A critical set of data we are still lacking is the exact expression pattern of iroC genes during development. It is unclear how much the widely used iroC lacZ and iroC Gal4 lines reflect the endogenous expression pattern. They clearly are not enough to differentiate the expression of these three genes. Available antibodies do not work well and are not specific enough to distinguish between eg. ara and caup. Thus, towards this end we are generating tagged versions of these proteins to unequivocally determine their expression pattern using CRISPR/Cas9 technology and hope to identify the subset of cells that are iroC positive.

In order to distinguish the expression changes of genes whether it is because of decrease in the expression or loss of the neurons totally or non-formation of a certain kind of a cell type, in situ hybridization experiments should be done on significantly differentially expressed OR genes.

In the meantime, to analyze the lineage of IroC-positive cells, lineage-tracing analysis could be done. For this purpose, we will use the IroC-Gal4 line crossed with UAS-FLP and UAS-(FRT.STOP)-mCD8::GFP line. When *IroC* is expressed the Gal4 will be expressed and

will bind to the UAS promoter to activate *flippase* expression. Flippase will then remove a STOP codon in front of mCD8::GFP allowing its expression. When the olfactory organs are stained with anti-GFP antibody, all of the cells that ever expressed IroC can be visualized and sensilla subsets which express IroC can be identified. This will help to see not only the present *IroC* expressing cells but also the cells that expressed *IroC* before.

To be able to determine the OR expression in mutant background, OR-Gal4 UAS-mCD8::GFP alleles should be added to *IroC* triple mutant flies. The OR alleles should be GFP fused.

Using several bioinformatic tools we have identified candidate IroC binding sites *de novo* from differentially expressed OR genes. In order to understand whether these motifs are IroC binding sites, minimal promoter constructs, containing these binding sites and their mutated versions can be generated by cloning these regions upstream of a Gal4 sequence and generating transgenic lines.

6. REFERENCES

Andreu, M. J., González-Pérez, E., Ajuria, L., Samper, N., González-Crespo, S., Campuzano, S., and Jiménez, G. (2012). Mirror represses pipe expression in follicle cells to initiate dorsoventral axis formation in Drosophila. *Development*, *139*(6), 1110–1114.

Andrews, S., and others. (2010). FastQC: a quality control tool for high throughput sequence data.

Bai, L., Goldman, A. L., and Carlson, J. R. (2009). Positive and negative regulation of odor receptor gene choice in Drosophila by acj6. *Journal of Neuroscience*, *29*(41), 12940–12947.

Bao, Z.-Z., Bruneau, B. G., Seidman, J. G., Seidman, C. E., and Cepko, C. L. (1999). Regulation of chamber-specific gene expression in the developing heart by Irx4. *Science*, *283*(5405), 1161–1164.

Barrios, N., González-Pérez, E., Hernández, R., and Campuzano, S. (2015). The homeodomain iroquois proteins control cell cycle progression and regulate the size of developmental fields. *PLoS Genetics*, *11*(8), e1005463.

Beckingham, K. M., Armstrong, J. D., Texada, M. J., Munjaal, R., and Baker, D. A. (2007). Drosophila melanogaster-the model organism of choice for the complex biology of multi-cellular organisms. *Gravitational and Space Research*, *18*(2).

Bellefroid, E. J., Kobbe, A., Gruss, P., Pieler, T., Gurdon, J. B., and Papalopulu, N. (1998). Xiro3 encodes a Xenopus homolog of the Drosophila Iroquois genes and functions in neural specification. *The EMBO Journal*, *17*(1), 191–203.

Bilioni, A., Craig, G., Hill, C., and McNeill, H. (2005). Iroquois transcription factors recognize a unique motif to mediate transcriptional repression in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14671–14676.

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., and Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, *7*(1), 12.

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., and Sherlock, G. (2004). GO:: TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, *20*(18), 3710–3715.

Brochtrup, A., and Hummel, T. (2011). Olfactory map formation in the Drosophila brain: genetic specificity and neuronal variability. *Current Opinion in Neurobiology*, *21*(1), 85–92.

Carrasco-Rando, M., Tutor, A. S., Prieto-Sánchez, S., González-Pérez, E., Barrios, N., Letizia, A., and Ruiz-Gómez, M. (2011). Drosophila araucan and caupolican integrate intrinsic and signalling inputs for the acquisition by muscle progenitors of the lateral transverse fate. *PLoS Genetics*, *7*(7), e1002186.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

Duffy, J. B. (2002). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis*, *34*(1–2), 1–15.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC*

Bioinformatics, *10*(1), 48.

Endo, K., Aoki, T., Yoda, Y., Kimura, K., and Hama, C. (2007). Notch signal organizes the Drosophila olfactory circuitry by diversifying the sensory neuronal lineages. *Nature Neuroscience*, *10*(2), 153–160.

Fuss, S. H., and Ray, A. (2009). Mechanisms of odorant receptor gene choice in Drosophila and vertebrates. *Molecular and Cellular Neuroscience*, *41*(2), 101–112.

Gómez-Skarmeta, J.-L., del Corral, R. D., de la Calle-Mustienes, E., Ferrés-Marcó, D., and Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell*, *85*(1), 95–105.

Gómez-Skarmeta, J. L., and Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. *Current Opinion in Genetics & Development*, *12*(4), 403–408.

Graham, L. A., and Davies, P. L. (2002). The odorant-binding proteins of Drosophila melanogaster: annotation and characterization of a divergent gene family. *Gene*, *292*(1), 43–55.

Hallem, E. A., and Carlson, J. R. (2004). The odor coding system of Drosophila. *TRENDS in Genetics*, *20*(9), 453–459.

Hildebrand, J. G., and Shepherd, G. M. (1997). Mechanisms of olfactory discrimination: converging evidence for common principles across phyla. *Annual Review of Neuroscience*, *20*(1), 595–631.

Jefferis, G. S. X. E., Marin, E. C., Komiyama, T., Zhu, H., Chihara, T., Berdnik, D., and Luo,

L. (2005). Development of wiring specificity of the Drosophila olfactory system. *Chemical Senses*, *30*(suppl_1), i94--i94.

Kenyon, K. L., Ranade, S. S., Curtiss, J., Mlodzik, M., and Pignoni, F. (2003). Coordinating proliferation and tissue specification to promote regional identity in the Drosophila head. *Developmental Cell*, *5*(3), 403–414.

Koestler, S. A., Alaybeyoglu, B., Weichenberger, C. X., and Celik, A. (2015). FlyOde-a platform for community curation and interactive visualization of dynamic gene regulatory networks in Drosophila eye development. *F1000Research*, *4*.

Kudoh, T., and Dawid, I. B. (2001). Role of the iroquois3 homeobox gene in organizer formation. *Proceedings of the National Academy of Sciences*, *98*(14), 7852–7857.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Mazzoni, E. O., Celik, A., Wernet, M. F., Vasiliauskas, D., Johnston, R. J., Cook, T. A., and Desplan, C. (2008). Iroquois complex genes induce co-expression of rhodopsins in Drosophila. *PLoS Biology*, *6*(4), e97.

McNeill, H., Yang, C.-H., Brodsky, M., Ungos, J., and Simon, M. A. (1997). mirror encodes a novel PBX-class homeoprotein that functions in the definition of the dorsal-ventral border in the Drosophila eye. *Genes & Development*, *11*(8), 1073–1082.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J. A., and Delerce, J. (2015). RSAT 2015: regulatory sequence analysis tools. *Nucleic Acids Research*, *43*(W1), W50--W56.

Morrison, C. M., and Halder, G. (2010). Characterization of a dorsal-eye Gal4 Line in Drosophila. *Genesis*, *48*(1), 3–7.

Münch, D., and Galizia, C. G. (2016). DoOR 2.0-comprehensive mapping of Drosophila melanogaster odorant responses. *Scientific Reports*, *6*, 21841.

Parrish, J. Z., Kim, M. D., Jan, L. Y., and Jan, Y. N. (2006). Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes & Development*, *20*(7), 820–835.

Ray, A., van Naters, W. van der G., and Carlson, J. R. (2008). A regulatory code for neuron-specific odor receptor expression. *PLoS Biology*, *6*(5), e125.

Ray, A., van Naters, W. van der G., Shiraiwa, T., and Carlson, J. R. (2007). Mechanisms of odor receptor gene choice in Drosophila. *Neuron*, *53*(3), 353–369.

Sato, K., Pellegrino, M., Nakagawa, T., Nakagawa, T., Vosshall, L. B., and Touhara, K. (2008). Insect olfactory receptors are heteromeric ligand-gated ion channels. *Nature*, *452*(7190), 1002–1006.

Smitka, M., Puschmann, S., Buschhueter, D., Gerber, J. C., Witt, M., Honeycutt, N., and Hummel, T. (2012). Is there a correlation between hippocampus and amygdala volume and olfactory function in healthy subjects? *Neuroimage*, *59*(2), 1052–1057.

Strausfeld, N. J., and Hildebrand, J. G. (1999). Olfactory systems: common design, uncommon origins? *Current Opinion in Neurobiology*, *9*(5), 634–639.

Swarup, S., Williams, T. I., and Anholt, R. R. H. (2011). Functional dissection of Odorant binding protein genes in Drosophila melanogaster. *Genes, Brain and Behavior*, *10*(6), 648–

657.

Theodosiou, N. A., and Xu, T. (1998). Use of FLP/FRT System to StudyDrosophilaDevelopment. *Methods*, *14*(4), 355–365.

Tichy, A. L., Ray, A., and Carlson, J. R. (2008). A new Drosophila POU gene, pdm3, acts in odor receptor expression and axon targeting of olfactory neurons. *Journal of Neuroscience*, *28*(28), 7121–7129.

Vosshall, L. B., Amrein, H., Morozov, P. S., Rzhetsky, A., and Axel, R. (1999). A spatial map of olfactory receptor expression in the Drosophila antenna. *Cell*, *96*(5), 725–736.

Wicher, D., Schäfer, R., Bauernfeind, R., Stensmyr, M. C., Heller, R., Heinemann, S. H., and Hansson, B. S. (2008). Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels. *Nature*, *452*(7190), 1007–1011.

Wilhelm, B. T., and Landry, J.-R. (2009). RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods*, *48*(3), 249–257.

Xu, J., Hao, X., Yin, M.-X., Lu, Y., Jin, Y., and Xu, J. (2017). Prevention of medulla neuron dedifferentiation by Nerfin-1 requires inhibition of Notch activity. *Development*, *144*(8), 1510–1517.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1), 134.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., and Bhai, J. (2017). Ensembl 2018. *Nucleic Acids Research*, gkx1098.