Developmental Strategy for Generating Sensory Neuron Diversity

by

Qingyun Li

Department of Biology
Duke University

Date:_______________________
Approved:

___________________________
Pelin Cayirlioglu Volkan, Supervisor

___________________________
Hiroaki Matsunami

___________________________
W. Dan Tracey

___________________________
Nina Tang Sherwood

___________________________
Gregory E. Crawford

___________________________
Vikas Bhandawat

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biology in the Graduate School
of Duke University

2015
ABSTRACT

Developmental Strategy for Generating Sensory Neuron Diversity

by

Qingyun Li

Department of Biology
Duke University

Date: ____________________________

Approved:

___________________________
Pelin Cayirlioglu Volkan, Supervisor

___________________________
Hiroaki Matsunami

___________________________
W. Dan Tracey

___________________________
Nina Tang Sherwood

___________________________
Gregory E. Crawford

___________________________
Vikas Bhandawat

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2015
Abstract

Sensory neuron diversity is a common theme in the animal kingdom. It provides the cellular infrastructure that supports the accurate perception of the external world. Among all sensory systems, the olfactory system demonstrates an extreme in the extraordinarily diversified neuronal classes it holds. The system-wide cellular diversity is in sharp contrast with the individual specialization of olfactory receptor neurons (ORNs) *per se*. How the nervous system, particularly the olfactory system, uses limited genetic information to generate a huge variety of neurons with distinct properties remains elusive.

The adult *Drosophila* olfactory system is an excellent model to address this question due to its conserved organizational principles and reduced complexity. The fly olfactory appendages contain 50 ORN classes, each of which expresses a single receptor gene from a family of ~80 genes. Stereotyped clusters of 1-4 ORN classes define about 20 sensilla subtypes, belonging to 3 major morphological types. All cellular components within a sensillum are born by a single sensory organ precursor (SOP) via asymmetric divisions. The molecular mechanisms that determine SOP differentiation potentials to develop into distinct sensilla subtypes and the associated ORN classes are unknown.

From a genetic screen, we identified two mutant alleles in the *rotund (rn)* gene locus, which has a critical function in diversifying ORN classes. Rn is required in a subset of SOPs to confer novel sensilla subtype differentiation potentials from otherwise
default ones within each sensilla type lineage. In \textit{rn} mutants, ORNs in \textit{rn}-positive sensilla subtypes are converted to lineage-specific default \textit{rn}-negative fates, resulting in only half of the normal ORN diversity. This work is described in Chapter 2.

Based on an unbiased time-course transcriptome analysis, we discovered two critical downstream targets of Rn, Bric-à-brac (Bab) and Bar. In light of the knowledge about leg development, we found these genes, along with Apterous (Ap) and Dachshund (Dac), are part of the conserved proximal-distal (PD) gene network that play a crucial role in patterning the antennal precursor field prior to proneural gene-mediated SOP selection. Interactions between these PD genes, likely under the influence of morphogen gradients, separate the developing antennal disc into 7 concentric domains. Each ring is represented by a unique combination of the aforementioned transcription factors, coding the differentiation potentials for a limited number of sensilla subtypes. In most cases, genetic perturbations of the network lead to predictable changes in the ratios of different sensilla subtypes and corresponding ORN classes. In addition, using CRISPR/Cas9 technology, we were able to add tags to specific \textit{rn} isoforms in the endogenous locus, and show the activation of Bab and repression of Bar by the direct binding of Rn to their DNA regulatory elements \textit{in vivo}. This work is presented in Chapter 3.

We propose a three-step mechanism to explain ORN diversification, starting from pre-patterning of the precursor field by PD genes, followed by SOP selection by
proneural genes, and ending with Notch-mediated neurogenesis. The final ORN outcomes are greatly determined by the pre-patterning phase, which may be modified during evolution to compensate for the special olfactory needs of individual species. In our model, each step serves a single purpose, which operates in a context-dependent manner. By changing contexts, reassembly of the same logical steps may guide neuronal diversification in parallel systems with completely different identities. This step-wise mechanism seems to be a common strategy that is used by many other systems to generate neuronal diversity.
Contents

Abstract ........................................................................................................................................ iv

List of Tables ................................................................................................................................ xi

List of Figures ............................................................................................................................... xii

Acknowledgements .................................................................................................................... xv

1. Introduction ........................................................................................................................... 1

   1.1 Neuronal diversity ............................................................................................................... 1

   1.2 Chemosensory systems and olfaction ........................................................................... 4

   1.3 Organization of the adult Drosophila olfactory system .............................................. 8

   1.4 Development of the Drosophila olfactory system ..................................................... 15

      1.4.1 Eye-antennal imaginal disc .................................................................................. 16

      1.4.2 Pattern formation in ventral discs ....................................................................... 16

      1.4.3 Sensory organ precursors and proneural genes ................................................. 21

      1.4.4 Neurogenesis and Notch-mediated binary fate determination ..................... 23

      1.4.5 OR gene choice ...................................................................................................... 26

      1.4.6 Wiring specificity .................................................................................................... 32


   2.1 Summary of the study ..................................................................................................... 36

   2.2 Results .............................................................................................................................. 37

      2.2.1 Rn is expressed in a subset of antennal ORN precursors .................................. 37

      2.2.2 Rn branches off novel ORN precursor fates from default ones in trichoid sensilla ..................................................................................................................... 39
2.2.3 Rn branches off novel ORN precursor fates from default ones in basiconic and coeloconic sensilla .................................................................41

2.2.4 Lineage specification by nesting the transcription factor functions drives precursor cell diversification ........................................................................42

2.2.5 A common motif upstream of rn-positive OR/IR genes is required for their expression ..................................................................................................................43

2.2.6 A decision landscape for antennal ORN precursor diversification ........44

2.3 Discussion ........................................................................................................47

2.4. Experimental procedures .............................................................................70

2.4.1 Fly genetics ..................................................................................................70

2.4.2 RT-PCR .......................................................................................................71

2.4.3 Immunohistochemistry ................................................................................73

2.4.4 Bioinformatics .............................................................................................74

2.4.5 Motif replacement transgene construction...............................................75

2.4.6 Yeast one-hybrid .........................................................................................76

2.4.7 Statistical analysis and mathematical modeling of precursor diversification ..77

2.5 Acknowledgements ..........................................................................................78

3. A Conserved Molecular Network Governing Sensory Neuron Diversification ....79

3.1 Summary of the study ....................................................................................79

3.2 Results ............................................................................................................80

3.2.1 Time-course RNAseq analysis ....................................................................80

3.2.2 Mis-regulation of developmentally critical genes in rn mutants revealed by RNAseq analysis .................................................................82

3.2.3 A common molecular network patterning the antennae and legs ..........84
3.2.4 Intercalation of pre-patterning domains .......................................................... 85

3.2.5 Functional involvement of PD genes in ORN diversification ............................. 87

3.2.6 Sensilla subtype fate mapping onto the pre-patterned domains ..................... 90

3.2.7 Separation of different precursor fates within a ring by Bab concentration gradients ................................................................. 94

3.2.8 Explanations of rn, Ap-overexpression and Bar-overexpression lines using the model proposed in this study .......................................................... 95

3.2.9 Conserved molecular network in patterning gustatory receptor neuron fates ........................................................................................................... 100

3.2.10 Binding of Rn to Bar and Bab in vivo using a CRISPR-mediated Rn-EGFP line ........................................................................................................ 102

3.3 Discussion ............................................................................................................. 105

3.4 Experimental procedures .................................................................................. 139

3.4.1 Fly genetics ...................................................................................................... 139

3.4.2 RNA extraction and library preparation ........................................................ 139

3.4.3 RNAseq analysis .............................................................................................. 140

3.4.4 Immunohistochemistry .................................................................................. 141

3.4.5 Real-time RT-PCR .......................................................................................... 142

3.4.6 CRISPR-based tagging of Rn ........................................................................ 145

3.4.6.1 Molecular cloning ....................................................................................... 145

3.4.6.2 Embryos Injection ..................................................................................... 152

3.4.6.3 Targeted event screening ........................................................................ 152

3.4.7 Chromatin immunoprecipitation .................................................................... 153

3.4.8 Statistical analysis .......................................................................................... 157
3.5 Acknowledgements .................................................................158

4. Concluding Remarks and Future Directions .................................159

4.1 Conclusions ...........................................................................159

4.2 Future directions ....................................................................161

4.2.1 What are the contributions of D/V and A/P axis determination and temporal dynamics of the PD network in ORN diversification? .............................................161

4.2.2 How is the pre-patterning information relayed to proneural gene expression? ........................................................................161

4.2.3 How are the differential potentials coded in SOPs and during divisions? .....162

4.2.4 How is the olfactory system evolved? ........................................166

4.2.5 Can we mathematically model the ORN diversification process? ..........166

Appendix A ..................................................................................168

Appendix B ..................................................................................172

References ....................................................................................184

Biography ....................................................................................197
List of Tables

Table 1: List of representative genes misregulated in rn mutants........................................113
Table 2: Fate mapping of pre-patterning factors with sensilla subtypes.................................125
Table 3: CRISPR injection scheme.........................................................................................136
Table 4: CRISPR screening results.........................................................................................137
Table 5: qPCR primers for olfactory receptors......................................................................142
Table 6: ChIP-qPCR primers.................................................................................................155
### List of Figures

Figure 1: Historical drawing of neurons in the chick cerebellum by Cajal. ........................................2

Figure 2: Zonal restriction of mammalian ORs. ..............................................................................6

Figure 3: Organization of the adult *Drosophila* olfactory system...............................................9

Figure 4: Scanning electron microscopy (SEM) pictures showing the anatomy of olfactory sensilla. .........................................................................................................................10

Figure 5: SEM pictures showing different morphological types of sensilla on the antenna. .................................................................................................................................11

Figure 6: Spatial organization of sensilla types and subtypes on the antenna. .........................12

Figure 7: Hypothetical steps of neuronal diversification.................................................................20

Figure 8: Notch-mediated neurogenesis in the olfactory sensilla. ..............................................24

Figure 9: Molecular characterization of rn mutants. .....................................................................52

Figure 10: Rotund is expressed in subsets of antennal SOPs. ....................................................53

Figure 11: Diversification of new trichoid precursor/ORN identities from default rn-negative ones by Rn.......................................................................................................................54

Figure 12: Diversification of new precursor/ORN identities from default rn-negative ones in coeloconic and basiconic lineages by Rn. .................................................................56

Figure 13: The effects of rn, lz and amos single mutants, as well as rn/lz and rn/amos double mutants on ORN specification. .................................................................57

Figure 14: Requirement of M1 upstream of rn-positive olfactory receptor genes for appropriate OR expression.................................................................58

Figure 15: Combinatorial modulation of ORN precursor specification landscape.................59

Figure 16: Molecular and phenotypic analyses in different rn allelic combinations, MARCM clones and overexpression lines.........................................................61

Figure 17: Rn expression pattern..................................................................................................63
Figure 18: OR expression in trichoid, coeloconic and intermediate sensilla showing sensory conversions in rn mutants. ..............................................................64

Figure 19: ORN specification phenotypes in the basiconic sensilla of rn mutants. ........66

Figure 20: The independent effects of rn and lz on ORN specification in coeloconic ac2 sensilla. .................................................................................67

Figure 21: Presence of M1 motif in the upstream of rn-positive ORs/IRs confirmed by the known motif finding method MAST. .................................................................68

Figure 22: Changes in ORN precursor lineages predicted by the mathematical model. 69

Figure 23: Lineage-specific sensilla subtype conversions in rn mutants....112

Figure 24: Time-course RNAseq analysis showing misregulations of genes critical to development.............................................................................................................115

Figure 25: Overlapping expression of Bab1 and Bab2........................................115

Figure 26: Expansion of Bar but not Ap outside the central fold of the larval antennal. 116

Figure 27: Intercalation of pre-patterning genes outside the central fold.............117

Figure 28: Intercalation of pre-patterning genes inside the central fold. ...............118

Figure 29: Olfactory receptor expression analysis by qPCR in Ap overexpression line. 119

Figure 30: Olfactory receptor expression analysis by qPCR in Bar overexpression line. 120

Figure 31: Expansion of Bar outside CF and repression of Rn inside CF by Ap overexpression. .................................................................................................................121

Figure 32: Limited expansion of Bar and repression of Dac by Ap overexpression....122

Figure 33: Elimination of the central fold by Bar overexpression. ......................123

Figure 34: Modifications of precursor pre-patterned domains in rn mutants, Ap/Bar overexpression lines. .................................................................................................124

Figure 35: Sensilla subtype diversification by cross-regulated transcription factors in concentric domains of the antennal disc. .................................................................126

Figure 36: Olfactory receptor expression analysis by qPCR in bab hypomorphic allele. 127
Figure 37: rn mutant, Ap and Bar overexpression phenotypes explained by the sensilla subtype diversification model. ..........................................................128

Figure 38: Regulation of GRN specification by the conserved molecular network. ....129

Figure 39: Molecular cloning design for CRISPR-induced tagging repair templates. ....130

Figure 40: Plasmid map for CRISPR-induced tagging repair template. ....................131

Figure 41: Flowchart of CRISPR-induced endogenous tagging. ..............................132

Figure 42: Expression pattern of DsRed selectable marker in the repair template. ....133

Figure 43: Correct targeting of the tags by CRISPR showing Rn expression pattern in the eye-antennal disc. ..................................................................................................................134

Figure 44: Faithful labeling of Rn expression by CRISPR-induced tags. ....................135

Figure 45: Normal function of Rn with CRISPR-induced endogenous tags. .............135

Figure 46: Direct binding of Rn to regulatory elements of Bab and Bar by ChIP. .......138
Acknowledgements

First and foremost, I would like to thank my advisor, Pelin Volkan for being a wonderful mentor who will be hard to match in the future. None of these accomplishments would be possible without your encouragement, patience, and generous support over the years. Thank you for taking me as your first graduate student, and giving me the opportunities to work on the projects that truly interest me. Thank you for offering the excellent mentorship with enough flexibility so that I can grow to be an independent scientist. Working with you is the best decision I’ve ever made in my life! Your vision and enthusiasm are and will be influencing my attitude about science for my entire career.

I would like to thank all my committee members, Hiro Matsunami, Dan Tracey, Nina Sherwood, Vikas Bhandawat, and Greg Crawford for your continued support. Your insights, advice and sometimes challenges are indispensable for me to make progresses step by step, from giving a good talk, to publishing my first paper, to choosing a career path....

I would like to thank our collaborators Dean Smith and Corbin Jones. The rotund mutants were isolated in the Smith lab, and that’s where the story begins. Thank Corbin for teaching me how to do bioinformatic analysis.

I would like to thank the current and former lab members for being wonderful colleagues and friends. Thank you for making the lab a collaborative and supportive
place to work and study. Thank you for the suggestions and help you gave along the way. I owe special thanks to Sumie Okuwa and Scott Barish, who greatly contributed to my work, both physically and intellectually.

I would like to thank members of the neighboring labs and the neurogenetics group. I’m particularly grateful to Jorge Marques, Janice Crawford, Regan Price, Serdar Tulu, Anna Chao, Nicholas Rizzo, Andrew George, Yongjian Qiu, Yiting Shi for your technical assistance. I’m also very thankful for Debra Silver and Cagla Eroglu for your suggestions and career advice.

I would like to thank all the great friends I made at Duke, inside and outside the department. I can’t even begin to imagine what my graduate school would have been like if you were not here in these 7 years. Too many good memories with you guys already became part of me. You are the best!

I would like to thank Anne Lacey, Kathleen Pryer, Sönke Johnson, Jill Foster, Jo Bernhardt, Andrew Turnier, Jim Tunney, Sam Johnson and all other Biology staffs who help me directly and indirectly during my time at Duke. It is because the strong departmental support you provide that all the excellent research can be done here.

I would like to thank the people in the fly community who provided reagents and technical supports during my research.

I would like to thank the funding agencies and donors who generously provided financial support for my work and me.
Finally, I would like to thank my family who has always been on my side, to encourage me, to comfort me, and to be proud of me. Even though you are all over thousands of miles away, I’ve never felt alone. You are the reasons why what I am doing is more meaningful. Particularly, my aunt, Liwei Li, who lives in the US, has always been the first person I asked help from. It won’t be enough just to say “thank you”, because you did so much more. To everyone in my family, this dissertation is my love to you in return!
1. Introduction

The goal of this PhD dissertation is to address a central question in developmental neurobiology: How is the nervous system able to generate such an astonishing variety of neurons, particularly sensory neurons, during development? We used the adult *Drosophila* olfactory system as a model to conduct detailed studies of neuronal development. In this chapter, I will justify the rationale for the research, introduce the organization of the system, and review some of the key findings in the field, in order to introduce the specific problem we were trying to solve: the molecular mechanism of olfactory receptor neuron precursor diversification.

1.1 Neuronal diversity

Neuronal diversity is a common theme in the nervous system. This is true for both central and peripheral nervous systems, and it is certainly not restricted to higher order vertebrate organisms. For example, the nematode *Caenorhabditis elegans* only has ~1000 cells, but the nervous system alone contains 302 neurons, which fall into 118 morphologically distinct classes (Hobert, 2010). The fruitfly, *Drosophila melanogaster*, has about 100,000 neurons, which can make classification itself a daunting task, depending on the criteria used. The number of neurons increases exponentially as the survival requirement for an organism to sense and respond to the environmental cues becomes more and more complex along phylogenies. It is estimated that a human brain has approximately 100 billion neurons with 100 trillion connections underlying all the body...
movements, sensations, emotions, and cognition that we are experiencing in any moment.

Figure 1: Historical drawing of neurons in the chick cerebellum by Cajal.

This picture drawn by the Spanish neuroscientist Santiago Remón y Cajal remains one of the best illustrations to demonstrate the extraordinary morphological diversity of neurons in the brain. Neurons with different shapes and sizes can be clearly seen. (Santiago Remón y Cajal, 1905)

The recognition of such great neuronal diversity can be traced back as early as over a century ago to the drawings of the father of modern neuroscience, Santiago Ramón y Cajal (Figure 1). Even in a tiny piece of brain tissue, the morphological differences among neurons (although the neuron theory was only accepted 50 years later) are astonishing.
Distinct morphology is often associated with specialized functions, thereby naturally separating neurons into categories. However, neuronal classification extends far beyond morphological differences. For instance, motor neurons and interneurons in the spinal cord can be easily distinguished based on their different arborization patterns, but motor neurons themselves, which look very similar to one another, can be further divided into subtypes, depending on the cell body positions, muscle targets, and gene expression (Jessell, 2000). Interneurons, on the other hand, can be either excitatory or inhibitory due to the nature of the neurotransmitters they secrete.

Sensory systems are prime examples of neuronal diversity, of which the olfactory system is an extreme. Mice have approximately 5 million neurons expressing over 1000 receptors in the periphery that are dedicated to detecting volatile chemicals, which makes the olfactory receptor gene family the largest in mammals (Niimura et al., 2014). Undoubtedly, this modality is absolutely essential for animals’ survival and reproduction, and such extraordinary diversity truly reflects the enormous sensory demand in a constantly changing chemical world. I will discuss the features of the olfactory system in greater detail in the next section, and it will remain the major topic for the rest of this dissertation.

Albeit the universality and significance of neuronal diversity, little is known about how the nervous system utilizes its limited genetic resources to spawn an astronomical number of diversified neurons with meticulous precision. Therefore,
understanding the fundamental principles that give rise to neuronal diversity is scientifically important and a necessary step before we can manipulate it to advance medical or agricultural applications.

1.2 Chemosensory systems and olfaction

Although primates such as humans have evolved to depend more on the visual system to conduct essential behaviors, the vast majority of organisms in the animal kingdom still primarily reply on chemical cues to navigate through their environments. Animals use two systems, the olfactory and gustatory system, to sense volatile and non-volatile chemicals, respectively. These two chemosensory systems facilitate food searching, toxin avoidance, and elicit social behaviors such as courtship, aggression and parenting (Ihara et al., 2013).

Unlike visual or auditory stimuli that are continuous and confined by particular physical properties, chemical cues can be much more diverse. Alternation of even just a few atoms or a single functional group can change a substance’s physical and/or chemical properties, leading to changes in organismal perception and response. Enantiomers, between some of which humans can discriminate, add further to the complexity of chemosensory systems, making the range of detectable chemicals nearly infinite (Laska, 1999).

In order to adapt to the ever-changing ecological niches, animals are equipped with a massive variety of chemosensory receptors, particularly the olfactory receptors
(ORs) that evolve rapidly (Guo and Kim, 2007; Jiang and Matsunami, 2015; Kambere and Lane, 2007; Niimura et al., 2014; Sánchez-Gracia et al., 2009). Human olfactory receptors were discovered over two decades ago as a large multigene family, which is known now to contain 396 intact members and about 400 pseudogenes (Buck and Axel, 1991; Jiang and Matsunami, 2015; Niimura et al., 2014). The most recent survey of 13 placental species shows that rodents have over 1000 intact ORs, while African elephants register a remarkable record of 1948 functional receptors (Niimura et al., 2014). The majority of mammalian odorant-responsive receptors are G protein-coupled receptors (GPCRs). They signal through the G protein, $G_{olf}$ that produces cAMP, which in turn activates cyclic nucleotide-gated ion channels causing $\text{Na}^+$ and $\text{Ca}^{2+}$ influx and subsequent neuron depolarization (Brunet et al., 1996).

Insects have a much simpler olfactory system with an order of magnitude fewer receptors (Sánchez-Gracia et al., 2009). For example, the commonly used genetic model, *Drosophila melanogaster*, has about 80 olfactory receptor genes expressed in the adult olfactory appendages. Sixty-two of these encode distinct seven trans-membrane proteins with reverse topology compared to GPCRs (Couto et al., 2005; Fishilevich and Vosshall, 2005). They are known as the conventional insect ORs. Individual ORs are thought to form a complex with the conserved co-receptor Or83b and function as odorant-gated non-selective cation channels (Larsson et al., 2004; Nakagawa et al., 2012; Sato et al., 2008; Wicher et al., 2008). Fifteen of the rest are related to ionotropic glutamate receptors, thus
named IRs (Benton et al., 2009). They form heteromeric complexes with up to three subunits from the same family to conduct currents upon ligand binding. In addition, co-expression of two gustatory receptors Gr21a and Gr63a in an olfactory neuron class represents a rare case, which is specifically tuned to detect carbon dioxide (Jones et al., 2007; Kwon et al., 2007). To avoid confusion, “olfactory receptors” will refer to all receptors that respond to odors, whereas the abbreviation “ORs” will denote conventional OR genes in this dissertation.

Figure 2: Zonal restriction of mammalian ORs.

(A) and (B) The mammalian olfactory epithelium is shown. Four broad zones in banded shapes and the patch (P) region are labeled. Each neuronal class expressing a single receptor is restricted to one specific zone. Within a zone, different neuronal classes are intermingled, as shown by the expression of MOR28 and 230-1. (Figure modified from Fuss, Stefan H., Ray, Anandasankar. Mechanisms of odorant receptor gene choice in Drosophila and vertebrates. Molecular and Cellular Neuroscience. 2009. 41: 101-112). (C) A scheme of the mammalian turbinates showing different neuronal classes are restricted to different broad zones. Colored dots represent different OR subfamilies. (Figure modified from Vassar, R., Ngai, J., Axel, R. Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. Cell. 1993. 74(2): 309-318)
Despite the divergence in the nature of receptor genes between vertebrate and insects, the organizational principle of the system is remarkably similar. In either case, individual olfactory receptor neurons (ORNs) typically express a single odorant-selective receptor gene from a large gene family pool, and neurons that express the same receptor constitute an ORN class that is intermingled with other classes in a restricted region of the peripheral tissue (Figure 2). Therefore, extreme sensory and functional specialization of individual ORNs is accompanied by a high level of diversity of ORN classes as a whole. In the CNS, neurons of the same class terminate their axons in a stereotypical area in the first information relay center, the olfactory bulb (in mammals, or the antennal lobe in flies), where they make synapses with the second order neurons in a structure called a glomerulus (Vassar et al., 1994; Vosshall et al., 2000). Since each odorant can activate multiple receptors and each receptor can be activated by multiple odorants, combinatorial activation of subsets of ORN classes and subsequent glomerular activity patterns are believed to be the initial step of odor information coding (Vassar et al., 1994).

In light of the above discussion, I would like to point out a number of reasons why we used the adult Drosophila olfactory system as a model to study neuronal diversification: (1) it is a complex system with sufficient diversity, which calls for sophisticated mechanisms of differentiation; (2) its numerical complexity is much reduced, which makes systems-level investigation possible; (3) olfactory diversity is the
cellular basis of a faithful perception of the external world, and thus important; (4) the organizational principle is conserved across species; (5) the evolvability of the system may link individual development to evolution and speciation; (6) flies are genetically tractable and research tools are abundant.

1.3 Organization of the adult Drosophila olfactory system

Adult *Drosophila* has two pairs of olfactory sensory appendages: the third segment of antenna (funiculus) and maxillary palp (Figure 3). Although the antenna itself is segmented, the olfactory funiculus possesses contiguous anatomy. The surface of these olfactory organs are covered by multiporous sensory hairs, called “sensilla”, which are intermingled with non-innervated hairs, called “spinules” (Figure 4, Figure 5). Each antenna and maxillary palp contains about 410 and 60 sensilla, respectively (Stocker, 1994). The base of a sensillum houses clusters of 1-4 (usually 2) ORN cell bodies, accompanied by the non-neuronal shaft, socket and sheath cells (Rodrigues and Hummel, 2008). The total numbers of ORNs on each side of a fly head are around 1300 from the antenna and 130 from the maxillary palp (Shanbhag et al., 1999; Stocker, 1994). Odorant receptors are expressed on ORN dendrites, which extend into the cilia bathing in the sensilla lymph. As mentioned above, each ORN typically only expresses a single receptor gene from a repertoire of 80 genes, and thus a total of 50 ORN classes are clustered in stereotypical combinations corresponding to sensilla types and subtypes.
Figure 3: Organization of the adult *Drosophila* olfactory system.

Schematic representation of the fly olfactory system is shown. Antennae and maxillary palps are covered with sensilla that house ORN cell bodies. Each ORN typically expresses one olfactory receptor gene on ORN dendrites. ORNs that express the same olfactory receptor target a class-specific glomerulus in the antennal lobe. Projection neuron (red fibers) synapse with ORNs in specific glomeruli, and send axons to mushroom bodies and lateral horn. ORNs stand for olfactory receptor neurons. The fly head picture is modified from the cover image of *Science*. 2010. 327 (5970).
Figure 4: Scanning electron microscopy (SEM) pictures showing the anatomy of olfactory sensilla.

Figure 5: SEM pictures showing different morphological types of sensilla on the antenna.

(A) The third segment of antenna (silver-stained). The boxed areas are shown in (B, C). (B) The dorsal-medial region of the antenna showing thin basiconica (TB), large basiconica (LB), and non-innervated spinule (Sp). (C) The ventral-lateral region of the antenna showing trichoidea (T), small basiconica (SB), coeloconica (C), and intermediate sensilla (I). (Figures modified from Shanbhag, S.R., Müller, B., Steinbrecht, R.A. Atlas of olfactory organs of Drosophila melanogaster. Arthropod Structure & Development. 1999. 28: 377-397.)
Figure 6: Spatial organization of sensilla types and subtypes on the antenna.

(A) Spatial organization of sensilla types on the antenna. Trichoid, basiconic (large, thin, and small), and coeloconic sensilla occupy distinct zones. Rare intermediate sensilla are not shown. (B) Sub-segmentation of the trichoid sensilla type into subtypes (a1-at4) is exemplified on the right. ORNs in each subtype express an invariable combination of OR genes (Or19a and Or19b are co-expressed; Or65a, Or65b and Or65c are co-expressed). (Figure modified from Li, Q. et al. Combinatorial rules of precursor specification underlying olfactory neuron diversity. Curr Biol. 2013. 23: 2481-2490.)

Antennal sensilla have 3 major morphological types: club-shaped basiconica (ab: antennal basiconic), spine-shaped trichoidea (at) with a prominent basal drum, and cone-shaped coeloconica (ac) with groove-like pores called cuticular fingers, in addition to the rare intermediate type (ai) that appears to be a hybrid between basiconic and trichoid sensilla (Shanbhag et al., 1999) (Figure 4, Figure 5). Basiconic sensilla have further morphological subdivisions into small, large and thin basiconica. On the antenna, each sensilla type occupies different zones with basiconica taking the dorso-medial region, tricoidea predominantly found in the ventral-lateral area, and coeloconica blended in between (Rodrigues and Hummel, 2008). Basiconica and trichoidea are the more abundant types, and also show sexual dimorphism where males have 30% more trichoidea at the expense of 20% fewer basiconica compared to females (Stocker, 1994).
Maxillary palps only have thin basiconic sensilla (pb: palp basiconic) (Rodrigues and Hummel, 2008).

Each morphologically distinct sensilla is further segmented into generally 4, sometimes 3 subzones based on sensilla subtypes, which are defined by the unique subsets of ORN classes that express invariable combinations of olfactory receptors (Figure 6). Numbers are used following sensilla type symbols, i.e. ab, at, ac, pb, to distinguish different subtypes. Basiconic and tricoid sensilla contain ORNs that express conventional insect OR genes, except one ORN in the large basiconic subtype ab1 that expresses CO$_2$-sensing gustatory receptors (Jones et al., 2007; Kwon et al., 2007). Most basiconic ORs are tuned to fruit odors such as alcohols, ketones and esters, while tricoid ORs are responsive to fly-derived pheromones (van der Goes van Naters and Carlson, 2007; Hallem and Carlson, 2006). Coeloconic sensilla generally contain IR-expressing ORNs that are tuned to carboxylic acids and amines (Ai et al., 2013; Silbering et al., 2011).

Because of the zonal localization of sensilla types/subtypes and their defined relationships with olfactory receptor genes, the expression of a given receptor is accordingly restricted to a specific zone, and thus all ORs collectively form a sensory map in the periphery (Couto et al., 2005; Fishilevich and Vosshall, 2005; Vosshall et al., 1999, 2000). Interestingly, despite the separation between Drosophila and mammals by hundreds of millions of years of evolution yielding drastic anatomical differences, the principle of zonal restriction of OR expression seems to be conserved (Figure 2, Figure 6)
(Fuss and Ray, 2009; Ressler et al., 1993; Vassar et al., 1993). This precise structural and molecular organization raises an interesting question of how it is established during development, and also makes the adult *Drosophila* peripheral olfactory system an attractive model to approach the problem of ORN diversification.

As the antennal ORN axons leave the periphery, they form three main fascicles that converge to become the antennal nerve. Then, they enter the antennal lobe via the ventral-lateral entry point, and defasciculate into a number of lateral and medial pathways across the antennal lobe. Maxillary palps ORNs invade the antennal lobe from labial nerve ventral to the lobe.

All ORNs that express the same receptor project their axon terminals to a single stereotypical glomerulus identified by its unique size, shape and location, where they form class-specific synapses with dendrites of monoglomerular projection neurons (PNs) (Figure 3). Most ORNs make bilaterally symmetric connections via the antennal commissure, which facilitate the integration of information from both lobes. Notably, the spatial relationship among ORN cell bodies in the antenna seems to be topographically preserved by ORN axons in the antennal lobe. For example, the laterally localized trichoid ORNs predominantly project to the lateral half of the lobe, while medially localized basiconic ORNs project to the medial region more often (Couto et al., 2005; Fishilevich and Vosshall, 2005).
Therefore, around 50 ORN classes corresponding to 50 glomeruli on each side of the brain separate olfactory information into 50 distinct channels during the first relay (Jefferis and Hummel, 2006). PNs on the other hand send axons to the mushroom body and lateral horn for further information processing (Lin et al., 2007; Wong et al., 2002). This wiring specificity requires sophisticated coordination between different compartments of the same neuron and all partners involved, which is subject to precisely developmental control (Hong and Luo, 2014; Jefferis and Hummel, 2006). As a result of this “one receptor-one neuron-one glomerulus” rule, ORN sensory identities are closely linked with glomerular identities, which can be reversely utilized to deduce ORN fate origins during precursor fate mapping.

1.4 Development of the Drosophila olfactory system

Due to the extremely stereotyped organization of the Drosophila olfactory system, it has been an ideal model to study neuronal fate specification and circuit assembly for many years. Despite the accumulating knowledge about the development of the system, a couple of big gaps still exist, one of which is how the 50 different ORN classes are generated from an initially homogenous population of precursor cells. Specifically, what are the factors that “paint” the precursors with distinct colors, which eventually lead to highly heterogeneous precursors giving rise to different terminal fates? If we could find such mechanisms, what can we learn from it about some general principles of neuronal diversification? These are the questions that my dissertation is hoping to address.
In the following sections, I would like to discuss the current understanding of ORN fate specification and olfactory receptor gene choice prior to this work. Towards the end, I will briefly review a slightly different but related topic, which is olfactory neuron wiring specificity.

1.4.1 Eye-antennal imaginal disc

Like all other adult tissues, the adult peripheral olfactory system develops from imaginal discs. The discs that give rise to antennae and maxillary palps are called the eye-antennal imaginal disc complex, a part of which, as the name indicates, also specifies the eye. Imaginal discs are single-layered epithelial cells that invaginate from the surface of the embryo to form sac-like structures and carried in the larval body until metamorphosis. They are in an oval shape with concentric folds in invariant positions, thereby giving them a translucent disc-like appearance (Figure 24G). Imaginal discs are named after the corresponding adult appendages, which can be classified as two big groups: the ventral appendages, such as the antennae, legs, mouthparts and analia, and the dorsal appendages including the wings and halteres. The former appendages are thought to share a common ancestral leg origin, and the latter ones are specialized organs for insects (Morata, 2001).

1.4.2 Pattern formation in ventral discs

All ventral discs are differentiated from the rest of the embryo by a central determinant factor, the homeodomain transcription factor distal-less (dll). It is in turn
activated in the lateral embryo by the combinatorial effect of *wingless* (*wg*), *decapentaplegic* (*dpp*), epidermal growth factor receptor (EGFR) signaling during early embryogenesis (Morata, 2001). Interestingly, *dll* has an evolutionarily conserved function to promote body wall outgrowth from insects to humans. The dorsal discs execute a different set of developmental programs, and use *vestigial* (*vg*) as the determinant (Williams et al., 1991). Meanwhile, homeotic genes are selectively activated in each individual disc according to their anterior-posterior (A/P) positions to further differentiate tissues with the same D/V identity (Morata, 2001). For example, the antennal disc is specified by *Dll*, Homothorax (*Hth*) and Extradenticle (*Exd*) (Dong et al., 2000, 2002). *Hth* is an A/P homeotic selector gene, which is required for the nuclear localization of its broadly expressed partner *Exd*. Ectopic co-expression of *Hth* and *Dll* is sufficient to confer antennal identity in other tissues. Likewise, the homeotic gene Antennapedia (*Antp*) induces leg fate, and forced expressed of *Antp* can transform antennae into legs (Casares and Mann, 1998; Dong et al., 2000; Postlethwait and Schneiderman, 1971; Schneuwly et al.; Yao et al., 1999). These results suggest that both primordia have the potential to develop into either antennae or legs (and other ventral appendages alike) probably by sharing similar downstream programs, but the selector genes sitting at the top of the hierarchy somehow determine what the final outcome would be. Consistent with this notion, the patterning strategies by morphogen gradients and transcription factor networks seem to be very similar among all ventral discs. These
processes have mostly been studies in the leg development with some comparisons to the antenna.

Specifically, segment polarity genes, *engrailed (en)*, is expressed in the posterior compartment of a disc and activates *hedgehog (hh)*, whose activity is blocked by *en* (Estella et al., 2012; Morata, 2001). Hh diffuses into the anterior compartment and induces its receptor Patched (Ptc) only in the anterior compartment. Ptc in turn sequesters Hh along the A/P boundary forming a border-high to anterior-edge-low gradient of Hh. Therefore, the Ptc-expressing cells and En-expressing cells separate the disc into anterior and posterior halves, respectively. In addition, Hh also activates *dpp* in the dorsal anterior and *wg* in the ventral anterior (Lecuit and Cohen, 1997). They diffuse in both anterior and posterior directions, but are confined within their own D/V domains by antagonizing each other’s function. The highest concentrations for both genes lie in the center, which activates late-expressed Dll and EGFR signaling ligand, *vein*, which diffuses radially (Kojima, 2004). The center becomes the future distal-most point and the concentration gradient of EGFR along with Dll trigger a transcription factor regulatory network, which sets up a series of domains along the proximodistal axis. Therefore, the morphogen gradients of Dpp and Wg guide both dorsoventral and proximodistal axes formation.

Notably, the proximodistal axis is a distinct feature for insect appendages or vertebrate limbs. Unlike body parts that have inherent A/P and D/V axes, the structures
along the length of the limbs need to be assembled *de novo*. Because the body plan of *Drosophila* is segmented, it allows for unambiguous fate mapping. *Drosophila* legs have long been used as a model to investigate the developmental mechanisms of proximodistal axis formation. As mentioned above, the picture that has emerged from decades of studies, shows a chain of events, starting from Hh-Wg-Dpp morphogen gradients to Dll and EGFR signaling to a dynamic transcription factor (TF) network (Estella et al., 2012; Galindo et al., 2002; Kojima, 2004). The TFs that are involved in this network include, but not limited to Aristaless (Al), BarH1/BarH2 (B-H1/B-H2, or Bar), Apterous (Ap), Bric-à-brac1/2 (Bab1/2, or Bab), Spineless (Ss), Rotund (Rn), and Dachshund (Dac). They are activated in distinct overlapping concentric domains relative to the center, partially due to a gradient of EGFR signaling. The cross-repressive functions between these factors play a major role in refining the boundaries between each ring, resulting in the expression of a unique combination of TFs that define each concentric ring (Giorgianni and Mann, 2011; Natori et al., 2012). This mechanism ensures that morphological features that are associated with each distinct proximodistal locus can be faithfully specified. However, how exactly these handful of master regulators modify the molecular architecture within the precursor cells and how these changes eventually lead to rather heterogeneous populations of terminally differentiated cells are unknown.
Since neurons are an integral part of a given appendage, it is plausible to suspect that the patterning mechanism plays a role in guiding the initially homogenous precursor cells to generate diversified neurons. Indeed, such strategy has been applied in other developing systems. For example, during spinal cord neuron diversification, patterning genes set up 11 domains along D/V axis prior to neuron fate specification, and this is believed to account for the majority of neuronal diversity seen in the system (Jessell, 2000; Ulloa and Briscoe, 2007). Considering the stereotyped organization of the fly olfactory system which calls for a deterministic mechanism, along with the fact that the patterning event also takes place before the neural precursor selection process during antennal disc development, the “patterning-precursor selection-neurogenesis” strategy seems to be an attractive model underlying the extraordinary diversity of olfactory receptor neurons (Figure 7). This hypothesis will be explored and discussed in greater detail in Chapter 3.

![Figure 7: Hypothetical steps of neuronal diversification.](image)

Schematic representation showing the three hypothetical steps to generate olfactory neuron diversity. The yellow “plate” represents the antennal disc, in which each circle is
an epithelial cell. During pre-patterning phase, clusters of cells (in different colors) in different domains have different differentiation potentials. In the proneural phase, only one cell from each cluster becomes a sensory organ precursor (SOP, highlighted, also see the next section). During neurogenesis, Notch is used in each division. Some neurons may die in the neuronal sub-lineage.

1.4.3 Sensory organ precursors and proneural genes

Once pre-patterning is finished, neuronal progenitors are selected from ectodermal cells by the expression of proneural genes prior to neurogenesis. This strategy seems to be commonly applied in both the central and peripheral nervous systems (Bertrand et al., 2002). The progenitors in the PNS are called sensory organ precursors (SOPs), and the ones in the CNS are called neuroblasts (NBs). Both types of progenitors are multipotent cells that are committed to generate neurons through series of asymmetric divisions, which would be discussed in the next section. Like other PNS neurons, olfactory receptor neurons along with the non-neuronal cells in a sensillum arise from a single SOP. Therefore, knowing how SOPs are specified is a critical step before we can understand how neuronal diversification occurs.

Generally, proneural genes are basic helix-loop-helix (bHLH) transcription factors that confer SOP identities to quiescent ectodermal cells and thus the potential to become neurons. Proneural genes are structurally and functionally conserved across species (Bertrand et al., 2002). The first proneural genes to be identified belong to the Achaete-Scute Complex that specifies external sensory organs in flies. Atonal (ato) was later discovered to specify photoreceptor R8 and both embryonic and adult chordotonal
organs. Proneural genes function through forming a heterodimer with another bHLH protein, Daughterless (Da) to bind an E box in the target genes, often including themselves (Cabrera and Alonso, 1991; Jan and Jan, 1993; Jarman et al., 1993).

The proneural genes that govern olfactory SOP selections are amos and ato. In addition to promoting neuronal fates, they also define precursor identities: SOPs that express amos differentiate into basiconic or trichoid sensilla in the antennae, whereas the ato-expressing SOPs become coeloconica and basiconica in the antennae and maxillary palps, respectively (Goulding et al., 2000; Gupta and Rodrigues, 1997; Jhaveri et al., 2000; zur Lage et al., 2003). Amos begins to be expressed at larva-pupal transition, and continues to be expressed till 16hr after puparium formation (APF), peaking around 8hr APF. Ato is activated in the 3rd instar larval stage, peaks at 4hr APF, and remains highly expressed until 8hr APF, and by 16hr APF Ato is no longer expressed (Gupta and Rodrigues, 1997; zur Lage et al., 2003). In both cases, the expression begins in clusters of cells covered by a pre-patterned domain and is gradually resolved into singly expressed cells through Notch-mediated lateral inhibition. Loss-of-function in amos or ato causes the total elimination of corresponding sensilla, while over-expression induces ectopic sensilla. On the other hand, mis-regulation in Notch leads to supernumerary sensory organs (Bertrand et al., 2002).

Proneural genes are thought to be spatially regulated by combinations of patterning genes and negative factors such as Extramachrochaete (Gupta and Rodrigues,
1997; Jhaveri et al., 2000). lozenge (lz) is one of the very few pre-patterning genes that were identified to function early in olfactory SOPs. lz activates amos and is expressed in a concentration gradient, where in combination with amos, high levels of lz specify basiconic sensilla and low levels of lz direct tricoid sensilla differentiation (Goulding et al., 2000; Gupta et al., 1998). engrailed (en) and dachshund (dac) may fulfill the criteria of pre-patterning genes as well, as they are expressed in specific domains prior to proneural gene expression. Disruption of these two genes causes reductions in subsets of ORN classes (Song et al., 2012). However, what are other factors involved and how these factors are organized in a logical way to contribute to the rich diversity of sensilla subtypes are unknown.

1.4.4 Neurogenesis and Notch-mediated binary fate determination

As discussed in the previous section, the 1-4 ORNs and 3 support cells in a sensillum constitute a sensory unit that is developed from a single SOP. The process that generates these individual components involves at least 3 rounds of Notch-mediated asymmetric divisions. The first noticed division indicated by positive phosphohistone-3 labeling occurs at 12hr APF, which immediately follows the end of the SOP selection process (Sen et al., 2003). Neuronal proliferation is prominent between 16-22hr APF.

The use of Notch signaling in diversifying a neuronal precursor or any other stem cell-like precursor is one of the most fundamental mechanisms in developmental biology (Artavanis-Tsakonas et al., 1999; Cau and Blader, 2009; Hori et al., 2013). It has
been well studied as a mechanism for cell fate specification in the model of mechanosensory bristle development. Notch instructs binary fate decisions at each progenitor bifurcation point, generating only one sensory neuron and three support cells per sensillum.

Figure 8: Notch-mediated neurogenesis in the olfactory sensilla.

Three rounds of asymmetric divisions of a SOP are shown. The names of intermediate precursors and neurons are labeled. Divisions are in asynchrony between pIIa and pIIb lineages.
Here in the olfactory sensilla, Notch plays a similar role in separating alternate fate choices from a precursor, except that more repetitions of the same signaling pathway are applied to accommodate a more complex structure (Rodrigues and Hummel, 2008) (Figure 8). During each round of asymmetric divisions, the Notch antagonist, Numb, is preferentially allocated to one of the two daughter cells, and turns off Notch, while the partner cell receives Notch ligand and transduces Notch signaling to activate target genes expression. The two intermediate progenitors from the first mitosis are named pIIa (Notch ON) and pIIb (Notch OFF), which correspond to non-neuronal and neuronal lineages, respectively. Due to the asynchrony of later divisions between these two lineages, 3, 4, or 6-cell stages can be detected before the final point. The pIIb precursor gives rise to pNa (Notch ON) and pNb (Notch OFF) sub-lineages, which in turn generate Elav-positive terminally differentiated cells: pNaa (Notch ON), pNab (Notch OFF), pNba (Notch ON), and pNbb(Notch OFF). Selective cell death occurs in these daughters, and the cells that survive account for the final outcome of 1-4 neurons per sensillum. Descendants of pIIa express the transcription factor Cut and eventually become socket, shaft and sheath cells (Endo et al., 2007, 2011).

Interestingly, due to the reiterative use of the same signaling and target molecules, an epigenetic mechanism has been proposed to explain the switching of Notch status from a Notch ON parent to daughters with different identities (Endo et al., 2011). Specifically, the chromatin modifying protein, Hamlet is expressed only in pNa
sub-lineage and functions to induce epigenetic silencing of Notch targets to allow new assignment of Notch status prior to the next round of signaling event. This poses a classic example of efficient usage of limited genetic information to generate large quantities of cell types. Indeed, perturbations of components involved in the Notch pathway including Hamlet are associated with fate switches between lineages resulting in dramatic reductions in cellular diversity (Endo et al., 2007, 2011; Sen et al., 2003). However, since SOP identities seem to be irrelevant to Notch function, it is still unclear what the nature of differentiation competency is in SOPs on which Notch signaling acts to separate the “encoded” fates.

1.4.5 OR gene choice

Following neurogenesis in the mid-pupal stage, ORN axons make class specific connections with PN dendrites that have already patterned the targeting field (this process will be discussed in the next section), and then receptor gene expression marks the end of ORN differentiation. One most striking feature of the olfactory system is its functional specialization by singular expression of a receptor gene from a large repertoire. This is postulated as the “one neuron-one receptor” rule, and brought up the OR gene choice problem which has led to extensive inquiries in both vertebrate and invertebrate systems (Fuss and Ray, 2009). Although Drosophila appears to follow this rule less strictly compared to the mammalian counterparts, more than 70% ORN classes do only express a single receptor gene. As mentioned in previous sections, stereotyped
combinations of ORN classes are housed in distinct sensilla subtypes, which display unique zonal distributions. Thus, each olfactory receptor gene restricts its expression to a specific domain. This pre-defined gene expression pattern implies a deterministic mechanism in fly OR choice.

To explore this possibility, earlier research was primarily focused on the analysis of cis-regulatory elements in OR genes. Because a short fragment (~500bp) in the promoter region of an OR is both necessary and sufficient to drive appropriate expression in most cases, it is believed that all of the important regulatory elements lie within this small region. Initially, ORs expressed in maxillary palps (MPs) were chosen due to MPs’ simpler structure with fewer genes and sensilla subtypes. Two cis-regulatory motifs, Dyad1 and Oligo-1 were discovered, which were shown to positively regulate MP-specific OR expression in the MPs and negatively regulate these genes in the antenna, respectively (Ray et al., 2007). These organ-level regulatory motifs have led to the speculation of a hierarchical logic in determining which OR should be expressed in a certain neuron. Consistent with this hypothesis, more systematic comparisons of the same regulatory regions among 12 Drosophila species showed both positive and negative regulatory motifs in each MP OR promoter to ensure the proper OR is activated, and only activated in the right neuron (Ray et al., 2008). Similar conclusions have been reached from the investigation of a subset of antennal ORs, specifically the alleged pheromone ORs (Miller and Carlson, 2010).
On the other hand, the trans-acting factors that bind to these elements and are responsible for the intricate regulations have proved to be difficult to find. The POU-domain transcription factor acj6 is the first example that was shown to activate subsets of ORs both in the maxillary palps and antennae, and repress a larval OR in a MP ORN through a common motif upstream of these genes (Bai and Carlson, 2010; Bai et al., 2009; Clyne et al., 1999). Another POU-domain factor pdm3 is also expressed in some MP ORNs, but only required for Or42a expression in that ORN class, which also requires acj6. Interestingly, acj6 is alternatively spliced, and it seems that different isoforms have quite different, sometimes opposite functions in OR gene choice, suggesting that combinations of POU-domain factors, including their distinct isoforms may form a code for class-specific OR expression (Bai and Carlson, 2010; Tichy et al., 2008). Scalloped is another transcription factor that was discovered to function in a specific MP ORN class to repress the expression of the neighboring OR in the same sensilla (Ray et al., 2008).

To add more onto this list, a recent RNAi-based genetic screen uncovered 7 transcription factors that function in antennal OR expression (Jafari et al., 2012). Each factor is required for a subset of ORs, and two of them, E93 and Acj6 also seem to repress ectopic expression of specific ORs. Because of the prevalence of essential binding motifs for these factors in the proximal upstream of most OR genes, the authors concluded that the restricted expression of a given OR is the result of broad activation by combinations of transcription factors in the proximal promoter region accompanied
by partial repression through the distal regulatory elements. However, it is unclear why these broadly expressed genes only affect subsets of neurons and why sensory conversions were rarely seen when the code was manipulated.

Epigenetic modifications have also been shown to control olfactory receptor expression. Myb-MuvB (MMB)/dREAM complex has been reported to regulate CO2 receptor Gr63a/Gr21a expression (Sim et al., 2012). Interestingly, depending on the subunit composition of the complex, it can either activate receptor expression or repress inappropriate expression in other tissues. This cell-type specific modulation of the complex activity is also associated with histone modifications showing counteractive effects, further proving an epigenetic involvement of OR regulation. Since this study was only focused on one ORN class, it is curious to see to what level this type of regulation can be generalized to the rest of the system.

Based on the stereotyped organization of fly olfactory system and results of the mentioned studies, it is a general consensus that OR choice in flies is determined mainly by combinatorial regulation of positive and negative transcription factors. The conventional view believes that these factors primarily function in the late stage of ORN specification to turn on or off a certain receptor gene, thereby explaining OR choice. However, because ORN sensory identities are pre-defined by SOP identities and neurogenesis, the later transcription factor network responsible for class-specific OR expression must be coordinated with the early patterning networks. How these late
functioning genes link to the earlier stages of fate determination is unknown. Also, to what extent and by what mechanisms the early events could directly contribute to the final receptor expression remains unanswered.

In contrast to the deterministic nature of fly OR choice, the mammalian olfactory system seems to have adopted a more stochastic mechanism. As the number of mammalian ORs is an order of magnitude higher and therefore might have exceeded the threshold of what a deterministic system could handle, it is logical to favor such an alternative strategy. Indeed, epigenetic mechanisms were shown to play a major part in governing monogenic OR expression in mice. Extensive gene duplication, conversion and recombination have led to large OR gene clusters in mammalian genomes (Kambere and Lane, 2007). Surprisingly, they form heterochromatic macrodomains that contain constitutively repressive histone modification marks, and aggregate in OR-specific foci in the nucleus (Clowney et al., 2012; Magklara et al., 2011). Such biochemical modifications and nuclear organization ensure global silencing of ORs prior to the randomly selective expression of a receptor by a developmentally regulated opposite process. Because the histone demethylase LSD1 can be both a transcription co-activator and co-repressor, it is responsible for removing the repressive mark from an OR to turn on gene expression at the initial choice, and removing the active mark from that OR to turn it off if a second choice is needed (Lyons et al., 2013). Interestingly, the feedback signal for receptor switching involves OR-induced unfolded protein response (UPR) that
regulates LSD1 level (Dalton et al., 2013). Proper expression of a given OR reduces LSD1 via adenylyl cyclase 3 locking the choice, and releases UPR to restore global translation. Otherwise, sustained LSD1 triggers new choices until a proper OR is chosen.

This epigenetic model fits well with the monogenic and monoallelic aspect of mammalian OR expression. However, it is still unclear how a specific OR is chosen to be activated in the first place, and whether or not this process is totally random. The restriction of mammalian ORs into zones argues against a completely random scenario (Figure 2). Indeed, when a non-OR gene or a pseudogenized OR is expressed, a second choice seems to be only made from genes in the same zone or sometimes the same gene family (Fuss and Ray, 2009; Sato et al., 2007; Serizawa et al., 2003; Shykind et al., 2004). Therefore, it has been proposed that mammalian OR choice is a combination of a deterministic process that restricts each class into zones and a stochastic process that generates intermingled expression within each zone (Fuss and Ray, 2009). Consistent with this hypothesis, transcription factor binding sites have been found in the promoter regions of the more ancient OR class/family that are restricted to specific domains in the olfactory epithelium (Hoppe et al., 2003, 2006). Interestingly, some of these factors are evolutionarily conserved, and are involved in patterning Drosophila antennal field. It is curious to see if any of the mechanisms used in specifying fly ORN fates also plays a role in the zonal separation of mammalian ORs.
1.4.6 Wiring specificity

A critical aspect of neuronal identities is the precise connectivity with other neurons. Once ORNs are born, they need to make correct connections with cognate projection neurons (PNs) in a one-to-one manner. In mammals, ORs themselves play an important role in instructing ORN targeting (Takeuchi and Sakano, 2014). In contrast, fly ORs begin their expression after connections are already made, implicating a fundamentally different OR-independent process. Due to the high levels of wiring specificity, the fly olfactory system has been used as an ideal model to study genetic mechanisms of neurite pathfinding and neuronal class-specific recognition.

It is now well accepted that making stereotyped connections between specific ORN classes and PN classes in the antennal lobe involves at least three steps: (1) PN dendrites initially pattern the targeting field; (2) ORN axons independently find the correct regions to target; (3) PNs and ORNs recognize each other to make class-specific connections (Hong and Luo, 2014). Many aspects of these processes are mediated by secreted or membrane-bound cell-surface molecules, and sometimes the same molecules are used by both partners in multiple contexts.

**PN dendritic targeting**

PN dendrites enter antennal lobe at puparium formation, and finish patterning the field around 18hr APF, before pioneer antennal ORNs begin to invade the lobe (Jefferis and Hummel, 2006). During this process, the cell adhesion molecule N-cadherin
(N-cad) and Ig family cell surface molecule Dscam induce intra-class attraction and repulsion, respectively, to ensure proper regional arborization with enough coverage in a confined manner (Zhu and Luo, 2004; Zhu et al., 2006). To guide class-specific targeting, transcription factors Acj6 and Drifter were found to couple PN lineages specification and targeting identities, which in turn may regulate cell surface molecules, such as Semaphorins (Komiyama et al., 2003). Sema-1a is a cell-autonomous receptor that forms a concentration gradient among PNs, which instructs dendrite coarse targeting along the dorsolateral (high Sema-1a)-ventromedial (low Sema-1a) axis (Komiyama et al., 2007). Interestingly, the ligands for the receptor, Sema-2a and 2b are secreted by degenerating larval ORNs and newly born adult PNs. They distribute opposing concentration gradients to Sema-1a in the antennal lobe (Sweeney et al., 2011). Besides these global cues, leucine-rich repeat transmembrane proteins Capricious (Caps) and Tartan (Trn) are expressed in partially overlapping subsets of PN classes. They function more locally to separate PNs into categories based on gene activities so that neighboring dendrites are targeted to distinct glomeruli (Hong et al., 2009).

ORN axonal targeting

The earliest antennal ORN axons arrive in the antennal lobe around 18hr APF. This is within the peak of ORN neurogenesis, but prior to the onset of OR expression. As mentioned earlier, the Notch signaling pathway plays a major part in diversifying neuronal sensory identities in the periphery during neurogenesis. Here, it also instructs
axon targeting (Endo et al., 2007). Likewise, Acj6 and Pdm3 not only function in ORN sensory fate determination but also in axon pathfinding, suggesting cell fate and connectivity could be coupled by the same set of factors (Komiyama et al., 2004; Tichy et al., 2008). Indeed, a Hedgehog (Hh)-mediated two-step mechanism was reported to coordinate the ORN cell body positions in the antenna with the targeting selection in the brain. ORNs from anterior versus posterior compartments expressing different levels of Patched receptor respond differently to the CNS-derived Hh (Chou et al., 2010). Since imaginal disc patterning is highly stereotyped and closely linked to ORN fates, it is plausible to imagine that Hh signaling couples ORN specification and axon targeting.

As ORNs enter the lobe, they are separated into two broad ventromedial and dorsolateral pathways. Notch is responsible for restricting Sema-2b only to the ventromedially projecting ORNs, which in turn are attracted by extrinsic Sema-2a and 2b. Remarkably, they are the same molecules that are concentrated in the ventromedial region, and guide PN dendrite targeting. Furthermore, ORN axonal interactions mediated by Sema-2b/PlexB facilitate the consolidation of axon bundles in the ventromedial region to reinforce the targeting choice (Joo et al., 2013). These global cues determine the general pathways ORN axons would take before sending processes into nearby target regions. Other guidance molecules are also found to be involved in ORN axon initial targeting. For examples, mutations in Dscam, the downstream factors of
Dscam Pak and Dock, and Robo receptors all lead to disruptions of ORN coarse targeting (Ang et al., 2003; Hummel and Zipursky, 2004; Jhaveri et al., 2004). Once ORN axons are in the roughly correct region, interactions between them become very important. For example, Sema-1a and N-cad both participate in ORN axon sorting to ensure class-specific axon convergence, which is likely mediated by intra-class interactions (Hummel and Zipursky, 2004; Lattemann et al., 2007). In addition, the correct targeting of late arriving MP ORNs is under the influence of a repulsive force from early arriving antennal ORNs via Sema-1a/PlexinA signaling (Sweeney et al., 2007).

**ORN-PN matching**

The last step to establish the wiring specificity is to match ORN axons and PN dendrites according to their classes. Recent findings show that this process is governed by EGF transmembrane protein Teneurin-mediated homophilic interactions (Hong et al., 2012). Compatible levels of Ten-m and Ten-a in PN and ORN pairs guide class-specific matching. Conversely, disruption of their endogenous levels causes mismatches between classes. Since this study was based on selective pairs, more molecules await to be discovered to explain the specific matching between 50 classes of ORNs and PNs.


2.1 Summary of the study

Background

Sensory neuron diversity ensures optimal detection of the external world and is a hallmark of sensory systems. An extreme example is the olfactory system, as individual olfactory receptor neurons (ORNs) adopt unique sensory identities by typically expressing a single receptor gene from a large genomic repertoire. In Drosophila, about 50 different ORN classes are generated from a field of precursor cells, giving rise to spatially-restricted and distinct clusters of ORNs on the olfactory appendages. Developmental strategies spawning ORN diversity from an initially homogeneous population of precursors are largely unknown.

Results

Here we unravel the nested and binary logic of the combinatorial code that patterns the decision landscape of precursor states underlying ORN diversity in the Drosophila olfactory system. Transcription factor Rotund (Rn) is a critical component of this code that is expressed in a subset of ORN precursors. Addition of Rn to pre-existing
transcription factors that assign zonal identities to precursors on the antenna, subdivides each zone and almost exponentially increases ORN diversity by branching off novel precursor fates from default ones within each zone. In \textit{rn} mutants, \textit{rn}-positive ORN classes are converted to \textit{rn}-negative ones in a zone-specific manner.

\section*{Conclusions}

We provide a model describing how nested and binary changes in combinations of transcription factors could coordinate and pattern a large amount of distinct precursor identities within a population to modulate the level of ORN diversity during development and evolution.

\section*{2.2 Results}

\subsection*{2.2.1 Rn is expressed in a subset of antennal ORN precursors}

To understand the regulation of olfactory receptor expression and ORN diversification, we analyzed two previously isolated mutants, \textit{tod}^{Z3-6129} and \textit{tot}^{Z3-4626} (Ha and Smith, 2006). In both alleles, the antennal odorant receptor Or67d expressed by ORNs singly housed in the trichoid subtype at1 sensilla is absent and the DA1 glomerulus targeted by these receptor neurons in the antennal lobe is missing (Figure 9A). Sequence information revealed that both \textit{tod}^{Z3-6129} and \textit{tot}^{Z3-4626} mutants have lesions in the \textit{rotund (rn)} gene, a Krüppel-like zinc finger transcription factor with several putative splice variants, the major three of which are called E, F and C isoforms (St Pierre et al., 2002) (Figure 9B). All three isoforms can be detected in the developing
pupal antennae by RT-PCR. The tod^{Z3-6129} allele harbors a lesion predicted to change the conserved arginine (R565) to glutamine (Q) in the shared zinc finger domain without lowering the overall RNA levels. The tot^{Z3-4626} allele contains a 37 nucleotide insertion 63 base pairs upstream of the C-isoform start codon, leading to significant reductions in F and C isoform transcripts (Figure 16A). We also used a previously reported hypomorphic allele, rn^{89GAL4} in which the inserted GAL4 disrupts the gene function and recapitulates the endogenous gene expression (St Pierre et al., 2002). This GAL4 line was shown to be a faithful expression reporter for rn in the antennal disc by RNA *in situ* hybridization, but not for the *roughened eye* (roe) gene, which is a part of the same gene locus but expresses a different transcript (St Pierre et al., 2002). Trans-heterozygous tod^{Z3-6129/rn^{89GAL4}}, tot^{Z3-4626/rn^{89GAL4}} and tod^{Z3-6129/tot^{Z3-4626}} all produce the same phenotype as tod^{Z3-6129} and tot^{Z3-4626} homozygous mutants (Figure 16B, Figure 16C). Therefore, we have identified tod^{Z3-6129} and tot^{Z3-4626} as two new alleles of rn, and we will hereafter refer to them as *rn*^tod^ and *rn*^tot^.

To understand the role of Rn in the specification of at1 ORNs, we analyzed the Rn expression pattern in developing antennae using *rn*^{89GAL4} enhancer trap driven GFP expression. In third instar larval antennal discs, a GFP signal was detected in the central ring, representing precursors of ORNs in the olfactory segment of antennae (zur Lage et al., 2003) (Figure 10A, Figure 17A). At 35 hours after puparium formation (APF), *rn* expression occupies the mediolateral region of the developing antennae. At 45 hours
APF, *rn* is expressed both in neurons and non-neuronal cells. *rn* expression cannot be detected in the adult olfactory appendages.

Using a FLP/FRT-mediated lineage tracing system (Experimental procedures 2.4.1), we identified 18 *rn*-positive ORN classes in the adult olfactory system clustered within 8 distinct antennal sensilla subtypes (Figure 10B, left in graph and top in table, Figure 17B). Subsets of all three major types of sensilla were found to be *rn*-positive. Among these subtypes, two belong to trichoid sensilla (at1 and at3), three are basiconic (thin: ab5; small: ab7, ab10), two are coeloconic (ac1 and ac4), and one is the rare intermediate type sensilla (ai1) (Benton et al., 2009; Couto et al., 2005; Endo et al., 2007; Fishilevich and Vosshall, 2005; Silbering et al., 2011). None of the maxillary palp ORNs expresses *rn* (Figure 17B).

### 2.2.2 Rn branches off novel ORN precursor fates from default ones in trichoid sensilla

Within the trichoid sensilla subtypes, *rn* is expressed in at1 and at3 lineage precursors, but not in at4 or at2 (Figure 10B, left in graph and top in table). Since at1 ORNs are abolished in *rn* mutants, we asked whether a similar effect is observed in at3 sensilla. at3 sensilla house three ORN classes that express Or2a, Or19a/b and Or43a (Couto et al., 2005; Fishilevich and Vosshall, 2005). We analyzed OR expression by at3 ORNs and found that the expression of all three OR genes and their target glomeruli were absent in *rn* mutants (Figure 11A). These results suggest that Rn regulates the specification of *rn*-positive ORNs in trichoid sensilla.
The loss of ORNs in at1 and at3 in rn mutants could arise either from the loss of these sensilla subtypes by cell death or from their conversion to other subtype(s). We tested for conversion to the at4 sensilla subtype because we found significant increases in the sizes of glomeruli innervated by rn-negative at4 ORNs in rn mutants (Figure 11B bottom right). at4 sensilla house the cell bodies of Or47b-, Or88a- and Or65a-expressing ORNs which target VA1v, VA1d, and DL3 glomeruli in the antennal lobe, respectively (Couto et al., 2005; Fishilevich and Vosshall, 2005). In rn mutants, we found that Or47b, Or88a and Or65a expression expands to medial zones within the antenna that are normally occupied by at1 and at3 sensilla in a cell-autonomous manner (Figure 11B, top and Figure 16C, Figure 18A, Figure 16D). Double-labeling of Or47b and Or88a or Or65a cell bodies showed that the OR pairings in at4 were retained in the expanded antennal zones in rn mutants (Figure 11B, bottom left and Figure 18A). The increases in the numbers of at4 ORNs in rn mutants strongly suggest that rn-positive at1 and at3 subtypes are converted to rn-negative at4 sensilla. If this is true, the converted Or47b ORN population should overlap with rn\textsuperscript{89GAL4}-driven reporter expression in rn mutants. Indeed, two classes of Or47b neurons were observed in the developing antenna of rn mutants: Or47b ORNs that are in the endogeneous at4 sensilla negative for the rn expression reporter, and the Or47b neurons positive for the rn reporter that are converted from at1 and at3 sensilla (Figure 11C). We also used the lineage-tracing method described above to analyze rn-positive ORN projections in rn mutant antennal
lobes. In agreement with the conversion model, we observed the innervation of VA1v, VA1d, and DL3 glomeruli targeted by at4 ORNs, which are normally $rn$-negative in the wild type (Figure 10B). OR expression in at2 ORNs is unaffected (Figure 11D). These results suggest that Rn diverges new trichoid ORN identities (at1 and at3) from a “default” at4 identity (Figure 11E). Overexpression of UAS-$rn$ transgene (St Pierre et al., 2002) in all trichoid and basiconic lineage precursors using amos-GAL4, leads to an increase of $rn$-positive at3 sensilla ORNs and a decrease in the $rn$-negative at4 ORNs (Figure 16E). However, we also detected a decrease in $rn$-positive at1 ORNs. Even though this might suggest that ectopic at3 fates can be induced by $rn$ overexpression, the decrease in at4 (and at1) might also arise from the toxicity of high levels of Rn as expressing this transgene in $rn$-positive lineages in the wild type background leads to a loss of $rn$-positive ORNs, and it is unable to rescue the OR expression phenotype in the mutant flies. (Figure 16B’).

2.2.3 Rn branches off novel ORN precursor fates from default ones in basiconic and coeloconic sensilla

Next we extended our study to non-trichoid sensilla, and found that $rn$ plays a similar role in specifying ORNs in these sensilla. In both coeloconic and basiconic cases, absent $rn$-positive sensilla were converted to certain default $rn$-negative sensilla subtype identities in a lineage-specific manner, as was confirmed by the lineage-tracing experiment in $rn$ mutants (Figure 12A, B, Figure 18, Figure 19 and Figure 10B). For the coeloconic subtypes, we observed the expansion of the default ac2 sensilla into antennal
zones that are normally occupied by ac1 and ac4 in *rn* mutants (Figure 12A, Figure 18C-F). Similar conversion effects were seen for basiconics within and among their morphological subdivisons in the absence of *rn* (Figure 12B, Figure 19). Intermediate sensilla have only one *rn*-positive subtype ai1, which is lost in *rn* mutants (Figure 11B). RT-PCR was also performed on selective OR genes to confirm our genetic findings (Figure 18G). Taken together, these results suggest that Rn is responsible for the divergence of novel ORN precursor identities from default sensilla subtype precursors (Figure 12C, D and Figure 11E).

### 2.2.4 Lineage specification by nesting the transcription factor functions drives precursor cell diversification

Since Rn expression is early, and partially overlaps with the pattern of the proneural gene *amos* and the SOP marker *senseless* (*sens*) (Figure 10A, Figure 17A), we reasoned that the conversions in *rn* mutants occur due to the alterations in lineage-specific SOP identities and the corresponding combinations of olfactory receptors that have the potential to be expressed in that lineage. Indeed, the development of both endogenous and ectopic at4 ORNs in *rn* mutants is *amos*-dependent, as *rn* and *amos* double mutants lack all basiconic and trichoid sensilla, including at4 (Figure 13). Furthermore, *rn* and *lozenge* (*lz*) double mutants show subtype conversion of at1 to at4 in both the endogenous trichoid zone and the duplicated ectopic zone, which is converted from basiconic to trichoid sensilla due to *lz* mutation (Gupta et al., 1998) (Figure 13, Figure 20). These results support lineage-specific precursor fate conversions in *rn*.
mutants, and reveal the bifurcating and nested nature of fate diversification among precursor cells within lineages.

2.2.5 A common motif upstream of rn-positive OR/IR genes is required for their expression

We predicted that the transcription factor Rn regulates the olfactory receptor gene combinations allowed in a precursor lineage, either directly or indirectly, by regulating the status of the promoters of olfactory receptor genes. We used de novo motif finding approach to search for a consensus sequence shared by the promoters of OR genes expressed in rn-positive but not rn-negative lineages (Experimental procedures 2.4.4). The top-scoring motif (M1) from three independent programs (MEME, PRIORITY and WEEDER) (Bailey et al., 2006; Narlikar et al., 2006; Pavesi et al., 2004) is [TG] GGTTGGGA [GAT] [AG] (Figure 14A). We confirmed the presence of this motif in rn-positive ORs/IRs and absence of it in rn-negative ones using known motif searching tool MAST (Figure 21A) (Bailey and Gribskov, 1998). To test if M1 is required for rn-positive OR expression, we generated Or49aAML-GAL4 and Or82aAML-GAL4 promoter reporter lines in which M1 is replaced by a random sequence. We found significant decreases in the cells labeled by these two transgenic reporter lines, suggesting the involvement of M1 motif in the activation of Or49a and Or82a expression (Figure 14B). We speculated that Rn may interact with this motif in a given precursor to determine the allowable receptor genes that can be expressed in ORNs from that lineage. However, we could not detect the binding of Rn to these promoters using yeast one-hybrid (Y1H) assay, even
though direct interactions with variable strengths between known DNA-binding proteins and corresponding DNA elements were seen as expected (Deplancke et al., 2006; Pruneda-Paz et al., 2009) (Figure 21B). A recent study used electrophoretic mobility-shift assay with in vitro translated Rn to demonstrate the binding of Rn to motif T13 that appears to be different from the M1 motif (Baanannou et al., 2013). We were unable to detect the binding of Rn to T13 in Y1H assay (Figure 21B). In addition, a sequence (856bp upstream of the Or49a start codon) almost identical to T13 is present in the Or49a promoter, which was not bound by Rn in our Y1H assay. It is possible that yeast lacks proteins required for the proper folding and processing of Rn or co-factors that facilitate the binding of Rn to this sequence. Further analysis of upstream regulatory sequences of OR/IR genes did not produce any correlation between the presence of T13 and the lineages of OR/IR genes (data not shown). These results suggest that Rn may not bind to OR/IR regulatory elements, but perhaps regulates or interacts with other factors to determine sensilla-specific combinations of OR expression.

### 2.2.6 A decision landscape for antennal ORN precursor diversification

How can the complex landscape of different ORN fates be carved from a homogeneous plane of precursors? The systematic analysis of *rn* mutant phenotype in ORN diversity suggests that Rn functions to increase the diversity from each lineage, but generates different ORN outcomes based on the historic and molecular contingency of the precursors. Our results along with previously identified factors expressed in subsets
of precursors (Goulding et al., 2000; Gupta and Rodrigues, 1997; Gupta et al., 1998; zur Lage et al., 2003; Song et al., 2012), suggest that the final patterning of antennal olfactory receptor choice is an outcome of combinations of transcription factors functioning in a nested fashion to bifurcate and diversify precursor cells identities (Figure 15A). These combinations encode sensilla type and subtype identities to determine olfactory receptor pairings allowed in a given precursor lineage by exerting their effects, directly or indirectly, on olfactory receptor genes. We found that Rn is a critical component of this nested control mechanism to diverge rn-positive sensilla subtypes from otherwise default rn-negative ones in each lineage-specific precursor, leading to a dramatic increase of ORN classes in the olfactory system. engrailed (en) and dachshund (dac) are two other factors with restricted and non-overlapping expression in a proportion of sensilla subtype precursors and adult ORNs, and their mutations are associated with a lineage-specific decrease in the expression of ORs (Song et al., 2012). Mapping en and dac onto the lineage tree based on their expression pattern showed that both genes act in a binary fashion similar to rn to diversify specific precursor lineages (Figure 15A). However, no fate conversion event was reported in these mutants, suggesting different functional modes may be adopted by en and dac. Particularly, we found en-negative ac4 or at1 ORNs was not increased in concomitant with the loss of neighboring en-positive lineages in en mutants (Figure 22A).
In order to simulate combinatorial rules giving rise to the diverse population of SOP identities, we generated a toy Waddington model describing the diversification landscape of early ORN precursor fates prior to the onset of asymmetric divisions by nesting the regulatory relationships between the transcription factors. This allowed us to create a decision landscape that can be modified by changing the state of each transcription factor in precursors from the matrix of transcription factors listed in the order of their temporal expression in the antennal disc. Therefore, a smooth plane of precursors can be modified into more complex landscapes where the addition of each factor creates a dip from a prior state that defines a new precursor fate (Figure 15B, Experimental procedures 2.4.7 and Appendix). This computational simulation not only allows us to visualize the outcome of ORN precursor diversification process as a whole, but also leads to predictions about the decision landscapes for different mutant conditions (single or double mutant combinations) through modifying input parameters in the transcription factor matrix (Figure 15B, Figure 13 and Figure 20). The model successfully reproduces ORN diversity landscapes of some experimentally testable single and double mutants (Figure 13, Figure 15, and Figure 22B). Systems level dynamic analysis of olfactory receptor gene regulation during development in the future will enable us to more thoroughly explore the decision cascade of ORNs from a diverse array of precursors to refine this model.
2.3 Discussion

Neuronal diversity is a common characteristic of all sensory systems throughout the animal kingdom. Among these, the olfactory system demonstrates an extreme case in its diversity of ORN classes. In Drosophila, each of the 50 adult ORN classes is defined by the unique expression of typically a single olfactory receptor from a pool of around 80 genes (Benton et al., 2009; Couto et al., 2005; Fishilevich and Vosshall, 2005). How this ORN diversity is generated from a field of homogeneous precursor cells during development remains elusive. Combinatorial control of transcription factors has been proposed as an important mechanism that complex systems utilize to create cellular diversity (Baumgardt et al., 2007; Chlon and Crispino, 2012; Puri and Mercola, 2012). Here we demonstrate the nested and binary combinatorial rules by which transcription factors interact with each other to guide decisions regarding ORN precursor identities. Our results suggest that nesting the regulatory relationship of transcription factor combinations allows the concurrent use of the same factors in parallel lineages to generate ORN diversity in a very efficient manner. Under this logic, binary lineage choices in precursor cells are made based on historical contingency, which could serve as an effective strategy for establishing cellular complexity in many other developing systems.

In both vertebrates and invertebrates, each ORN class is spatially restricted to specific zones within the peripheral olfactory organs. In Drosophila, antennal ORNs are
housed in 3 morphologically and topographically different types of sensilla occupying distinct zones, while maxillary palps have only a single type of sensilla (Couto et al., 2005; Fishilevich and Vosshall, 2005). Each of the sensilla type zones on the antenna are subdivided into subzones that are defined by sensilla subtypes, which have similar morphology but differ in the set of olfactory receptors expressed in the ORNs they house. It has been shown that the decision for a given palp-specific olfactory receptor gene to be expressed in maxillary palp ORNs but not in antennal ORNs requires both positive and negative regulatory elements around that gene (Bai and Carlson, 2010; Ray et al., 2008). For antennal ORNs, the proneural genes, amos and atonal, and the prepatterning gene lozenge were found to assign sensilla type identities to the precursors and determine olfactory receptors expressed by the neurons housed in these sensilla (Figure 15) (Goulding et al., 2000; Gupta and Rodrigues, 1997; zur Lage et al., 2003). The loss of amos or atonal leads to the complete loss of basiconic and trichoid or coeloconic sensilla type(s), respectively, and corresponding ORNs. lozenge diversifies sensilla type identities within the amos lineage, where high levels of Lozenge are associated with basiconic sensilla fates, versus low levels of Lozenge generates ORNs in trichoid sensilla (Bhalerao et al., 2003; Gupta et al., 1998). Hypomorphic alleles of lozenge result in basiconic to trichoid sensilla type conversions. lz is also required for the expression of amos, suggesting the existence of regulatory loops among transcription factors in the same network (Goulding et al., 2000). Our results explained following sensilla type
specification in the antenna how the next level of diversification occurs. Rn is expressed in a subset of antennal sensilla precursors, and splits precursors of each zone into \textit{rn}-positive and \textit{rn}-negative subtypes. In \textit{rn} mutants, ORN diversity decreases almost by half as ORN classes from \textit{rn}-positive subtypes are switched to \textit{rn}-negative identities within the same zone. Our results suggest that Rn is required to branch off novel precursor identities from default ones resulting in the generation of new ORN classes in a zone-specific manner. It should be noted that some \textit{rn}-negative sensilla subtypes, for example, at2 and ac3, neither decrease nor increase in their numbers in \textit{rn} mutants, suggesting that there are additional factors driving the diversification of the ORN classes in these sensilla. Similarly, further diversification of \textit{rn}-positive ORN precursors should also be under the control of additional factors, such as En, operating in concert with Rn function.

Our results along with others suggest a two-step mechanism for ORN diversification: (1) successive restrictions on precursor differentiation potentials by spatio-temporal factors, such as proneural/pre-patterning genes and \textit{rn}; (2) segregation of restricted fates through Notch-mediated asymmetric divisions and local transcription factor networks for directly turning on olfactory receptor expression. Hypothetically, the sensilla precursor differentiation potentials can be represented by distinct sets of olfactory receptor genes being organized into euchromatic regions in a lineage-specific manner. The aforementioned combinations of transcription factors may influence the
dynamics of such epigenetic states, resulting in limited combinations of receptors transcriptionally accessible for later stages of ORN differentiation. Examples of chromatin modulation in OR expression has been shown in both flies and mice (Clowney et al., 2012; Magklara et al., 2011; Sim et al., 2012). Once precursor potentials are set, Notch signaling pathway could continue to bifurcate alternate sensory identities into ORNs generated through asymmetric precursor cell divisions (Endo et al., 2007, 2011). Transcription factor networks expressed later in development, including the well-characterized Acj6, Pdm3 and Scalloped could then directly regulate olfactory receptor expression during these divisions based on their genomic accessibility, giving rise to terminally differentiated ORNs (Bai and Carlson, 2010; Clyne et al., 1999; Jafari et al., 2012; Ray et al., 2008; Tichy et al., 2008).

In comparison with the *Drosophila* olfactory system, mammals exhibit remarkable organizational similarities in the olfactory circuitry, even though the numerical complexity far exceeds their insect counterparts. For example, the zonal pattern of olfactory receptor expression in the mammalian olfactory epithelium is analogous to the topographic segregation of sensilla type-dependent olfactory receptor expression in the antennae (Fuss and Ray, 2009). A number of transcription factors were reported to regulate the zone-specific expression of a subset of olfactory receptors, yet no mutants resulting in ORN sensory conversion have been described (Hirota et al., 2007; Kolterud et al., 2004; McIntyre et al., 2008). Despite the consensus on the stochastic nature of
mammalian olfactory receptor expression within each zone, it would be interesting to see whether the zones are defined by a similar developmental strategy.

The model we present here also provides us with an ancestral precursor decision landscape that reveals the interaction pattern among factors to maintain and modify phenotypic complexity/diversity within sensory neural circuits in evolutionary timescales. New regulatory nodes might be added to the combinatorial code at distinct stages of precursor cell development to change ORN specification programs. For example, addition of a mir-279, a negative regulator of the transcription factor nerfin-1 expressed in maxillary palp ORN precursors, results in the elimination of CO2-sensory ORNs from specific maxillary palp sensilla (Cayirlioglu et al., 2008). Furthermore, olfactory receptor genes have been shown to be fast evolving across and within genomes (Kambere and Lane, 2007; Sánchez-Gracia et al., 2009). Incremental addition of individual regulatory modules to pre-existing lineage-specific combinations operating in binary ON/OFF mode could facilitate the coordination of novel ORN fates with the evolution of receptor genes, which can be modified in response to changes in the quantity, quality, and context of the olfactory environment.
Figure 9: Molecular characterization of rn mutants.

(A) Loss of Or67d-expressing ORNs in rn mutants. Z-projections of confocal stacks for antennae and antennal lobes are shown in top and bottom panels, respectively. Anti-Bruchpilot staining lights up the antennal lobe neuropil (red). Or67dGAL4 mediated CD8 GFP expression is in green. Graph shows the quantification of cell body counts. n=10-15, *** P<0.001. Error bars indicate SEM. (B) rn genomic locus describing different splice isoforms and mutant alleles used in our studies. See also Figure 16.
Figure 10: Rotund is expressed in subsets of antennal SOPs.

(A) *rn* expression in the developing olfactory system revealed by *rn*^{89GAL4}-driven UAS-CD8 GFP (green). Senseless (red) and Amos (blue) antibody staining on 0 hr APF antennal disc. Arrowheads point to *rn*-positive ORN cell bodies in the 35 and 45 hrs APF antennal images, and their targeting in the pupal antennal lobe. Arrows indicate *rn*-negative cell bodies. Elav (red) labels neurons. N-Cadherin (red) antibody labels the pupal antennal lobe neuropil. The middle panels in 35 and 45 hrs APF are single confocal sections; all others are Z-projections. Scale bars in all images are 10mm except for the bright field antennal disc image, which is 50mm. (B) Lineage-tracing using tub-FRT-nYFP-FRT-CD2 RFP flip-out cassette driven by *rn*^{89GAL4} UAS-FLP, labels *rn*-positive ORNs (red) in wild type (left) and in *rn* mutant (right) adult antennal lobes. Only anterior and posterior confocal slices from female brains are shown. N-Cadherin is in blue. Red glomeruli in the antennal lobe schemes represent target sites for *rn*-positive ORNs. Posterior glomeruli are outlined with dotted lines. The table shows the list of *rn*-positive ORNs and corresponding sensilla types/subtypes in wild type (top) and *rn* mutants (bottom). Question marks indicate unmapped *rn*-positive glomeruli. See also Figure 17.

<table>
<thead>
<tr>
<th>Sensilla type</th>
<th>Subtype</th>
<th>Rotund reporter positive ORNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>trichoid</td>
<td>at1</td>
<td>Or67d</td>
</tr>
<tr>
<td></td>
<td>at3</td>
<td>Or2a, Or19a/b, Or43a</td>
</tr>
<tr>
<td>basiconic</td>
<td>ab5 (thin)</td>
<td>Or82a, Or47a/33b</td>
</tr>
<tr>
<td></td>
<td>ab7 (small)</td>
<td>Or98a, Or67c</td>
</tr>
<tr>
<td></td>
<td>ab10 (small)</td>
<td>Or49a/Or85f, Or67a</td>
</tr>
<tr>
<td>coeloconic</td>
<td>ac1</td>
<td>IR31a, IR75d, IR92a/IR76b</td>
</tr>
<tr>
<td></td>
<td>ac4</td>
<td>IR84a, IR75d, IR78a/IR76b</td>
</tr>
<tr>
<td>intermediate</td>
<td>ai1</td>
<td>Or13a, 7 (V37m)</td>
</tr>
<tr>
<td>trichoid</td>
<td>at4</td>
<td>Or47b, Or98a, Or65a/b/c</td>
</tr>
<tr>
<td>basiconic</td>
<td>ab1^* (large)</td>
<td>Gr21a, Or92a, Or10a/Gr10a, Or42b</td>
</tr>
<tr>
<td></td>
<td>ab6^* (small)</td>
<td>Or67b, Or69a/A/B</td>
</tr>
<tr>
<td>coeloconic</td>
<td>ac2</td>
<td>IR75a, IR75d, IR41a/IR76b</td>
</tr>
</tbody>
</table>

*Only seen in female mutants.
Figure 11: Diversification of new trichoid precursor/ORN identities from default rn-negative ones by Rn.

(A) Loss of OR expression in at3 sensilla in rn mutants. Data is presented similarly as in Figure 9A. (B) at4 OR expression is expanded to new antennal territories in rn mutants (top). Double labeling of Or88a and Or47b ORNs (bottom left) shows correct at4 OR pairings in expanded zones. Or47b, Or88a and Or65a glomerular volumes are increased in rn mutants. Quantification of Or47b volume is shown with Gr21a glomerulus as the reference (bottom right). Or47b, Or47a and Gr21a glomeruli were labeled by synaptotagmin GFP transgenes on the same chromosome (Cayirlioglu et al., 2008).
Selected glomeruli were outlined. Note that Or47a glomerulus is absent in the mutant (see below). n=10-12. *** P<0.001. (C) rn8GAL4 UAS-Redstinger (red) expression normally is excluded from Or47b ORNs (green) in at4 sensilla. In rn mutants, an additional population of Or47b ORNs that is labeled by rn reporter is apparent, indicating sensory conversion from at1 and at3 to at4. Single confocal sections are shown. Bottom panels show higher magnification. (D) Or23a expression in rn-negative at2 trichoid sensilla ORNs is unaffected in rn mutants. (E) Schematic summary depicting conversion of at1 and at3 ORNs to an at4 sensilla identity on the left, and lineage tree for trichoid sensilla subtype precursor diversification on the right. The purple square designates rn-positive precursor identities diverged from the default rn-negative at4. See also Figure 16 and Figure 18A, Figure 18G.
Figure 12: Diversification of new precursor/ORN identities from default rn-negative ones in coeloconic and basiconic lineages by Rn.
(A) Loss of IR expression in \textit{rn}-positive coeloconic ORNs in mutants (left). An ectopic population of \textit{rn}-negative ac2 ORNs expands to new antennal territories. IR41a glomerulus appears larger in \textit{rn} mutants (right). (B) Loss of OR expression in \textit{rn}-positive basiconic ORNs in mutants (left). \textit{rn}-negative ORNs from ab1 and ab9 sensilla show a slight increase in female mutants (right, images were from female flies). In (A) and (B), data is presented similarly as in Figure 9A. (C) and (D) Lineage trees for coeloconic and basiconic sensilla subtype precursor diversification, respectively. The purple squares designate \textit{rn}-positive precursor identities diverged from the default \textit{rn}-negative ones. For basiconics, we think that \textit{rn}-positive ab5 (thin basiconic) diverged from the default \textit{rn}-negative ab1 (large basiconic) identity across morphological subdivisions. At this point developmental markers to confirm this diversification pattern are unavailable. See also Figure 18 and Figure 19.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{The effects of \textit{rn}, \textit{lz} and \textit{amos} single mutants, as well as \textit{rn}/\textit{lz} and \textit{rn}/\textit{amos} double mutants on ORN specification.}
\end{figure}

\textit{rn} and \textit{lz} incrementally affect the number of Or47b neurons in at4 sensilla which are dependent on \textit{amos}. \textit{lz} functions similarly in at1 and at4 as both Or67d and Or47d neurons expand to new antennal zones in \textit{lz} single mutants. The specification of Or67d neurons also requires \textit{rn}. For antennal lobes, Or47b glomerular sizes: \textit{rn/-} /\textit{lz/-} > \textit{rn/-} or
lz/- > WT. Or67d glomerulus in lz/- is bigger than that in WT. In amos-/- and rn-/- amos-/-, many glomeruli in the anterior region of antennal lobe are missing, including Or47b. In rn-/- and rn-/- lz/-, Or67d glomerulus is missing. Graph shows the quantification of cell body counts. n= 10-15. Differences between samples from different genotypes are all statistically significant except for the labeled pairs. P<0.001. Error bars indicate SEM. See also Figure 20.

Figure 14: Requirement of M1 upstream of rn-positive olfactory receptor genes for appropriate OR expression.

(A) M1 consensus motif upstream of rn-positive ORs discovered by de novo motif finding approach (MEME). Axis indicates number of base pairs upstream from TSS of each OR gene, and red boxes indicate locations of motif occurrences. Boxes above line indicate that original version of motif was found, boxes below line indicate that reverse complement was found. (B) Replacement of M1 motif with a random sequence in rn-positive Or49a or Or82a promoter GAL4 reporter transgenes dramatically decreases their antennal expression. Four independent lines (in lighter shades) were randomly chosen for each gender in both Or49a and Or82a promoter GAL4 transgenes. n= 10-15, *** P<0.001; * P< 0.05. Error bars indicate SEM. See also Figure 21.
Figure 15: Combinatorial modulation of ORN precursor specification landscape.

(A) Representation of the transcription factors driving ORN precursor diversification. Diamonds represent transcription factors: en (engrailed), rn (rotund), ato (atonal), amos, lz (lozenge), dac (dachshund). For lz, “++” indicates a higher level of expression than “+”. Squares denote precursor differentiation potentials based on sensilla type/subtype relationships obtained from our results. Sensilla identities determined by lineage specific
transcription factors are shown in color, while the factors differentiating the precursors in black are unaccounted for. Addition of Rn to already existing combinations of transcription factors in each lineage hierarchically diversifies new precursor identities from default ones. (B) A toy Waddington model generated in Matlab (see 2.4.7 Experimental procedures and Appendix) automating the branching relationships in (A) to simulate the diversification of ORN precursor landscape. Simulation for rn mutation reshapes the landscape as expected decreasing the complexity of the landscape (lower image). Each dip/layer designates the emergence of a new precursor identity from existing states generated by the addition of a new factor. The novel ORN pairings depend on what other factors are present in the combinatorial code. Color-coding is the same as in (A). See also Figure 22.
Figure 16: Molecular and phenotypic analyses in different m Allelic combinations, MARCM clones and overexpression lines.
(A) RT-PCR analysis of \textit{rn} isoforms from wild type and \textit{rn} homozygous larval samples. (B) Or67d expression is lost in trans-heterozygous mutants. (B’) Expressing UAS-rn using \textit{rn}^{89GAL4} does not rescue the loss of Or67d expression in the mutants and causes a dominant phenotype in \textit{rn} heterozygous background. (C) Or47b expression is expanded to the medial zone in different alleles and allelic combinations. (D) Rotund functions in a cell-autonomous manner. \textit{rn} mutant MARCM clones phenocopy OR expression in genetic mutants. In reverse MARCM analysis, OR expression in wild type siblings of \textit{rn} homozygous mutant clones appear normal. (E) Overexpression of \textit{rn} by amos-GAL4 driven UAS-rn in all basiconic and trichoid precursor lineages causes decreases in the numbers of at4 (represented by Or47b) and at1 (represented by Or67d) ORNs, and an increase in the number of at3 (represented by Or2a) ORNs. The at3 domain in the overexpression line seems to expand to more lateral and medial zones in the antenna. The sizes of glomeruli targeted by corresponding ORNs change in accordance with the numbers of cell bodies. Graphs show the quantification of cell body counts. \( n = 10-15 \), ** \( P<0.01 \), and *** \( P<0.001 \). Error bars indicate SEM. For Or2a, only female quantification is shown due to the extremely weak signal intensity in male flies. For each OR, Z-projections of confocal stacks for antennae and antennal lobes are shown in top and bottom panels, respectively. Anti-Bruchpilot staining is used to visualize the antennal lobe neuropil (red). OR expression is in green.
Figure 17: Rn expression pattern.

(A) Developmental analysis of \textit{rn} expression (green) in early pupal antennal discs labeled with anti-Amos (blue) and anti-Senseless antibodies (red). \textit{rn}, Amos and Senseless expression partially overlap. Amos-positive staining indicates trichoid and basiconic sensilla precursors, and Senseless staining (Sens) signifies sensory organ precursors. Time points indicate hours after puparium formation. (B) Z projections of confocal stacks of antennae and maxillary palps from lineage-tracing experiments. \textit{rn}-positive cells were found in the medial lateral region of antennae both in the control and mutant flies. No \textit{rn}-positive cells were seen in maxillary palps. Images are from female flies.
Figure 18: OR expression in trichoid, coeloconic and intermediate sensilla showing sensory conversions in rn mutants.
(A) Top, double labeling of Or47b and Or65a cell bodies shows the pairing between these two ORNs in wild type and mutants. All the cells that are labeled with the Or65a reporter are paired with Or47b ORNs in both wild type and mutant antennae. Bottom, double labeling of Or47b and Or65a axon terminals shows that Or65a glomerulus is immediately adjacent to Or47b and Or88a glomeruli in the mutants, while they are separated by Or67d glomerulus in the control flies. (B) Effect of rn on intermediate sensilla ORNs. Or13a expression in rn-positive intermediate sensilla is lost in rn mutants. Z-projections of confocal stacks for antennae and antennal lobes are shown in top and bottom panels, respectively. Anti-Bruchpilot staining is used to visualize the antennal lobe neuropil (red). OR expression is shown in green. Graph shows the quantification of cell body counts. n= 10-15, *** P<0.001. Error bars indicate SEM. (C-F) Conversion of ac1 and ac4 to ac2 in coeloconic sensilla. (C) Loss of IR expression in rn-positive ac1 and ac4 sensilla ORNs in rn mutants. (D) Expansion of IR75a expressing ORNs in rn-negative ac2 sensilla to new antennal territories. IR75a glomerulus appears larger in rn mutants. (E) Enhancer trap line 72OK-GAL4 mediated CD8 GFP expression in wild type and mutants. Loss of two rn-positive glomeruli and a single enlarged rn-negative glomerulus are seen in rn mutants. (F) Or35a expression in rn-negative ac3 sensilla is not affected in rn mutants. (G) RT-PCR on selective ORs using RNA from adult rn heterozygous or trans-heterozygous antennae confirms the results from genetic experiments. Or67d and IR92a were chosen to represent rn-positive trichoid and coeloconic sensilla ORNs, respectively, which were lost in rn mutants. Or47b and IR41a were chosen to represent the default rn-negative trichoid and coeloconic sensilla ORNs, respectively, which showed increases in rn mutants. Notably, sexual dimorphism in OR expression is apparent.
Figure 19: ORN specification phenotypes in the basiconic sensilla of rn mutants.
(A) Loss of Or82a (ab5), Or98a (ab7), Or49a, Or85f and Or67a (ab10) expression in \(rn\)-positive basiconic ORNs. \(rn\)-negative ORNs expressing Or92a, Or10a and Or42a from ab1 show increases in the numbers of cell bodies and the sizes of the glomeruli in female mutants (images are from female flies). (B) OR expression in \(rn\)-negative ab2, ab3, ab4, ab6, and ab8 basiconic sensilla ORNs is unaffected in \(rn\) mutants. Graphs show the quantification of cell body counts. \(n=10-15\), ** \(P<0.01\), and *** \(P<0.001\). Error bars indicate SEM. For each OR, Z-projections of confocal stacks for antennae and antennal lobes are shown in top and bottom panels, respectively. Anti-Bruchpilot staining is used to visualize the antennal lobe neuropil (red). OR expression is in green.

Figure 20: The independent effects of \(rn\) and \(lz\) on ORN specification in coeloconic ac2 sensilla.

Antennal lobe images from \(rn\), \(lz\) single mutants and \(rn/lz\) double mutants labeled with IR75a (ac2) reporter (green) are shown. IR75a glomerulus in \(rn/-\) or \(rn/- lz/-\) is bigger than that in WT or \(lz/-\), consistent with \(lz\) not functioning in coeloconic sensilla (atonic lineage).
Figure 21: Presence of M1 motif in the upstream of rn-positive ORs/IRs confirmed by the known motif finding method MAST.

(A) Results obtained by scanning the promoters (1000bp upstream of TSS) of ORs, and 1-1000bp or 1000-2000 bp upstream of TSS for IRs with MAST, using M1 motif found by MEME. Upper axis indicates the number of base pairs upstream from TSS of each OR gene, bottom axis indicates the number of base pairs upstream of IR TSS, and red boxes indicate locations of motif occurrences. Boxes above the line indicate that original motif was found, and boxes below the line indicate that reverse complement was found. For sequences with multiple motif occurrences, green star denotes most significant motif, whose sequence is given to the right. Only two rn-negative receptors Gr21a and Or85a (boxed) contain marginal hits, and the rest ORs/IRs shown are rn-positive. Only genes for which MAST E-value < 1 are shown. No significant hits were found for 1-1000bp upstream of IRs (not shown). (B) Quantification of β–Galactosidase activity in yeast one-hybrid ONPG assay showed no direct binding of Rn to M1 motif. The same 1.8kb Or49a or Or82a promoter fragments containing M1 motif (see also Figure 14B) used in the promoter replacement experiments were cloned as baits. Full length Rn was used as prey. p53 binding site (p53BS) with AD-p53 and Pfog-3 with AD-TRA-1 were used as positive controls for strong and medium interactions, respectively. A 557bp fragment containing T13 motif was cloned into the bait vector for testing the binding of Rn to T13. Results for 1, 2, 3, 4, 9 and 10 represent the average activity from 8 colonies pooled together. Error bars indicate standard deviation.
Figure 22: Changes in ORN precursor lineages predicted by the mathematical model.
(A) No expansion of potential default en-negative ORNs neighboring the en-positive lineages shown in lineage diversification tree (see also Figure 15) was observed, implicating a different function mode employed by en compared to rn and separate factors governing these fates. IR84a and Or67d were chosen as representative ac4 and at1 ORNs, respectively. The graph shows the quantification of cell body counts. n= 10-15. Error bars indicate SEM. Mutants show no significant differences compared to controls. (B) Changes in ORN precursor lineages predicted by the mathematical model. Diversification landscapes for lz, amos, ato single mutants as well as lz/rn, amos/rn, rn/ato double mutant combinations are shown. Output landscapes predicted by the program mimic the experimental outcomes of the mutant combinations except for ato/- rn/- for which experimental results are missing due to technical difficulties.

2.4. Experimental procedures

2.4.1 Fly genetics

Stocks:

todZ3-6129 and totZ3-4626 mutant alleles were previously isolated in the Smith lab (Ha and Smith, 2006). rnGAL4, UAS-rn, UAS-CD8 GFP, UAS-CD2, UAS-FLP, UAS-RedStinger, amos deficiency line Df(2L)M36-S6, en1, en5d, the hypomorphic allele lzS and flies used in MARCM analysis were obtained from the Bloomington Stock Center. amos3 allele and amos-GAL4 line was provided by Andrew Jarman. All OR-CD8 GFP, OR-GAL4 and IR-GAL4 lines were generously provided by the labs of Leslie Vosshall, Barry Dickson, and Richard Benton. Most of these lines have been previously confirmed for faithful expression of the reporters using OR and IR in situ probes (Couto et al., 2005; Fishilevich and Vosshall, 2005). Or67dGAL4 knock-in flies (Kurtovic et al., 2007) were used instead of Or67d-GAL4 due to unfaithful expression of this transgene in additional ORNs. The 72OK-GAL4 line was previously described (Hummel and Zipursky, 2004). See also fly
genotypes used in each experiment.

**Lineage tracing:**

To identify the population of ORNs that express *rn*, we made use of an FLP/FRT-mediated lineage tracing system. This technique utilizes a constitutively active nuclear YFP reporter flanked by FRT sites and a downstream CD2 RFP marker. In the absence of the flipase enzyme, YFP is ubiquitously expressed through the *tubulin* promoter. Providing *rnt⁸GAL4*-driven UAS-FLP leads to recombination between the FRT sites excising the nuclear YFP cassette and allows the expression of the CD2 RFP transgene independent of the GAL4 driver in the cells that used to have an active *rn* promoter.

**Genotypes for fly genetics:**

See Appendix A.

**2.4.2 RT-PCR**

Third instar larvae (around 10 for each genotype), 40 hrs APF pupal antennae (50 for each genotype) or adult antennae (200-300 for each genotype) from wild type and homozygous *rn* mutants were collected. RNA was extracted with an RNeasy kit (Qiagen), treated with on-column DNase digestion (Qiagen), and then reverse transcribed into cDNA using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). PCR was initially performed using treated RNA as templates to exclude the possibility of genomic DNA contamination, and then with cDNA following the standard protocol. *rp49* was used as an internal control.
Primer sequences for RT-PCR

Extron specific primers used in Figure 16A:

E-isoform:

TAATCCTCAACGACGGCTCCAAGC
ATCCGCATAACAGTTGCACCGG

F-isoform:

ATTCCGGCGTTGGGTCTCTCTTCG
ATGCACGAACTCATTTGCAAATACAG

C-isoform:

TACCATCACAACTCGCCGCTGATT
TGCTGGTCACTATGATGTGCTGGT

C2H2 domain shared by all isoforms:

ATCTGCCAGCGAAAGTTCACGC
ATGTGCGTCTTCAGGTGCTTCG

rp49:

ATACAGGCCCCAGATCGTGAAG
ACGTGTGCAACCAGGAACCTCT

OR gene primers used in Figure 18G:

Or67d

TTGCCTTTCTTCTTCGCTGT
TCTCCCTTAGATCGTATTCTCTAT
Or47b
CGCCTTTCTGAGTTTGCTGTG
GTGCCATCCCCGAAATCCAA
IR92a
ACGAAAAACAGGACGATGG
CTCGACCTCGGTGACAAT
IR41a
ACCCACTGCGAGACATTTGTA
GGATTTCCGGCTTTGGCC

2.4.3 Immunohistochemistry

Staining of adult and pupal brains was performed as previously described (Cayirlioglu et al., 2008). Primary antibodies were used in the following dilutions: rabbit α-GFP 1:1000 (Invitrogen), rat α-Ncad 1:20, rat α-Elav 1:100, mouse α-Bruchpilot 1:20 (Developmental Studies Hybridoma Bank), mouse α-CD2 1:1000 (Serotec), guinea pig α-Senseless 1:6250, rabbit α-Amos 1:1250. The following secondary antibodies were used: goat α-rabbit-FITC 1:1000, goat α-mouse-Cy3 1:100, goat α-rat-Cy3 1:200, goat α-guinea pig-Cy3 1:500, goat α-rat-Cy5 1:200, goat α-rabbit-Cy5 1:500 (Jackson ImmunoResearch), Alexa 568 goat α-mouse IgG highly cross-adsorbed 1:300 (Invitrogen). For Or47b and Or88a double labeling experiments, 14-16 um antennal sections were obtained by
cryostat and stained with GFP and CD2 antibodies. Confocal images were taken by an Olympus Fluoview FV1000. All images shown were taken from male flies unless otherwise mentioned. Glomerular volumes were calculated using Imaris MeasurementPro Software.

2.4.4 Bioinformatics

De Novo Motif Finding:

Three de novo motif finding programs (MEME, PRIORITY, and WEEDER) were run on the set of rr-positive OR promoters (Bailey et al., 2006; Narlikar et al., 2006; Pavesi et al., 2004). Promoters were defined as starting at transcription start site (TSS) and continuing upstream for 1000 bp or until meeting the coding sequence of another gene. All sequences were downloaded from Flybase version FB2013_02. MEME and PRIORITY were run with a position-specific prior (Bailey et al., 2006; Narlikar et al., 2006), constructed by comparing between rr-positive and negative promoter sequences. All three motif finders succeeded in discovering a motif occurring in all 11 rr-positive OR promoters. All motifs occurred within 1000 bp of TSS.

Known Motif Finding:

To confirm enrichment of motif within rr-positive promoters, MAST (Bailey and Gribskov, 1998) was used to scan from 0 to 1000 bp upstream of TSS of rr-positive and negative OR. For IR genes, 1-1000 bp and 1000-2000 bp upstream of TSS of rr-positive IR genes were scanned, using position-specific scoring matrix (PSSM) of motif found by
MEME. All 11 motif occurrences were recovered in \(rn\)-positive sequences. Two marginal occurrences (within E-values above all \(rn\)-positive sequences) were found in \(rn\)-negative OR and IR sequences. Two occurrences were found in \(rn\)-positive IR sequences when the 1000-2000 bp range was scanned.

**Command line parameters for all programs:**

MEME: minw (minimum width of motifs to look for): 4, maxw (maximum width of motifs to look for): 20, bfile (background file of nucleotide frequencies): second order Markov model generated from *Drosophila melanogaster* intergenic regions, revcomp: search for motifs on both forward and reverse strands, nmotifs (number of motifs to look for): 5.

PRIORITy: default settings of 50 trials of Gibbs sampling, with 10,000 iterations per trial.

Weeder: DM, Markov background file derived from *Drosophila melanogaster*, S, consider both strands of input sequence, M, motifs can appear more than once in a sequence, medium, search of motifs of length 6 or 8 basepairs.

MAST: bfile (background file of nucleotide frequencies): secondorder Markov model generated from *Drosophila melanogaster* intergenic regions.

**2.4.5 Motif replacement transgene construction**

To test the biological significance of M1, motif replacement constructs were made for \(rn\)-positive Or49a and Or82a promoters. These two genes were arbitrarily chosen
from all \( nr \)-positive ORs to facilitate the mutagenesis procedure and data analysis, as in both cases a single occurrence of M1 resides in (or close to) the middle of a less than 2kb promoter region, which was previously shown (and in this study) to faithfully represent the corresponding gene expression pattern (Fishilevich and Vosshall, 2005). For Or49a, 1844bp upstream of the start codon with the motif TGGTTGGGACA (-564) was replaced with GACCTGCAGTGC, by ligating a 1280bp fragment and a 553bp fragment with a PstI site. The motif-deleted construct was then subcloned into pCaSpeR-AUG-Gal4-X using EcoRI/BamHI sites (Vosshall et al., 2000). Similarly, for Or82a, 1865bp upstream of the start codon with the motif CCTTCCAAGCC (-967) was replaced with GGCTGCAGAG, by ligating an 898bp fragment and a 956 fragment with a PstI site. The modified construct was subcloned into the same vector using EcoRI/BglII sites.

2.4.6 Yeast one-hybrid

Yeast one-hybrid assay was performed as previously described (Deplancke et al., 2006; Pruneda-Paz et al., 2009). The same Or49a or Or82a promoter region used for the motif replacement experiment was cloned as bait into pMW2 and pMW3 vectors. The full-length rotund was cloned as prey into the pDEST-AD vector (Deplancke et al., 2006). Prior to transforming the prey construct, HIS3 and lacZ reporter were tested to select clones with low or no self-activation. ONPG assay for \( \beta \)-galactosidase activity was conducted following the standard protocol (Pruneda-Paz et al., 2009). Positive controls were provided by Albertha Walhout.
2.4.7 Statistical analysis and mathematical modeling of precursor diversification

Statistical analysis:

Statistical analysis of cell body counts and glomerular volume measurement was by unpaired, two-tailed Student’s t test. For all tests, ** P<0.01, *** P<0.001.

Mathematical modeling:

For modeling the diversification landscape, a main function is called ORN_map_landscape.m was generated in Matlab, which sets up the grid and then changes the value of different coordinates of it depending on how many factors are affecting a given column (representing different precursor cells, each of which at the end of the landscape will have a 2 box width) in the grid. The rows correspond to levels of diversification on the tree diagram. A series of nested “if” tests determine which points in the matrix gain a value (dip down), which allows you to enter a matrix of present transcription factors and the grid will change, along with the color of the labels. The input matrix will have a 1 if a factor is present and a 0 if the factor is missing. The program then tests if a factor is present and if it is, then it makes dips for that factor. It then continues to test for the other factors and acts the same way. Example: If atonal is present, it will make the coeleconic dips. Then it will test for rn and if rn is present, it will test for en. If any of the factors are missing, the program will stop generating dips below that layer it in the tree.

A second function called update_grid.m that changes the values in the grid
matrix. ORN_map_landscape.m calls the update_grid.m program each time a new dip in the grid is needed.

The height of the landscape is represented by the values of the elements in the matrix. The matrix is only two dimensions so it represents the x- and y-axis. The values on the matrix represent the third dimension, the z-axis and give the final graph the different heights and dips.

Matlab code: See Appendix B.

2.5 Acknowledgements

We are grateful H. Bellen, Andrew Jarman and Richard Benton for providing fly stocks and antibodies. We thank Duke Light Microscope Facility and Model Systems Genomics, for great technical assistance, Uwe Ohler, Amy Schmid and Katia Koelle for discussions on modeling and Hiro Matsunami, Dave McClay, Amy Bejcovec, Paul Durst and members of the Volkan laboratory for discussions and comments on the manuscript. This work was supported by NIH R01DC02539 funds to D.P.S. and Duke University start up funds and Whitehall Foundation to P.C.V.
3. A Conserved Molecular Network Governing Sensory Neuron Diversification

3.1 Summary of the study

Background

Sensory neuron diversity underlies essential behavioral responses to complex environmental cues. Olfactory receptor neurons (ORNs) are extraordinarily diverse in the receptor genes that they express. This organ-level diversity is achieved by an ensemble of highly specialized individual neurons expressing single receptors from a large pool. In *Drosophila*, 50 ORN classes are grouped in stereotyped combinations that correspond to distinct sensilla subtypes. Each sensillum arises from a single sensory organ precursor (SOP) representing a unique subtype lineage with predefined sensory differentiation potentials. The mechanisms that diversify sensilla subtype-specific SOPs and thus ORN classes are largely unknown.

Results

Previously, we reported a critical component of the SOP diversification program, Rotund (Rn), which functions to increase ORN diversity by diverging novel developmental trajectories from existing precursors within each independent sensilla type lineage. Here, we show that the divergence occurs through direct Rn-dependent repression of the homeodomain transcription factors BarH1/H2. Further, we show that Rn and BarH1/H2, along with Bric-à-brac (Bab), Apterous (Ap) and Dachshund (Dac), constitute a conserved transcription factor (TF) network that patterns the developing
olfactory tissue, where SOPs with diverse ORN differentiation potentials are selected from concentric rings defined by unique combinations of these TFs along the proximodistal axis. The combinatorial code that specify each precursor within the field is set up by cross-regulatory interactions among different factors in the network and possibly under the influence of epidermal growth factor receptor (EGFR) signaling. Modifications of this network generally lead to predictable changes in the diversity of sensilla subtypes and ORN pools.

**Conclusions**

We proposed a molecular map that defines the majority of distinct SOP fates and corresponding sensilla subtype identities. This pre-patterning stage is followed by proneural gene-mediated SOP selection and Notch-mediated neurogenesis, forming a three-step mechanism that can largely explain the sensory neuron diversity observed in the olfactory system. Our model points to a conserved strategy that many other developing systems may use to generate neuronal diversity, and also sheds light on possible mechanisms underlying the unusual evolvability of the chemosensory systems.

**3.2 Results**

**3.2.1 Time-course RNAseq analysis**

From previous studies, we have shown that the Krüppel-like transcription factor Rotund (Rn) cell-autonomously diversifies ORN classes by branching off novel sensilla subtype lineages from parallel default ones. In \( r\) mutants, ORN diversity is reduced
almost by half. Neurons from at4 sensilla in the trichoid zone, ac2 in the coeloconic zone, and ab1 and ab9 in the basiconic zone are all expanded at the expense of ORNs in lineage-specific \textit{rn}-positive sensilla subtypes (Figure 23) (Li et al., 2013).

In order to understand the mechanism by which Rn modulates ORN precursor identities, we took an unbiased approach by carrying out a time-course RNAseq analysis in the wild type (\textit{w^{1118}}), heterozygous and homozygous \textit{rn} mutant flies (Figure 24B, Experimental procedures 3.4.2, 3.4.3). Sequencing libraries were made from developing and adult antennae at the 3rd instar larval stage (T1), 8hr APF (after puparium formation; T2), 40hr APF (T3), and the adult stage (T4). We extracted RNA only from the antennal portion of the larval eye-antennal discs in order to remove contamination by transcripts from the developing eye. These four stages were chosen because each represents an important developmental landmark. Specifically, the imaginal disc epithelium is presumably pre-patterned during the third instar larval stage when Rn is also highly expressed (Goulding et al., 2000; Gupta et al., 1998; Li et al., 2013; Sen et al., 2010). SOP selection from pre-patterned fields happens after puparium formation, and both preneural genes, Amos and Atonal, reach their peak expression around 8hr APF (Gupta and Rodrigues, 1997; zur Lage et al., 2003). Most ORNs are born and some of the earliest olfactory receptors begin to be expressed at 40hr APF. Finally, the adult stage represents the fully developed system with terminally differentiated ORNs and all receptors expressed. Paired-end reads were performed for two biological replicates per
stage per genotype. For heterozygous and homozygous mutants, additional single-read sequencing was conducted on T1, T3 and T4 stages, although the data analysis was primarily done with the complete sets of results from paired-end sequencing (Figure 24B).

3.2.2 Mis-regulation of developmentally critical genes in \textit{rn} mutants revealed by RNAseq analysis

To find genes that potentially function downstream of Rn, we focused our analysis on the three early stages. Because Rn is only expressed during larval and pupal periods (Li et al., 2013), we reasoned that the genes that are under more direct control would show differential expression in one or more of these early time points. For each stage, we first filtered out ORs/IRs/GRs, and then narrowed it down to genes that show differences for both homozygous vs. \textit{w}^{1118} and homozygous vs. heterozygous comparisons. Because the heterozygous background may have some dominant effects due to the presence of the balancer chromosome and the heterozygous flies do not show any OR phenotypes (Li et al., 2013), saving genes that pass both comparisons would help remove irrelevant genes caused by the balancer and meanwhile enhance the confidence of our discovery. We particularly chose to use a less stringent cutoff (unadjusted $p < 0.1$) to include as many candidates as possible at the early phase of the analysis. Genes with low expression levels in all three genotypes (normalized count < 20) and the ones that showed different directions of misregulation between homozygous vs. \textit{w}^{1118} and homozygous vs. heterozygous were excluded. A Venn diagram generated from
the final lists for each stage reveals that some genes are misregulated only in one particular stage, while others show misregulation in multiple stages (Figure 24A). Both up and down-regulated genes were identified in the mutants (Figure 24C).

Based on GO term analysis, we listed genes that may have potential functions in development, such as transcription factors and signaling molecules (Table 1). In addition, functional clustering analysis using the online tool, DAVID (Huang et al., 2009a, 2009b), for each category in the Venn diagram uncovered a functional cluster including homeodomain (-like) proteins BarH1/2 (B-H1/2, Bar or B), Apterous (Ap), Bric-à-brac1 (Bab1) and Rn as being modified in rn mutants. We speculated that these genes might have regulatory relationships with Rn due to their known functions in imaginal disc morphogenesis based on the GO germs. Interestingly, they all showed changes in transcription levels only during early developmental stages (Figure 24D). It is worth mentioning that B-H1 and B-H2 are functionally redundant, and the same is largely true for Bab1 and Bab2 (Figure 25) (Couderc et al., 2002; Kojima et al., 2000). Because only Bab1 but not Bab2 was included in the functional clustering analysis, we re-examined the RNAseq datasets for Bab2. We found that Bab2 has an overall higher level of expression than Bab1, and the same trend of misregulation exists for Bab2 (Figure 24D), even though the p values are approaching but still below the arbitrary 0.1 cutoffs in multiple cases. As a result, we decided to focus on B-H1/2, Ap and Bab1/2 first and explore their functions in ORN diversification further.
3.2.3 A common molecular network patterning the antennae and legs

The developmental functions of B-H1/2, Bab1/2, Ap and Rn have been mainly studied in the context of Drosophila leg development. Due to their segmental feature, fly legs have been used as an excellent model to study the molecular mechanism of proximodistal (PD) axis formation, where these genes play a critical role (Estella et al., 2012; Kojima, 2004). Briefly, the PD genes, under the influence of morphogen gradients, form a dynamic transcription factor network via cross-regulation to define a number of concentric domains on the disc, which in turn determine individual segment identities (also see Chapter 1.4.2) (Estella et al., 2012; Kojima, 2004; Morata, 2001). We systematically examined the spatial patterns of these factors in the developing antennae, and found that each factor is expressed in a concentric ring along the proximal-distal axis. The gene expression patterns are remarkably similar between the legs and antennae, suggesting that these two organs share the same molecular tool kits that pattern the discs (Figure 24G) (Baanannou et al., 2013; Giorgianni and Mann, 2011; Natori et al., 2012; Pueyo et al., 2000).

Many genetic studies have been done on the cross-regulations between individual PD genes during leg development. Because antennae and legs are homologous and thought to share a common ancestral leg origin, it is plausible to think that a similar regulatory network may exist in the antenna (Casares and Mann, 1998, 2001; Pueyo et al., 2000). We found that the regulatory relationships of PD genes from
the leg-patterning network can explain our RNAseq results on the misregulation of B-H1/2, Bab1/2 and Ap (Figure 24D, Figure 35, right bottom), confirmed by *in vivo* stainings (Figure 24E, F, Figure 26). B-H1 is normally expressed in the center of the disc, bounded by the central fold. In *rn* mutants, B-H1 is expanded outside of the central cells into cells that are normally Rn-positive and Bar H1/2-negative, but the expansion is confined within the distal boundary of Dac (Figure 26C, D). The ectopic cells that are labeled with B-H1 antibody in *rn* mutants are positive for the *rn* promoter reporter, suggesting that this *rn*-positive precursor domain may have switched fates resulting from the loss of Rn and the expansion of Bar-H1/2 (Figure 26A). On the other hand, Bab2 expression is significantly reduced in *rn* mutant, and the reduction seems to be more severe outside the central fold where Rn is mostly expressed (Figure 24E, F). Consistent with the RNAseq results, we did not detect any changes in *ap* expression in the third instar larval stage (Figure 26B, Figure 28E). Taken together, these results suggest that a common PD gene regulatory network module operates in parallel during leg and antennal development.

### 3.2.4 Intercalation of pre-patterning domains

We noticed that the expression domains of several of these factors overlap in the third instar antennal disc, and therefore we wanted to dissect the spatial relationships between these factors more carefully. To simplify the descriptions, we use the central fold (CF) as a landmark, which is usually observed as an unstained dark circle in a
superficial section of confocal images, to separate the disc into inside and outside CF regions (Figure 27A, B). Outside CF, Dac, Rn and Bab are expressed from more proximal to more distal area in the disc (Figure 27B). Due to the substantial overlap in their expression patterns, these three factors divide the region into four concentric rings. We number the rings starting with the outermost one being R(1), and therefore R(1) to R(4) are assigned to this region (Figure 27B, D). It is worth mentioning that Bab is expressed in a gradient, consistent with the previously reported phenomenon in the context of leg development (Godt et al., 1993). Our results show that the highest concentration of Bab is found near the central fold, and its gradient decreases toward both outermost and innermost areas of the disc (Figure 25, Figure 27B, Figure 28A, B).

Three more rings can be found inside the central fold. R(5) is the only ring that shows quadruple labeling by 4 factors (Rn, Bab, Ap, and Bar) (Figure 28, Figure 27E). Interestingly, this ring corresponds to the only region that expresses Rn inside of the central fold. Bar expression is either absent in the center-most region, or its level may be under the detection threshold (Figure 28E). Taken together, the partial overlapping patterns of Dac, Rn, Bab1/2, B-H1/2 and Ap expression demarcate 7 concentric ring domains in the third instar antennal disc, and each ring is marked by a unique combination of pre-patterning factors (Figure 27C, D, E).
3.2.5 Functional involvement of PD genes in ORN diversification

As mentioned earlier, \( rn \) mutation causes limited expansion of Bar and reduced expression of Bab in the third instar antennal disc (Figure 24E, Figure 26C, D). We constructed a similar scheme to depict the spatial relationships of these PD factors in \( rn \) mutants (Figure 34A). In this model, the TF combinations in domains where Rn is expressed, namely R(2) to R(5), are altered due to the changes to the expression of Bar, Bab and Rn. As a result, R(1) would be expanded into R(2) and the proximal portion of R(3). Similarly, R(6) would be expanded proximally, as R(5) has the same code to R(6) in \( rn \) mutants. The rest of Rn-positive domains contain new codes that are not found in the wild type discs. Since \( rn \)-positive precursors and corresponding ORNs are lost in the mutants with concomitant increases of some \( rn \)-negative ones (Li et al., 2013), we hypothesized that the changes in ORN classes are caused by the modifications in the network components which disrupt the code that defines each ring of future S0Ps.

To test this hypothesis, we analyzed OR expression as a read-out of ORN classes in flies overexpressing Bar or Ap in \( rn \)-positive fields, using quantitative RT-PCR for a panel of 20 olfactory receptor genes representing each of the antennal sensilla subtypes (Table 5). Ap was previously shown to protect Bar from being repressed by Rn during leg development (Natori et al., 2012). Since the antennae and legs seem to share a similar network, we predicted that overexpressing Ap in \( rn \)-positive cells in the antennal discs should release Bar inhibition in this zone. Indeed, Ap overexpression results in the
ectopic expression of Bar outside of its normal territory in the antennal disc (Figure 31). Therefore, either manipulation should at least partially recapitulate the adult ORN phenotypes in *rn* mutants, which also has expanded Bar expression. Indeed, quantitative RT-PCR from adult antennae in these two genetic backgrounds showed that the changes in the expression of the majority of receptors tested mirror the ORN fate changes observed in *rn* mutants (Figure 29, Figure 30).

We also examined the patterns of precursor fields in these overexpression lines (Figure 34B, C). Interestingly, the expansion of Bar in the Ap overexpression line is also bounded by the distal limit of Dac, as seen in *rn* mutants, possibly due to a repressive effect of Dac or/and a lack of activation (Figure 31, Figure 32, Figure 26). However, unlike in *rn* mutants, Bar does not fully extend to the boundary, and hence, these proximal cells in R(4) are positive for Rn but negative for Dac and Bar (Figure 32C). They also express Bab and Ap, making them a separate subpopulation within R(4) (Figure 34B). In addition, we saw a loss of Rn expression in R(5), leading to the domains within the central fold devoid of Rn expression (Figure 31A, B). The simplest interpretation is that increased levels of Ap repress Rn expression in a context-dependent manner. Moreover, we found that Dac expression is decreased in R(2) and R(3) that express Rn (Figure 32A, B), suggesting that *rn* promoter-mediated Ap expression represses Dac in this overlapping domain. Because Dac represses Bab (Chu et al., 2002), the reduction in Dac expression should theoretically cause an increase in Bab expression, although we
cannot detect any obvious changes for Bab (Figure 31). This may be due to its overall low concentration in this region by repression from other factors (Chu et al., 2002).

When we looked at the imaginal discs of the Bar overexpressing larvae, we found that the central fold (CF) disappeared (Figure 33), which is consistent with the previous finding that CF formation in the leg disc requires differential expression of Bar at the CF boundary during the early third instar stage (Kojima et al., 2000). Nonetheless, Bab concentration gradients are unaffected (Figure 33A). Similarly, Dac shows normal expression, despite the reported function of Bar to repress Dac in the distal area, which suggests that the repression may be time-sensitive and/or context-dependent (Giorgianni and Mann, 2011; Kojima et al., 2000). An alternative explanation is that the driver is not strong enough to induce effective levels of Bar for its repressive function as manifested by the weak staining of Bar in the proximal region (Figure 33B). We are in the process of assessing Ap and Rn expression in the Bar overexpression line. Because Bar activates Ap, we expect to see an expansion of Ap along with Bar expression to cover the Rn domain. However, if we do not see such an expansion, it would likely mean that some factor or factors in the proximal region prevent Ap from being expressed there, or the timing and levels of Bar is inappropriate to activate Ap. Similarly, analyzing Rn expression using the tagged line would provide a more definitive answer for its expression domain in relation to other factors.
Based on these analyses, we drew similar illustrations for precursor domains in the Ap and Bar overexpression backgrounds (Figure 34B, C). They reveal different patterns of gene expression for a number rings compared to the \(rn\) mutants, which may account for their differences in adult ORN classes as shown by the qPCR results (Figure 29, Figure 30). We conclude that the PD genes function in combinations to modulate precursor and ORN fates.

### 3.2.6 Sensilla subtype fate mapping onto the pre-patterned domains

Next we wanted to ask which precursor identities are generated from each of these 7 domains. Since all of the components within a sensillum arise from a single sensory organ precursor (SOP), we wanted to know the sensilla subtype identities of SOPs from each concentric domain. To do this, we used promoter-driven reporter lines for each individual gene to label ORN axons. Because ORN sensory identities are closely linked with the glomerular identities in the brain, we can infer which ORN classes express the given factor from the glomerular labeling pattern in this analysis. These fate mapping data are analyzed together with expression domains of each factor and genetic mutant phenotypes to generate the sensilla subtype fate map.

From the qPCR results and previous studies (Li et al., 2013), we noticed that the only sensilla subtypes that are expanded in the \(rn\) mutants as well as in the Ap/Bar overexpression lines are at4 and ac2, with the exception of Or47b in at4 in Bar-overexpression, which showed a slight decrease (Figure 23, Figure 29, Figure 30).
However, even for this outlier, imaging of adult antennae with an Or47b reporter reveals ectopic Or47b ORNs in the medial region of the antenna, similar to those observed in rn mutants (data not shown). Similarly, the only precursor domain that is expanded in all three genotypes is R(6) (Figure 34). We therefore deduced that the R(6) region generates precursors that give rise to ORNs in at4 and ac2 sensilla (Figure 35). This is consistent with the fate mapping results, such that some at4 ORNs are positive for ap, bar and bab reporters, but all at4 ORNs are negative for rn and dac (Table 2). We reason that dynamic changes of gene expression during asymmetric precursor cell divisions can lead to only subsets of ORNs expressing a certain factor. We applied the same reasoning to other similar situations discussed below. In addition, ORNs in ac2 are bab-positive (Table 2), consistent with the origin of R(6) which expresses Bab (Figure 27D, E). Even though we do not have direct evidence to show that ac2 precursors are positive for either ap or bar during development, presumably due to gene expression dynamics or/and limitations of the reporters, we believe ac2, behaving in the same way as at4 in genetic manipulations, arises from R(6) (Figure 35).

Because ab1 and ab9 are expanded in rn mutants (Figure 23), and the representative ORs in these sensilla showed no change in Ap/Bar overexpression antennae (Figure 29, Figure 30), these two sensilla fates fit precursors from R(1), which is expanded only in rn mutants but not the other two backgrounds (Figure 34). Consistently, ab1 and ab9 are labeled by the dac reporter (Table 2), and it has been
shown that Dac is required for the specification of ab1 and ab9 (Song et al., 2012). However, they are also positive for bab but negative for rn (Table 2). This is a non-existing combination according to the spatial relationships between these three factors (Figure 27D, E). We attribute this discrepancy to either an artifact of the reporter or late expression of Bab unrelated to precursor fate determination. Therefore, we map ab1 and ab9 to R(1) (Figure 35).

at2, ab2, ab4, ab8, ab6, and ac3 are rn-negative and unaffected in rn mutants (Figure 23). Because all of them contain at least one ORN class that is positive for ap and/or bab, but they are all negative for dac, we assigned them to R(7) (Table 2, Figure 27D, E, Figure 35). ab3, similar to ab1 and ab9, is positive for bab and dac, and requires Dac for its specification (Song et al., 2012), which would place it within R(1). However, overexpression of Bar shows that ab3 behaves the same way as the sensilla subtypes in the center R7 (Figure 30). It is possible that there are two populations of ab3 determined by different sets of factors, one specified in R(7), and the other one specified in the outermost R(1) (Figure 35).

Next we wanted to map all rn-positive sensilla in the middle 4 rings. Since the only Rn-positive, Bab-negative and Dac-negative domain is R(2) (Figure 27D, E), we assigned this ring with the ab10 fate, which is the only sensilla subtype that meets the same criteria (Table 2, Figure 35). Likewise, ab7 and ai1 are the only two sensilla subtypes that are triple labeled by rn, bab and dac reporters, and are mapped to R(3).
at3, ab5 are labeled by rn and bab but not any other reporters, we assigned them to R(4). (Table 2, Figure 27D,E, Figure 35). Finally, ac1 and ac4 are mapped to the quadruple positive R(5) (Figure 35). Consistent with this mapping, these two sensilla are positive for rn and bab, and one ORN in ac1 is unambiguously labeled by the ap reporter, while one ORN in ac4 is labeled by the bar reporter (Table 2). Even though it seems no ORN classes in these two sensilla are labeled by all 4 factors simultaneously, the expression of these factors may be only required for precursor fate specification in a narrow window of the larval stage, and later restricted to specific lineages and daughter cells for downstream functions. Consistent with this explanation, ap expression is highly dynamic, and disappears in the adult stage (data not shown). Also, the reporter lines we used may not fully capture the gene expression dynamics presumably due to complex enhancer elements in the control region.

These series of analyses provide us with a sensilla subtype fate map on the concentric domains of the larval antennal disc. Each ring is labeled by a unique set of PD transcription factors, and these factors in turn determine a specific subset of sensilla subtypes (Figure 35). We used this model to explain the majority of the phenotypes observed in rn mutants and Ap/Bar overexpression lines (see 3.2.8 for detailed explanations).
3.2.7 Separation of different precursor fates within a ring by Bab concentration gradients

Bab is partially activated by Rn (Baanannou et al., 2013), and it is significantly downregulated in \( rn \) mutants (Figure 24E, F). It is plausible to think that Bab functions downstream of Rn to specify \( rn \)-positive precursor fates. If this is the case, we should see reduced expression of the receptors from the 8 \( rn \)-positive sensilla subtypes in a bab mutant. To our surprise, only 2 of the 8 receptors tested showed reductions, and other 2 were even increased in the bab\(^{PR72} \) hypomorphic allele (Figure 36).

We noticed varied changes for sensilla subtypes from the same ring (Figure 36). For example, for sensilla specified in R(7), ab2 and ab6 are reduced, whereas at2 is increased in the mutant. The simplest interpretation is that different levels of Bab are required to distinguish these fates in the same ring. When the overall level of Bab is decreased, some sensilla requiring higher Bab may be converted to the ones that require lower Bab. Similarly, only ac1 from R(5) is reduced, and one explanation is that the lowered Bab expression is still above the threshold for specifying ac4, but not for ac1. Alternatively, ac1 (requiring higher Bab) may be converted to ac4 (requiring lower Bab), and compensate for the loss of the endogenous ac4, which may die due to the loss of Bab. The same reasoning can be applied to ai1 vs. ab7 from R(3). For sensilla from R4, we saw increases in at1 and ab5, but no significant change in at3, albeit a trend towards downregulation, in the mutant. We then counted the number of Or19b neurons housed in at3 sensilla, and found that it is significantly reduced (data not shown). This
discrepancy may be due to an artifact in the qPCR result, which is sensitive to random fluctuation of gene expression levels, especially when the changes in the numbers of cells are small. In contrast, the bab-positive at4 and ac2 appear to be normal in this hypomorphic allele. This could either because these two sensilla are specified with wider ranges of Bab levels, or some other factors are needed to differentiate the two fates.

To further prove that one major function of Bab is to use its concentration gradients to distinguish alternate fates specified by the same code within a ring, we need to do the converse experiments by overexpressing Bab. Preliminary data suggest that sensilla subtypes show the opposite changes in this genetic background compared to the mutant allele (data not shown). A detailed analysis is currently in progress. Besides Bab, it is likely that other factors are required to differentiate alternate precursor fates, especially in those Bab-negative rings.

3.2.8 Explanations of rn, Ap-overexpression and Bar-overexpression lines using the model proposed in this study

First we wanted to use our model to explain the rn mutant phenotype. In the rn mutant imaginal discs, Bar expression is de-repressed and expanded outside the central fold, but this expansion respects the boundary between R(3) and R(4) (Figure 26C, D, Figure 37A). Consequently, the new proximal limit of Bar is abutted by the distal limit of Dac. Another prominent phenotype is the reduction of Bab expression due to the lack of Rn activation, resulting in a shallower concentration gradient (Figure 24E). The
combined effects, along with the loss of Rn, alter the code for each ring (Figure 37A). Particularly, R(2) and the proximal part of R(3) would become Dac only, converting the precursors’ potentials to an R(1) potential. We found that this conversion in precursor fates can be used to explain the loss of ai1, ab7 and ab10 subtypes and the expansion of ab1 and ab9 observed in rn mutants (Figure 23). Although we do not have markers to show the origins of sensilla subtype conversions, we suspect that ab7 and ab10 may be converted to ab9 within the small basiconic sub-lineage, and ai1 might be converted to ab1. The precursors in the distal part of R(3) may die because of an invalid code.

Moreover, the code in R(5) is changed to the same one as in R(6), particularly because the decrease in Bab brings its level in R(5) closer to the level in R(6), while R(6) itself is Rn-independent and thus unaffected. This change leads to the loss of R(5) and expansion of R(6) (Figure 37A). Meanwhile, the ectopic Bar in R(4) changes its code to Bab and Bar. However, we could not detect Ap expansion in R(4) in the larval stage, presumably due to a delay in Bar activation in this ectopic zone (Figure 26B). We want to argue that the presence of Ap in R(6) and R(7) is dispensable and mainly due to a lack of repression in the center. Therefore, Bab and Bar expression in R(4) in rn mutants is equal to the code in R(6), and is sufficient to specify ac2 and at4 fates. Consistent with this argument, Ap function is not required for the specification of ac2 or at4 sensilla subtypes (data not shown), which are generated from R(6) according to our model (Figure 37A). Based on this reasoning, it would appear that the major function of Ap is to protect Bar
from being repressed by Rn in the quadruple positive zone R(5). This function does not occur at the transcriptional level, but is perhaps mediated by formation of a complex with Bar and/or Rn to counteract Rn’s repressive function. Consistently, Ap has been shown to interact with Bar in a protein complex (Natori et al., 2012; Pueyo and Couso, 2004). Therefore, our model can explain the loss of at1, at3, ab5, ac4 and ac1 sensilla subtypes from R(4) and R(5), and the expansion of ac2 and at4 normally specified in R(6) (Figure 37A). Again, we speculate that the conversions occur within sensilla type lineages: at1 and at3 are converted to at4; ac1 and ac4 are converted to ac2; ab5 might be lost due to an incompatible Bab level. Taken together, our model provides an explanation for lineage-specific conversions in sensilla subtypes found in \( rn \) mutants by changes in the network code (Figure 23).

When we used \( rn^{89GAL4} \) to drive UAS-Ap, we found Rn expression in R(5) is lost (Figure 31A, B). Due to the code change, R(6) is expanded into R(5), leading to sensilla subtype conversions from ac4 and ac1 to ac2 and at4 (Figure 37B). In addition, we found Bar is expanded close to the R(3) and R(4) boundary (Figure 32). As a result, the cells in the proximal portion of R(4) express Rn, Bab and Ap (Figure 37B). We think that this subpopulation of cells may be converted to fates in R(7) in this line, because: 1) qPCR results showed that some sensilla subtypes from R(7) are increased (Figure 29); 2) all sensilla subtypes from R(4) are significantly reduced; 3) the new code is similar to that in R(7), except that it contains Rn (Figure 37B). Our explanation for why the presence of Rn
does not interfere with the rest factors specifying R(7) fates is that Ap may interact with
Rn to counteract its function. On the other hand, the cells in the distal half of R4 become
quadruple positive for factors normally found in R(5), resulting in sensilla subtype
conversions from at1, at3 and ab5 to ac4 and ac1 (Figure 37B). We noticed that ac4 and
ac1 are changed in opposite directions (Figure 29). This is likely caused by endogenous
cells in R(5) being further converted to the fates in R(6) as mentioned earlier. The
differences in cell numbers that are converted during this relay may account for the
different phenotypes for these two sensilla subtypes. Furthermore, sensilla subtypes
from R(6) are increased and the corresponding glomeruli appear bigger in the
overexpression line (Figure 29, data not shown).

We have shown that Dac levels are reduced in R(2) and R(3), and we deduced
that Bab expression should be elevated, despite that this elevation may be under the
detection threshold in our assay (Figure 32A, B, also see Chapter 3.2.5). Based on our
findings about the function of Bab concentration gradients (see Chapter 3.2.7), we think
that this hypothetical de-repression of Bab may cause ai1, which requires lower levels of
Bab, to be converted to ab7, requiring higher levels of Bab, in R(3) (Figure 29, Figure
37B). However, it is unclear why the overexpression of Ap in this domain does not affect
ab7 specification. Finally, ab10 sensilla are dramatically reduced in Ap overexpression
(Figure 29). This may be due to some toxic effect of this non-existing code, such as the
blocking of Rn function by Ap leaving only low levels of Dac, which is insufficient to
specify any fate. Taken together, the proposed model can by and large explain the ORN phenotypes seen in Ap overexpression flies (Figure 37B).

Next we wanted to use our model to explain the changes in sensilla subtype composition in Bar overexpression flies. Since the analysis on the expression patterns of PD genes is not fully complete, some of the following statements were based on our best estimation. Provided that Ap is activated in ectopic domains in response to Bar overexpression, precursors in R(4) with the new code of Rn, Bab, Bar and Ap would be converted to the fates normally specified in R(5), leading to sensilla subtype conversions from at1, at3 and ab5 to ac4 and ac1 (Figure 30, Figure 37C). The overexpression of Bar in R(5) may inhibit Rn expression similar to the context-dependent repression of Rn by Ap in this ring (Figure 31A, B). Under this assumption, the precursors in R(5) are converted to R(6), causing the endogenous ac4 and ac1 to be converted to ac2 and at4 (Figure 37B). Accordingly, we saw increases in the transcription of receptors in ac2 and at4 subtypes (Figure 30). Due to the large numbers of at1, at3 and ab5 being converted to ac4 and ac1, the overall numbers of ac4 and ac1 can be increased, despite the conversion of the endogenous population. Indeed, we found that receptors expressed by these two sensilla are up-regulated in Bar overexpression (Figure 30). We also found that a number of sensilla subtypes from R7 are increased (Figure 30). Although we do not have a satisfying explanation for it at this moment, we suspect that the ectopic sensilla may originate from the Bab-positive R(3), in which this new code somehow is competent to
induce the fates normally specified in the center of the disc (Figure 37C). Lastly, ab10 from R(2) may be lost due to the toxicity of the new code in this ring, which is consistent with the down-regulation of the tested receptor in ab10 (Figure 30, Figure 37).

An interesting phenomenon is that, even though Bar expansion seems to be a critical contributor to the \( rn \) mutant phenotype, and indeed many sensilla subtypes in the Ap/Bar overexpression lines show similar changes to those in \( rn \) mutants (Figure 29, Figure 30), the causes are very different in these situations based on our model (Figure 37). Because the intercalated patterns of these transcription factors, changes to one factor often lead to a domino effect involving many neighboring domains. In this sense, the similar phenotypes described above are essentially incidental and occur through distinct mechanisms.

### 3.2.9 Conserved molecular network in patterning gustatory receptor neuron fates

An interesting anatomical analogy between the legs and antennae is that the distal portions of both appendages are chemosensory organs covered by sensilla. The legs, being part of the gustatory system, also display neuronal diversity, characterized by a huge variety of gustatory receptors expressed on the legs. Unlike ORNs, individual gustatory receptor neurons (GRNs) express multiple receptors, and a given GRN class can be found in different sensilla with different GRN clusters (Ling et al., 2014).

Because the legs and antennae use the same molecular network to pattern the chemosensory portions of the appendages, we want to ask if a similar genetic program
operates to diversify GRN fates. We tested this hypothesis in *rn* mutants. It has been reported that *rn* is required for the development of tarsal segment 3 (ta3), and this segment is lost in *rn* mutant (St Pierre et al., 2002). However, we do not have a reporter line that uniquely label ta3, and in fact, the only GRN class known to reside in ta3 is only found in the forelegs and it is also present in all other segments except ta1 (Ling et al., 2014). Indeed, using markers, Gr5a-Gal4 and Gr61a-Gal4 to label this sugar sensing GRN class yields variable patterns of receptor expression in the *rn* mutant forelegs (data not shown). Therefore, we cannot conclude with certain about the fate of GRNs in ta3.

On the other hand, these two GRs are also expressed in the mid and hind legs, in which they are restricted to the GRNs in ta4 and ta5 (Ling et al., 2014). In both cases, we could reproducibly detect an extra neuron proximal to the endogenous neurons that express the receptor in the mid or hind legs of *rn* mutant (Figure 38). Using a bitter receptor Gr58c that is expressed by a partner neuron in the same sensilla, we observed a similar ectopic neuron (Figure 38). In contrast, the Gr43a-expressing neurons that coexpress Gr61a but do not pair with the Gr58c-expressing bitter neuron appear to be unchanged in the mutant (data not shown). These results suggest that the sensilla, 5b and 4s, that house the Gr5a/Gr61-expressing sugar neurons and the Gr58c-expressing bitter neuron are expanded towards the proximal region of the legs in *rn* mutant (Figure 38). We think that it is more likely to be the expansion of 4s due to its closer distance to the ectopic neuron. However, we cannot distinguish between these sensilla with the
reporters currently available. Taken together, we speculate that the same molecular
network is used in parallel to diversify chemosensory neurons in the antennae and legs.

3.2.10 Binding of Rn to Bar and Bab in vivo using a CRISPR-mediated Rn-EGFP line

Next we wanted to know if the genes in the network directly regulate each other
by binding to promoters. We focused on the function of Rn, as this may help explain the
misregulation of B-H1/2 and Bab1/2 in rn mutant. In addition, the molecular function of
Rn is poorly understood compared to other factors. Previous in vitro assays have shown
that Rn binds to a T-rich motif (T13) in the LAE enhancer upstream of Bab2 to activate
its expression during leg and antennal development (Baanannou et al., 2013). However,
in vivo evidence for Rn binding targets has been missing due to the lack of a high-quality
antibody. We decided to generate a fly line that has the endogenous rn locus tagged
with an epitope, using CRISPR/Cas9-mediated homologous recombination, which
allows us to circumvent the problem of antibody quality (Experimental procedures
3.4.6). Then we can use antibodies against the epitope to do chromatin
immunoprecipitation (ChIP) followed by qPCR to test binding of Rn to bab or bar
regulatory elements in the antennal discs.

rn locus encodes three isoforms, and only the E/F isoforms are relevant to ORN
diversification (Li et al., 2013; St Pierre et al., 2002). Because all three isoforms share the
same 3’ region, we decided to add an EGFP or 3XFLAG tag to this end, which would
label all three transcripts without potentially disrupting any isoform function by the
insertion of the tag. Two CRISPR cutting sites were made to replace the last common exon with the same sequence (stop codon removed) connected to a tag via a flexible linker. Meanwhile, a DsRed selectable marker was incorporated in the repair template to facilitate genetic screening (Figure 39, Figure 40, Figure 42). Homologous recombination was mediated by a 1kb arm beyond each cutting site. Through this manipulation, we were able to obtain EGFP or 3X-FLAG tagged Rn stocks (Figure 41, Figure 43, Table 4). The labeling is faithful and robust as seen by the co-staining with the *rn* reporter (Figure 44) (Li et al., 2013). Further, the flies homozygous for the tagged chromosome are healthy and normal for ORN diversification, suggesting that adding a tag does not affect Rn function (Figure 45). We could not detect any mutations in the potential off-target sites by sequencing.

We also attempted to make an E/F isoform-specific tagging line by injecting the same guide RNAs and repair template mixture to the *roe*<sup>3</sup> heterozygous stock. Because *roe*<sup>3</sup> harbors an amber stop codon in the first exon, a successful recombination onto the mutant chromosome would lead to a truncated C isoform without the tag and normally tagged E/F isoforms (Figure 39). As expected, we obtained the 3XFLAG tagged E/F isoform-specific line from the screening. However, the EGFP tagging was unsuccessful presumably due to the heterozygous background, a larger tag, and very low survival/fertility rates (Table 4).
In order to generate an E/F isoform-specific EGFP-tagging line, we made a second attempt by using CRISPR technology to add stop codons to the first exon of C isoform in the homozygous stock that already has the EGFP tag on all three isoforms (Figure 39). Eventually, we were able to get this line, which we used to do ChIP (Figure 43). Consistent with previous reports, from 5 CRISPR screenings we did here, founder lines have relatively low survival (55.6% on average) and fertility (34.5% of survived on average) rates. Nonetheless, within the lines that are fertile, the germline transmission rate is decent (18.5% on average) (Table 3, Table 4). Taken together, we showed an example how to use CRISPR technology to make isoform specific tagging in an endogenous gene locus.

Lastly, we would use the Rn-E/F-EGFP line to probe the binding targets of Rn. qPCR primers were designed in the first 2kb upstream of the transcription start sites (TSS) in the bar loci (Table 6). A primer set covering the T13 motif in the Bab2 enhancer was used as a positive control, while the M1 motif region from Or82a promoter was used as a negative control (Baanannou et al., 2013; Li et al., 2013). ChIP on antennal disc tissues was able to detect direct binding of Rn to the published bab2 enhancer and the promoter region of B-H2 using the Rn-EGFP line, and further confirms that Rn does not bind to OR promoters (Figure 46). These results suggest that the intercalated TF domains may be formed at least partially by directly cross-regulating each other through binding to the promoters/enhancers.
3.3 Discussion

How neuronal diversity in the brain is generated from a limited genomic toolkit remains largely unknown. An extraordinary example of neuronal diversity is seen in the *Drosophila* olfactory system, where selective expression of generally a single olfactory receptor gene from a repertoire of approximately 80 possible genes generates 50 different classes of ORNs. ORN classes are found in invariable clusters of 1-4 neurons in individual sensilla, which can be classified into types based on their morphology and subtypes based on the specific combination of ORN classes they house (Couto et al., 2005; Fishilevich and Vosshall, 2005; Rodrigues and Hummel, 2008). Previously, we reported that a critical component of the diversification program, Rotund (Rn), functions to diverge new sensilla sub-lineages from default ones within each parallel sensilla type lineage (Li et al., 2013). Here, we show through transcriptome analysis and chromatin immunoprecipitation of endogenously GFP-tagged Rn that Rn is a component of a cross-regulatory transcription factor (TF) network module, likely under the influence of epidermal growth factor receptor (EGFR) signaling, that patterns the ORN precursor field along the proximodistal axis prior to neurogenesis. The interactions between different components of the TF network module partition the precursor field into concentric domains, representing clusters of epithelial cells with distinct differentiation potentials. Unique combinations of TFs from the network define each domain, from which sensilla subtype lineage-specific SOPs are selected by proneural gene expression
and lateral inhibition. Genetic manipulations of the factors cause alterations of the code, leading to mostly predictable shifts between sensilla subtypes and neuronal identities. We think that this early TF network plays a major role during neuronal diversification by pre-patterning the disc causing restriction of identities of cells in the precursor field. The addition of new factors to this network, such as Rn, re-routes developmental trajectories that increase neuronal diversity. Naturally occurring changes to this pre-patterning phase at a larger time scale may also accommodate rapid olfactory evolution.

We proposed a conserved stepwise strategy to explain the majority of ORN diversity. First, the pre-patterning phase generates distinct pools of epithelial cells; then, sensory organ precursors are selected by turning on proneural genes; finally, these precursors undergo neurogenic divisions that allocate alternate fates into daughter cells via Notch signaling. This three-step mechanism provides an efficient solution that chemosensory systems may use to generate sensory neuron diversity. One salient aspect of this cellular diversification strategy is its modularity. Each step is singular in its purpose, yet could function broadly across systems in a context-dependent manner. For example, Rn may be used by both the antenna and leg is to increase the complexity of the patterned precursor field, but generates distinct neuronal differentiation potentials. Similarly, Notch is reiteratively used following SOP selections, with the purpose of segregating binary cell fates during asymmetric divisions, although the exact fates being segregated are very different (Endo et al., 2007). Therefore, this step-wise mechanism
simplifies the overall difficulty of the neuronal diversification problem by logically deconstructing similar differentiation processes into smaller single-purpose steps.

The overall neuronal identities within the tissues (e.g. olfactory vs. gustatory) are likely to be determined by Hox genes during embryogenesis at the top of the hierarchy in the cascade. For instance, olfactory lineages are determined by the gene Homothorax in the antennal disc, whereas Homothorax is inhibited by Antennapedia in the leg discs conferring gustatory identities (Morata, 2001). Regardless of the particular Hox gene, the PD gene network module seems to perform similar diversification commands in both chemosensory systems in a context-dependent manner. One difference between the leg (gustatory) and antennal (olfactory) appendages is that SOPs in the antenna appear not to respect the original domains in which they are patterned, and migrate to class-specific locations within the antenna (Song et al., 2012). This feature of the olfactory system is consistent with the lack of the extensive folds in the disc and joints in the adult-form appendage. It seems that the PD gene network establishes different precursor identities but likely does not directly govern either migration of SOPs or the final location of ORNs.

Even though our findings are in the PNS, there are similar examples of stepwise patterning and diversification in the CNS. For example, different neuroblast lineages in the Drosophila embryonic CNS are first specified by spatially restricted factors within specific positions in an orthogonal grid (Bayraktar and Doe, 2013). Anterior-posterior
axis specification is controlled by Hox-segment polarity genes, which determine the overall fate, just as in the PNS (leg vs. antenna). Dorsoventral (D/V) patterning is controlled by cross-regulatory transcription factors, which are turned on in response to different concentrations of morphogens such as Hedgehog and Dpp. Similar to olfactory SOPs, patterning of the neuroepithelium is followed by the expression of proneural genes and selection of neuroblasts, which undergo asymmetric divisions and neurogenesis. The division patterns and factors that asymmetrically segregate into each daughter cell are remarkably similar regardless of the neuroblast lineages to which they belong (Bayraktar and Doe, 2013; Jacob et al., 2008). This is analogous to how Notch-mediated binary fate determination is reiteratively used to segregate alternate potentials of SOPs across lineages.

Examples of similar neuronal diversification cascades are also seen in the vertebrate CNS and PNS. One classic example is in the case of spinal cord neuron diversification, where morphogen gradients (BMP/Shh) along D/V axis of the neural tube activate different sets of transcription factors in the precursors to set up a number of domains prior to neurogenesis, thereby diversifying neuronal subtypes (Jessell, 2000; Ulloa and Briscoe, 2007). Recently, radial glia that give rise to neurons of the cortex were also found to be heterogeneous (Bayraktar et al., 2015). Combinatorial TF network modules confer positional and temporal information to each cell in order to create a diverse progenitor population. Segmental patterning of these neural stem cells
contributes to neuronal and glial diversity (Bayraktar et al., 2015). Similarly, cortical projection neuron fates can be switched among lineages when the corresponding gene network is modified, leading to different proportions of brain regions in the neocortex (Greig et al., 2013). These results, in light of our findings, suggest a strategy composed of modular and simple commands functioning in a nested manner in multiple developmental contexts to increase neuronal diversity.

There are also parallels between our findings in the fly olfactory system and the much more complex vertebrate olfactory system. Even though stochastic selection has been proposed as a mechanism for selection of specific OR genes by different ORNs, the restriction of mammalian OR expression into distinct zones suggests that a deterministic mechanism may also be at play (Fuss and Ray, 2009). For example, each zone within the olfactory epithelia expresses distinct sets of OR genes, and deletion of an OR gene is associated with a choice switch only made from genes in the same zone or sometimes the same gene cluster (Fuss and Ray, 2009; Sato et al., 2007; Serizawa et al., 2003; Shykind et al., 2004). Therefore, mammalian OR choice likely contains two modes of regulation: zonal restriction by a deterministic process and random selection within a zone by a stochastic process. (Fuss and Ray, 2009). Consistent with this hypothesis, some OR classes that are restricted to specific domains in the olfactory epithelium were shown to contain known TF binding sites (Hoppe et al., 2003, 2006). Interestingly, some TFs are evolutionarily conserved, belonging to the same PD gene network mentioned in this
study. We are curious to see if any of the mechanisms used in diversifying fly ORNs are also used in the mammalian system during the step of OR zonal separation.

On the evolutionary scale, clear analogy exists between ORN precursor diversification process and the segment diversification during early embryogenesis along the myriapods-insect lineage. The addition or eliminations of TFs governing either process might reflect, or likely instruct the generation of new fates. Based a modern point of view about “the law of development” postulated almost two centuries ago (Abzhanov, 2013), the increasing individual complexity of patterns and accompanying sequences of dynamic regulation in gene expression might mirror the history of the modifications in the involving partners over time. For example, a ground state might represent expression domains of Ap, Bar and Dac during antennal disc development, and Rn might be added to this network later in development or evolution. This would explain the expansion of specific ORN pools in default sensilla subtypes within each morphological type in \( rn \) mutants as well as \( Bar/ap \) overexpression resulting in a dramatic decrease in ORN diversity. Indeed, the Rn gene product is expressed later relative to Dac/Bar/Ap (Natori et al., 2012), and seems to be unique to Arthropods and especially insects. Thus, it is plausible to speculate that Rotund is a newer addition to the network, which evolved to generate novel olfactory neurons in order to compensate the special olfactory needs by the ancestral forms of the organisms.
And finally, what do neuronal identity and diversity mean for ORNs? We associate terminal differentiation of ORNs with the specific selection of an OR gene for expression. It is possible that OR expression and ORN diversity are regulated by a set of “terminal selector genes,” similar to those proposed by Oliver Hobert, where a TF, or combination of TFs, directly regulates the expression of genes required for terminal differentiation and function (Hobert, 2010). So far, a list of postmitotic TFs have been shown to regulate OR expression directly, by associating with OR promoters (Bai and Carlson, 2010; Jafari et al., 2012; Ray et al., 2008; Tichy et al., 2008). However, loss of these TFs only affect the expression of specific subsets of OR genes, and yet many OR genes have binding sites for these factors. What then determines the specificity of each TF’s function remains unknown. It is proposed that these terminal selector genes may act in unique combinations within ORN classes to specifically control the expression of different ORs, although the expression patterns of these TFs has not been defined among ORN classes. It is possible that chromatin states around OR promoters in different ORN classes govern how these TFs function within each class. Regardless of the exact mechanisms in between, it is likely that the expression/function of the terminal selector genes is regulated, at least in part, by the developmental context established by the pre-patterning network that we have described. Establishing a clear link between the early patterning networks and the late terminal selector TF network will be critical to resolving these paradoxical results. Understanding the diversification process from
homogeneous fields of precursors to diverse, terminally differentiated neuronal pools will provide key insights into how cascades of master regulatory transcription factor networks can generate the cellular complexity seen in multicellular organisms.

Figure 23: Lineage-specific sensilla subtype conversions in rn mutants.

Schematic representation showing ORNs in rn-positive sensilla subtypes are converted to ORNs in rn-negative ones within each sensilla type lineage. Amos and high Lozenge determine basiconic lineage; Amos and low Lozenge determine trichoid lineage; Atonal determines coeloconic lineage. The hierarchy of the genes is about the logic of the organization, but not the timing or regulatory relationships of expression. Each box represents a sensilla subtype with the olfactory receptor genes expressed by neurons housed in the sensilla. Co-expressed genes are boxed in dashed line. Arrows denote sensilla conversions in rn mutants. Questions marks are uncertain directions of conversions due to the lack of developmental markers. The rn-positive intermediate sensilla ai1 is not shown.
Table 1: List of representative genes misregulated in rn mutants.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Genes</th>
<th>laval</th>
<th>P_8hrAPF</th>
<th>P_40hrAPF adult</th>
<th>Molecular Features</th>
<th>Functional Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>transcription factors</td>
<td>BarH1/2</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>homeobox domain</td>
<td>TF; eye-antennal disc morphogenesis</td>
</tr>
<tr>
<td></td>
<td>bab1</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>BTB/POZ-like</td>
<td>TF; eye-antennal disc morphogenesis</td>
</tr>
<tr>
<td></td>
<td>ap</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>homeobox domain</td>
<td>TF; imaginal disc development</td>
</tr>
<tr>
<td></td>
<td>E(spl)m5- HLH</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>bHLH domain</td>
<td>TF; cell fate commitment</td>
</tr>
<tr>
<td></td>
<td>ely2b</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>ENY2 family</td>
<td>transcription (co)activator; nuclear pore; SAGA complex</td>
</tr>
<tr>
<td>inter(intra)-cellular</td>
<td>Wnt4</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>Wnt</td>
<td>pattern specification; sensory organ development</td>
</tr>
<tr>
<td>signaling molecules</td>
<td>ana</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>TGF-β</td>
<td>negative regulation of neuroblast proliferation</td>
</tr>
<tr>
<td></td>
<td>trrol</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>EGF-like</td>
<td>downstream of ana; neuroblast proliferation</td>
</tr>
<tr>
<td></td>
<td>pyr</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>FGF family</td>
<td>cardioblast differentiation</td>
</tr>
<tr>
<td></td>
<td>Mip</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>B-type prepro-allatostatin</td>
<td>neuropeptide hormone; GPCR signaling</td>
</tr>
<tr>
<td></td>
<td>Ac3</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>adenylate cyclase</td>
<td>intracellular signal transduction</td>
</tr>
<tr>
<td></td>
<td>Pkc53E</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>Ser/Thr kinase</td>
<td>intracellular signal transduction</td>
</tr>
<tr>
<td>signaling receptors</td>
<td>Toll-9</td>
<td>↓</td>
<td>↓</td>
<td></td>
<td>TIR domain</td>
<td>signaling receptor; immune response</td>
</tr>
<tr>
<td></td>
<td>Sdr</td>
<td>↓</td>
<td>↓</td>
<td></td>
<td>EGF receptor</td>
<td>IGF binding; negative regulation of growth</td>
</tr>
<tr>
<td></td>
<td>Eip758</td>
<td>↓</td>
<td>↓</td>
<td></td>
<td>nuclear receptor</td>
<td>hormone receptor; response to ecdysone; DNA binding</td>
</tr>
<tr>
<td>heat shock proteins</td>
<td>Hsp23</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>Hsp20 domain</td>
<td>response to heat, hypoxia; protein binding</td>
</tr>
<tr>
<td></td>
<td>Hsp68</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>Hsp70 family</td>
<td>protein folding; neurogenesis; rn interactor</td>
</tr>
<tr>
<td></td>
<td>Hsp70Aa/b</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>Hsp70 family</td>
<td>response to heat, hypoxia</td>
</tr>
<tr>
<td></td>
<td>Hsp70Ba/b/b/c</td>
<td>↑</td>
<td>↑</td>
<td></td>
<td>Hsp70 family</td>
<td>response to heat, hypoxia</td>
</tr>
<tr>
<td>cytoskeletal proteins</td>
<td>Map205</td>
<td>↓</td>
<td>↓</td>
<td></td>
<td>205 kDa MAP</td>
<td>microtubule binding</td>
</tr>
<tr>
<td></td>
<td>Unc-115a/b</td>
<td>↓</td>
<td>↓</td>
<td></td>
<td>Zinc finger, LIM-type</td>
<td>actin binding; axon guidance</td>
</tr>
</tbody>
</table>

↑ Up-regulation in rn mutants  ↓ Down-regulation in rn mutants  ◌ No changes in rn mutants  ○ Little expression in both genotypes
A

T1: 3rd instar  
T2: 8hr APF

143
15
13
5

T3: 40hr APF

269

B

Experimental Design

<table>
<thead>
<tr>
<th>Stages</th>
<th>Genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w1118</td>
</tr>
<tr>
<td>T1</td>
<td>2 (2)</td>
</tr>
<tr>
<td>T2</td>
<td>2 (2)</td>
</tr>
<tr>
<td>T3</td>
<td>2 (2)</td>
</tr>
<tr>
<td>T4</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

C

Summary of mis-regulated genes

<table>
<thead>
<tr>
<th>Directions</th>
<th>Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
</tr>
<tr>
<td>Up</td>
<td>73</td>
</tr>
<tr>
<td>Down</td>
<td>115</td>
</tr>
<tr>
<td>Total</td>
<td>188</td>
</tr>
</tbody>
</table>

D

BarH1

BarH2

Ap

Bab1

Bab2

E

rm +/-  
rm +/-

F

Bab2 expression level

G

Bab2

Proximal

Detail
**Figure 24:** Time-course RNAseq analysis showing misregulations of genes critical to development.

(A) Venn diagram showing the numbers of gene misregulated in *rn* mutants in the three early stages. (B) The experimental design of the RNAseq experiments. Numbers indicate total biological replicates. The numbers in parentheses denote the samples sequenced with paired-end reads as a single batch. The differences are for single-read sequencing, which was conducted on *rn*+/− and *rn*−/− in T1, T3, and T4 as a single batch. (C) Summary of misregulated genes based on the directions. (D) Expression levels of B-H1/2, Bab1/2 and Ap between the control and mutants in different stages. Normalized expression data by DESeq2 was used. (E) Confirmation of RNAseq results on Bar and Bab expression. Antibodies to B-H1 (red), and Bab2 (blue) were used. Expansion (bracketed) of B-H1 outside CF (dashed line) is apparent. Bab2 expression shows concentration gradients. In these composite images, two circles (arrow heads) on the ridges of the central fold show the highest concentrations, but the high level of expression is continuous in the central fold (see Figure 27A, Figure 28B for a bottom section). Bab2 expression is weaker in *rn* mutants, although the overall pattern is unchanged. The *rn*89GAL4 reporter labels cells that have active *rn* promoter (green). CF, central fold (arrow). (F) Quantification of Bab levels in (E). n=10, *** p<0.001. (G) Expression patterns of PD genes in the third instar larval antennal disc. They are expressed in concentric rings along proximodistal axis. Anti-B-H1, anti-Bab2, anti-Dac were used to visualize Bar, Bab and Dac. aprK568 (stained with anti-β-gal) and Rn-EGFP (CRISPR line) were used to visualize Ap and Rn.

**Figure 25:** Overlapping expression of Bab1 and Bab2.
(A) Z-projections of confocal images on the third instar larval antennal disc. Bab1 expression is visualized by staining the enhancer trap line \textit{bab1}^{A128} with β-gal antibody (Godt et al., 1993). Bab2 is stained with its antibody (Couderc et al., 2002). These two proteins are partially redundant and overlapping in expression. Both genes show gradient expression and are confined within the boundary set by Dac (red). (B) Single slices of confocal images show the boundary between Bab and Dac. Boxed area is shown on the right. Weak Bab2 slightly overlaps with Dac, while the overlapping between \textit{bab1}^{A128} and Dac is not obvious, presumably due to the level of expression below the detectable range.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure26.png}
\caption{Expansion of Bar but not Ap outside the central fold of the larval antennal.}
\end{figure}

(A) Bar (red) is expanded in the third instar antennal disc of \textit{rn} mutants (same as in Figure 24E). The ectopic domain (bracketed) is labeled by the \textit{rn} reporter (green), suggesting fate conversions may occur. (B) In contrast, \textit{ap} expression (red) visualized by the enhancer trap line \textit{aprK568} remains inside the central fold (compared to Figure 28E and F). CF, central fold (arrow and dashed line). (C) The expansion of Bar is restricted within the distal boundary of Dac (blue).
Figure 27: Intercalation of pre-patterning genes outside the central fold.

(A) The side view of the antennal disc showing the prominent central fold. Bab expression is highest near the central fold and decreases towards the center and the proximal edge. Bab (labeled with anti-Bab2 in red) and Rn (labeled with tagged EGFP in green) partially overlap. Three dashed lines indicated the optical sections, denoted as “top”, “middle” and “bottom”, where most of the images in the document were taken. The puncta of Bab2 staining partially overlaps with DAPI signals. (B) Co-labeling of the disc by Dac and Bab2 antibodies and Rn-EGFP. Individual channels are shown on the right. The boxed area is shown below. Four rings with distinct labeling patterns are marked and depicted in (D). (C) A cartoon showing the side view of 7 concentric domains in the antennal disc. Each ring is color coded as in (D) and (E). (D) and (E) Schematic representation of intercalated domains of pre-patterning genes in the antennal disc as shown in Figure 27B and Figure 28. Each ring is labeled by a combination of factors, and each factor is expressed by a number of rings. Bab shows concentration gradients both inside and outside the central fold. CF, central fold.
Figure 28: Intercalation of pre-patterning genes inside the central fold.

(A) and (C) Middle sections (Figure 27A) of the third instar larval antennal disc stained with Bab2, B-H1, GFP (Rn-EGFP), β-gal (ap\(r^{K568}\)) antibodies. Individual channels are shown on the right of each image. Boxed areas are shown below. Rn expression is seen as a circle (arrow) inside the central fold in this view. Ring 5 is labeled by all four factors. Ring 6 is positive for Bab, Bar and ap. Ring 7 is positive for Bab and ap, although ap expression is out of focus in this plane (the section is deeper than that in A. See F for co-
labeling of the two in a top section). (B) and (D) Bottom sections of the same samples in (A) and (C), respectively. The co-expression of Bar and ap appears as a circle. High levels of Bab are continuous in the central fold. (E) A top section view showing co-expression of Bar and ap in R(6), and negative for Bar in R(7). (F) A top section view showing co-expression of Bab and ap in R(7). The boxed area and spit channels are shown on the right. CF, central fold (dashed line). Also see Figure 27E.

Figure 29: Olfactory receptor expression analysis by qPCR in Ap overexpression line.

At least one receptor from each sensilla subtype is analyzed. Receptor genes and sensilla subtypes are listed below the graph. Corresponding rings (Figure 35) from which sensilla are born are labeled on the top (see below for the mapping process). The receptors that show the same phenotype as in rt mutants are blocked in green shade. The level of expression for each gene is normalized to the average expression of all receptors tested. At least three biological replicates were done for each genotype. * p<0.05, ** p<0.01, *** p<0.001.
Figure 30: Olfactory receptor expression analysis by qPCR in Bar overexpression line.

At least one receptor from each sensilla subtype is analyzed. Receptor genes and sensilla subtypes are listed below the graph. Corresponding rings (Figure 35) from which sensilla are born are labeled on the top (see below for the mapping process). The receptors that show the same phenotype as in *rr* mutants are blocked in green shade. The level of expression for each gene is normalized to the average expression of all receptors tested. At least three biological replicates were done for each genotype. * \( p<0.05 \), ** \( p<0.01 \), *** \( p<0.001 \).
Figure 31: Expansion of Bar outside CF and repression of Rn inside CF by Ap overexpression.

(A) and (A’) Control disc showing normal expression of Rn (Rn-EGFP), Bab (anti-Bab2) and Bar (anti-B-H1) in the top (anterior) and bottom (posterior) section, respectively. Bar is always restricted within the central fold (CF), and appears as a circle in the bottom section (A’). (B) and (B’) Ap overexpression disc stained for the same factors in the top (anterior) and bottom (posterior) section, respectively. Bar no longer respects the CF boundary in this manipulation, and expanded outside CF (arrow head). Meanwhile, the expression of Rn within CF in R(5) disappeared. (C) A side-view scheme illustrating the relative positions of confocal sections shown in the images. Bar expression is labeled in red. Individual channels are shown on the right of each image. Central fold is circled by dashed line.
Figure 32: Limited expansion of Bar and repression of Dac by Ap overexpression.

(A) Z-projection of the confocal image showing spatial relationship between Bar and Dac in the control disc. Bar is confined within the CF and a gap between Bar and Dac can be clearly seen. (B) In the \textit{rn89GAL4}-driven Ap overexpression line, ectopic Bar and Dac are adjacent. Dac expression is repressed where Ap is overexpressed (R(2/3), bracketed). CF, central fold (dashed line). Relevant rings are marked (see Figure 34). (C) In a single section of the same sample in (B), cells express no Bar (asterisks) and weak Bar (crosses) can be seen. Bar does not fully extend to cover the whole gap between endogenous Bar and Dac. (D) Relevant regulatory relationships between pre-patterning genes in Ap overexpression line are shown. Boxed areas are shown on the right of each image.
Figure 33: Elimination of the central fold by Bar overexpression.

(A) In $rn^{89GAL4}$-mediated B-H1 overexpression, the central fold is not formed. Besides that, the relative spatial relationship between Bab (anti-Bab2) and Dac (anti-Dac) is unchanged. The levels of expression also seem to be normal. The highest level of Bab still occurs in the center of the band, as if the disc was flattened. (B) B-H1 expression is weaker near the Dac domain. Dac expression appears to be unaffected.
Figure 34: Modifications of precursor pre-patterned domains in \(rn\) mutants, \(Ap/Bar\) overexpression lines.

Schemes showing expression patterns of PD genes in \(rn\) mutants (A), \(Ap\) overexpression (B), or \(Bar\) overexpression (C) lines, in relation to the 7 rings defined in Figure 27D, E. For \(rn\) mutants in (A), the relevant genes are shown below. Each ring is color-coded based on the TF combinations in the wild type. All new combinations are labeled in grey. Factors that are changed (patterns or levels) compared to the wild type are labeled in brown. Expression patterns of \(Ap\) and \(Rn\) in the \(Bar\) overexpression line are not determined (question marks).
Table 2: Fate mapping of pre-patterning factors with sensilla subtypes.

<table>
<thead>
<tr>
<th>Sensilla</th>
<th>OR</th>
<th>Glomerulus</th>
<th>Ap (developing brain)</th>
<th>Bar (adult)</th>
<th>Bab (adult)</th>
<th>Rn (Li et al.)</th>
<th>Dac (Song et al.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>at1</td>
<td>Or67d</td>
<td>DA1</td>
<td>+/−</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at2</td>
<td>Or23a</td>
<td>DA3</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or93c</td>
<td>DC3</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>at3</td>
<td>Or2a</td>
<td>Da4en</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or19a/b</td>
<td>DC1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or41a</td>
<td>DA4i</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>at4</td>
<td>Or47b</td>
<td>VA1iv</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or88a</td>
<td>VA1d</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or65a/b/c</td>
<td>DL3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab1</td>
<td>Or21a</td>
<td>V</td>
<td>+/−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or92a</td>
<td>VA2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or10a/Gr10a</td>
<td>DL1</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or42b</td>
<td>DM1</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab2</td>
<td>Or59b</td>
<td>DM4</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or85a</td>
<td>DM5</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab3</td>
<td>Or22a/b</td>
<td>DM2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or85b</td>
<td>VM5sd</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab4</td>
<td>Or7a</td>
<td>DL5</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or56a/Or33a</td>
<td>DA2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab5</td>
<td>Or47a</td>
<td>DM3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or82a</td>
<td>VA6</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab6</td>
<td>Or49b</td>
<td>VA5</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab7</td>
<td>Or67c</td>
<td>VC4</td>
<td>+/−</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or89a</td>
<td>VM5v</td>
<td>+/−</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab8</td>
<td>Or9a</td>
<td>VM3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or43b</td>
<td>VM2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab9</td>
<td>Or67b</td>
<td>VA3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or95a/A/B</td>
<td>D</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ab10</td>
<td>Or49a/Or85f</td>
<td>DL4</td>
<td>+/−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or67a</td>
<td>DM6</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac1</td>
<td>IR31a</td>
<td>VL2p</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR92a/IR76a</td>
<td>VM1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR76d</td>
<td>VL1</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>VM6</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac2</td>
<td>IR75a</td>
<td>DP11</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR44a/IR76a</td>
<td>VC3m/VC5?</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR76d</td>
<td>VL1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac3</td>
<td>IR76a/b/c</td>
<td>DL2d/DL2v?</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or35a/IR76a</td>
<td>VC3</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ac4</td>
<td>IR84a</td>
<td>VL2a</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR76a/IR76b</td>
<td>VM4</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IR76d</td>
<td>VL1</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ai1</td>
<td>Or13a</td>
<td>DC2</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Or46a/B?</td>
<td>VA7m</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pupal brains (50-75hr APF) were used to examine ap-positive ORN identities by ap<sup>md544</sup>-driven GFP. Adult brains were used for analyze the Bar (by bab<sup>Pgal4-2</sup>-driven GFP) and Bab (by NP4099-driven GFP) data. Weak expression is indicated by “+/−”. Rn and Dac expression data are taken from published results (Li et al., 2013; Song et al., 2012).
Figure 35: Sensilla subtype diversification by cross-regulated transcription factors in concentric domains of the antennal disc.

A model showing sensilla subtype identities determined by unique combinations of transcription factors that are organized in 7 concentric domains in the antennal disc. The model is generated by integrating the data of TF expression domains (Figure 27D, E), sensilla fate mapping (Table 2), and genetic phenotypes of various manipulations (Figure 29, Figure 30, Figure 34). In R(3), R(4), R(5) and R(7), different levels of Bab are required to distinguish sensilla subtype fates within a ring. R(1) and R(6) may use other mechanisms to distinguish the alternate fates within a ring. The regulatory relationships between pre-patterning factors are shown in the graph, which is created based on the published data during leg development (Estella et al., 2012; Kojima, 2004; Kojima et al., 2000; Natori et al., 2012) and the results shown in this study.
Figure 36: Olfactory receptor expression analysis by qPCR in bab hypomorphic allele.

Homozygous bab⁰(PR72) allele is used to compare with the wild type. At least one receptor from each sensilla subtype is analyzed. Receptor genes and sensilla subtypes are listed below the graph. Corresponding rings (Figure 35) from which sensilla are born are labeled on the top. The level of expression for each gene is normalized to the average expression of all receptors tested. At least three biological replicates were done for each genotype. * p<0.05, ** p<0.01, *** p<0.001.
Figure 37: rn mutant, Ap and Bar overexpression phenotypes explained by the sensilla subtype diversification model.

Using the model proposed in Figure 35 to explain the rn mutant phenotype (A), Ap and Bar overexpression phenotypes, (B) and (C), respectively. Each ring is color-coded based on the combinations of TFs in the wild type. In the wild type, sensilla subtype fates that would be affected due to the genetic manipulations are labeled in green, and the respective fate conversions are indicated by arrows. Factors that are changed (patterns or levels) compared to the wild type are labeled in brown. In (A), Bab*: downregulation of Bab. In (B), Dac*: downregulation of Dac; Bab*: upregulation of Bab; Rn*: proposed functional blocking of Rn by Ap overexpression. In (C), Ap and Rn expression are labeled with question marks, indicating not yet tested. No sensilla are formed from the grey zones.
Gustatory receptor neuron (GRN) and ORN specification may be regulated by the same molecular network. In \( rn \) mutant, tarsal segment (ta3) is missing, and other segments are fused. Ectopic expression of Gr5a, Gr61a and Gr58c can be seen in the distal region of the second tarsal segment, suggesting a duplication of 4s/5b sensilla. Neurons labeled by the red asterisk in the pictures are the ectopic neuron only seen in the mutant. Different sections of confocal images are shown when neurons are not on the same focal plane. All legs are either from the mid or from the hind leg. Cartoon illustrating the \( rn \) phenotype.

Figure 38: Regulation of GRN specification by the conserved molecular network.
in GRNs is shown below. The ectopic sensillum is labeled in red. The scheme for GRN classes and the receptors expressed in the m5b/m4s and m5v sensilla are modified from Ling, F., et al. The molecular and cellular basis of taste coding in the legs of Drosophila. The journal of neuroscience. 2014. 34. 21(7148-64). B, bitter neuron; S, sugar neuron. Receptors that are used in the study are circled in green. In the wild type, there are a pair of sensilla for each sensilla type, and only one for each pair was drawn.

![Diagram of GRN classes and receptor expression](image)

**Figure 39: Molecular cloning design for CRISPR-induced tagging repair templates.**

*rn* locus on the genome is oriented from right to the left. E/F (yellow), and C (blue) isoforms share the last two exons. Introns and intergenic regions are shown as a line. Individual elements are depicted in the scheme. All homology arms (in grey shade) are around 1kb (see Experimental procedures 2.4.6). Only the construct for EGFP-tagging is shown. To add the 3XFLAG tag, the design is the same except for the tag. To generate isoform specific tagging, the *roe* allele is used, and indicated by asterisk. HR, homologous recombination.
Figure 40: Plasmid map for CRISPR-induced tagging repair template.

Different parts of the construct are shown on the map, including the cloning sites. The backbone is pBluescript II SK. Only the EGFP-tagging construct is shown, and others are very similar (see Experimental procedures 2.4.6.1).
Figure 41: Flowchart of CRISPR-induced endogenous tagging.
Figure 42: Expression pattern of DsRed selectable marker in the repair template.

The addition of a DsRed cassette can facilitate the screening process (Gratz et al., 2014; Horn et al., 2000). The Cas9 transgenic fly we used is labeled by a GFP marker driven by the same 3XP3 promoter. The candidate founder lines show both DsRed and GFP signals in the same regions. The reporter can be seen throughout development and adult, except a brief period in mid-pupa. Particularly, the anal plate (arrow) is strongly labeled in the larval stage. Dorsal organs (arrow) are labeled in the early pupal stage. Eyes and ocelli are fluorescent in the late pupal and adult stage.
Figure 43: Correct targeting of the tags by CRISPR showing Rn expression pattern in the eye-antennal disc.

(A) Live GFP image of eye-antennal disc tissue from the Rn-EGFP line. All three isoforms are tagged. E and F (and perhaps C) are expressed in the antennal portion, while C isoform is predominantly expressed in the eye part. (B) Live GFP image from the isoform-specific tagging RnE/F-EGFP line. The expression in the eye disappeared due to the stop codons introduced in the C isoform. (C) Rn-3XFLAG line stained with anti-FLAG. Expression pattern is the same as in (A). (D) RnE/F-3XFLAG line stained with anti-FLAG. Expression pattern is the same as in (B). C isoform is not translated due to the amber stop codon from rae3.
Figure 44: Faithful labeling of Rn expression by CRISPR-induced tags.

Rn-EGFP and Rn-3XFLAG labeling cells completely overlap, which also overlap with cells labeled by a faithful Rn E and F specific reporter, *rn*89GAL4. Pupal antennal discs (5hr APF) were stained.

Figure 45: Normal function of Rn with CRISPR-induced endogenous tags.

Rn-EGFP and Rn-3XFLAG CRISPR lines are healthy and show normal OR expression in the antenna, suggesting tagged Rn has normal functions. If Rn function were comprised due to the tagging, phenotypes similar to *rn* mutants would be seen (*rn* mutant images are the same as in Figure 9A, Figure 11B).
Table 3: CRISPR injection scheme.

<table>
<thead>
<tr>
<th>Injected flies*</th>
<th>Donor</th>
<th>chiRNA 1</th>
<th>chiRNA 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>vas-Cas9.RFP(-)</td>
<td>rncR-EGFP</td>
<td>rnc-chiRNA-intron</td>
<td>rnc-chiRNA-inter</td>
</tr>
<tr>
<td>(FM7a,Tb)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vas-Cas9.RFP(-)</td>
<td>rncR-FLAG</td>
<td>rnc-chiRNA-intron</td>
<td>rnc-chiRNA-inter</td>
</tr>
<tr>
<td>(FM7a,Tb)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vas-Cas9.RFP(-);</td>
<td>rncR-EGFP</td>
<td>rnc-chiRNA-intron</td>
<td>rnc-chiRNA-inter</td>
</tr>
<tr>
<td>roe[3]/TM3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vas-Cas9.RFP(-);</td>
<td>rncR-FLAG</td>
<td>rnc-chiRNA-intron</td>
<td>rnc-chiRNA-inter</td>
</tr>
<tr>
<td>roe[3]/TM3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vas-Cas9.RFP(-);</td>
<td>rncC-STOP</td>
<td>rncC-chiRNA-5'</td>
<td>rncC-chiRNA-3'</td>
</tr>
<tr>
<td>rEgFP-DsRed(-)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

vas-Cas9.RFP(-) chromosome is from Bloomington stock #55821; roe[3] chromosome is from #7411. The majority of the injected embryos from first and second experiments are homozygous for vas-Cas9.RFP- (thus parentheses for FM7a); all the rest are homozygous for vas-Cas9.RFP (-). See Experimental procedures 3.4.6 for details.
Table 4: CRISPR screening results

<table>
<thead>
<tr>
<th>ID</th>
<th>Injected flies¹</th>
<th>Donor</th>
<th>Larvae</th>
<th>Founder fertility⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>vas-Cas9.RFP(−)</td>
<td>rncR-EGFP</td>
<td>173</td>
<td>20(2)/44(3) 22(6)/59(10)</td>
</tr>
<tr>
<td>2</td>
<td>vas-Cas9.RFP(−)</td>
<td>rncR-FLAG</td>
<td>151</td>
<td>17(2)/49(3) 22(4)/37(6)</td>
</tr>
<tr>
<td>3</td>
<td>vas-Cas9.RFP(−); roe[3]/TM3</td>
<td>rncR-EGFP</td>
<td>188</td>
<td>8/31 18/42</td>
</tr>
<tr>
<td>4</td>
<td>vas-Cas9.RFP(−); roe[3]/TM3</td>
<td>rncR-FLAG</td>
<td>158</td>
<td>3/43 16/40</td>
</tr>
<tr>
<td>5</td>
<td>vas-Cas9.RFP(−); rnc-EGFP DsRed(−)</td>
<td>rnc-STOP</td>
<td>174</td>
<td>21/60 15/64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Founders yielding DsRed+ offspiring F₁⁵</th>
<th>Founders with DsRed+ offspring</th>
<th>Overall germline transmission⁰</th>
</tr>
</thead>
<tbody>
<tr>
<td>3(1)/20(2) 1/2</td>
<td>3(1)/20(2); 15.0% 1/22(6); 4.5%</td>
<td></td>
</tr>
<tr>
<td>4(1)/17(2) 3/4</td>
<td>4(1)/17(2); 23.5% 3/22(4); 13.6%</td>
<td></td>
</tr>
<tr>
<td>0/3</td>
<td>0/3; 0% 1/16 6.3%</td>
<td></td>
</tr>
<tr>
<td>13/21 7/15 11/13 7/7</td>
<td>11/21; 52.4%; 7/15 46.7%</td>
<td></td>
</tr>
</tbody>
</table>

**Table Notes:**

a. vas-Cas9.RFP(−) chromosome is from Bloomington stock #55821; roe[3] chromosome is from #7411. The majority of the injected embryos from ID 1 and 2 are homozygous for vas-Cas9.RFP−; all the rest are homozygous for vas-Cas9.RFP (−).

b. The format is the number of fertile flies over the number of survived adult flies. The numbers in parentheses are for founders of vas-Cas9.RFP (−)/FM7a,Tb genotype.

c. The format is the number of flies in the category over the number of fertile flies. The numbers in parentheses are for founders of vas-Cas9.RFP (−)/FM7a,Tb genotype. At least 60 or all F1 flies from each fertile founder were screened for DsRed.

d. The format is the number of flies in the category over the number of founders yielding DsRed+ F1. The numbers in parentheses are for founders of vas-Cas9.RFP−/FM7a,Tb genotype. At least 60 or all F1 flies from each fertile founder were screened for DsRed, and about 20 individual DsRed+ F1 (unless fewer flies were recovered which would all be used) from each candidate founder were crossed. For ID 1 and 2, all fertile DsRed+ F1 were PCR screened and stained to test the presence of tags. For each founder lineage, at least 1 F1 fly with positive results for all the tests was sequenced for the targeted region to confirm a clean homologous recombination event. For ID 3,4,5, F2 larvae were stained to check the tags before sequencing to confirm targeted events. Potential off-targets regions were PCR amplified for sequencing.

e. The percentage is calculated as the proportion of fertile founders that yield targeted tagging event. The numbers in parentheses are for founders of vas-Cas9.RFP−/FM7a,Tb genotype.
Figure 46: Direct binding of Rn to regulatory elements of Bab and Bar by ChIP.

For B-H2 promoter binding, the numbers indicate distance in base pairs from TSS. Or82a is used as a negative control. Binding of Rn to Bab2 T13 motif has been shown by EMSA (Baanannou et al., 2013). The binding is tested *in vivo* here, and used as a positive control. The current data is shown by using the Rn-EGFP line. At least three biological replicates have been done. **p<0.01.
3.4 Experimental procedures

3.4.1 Fly genetics

Stocks:

bab\textsuperscript{A128} and Df(3L)bab\textsuperscript{PR72} were from Frank Laski. UAS-BarH1\textsuperscript{M13} was from Tetsuya Kojima. \textit{rn}\textsuperscript{btl}, \textit{rn}\textsuperscript{btl}, tub-FRT-nYFP-FRT-CD2 RFP was previously described (Li et al., 2013). OR-CD8 GFP, OR-GAL4, IR-GAL4, GR-GAL4 lines were from Leslie Vosshall, Barry Dickson, Richard Benton and John Carlson respectively (Couto et al., 2005; Fishilevich and Vosshall, 2005; Ling et al., 2014; Silbering et al., 2011). \textit{Or67d}\textsuperscript{GAL4} knock-in stock was a pulished line to show faithful expression of Or67d (Kurtovic et al., 2007).

\textit{rn}\textsuperscript{89GAL4}, bab\textsubscript1Agd4-5 (#6802), bab\textsubscript1Pgal4-2 (#6803), ap\textsuperscript{md544}, ap\textsuperscript{K58h}, UAS-ap, UAS-CD8 GFP, UAS-FLP were all from Bloomington Stock Center. NP4099 (\textit{barGAL4}) was from Drosophila Genetic Resource Center. For \textit{bab} expression, we tested both enhancer trap lines. The BL#6802 labels both ORNs and PNs, which would interfere our interpretation of ORN contribution to the labeling in the brain. Therefore, we only used the data collected from the nearly ORN-specific line (BL#6803) to do the fate mapping for \textit{bab}.

Genotypes for fly genetics:

See Appendix A.

3.4.2 RNA extraction and library preparation

For the RNAseq analysis, wandering third instar larvae antennal discs (~70 for each genotype), 8hr APF pupal antennae (~50 for each genotype), 40hr APF pupal
antennae (~50 for each genotype), and adult antennae (150 males and 150 females) from

\textit{w\textsuperscript{1118}, r\textsubscript{MN}TM6b, and r\textsubscript{MN}TM6b flies were dissected. RNA was extracted with RNeasy kit (Qiagen) following manufacture instructions, and was treated with on-column DNase digestion (Qiagen). All samples were diluted to 20ng/ul in 55ul volume with H\textsubscript{2}O, out of which 3.5ul was used for quality control using Bioanalyzer (Duke Microarray Core Facility). The concentrations were measured again with Qubit 2.0 (Life Technologies), and 700ng RNA was diluted to 50ul total volume with H\textsubscript{2}O. RNA sequencing libraries were prepared with TruSeq Stranded mRNA Sample Prep Kit (Illumina) following the manufacture instructions. For the RNA fragmentation step, 94°C, 2min was used with the intention to obtain a median size ~185bp. PCR amplification was done with 15 cycles. A total of 24 multiplexed libraries (barcoded) were accessed for quality and mixed altogether before separating to two identical pooled libraries, which are subject to cluster generation followed by Illumina 50bp paired-end sequencing by UNC High-Throughput Sequencing Facility (HTSF).

\textbf{3.4.3 RNAseq analysis}

\textit{Drosophila melanogaster} transcriptome (r5.57) was downloaded from flybase and bwa indexed was created with bwa-0.7.8. Each sequencing file was aligned to the transcriptome, and .sam files for each sample were generated by putting two alignments from both reads together. At least over 80% of the total reads were able to align to the reference. After that, count tables were made for each sample with a customized python
script (Corbin Jones), and further consolidated into a matrix containing transcript ID and read counts from all genotypes for each stage with a Ruby script (Corbin Jones). These matrices were used as inputs for differential expression analysis using customized DESeq2 R script. The sequencing data for each stage was treated individually. Read counts were normalized to allow direct pairwise comparisons between genotypes by the built-in algorithms of DESeq2 package, which is based on a negative binomial model (Love et al., 2014). Once a short list of potential targets was generated (also see 3.2.1), each category from the Venn diagram was run on functional clustering using DAVID, and particular attention was paid to genes falling into intersections. (Huang et al., 2009a, 2009b). Some known developmentally regulated genes and genes that are misregulated across all stages were listed for further analysis (Table 1).

3.4.4 Immunohistochemistry

Samples were fixed with 4% paraformaldehyde, washed with phosphate buffer with 0.2% Triton X-100, and staining as previously described (Cayirlioglu et al., 2008). Primary antibodies were used in the following dilutions: rabbit α-GFP 1:1000 (Invitrogen), chicken α-GFP 1:700 (Aves Labs), rat α-Ncad 1:20 (Developmental Studies Hybridoma Bank), mouse α-Bruchpilot 1:20 (Developmental Studies Hybridoma Bank), mouse α-CD2 1:1000 (Serotec), mouse α-Dac 2-3 1:20 (Developmental Studies Hybridoma Bank), rabbit α-β galactosidase 1:800 (Invitrogen), mouse α-β galactosidase 1:800 (Promega), rat α-Bab2 1:1500 (Frank Laski), rabbit α-Bar-H1 1:100 (Tetsuya Kojima).
The following secondary antibodies were used: Alexa 488 goat α-rabbit 1:1000, Alexa 488 goat α-chicken 1:1000, goat α-mouse-Cy3 1:100, goat α-rat-Cy3 1:200, goat α-rabbit-Cy3 1:200, Alexa 568 goat α-mouse IgG highly cross-adsorbed 1:300, Alexa 647 goat α-rat 1:200, Alexa 633 goat α-mouse 1:200, Alexa 647 goat α-mouse 1:200. Confocal images were taken by an Olympus Fluoview FV1000 or Zeiss LSM 510 (Light Microscopy Core Facility).

### 3.4.5 Real-time RT-PCR

Antennae from approximately 50 flies were dissected for each genotype and at least three biological replicates were analyzed for each genotype. RNA was extracted with an RNeasy kit (Qiagen), treated with on-column DNase digestion (Qiagen), and then reverse transcribed into cDNA using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen). qPCR was performed using the FastStart Universal SYBR Green Master Mix (Roche) or the FastStart Essential DNA Green Master Mix using standard protocol. Expression for each gene was analyzed in triplicate. Ct values were used to calculate dilution factors for each gene based upon standard curves created for each gene. Dilution factors were then normalized to the average factor of all ORs tested.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Or67d-qPCR-F</td>
<td>GCATCAGCTGTATACTAGAATGCTT</td>
</tr>
<tr>
<td>Or67d-qPCR-R</td>
<td>GGGCCAGGCTTTCATAAAGAT</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Or23a-qPCR-F</td>
<td>ACTGTACCTGATCTCCGAGC</td>
</tr>
<tr>
<td>Or23a-qPCR-R</td>
<td>GTCACATCGAGTAATCTATACAGCG</td>
</tr>
<tr>
<td>Or2a-qPCR-F</td>
<td>CCTTCTACGATTGCAACTGGAT</td>
</tr>
<tr>
<td>Or2a-qPCR-R</td>
<td>AACCTCATGACCTGCTCGAAG</td>
</tr>
<tr>
<td>Or47b-qPCR-F</td>
<td>CAAATCTCAGCCTTCTGCGG</td>
</tr>
<tr>
<td>Or47b-qPCR-R</td>
<td>GATACTGGCACAGCAAACTCA</td>
</tr>
<tr>
<td>Gr21a-qPCR-F</td>
<td>CCAACATGTACGGCATGTACT</td>
</tr>
<tr>
<td>Gr21a-qPCR-R</td>
<td>ACAGACCCACCTCCTTGTA</td>
</tr>
<tr>
<td>Or92a-qPCR-F</td>
<td>TGATATCTCTCAAGCTCTCGGAT</td>
</tr>
<tr>
<td>Or92a-qPCR-R</td>
<td>TAGGCGGTCTTATAGAGGC</td>
</tr>
<tr>
<td>Or85a-qPCR-F</td>
<td>GAGCGACGATACAGAAACCAC</td>
</tr>
<tr>
<td>Or85a-qPCR-R</td>
<td>AAGCGAACTTGCCATCTTTAT</td>
</tr>
<tr>
<td>Or22a-qPCR-F</td>
<td>CCGATCGTCGCTACAAATCC</td>
</tr>
<tr>
<td>Or22a-qPCR-R</td>
<td>ATGCCAGCTTCACCATTAGCC</td>
</tr>
<tr>
<td>Or56a-qPCR-F</td>
<td>TTGACAGTTGCGGCTTCCAAG</td>
</tr>
<tr>
<td>Or56a-qPCR-R</td>
<td>AAGCAAGGCTCAGTTCATCG</td>
</tr>
<tr>
<td>Or82a-qPCR-F</td>
<td>CTCCAATTGGCATCTGGCTT</td>
</tr>
<tr>
<td>Or82a-qPCR-R</td>
<td>GTTCGACAGATCCCAACGAAA</td>
</tr>
<tr>
<td>Or49b-qPCR-F</td>
<td>ACAAGGTGGGAAAAGTTGATG</td>
</tr>
<tr>
<td>Or49b-qPCR-R</td>
<td>AATGGCAGGACTCTTTCGCT</td>
</tr>
<tr>
<td>Or98a-qPCR-F</td>
<td>ATTCAAGCCGCAGTTACAAGT</td>
</tr>
<tr>
<td>Or98a-qPCR-R</td>
<td>TGCCAGCTTAGCCACCTTAAT</td>
</tr>
<tr>
<td>Or9a-qPCR-F</td>
<td>GAGACCAACTGGACCAGCTT</td>
</tr>
</tbody>
</table>

143
<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Or9a-qPCR-R</td>
<td>GAACAATCGTCGAGAAGGTGG</td>
</tr>
<tr>
<td>Or67b-qPCR-F</td>
<td>AGATCGTTTGCATGCTGTT</td>
</tr>
<tr>
<td>Or67b-qPCR-R</td>
<td>GCCGGGTAAGTCAAGGTCAT</td>
</tr>
<tr>
<td>Or67a-qPCR-F</td>
<td>ATTCAAGCCCATAACGCACC</td>
</tr>
<tr>
<td>Or67a-qPCR-R</td>
<td>TCATCAGATCAGTGAGTCAAGT</td>
</tr>
<tr>
<td>IR31a-qPCR-F</td>
<td>CGAGATCTGTGATCTGCATG</td>
</tr>
<tr>
<td>IR31a-qPCR-R</td>
<td>CCTGGGCATTACACATAGCTG</td>
</tr>
<tr>
<td>IR41a-qPCR-F</td>
<td>CCAAATGTGATTCATCTGCAC</td>
</tr>
<tr>
<td>IR41a-qPCR-R</td>
<td>ACCACGATACATAGCTCCAA</td>
</tr>
<tr>
<td>Or35a-qPCR-F</td>
<td>CGACTTGGCCTTTACTACGGA</td>
</tr>
<tr>
<td>Or35a-qPCR-R</td>
<td>AGGGCTTGTGCTTCATCTCA</td>
</tr>
<tr>
<td>IR84a-qPCR-F</td>
<td>CAGTTGGTCAGGTGTGATGG</td>
</tr>
<tr>
<td>IR84a-qPCR-R</td>
<td>AAAGTGGATGTCTGGGTGTG</td>
</tr>
<tr>
<td>Or13a-qPCR-F</td>
<td>CAATCGTTACGCCAAACAAC</td>
</tr>
<tr>
<td>Or13a-qPCR-R</td>
<td>ATCGAGGTACTTTAGAATGGCCG</td>
</tr>
<tr>
<td>ACT5C-qPCR-F</td>
<td>GGCGCAGAGCAAGCGTGGTA</td>
</tr>
<tr>
<td>ACT5C-qPCR-R</td>
<td>GGCTGCACACGCAGCTCAT</td>
</tr>
<tr>
<td>GAPDH2-qPCR-F</td>
<td>CGTTCATGCCACCACCGCTA</td>
</tr>
<tr>
<td>GAPDH2-qPCR-R</td>
<td>CCACGTCATCAGCCACAA</td>
</tr>
</tbody>
</table>
3.4.6 CRISPR-based tagging of Rn

3.4.6.1 Molecular cloning

In order to make 3’ end isoform-specific (Rn E and F) tagging, we initially designed to make CRISPR flies with all three isoforms (C, E and F) tagged with either EGFP or 3XFLAG. Two cutting sites were selected, one in the last intron (referred as intron) and the other one in the intergenic region downstream of Rn coding region (referred as inter) (Figure 39). Twenty base-pair long target sites were selected using the flyCRISPR Optimal Target Finder tool on the flyCRISPR website (Gratz et al., 2014). We set the parameter for initial nucleotide selection as “All CRISPR targets”, and would simply add a G at the beginning, if it does not include one, to facilitate its transcription by the U6 promoter (as seen below by the g/c for the intergenic target we used). We then evaluated potential off-targets for all the candidates by setting “Maximum” for “stringency”, and “NGG and NAG” for “PAM”. The sites with minimal or zero potential off-targets were confirmed with no mutations in the fly stocks to which we would do injections. One site from each region was cloned into the pU6-BbsI-chiRNA plasmid following the U6-gRNA (chiRNA) cloning protocol on the same website. The final constructs are named rn-chiRNA-intron and rn-chiRNA-inter. The oligos are listed below:

rn-chiRNA-intron-sense: 5’-cttcGTGTGAAGAATCGAAGAGA-3’
rn-chiRNA-intron-antisense: 5′-aaacTCTCTTCGATTCTTCACAAC-3′

rn-chiRNA-inter-sense: 5′-cttcgATATTCCGAGACACAGGGGA-3′

rn-chiRNA-inter-antisense: 5′-aaacTCCCTGTGTCTCGGAATATc-3′

To make the repair template plasmid, we used ~1kb homology arm from either side of the cutting site. Specifically, for the fragment including the homology arm upstream of the intron cutting site (5HDR), we amplified a 1,374bp region, which contains a 1,041bp homology arm and the following 333bp right before the stop codon. A KpnI and an XhoI cutting sites were added in order for it to be cloned into the pBluescript II SK vector. Primers used:

5HDR-mCR-F: ggtaccCATCAGCGCAACAACCTGG

5HDR-mCR-R: ctcgagTCCCTTGTCCTTCCCAGGA

To protect this fragment from being cut by the intron chiRNA, we mutated 3 nucleotides in the CRISPR recognition site using a Site-Directed Mutagenesis Kit (Agilent Technologies), so that it changes from CCG tct | cttcggattcttcacaac to CCG tct | tctctattcacaac (PAM is capitalized; the presumable cutting site is denoted as “|”; mutated nucleotides are labeled in red). This mutation should not affect homologous
recombination, because the manipulation was done only in the place that would be replaced. In addition, since these mutations are immediate to the cutting site, they should effectively block chiRNA recognition, thereby protecting the template (Gratz et al., 2014; Jinek et al., 2012). We also made sure the changes would not affect splicing signals due to its proximity to the acceptor site.

To make the EGFP tagging, the coding region of EGFP with the start and stop codons was amplified from the pTGW vector flanked by XhoI and SalI sites. Primers used:

EGFP-cDNA-N: ctgagATGGTGAGCAAGGCGAG
EGFP-cDNA-C: gtcgacTTACTTGTACAGCTCGTCCATGC

A flexible (GGGGS)3 linker was incorporated between the 5HDR fragment and EGFP to facilitate protein folding. To do this, 5’ phosphorylated oligos with XhoI overhangs were synthesized and annealed:

3G4S-XhoI-F:

tcgagGGAGGAGGCGGCTCCGGAGGCGGGAGGATCCGGCGGAGGTGGCTCCG

3G4S-XhoI-R:

tcgagGGAGCCACCTCCGGCGGATCTCCGGCCTCCGGAGGCGGGCCTCCG

147
The same linker was added between the 3XFLAG tag and 5HDR, except that the 3XFLAG coding sequence and the linker were synthesized as a single piece flanked by XhoI overhangs. These two 5’ phosphorylated oligos were annealed (3XFLAG coding region was optimized for fly codon usage and is labeled in red):

3G4S-3XFLAG-F:

tcgagGGAGGAGGCGGTCCCGAGGCGGAGGATCCGCGGAGGTGGCTCCGACTA CAAAGACCATGACGGTATTATAAGATCATGACATCGATTACAAAGGATGACG ATGACAAAGTGAG

3G4S-3XFLAG-R:

tcgcTTACTTGTCACTCGTCATCCTTGTAAATCGATGTACATGCTTATTTATAATC ACCGTCATGGTGTTGTAGTCGGACCCACCTCCGCGGATCCTCCCCCCCTCCGGAG CCGCCTCCTCCc

For the homology arm downstream of the intergenic cutting site (3HDR), a 1,388bp fragment immediately starting from the cutting site was amplified. It was flanked by NotI and SacI sites. Primers used:

3HDR-rmCR-F: gcggccgCCTGTGTCTCGGAATATCATTTTAGG
3HDR-mCR-R: \textit{gagctcATTGCAAGGGTCTGAAC}T

The leftover sequence between the 5HDR and 3HDR fragments (mainly 3’ UTR of Rn) was cloned as it is with SalI and HindIII on either side:

3UTR-rnCR-F: \textit{gtcgacCTAGGGGCCTACTTCTAGATGG}

3UTR-rnCR-R: \textit{aagcttGGAAGGATAACATTTAATTTACTTTATTACG}

To aid the genetic screening process, we added a 3XP3 DsRed selectable marker which can be excised eventually by crossing to a HS-Cre line (Gratz et al., 2014; Horn et al., 2000). This cassette was cut out from the pHD-DsRed-attP vector using SpeI and NotI restriction enzymes. All 6 fragments for the Rn-EGFP construct (or 5 fragments for the Rn-3XFLAG construct) were sequentially inserted into the pBluescript II SK vector. The final constructs are named pBS-RnEGFP-crHDR_Cas9 (rnCR-EGFP for short) and pBS-Rn3XFLAG-crHDR_Cas9 (rnCR-FLAG for short) (Figure 40). They were fully sequencing before being injected to vas-Cas9.RFP(-)/FM7a embryos, which has a wild type third chromosome (see below. Also note that Rn is on the third chromosome.). These two injections yielded Rn-EGFP, and Rn-3XFLAG lines, in which all three isoforms are tagged.
Our initial strategy to make E/F isoform-specific tagging includes injecting similar constructs into the vas-Cas9.RFP(-); roe<sup>3</sup>/TM3 flies. roe<sup>3</sup> contains an premature amber stop codon in the C isoform, and therefore only the E and F isoforms will be fully translated with the tags (St Pierre et al., 2002). The Constructs were made in the same way as described above, except that the relevant fragments were amplified from the roe<sup>3</sup> chromosome. From screening, we were able to get correctly targeted RnE/F-3XFLAG line.

After we failed to obtain a correct target event for the E/F-specific EGFP tagging, we decided to generate a stop codon at the beginning of the C isoform coding region in the Rn-EGFP line we just made by using the CRISPR methodology. Two cutting sites were selected in the first exon for the C isoform using the same criteria, and the oligos were annealed and cloned as described above. This construct is named as rnC-chiRNA-5’ and rnC-chiRNA-3’. The sequences for the oligos:

rnC-chiRNA-5’-F: `tttcGGCGGAATCTCCCCAATCAG`

rnC-chiRNA-5’-R: `aaaaCTGATTGGGGAGATTCCGCC`

rnC-chiRNA-3’-F: `tttcGATCCGGGACTTGCGGCCCC`

rnC-chiRNA-3’-R: `aaaaGGGGCCGCAAGTCCCGGATC`
Again, we used ~1kb homology arm from either side of the cutting site. For amplifying the 5HDR fragment, a 10bp sequence (labeled in red) between the SpeI cloning site and 5’ end of the cut leftover was included in the reverse primer. Along with SpeI site, this 16bp fragment encodes 4 stop codons (highlighted in yellow) in all three frames, including 2 stop codons in the frame that we intended to manipulate. This homology arm is flanked by KpnI and SpeI. Primers used:

**rnC-5HDR-KpnI-F:** `ggtaaccATGTCTGCGCCTGAATGACT`

**rnC-5HDR-SpeI-R:** `tagtaatagtaCAGCGGCGAGTTGTGATGGTAG`

To make the 3HDR, a 1,148bp fragment flanked by NotI and SacI was amplified using the following primers:

**rnC-3HDR-NotI-F:** `gcgGCCGCAAGTCCCCGATCTAC`

**rnC-3HDR-SacI-R:** `gagctcGATGCTGCACTTGTACGGA`

Finally, the sequences on both sides of 3XP3 DsRed from the rnCR-EGFP construct (described above) were replaced by rnC-5HDR and rnC-3HDR, respectively, using the KpnI/SpeI and NotI/SacI sites. This yields the rnC-STOP construct, which was used as the repair template to be injected into the vas-Cas9.RFP(-); rn-EGFP DsRed(-)
Because the 3XP3 DsRed cassette was excised from the injected embryos after the initial screening, the newly incorporated 3XP3 DsRed cassette can be used again as a selectable marker for efficient genetic screening.

3.4.6.2 Embryos Injection

The injection mixtures were prepared as instructed on the flyCRISPR website. Specially, 100 ng/µL each pU6-BbsI-chiRNA and 500 ng/µL repair template were mixed and injected into embryos following the scheme outlined in Table 3.

3.4.6.3 Targeted event screening

Single G₀ founders were crossed to double balancer flies, and G1 flies were screened for the presence of 3XP3 DsRed (Figure 42). For each founder line that yielded DsRed (+) progenies, 20 individual G1 flies were crossed to the double balancer stock, unless fewer flies were produced. DNA of G2 larvae or adult flies from each G1 lineage was extracted and PCR screened for the presence of EGFP/3XFLAG. Meanwhile, G2 larvae were dissected to check for live GFP (or stained with FLAG antibody) in the eye-antennal discs. The targeted area was sequenced for the G1 lineages that show positive PCR results and EGFP/FLAG signals to confirm that the targeted event is 100% correct. The potential off-target regions were sequenced for all correctly targeted lines. A couple of healthy G2 lines that passed all these tests were crossed to the HS-Cre line to excise the 3XP3 DsRed cassette, from which we obtained the lines with Rn (or E/F isoform only) tagged by EGFP or 3XFLAG in the endogenous locus (Figure 41).
3.4.7 Chromatin immunoprecipitation

This procedure is modified based on a published protocol (Jusiak et al., 2014).

For each genotype, around 400 disc-brain complexes from wandering third instar larvae were dissected in freshly made ice-cold ChIP dissection buffer (1X PBS with 1 mM PMSF and protease inhibitor). The samples were immediately cross-linked with 1% formaldehyde in dissection buffer for 10 min at room temperature. To quench cross-linking, glycine was added to 125 mM final concentration, and the samples were incubated for 5 min followed by washing with the dissection buffer for 4 times. The eye-antennal discs were then dissected out from the fixed disc-brain complexes in cold dissection buffer. The discs were homogenized in 450 ul ChIP lysis buffer with a pestle (50 mM K-HEPES, pH 7.8, 140 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.5% SDS, 1 mM PMSF, protease inhibitor), then passed through a 25-gauge needle for 10 times, and a 27-gauge needle for 10 times. The samples were incubated for 20 min at 4°C on a nutator. The solution was split into 2 TPX tube (210 ul each), and sonicated in a Bioruptor machine for 13 min (high frequency; 30 sec ON/30 sec OFF). Cell debris was removed by centrifuge 13,200 rpm, 4°C, 10 min. Supernatant from both tubes was combined, and 30 ul was taken to check sonication efficiency. A smear with 100-1000 bp fragments and an average size around 400 bp indicates sufficient sonication. The rest of the sonicated chromatin was diluted with ChIP dilution buffer (the same as lysis buffer except no SDS) to increase the volume to 2.07 ml. The diluted
chromatin was pre-cleared with pre-washed Dynabeads Protein G (Life Technologies) for 1hr at 4°C on a nutator. The pre-cleared chromatin was split into 2 tubes (1ml/tube), and another 20ul (2%) was saved as input and stored at -20°C. 5ug Anti-GFP antibody (Ab290) or equal amount of normal rabbit IgG were added to either tube, followed by overnight incubation at 4°C. Pre-washed beads were added to both tubes, and the samples were incubated for 2 hours at 4°C on a nutator. Beads was briefly rinsed with wash buffer I (50mM K-HEPES, pH7.8, 140mM NaCl, 1mM EDTA, 1mM EGTA, 1% Triton X-100, 0.1% Na-deoxycholate, 0.1% SDS), and washed 1X with wash buffer I, 1X with wash buffer II (the same as buffer I, except that NaCl is 500mM), 1X with wash buffer III (250mM LiCl, 0.5% Igepal CA-630, 0.5% Na-deoxycholate, 1XTE), 2X with the TE buffer, at 4°C, 5min/each wash. The chromatin was eluted 2X with pre-warmed elution buffer (1% SDS, 100mM NaHCO3). For each elution, beads were incubated in 100ul solution for 10min at 65°C, with gentle vortexing every 2-3 min. Two elutes were combined, and elution buffer was added to the input. To reverse cross-link, 5M NaCl was added to each tube, followed by overnight incubation at 65°C. The ChIP-ed DNA was treated with RNase and proteinase K, and extracted by PCR purification columns with 50ul H2O elution (Qiagen). The purified DNA was tested for enrichment of DNA fragments by qPCR. For each target gene, up to 150bp amplicons were selected every ~300bp in the first 2kb fragment upstream of the coding region. To test direct binding of Rn to the published 13bp T13 motif within the Bab2 LAE (leg and antennal enhancer) in
*vivo*, a primer pair covering this region was designed for ChIP-qPCR analysis (Baanannou et al., 2013). This result, provided being positive, would serve as a positive control. To confirm that Rn does not bind to the M1 motif upstream of *rn*-positive OR promoters, primer sets covering the motif from Or49a and Or82a promoters were designed and used in ChIP-qPCR analysis. This experiment can also be used as a negative control.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bab2_ChIP_T13_F</td>
<td>TATTTGCGTGAGCCTTC</td>
</tr>
<tr>
<td>Bab2_ChIP_T13_R</td>
<td>TAACGATTGCAGCGATTT</td>
</tr>
<tr>
<td>Or49a_ChIP_M1_F</td>
<td>TATGGTCCGATTCCTCGA</td>
</tr>
<tr>
<td>Or49a_ChIP_M1_R</td>
<td>TTCAGCGGCTCTGTATG</td>
</tr>
<tr>
<td>Or82a_ChIP_M1_F</td>
<td>CACAGTACATACAGCCATACAG</td>
</tr>
<tr>
<td>Or82a_ChIP_M1_R</td>
<td>CGCTTCCCTCTTGCTTGTT</td>
</tr>
<tr>
<td>B1_ChIP_96_F</td>
<td>CTTGACATCTATAGGAACTCCAG</td>
</tr>
<tr>
<td>B1_ChIP_96_R</td>
<td>GGTTCACGCTTTCTCCCTTT</td>
</tr>
<tr>
<td>B1_ChIP_419_F</td>
<td>GAGAGTGACAGCGATGAG</td>
</tr>
<tr>
<td>B1_ChIP_419_R</td>
<td>CGGCTGGCAACAATTT</td>
</tr>
<tr>
<td>B1_ChIP_519_F</td>
<td>GACCGACATTTTGCTATAAA</td>
</tr>
<tr>
<td>B1_ChIP_519_R</td>
<td>ATCGGCTGCATCTCCTCTC</td>
</tr>
<tr>
<td>B1_ChIP_1010_F</td>
<td>CCAAATGTCGAGCGGTATAAA</td>
</tr>
</tbody>
</table>
B1_ChIP_1010_R AGTCCCATTTTGCTTGAC
B1_ChIP_1342_F CGCTTCAGTCCGCTTTAC
B1_ChIP_1342_R TGAACCAACCTGGAGAGAA
B1_ChIP_1702_F CGGTCTATGTCTCTGTCTT
B1_ChIP_1702_R GTCAGCAGTGAAGAGTTTAG
B1_ChIP_2036_F CAGTGCAAAACAGCTACTC
B1_ChIP_2036_R CCATGTGAGGCTTCTATC
B2_ChIP_0_F CTTGTTTGTGCTGCTGTTG
B2_ChIP_0_R TTCGACGGAAACTGAAACTT
B2_ChIP_132_F GCGACAGAGAGGAAACTCA
B2_ChIP_132_R CCACCTGTGCCTGCTCTTT
B2_ChIP_399_F CCCTCAAAGATAACGAACCG
B2_ChIP_399_R CGAACTACAACCGCACAAA
B2_ChIP_728_F TGAGTTTCAAGCTGCCATAA
B2_ChIP_728_R TTGTACAGGAATGACAGAACAT
B2_ChIP_942_F CAAAGCGAGACAGAGATG
B2_ChIP_942_R GGATGCTGTCAGCCCTGT
B2_ChIP_1277_F GGAGCACGCTTTGTAGTC
B2_ChIP_1277_R CTGCTAATGTGCAAGTGA
B2_ChIP_1588_F CAGGTATTTACAGGCAATAC
B2_ChIP_1588_R CCAAGCATTTACGACAGA
B2_ChIP_1889_F CGAGAGACAGTGACAGA
B2_ChIP_1889_R GCACCAGAAATCTAAGAA

156
Note: B1 and B2 stand for Bar-H1 and Bar-H2, respectively. T13 and M1 are the motif names (Baanannou et al., 2013; Li et al., 2013). The numbers correspond to the lengths between transcription start sites (TSS) and the beginning of the amplicons.

### 3.4.8 Statistical analysis

Statistical analysis of Bab2 expression levels and qPCR results was by unpaired, two-tailed Student’s t test. For all tests, * $p<0.1$, ** $p<0.01$, *** $p<0.001$. 

| Bab1_ChIP_0_F   | CGTCTGAATTAGACTGAAATCGT |
| Bab1_ChIP_0_R   | GCACATCCACATCCACATC     |
| Bab1_ChIP_290_F | GCGACGAAGTTGCGAATAG      |
| Bab1_ChIP_290_R | GCTTCGCGTTGTCTCTCTCT    |
| Bab1_ChIP_587_F | GTTCTCCTCTCGTTCTCT      |
| Bab1_ChIP_587_R | GCAACTTCCGTTCTCTCA       |
| Bab1_ChIP_797_F | GTGCGGATTCCGACCTTAG      |
| Bab1_ChIP_797_R | ATTCCTGAAGACCACCTAGA     |
| Bab1_ChIP_1105_F| GCTTGTACATTTGTAGTAGAGC   |
| Bab1_ChIP_1105_R| AGTAATGGCTTGGAAGATAG      |
| Bab1_ChIP_1411_F| CGAATTGGCTCTCTCTCTCTCT   |
| Bab1_ChIP_1411_R| ATGCCCATCACCACCTACT       |
| Bab1_ChIP_1755_F| CCGGACGAAGCTACATATCA     |
| Bab1_ChIP_1755_R| CCATTGCGCAGTCACGAA       |
| Bab1_ChIP_1972_F| TTCAGTATCCTGTGGCATCTC    |
| Bab1_ChIP_1972_R| TGGCCATTAGAGTAGGCA        |
3.5 Acknowledgements

We owe special thanks to Tetsuya Kojima, Frank Laski, Leslie Vosshall, Barry Dickson, Richard Benton and John Carlson for providing fly stocks and antibodies. We thank Bloomington Stock Center and Drosophila Genetic Resource Center for their services. We are grateful to Jeff Sekelsky, Kate O’Connor-Giles and Scott Gratz for sharing fly CRISPR regents and giving technical assistance. We thank Jamie Roebuck (Model System Injections) for doing the embryo injections. We acknowledge the services provided by Duke Light Microscopy Core Facility, Duke Microarray Core Facility, UNC High-Throughput Sequencing Facility. Finally, we want to thank members of the Volkan laboratory for discussions and critical reading of the manuscript.
4. Concluding Remarks and Future Directions

4.1 Conclusions

As discussed in the previous chapters, my PhD research started with discovering the critical function of Rn in diversifying new ORN fates during development (Chapter 2). Rn is part of a nested hierarchical gene network that generates new sensilla subtype fates from each parallel sensilla type lineage. Later on, we found that this hierarchical relationship is on the logical level, rather than on the molecular level. In fact, Rn is a pre-patternning factor that is expressed prior to the onsets of proneural gene expression. Therefore, it seems that the subtype differentiation by Rn occurs before sensilla type determination by proneural gene-mediated SOP selection.

We continued to uncover a molecular mechanism in which Rn and other factors pattern an initially homogenous precursor field in the developing fly olfactory system (Chapter 3). This patterning process involves morphogen gradients and cross-regulations between transcription factors (TF) that have been shown to induce proximodistal axis formation in limbs. The result is a number of intercalated TF expression domains organized in concentric rings, each of which is represented by a unique combinatorial code of TFs. We provided evidence to show that each concentric domain specifies a limited number of SOP differentiation potentials, corresponding to sensilla subtypes, which can be further separated by TF concentration gradients or some unknown factors. Therefore, the pre-patterned precursor fields serve as distinct pools of
epithelial cells from which individual SOPs, along with their sensilla type identities, are determined.

Based on these findings, we proposed a three-step mechanism, which can explain the extraordinary ORN diversity in the olfactory system. First, the imaginal disc is pre-patterned into different domains by morphogen and transcription factors; then, SOPs with different developmental potentials are selected from each domain by proneural gene expression and lateral inhibition; finally, alternate ORN fates are segregated from individual SOPs via Notch-mediated asymmetric divisions.

By comparing the ORN diversification mechanism described in this dissertation with similar events in other systems, such as fly CNS, vertebrate spinal cord and neocortical neurons, we think that this study reveals a common strategy by which a developing nervous system utilizes to generate neuronal diversity. This strategy is comprised of single-purpose logical steps, which can be reassembled to guide context-dependent neuronal diversification processes. It not only simplifies the overall developmental difficulty to generate complexity required by multiple parallel systems within an organism, but also provides an entry point from which evolutionary plasticity may arise.
4.2 Future directions

4.2.1 What are the contributions of D/V and A/P axis determination and temporal dynamics of the PD network in ORN diversification?

Although our model of sensilla fate mapping can explain the majority of sensilla subtype specification and consequent ORN diversity, it is likely incomplete. First, some sensilla subtypes within a ring are specified by the same set of factors, and need to be further differentiated. For example, it is unclear how ab1 and ab9, or all the subtypes within the center are distinguished. One possibility is that factors that set up the dorsal/ventral axis or anterior/posterior axis are used in this process. Further work should address the contribution of these axis determination events to ORN diversification. Alternatively, the temporal regulation of the same gene network may have additional functions in differentiating precursor fates. It is known that these PD genes are expressed in asynchrony during leg development, and the expression patterns of some factors, such as Dac and Bar, show dynamic changes over time (Estella et al., 2012; Kojima, 2004; Natori et al., 2012). The current study mainly focused on the later phase of the pre-patterning stage, and hence the functional relevance of the temporal aspect of the TF network deserves further investigation.

4.2.2 How is the pre-patterning information relayed to proneural gene expression?

Once the precursor field is patterned, proneural genes are turned on to generate SOPs. However, little is known about how different fates of epithelial cells are
coordinated with proper proneural gene expression. It is possible that the pre-patterning factors themselves directly activate proneural genes. One example is seen in the case of *amos* being regulated by *lozenge* (Goulding et al., 2000), although whether or not this regulation is direct is unclear. To what extent that other factors have similarly functions has not been studied. It is plausible that combinations of different factors are competent to activate either *amos* or *atonal*, or perhaps only a few factors play a dominant role in turning on the corresponding proneural gene based on the predetermined fate. Genetic analysis can be done to evaluate these possibilities, followed by biochemical assays to directly test the association between pre-patterning factors and proneural gene promoters.

**4.2.3 How are the differential potentials coded in SOPs and during divisions?**

When individual SOPs are specified before neurogenesis, they carry different developmental potentials. These potentials include the information about what types and subtypes of sensilla the SOPs are going to become, what ORN classes they will generate, and what receptor genes need to be expressed in the future. We stated that combinations of pre-patterning factors and proneural genes by and large confer all the information. However, how the information is memorized by the cell at the molecular level, and how the memory is transmitted during divisions are unknown. One simple scenario is that the same early factors maintain the deterministic functions till the final stage of ORN differentiation by serving as a bridge. They can be distributed evenly, or
unevenly to the daughter cells during asymmetric divisions, so that the potentials are passed down through lineages. Consistent with this notion, factors such as Ap, Bar are only expressed in some but not all postmitotic neurons from a sensillum. Likewise, some sensilla subtypes contain one Bab-positive and one Bab-negative neuron in the adult stage. Interestingly, in most cases, it is the Notch-off neuron between the two that expresses Bab, suggesting Notch signaling may be involved in segregating these early factors during cell divisions, although the developmental relevance of such separation is unclear. Nonetheless, assuming the postmitotic neurons inherit some of these “bridge” factors as a mechanism to retain part of patterning information, it still does not answer how the information provided by other transient factors is embodied and transmitted at the molecular level. Therefore, it is unlikely that passing down the same factors would be the sole mechanism for the question. Besides, it also seems that this passive transmission would put a lot of burdens on the last stage of development, making it hard to imagine how a small number of TFs alone could confer the developmental specificity and at the same time handle many other responsibilities of making a neuron in a very narrow window.

Another hypothesis is that the early patterning factors initiate a series of downstream events that are then relayed to a later stage regulatory network, which can directly regulate OR expression and ORN specification. This idea liberates the precursors from constantly requiring the early factors, and relies on the gradual
changing of transcriptome as a way to explain information transmission. Consistent with this hypothesis, our data show that the transcription profiles across developmental stages are highly dynamic. In addition, many developmentally regulated genes show differential expressions in \textit{rn} mutants even in early stages, suggesting that pre-patterning factors, such as \textit{rn}, actively regulate other signaling pathways that likely amplify or solidify patterning information. More bioinformatic work should be done to analyze the dynamic patterns of gene clusters, accompanied by functional analysis of individual pathways to assess their involvement in intermediate precursor and ORN fate specification.

A third model takes epigenetic mechanisms into account. Epigenetic modulations of chromatin status have been shown to play an important role in numerous developmental processes, including the development of the olfactory system (Alexander and Lomvardas, 2014; Endo et al., 2011; Lyons et al., 2013; Magklara et al., 2011; Sim et al., 2012). It is possible that the pre-patterning TFs recruit epigenetic modifying factors to change the open and close states of the chromatin around genes critical for different fates. These modifications can be inherited during cell divisions, and affect the genomic accessibility of later factors. For example, olfactory receptor gene loci may be progressively modified, which influences the effectiveness of late receptor gene regulatory network depending on the context. This idea can explain why many postmitotic factors have broad expression patterns, and why mutations in these factors
often lead to inactivation of OR genes, rather than fate switching. To test this hypothesis, precursors with different developmental potentials and terminally differentiated neurons can be sorted and subject to epigenetic assays, such as ChIP-seq on histone marks, and ATAC-seq. Then the dynamic changes of epigenetic profiles can be correlated with transcription profile data to uncover potentially important events that take place at the chromatin level.

It is worth mentioning that these three mechanisms discussed above are not necessarily mutually exclusive. Rather, it is likely that the multiple modes of regulation collaboratively participate in different aspects of ORN diversification. Future work should also address the relationship between the transcriptional and epigenetic mechanisms, provided that both of them are involved.

As a matter of fact, understanding how a handful of master regulators determine the complicated process of cell differentiation is a major question in many other fields, particularly in stem cell biology. With the knowledge we added in this study, the fly olfactory system now should be an excellent model to approach such a question. This system is equipped with sufficient complexity, and yet simple enough to make tracking individual early and late fates possible. It also has a defined developmental end, marked by the expression of a unique receptor, which dramatically simplifies the analysis by having only one focal point per cell type. I hope that the advantages of future studies at this frontier will advance the stem cell biology field and beyond.
4.2.4 How is the olfactory system evolved?

The olfactory system is fast evolving, as manifested by the rapid expansions and alterations of olfactory receptor genes across genomes. One advantage using the fly olfactory system to study evolution is that sibling *Drosophila* species are fully sequenced and relatively well annotated. Differences in the ratios of distinct sensilla subtypes between species have been reported (Dekker et al., 2006), calling for the need to understand the underlying mechanisms that drive these evolutionary changes. As seen in the drastic shifting of sensilla subtypes by simple genetic manipulations, it is plausible to think that species-dependent phenotypic changes may be caused by naturally occurring modifications in cis-regulatory elements of pre-patterning factors and the ensuing conversions of precursor domains. In the future, cross-species transcriptome analysis at different developmental stages should be conducted. These datasets, in combination with the cis-regulatory element analysis, may provide some mechanistic views on the molecular underpinnings of olfactory system evolution.

4.2.5 Can we mathematically model the ORN diversification process?

As described in this dissertation, the ORN diversification process takes multiple steps, each of which may involve many factors. These factors are spatio-temporally regulated, and in some cases, the levels of expression are also important. Now we know a lot about the regulatory relationships between them, and we start to know how one event leads to another. In the future, we should analyze this information more carefully.
at the quantitative level to obtain precise data about the timing and levels of gene expression. With these data, we can start to build a mathematical model to simulate the whole process. Such a model would not only help us visualize how ORN diversification progressively occurs, but also provide experimentally testable predictions, which in turn can be used to refine the model. This would hopefully lead to the true understanding of the system.
Appendix A

Genotypes for fly genetics in Chapter 2.4.1 and Chapter 3.4.1:

Figure 9. (A) +/-: UAS-CD8 GFP/+; Or67d^GAL4^{FRT82} rntot/+; -/-: UAS-CD8 GFP/+; Or67d^GAL4^{FRT82} rntot/rntot.

Figure 10. (A) UAS-CD8 GFP/+; rns9^GAL4/+. (B) +/-: UAS-FLP/+; tub-FRT-nYFP-FRT-CD2 RFP rns9^GAL4/+; -/-: UAS-FLP/+; tub-FRT-nYFP-FRT-CD2 RFP rns9^GAL4/rntot.

Figure 11. (A) OR-CD8 GFP/+; rntot/+; -/-: OR-CD8 GFP/+; rntot/rntot. (B) (bottom left), +/-: Or47b-CD8 GFP Or88a-GAL4 UAS-CD2/Or47b-CD8 GFP Or88a-GAL4 UAS-CD2; rntot/+; -/-: Or47b-CD8 GFP Or88a-GAL4 UAS-CD2/Or47b-CD8 GFP Or88a-GAL4 UAS-CD2; rntot/rntot. (B) (bottom right), +/-: Or47a sytGFP Or47b sytGFP Gr21a sytGFP/+; rntot/+; -/-: Or47a sytGFP Or47b sytGFP Gr21a sytGFP/+; rntot/rntot. (C) +/-: Or47b-CD8 GFP/UAS-Redstinger; rns9^GAL4/+. -/-: Or47b-CD8 GFP/UAS-Redstinger; rns9^GAL4/rntot.

Figure 12. (A) and (B) +/-: IR/OR-GAL4 UAS-CD8 GFP (or OR-CD8 GFP)/+; rntot/+; -/-: IR/OR-GAL4 UAS-CD8 GFP (or OR-CD8 GFP)/+; rntot/rntot.

Figure 13. (top) WT: amos^3/CyO; Or47b-CD8 GFP FRT82 rntot/TM6. rntot/-.: amos^3/CyO; Or47b-CD8 GFP FRT82 rntot/rntot. lz/-/-: lz^C/Y; Or47b-GAL4 UAS-CD8 GFP/+; rntot/TM6b. rntot/lz/-/-: lz^C/Y; Or47b-GAL4 UAS-CD8 GFP/+; FRT82 rntot/rntot. amos/-/-: Df(2L)M36-S6/amos^3; Or47b-CD8 GFP FRT82 rntot/TM6b. rntot/-/-: Df(2L)M36-S6/amos^3; Or47b-CD8 GFP FRT82 rntot/rntot. (bottom) WT: UAS-CD8 GFP/CyO;
Figure 16. (B) +/-: UAS-CD8 GFP/+; Or67d^{GAL4} rm/+; -/-: UAS-CD8 GFP/+; Or67d^{GAL4} rm/rn. (B’) Left: UAS-CD8 GFP/UAS-rn; Or67d^{GAL4} rm^{89GAL4}/rn^{tod}. Right: UAS-CD8 GFP/UAS-rn; Or67d^{GAL4} rm^{89GAL4}/+. (C) Or47b-CD8 GFP/+; rn/+; Or47b-CD8 GFP/+; rm/rn. (D) Or67d: (MARCM WT) eyFLP/+ or y; UAS-CD8 GFP/+; Or67d^{GAL4} FRT82 [P87E]/FRT82 GAL80. (MARCM rm/-) eyFLP/+ or y; UAS-CD8 GFP/+; Or67d^{GAL4} FRT82 rm^{tot}/FRT82 GAL80. (reverse MARCM) eyFLP/+ or y; UAS-CD8 GFP/+; Or67d^{GAL4} FRT82 [P87E]/FRT82 rm^{tot} GAL80. Or47b: (MARCM WT) eyFLP/+ or y; Or47b-GAL4 UAS-CD8 GFP/+; FRT82 [P87]/FRT82 GAL80. (MARCM rm/-) eyFLP/+ or y; Or47b-GAL4 UAS-CD8 GFP/+; FRT82 rm^{tot}/FRT82 GAL80. (reverse MARCM) eyFLP/+ or y; Or47b-GAL4 UAS-CD8 GFP/+; FRT82 [P87]/FRT82 rm^{tot} GAL80. (E) UAS-rn/+; Or47b-CD8 GFP/+; amos-GAL4/UAS-rn; Or47b-CD8 GFP/+. UAS-rn/+; Or67d^{GAL4} UAS-CD8 GFP/+; amos-GAL4/UAS-rn; Or67d^{GAL4} UAS-CD8 GFP/+; UAS-rn/+; Or2a-CD8 GFP/+; amos-GAL4/UAS-rn; Or2a-CD8 GFP/+.

Figure 17. (A) UAS-CD8 GFP/+; rm^{89GAL4}/+; (B) +/-: UAS-FLP/+; tub-FRT-nYFP-FRT-CD2 RFP rm^{89GAL4}/+. -/-: UAS-FLP/+; tub-FRT-nYFP-FRT-CD2 RFP rm^{89GAL4}/rn^{tod}.

Figure 18. (A) top, +/-: Or47b-GAL4 UAS-Redstinger/Or65a-CD8 GFP; rm^{tod}/rm^{tot}. bottom, +/-: Or47b-GAL4 UAS-
CD2/Or65a-CD8 GFP; rn\textsuperscript{tod}/+. -/-: Or47b-GAL4 UAS-CD2/Or65a-CD8 GFP; rn\textsuperscript{tod}/rn\textsuperscript{tot}. (B) +/-: Or13a-CD8 GFP/+; rn\textsuperscript{tod}/+. -/-: Or13a-CD8 GFP/+; rn\textsuperscript{tod}/rn\textsuperscript{tot}. (C) +/-: UAS-CD8 GFP/+;

IR92a-GAL4 FRT82 rn\textsuperscript{tot}/+. -/-: UAS-CD8 GFP/+; IR92a-GAL4 FRT82 rn\textsuperscript{tot}/rn\textsuperscript{tod}. +/-: IR84a (or IR76a)-GAL4/UAS-CD8 GFP; rn\textsuperscript{bod}/+. -/-: IR84a (or IR76a)-GAL4/UAS-CD8 GFP;

rn\textsuperscript{tod}/rn\textsuperscript{tot}. (D) +/-: UAS-CD8 GFP/+; IR75a-GAL4 FRT82 rn\textsuperscript{tot}/+. -/-: UAS-CD8 GFP/+;

IR75a-GAL4 FRT82 rn\textsuperscript{tot}/rn\textsuperscript{tod}. (E) +/-: 72OK-GAL4 UAS-syTGF/+; rn\textsuperscript{tod}/+. -/-: 72OK-GAL4 UAS-syTGF/+; rn\textsuperscript{tot}/rn\textsuperscript{tod}. (F) +/-: Or35a-CD8 GFP/+; rn\textsuperscript{tod}/+. -/-: Or35a-CD8 GFP/+;

rn\textsuperscript{tot}/rn\textsuperscript{tod}.

Figure 19. +/-: OR-CD8 GFP (or OR-GAL4 UAS-CD8 GFP)/+; rn\textsuperscript{tod}/+. -/-: OR-CD8 GFP (or OR-GAL4 UAS-CD8 GFP)/+; rn\textsuperscript{bod}/rn\textsuperscript{tot}.

Figure 20. WT: UAS-CD8 GFP/CyO; IR75a-GAL4 FRT82 rn\textsuperscript{tot}/TM6b. rn/-/: FM6/Y;

UAS-CD8 GFP/+; IR75a-GAL4 FRT82 rn\textsuperscript{tot}/rn\textsuperscript{tot}. lz/-/: lz\textsuperscript{56}/Y; UAS-CD8 GFP/+; IR75a-GAL4 FRT82 rn\textsuperscript{tot}/TM6b. rn/-: lz\textsuperscript{56}/Y; UAS-CD8 GFP/+; IR75a-GAL4 FRT82 rn\textsuperscript{tot}/rn\textsuperscript{tot}.

Figure 22(A). en+/-: en\textsuperscript{1} or en\textsuperscript{54}/+; IR84aGAL4 UAS-CD8 GFP/+. en\textsuperscript{1} or en\textsuperscript{54}/+;

Or67d\textsuperscript{GAL4} UAS-CD8 GFP/+. en/-: en\textsuperscript{1}/en\textsuperscript{54}; IR84aGAL4 UAS-CD8 GFP/+. en\textsuperscript{1}/en\textsuperscript{54};

Or67d\textsuperscript{GAL4} UAS-CD8 GFP/+.

Figure 24(E), Figure 26(A). rn+/-: UAS-CD8 GFP/+; rn\textsuperscript{89GAL}/TM6b. rn/-/: UAS-CD8 GFP/+; rn\textsuperscript{89GAL}/rn\textsuperscript{tod}

Figure 26(B). apr\textsuperscript{K568}/UAS-CD8 GFP; rn\textsuperscript{89GAL}/rn\textsuperscript{tod}

Figure 26(D). rn\textsuperscript{tot}/FRT rn\textsuperscript{tot}
Figure 27. Figure 28(A) (B). Rn-EGFP.

Figure 28(C-E). ap^K568/+; Rn-EGFP

Figure 28(F). ap^K568/+ 

Figure 29. UAS-Ap or UAS-GFP/+; rn^89GAL4/+ 

Figure 30. UAS-BarH1^M13 or UAS-GFP/+; /rn^89GAL4. 

Figure 31(A) (A’) and Figure 32(A). UAS-Ap/+; Rn-EGFP/TM6b 

Figure 31(B) (B’) and Figure 32(B) (C). UAS-Ap/+; Rn-EGFP/rn^89GAL4 

Figure 33. UAS-CD8GFP/+; rn^89GAL4/UAS-BarH1^M13 

Figure 36. w^1118. bab^PR72 

Figure 38. rn^+/--: GR-GAL4/UAS-CD8 GFP; rn^tot/TM6b. rn^-/-: GR-GAL4/UAS-CD8 GFP; rn^tot/FRT rn^tot. 

Figure 44. top: Rn-EGFP/Rn-3XFLAG. bottom: UAS-CD8 GFP/+; Rn-3XFLAG/rn^89GAL 

Figure 45. top: UAS-CD8 GFP/+; Or67d^GAL4/+; UAS-CD8 GFP/+; Or67d^GAL4 Rn-EGFP/Rn-EGFP. UAS-CD8 GFP/+; Or67d^GAL4 Rn-3XFLAG/Rn-3XFLAG. bottom: Or47b-GAL4 UAS-CD8 GFP/+. Or47b-GAL4 UAS-CD8 GFP/+; Rn-EGFP/Rn-EGFP. Or47b-GAL4 UAS-CD8 GFP/+; Rn-3XFLAG /Rn-3XFLAG. 

Figure 46. Rn-EGFP/Rn-EGFP
Appendix B

Matlab code for making the developmental landscape in Chapter 2.4.7:

```matlab
function ORN_map_landscape(ORNmatrix)

%ORN_map_landscape

% Input: matrix of which tx factors are present (1 for present, 0 for
% missing)
% [rn, lz++, lz+, en, atonal, amos, dac]
% Output: graphic landscape of potential precursor cells

if nargin==0

    ORNmatrix = [1 1 1 1 1 1 1];

end

%%% initiate
% set order

amos_start = 2;

tlb_start = amos_start+1;

sb_start = tlb_start+24;

lz_start = sb_start+14;

atonal_start = lz_start+15;

% set general constants
```
LEVELS = 7;
level = zeros(1,LEVELS);
for i=0:LEVELS-1
    level(i+1) = 4*i + 2;
end
WIDTH = atonal_start+14;
LENGTH = level(LEVELS)+4;
ORNgrid = zeros(LENGTH,WIDTH);

%%% create ORNgrid

%%% atonal

% atonal
if ORNmatrix(5)
    ORNgrid = grid_update(ORNgrid, level(2), LENGTH,atonal_start,atonal_start+13,-1); % ac2
    ORNgrid = grid_update(ORNgrid, level(3), LENGTH,atonal_start+11,atonal_start+13,-1); % ac3
% rn
if ORNmatrix(1)
    ORNgrid = grid_update(ORNgrid, level(4),
LENGTH,atonal_start+3,atonal_start+9,-1); % ac 1 and 4

ORNgrid = grid_update(ORNgrid, level(5),
LENGTH,atonal_start+3,atonal_start+5,-1); % ac4

% en

if ORNmatrix(4)

ORNgrid = grid_update(ORNgrid, level(5),
LENGTH,atonal_start+7,atonal_start+9,-1); % ac1

end

end

end

% amos

if ORNmatrix(6)

% lz+

if ORNmatrix(3)

ORNgrid = grid_update(ORNgrid, level(2),
LENGTH,amos_start,amos_start+52,-1); % trichoid

ORNgrid = grid_update(ORNgrid, level(3),
LENGTH,lz_start+11,lz_start+13,-1); % at2

% rn
if ORNmatrix(1)

ORNgrid = grid_update(ORNgrid, level(4),
LENGTH,lz_start+3,lz_start+9,-1); % at 1 and 3

ORNgrid = grid_update(ORNgrid, level(5),
LENGTH,lz_start+3,lz_start+5,-1); % at1

% en

if ORNmatrix(4)

ORNgrid = grid_update(ORNgrid, level(5),
LENGTH,lz_start+7,lz_start+9,-1); % at3

end

end

end

%lz++

if ORNmatrix(2)

ORNgrid = grid_update(ORNgrid, level(3),
LENGTH,tlb_start,sb_start+13,-1); % basiconic

ORNgrid = grid_update(ORNgrid, level(4),
LENGTH,sb_start,sb_start+13,-1); % small

ORNgrid = grid_update(ORNgrid, level(4),
LENGTH,tlb_start,tlb_start+22,-1); % large and thin

ORNgrid = grid_update(ORNgrid, level(5),

175
LENGTH, tlb_start, tlb_start+6,-1); % 2 and 3

ORNgrid = grid_update(ORNgrid, level(6),
LENGTH, tlb_start+4, tlb_start+6,-1); % ab2

% rn

if ORNmatrix(1)

ORNgrid = grid_update(ORNgrid, level(6),
LENGTH, sb_start+3, sb_start+9,-1); % 7 and 10

ORNgrid = grid_update(ORNgrid, level(7),
LENGTH, sb_start+3, sb_start+5,-1); % ab7

ORNgrid = grid_update(ORNgrid, level(7),
LENGTH, sb_start+7, sb_start+9,-1); % ab10

ORNgrid = grid_update(ORNgrid, level(6),
LENGTH, tlb_start+12, tlb_start+14,-1); % ab5

end

% en

if ORNmatrix(4)

ORNgrid = grid_update(ORNgrid, level(5),
LENGTH, tlb_start+16, tlb_start+22,-1); % 4 and 6

ORNgrid = grid_update(ORNgrid, level(6),
LENGTH, tlb_start+16, tlb_start+18,-1); % ab4
ORNgrid = grid_update(ORNgrid, level(6), LENGTH, tlb_start+20, tlb_start+22, -1); % ab6
ORNgrid = grid_update(ORNgrid, level(5), LENGTH, sb_start+11, sb_start+13, -1); % ab8
end
% dac
if ORNmatrix(7)
ORNgrid = grid_update(ORNgrid, level(6), LENGTH, tlb_start, tlb_start+2, -1); % ab3
end
else
ORNgrid = grid_update(ORNgrid, level(3), LENGTH, tlb_start+30, tlb_start+38, -1); % at2
% rn
if ORNmatrix(1)
ORNgrid = grid_update(ORNgrid, level(4), LENGTH, tlb_start+10, tlb_start+28, -1); % at 1 and 3
ORNgrid = grid_update(ORNgrid, level(5), LENGTH, tlb_start+10, tlb_start+18, -1); % at1
% en
if ORNmatrix(4)

ORNgrid = grid_update(ORNgrid, level(5), LENGTH, tlb_start+20, tlb_start+28, -1); % at3

end

end

end

end

%% plot

surf(ORNgrid, 'EdgeColor', 'black', 'EdgeAlpha', .2)

alpha(.5)

set(gca, 'XTickLabel', [], 'YTickLabel', [], 'ZTickLabel', [])

axis([0 WIDTH+2 0 LENGTH+2 -5 0])

view(191,68)

colormap(flipud(gray))

% label precursers cells

level=level+2;

rn_color=[.6 0 .8];
atonal_color = [1 1 0];
lz_color = [0 .2 .8];
amos_color = [0 1 0];
en_color = [.9 .2 .2];
dac_color = [1 .5 0];
unknown_color = [0 0 0];

% atonal
if ORNmatrix(5)
    text(atonal_start+9, level(2), -9, 'coeloconic', 'Color', atonal_color)
    text(atonal_start+6, level(3), -9, 'ac2', 'Color', atonal_color)
    text(atonal_start+2, level(7), -9, 'ac2', 'Color', atonal_color)
    text(atonal_start+13, level(7), -1.9, 'ac3', 'Color', unknown_color)

% rn
if ORNmatrix(1)
    text(atonal_start+5, level(7), -2.9, 'ac4', 'Color', unknown_color)
    text(atonal_start+8, level(4), -1.9, 'ac1&4', 'Color', rn_color)

% en
if ORNmatrix(4)
    text(atonal_start+9, level(7), -2.9, 'ac1', 'Color', en_color)
end
end
end

% amos

if ORNmatrix(6)
  text(amos_start+30, level(2), -9, 'trichoid', 'Color', amos_color)

if ORNmatrix(3)
  text(lz_start+6, level(3), -9, 'at4', 'Color', amos_color)
  text(lz_start+2, level(7), -9, 'at4', 'Color', amos_color)
  text(lz_start+13, level(7), -1.9, 'at2', 'Color', unknown_color)

% rn

if ORNmatrix(1)
  text(lz_start+5, level(7), -2.9, 'at1', 'Color', unknown_color)
  text(lz_start+8, level(4), -1.9, 'at1&3', 'Color', rn_color)

% en

if ORNmatrix(4)
  text(lz_start+9, level(7), -2.9, 'at3', 'Color', en_color)

end
end

180
% lz++

if ORNmatrix(2)

text(tlb_start+20, level(3), -1.9, 'basiconic ', 'Color', lz_color)

text(sb_start+10, level(4), -2.9, 'small basiconic ', 'Color', unknown_color)

text(sb_start+6, level(5), -2.9, 'ab9', 'Color', dac_color)

text(tlb_start+18, level(4), -2.9, 'thin and large basiconic ', 'Color', unknown_color)

    text(tlb_start+12, level(5), -2.9, 'ab1', 'Color', dac_color)

    text(tlb_start+10, level(7), -2.9, 'ab1', 'Color', dac_color)

    text(tlb_start+5, level(5), -3.9, 'ab2&3', 'Color', unknown_color)

    text(tlb_start+6, level(7), -4.9, 'ab2', 'Color', unknown_color)

% rn

if ORNmatrix(1)

    text(sb_start+8, level(6), -3.9, 'ab7&10', 'Color', rn_color)

    text(sb_start+5, level(7), -4.9, 'ab7', 'Color', unknown_color)

    text(sb_start+9, level(7), -4.9, 'ab10', 'Color', unknown_color)

    text(tlb_start+14, level(7), -3.9, 'ab5', 'Color', rn_color)

end
% en

if ORNmatrix(4)

    text(tlb_start+20, level(5), -3.9, 'ab4&6', 'Color', en_color)

    text(tlb_start+18, level(7), -4.9, 'ab4', 'Color', unknown_color)

    text(tlb_start+22, level(7), -4.9, 'ab6', 'Color', unknown_color)

    text(sb_start+13, level(7), -3.9, 'ab8', 'Color', en_color)

end

% dac

if ORNmatrix(7)

    text(tlb_start+2, level(7), -4.9, 'ab3', 'Color', dac_color)

end

else

    text(tlb_start+16, level(3), -9, 'at4', 'Color', amos_color)

    text(tlb_start+5, level(7), -9, 'at4', 'Color', amos_color)

    text(tlb_start+35, level(7), -1.9, 'at2', 'Color', unknown_color)

% rn

if ORNmatrix(1)

    text(tlb_start+15, level(7), -2.9, 'at1', 'Color', unknown_color)

    text(tlb_start+18, level(4), -1.9, 'at1&3', 'Color', rn_color)

end

% en
if ORNmatrix(4)

    text(tlb_start+25, level(7), -2.9, 'at3', 'Color', en_color)

end

end

end

end

end

end

end

set(findall(gcf,'type','text'),'fontSize',20,'fontWeight','bold')\
References


Estella, C., Voutev, R., and Mann, R.S. (2012). A Dynamic Network of Morphogens and Transcription Factors Patterns the Fly Leg (Elsevier Inc.).


Biography

I, also known by my nickname Tristan, was born in the city of Shenyang, Liaoning, China on June 8, 1983, the single child of Shili Li and Ying Guo. I obtained a Bachelor of Science degree from the College of Biological Sciences at China Agricultural University in 2006. After that, I worked at National Institute of Biological Sciences at Beijing as a technician in the laboratory of Dr. Ligeng Ma for two years, during which I studied plant development. I came to the United States in July 2008, starting to pursue a doctoral degree in the Department of Biology at Duke University. I officially joined the Volkan laboratory in the December of 2009, when I began to study the development of the olfactory system in Drosophila. During my PhD period, I published a first author paper in the journal of Current Biology, under the title of “Combinatorial rules of precursor specification underlying olfactory neuron diversity.” At the time this dissertation is being written, I am preparing a manuscript based on the third chapter of this document, in which I will be the leading author of the collaborated work with Scott Barish. I was the recipient of Ray J. Tysor Graduate Fellowship, Departmental Semester Fellowship, Howard Hughes Vertical Integration Partnership Program fellowship, Graduate School Travel Award, Biology Grant-in-Aid Award and Biology Graduate Awards. I have been a teaching assistant for 7 semesters at Duke and trained a number of undergraduate and graduate students in the lab. I held the Certificate in College Teaching from Duke University.