**Title**
The regulatory architecture of olfactory receptor gene enhancers



**Author**
Markenscoff-Papadimitriou, Eirene

**Publication Date**
2014

Peer reviewed|Thesis/dissertation

The regulatory architecture of olfactory receptor gene enhancers

by

Elena Markenscoff-Papadimitriou

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

[redacted]

*To my loving mentor Sydney Kustu*

**Acknowledgments**

I am incredibly lucky to have been shepherded across the rocky island terrain of graduate school by Stavros. Stavros makes biology –the intractable problems, impossible experiments, and endless hours at the bench – effortless, fun, and exciting. He is an example to me on how to be an accomplished scientist and a pretty chill dude at the same time. I continue to be charmed and inspired by him. His confidence and optimism about science is overflowing and I feel like I have drunk enough of it to keep me going for the rest of my life ("That's what she said!").

I also have to give credit to Stavros for attracting some truly amazing people to his lab. I learned a lot from these people and will remember our time together as some of the most fun and formative years of my life. Dave, Brad, and Josie taught me so much about genetics, bioinformatics, and molecular biology, respectively, that the end result of my PhD would have been impossible without their guidance and knowledge. Angeliki Magklara was like a mother to us, and she did some very important work that changed the way we understand olfactory receptor gene regulation. Colleen and Kevin made some very important contributions to my project and I am indebted to them as well. Tracie Goh provided excellent technical support, and having her working on this project by my side has been invaluable.

I am also very lucky to have gotten to know and work with the people in the Ahituv lab. Nadav was always ready to do whatever he can to further my project, and I am very grateful to him. Karl was the ideal collaborator who injected thousands of zebrafish embryos for me, and I am very thankful for his important contribution to this work.

Other people at UCSF have shaped my time here, like my friends Jessica Tollkuhn and Dev Manoli. John Rubenstein was always available for me –except when he was away traveling the world - and provided the moral support, and comic relief, to get me through the PhD.

Jesse Gray and Michael Greenberg introduced me to neurobiology, and I am grateful to them for teaching me so much in a short amount of time, encouraging me to do a neuroscience PhD, and being such inspiring mentors.

I was raised by my mother Xanthippi who is a formidable role model of a successful woman scientist, and whose wisdom and love and encouragement has lighted my way through grad school. My father Christos is also an inspiration to me as a brilliant scientist with a bottomless curiosity and optimism. He has been my best friend and partner in crime in this endeavor, sharing all the lows and highs with me, listening patiently and lovingly to my every complaint, and he is the one person who I can credit for my graduation. My little sister Isabel, wise beyond her years, has been a source of comfort and joy for me. She gives so much to the world, it humbles and inspires me. My adoring boyfriend Boris and I have been on this thrilling and sometimes difficult journey together the last nine years that culminated in our PhDs. He deserves a medal for putting up with me all these years.

Ten years ago I met Sydney Kustu and the course of my life changed. I worked in Sydney's lab at Berkeley for a summer in college and she opened my eyes to the wonders of biology. She gave me the confidence and motivation I was lacking at the time to study biology. She taught me by example to be a fearless scientist, to think outside the box, to stay humble and curious in the face of biology, and to be a generous and human person.

The last ten years she was like a mom to me, encouraging me and comforting me at every stage of my personal and professional life. She also took my family and the people I love into her home and her love. I have only recently come to really appreciate the lasting contribution she made to biology with her work. The staggering number of early discoveries she made in the field of transcriptional regulation in E.coli forms the basis of our understanding of eukaryotic gene regulation, and of this thesis. I wish she could read it, and my work will always be dedicated to her.

**Contributions**

Parts of Chapters 2, Chapter 3, and Chapter 4, comprise a paper to be published in *Cell*: William E. Allen, Bradley M. Colquitt, Tracie Goh, Karl K. Murphy, Kevin Monahan, Colleen P. Mosley, Nadav Ahituv, and Stavros Lomvardas. Enhancer interaction networks as a means for singular olfactory receptor expression.

William Allen provided the analysis of the Hi-C data. Bradley Colquitt performed much of the ChIP-seq principal component analysis and DNAse footprinting analysis. Tracie Goh performed in situ hybridizations, mouse genotyping and breeding, and cloning experiments. Colleen Mosley performed the initial H3K79me3 ChIP-seq experiments. Kevin Monahan performed crosslinked ChIPs. The zebrafish enhancer screen was performed in Nadav Ahituv's lab, Karl Murphy injected constructs into zebrafish oocytes. Stavros Lomvardas oversaw the design of the experiments and the writing of the manuscripts. Eirene Markenscoff-Papadimitriou designed, performed, and analyzed all other experiments and wrote parts of the manuscripts and all of the thesis.

ABSTRACT OF THE DISSERTATION

by Eirene Markenscoff-Papadimitriou

Doctor of Philosophy in Neuroscience

University of California, San Francisco, 2014

The ability to detect odorants relies on the generation of thousands of different olfactory receptor (OR) neurons during the development of the olfactory epithelium (OE), the primary sensory organ of the mammalian olfactory system. The identity of each neuron is determined by the expression of a single OR gene from over ~1500 genes scattered across the genome. The mechanism of this choice is an intriguing and unsolved problem in the field of gene regulation. This thesis presents a large-scale investigation of the regulatory architecture of distal enhancer elements that regulate ORs and yields new insight into the mechanism of OR choice.

In the first part of the dissertation I identify 35 possible *cis*-regulatory elements for ORs and I characterize their function in the olfactory epithelium. These sequences exhibit canonical chromatin hallmarks of enhancers, as well as enrichment for the repressive histone modification H3K79me3. Transgenic analysis of these elements in zebrafish and mouse shows broad expression in olfactory neurons. Genetic deletion of one element, Lipsi on chromosome 2, demonstrates its requirement for the expression of proximal olfactory receptors in the mouse olfactory epithelium.

In the second chapter I describe experiments probing the three-dimensional nuclear organization of olfactory receptor enhancers. Intriguingly, these elements form long-range interactions in the nuclei of olfactory neurons which may be functional for OR

expression. Multiple OR enhancers from different chromosomes interact frequently with a transcribed olfactory receptor gene. Global analysis of enhancer-enhancer interactions show that interchromosomal interactions are extensive in olfactory neurons and are specific to OR enhancer sequences. Genetic disruption of these long-range enhancer interactions disrupts the singular expression of OR genes in OSNs. These observations support a model whereby the assembly of an enhanceosome consisting of enhancers from different chromosomes drives singular OR expression.

The third section of the dissertation explores the transcription factors that bind these enhancers and mediate their function in OR gene expression. Ebf4 binds an O/E-like motif on enhancers which is necessary and sufficient for enhancer activity. BPTF, part of the NURF chromatin remodeling complex, binds enhancers, is required for OR gene expression, and facilitates enhancer-enhancer interactions in olfactory neuron nuclei.

# Table of Contents

## List of Figures



### List of Tables

# Chapter 1: Introduction

**Development, gene regulation, and enhancers**

Every cell in our body contains the same genetic material and is a distant ancestor of an original fertilized egg. How is the staggering diversity of cell types generated from a common genetic code? How is an olfactory neuron, whose specialized function is to detect a particular odorant molecule in the air and relay that information to the brain, generated from the same DNA as a cardiac muscle cell whose job is to pump the heart steadily throughout life? Obviously, layers and layers of control must be in place to ensure that the appropriate genes are active in the appropriate cells at the appropriate time and place – this is what gene regulation is about. The system requires robustness, but at the same time must be sensitive to changes in the environment. During the development of the organism, cells make fate commitment decisions that alter their identity – how are these decisions made permanent? All of these requirements pose an incredible challenge to the cell which is outfitted with one genome – composed of thousands of genes which can be deployed in myriad combinations – and one mistake can be lethal or lead to anatomical and physiological abnormalities.

The answer is that nature has outsourced the problem. The human genome contains tens of thousands of enhancers, which are short sequences of DNA scattered across the genome and whose function is to control the expression of nearby or distant genes (Rada-Iglesias et al., 2011; Visel et al., 2009). Turning on a single transcription factor can thus activate a discrete set of enhancers, thus deploying a whole gene program. For genes that are involved in more delicate tasks such as finger/digit development, tens of enhancers negotiate their level of expression in assemblies they form in the nucleus (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011). Furthermore, the

genome is packaged in such a way in the nucleus so that covalent and non-covalent chemical modifications on DNA or the histone proteins that DNA is wrapped around can switch enhancers and promoters on and off. Expressing such a chromatin modifier is a good way to make a permanent change to enhancers that are ready to accept it. This kind of control is referred to as "epigenetic," wherein a non heritable change is effected without altering the DNA sequence itself so that it is interpreted differently by a cell.

**Enhanceosomes and promoter-enhancer interactions**

Control of gene expression requires precise mechanisms to allow genes to be turned on at the correct time and place. One of the ways in which this precision is achieved is by the assembly of multi protein complexes that can activate or repress transcription at enhancers. Enhancers ostensibly serve two main purposes: to tether activators nearby gene promoters, and to nucleate the assembly of complexes that activate transcription (Michael Levine, Cattoglio, & Tjian, 2014). The general mechanisms by which enhancers activate transcription at target gene promoters were discovered in prokaryotes in pioneering studies on the transcription factor NtrC and RNA polymerase containing the sigma54 holoenzyme. Phosphorylation of NtrC leads to the assembly of a large multi-unit structure on the enhancer DNA (Porter, North, Wedel, & Kustu, 1993). This NtrC complex acts as a machine, using ATP hydrolysis to isomerize the sigma-54 closed complex into a transcriptionally active, open complex(Wedel, Weiss, Popham, Dröge, & Kustu, 1990). The activation of sigma-54 happens via a DNA loop that delivers NtrC to the promoter (Wedel et al., 1990).

The same general principle of enhancer activation illustrated in the NtrC complex is seen at work in the assembly of the human interferon beta enhanceosome (Agalioti et al.,

2000). Tight control of the interferon beta gene is achieved by the requirement of a multi-protein complex at the 60 base pair enhancer sequence. The assembly of the three major components–NFkB, c-jun, and IRF3- occurs via multiple protein-protein interactions that synergistically form an activating complex competent to initiate transcription (Agalioti et al., 2000; Lomvardas & Thanos, 2002). In both these cases, the elaborate assembly of an activating complex at distant enhancer DNA sequence allows for tight control of transcription initiation from the target gene promoter.

In the above examples, enhancers allow precise control of gene activation in response to external stimuli by providing a DNA sequence that nucleates the creation of a protein complex that can loop to activate a promoter. Enhancers also exercise control of gene expression during development, when gene transcription is key to ensuring the correct generation of body plans and cell types. For example, the Sonic Hedgehog gene (Shh) is responsible for the patterning of the developing nervous system limb (Mariani & Martin, 2003). An enhancer located at 1 megabase distance from the Sonic Hedgehog gene drives expression in the limb, and mutations of this enhancer lead to ectopic expression of Shh and polydactyly phenotypes (Lettice, 2003).

This remarkable instance of a "limb" enhancer for Shh exercising precise anatomical control of gene expression during development is echoed in the enhancer organization of other important development genes. In Drosophila, the *even-skipped* gene is regulated by five enhancers which coordinate its expression in seven different body segments of the developing *larvae* (Michael Levine & Tjian, 2003). Similarly, expression of the *Pitx* gene, a homeodomain gene that controls the development of the stickleback fish, is regulated by several enhancers, each of which controls Pitx expression in a different

anatomical region (Chan et al., 2010). The evolution of an individual enhancer for Pitx lead to the formation of the pelvic fin in stickleback development.

In mammals, an elaborate enhancer architecture for hox genes has been illustrated. HoxD genes are located in a single gene cluster and are responsible for limb and digit development. Regulation of the hox genes by enhancers located on either side of the gene cluster precisely controls hox gene expression in the developing limb and digit (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011; Noordermeer et al., 2014; Soshnikova & Duboule, 2009).

Synergy is the key concept underlying the mechanism of enhancer function: transcription factors synergistically assemble at enhancer sequences to allow precise activation of transcription at distant gene promoters in response to extracellular signals. Multiple enhancers work synergistically during development to coordinate the correct anatomical expression of target genes.

**Spatiotemporal regulation of enhancers**

Distant regulatory elements can be identified by their tissue-specific histone modifications and transcription factor occupancy (Barski et al., 2007; Heintzman et al., 2009; Visel et al., 2013). Mammals have thousands of distant enhancers (Shen et al., 2012) but within a cell type their activity can be controlled by histone modification and transcription factor occupancy (Heintzman et al., 2009)

A primary identifier of distant enhancers is enrichment for H3K4me1, which by ChIP-ChIP experiments was found to be enriched on enhancers that are transcriptionally active in a cell-type specific manner (Heintzman et al., 2009). Another histone mark, H3K27ac,

is also found enriched on active enhancers in a tissue; H3K4me1 is considered as a mark that "primes" an enhancer, while H3K27ac is subsequently deposited to indicate an active enhancer (Rada-Iglesias et al., 2011). DNAse hypersensitivity is also a reliable indicator of active enhancers, as is transcription at enhancers (eRNAs) due to the presence of RNAPII (Kim et al., 2010). Together, these chromatin identifiers can indicate active regulatory elements in the intergenic regions of the genome.

A primary example of how the spatiotemporal regulation of enhancer chromatin controls developmental gene expression is the HoxD enhances mentioned above that are active during limb formation. The HoxD gene cluster in the mouse is regulated in such a way so that genes of the 3 prime end are expressed earlier in development during the formation of the arm, while the HoxD genes on the 5 prime end of the cluster are expressed later during the formation of the limb bud (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011; Noordermeer et al., 2014). These two sides of the cluster correspond to different topologically associated domains in the nucleus (Dixon et al., 2012).

The temporal control of activation of Hox genes is controlled by the chromatin modification status of the enhancers that control these genes within each topological domain.(Soshnikova & Duboule, 2009) Enhancers located on the 3 prime end are enriched for H3K27ac during the formation of the arm, while the 5 prime end of the cluster is enriched for the repressive H3K27me3 mark. Afterwards the modifcations are switched, and the 3 prime end is enriched for H3K27me3, while the enhancers on the 5 prime end get derepressed. At this stage, these enhancers located in the gene desert on the 5 prime end of the cluster interact with each other to yield the robust activation of 5

prime HoxD genes in the limb bud (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011)

**Olfactory receptor gene expression**

The mouse genome contains more than 1400 olfactory receptor (OR) genes (Buck & Axel, 1991) and comprises the largest gene family in the mouse genome. ORs are scattered across 40 gene clusters located across the genome (Zhang & Firestein, 2002). Each olfactory sensory neuron (OSN) in the olfactory epithelium expresses just one OR gene and in fact only one OR allele ( a Chess, Simon, Cedar, & Axel, 1994). Expression of an OR is not evenly distributed in 1/1400 olfactory neurons: the number of OSNs expressing a given OR varies from gene to gene, and the frequency of expression of a single OR is stereotyped from animal to animal (Feinstein, Bozza, Rodriguez, Vassalli, & Mombaerts, 2004; Iwema & Schwob, 2003; Lee, Cheng, & Gong, 2008). Furthermore, the expression of an OR is restricted topographically within a zone of the olfactory epithelium. The "zonal" expression of an OR across the dorsal-ventral axis of the olfactory epithelium is stereotyped from animal to animal (Vassar, Ngai, & Axel, 1993). Within the confines of a zone of the epithelium, the expression of an OR is seemingly random or stochastic.

Several models have been proposed to explain the monoallelic, stochastic mode of OR expression. It was first thought that irreversible DNA changes – as the recombination of antigen receptor genes - could account for the exclusive nature of OR expression. However, monoclonal mice generated from post-mitotic olfactory sensory neurons expressing a single OR demonstrated that this is not the case, as the olfactory epithelium of these mice expressed the entire repertoire of ORs (Eggan et al., 2004).

A second model for OR choice involves distant locus control regions (LCRs) in the choice of olfactory receptors. A distant regulatory element called H, located 75 kB upstream of an olfactory receptor gene cluster on chromosome 14, was required for transgenic expression of three proximal OR genes, which suggests that stochastic interactions by DNA looping of the H enhancer to the proximal ORs may be the basis of olfactory receptor choice (Serizawa et al., 2003a). The endogenous deletion of H results in the loss of expression of the proximal OR genes, supporting the idea that distant elements regulate the choice of olfactory receptor (Fuss, Omura, & Mombaerts, 2007a; Nishizumi, Kumasaka, Inoue, Nakashima, & Sakano, 2007).

Another model proposes that the nuclear organization of olfactory receptor loci is important in regulating the choice of olfactory receptor. Recent work has shown that olfactory receptors are distributed in a non-random fashion within the nucleus of the developing OSN, where the silent olfactory receptors reside in heterochromatic foci in the nucleus, while the active allele escapes this repression and is found in a euchromatic nuclear territory (Clowney et al., 2012). This unique architecture of olfactory receptors is achieved by the down-regulation in OSNs of the nuclear lamin b recepter gene, which tethers heterochromatin to the nuclear membrane. Disruption of the nuclear organization by exogenous expression of Lamin b receptor results in a loss of singular OR expression, suggesting that the singular expression of an OR may be due to the association of the expressed allele with a "transcription factory" within the nucleus.

Though the mechanism of singular OR choice is still not well understood, the mechanism for the stabilization of the initial choice has been recently delineated. A negative feedback system ensures stable expression of an OR upon activation and prevents

expression of other ORs (Serizawa et al., 2003a). The feedback depends on expression of a functional full-length OR protein that triggers the unfolded protein response in the endoplasmic reticulum and leads to the terminal differentiation of the neuron that expresses a single receptor (Dalton, Lyons, & Lomvardas, 2013; Lyons et al., 2013). The inability to further express other receptors is due to the pre-existing epigenetic silencing of olfactory receptors by H3K9me3 and H4K20me3, markers of constitutive heterochromatin (Magklara et al., 2011a). H3K9me3 and H4K20me3 spread across large domains over olfactory receptor gene loci, covering the promoter sequences, and only the expressed allele is enriched for euchromatic signature histone marks H3K4me3. The expression of an olfactory receptor is afforded by the transient activity of Lsd1, which demethylates one allele but is then turned off by the negative feedback signal from the expressed protein (Dalton et al., 2013; Lyons et al., 2013). This unique mechanism describes how through a negative feedback signal elicited by expression of a single OR and the epigenetic silencing of all OR alleles can lead to the stabilization of a single OR.

**Cis-regulatory elements of ORs**

Initial understanding of the *cis*-regulatory sequences of ORs came from transgenic studies which showed that DNA sequences residing upstream of OR genes M50, M71, and M4 are sufficient to drive OR expression (Qasba & Reed, 1998; Vassalli, Rothman, Feinstein, Zapotocky, & Mombaerts, 2002a). Later experiments identified the H element which is also sufficient and necessary for expression of transgenic and endogenous choice of proximal ORs (Fuss et al., 2007a; Serizawa et al., 2003a). These experiments indicate the importance of *cis*-regulatory sequences in the expression of ORs.

Analysis of promoter sequences of ORs have revealed several transcription factor binding sites, though the functional role of these sequences in OR gene choice has not been established. The Olf1 or O/E-like sequence is found on several olfactory specific genes such as *Adcy3, Golf, OMP,* and *Cnga2* (Kudrycki et al., 1993; M. M. Wang & Reed, 1993). This consensus sequence is also found on several OR promoters (Clowney et al., 2011a; Glusman et al., 2000; Michaloski, Galante, & Malnic, 2006; Sosinsky, Glusman, & Lancet, 2000; Vassalli, Rothman, Feinstein, Zapotocky, & Mombaerts, 2002b) (Glusman 2000, Sosinsky 2000, Vassalli 2002,Michaloski et al, Clowney et al).

The LIM-homeodomain protein Lhx2 binds a homeodomain site in the promoter of M71 olfactory receptor (Hirota & Mombaerts, 2004a)and Lhx2 knockout mice do not express olfactory receptors, indicating a possible role of this factor in olfactory receptor gene choice via binding at a subset of OR promoters (Hirota, Omura, & Mombaerts, 2007b). The comprehensive mapping of more than a thousand OR promoter sequences did not reveal any new common or unique transcription factor binding sites besides the O/E site and homeodomain sites (Clowney et al., 2011a).

Two distant *cis*-regulatory elements, H and P, have been identified that regulate ORs (Bozza et al., 2009; Fuss et al., 2007a; Khan, Vaes, & Mombaerts, 2011). These sequences are also enriched for homeodomain sites and contain O/E-like sites (Vassalli, Feinstein, & Mombaerts, 2011). However, the function of these motifs in singular olfactory receptor gene choice remains unexplored.

**Stochastic, monoallelic expression in mammals**

Other examples of *cis*-regulation of monoallelic, stochastically expressed genes may shed light into the possible mechanisms of OR gene choice. The protocadherin genes Pcdha, b, gamma mediate cell-cell recognition in neurons, and dendrite self-avoidance, and are expressed in a monoallelic, stochastic fashion (Lefebvre, Kostadinov, Chen, Maniatis, & Sanes, 2012). Like ORs, the protocadherin genes are located in large gene clusters. Monoallelic expression of pcdha and gamma genes is accomplished by the expression of different variable exons from each allele – these variable exons encode the extracellular domains of the protocadherin, which mediate cell-cell recognition, while the constant exons encode the intracellular domains (A. Chess, 2005). Each variable exon contains its own promoter and upon activation is spliced to the constant exons. Though both alleles of a protocadherin gene are expressed (unlike ORs, where one allele is silenced), different splice forms are expessed from each allele to generate a combinatorial pattern of expression.

The mechanism of protocadherin promoter choice is, like OR choice, not well understood. However, it is known that two neuronal specific *cis*-regulatory elements, HS1-5 and HS7 play a role in the promoter choice (Ribich, Tasic, & Maniatis, 2006) (Ribich et al). Deletion of these regulatory elements leads to a specific loss of some variable exon expression, suggesting that the activation of a promoter may be controlled by its interaction with a distant *cis*-regulatory element (Kehayova, Monahan, Chen, & Maniatis, 2011).

# Chapter 2: Discovery of olfactory receptor enhancers

## Introduction

Transcriptional specificity is achieved by the combinatorial control of gene expression, according to which distinct sets of transcription factors bind cooperatively to *cis*-regulatory sequences and synergistically activate genes in the proper cell types or in response to specific stimuli. Although core and proximal promoter sequences often play an instructive role in restricting gene expression programs (Deato et al., 2008), the remarkable specificity of tissue- and differentiation-dependent transcription appears to be encoded primarily in distant enhancer elements. Sometimes located hundreds or thousands of kilobases from the genes they regulate, distant enhancers can provide tighter transcriptional control and modularity in gene expression. Therefore, identifying distant enhancers is required to understand the molecular mechanisms of cell differentiation and fate commitment.

Predicting enhancer sequences can be extremely challenging, not only because their distance from a target promoter is variable, but also because enhancers may reside within introns or even coding exons of other genes (Birnbaum et al., 2012). Recent advances in high throughput genomics have revealed general hallmarks of enhancer activity, such as DNase I hypersensitivity, P300 binding, and enrichment for the histone modifications H3K4me1 and H3K27ac, which can be used to identify enhancers that drive transcription in a tissue- and differentiation-dependent manner (Heintzman et al., 2009; Rada-Iglesias et al., 2011; Visel et al., 2009). Despite increasing sequencing capabilities and an ever-expanding list of detectable epigenetic modifications, few qualitative epigenetic differences have been described for enhancer signatures that account for different modes of gene expression. For example, the chromatin state of distal elements engaged in

stochastic or monoallelic gene expression has not been well characterized. In fact, such information exists only for two distal enhancer elements that regulate the expression of clustered protocadherin genes and were identified based on their DNase I hypersensitivity in neurons (Ribich et al., 2006).

Because monoallelic gene expression is far more common in mammals than previously thought (Gimelbrant, Hutchinson, Thompson, & Chess, 2007), we sought to characterize the regulatory landscape of intergenic enhancers that regulate such transcription programs. An ideal model system for this type of analysis is the mouse olfactory epithelium, which is defined by the singular expression of one olfactory receptor (OR) allele out of more than two thousand available OR alleles in each olfactory sensory neuron (Buck & Axel, 1991). The stable expression of a single OR is maintained by an OR-elicited feedback signal (Lewcock & Reed, 2004; Serizawa et al., 2003a; Shykind et al., 2004a) that uses components of the unfolded protein response pathway to shut down Lsd1, the histone demethylase responsible for de-silencing one of the previously silenced OR alleles (Dalton et al., 2013; Lyons et al., 2013). Despite progress in understanding the molecular underpinnings of this feedback mechanism, very little is known about how singular OR expression is achieved and what role distal elements play in OR choice. Thus far, only two enhancers have been identified, the H and P element, which regulate ~0.5% of the total OR repertoire (Fuss, Omura, & Mombaerts, 2007b; Khan et al., 2011). Given the extraordinary number of OR family members dispersed across the mouse genome, a comprehensive characterization of the OR enhancer landscape is necessary to understand the regulatory principles of OR choice.

Here, we provide the first attempt at a high throughput identification and characterization of OR enhancers *in vivo*. DNase I hypersensitivity (DHS)-seq and ChIP-seq experiments uncovered 35 predicted enhancer elements linked to OR gene clusters. Reporter assays in zebrafish embryos confirmed the function of 11 novel enhancer elements that drive GFP expression specifically in olfactory neurons. Transgenic reporter assays and genetic deletion experiments in mice further validated that these DNA elements regulate OR expression. Importantly, our analysis revealed a novel epigenetic signature for distant OR enhancers that distinguishes them from other enhancers in the olfactory epithelium: functional OR enhancers are uniquely surrounded by high levels of H3K79me3 and H3K27me3, and have moderate levels of H3K4me1 and H3K27ac, the two *bona fide* enhancer marks.

**Results**

H, the prototypical OR enhancer (Serizawa et al., 2003), is required for the expression of only three linked OR genes (Fuss et al., 2007, Kahn et al., 2011), suggesting that its function is limited to a small fraction (~0.5%) of olfactory sensory neurons (OSNs). However, H interacts with active OR alleles in *cis* or *trans* in ~1/3 of OSNs (Lomvardas et al., 2006) suggesting that even if it is required for transcription in only a small fraction of OSNs, H (and possibly other OR enhancers) may sustain transcriptional competence, or even become physically engaged with OR transcription, in most neurons. We hypothesized that H and H-like elements may exist in an "epigenetically active" state at a high enough cellular frequency for the detection of the hallmarks of enhancer activity in mixed cell populations from the main olfactory epithelium (MOE). Such hallmarks of

active enhancer elements include enrichment for H3K4me1, H3K27ac, and Dnase I hypersensitivity (DHS) (Heintzman & Ren, 2009; Rada-Iglesias et al., 2011).

To examine the epigenetic state of the H enhancer and to use this information for the identification of novel OR enhancers, we performed ChIP-seq and DHS-seq using native chromatin preparations from the total MOE (Magklara et al., 2011a). Table 1 lists the number of sequencing reads for each experiment. Visual inspection of our data verifies DHS and enhancer mark enrichment for known enhancers of OSN-transcribed genes, such as the protocadherin alpha enhancers, (Figure 1A). Remarkably, the H enhancer sequence is also enriched for H3K4me1, H3K27ac, and has a well-defined DHS peak (Figure 2A), suggesting that we can employ this analysis towards the identification of additional OR enhancers. Indeed, the second described OR enhancer, the P element, also has these characteristics within this mixed cell population (data not shown).

To identify additional elements that share the chromatin pattern of the H enhancer, we performed a computational search for intergenic ChIP-seq and DHS-seq peaks using the SICER algorithm (Zang et al., 2009). To remove enhancers that might generally specify neuronal cell types, we filtered out regions that overlapped with H3K4me1 and H3K27ac peaks from cerebellum ChIP-seq experiments we performed in parallel (Figure 1B). However, even upon this filtering, there are 4750 intergenic sequences that have these characteristics in a MOE-specific fashion. One possible solution to reduce the number of intergenic enhancers that may be involved in OR gene regulation is to restrict our search to only intergenic regions residing within OR gene clusters. Because OR clusters are inaccessible to DNase I and devoid of activating histone marks (Magklara et al., 2011,

Clowney et al., 2012), this strategy reduces the number of positive hits to 35, with an average distance of ~35Kb from the nearest OR (Figure 1B). Under these thresholds both the H and P elements are included in this list. The genomic distribution of these elements in the context of OR gene clusters is shown in Figure 1C and genomic coordinates are shown in Table 2.

As a complementary approach we searched for other epigenetic marks that may be used to distinguish OR enhancers from the rest of the MOE-specific regulatory elements. Two histone modifications associated with repression, H3K79me3 and H3K27me3, (Barski et al., 2007; Ernst et al., 2011) have a unique distribution at the H locus; they are missing from the actual enhancer sequence but are enriched in the flanking sequences (Figure 1A). Visual inspection of the remaining OR enhancer candidates shows that this pattern is shared among 23 of the 35 potential enhancer elements (examples shown in Figure 2B,C). Aggregate plots comparing ChIP-seq and DHS-seq reads on OR proximal elements and predicted MOE enhancers located outside OR clusters highlight the striking specificity of this epigenetic signature (Figure 2D). Indeed, while 65% of predicted OR enhancers overlap with regions of H3K79me3 enrichment, only 2.6% of predicted MOE enhancers do (Figure 1B).

Generating independent mouse transgenic lines for each of the 35 predicted OR enhancers is not trivial, thus we sought a different functional assay that is compatible with high throughput screening *in vivo*. We performed transient reporter assays in zebrafish embryos and scored for MOE-specific reporter expression. Zebrafish have a less complex OR repertoire than mammals (~100 functional ORs) but share several principles of mammalian OR gene regulation, including monogenic OR expression and a

feedback system that prevents coexpression of multiple ORs (Ferreira et al., 2014; Niimura & Nei, 2005; Sato, Miyasaka, & Yoshihara, 2007). Although the predicted OR enhancer sequences are not conserved to zebrafish (Figure 3C), the H element supports reporter expression in zebrafish OSNs (Nishizumi et al., 2007), suggesting that this assay is appropriate for the functional identification of OR enhancers.

We cloned the DHS peaks of 32 novel candidate enhancer sequences into a Tol2 retrotransposon-based reporter vector with a minimal promoter and GFP (Fig 2D, Table 3). Each construct was injected into one-cell stage zebrafish oocytes and GFP expression was monitored during embryogenesis at 24 and 48 hours post fertilization. Because enhancer complexes were previously described as a "regulatory archipelago" (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011), we named the potential OR enhancers after Greek islands, a nomenclature that will be followed throughout the manuscript. Remarkably, specific GFP expression was observed in OSNs of the olfactory epithelium for 12 (11 plus H) of the sequences screened (Figure 2D, Figure 3A, Figure 3C). Empty vector injections did not drive expression in the OE (data not shown).

Interestingly, OR genes proximal to these 11 elements are more highly expressed than the average OR by RNA-seq of mouse OSNs, supporting an activating role of these sequences in OR transcription (Figure 5A). To further validate that the zebrafish reporter assay is predictive of enhancer activity in mouse OSNs, we generated transgenic beta galactosidase reporter lines using a reporter vector driven by the hsp68 promoter. We tested two zebrafish-positive elements, Sfaktiria and Kefallonia, referred to as Sfaktiria-lacZ and Kefallonia-lacZ in the rest of the manuscript. Whole mount x-gal staining of

these transgenic mice shows widespread reporter expression specifically in the MOE at frequencies similar to the H-lacZ transgenic, which we generated as positive control (Figure 4A-D, Figure 5B). Immunofluorescence (IF) for beta-galactosidase in the olfactory epithelium of the Sfaktiria-lacZ mouse shows broad expression in the neuronal layer (Figure 4E). If in the olfactory bulb of the Sfaktiria-lacZ and Lipsi-lacZ mouse shows beta-galactosidase positive axons targeting multiple glomeruli in the olfactory bulb and expressing the glutamate transporter vglut2 (Figure 4F, 4G, 5E), showing that this enhancer drives expression in mature OSNs. In contrast, there is no beta-galactosidase IF signal in Neurogenin-1 positive neurons (Figure 5D), suggesting that enhancer activity is synchronous to OR expression. Similar results were obtained from Kefallonia-lacZ transgenics (data not shown). To examine if two different OR enhancers can be simultaneously active in OSNs, we crossed the Sfaktiria-lacZ transgenics to MOR28iresGFP knock-in mice. In these mice, GFP positive neurons express MOR28 (Olfr1507), which is under the control of the H enhancer. The MOE and the olfactory bulb of these mice shows coexpression of GFP and beta-galactosidase (Figure 4H), suggesting that Sfaktiria and H can be both transcriptionally engaged in the same OSNs.

These transgenic reporter assays demonstrate an OSN-specific enhancer activity for the candidate OR enhancers. Previous deletion of two of these elements resulted in transcriptional downregulation of proximal OR genes. To further examine whether verified OSN enhancers regulate OR expression we attempted to knock out two of these sequences, Sfaktiria and Lipsi. Although we successfully targeted both of these elements in ES cells (data not shown), only the Lipsi KO chimeras transmitted the targeted allele. Lipsi is located on chromosome 2 between ofr362 and olfr364, and our targeting strategy

deleted, by homologous recombination, 1000bp of conserved sequence corresponding to the DHS peak at this location (Figure 4I). qRT-PCR analysis on RNA prepared from wild type and Lipsi knockout littermates shows marked reduction in expression of the eight linked ORs that reside within this genomic cluster on chromosome 2, while ORs from a distant genomic cluster in the same chromosome or ORs from different chromosomes are unaffected (Figure 4J). RNA ISH analysis in MOE sections from Lipsi knockout and wild type littermates confirms that expression of ORs from this genomic cluster is abolished (Figure 4K, Figure 5E). Thus, deletion of an enhancer element from a set of epigenetically identified enhancers that includes the H and P elements results in transcriptional deficits of proximal OR genes, suggesting that the intergenic sequences identified may be *bona fide* OR enhancers.

After establishing that the zebrafish reporter assay is appropriate for the screening of potential mouse OR enhancers, we sought to identify differences in the epigenetic and genetic properties of these elements that may be predictive of their activity. To capture histone modification variation across enhancers, we summarized their levels around DHS peaks using principal component analysis. The first principle component (PC1) of each modification robustly represents its overall level at each enhancer locus. Ordering the 35 candidate OR enhancers by H3K79me3 PC1 grouped together enhancers that were active in zebrafish OSNs, as well as the H and P elements (Figure 6B). This set of enhancers had high levels of flanking H3K79me3. Surprisingly, ordering by H3K4me1 PC1 shows that enhancers that were active in zebrafish OSNs had lower levels of H3K4me1 (Figure 6C). Analysis of the other chromatin states did not find a correlation with enhancer function (Figure 6D,E). Furthermore, hierarchical clustering by average levels of

H3K27ac, H3K4me1, DHS, H3K79me3, and H3K27me3 results in the segregation of OR enhancer candidates into two main clusters: one containing sequences enriched for H3K27ac and H3K4me1, and another comprised of sequences enriched for DHS, H3K79me3, and H3K27me3 (Figure 6A). The latter cluster contains the majority of the sequences that were functional in the enhancer assay, as well as H and P.

To further examine whether these epigenetic characteristics are predictive of enhancer activity, we performed a logistic regression of zebrafish enhancer activity against each histone modification PC1 and found that high H3K79me3 levels are significantly predictive of enhancer activity (Figure 6G). Random forest analysis performed by training the algorithm on the ChIP-seq and DHS-seq data found that the variable most strongly associated with enhancer activity is H3K79me3 (Figure 6F), providing additional support for the positive association between OR-proximal enhancer activity and H3K79me3 enrichment at the regions flanking these enhancers.

**Discussion**

Our experiments revealed 35 intergenic OR-linked sequences that share common epigenetic properties with the H element, the prototypical OR enhancer. In addition to the increased Dnase I hypersensitivity and the histone modifications found on most enhancers, these sequences are characterized by high flanking levels of H3K79me3, a modification considered repressive in mammals (Jones et al., 2008) but activating in drosophila embryos, where it is enriched on developmentally regulated enhancers (Bonn et al., 2012). Our reporter screen showed that 12 of these 35 elements regulate OSN-specific expression in zebrafish OSNs and revealed a positive association between flanking H3K79me3 enrichment and enhancer activity. The strong correlation between

flanking H3K79me3 levels and enhancer activity in zebrafish OSNs suggests that the distinct epigenetic signature of potential OR enhancers is established by sequence specific transcription factors that remained conserved, both in terms of DNA binding specificity as well as expression patterns, for the 420 million years that separate the two species.

We demonstrated the activity of three of these elements (Lipsi, Sfaktiria, Kefallonia) as OSN enhancers in the mouse, and we showed that Lipsi is necessary for the expression of proximal OR genes. Despite the many similarities between the rest of these elements and H, P, and Lipsi, the remaining OSN enhancers can only be considered putative OR enhancers at the moment. Because zebrafish reporter assays for mammalian enhancers generate false negatives (Ariza-Cosano et al., 2012; Booker et al., 2013; McGaughey et al., 2008), it is also likely that many of the DNA elements that did not activate transcription in zebrafish OSNs are mammalian or mouse-specific OSN enhancers that regulate proximal ORs, especially in light of their extensive interactions with verified enhancers. Finally, it is worth noting that our screen is far from being saturated, since enhancers that are transcriptionally engaged in a smaller cell population would not meet our computational thresholds and enhancers with a different epigenetic signature would be ignored. However, although we cannot claim that we discovered the complete repertoire of intergenic regulatory elements for the ~1400 OR genes, we increased by an order of magnitude the number of putative enhancers that participate in OR regulation and we revealed some of their genetic and epigenetic underpinnings.

The results presented here suggest that the number of intergenic OR enhancers is much lower than the number of OR genes. 35 intergenic sequences share similar epigenetic

characteristics, and 13 of these, including H and P, are functional in zebrafish or mouse. Even allowing for false negatives in the zebrafish screen, our data would predict that each intergenic enhancer regulates 30-40 OR genes, which is significantly higher than the number of OR genes affected by H, P, and Lipsi deletion (Fuss et al., 2007b; Khan et al., 2011). Of course we cannot exclude the existence of a large number of OR enhancers with a completely different chromatin signature. Alternatively, it is possible that the majority of OR genes may be transcribed from their endogenous loci without the use of distant regulatory elements. Given that most OR promoters are homogeneous and evenly heterochromatinized in the MOE, it is difficult to envision why a few OR genes would require distant enhancers while the rest do not. Thus, if our enhancer screen is indeed saturating, then an unknown mechanism may be employed for the singular and robust expression of one out of more than 2,000 alleles.

We previously showed that the H element interacts with active OR alleles in *cis* and in *trans* (Lomvardas 2006), and that disruption of these *trans* interactions results in reduced OR expression and loss of transcriptional singularity (Clowney et al., 2012). Interchromosomal interactions are extensive in olfactory neurons, where ORs from multiple chromosomes aggregate in a few nuclear foci and segregate from the rest of the genome (Clowney et al., 2012). This unique nuclear architecture may favor enhancer-enhancer and enhancer-promoter interactions in both *cis* and *trans*, which could explain the expression of OR alleles that lack a proximal enhancer, and could also account for the fact that enhancer-proximal ORs are expressed at higher frequencies than the rest of the repertoire. The identification and characterization of such a large number of OR-proximal

enhancers and of transcription factors that bind to them makes experimental testing of this regulatory model possible in the near future

**Methods**

DHS-seq. Nuclei from the olfactory epithelium of five 6-8 week old wildtype mice were isolated as described (Magklara et al., 2011b) and resuspended in nuclease digestion buffer (0.32M sucrose, 50mM Tris-HCL pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$, 0.1mM PMSF). 3 million nuclei were brought to 250mL nuclease digestion buffer, pre-warmed for one minute at 37C, and incubated with 20 U Dnase I (Ambion) for 2 minutes. Reactions were stopped with 30mL 0.5 M EDTA and placed on ice. Reactions were then incubated with 200 μg proteinase K (Ambion) at 55°C overnight, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, incubated with 10 μg RNase A (Roche) at 37°C for 30 minutes, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, and one half of each reaction was run on a 1% agarose/1x TAE gel. Regions 100-500 bp were excised and purified using Qiagen Gel Extraction Kit. Illumina sequencing libraries were then prepared using standard protocols, and amplified for 15 cycles with Phusion DNA Polymerase (NEB). DNA polymerase (NEB) and Illumina TruSeq primers.

ChIP-seq. Nuclei from the olfactory epithelium of 6-8 week old wildtype mice were isolated and native chromatin ChIP was performed as described (Magklara et al., 2011a) using 1 mg anti-H3K27ac (Millipore, cma309), anti-H3K4me1 (Abcam, ab8895), anti-H3K79me3 (Abcam, ab2621) and anti-H3K27me3 (Millipore, 07-449) antibodies. For

library preparation, ChIP DNA was sonicated for 180 seconds on a Covaris S220 and prepared for sequencing using the Ovation Ultralow Library Kit (Nugen). Sequencing data from duplicate experiments were combined.

ChIP-seq and DHS-seq analysis.

Sequencing reads were mapped to the mouse genome using Bowtie2. Peak finding algorithm SICER (Zang et al., 2009) was run on reads that mapped to OR clusters or the whole genome, to identify OR and MOE potential enhancers, respectively. OR clusters were defined as genomic regions containing one or more OR genes extending to the nearest non-OR Refseq gene. Peaks located within 5kB of a Refseq gene or an OR promoter (Clowney et al., 2011b) were subtracted. DHS peaks that intersected H3K4me1 or H3K27ac peaks were selected, and DHS peaks that intersected with cerebellum H3K4me1 or H3K27ac peaks were subtracted

Transgenic zebrafish and mouse assays. Enhancer candidate sequences were amplified from mouse genomic DNA using primers targeting DHS peaks. PCR products were cloned into the e1b-GFP-tol2 vector containing a minimal promoter followed by GFP and injected using standard protocols into at least 75 one-cell stage zebrafish oocytes per construct. GFP expression was observed at 24 and 48 hpf. Embryos containing at least one GFP positive OSN were counted, embryos with GFP expression in other tissues were excluded. For transgene reporter mice, the same PCR products were cloned into a vector containing the Hsp68 minimal promoter followed by the lacZ reporter gene (Pennacchio et al., 2006). Transgenic mice were generated using the Gladstone Transgenic Core.

Founders were genotyped by PCR and stained for lacZ expression using whole mount x-gal staining using standard protocols. All animal work was approved by the UCSF IACUC.

In Situ Hybridization (ISH) and Immunofluorescence (IF): Individual OR coding sequences were amplified using the following primers:

| | Forward primer | Reverse primer |
|---|---|---|
| olfr361 | ggctactaaaaacaggacagaagt | tgggaaatccaagctggacactg |
| olfr362 | attggggaatgtgacttt | ccactctcccctgtctcagta |
| olfr364 | cagcagcacttcagacttcatct | agtgaagcctgtcccttaacttcc |
| olfr365 | acctcagacttcatcttgct | ccattaacttcctcaaacctc |
| olfr366 | aaaaccagagccacatgacaga | tagccagttttctcaaggcct |
| olfr367 | acctgagacttcagtttgcactg | tccccaagccaaatagtttcca |

ISH was performed as previously described (Lyons et al., 2013). ISH experiments were quantified by counting cells over consecutive sections of the MOE.

IF experiments were performed on pre-fixed MOE as previously described (Clowney et al., 2012) using anti-betagalactosidase (abcam, ab9361), anti-GFP (abcam, ab290), anti-vglut2 (Millipore, AB2251). Confocal images were collected with the Zeiss LSM 700 and brightfield images were collected on the Zeiss Axioskop Plus. ImageJ (NIH) was used for image processing.

Lipsi knockout mice. Targeting vector was made by replacing the Lipsi enhancer (mm9, chr2:36983201-36983999) with a *floxed* neo cassette flanked by 5.6 kb and 6kb homology arms. Gene targeting was done by Biocytogen. The targeting construct was electroporated into C57/Bl6 embryonic stem cells. G418-resistant colonies were picked

and analyzed for recombination events by genotyping PCR and by Southern Blot using centromeric and telomeric external probes made with the following primers:

Telomeric-F: ctgtggaatagtgagggatctt
Telomeric-R: gtactagagcaggaatgtggc
Centromeric-F: atcaaccaaagagcacacatgg
Centromeric-R: ggcatttgatgacattgcagat

Clones containing the deletion were injected into mouse blastocysts and, upon germline transmission from chimeras, animals containing the Lipsi allele were genotyped using the following primers:

WT-F: ccagatcctgaataactagtgca,
WT-R: tgacctcatggaagatctggc,
Neo-R: cagaggccacttgtgtagcg.

RT-PCR. RNA was extracted from total MOE using Trizol (Invitrogen) and treated with Turbo DNAse (Invitrogen). 1 mg RNA was used for first strand synthesis using Superscript III (Invitrogen). qPCR was performed using Maxima SYBR Green (Fermentas). The following primers were used to target individual ORs:

RT-PCR primers

| | Forward primer | Reverse primer |
|---|---|---|
| olfr360 | cacattctggttgtgctgct | cataggggagatgagcgtgt |
| olfr361 | tgacagtgggcagtatggaa | accacccatgaggtcaatgt |
| olfr362 | cagcctccctcacagaactc | gggcacctttcacatcctta |
| olfr364 | gtggtgatgaggtcacaacg | gggaagcacagaatgagagg |
| olfr365 | gcctctcattctgtgcttcc | acagccagggtctcagtcat |
| olfr366 | gtcccatcttcactccctca | gcaggagagtgtcagcagtg |
| olfr367 | tccagccatcatcctcctac | cccaagccaaatagtttcca |
| olfr368 | ccctggaagtcaacagcttc | ccaacactgtggttgtctgg |
| olfr1507 | atatagttgcttctttgatttagattacca | gctgaaccccataatctacacctt |
| olfr1509 | tgcagaaggtcgtcggaaag | atctggccgagtgtagagga |

| | | |
|---|---|---|
| olfr1015 | gcccatactgccaacacttt | gaagcccaccatactggaaa |

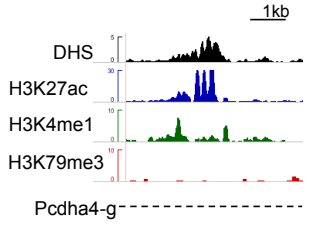
Principal component analysis. 200 25 bp windows were generated starting from the midpoint of a given genomic region (for example DNase hypersensitivity). For a given set of regions, these windows were arranged into a matrix with regions for rows and positions flanking region as columns. Ordering of rows was achieved by performing principal component analysis (R command *prcomp*) and ordering by PC1.

Logistic regression. Associations between enhancer activity in zebrafish and chromatin features (histone modifications and DNase hypersensitivity) were assessed via logistic regression. The first principal component of each chromatin feature matrix (computed as described in methods for 'Principal component analysis') was used to provide a summary value for each feature. These values were then scaled by mean and standard deviation. Logistic regression was performed using the R command glm(zebrafish~., family=binomial).

Random Forest. Associations between enhancer activity in zebrafish and chromatin features (histone modifications and DNase hypersensitivity) were also determined using random forest (RF) regression (R package *party*). Input variables consisted of the chromatin feature data as described in 'logistic regression'. In addition, binary variables of TRANSFAC motif occurrence were included in the input matrix for 169 total variables. Response variable zebrafish activity was coded as binary yes/no. Four thousand RF trees were constructed (control = cforest_unbiased(ntree=4000, mtry=3); cforest1 = cforest(zebrafish~., data =data, controls=control)), and conditional variable importances

as defined in (Strobl, Boulesteix, Kneib, Augustin, & Zeileis, 2008) were determined using: varimp(cforest1, conditional = TRUE). This process was repeated ten times using different seeds, and variable importances were averaged across these runs. Significant variables were designated as those with importances above the absolute value of the lowest negative importance, as suggested in (Strobl, Malley, & Tutz, 2009).

A




B

| | OR enhancers | MOE enhancers |
|---|---|---|
| DHS | 634 | 25,562 |
| H3K4me1 | 1,080 | 71,552 |
| H3K27ac | 941 | 56,760 |
| Intersect DHS H3K27ac/H3K4me1 | 69 | 10,095 |
| Subtract cerebellar | **35** | **4750** |
| Intersect H3K79me3 | 23 | 126 |

C

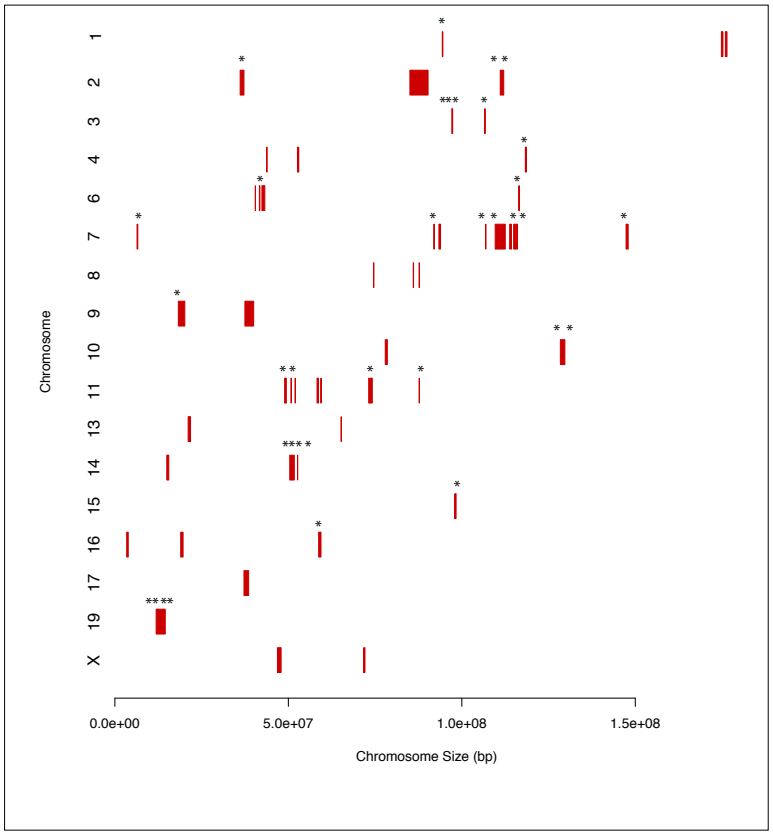


Figure 1: Identification of 35 putative enhancers

**(A)** DHS-seq and ChIP-seq tracks at protocadherin a enhancer HS4. **(B)** Analysis pipeline for generating OR enhancer candidates (column 1) and MOE enhancer candidates (column 2). Analysis for OR enhancers is restricted to genomic coordinates of the 52 OR clusters, extending to within 5kB of nearest non-OR gene. Intergenic peaks were selected to be at a distance of 5kB or more from a Refseq gene or OR promoter. First three rows list the number of intergenic peaks from DHS-seq and H3K27ac and H3K24me1 ChIP-seq data. Fourth row lists the number of DHS peaks that intersected either H3K27ac or H3K4me1 peaks. Fifth row subtracts DHS peaks that intersected either H3K27ac or H3K4me1 SICER peaks from cerebellum ChIP-seq dataset. These are the sequences referred to as "OR enhancers" or "MOE enhancers", respectively, in the text. Sixth row is number of OR enhancers or MOE enhancers that intersect an H3K79me3 peak. **(C).** Map of 35 OR enhancer candidate locations (asterisk) and OR clusters (red bars) generated using the CEAS tool (Shin, Liu, Manrai, & Liu, 2009). X axis is chromosome size.

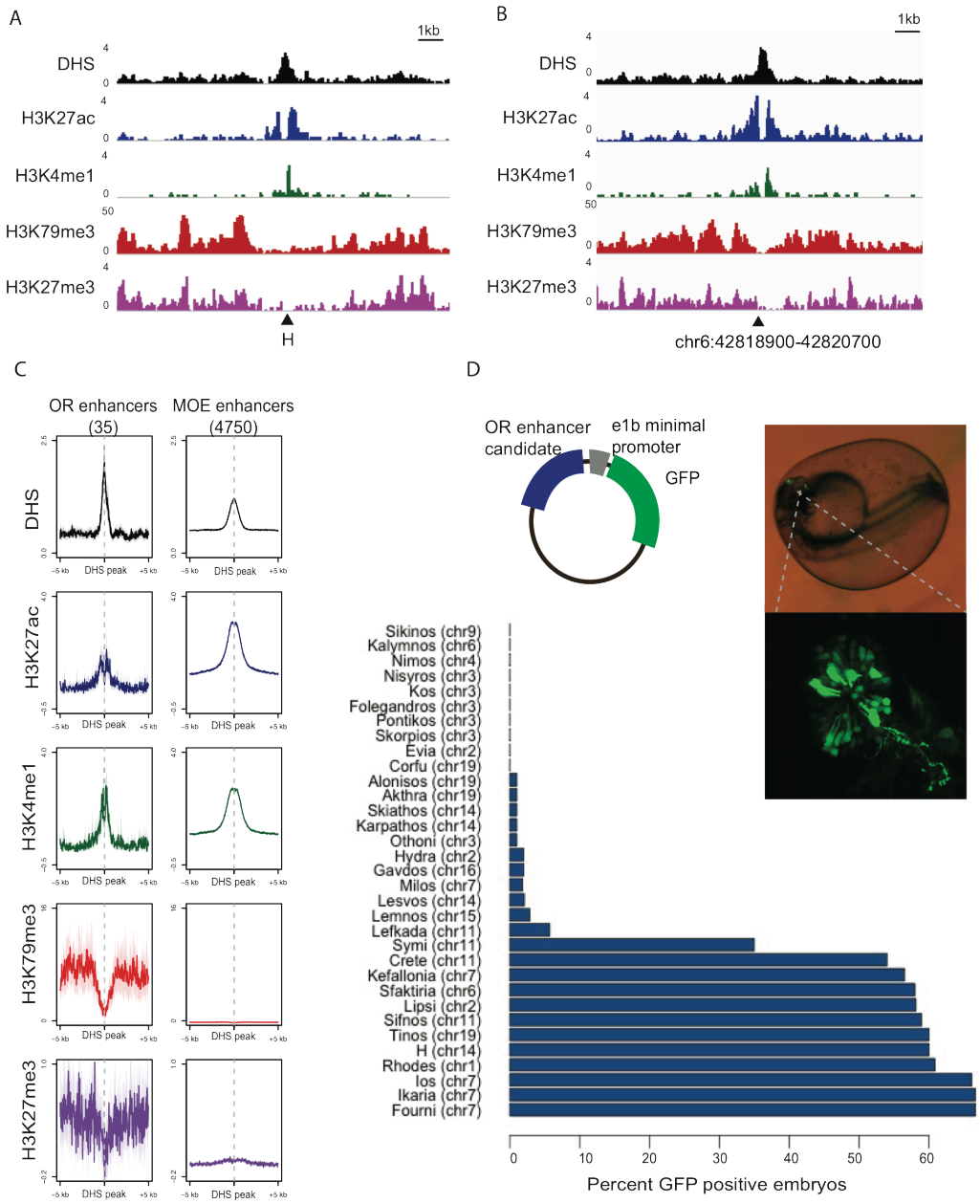
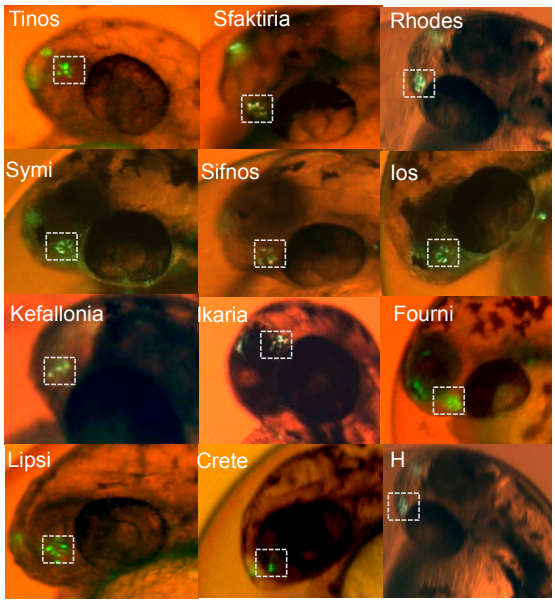

A
1kb
DHS
H3K27ac
H3K4me1
H3K79me3
H3K27me3
H
B
1kb
DHS
H3K27ac
H3K4me1
H3K79me3
H3K27me3
chr6:42818900-42820700
C
OR enhancers (35)
MOE enhancers (4750)
DHS
H3K27ac
H3K4me1
H3K79me3
H3K27me3
DHS peak
D
OR enhancer candidate
e1b minimal promoter
GFP
Sikinos (chr9)
Kalymnos (chr6)
Nimos (chr4)
Nisyros (chr3)
Kos (chr3)
Folegandros (chr3)
Pontikos (chr3)
Skorpios (chr3)
Evia (chr2)
Corfu (chr19)
Alonisos (chr19)
Akthra (chr19)
Skiathos (chr14)
Karpathos (chr14)
Othoni (chr3)
Hydra (chr2)
Gavdos (chr16)
Milos (chr7)
Lesvos (chr14)
Lemnos (chr15)
Lefkada (chr11)
Symi (chr11)
Crete (chr11)
Kefallonia (chr7)
Sfaktiria (chr6)
Lipsi (chr2)
Sifnos (chr11)
Tinos (chr19)
H (chr14)
Rhodes (chr1)
Ios (chr7)
Ikaria (chr7)
Fourni (chr7)
0
10
20
30
40
50
60
Percent GFP positive embryos

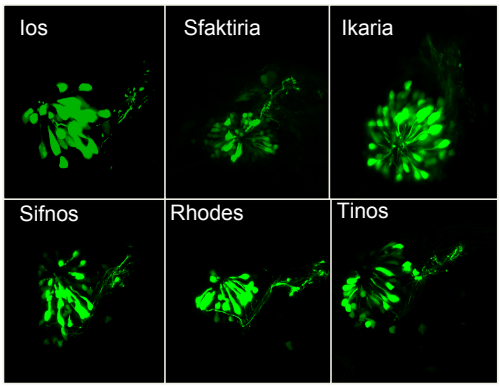
Figure 2: An epigenetic signature for olfactory receptor enhancers

**(A**) Sequencing tracks for DNase I Hypersensitivity-seq (DHS) and ChIP-seq. Each row displays the number of reads for each track. Triangle indicates H enhancer. (**B)** DHS-seq and ChIP-seq tracks over potential OR enhancer Sfaktiria (coordinates are mm9). **(C)** Aggregate plots of ChIP-seq and DHS-seq reads over all potential OR enhancers (left) and all potential MOE enhancers (right). Y axis is RPKM, error is bootstrapped 95% confidence intervals. X axis is centered at DHS peaks. **(D).** Schematic depicting the E1b-tol2 expression construct. **(E).** Zebrafish embryo injected with Sfaktiria-gfp construct at 24 hours post fertilization (hpf). Olfactory epithelium is indicated with dotted lines. Below, OSN cell bodies and axons expressing GFP. **(F).** Results of zebrafish enhancer screen. Percent of injected zebrafish embryos with GFP positive olfactory neurons at 48 hpf for each OR enhancer candidate.

A


Tinos
Sfaktiria
Rhodes
Symi
Sifnos
Ios
Kefallonia
Ikaria
Fourni
Lipsi
Crete
H

B


Ios
Sfaktiria
Ikaria
Sifnos
Rhodes
Tinos

C

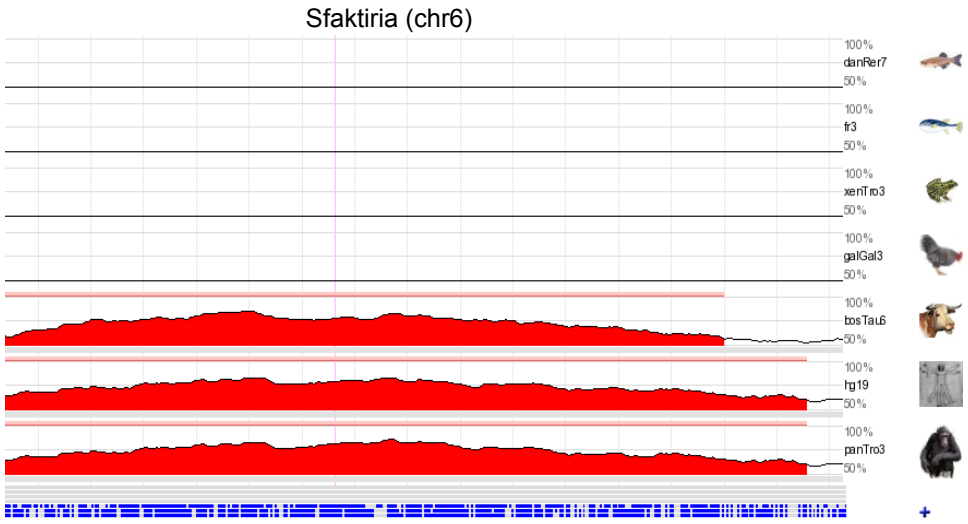

Sfaktiria (chr6)
100%
danRer7
50%
100%
fr3
50%
100%
xenTro3
50%
100%
galGal3
50%
100%
bosTau6
50%
100%
hg19
50%
100%
panTro3
50%
+

Figure 3: Enhancer transgenes drive reporter expression in zebrafish OSNs

**(A).** Representative zebrafish embryos from oocyte injections of 11 candidate OR enhancers and H enhancer. GFP expression in the olfactory epithelium is indicated. (**B).** Representative images of zebrafish OSNs expressing GFP at 48hpf from 6 OR enhancer transgene injections. (**C)** Evolutionary conservation (% sequence identity, Y axis) of the mouse Sfaktiria sequence is shown using the Evolutionary Conserved Regions browser (ECR, http://ecrbrowser.dcode.org) for vertebrates: zebrafish (danRer7), fugu (fr3), frog (xenTro3), chicken ( galGal3), cow (bosTau6), chimpanzee (panTro3), and human (hg19).

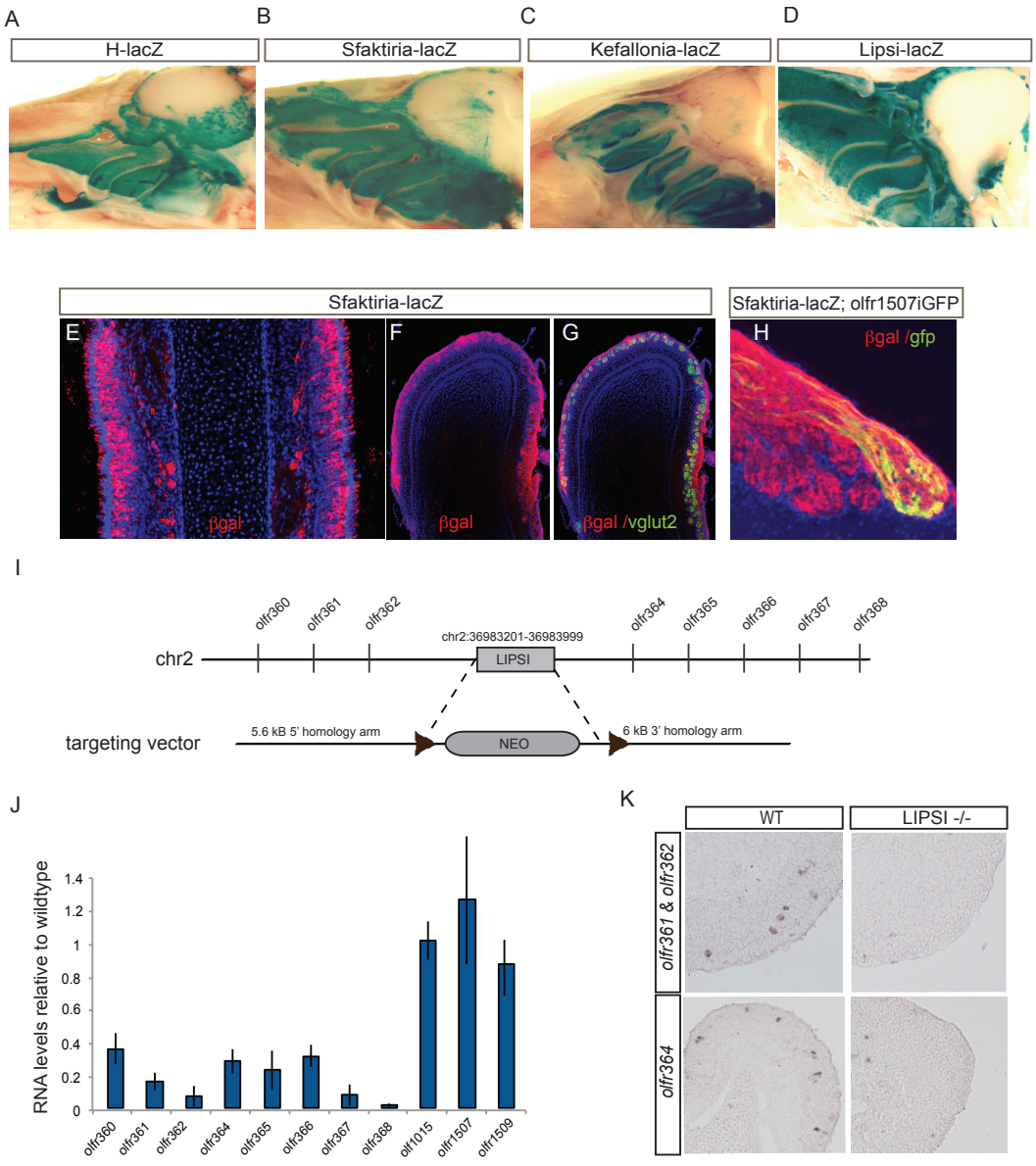
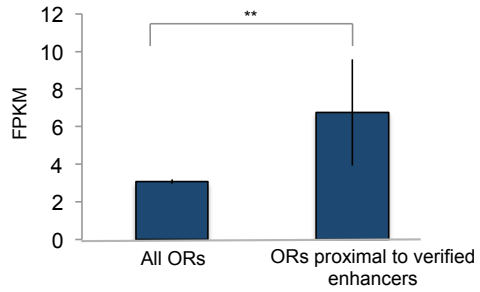

A
B
C
D
H-lacZ
Sfaktiria-lacZ
Kefallonia-lacZ
Lipsi-lacZ
Sfaktiria-lacZ
Sfaktiria-lacZ; olfr1507iGFP
E
F
G
H
βgal
βgal
βgal /vglut2
βgal /gfp
I
chr2
olfr360
olfr361
olfr362
chr2:36983201-36983999
LIPSI
olfr364
olfr365
olfr366
olfr367
olfr368
targeting vector
5.6 kB 5' homology arm
NEO
6 kB 3' homology arm
J
RNA levels relative to wildtype
0
0.2
0.4
0.6
0.8
1
1.2
1.4
olfr360
olfr361
olfr362
olfr364
olfr365
olfr366
olfr367
olfr368
olf1015
olfr1507
olfr1509
K
WT
LIPSI -/-
olfr361 & olfr362
olfr364

Figure 4: Genetic verification of enhancer function in mouse OSNs

**(A-D)** Whole mount x-gal staining of MOE from H-lacZ, Sfaktiria-lacZ, Kefallonia-lacZ, and Lipsi-lacZ enhancer transgenic mice, respectively. **(E)** IF for bgal (red) in Sfaktiria-lacZ olfactory epithelium. DAPI nuclear stain (blue)**. (F)** IF for bgal (red) in Sfaktiria-lacZ olfactory bulb. **(G)** IF for bgal (red) and vglut2 (green) in Sfaktiria-lacZ olfactory bulb. **(H)** IF for bgal (red) and GFP (green) in Sfaktiria-lacZ ; olfr1507iresGFP olfactory bulb. DAPI nuclear stain (blue). **(I)** Targeted deletion of Lipsi allele on chromosome 2. 998 bp were replaced with a floxed neo cassette via homologous recombination. Coordinates are mm9. **(J)** RT-qPCR from Lipsi KO and Lipsi WT MOE. RT-qPCR levels for each primer set were normalized to OMP, and the results are shown as fold difference of Lipsi KO over Lipsi WT. Error bars represent SEM over duplicate experiments.

**(K)** ISH for *Olfr361* and *Olfr362* pooled probes and *Olfr364* in Lipsi KO and WT MOE at P

A

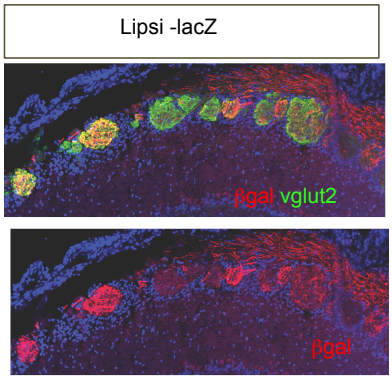



B

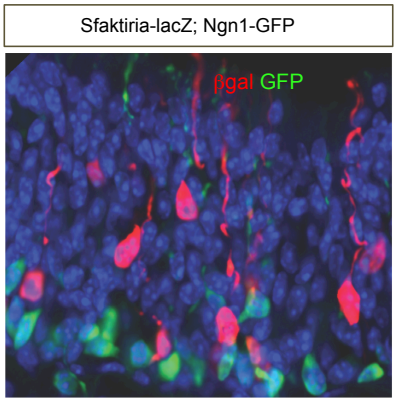
| | LACZ+ FOUNDERS |
|---|---|
| H-lacZ | 2/5 |
| Sfaktiria-lacZ | 2/5 |
| Kefallonia-lacZ | 1/3 |
| Lipsi-lacZ | 2/5 |

E

C




D




E

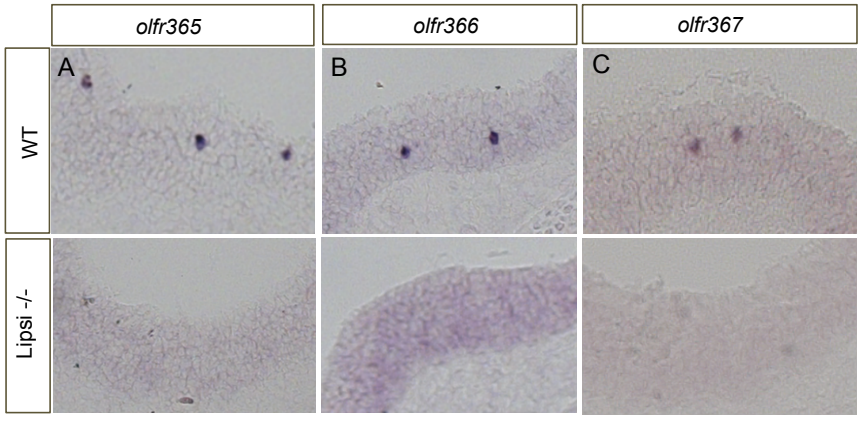


Figure 5: Characterization of enhancer function in mouse OSNs

**(A)** Average expression of ORs nearest to zebrafish-verified enhancers compared to expression of all ORs. RNA-seq data is from mature OSNs isolated by FAC-sorting of OMP-GFP mice. Y axis is FPKM (fragments per kilobase of transcript per million fragments mapped). **(B)** Table indicating number of transgenic founder mice for each OR enhancer transgene. LacZ expression was assayed by whole-mount x-gal staining of the MOE. **(C)** IF for bgal (red) and vglut2 (green) in Lipsi-lacZ olfactory bulb. (**D)** IF for bgal (red) and GFP (green) in Sfaktiria-lacZ; Neurogenin1-GFP mice MOE at P4. **(E).** In situ hybridization for *olfr365, olfr366, and olfr367* in WT and Lipsi KO MOE at P0.

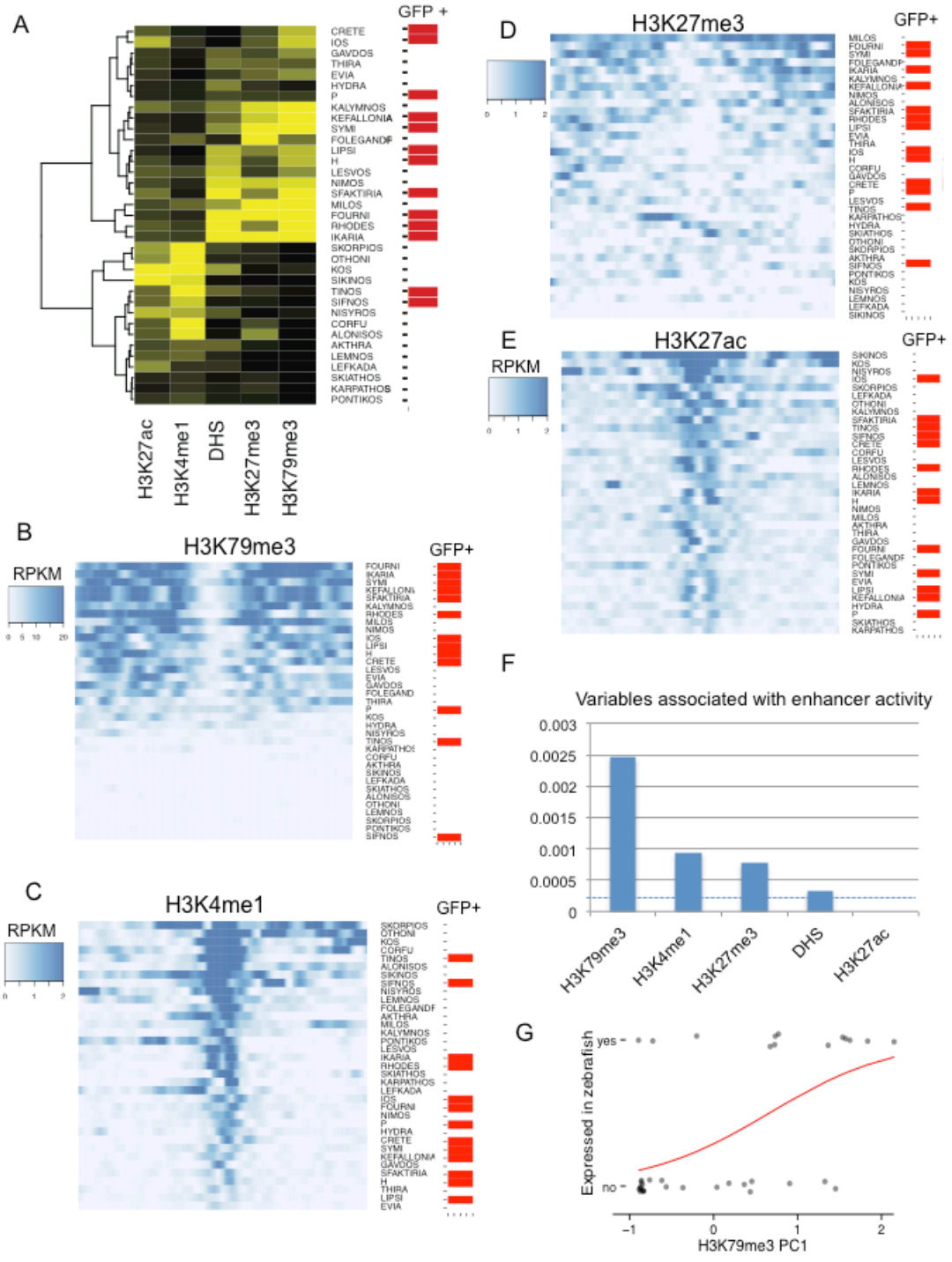

A
GFP +
CRETE
IOS
GAVDOS
THIRA
EVIA
HYDRA
P
KALYMNOS
KEFALLONIA
SYMI
FOLEGANDR
LIPSI
H
LESVOS
NIMOS
SFAKTIRIA
MILOS
FOURNI
RHODES
IKARIA
SKORPIOS
OTHONI
KOS
SIKINOS
TINOS
SIFNOS
NISYROS
CORFU
ALONISOS
AKTHRA
LEMNOS
LEFKADA
SKIATHOS
KARPATHOS
PONTIKOS
H3K27ac
H3K4me1
DHS
H3K27me3
H3K79me3
B
H3K79me3
RPKM
GFP+
C
H3K4me1
RPKM
GFP+
D
H3K27me3
GFP+
E
H3K27ac
RPKM
GFP+
F
Variables associated with enhancer activity
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
H3K79me3
H3K4me1
H3K27me3
DHS
H3K27ac
G
Expressed in zebrafish
yes
no
−1
0
1
2
H3K79me3 PC1

Figure 6: H3K79me3 is associated with enhancer activity in zebrafish asay

**(A)** Hierarchical clustering of OR enhancer candidates by DHS and ChIP-seq histone modifications levels. Red ticks indicate H, P, and the OR enhancers that drove GFP expression in zebrafish. **(B-E)** Heatmaps of H3K79me3, H3K4me1, H3K27me3, H3K27ac, respectively, RPKM over OR enhancer candidates. Y axis is OR enhancer candidates ordered by principal component 1 of each chromatin modification matrix. X axis is centered at DHS peak. Red ticks indicate H, P, and the 11 sequences that drove GFP expression in zebrafish OSNs. (**F)** Regression analysis using random forest supervised learning (see Supplemental Experimental Procedures). Chart shows variables associated with enhancer activity in zebrafish reporter assay. Y axis is the importance score for each variable, the dotted line marks the significance threshold (0.00025).

**(G)** Logistic regression curve fitting each OR enhancer's performance in the zebrafish expression assay (Y axis, yes or no) to H3K79me3 pattern (X axis, Principal Component 1), p=0.03.

| Sample | Sequencing type | # cycles | # mapped reads |
|---|---|---|---|
| MOE H3K4me1 rep1 | Paired-end | 2x50 | 150,876,233 |
| MOE H3K4me1 rep2 | Paired-end | 2x50 | 191,665,025 |
| MOE H3K27ac rep1 | Paired-end | 2x50 | 75,308,226 |
| MOE H3K27ac rep2 | Paired-end | 2x50 | 80,653,918 |
| MOE H2K27me3 rep1 | Paired-end | 2x50 | 52,632,918 |
| MOE H3K27me3 rep2 | Paired-end | 2x50 | 40,300,679 |
| MOE H3K79me3 rep1 | Paired-end | 2x50 | 50,766,610 |
| MOE H3K79me3 rep2 | Paired-end | 2x50 | 40,112,431 |
| MOE DNAse HS rep1 | Paired-end | 2x50 | 114,301,834 |
| MOE DNAse HS rep2 | Paired-end | 2x50 | 67,428,929 |
| Cerebellum H3K27ac | Single read | 1x50 | 24,562,849 |
| Cerebellum H3K4me1 | Single read | 1x50 | 31,432,377 |

Table 1: Summary of sequencing experiments: Type of sequencing, paired end, and total number of mapped and filtered reads for each experiment

| OR enhancer | nearest OR | mm9 coordinates | GFP + embryos |
|---|---|---|---|
| RHODES | olfr1413 | chr1:94441500-94443500 | yes (n=56) |
| LEFKADA | olfr1371 | chr11:52014700-52015900 | |
| SIFNOS | olfr464 | chr11:87749900-87751500 | yes (n=65) |
| CRETE | olfr139 | chr11:73848400-73850199 | yes (n=55) |
| SYMI | olfr51 | chr11:50812200-50813599 | yes (n=20) |
| KARPATHOS | olfr726 | chr14:50681500-50682900 | |
| LESVOS | olfr749 | chr14:51377700-51379500 | |
| SKIATHOS | olfr1513 | chr14:52957700-52959300 | |
| H | olfr1507 | chr14:53166900-53168700 | yes (n=60) |
| LEMNOS | olfr279 | chr15:98340700-98342300 | |
| GAVDOS | olfr181 | chr16:58956100-58957500 | |
| TINOS | olfr1505 | chr19:14170300-14171700 | yes (n=60) |
| AKTHRA | olfr1505 | chr19:14300700-14301900 | |
| ALONISOS | olfr1505 | chr19:14314300-14315900 | |
| CORFU | olfr1501 | chr19:14321300-14322900 | |
| LIPSI | olfr362 | chr2:36982900-36984500 | yes (n=50) |
| EVIA | olfr1301 | chr2:111566700-111567700 | |
| HYDRA | olfr1318 | chr2:112049700-112051300 | |
| OTHONI | olfr1402 | chr3:97111500-97113500 | |
| SKORPIOS | olfr1402 | chr3:97113500-97114900 | |
| PONTIKOS | olfr1402 | chr3:97141900-97143500 | |
| FOLEGANDROS | olfr1402 | chr3:97294700-97296500 | |
| KOS | olfr266 | chr3:106678500-106680100 | |
| NISYROS | olfr266 | chr3:106683500-106684900 | |
| NIMOS | olfr62 | chr4:118314100-118315900 | |
| SFAKTIRIA | olfr448 | chr6:42818900-42820700 | yes (n=58) |
| KALYMNOS | olfr215 | chr6:116563100-116564900 | |
| KEFALLONIA | olfr1350 | chr7:6497300-6499100 | yes (n=52) |
| THIRA | olfr309 | chr7:93443100-93444300 | |
| P | olfr713 | chr7:114190700-114192100 | |
| IOS | olfr514 | chr7:115939700-115941900 | yes (n=82) |
| FOURNI | olfr523 | chr7:147372100-147374100 | yes (n=62) |
| MILOS | olfr545 | chr7:109661022-109662409 | |
| IKARIA | olfr521 | chr7:106936000-106937199 | yes (n=52) |
| SIKINOS | olfr24 | chr9:18525100-18526900 | |

Table 2: Candidate OR enhancer locations (mm9 genome coordinates), nearest olfactory receptor gene, and results of zebrafish expression assay. Number of zebrafish embryos counted that expressed GFP under control of each enhancer is indicated.

| | Forward Primer | Reverse Primer |
|---|---|---|
| Sikinos (chr9) | caccTTTCTGCATCTCGTCCCTTT | GGGTGGGGGAACTTCTTAGT |
| Kalymnos (chr6) | caccTCTGAGGACCAATGAGAAGGG | GAGAGGCTTTTTCCCACCCAA |
| Nimos (chr4) | caccTGTTATGGGAATGCTGACCA | ACTAGCCAACTCAGCCCTCA |
| Nisyros (chr3) | caccGATTGTGGGAGGGGGTAACT | CACCACGATCAAAGCAACAC |
| Kos (chr3) | caccGAAGGCCTCTCAAAACTGACA | GCCTCAGGACTGCTTCTAGG |
| Folegandros (chr3) | caccAGGCAGCATTGTTGTTGTTG | TCATCACTCTCTGTCCCCATC |
| Pontikos (chr3) | caccTTAGGCAAAGATCGGCAAAG | TGCCATCCTGGAGTAGGAAC |
| Skorpios (chr3) | caccCATCCTGAAAAGCATCGTCA | TCTTTGGCTTCTGCCTCAGT |
| Evia (chr2) | caccTGCCAGAGGTAAATCCCATC | GATTTTTAGACAACTGGGACAGC |
| Corfu (chr19) | caccGCAGCCCTGGCATTTTAGTG | TCTCAATCCATTTGCCCCAGT |
| Alonisos (chr19) | caccCGACCCCATCATTTAACACA | GGGTGGGGTCTTCATTTGTA |
| Akthra (chr19) | caccTTTATTTGACGCCCCCTTCT | TGAACCAGAAATTGGCACTG |
| Skiathos (chr14) | caccCTTCCCTACCTCCCCATCAT | TGCATTCGTAGCAGCAAGAC |
| Karpathos (chr14) | caccCAGAAGGTTCCACCCTAAGC | ACGAATCGACAGACACATGG |
| Othoni (chr3) | caccGAATGGTCCCTGACACCCAATAAAG | CTGCTCCTGACTCCATTTCC |
| Hydra (chr2) | caccGCCTACCCTTGTGGAGCATA | CTGATGCTGAACCTGTCCTG |
| Gavdos (chr16) | caccTGTAAATCCTCACAAAATATCTCCAA | TTGATCTACGAAGAATCACTGAAGA |
| Milos (chr7) | caccCCTTCTCTGAAATGCGAACA | GGGGCTAACAATGCCAAATA |
| Lesvos (chr14) | caccTGCCTGTAGGCAAGACTGTG | ATTGGCCAAAACTGCAAGAC |
| Lemnos (chr15) | caccGGCTGTGTTTTGCCCATAGT | TGTAGCTCAGAGGCAGAGCA |
| Lefkada (chr11) | caccGGGAAAACCATGCAGAGCTA | GCCTAAATTGCGTTCTGAGC |
| Symi (chr11) | caccGAATTTGGAGGGAGGAAAGG | CGACAGAACAAGATGCACCA |
| Crete (chr11) | caccGGTAGATGGTGGCACAGGAC | ACACACAGGAGGGAGAATGG |
| Kefallonia (chr7) | caccGGGAAGTATTGCTGCCAAGA | GTGTGTGTGTGTGCCTTTAATGt |
| Sfaktiria (chr6) | caccCAGGCATGCTCTTCCTACCT | GGGGAGGTTTGTATGTGCTT |
| Lipsi (chr2) | caccTGGCATCTAAAAACTATGGATAGAAA | TCTAGAGTTGAAATATCATGGGAAG |
| Sifnos (chr11) | caccGCAGCTGTTCCCTCTGTTTC | AGAGCTGTGTGCCAGGAAAT |
| Tinos (chr19) | caccTCTCCTCCACGGTCCTGTGCT | TGACAGGTCATGGCAGACCCC |
| H (chr14) | caccTTGTCCTTCAAAAAGCCAAAA | GCCCATTAGCAACCTTACGA |
| Rhodes (chr1) | caccGGCAACATCTCCAATATGAGG | GTGCTCATGCACACACACAC |
| Ios (chr7) | caccTTAGCCCCAGGGAGAAACTT | CCCTTATGTCAGAAGTGGAGATG |
| Ikaria (chr7) | caccCAGAGGTGCTCAGGCAGTG | GCCCAGCCTAGGGCTTAGT |
| Fourni (chr7) | caccGCATTTGTTGAGGGGACATTTTTC | CATCTCAAGCTGCCTTT |

Table 3: Oligos used to clone candidate OR enhancers into pENTR gateway construct.

# Chapter 3: An enhancer interaction network associated with transcription

## Introduction

The olfactory system has the ability to detect and distinguish among an astounding number of olfactory stimuli (Bushdid et al., 2014). This vast receptive field is afforded by the large repertoire of olfactory receptors (OR), which, in most mammals, are encoded by more than a thousand genes located in numerous genomic clusters throughout the genome (Buck and Axel, 1991; Sullivan et al., 1996; Zhang et al., 2004). Olfactory receptors are expressed in olfactory sensory neurons (OSNs) in a monogenic, monoallelic, and seemingly stochastic fashion (Chess et al., 1994). For each OSN, the identity of the expressed OR determines the spectrum of chemicals that it responds to and its connectivity to the brain (Barnea et al., 2004; Wang et al., 1998). The dual role of ORs in odor detection and axon guidance makes the singularity of their expression critical for olfactory perception; were multiple ORs coexpressed in each OSN, the topographic map of OSN projections to the olfactory bulb would be perturbed, likely resulting in reduced olfactory sensitivity and resolution.

The continuous transcription of a single OR is maintained by an OR-elicited feedback signal that stabilizes the expression of the chosen OR and prevents the activation of additional ones (Ferreira et al., 2014; Lewcock and Reed, 2004; Serizawa et al., 2003; Shykind et al., 2004). This feedback uses components of the unfolded protein response (UPR) to detect the newly translated OR in the endoplasmic reticulum and to induce transient translation of transcription factor Atf5 (Dalton et al., 2013). Atf5 orchestrates, among others, the expression of Adcy3, the major adenylyl cyclase in the OSNs that is necessary for stable OR transcription and OSN differentiation. Adcy3 expression makes OR choice permanent by signaling for the downregulation of Lsd1, a lysine demethylase

with dual co-activator and co-repressor activities that regulate OR expression (Lyons et al., 2013). Lsd1 activates OR transcription by demethylating lysine 9 of histone H3, but it can also repress OR transcription by demethylation of lysine 4 of histone H3. Consequently, although Lsd1 is required for initiation of OR transcription, the timely downregulation of its expression is essential for the stable commitment to the chosen OR (Lyons et al., 2013). The feedback system that targets Lsd1 is only possible because OR silencing, via the hallmarks of constitutive heterochromatin, occurs early, during OSN differentiation and before the onset of OR transcription (Magklara et al., 2011).

Remarkably, although mutations in any component of this feedback signal destabilize OR choice and, likely, cause gene switching, they do not result in OR coexpression, suggesting the OSN is able to transcribe at high levels only one OR allele at a time. Furthermore, preventing the heterochromatic silencing of OR genes by conditional deletion of H3K9 methyltransferases G9a and Glp results in only low frequency OR coexpression (Lyons et al in revision). Thus, de-silencing an OR allele, albeit necessary, is not sufficient for its robust transcription, suggesting that the role of this feedback loop is to maintain the singularity of OR expression but not to generate it (Magklara and Lomvardas, 2013; Rodriguez, 2013).

Some insight to the puzzle of OR expression came from the observation that most OR loci converge into a few, OR-specific heterochromatic foci following the downregulation of Lamin b receptor (Lbr) in the OSN lineage (Clowney et al., 2012). The active OR allele in each OSN resides outside of these foci, supporting a role of the spatial compartmentalization between active and inactive OR alleles to the singular OR expression(Armelin-Correa et al., 2014). Indeed, disrupting the nuclear architecture of

OSNs violates the "one receptor per neuron" rule, causing co-expression of at least 60 OR alleles per neuron, albeit at reduced levels, and disruption of the topographic map in the olfactory bulb (Clowney et al., 2012). Thus, similar to the observation that loss of heterochromatic silencing is not sufficient for robust OR expression, escape from repressive OR foci *per se* is also not adequate for OR transcription at high levels. Therefore, robust OR transcription may depend on two mechanistically distinct molecular events: first the desilencing of an OR allele, and second its association with an extremely limited activator, a transcriptional "selector" that can only enhance the transcription of one OR allele by virtue of its singularity (Magklara and Lomvardas, 2013).

We previously hypothesized that an intergenic OR enhancer, H, could provide this singularity because it frequently associates with transcriptionally active OR alleles from the same or different chromosomes (Lomvardas et al., 2006). However, deletion of the H enhancer affects only the expression of three linked and proximal ORs (Fuss et al., 2007; Khan et al., 2011; Nishizumi et al., 2007). Redundancy for the function of H as a *trans* enhancer, provided by additional H-like elements, could explain why the physical association with H appears to be genetically superfluous for the transcription of most ORs (Williams et al., 2010). This model, which predicts an intricate network of genomic interactions (Bargmann, 2006), ascribes two distinct roles to each intergenic OR enhancer: a critical function as a *cis* regulatory element, which may open up the local chromatin architecture orchestrating the first step of OR choice, and a redundant function as a *trans* enhancer which, together with other enhancers, facilitates high rates of OR transcription.

Circularized chromosome conformation capture sequencing (4C-seq) analyses of sorted OSNs, combined with two- and three-color DNA fluorescent in situ hybridization (FISH) experiments, demonstrate the convergence of multiple enhancers over the chosen OR. Hi-C analysis with chromatin from the whole MOE, along with DNA FISH experiments, revealed extensive inter- and intra-chromosomal interactions between most of these elements in olfactory neurons. These interactions occur outside the heterochromatic OR foci in olfactory neurons, indicating that they are not a consequence of the linked ORs aggregating in the nucleus. Finally, disruption enhancer interactions by ectopic expression of Lbr in mature OSNs results in significant decrease of *trans* enhancer interactions accompanying the significant reduction of OR transcription. Our experiments, which reveal the interaction landscape of olfactory receptor enhancers, are consistent with a model of singular OR choice that depends upon the convergence of multiple enhancers in a three-dimentional enhanceosome (Thanos and Maniatis, 1995).

**Results**

Our data thus far provide a comprehensive epigenetic and genetic characterization of intergenic DNA elements that may act as OR enhancers. To examine whether these elements, like the H enhancer, also associate with active OR genes in *trans* (Lomvardas et al., 2006), we performed 4C on an isolated population of OSNs expressing the same OR, olfr1507. We chose olfr1507 for this experiment because its expression depends on the linked H enhancer residing ~75Kb upstream of this gene. Thus, it provides an ideal gene locus for testing the hypothesis that a *cis* enhancer may act in concert with *trans* enhancers. 4C was performed on fluorescence activated cell (FAC) sorted neurons from olfr1507iresGFP knock-in mice, and libraries were amplified with inverse PCR primers

anchored at the olfr1507 promoter as previously described (Clowney et al., 2012). 4C libraries from GFP positive and negative cells were analyzed by qPCR to quantitate the relative enrichment of various DNA loci. Strikingly, several of the newly identified sequences are enriched in this library, at levels approaching the enrichment levels of H (Figure 7A). Enrichment is significantly reduced in GFP negative cells, which suggests that these associations are restricted to cells that transcribe olfr1507. Two color DNA FISH analysis (Figure 7B) verified that the olfr1507 locus frequently co-localizes in *trans* with the three most highly enriched loci, Lipsi (chr2), Sfaktiria (chr6), and Crete (chr11), in OSNs immunolabeled with an anti-olfr1507 antibody (~63% of olfr1507+ OSNs, n=124). Similar results were obtained by performing FISH in olfr1507iresGFP MOE using anti-GFP immunolabeling (Figure 8A). Three-color DNA FISH (Figure 7C) revealed that olfr1507 co-localizes with both Lipsi and Crete in 16% of olfr1507+ OSNs, a highly significant increase over the frequency of co-localization of the three loci in olfr1507- OSNs (0.2%, n=406, p=2E-10, chi-square test). Thus, as the network of putative enhancer interactions becomes more complex, it becomes further associated with OR transcription.

To explore the long-range interactions of olfr1507 in an unbiased fashion we generated 4C-seq libraries generated from GFP positive and negative cells from olfr1507iresGFP mice (Figure 8B). In agreement with our qPCR analysis, 4C-seq revealed multiple contacts between olfr1507 and fifteen candidate OR enhancers, of which nine are functional in the zebrafish reporter assay (Figure S8C-E, Table 4). Interactions with OR enhancers were significantly ($p<<0.01$, Wilcoxon t-test) stronger in olfr1507-expressing OSNs compared to the negative population (Figure 7E). The network of

interchromosomal interactions between olfr1507 and predicted OR enhancers is depicted in Figure 7F.

The analysis of olfr1507 cells suggests that interactions between OR enhancers are extensive in OSN nuclei and may be irrespective of the specific OR allele that each cell expresses. To examine this possibility we performed Hi-C analysis using crosslinked chromatin prepared from the whole MOE. To specifically interrogate the interactions of OR enhancer candidates, we performed a modified Capture-C protocol (Hughes et al., 2014) and enriched our library using oligos that tile the 35 OR enhancers (see extended experimental methods). Our analysis revealed that 32/35 elements associate in high frequency with other enhancers from this repertoire. These interactions appear to be highly specific; there is a significant ($p<<0.01$ Wilcoxon t-test) 20-fold enrichment for reads that span two different potential OR enhancers compared to reads that span an OR enhancer and one of the other MOE enhancers identified by our epigenetic analysis (Figure 9A). Moreover, within the observed repertoire there are "promiscuous" enhancers that form frequent interactions with many other elements (Evia, Gavdos), while others form fewer interactions (Nimos, Skiathos). A contact matrix depicting the pairwise frequencies of these interactions organized by hierarchical clustering reveals the existence of four clusters of potential enhancers exhibiting similar frequencies of interactions (Figure 9B). Enhancers located on chromosomes 2,3,7, and 16 make the most frequent contacts with each other and with enhancers from other chromosomes (Figure 9C).

To obtain cellular resolution and to independently verify the Hi-C data, we performed an extensive two color DNA FISH analysis on sections of the MOE using BAC probes for

candidate OR enhancer loci (Figure 9E, Figure 10). DNA sequences that interact infrequently by Hi-C, like Nimos, exhibit a low frequency of co-localization with other OR enhancers both in OSNs and in sustentacular cells, a non-neuronal cell type of the MOE that we used as an internal control (Figure 9D). In contrast, increased Hi-C interactions correspond to enhancer-enhancer co-localizations by DNA FISH that are restricted to OSNs. On average, enhancer-enhancer co-localizations are six times more prevalent in OSNs than in sustentacular cells (n=3,264 nuclei, $p=10^{-44}$ chi-square test), discounting interactions that were infrequent in the Hi-C analysis. As in the Hi-C experiment, enhancer-enhancer interactions are specific to OR enhancers, as co-localizations with a predicted MOE enhancer located nearby Ebf3 are not significantly increased in OSNs compared to sustentacular cells (Figure 9D).

The only exception to the correlation between the Hi-C and DNA FISH analysis comes from Symi, an element on chromosome 11, which co-localizes frequently with other enhancers by DNA FISH but not by Hi-C. The local chromatin architecture may affect restriction enzyme accessibility and thus bias chromosome conformation capture experiments, providing a possible explanation for the difference between the two assays. It is also worth noting that the co-localization frequencies between Ios and Ikaria were not significantly reduced in sustentacular cells, probably due to the genomic linkage of these DNA elements which both reside on chromosome 7 at 9MB distance. Interestingly, sequences engaged in interchromosomal contacts include those that were not functional in the zebrafish reporter assay, suggesting that they may have functional regulatory roles in the mouse. In any case, even without *a priori* knowledge of the molecular identity of

each neuron, potential intergenic enhancers interact frequently with each other specifically in OSN nuclei.

An important question emerging from our analyses is whether frequent enhancer interactions are a consequence of OR gene aggregation in OSN nuclei. The fact that we detect enhancer co-localization with olfr1507 more frequently in olfr1507+ cells suggests that these interchromosomal associations would preferentially occur outside the repressive OR foci, which is where the transcriptionally active ORs also reside (Clowney et al., 2012). Two-color DNA FISH shows that 60% of H or Lipsi alleles co-localize with a complex DNA FISH probe that recognizes most OR loci (panOR probe), a result which should be expected based on the close proximity of these two OR enhancers with linked OR genes (Figure 11A). However, three-color DNA FISH shows that H and Lipsi co-localization occurs outside the OR foci ($p<10^{-12}$, n=1530, chi-square test), within euchromatic nuclear territories (Figure 11 A, 11B). Thus, enhancer interactions in *trans* are not a simple product of the convergence of linked ORs.

It is possible that the differentiation dependent aggregation of OR loci instead facilitates enhancer interactions by bringing the elements into close proximity. To test this we analyzed the frequency of interchromosomal associations between potential OR enhancers in OSNs that express Lbr. We previously showed that ectopic Lbr expression in mature OSNs disrupts the aggregation of OR genes and the interaction of the H enhancer with OR genes in *trans* but not in *cis* (Clowney et al., 2012). 4C-qPCR analysis of control and Lbr-expressing FAC-sorted OSNs revealed significant reduction in the frequency of *trans* interactions between potential OR enhancers upon Lbr expression (Figure 11C), supporting a role for the unusual nuclear architecture of OSNs, and

possibly for the aggregation of OR loci, in the frequent association of predicted OR enhancers from different chromosomes. Importantly, since OR transcription is significantly reduced in Lbr-expressing OSNs (Clowney et al., 2012), these data provide an independent genetic manipulation whereby the specific disruption of *trans* enhancer interactions reduces OR transcription rates. Thus, the same process that contributes to the effective silencing of OR transcription, the aggregation of OR genes in heterochromatic foci, may also facilitate the activation of a single allele by increasing the probability of enhancer-enhancer interactions. Of course, indirect effects could explain the transcriptional effects of the nuclear re-organization induced by Lbr expression in mature OSNs.

**Discussion**

Our experiments revealed an unprecedented network of genomic interactions occurring predominantly between putative OR-associated enhancers from different chromosomes. Interactions identified by Hi-C were verified by extensive DNA FISH experiments, which demonstrated that sequences from different chromosomes co-localize in up to 35% of the OSN nuclei. 4C-seq from FACsorted OSNs and three-color DNA FISH experiments suggest that more than one enhancer co-localizes with an OR gene in the neurons that transcribe that OR. Due to technical limitations we were unable to perform three-color RNA/DNA FISH. However, our previous work showed that H interacts in *cis* and in *trans* with the transcriptionally active OR allele (Lomvardas et al., 2006), thus it is unlikely that other OR enhancers would interact with the non-chosen allele. Moreover, allele specific ChIP-qPCR analysis in olfr1507+ OSNs has shown that the inactive olfr1507 allele is epigenetically similar to other ORs in these neurons (Magklara et al.,

2011). In other words, in olfr1507 expressing OSNs, while the transcriptionally active allele is euchromatic and accessible, the inactive allele is indistinguishable from all other silent OR alleles in this neuron and resides in heterochromatic territories (Armelin-Correa et al., 2014). Thus, given that *trans* enhancer interactions occur preferentially outside of the OR foci, the parsimonious assumption is that multiple OR enhancers coalesce over the transcriptionally active OR allele.

A significant question emerging from our observations regards the functional significance of the convergence of multiple enhancer elements over the chosen OR. Recent data revealed that many developmentally regulated genes require numerous enhancers for their proper expression. Experiments in the developing mouse embryo have shown that multiple enhancer sequences act in a coordinated fashion to activate Hox genes in various tissues, and similar observations have been made for the activation of protocadherin gene clusters (Andrey and Duboule, 2014; Delpretti et al., 2013; Guo et al., 2012; Montavon et al., 2011; Noordermeer et al., 2014). Thus, enhancers may act in an additive or even synergistic fashion to increase transcription rates. Given that ORs represent the most abundant protein coding mRNAs in OSNs (Tom Maniatis, George Mountoufaris, personal communication and our own unpublished observations), it is possible that the convergence of multiple enhancers contributes to the robustness of OR transcription. Similar coordinated action between multiple loci has been described for the activation of the IFNbeta gene in human cells, which is also expressed in a stochastic fashion and at high rates upon virus induction (Apostolou and Thanos, 2008). Finally, genetic experiments suggest that a DNA sequence may act as a *trans* enhancer during the

stochastic photoreceptor gene choice in drosophila ommatidia (Johnston and Desplan, 2014).

Beyond the striking extent of the interchromosomal associations between putative OR enhancers and OR genes, we do not have direct evidence of a role of these interactions in OR transcription. Previous work showed that deletion of H does not affect expression of more than 3 proximal OR alleles, despite the physical association of this enhancer with multiple other ORs. Similarly, deletion of P or Lipsi appears to affect the expression of linked OR alleles only, a result also observed in H,P double KO mice (Khan et al., 2011). However, with at least 15 enhancer elements interacting frequently with olfr1507 promoter in olfr1507-expressing OSNs, it should be expected that the deletion of a single *trans* enhancer does not have detectable transcriptional consequences. Moreover, it is possible that upon deletion of one *trans* enhancer, a different one from the total of 35 putative enhancers participates in this enhancer complex. The detection of a transcriptional phenotype is thus rendered extremely challenging by straightforward genetic approaches, since an unattainable number of homozygote deletions may be necessary for detectable effects on OR transcription. In contrast, if each of these elements is required for a critical step in the transcriptional activation of *cis*-linked OR genes, such as mediating the recruitment of Lsd1 and the desilencing of the linked ORs, then each individual deletion may be sufficient for a detectable transcriptional effect in *cis* (Figure 11D-F). Alternatively, it is possible that the role of these sequences is to act as positioning signals that increase the recruitment frequencies of the linked ORs to an interchromosomal enhancer complex. In this case, as it was previously proposed (Khan et al., 2011), describing these DNA elements as enhancers may not be appropriate. In this

vein, the DNA elements that did not activate reporter expression in zebrafish OSNs may be necessary for the activation of their proximal ORs in the mouse even if they cannot activate transcription on their own.

To test this model of *trans* enhancement, we performed a genetic manipulation that could impair the ability of these putative OR enhancers to interact with each other in *trans*. Expression of Lbr in mature OSNs resulted in significant downregulation of OR transcription, to the extent that ORs are not detectable by RNA ISH or IF in the mutant OSNs. Although in both cases we cannot directly attribute the downregulation of OR transcription to the reduced frequency of *trans* interactions, these results are consistent with our model. Furthermore, recent experiments showed that an increase in the number of homeodomain and O/E sites on the promoters of transgenic ORs increases the frequency by which they are transcribed (Vassalli et al., 2011). Since the only way to increase the local concentration of binding sites for these transcription factors near an OR promoter is to recruit sites from other genomic regions, this result is consistent with our proposed hypothesis.

We wish to emphasize the existence of alternative, equally interesting interpretations of our data that do not invoke coordinated action between distant OR enhancers. For example, it is possible that the convergence of OR enhancers reflects the existence of specialized nuclear bodies, or factories with high affinity for these enhancers and for the protein complex that supports OR transcription. In this case the convergence of multiple enhancer elements would not be the cause of OR transcription but it could be a consequence of the spatial restrictions governing OR choice. Thus, instead of cooperating, the co-localized enhancers may compete with each other for the recruitment

of linked ORs to this nuclear factory. Such a nuclear body governing the stochastic and singular expression of more than a thousand genes is used for the regulation of vsg genes in trypanosome (Navarro and Gull, 2001). Moreover, we cannot exclude that the nature of these interactions is repressive rather than activating. Sequestering these putative enhancers into distinct nuclear territories may prevent them from interacting with their proximal ORs, essentially "decommissioning" a large number of elements in each OSN in the same fashion that OR aggregation contributes to the singularity of their expression. The observation that in Lbr-expressing OSNs we detect coexpression of multiple OR genes is consistent with this idea. Enhancer convergence may serve two functions: to eliminate the possibility of spurious, basal transcription of multiple ORs caused by the concomitant interaction of multiple enhancers with linked ORs, and to ensure that the transcription of the chosen OR allele occurs at an excessive rate.

In summary, our data are consistent with a model in which the robust transcription of an OR requires an enhancer in *cis* and numerous enhancers in *trans*. High levels of OR expression are necessary for activation of the Perk pathway via ER-stress, and it is likely that only ORs expressed above a certain threshold can elicit this feedback, a prediction consistent with the observation that transgenic ORs expressed at low levels from heterologous promoters can be coexpressed with endogenous ORs (Zhou and Belluscio, 2012) (Zhao and Reed, 2001). Since the vetting mechanism that stabilizes OR choice screens for both the quality and quantity of OR protein, ORs transcribed at suboptimal levels will be turned off by sustained Lsd1 expression. If the number of enhancers associating with an OR promoter indeed determines expression levels, then stable OR expression will occur only once a sufficient number of enhancer elements associate with

an OR promoter. Simple modeling of the observed experimental frequencies of pairwise enhancer interactions predicts that the co-localization of 16 different enhancers from a repertoire of 35 will occur only once in each OSN nucleus (see supplemental experimental conditions). Thus, depending on the actual number of enhancers needed to achieve feedback-eliciting levels of OR transcription, the limited, or even unique generation of a multienhancer complex may provide the elusive singularity of OR choice.

It seems counterintuitive that a sensory system critical for the animal's survival and reproduction would rely on a molecular mechanism as inefficient and probabilistic as the interchromosomal convergence of a large number of enhancer elements. However, unlike most developmental systems that are built upon tight spatiotemporal regulation, the peripheral olfactory system may be able to tolerate such a variable and often non-productive process because an efficient feedback mechanism is in place to ensure that terminal differentiation of olfactory neurons occurs only upon OR choice. This mechanism is compatible with the rapid evolution of the OR gene family, which is characterized by significant copy number variations among closely related species and significant polymorphisms within species, in accordance with the essential function of this gene family in adaptation. It remains to be seen if other fast evolving gene families involved in the perception of- and the protection from- the constantly changing external environment (Clowney et al., 2011) may be employing similar radical mechanisms for stochastic and mutual exclusive gene expression.

**Methods**

4C-qPCR: Olfactory epithelia from 35 4-6 week old olfr1507-ires-GFP heterozygote knockin mice were dissected and dissociated for half an hour using papain (Worthington). After stopping the reaction with albumin, dissociated cells were washed in 1xPBS. Cells were then fixed for 7 minutes at room temperature with methanol-free formaldehyde (Pierce 28906) diluted to 2% in PBS. Fixation was quenched by adding 1/10$^{th}$ volume of 1.25M glycine. Fixed tissue was collected by centrifugation at 200g for 5 minutes at 4°C, then washed twice with cold PBS. 8000 GFP positive and negative cells were collected by FAC sorting into 1xPBS medium, and spun down. Sorted cells were resuspended in 330uL lysis buffer (10mM Tris pH 7.5, 1mM EDTA, 1mM EGTA, 1% Triton, 1% NP-40, 0.125 mM PMSF, protease inhibitors (Roche)) and rotated at 4C for one hour. After lysis, chromatin was spun down, washed once with cold 1X NEB buffer 4, and resuspended in 50 uL 1.2x NEB4 with 0.3% SDS, and incubated at 37C with gentle shaking for one hour. Triton (1.8%) was added and tubes were incubated for another hour at 37C. Next we added 10 units MboI (NEB) and digested overnight at 37C. Digest was diluted into 250uL 1x T4 Ligase buffer (NEB) and treated with 2000 units T4 DNA ligase overnight at 16C and then 2 hours at room temperature. Ligated chromatin was reverse crosslinked with proteinase K (overnight, 65C), purified by phenol chloroform extraction, and ethanol precipitated, and pellet was resuspended in 50uL 1xTE. To generate 4C library we amplified from 10uL 3C DNA using inverse PCR primers adjacent to MboI sites in the olfr1507 promoter (For: TGAGCAAATACCATGCTGAAA, Rev: ACTAGTTACAGCAGTTACTTTGATTGA) Samples were amplified using Phusion Polymerase (NEB) for 25 cycles and purified

using Ampure XP beads (Beckman Coulter). qPCR was performed using the following primers:

| | Forward primer | Reverse primer |
|---|---|---|
| olfr1507 cds | ggggttctgagtaagtcggtg | ccagcgatgaggagggattc |
| H | gggtccctgaggaattcagt | agggtgcctctagtggttca |
| Lipsi | ttgtacctggggagttctgg | agttgctcctgatggccta |
| Sfaktiria | aggcaactttccagctgtgt | agttgaaaggctgcccagta |
| Crete | atctctgtggggatcattgc | acacacaggagggagaatgg |
| Ikaria | gctccctgctgctttaattg | ctagcaactggggagtggag |
| Lesvos | gtcccaattaaaagtgctccag | ttgccttttctttctgcttagg |
| Evia | tgcctgccaagaatctgtct | tagatgcaatccccatgtga |

4C-seq libraries were prepared for Illumina sequencing using Ovation Ultralow Library Kit (Nugen) and amplified for 10 cycles. Paired end reads containing 4C inverse PCR primer sequences were mapped using Bowtie2 (Langmead & Salzberg, 2012). After filtering, the total read counts for each enhancer region were used to generate a circos plot (http://circos.ca/), where the thickness of the line encodes the reads normalized to the total number of mapped reads.

DNA FISH probes: BAC probes were prepared by standard Nick Translation. PanOR probe was generated as described (Clowney et al., 2012) from genomic DNA eluted from a custom array tiling ~40Mb of class II OR clusters. Amplified library was labeled with High Prime (Roche) for Dig or Biotin and nicked to 200-500bp by Nick Translation (Roche). Probes were purified by G50 column, precipitated in the presence of 20x CotI, and resuspended in CamBio hybridization buffer. The following BACs were used to generate enhancer probes:

| | BAC clone |
|---|---|
| olfr1507 - H | RP24-290L16 |
| Lipsi | RP24-353M9 |
| Sfaktiria | RP23-65G23 |
| Crete | RP24-156B9 |
| Ikaria | RP24-212K15 |
| Symi | RP24-323I2 |
| Evia | RP23-97N5 |
| Ebf3 enhancer | RP23-200O13 |
| Ios | RP23-336P21 |
| Nimos | RP24-319L11 |

DNA FISH: Whole heads of P4 mice were directly embedded in OCT (Sakura) and frozen. 5 um cryosections were cut, air dried for 30 minutes, and fixed in cold 4% PFA for 5 minutes. Sections were permeabilized with PBS-0.1% Triton (PT), DNA was fragmented with 0.1M HCl for 5 minutes at RT, digested with 3mg/ml RNAse A in PT for one hour at 37C, and dehydrated at RT in ethanol series (70, 95, 100%), and slides were baked at 45C for ten minutes. Sections were denatured at 85C for minutes in 2xSSC, 75% formamide (Invitrogen), and immediately dehydrated in an ice-cold ethanol series (70, 95. 100%) and baked again at 45C for ten minutes. Probes were applied at 5 ng/uL concentration (BACs) or 25 ng/uL (complex panOR probe) under 8mm circular coverslips, sealed with rubber cement and incubated overnight at 37C. Slides were washed in 2xSSC, 55% formamide, 0.1% NP-40, three times for 15 minutes at 42C, rinsed with PT, blocked in TNB (Promega TSA kit), incubated 2 hours at room temperature with anti-dig or anti-biotin conjugated to DyLight fluors (Jackson Immunoresearch). Then the slides were washed in PT+8% formamide and mounted.

Immuno-DNA FISH: IF was performed as described below on 5 um sections, in the absence of DAPI. After IF, slides were post-fixed in 4% PFA for 5 minutes and rinsed. FISH protocol was continued from the HCl step.

Image analysis: Confocal images were collected on a Zeiss LSM700. All images are confocal slices. For FISH counts, stacks were collected with images 1uM apart and cells with OSN-typical centromeric focus were analyzed. Nuclei were counted if FISH signal was detectable for all BAC probes. BAC signals were colocalized if there was pixel overlap. BAC signals were colocalized with panOR probe if they were contained within discrete panOR foci in the Z-slice in which they were brightest.

Modeling of DNA FISH data: We performed least-squares regression analysis on enhancer pair co-localization frequencies to determine the probability of each individual enhancer to be co-localized in an OSN nucleus. To compute this we used the following formula: $x_i=(S\ x_j\ a_{ij})\ /\ (S\ x_j^2)$ , where $x_i$ and $x_j$ is the probability of each enhancer in a pair to be co-localized, and the observed frequency of colocalization $a_{ij} = x_i\ x_j$ . The average x for all enhancers tested is .47

Hi-C: 3C was performed as described (Lomvardas et al., 2006) using olfactory epithelium from 6-8 week old wildtype mice and MboI restriction enzyme. 3C DNA was sonicated for 180 seconds on a Covaris S220 and prepared for Illumina sequencing using Ovation Ultralow Library Kit (Nugen) and 13 cycles amplification. 200ng sequencing library was hybridized to biotinylated Seqcap EZ choice probes Nimblegen, Roche)

designed to target and tile OR enhancer regions. Repetitive sequences were blocked using Mouse Hyboc (Applied Genetics Laboratores) and adapter sequences were blocked using custom oligos. Following a double capture protocol per the manufacturer's instructions, probes were pulled down using Streptavidin M270 beads. After elution, the sequence capture library was amplified for 18 cycles using Illumina Truseq primers and Phusion DNA Polymerase (NEB).

Reads were mapped using Bowtie2 (Langmead & Salzberg, 2012) and analyzed using hiclib software package (Imakaev et al., 2012). Briefly, sub sequences from each end of a mate pairs were independently mapped in increments of 5 bp from 25 bp to 75 bp. Reads that were not mappable with a shorter subsequence were tried again until the maximum subsequence length was reached. After mapping, reads were filtered so that reads within 5 bp of the MboI restriction site, reads mapping to the same position, and reads coming from extremely small (< 100 bp) or large ( > 10^5 bp) were removed from the dataset. After filtering, the total read counts between each pair of enhancer regions was used to generate a circos plot (http://circos.ca/), where the thickness and width of the line between each pair of enhancers encoded the reads between those enhancers normalized to the maximum reads between any pair of enhancers.

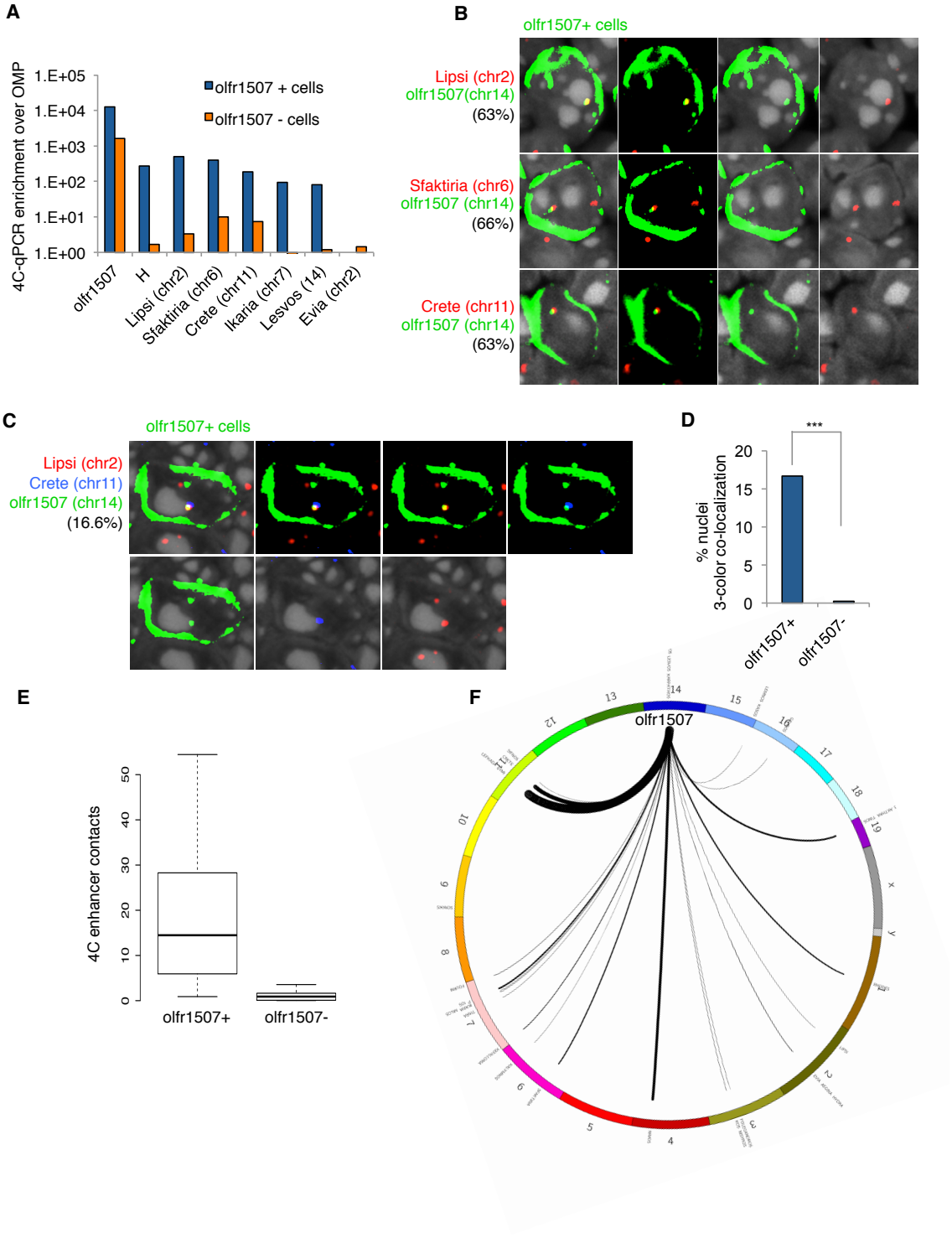

A
4C-qPCR enrichment over OMP
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
olfr1507 + cells
olfr1507 - cells
olfr1507
H
Lipsi (chr2)
Sfaktiria (chr6)
Crete (chr11)
Ikaria (chr7)
Lesvos (14)
Evia (chr2)
B
olfr1507+ cells
Lipsi (chr2)
olfr1507(chr14)
(63%)
Sfaktiria (chr6)
olfr1507 (chr14)
(66%)
Crete (chr11)
olfr1507 (chr14)
(63%)
C
olfr1507+ cells
Lipsi (chr2)
Crete (chr11)
olfr1507 (chr14)
(16.6%)
D
***
% nuclei
3-color co-localization
20
15
10
5
0
olfr1507+
olfr1507-
E
4C enhancer contacts
0
10
20
30
40
50
olfr1507+
olfr1507-
F
olfr1507

Figure 7: Multiple enhancers interact in *trans* with a transcribed OR

**(A)** 4C from olfr1507-ires-GFP FAC sorted cells, library constructed by inverse PCR off olfr1507 promoter. qPCR enrichment of candidate OR enhancers relative to OMP in GFP positive and negative cells. Error bars display SEM between triplicates. (**B)** IF staining for olfr1507 (green) and two-color DNA FISH for olfr1507 DNA (green) and Lipsi, Sfaktiria, and Crete DNA (red). DAPI is nuclear stain (gray). Percent of olfr1507 positive cells containing co-localized red and green probes is indicated for each combination. **(C)** IF staining for olfr1507 (green) and DNA FISH for olfr1507 (green), Lipsi (red) and Crete (blue) DNA. DAPI is nuclear stain (gray). Percent of olfr1507 positive cells containing three color co-localization is indicated. **(D)** Quantification of C. Percent olfr1507 positive and negative OSNs containing three-color co-localizations. **(E)** 4C-seq from olfr1507iresGFP positive and negative cells, library constructed by inverse PCR from olfr1507 promoter. Average number of contacts for OR enhancers are plotted. Y-axis is RPKM that span two different putative enhancers at an expected ligation site. **(F)** Circos plot (Krzywinski et al., 2009) of OR enhancers and 4C-seq contacts with olfr1507 promoter in olfr1507+ cells. Lines are weighted by frequency of interaction (RPKM that span two different putative enhancers at an expected ligation site).

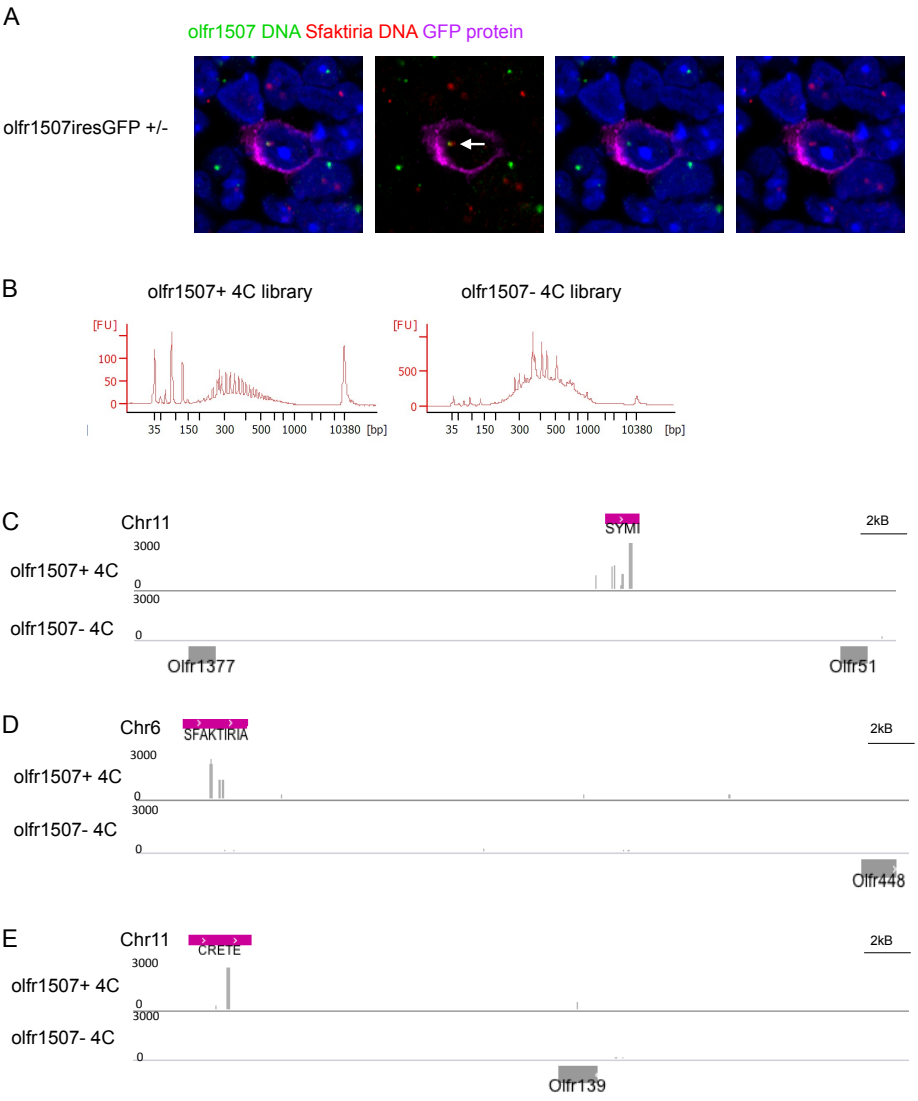

A
olfr1507 DNA Sfaktiria DNA GFP protein
olfr1507iresGFP +/-
B
olfr1507+ 4C library
olfr1507- 4C library
[FU]
100
50
0
500
35 150 300 500 1000 10380 [bp]
C
Chr11
SYMI
2kB
olfr1507+ 4C
olfr1507- 4C
3000
0
Olfr1377
Olfr51
D
Chr6
SFAKTIRIA
Olfr448
E
Chr11
CRETE
Olfr139

**Figure 8: Multiple enhancers interact in *trans* with a transcribed OR**

**(A)** IF-FISH in olfr1507-ires-GFP MOE. IF for GFP (magenta), FISH for olfr1507 DNA (green) and Sfaktiria (red). Arrow indicates co-localized probes. DAPI is nuclear stain (blue). (**B)** Bioanalyzer trace of 4C-seq libraries generated from olfr1507+ and olfr1507- cells. Y axis is fluorescence units, X axis is base pairs. (**C-E)** 4C-seq reads at candidate OR enhancer loci Symi, Sfaktiria, and Crete. Sequencing tracks from olfr1507+ and olfr1507- libraries.

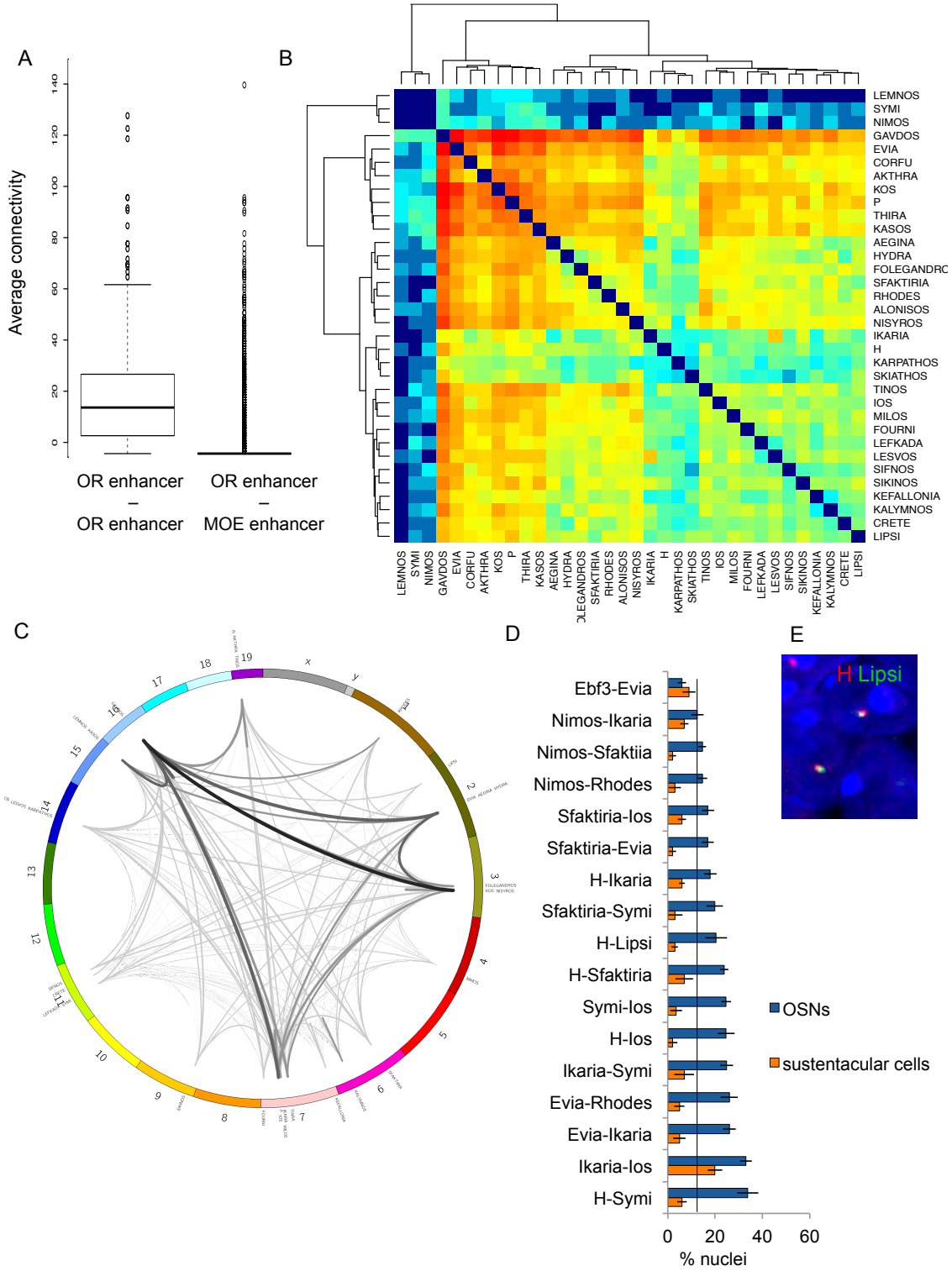

A
B
C
D
E
Average connectivity
0
20
40
60
80
100
120
140
OR enhancer
–
OR enhancer
OR enhancer
–
MOE enhancer
LEMNOS
SYMI
NIMOS
GAVDOS
EVIA
CORFU
AKTHRA
KOS
P
THIRA
KASOS
AEGINA
HYDRA
FOLEGANDROS
SFAKTIRIA
RHODES
ALONISOS
NISYROS
IKARIA
H
KARPATHOS
SKIATHOS
TINOS
IOS
MILOS
FOURNI
LEFKADA
LESVOS
SIFNOS
SIKINOS
KEFALLONIA
KALYMNOS
CRETE
LIPSI
Ebf3-Evia
Nimos-Ikaria
Nimos-Sfaktiia
Nimos-Rhodes
Sfaktiria-Ios
Sfaktiria-Evia
H-Ikaria
Sfaktiria-Symi
H-Lipsi
H-Sfaktiria
Symi-Ios
H-Ios
Ikaria-Symi
Evia-Rhodes
Evia-Ikaria
Ikaria-Ios
H-Symi
0
20
40
60
% nuclei
OSNs
sustentacular cells
H Lipsi

Figure 9: An intricate network of enhancer interactions in mouse OSNs

**(A)** Average Hi-C connectivity of OR enhancers with other OR enhancers compared to average connectivity of OR enhancers to MOE enhancers. Y-axis is reads spanning two different genomic regions normalized to the total number of reads. (**B)** Contact matrix depicting interaction frequency between candidate OR enhancers (red highest, blue lowest interaction frequency). Normalized read counts spanning two enhancer regions were divided into 20 bins, with 5 bins representing each color shade. Interactions between enhancers are hierarchically clustered. (**C)** Circos plot of OR enhancer chromosomal locations and Hi-C contacts. Lines are weighted according to frequency of enhancer-enhancer interactions. (**D)** Results of DNA FISH screen. X-axis is percent nuclei containing two-color co-localization between OR enhancer candidates, Y-axis is enhancer candidate pairs tested. OSN and sustentacular cell nuclei indicated. Vertical line is baseline OSN co-localization frequency. Error bars are SEM from multiple sections of the MOE. (**E)** Representative DNA FISH for H (red) and Lipsi (green) in OSN nucleus. The blue nuclear stain is DAPI.

**A**

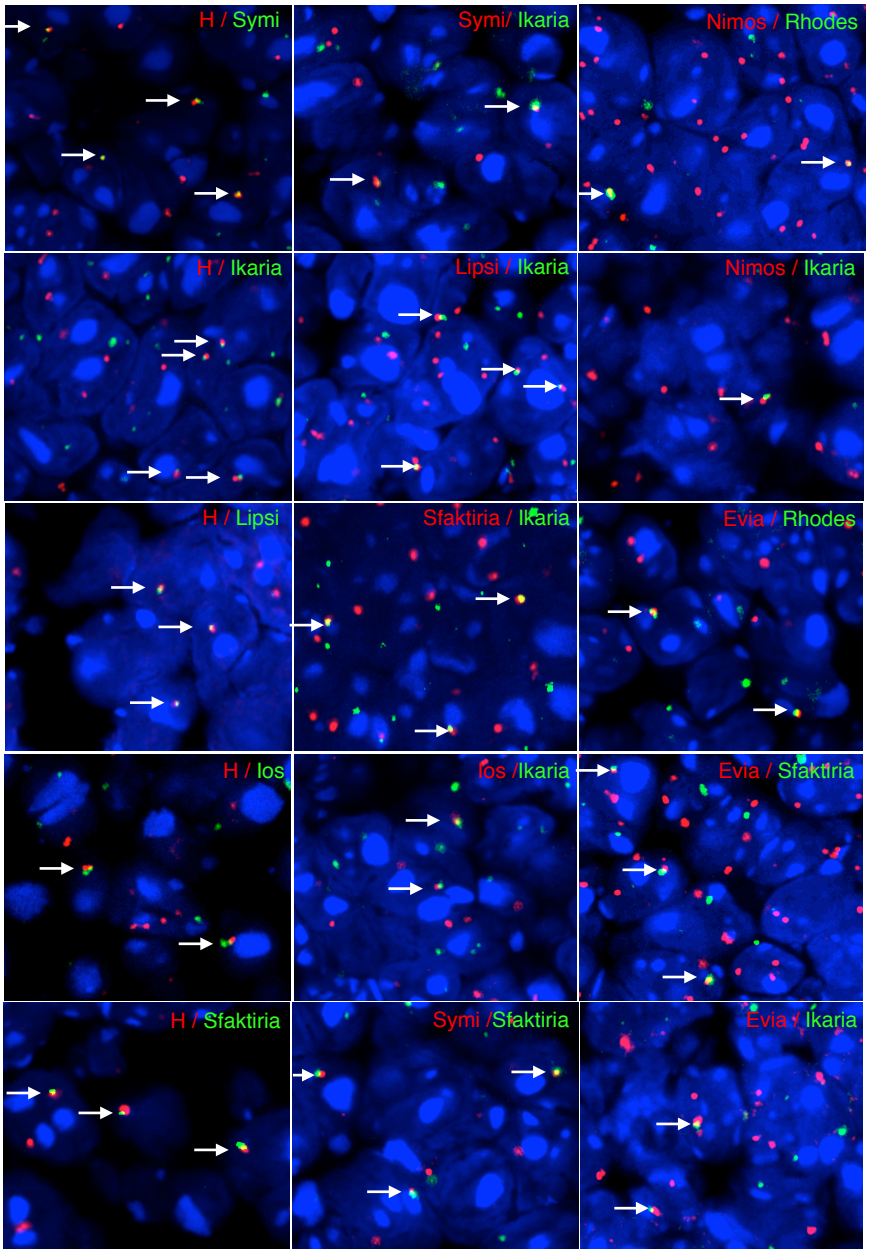

H / Symi
Symi/ Ikaria
Nimos / Rhodes
H / Ikaria
Lipsi / Ikaria
Nimos / Ikaria
H / Lipsi
Sfaktiria / Ikaria
Evia / Rhodes
H / Ios
Ios /Ikaria
Evia / Sfaktiria
H / Sfaktiria
Symi /Sfaktiria
Evia / Ikaria

Figure 10: Extensive enhancer pair co-localizations in OSNs

**(A)**. DNA FISH of OSN nuclei in sections of P4 MOE. Enhancer BAC probes in red and green are indicated, arrows indicate probe co-localizations. DAPI (blue) is nuclear stain..

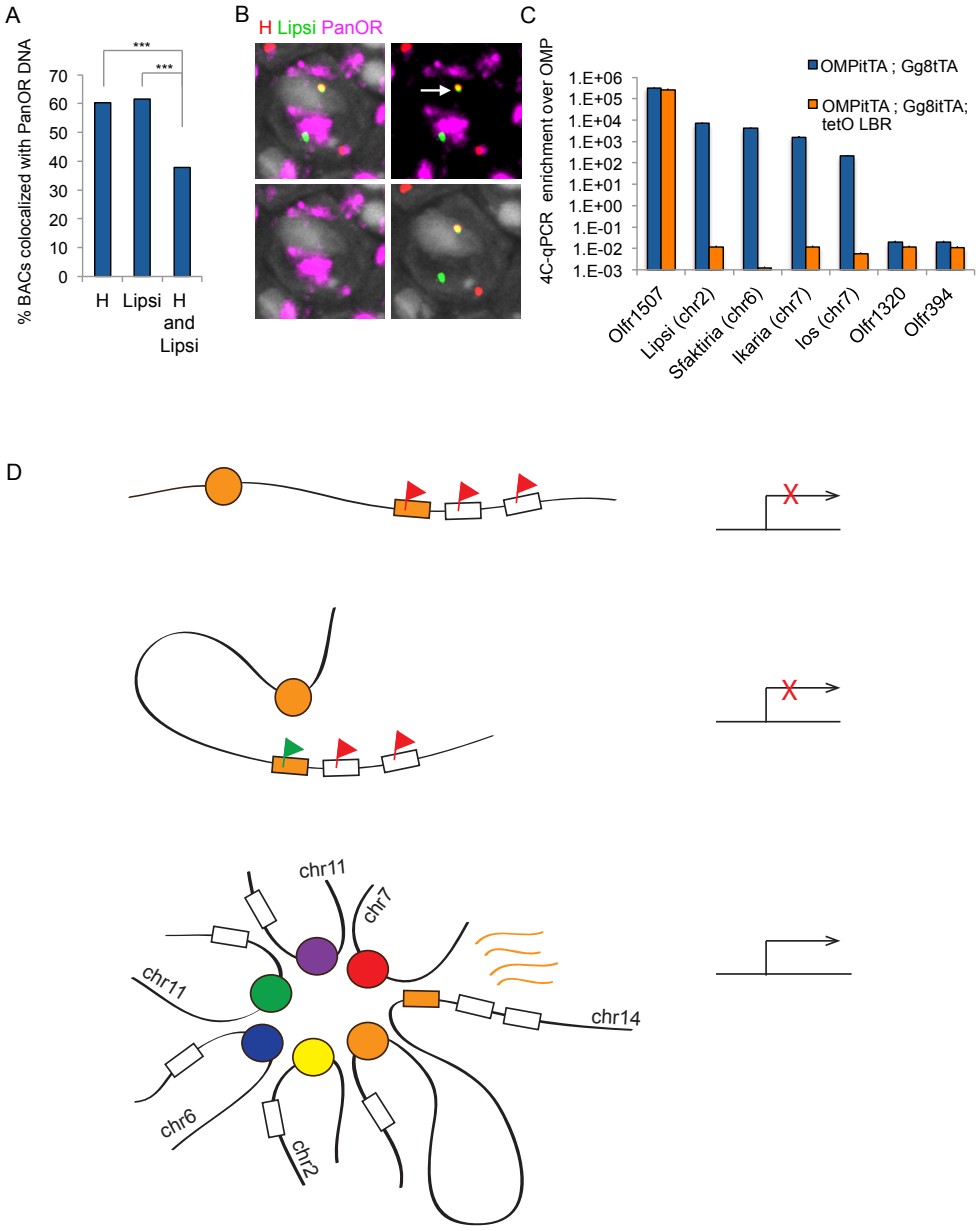

A
% BACs colocalized with PanOR DNA
70
60
50
40
30
20
10
0
***
***
H
Lipsi
H and Lipsi
B
H Lipsi PanOR
C
4C-qPCR enrichment over OMP
1.E+06
1.E+05
1.E+04
1.E+03
1.E+02
1.E+01
1.E+00
1.E-01
1.E-02
1.E-03
OMPitTA ; Gg8tTA
OMPitTA ; Gg8itTA; tetO LBR
Olfr1507
Lipsi (chr2)
Sfaktiria (chr6)
Ikaria (chr7)
Ios (chr7)
Olfr1320
Olfr394
D
X
X
chr11
chr7
chr11
chr14
chr6
chr2

Figure 11: OR gene aggregation promotes enhancer interactions

**(A)** Percent of nuclei in which individual BAC probes for H and Lipsi are co-localized with PanOR probe (two color DNA FISH), and in which co-localized H and Lipsi BAC probes are co-localized with PanOR probe (three color DNA FISH). **(B)** Three-color DNA FISH with PanOR probe (magenta), H BAC (red) and Lipsi BAC (green). DAPI is nuclear stain (gray). Arrow indicates co-localized enhancer probes located outside of OR foci. **(C)** 4C-qPCR analysis in Lamin b receptor (LBR) over-expressing mice (orange bars) and control mice (blue bars). 4C library is generated by inverse PCR from the H enhancer. Enrichment of candidate OR enhancers and olfactory receptor promoter sequences are normalized to control gene OMP. Error bars are SEM from duplicates.

**(D-F)** A model for different functions of an OR enhancer in *cis* and in *trans*. *S*chematic representations of different states in the OSN nucleus (left) and corresponding transcriptional outputs (right). An OR gene (orange box) located proximal to an enhancer (orange circle) is repressed by H3K9me3 (red flag, D). The *cis*-proximal enhancer may facilitate de-repression of the OR chromatin landscape (green flag, E) but is not sufficient for OR transcription. Multiple *trans*-interacting enhancers (colored circles) aggregate around the transcribed *cis*-proximal OR (orange box, F).

| OR enhancer | chromosome | # normalized reads | GFP+ embryos |
|---|---|---|---|
| Symi | 11 | 115.8 | yes |
| Crete | 11 | 54.4 | yes |
| Nimos | 4 | 35.4 | |
| Corfu | 19 | 21.1 | |
| Rhodes | 1 | 16.5 | yes |
| Ios | 7 | 16.4 | yes |
| Sfaktiria | 6 | 12.5 | yes |
| Alonisos | 19 | 9.8 | |
| Kefallonia | 7 | 7.5 | yes |
| Fourni | 7 | 4.4 | yes |
| Sifnos | 11 | 1.3 | |
| Ikaria | 7 | 1.1 | yes |
| Folegandros | 3 | 1.0 | |
| Karpathos | 14 | 1.0 | |
| Lipsi | 2 | 0.9 | yes |

Table 4: 4C-seq contacts in olfr1507+ and olfr1507- cells

# Chapter 4: The transcription factor landscape of olfactory receptor enhancers

## Introduction

In the first chapter I identified a set of 35 candidate OR enhancers that are distinguishable by a unique epigenetic signature, are broadly active in olfactory sensory neurons, and may be required for the choice of proximal OR genes. A next question is whether these enhancers also exhibit any identifiable transcription factor motifs in their DNA sequence. Identifying transcription factors that bind these enhancers will no doubt lead to a better understanding of olfactory receptor gene choice and the role that distant OR enhancers play in the choice. Identifying TF binding sites on these elements will also help elucidate what factors might be mediating the long-range interactions between OR enhancers that we described in Chapter 3.

Much work has been done dissecting the sequences of OR promoters to find regulatory motifs. Genome-wide analyses of OR promoter sequences find homeodomain transcription factor motifs and O/E motifs on promoters (Magklara et al., 2011b; Michaloski et al., 2011, 2006; Young, Luche, & Trask, 2011). Homoedomain transcription factor motifs match binding sites for Lhx2, and the M71 promoter is bound by Lhx2 (Hirota & Mombaerts, 2004a). Lhx2 is required for the expression of class II OR genes (Hirota, Omura, & Mombaerts, 2007a) . O/E motifs on promoters match O/E motifs on other olfactory genes, such as OMP and Adcy3 (M. M. Wang & Reed, 1993; S S Wang, Tsai, & Reed, 1997) and have been shown by yeast-two hybrid assays to be bound by Ebf transcription factors. Other than homeodomain and O/E motifs, no new transcription factor motifs have been identified by thorough examination of OR promoters.

The H and P enhancer sequences have been analyzed and found to contain a homeodomain transcription factor motif TAATG (Vassalli et al., 2011). Transgenic analysis of the H sequence in zebrafish showed that three homeodomain motifs in the H element are required for enhancer activity in the olfactory epithelium (Nishizumi et al., 2007). However, no other sequences have been characterized. The lack of information on the transcription factors that bind both proximal and distal OR *cis*-regulatory sequences is the motivation for the detailed analysis of transcription factor motifs on the 35 candidate enhancer sequences we have mapped.

**Results**

We took an unbiased computational approach to identifying possible transcription factor motifs on the OR enhancer sequences. In vivo footprinting by DHS-seq was recently successfully used to identify, in a high throughput fashion, regulatory sequences and factors that control cell-type specific differentiation (Neph et al., 2012; Thurman et al., 2012). Due to the extreme heterogeneity of cell types in the MOE, DNase I protected footprints were largely undetectable in the DHS-seq data as a result of poor sequencing coverage of the OR-proximal OSN enhancers. To overcome this limitation, we devised an enrichment strategy based on targeted sequence capture that would allow deep sequencing coverage of the OR enhancers, using capture probes tiling the OR enhancer candidates (Figure 12A, see methods section). After enrichment, DHS reads mapped primarily to the targeted regions and read coverage of the OR enhancers increased by ~10,000 fold (Figure 12B). Mapping the DNase I cleavage sites across the H enhancer reveals several sequences that are protected from DNase (Figure 12B) and that can be identified as footprints using a footprint calling algorithm (Neph et al., 2012). Across the

35 potential OR enhancers, 1040 footprints were identified with significant footprint occupancy scores (FOS<0.5).

To uncover transcription factor (TF) motifs in these DNase protected sequences we performed a *de novo* motif search using MEME (Bailey et al., 2009). First, we searched for motifs enriched in the verified OR enhancers. MEME identified four distinct motifs which matched predicted binding sites for Atf5 (Transfac P=0.003), Evx2/Lhx2 (Uniprobe P=0.001), Olf/Ebf (Uniprobe P=0.001) and Hdx (Uniprobe P=0.0001) (Figure 13A). Individual Atf5, Evx/Lhx, Hdx, and Ebf motifs correspond to DNAse protected footprints, examples are shown in Figure 12C and Figure 13B. Further, analysis of motif sites across all the verified OR enhancers in aggregate demonstrates widespread protection of the consensus sequences, as shown in Figure 12D.

Three of the TFs predicted to bind on these protected motifs have reported roles in OR regulation. Atf5 is part of the feedback mechanism that ensures stable and robust OR expression (Dalton et al., 2013). Lhx2 is a homeodomain transcription factor that is known to be required for the expression of a majority of OR genes (Hirota and Mombaerts, 2004). The protected Lhx motif is unlike the homeodomain motif in OR promoters that binds Lhx2 *in vitro* (Hirota et al., 2007), so we performed chromatin immunoprecipitation experiments to validate Lhx2 binding (Figure 13C). Ebf family members bind to O/E sites (M. M. Wang & Reed, 1993; S S Wang et al., 1997) which are found on OR promoters, the H and P elements, and promoters of several OSN-expressed genes (Vassalli et al., 2011). Because the consensus sequence in the OR enhancer footprints is similar but not identical to the O/E motif in promoters, we call it an O/E-like motif. Finally, Hdx is a homeodomain transcription factor; however Hdx transcript is not

detectable in the MOE by RNA-seq (data not shown), which leads us to believe that another homeodomain protein, such as Emx2 (McIntyre, Bose, Stromberg, & McClintock, 2008), may be binding these sequences.

To discover more transcription factor motifs we took a complementary approach to the *de novo* analysis described above. We compiled a list of all TF motifs from the Transfac database present on the 35 candidate OR sequences and calculated their footprint occupancy scores (FOS). Eleven TF motifs had median footprint scores less than one, indicating DNase I protection (Figure 14A). Homeodomain TF motifs comprise the majority of the footprinted motifs. Three TF motifs (Nobox, Foxj2, and C/EBPgamma) have p-values < 0.05 (one-sample, one-sided Wilcoxon) and exhibit protection of the consensus sequence in aggregate across all OR enhancers (Fig14B). C/EBPgamma, which is part of the family of CCAAT/enhancer element-mediated transcription factors, is the most highly expressed C/EBP family member in the MOE (Figure 14C) and a heterodimer for Atf5 (Ravasi et al., 2010). FoxJ2 is a forkhead transcription factor that is also expressed in the MOE (data not shown): FoxJ motifs are found in OR promoters (Clowney et al., 2011), and FoxJ1 KO mice exhibit glomerular defects in the olfactory bulb (Jacquet et al., 2011; Kolterman et al.). Finally, reduced (but not significant, p = 0.08) FOS was observed for a consensus sequence recognized by bromo- and PHD-finger domain transcription factor (BPTF), which is also expressed in the MOE but has not been implicated in OR transcription.

The in vivo footprinting analysis presented above indicates that OR-proximal OSN enhancers share several sequences that may be instructive for OR transcription. Of great interest is the motif predicted to be bound by Ebf family members, because the protected

consensus sequence enriched on active OR enhancers differs by two nucleotides from the canonical Ebf binding site (M. M. Wang & Reed, 1993) (Figure 15A). The canonical O/E motif that is enriched on OR promoters (Clowney et al., 2011a) is also present on verified OR enhancers (Figure 15B). To test the functional significance of this sequence polymorphism we cloned the canonical O/E and the O/E-like motifs into the zebrafish reporter construct. Remarkably, three copies of the novel O/E-like motif was sufficient to drive GFP expression in a large number of zebrafish OSNs, with minimal expression in other cell types of the zebrafish embryo (Figure 15C). The canonical O/E site did not drive GFP expression in zebrafish OSNs, nor did a mutant O/E-like motif in which we altered the two non-canonical nucleotides (Figure 15D).

To test the requirement for the O/E-like motif in enhancer activity, we generated two additional reporter transgenic mice, Lipsi-lacZ and Lipsi-lacZ with a mutated O/E-like motif (referred to as Lipsi (O/E-like mutant)-lacZ). Lipsi contains a well-protected O/E-like motif (Figure 15E), as well as an O/E motif, allowing us to test if the O/E-like site, in particular, is required for enhancer activity. Whole-mount x-gal staining shows widespread reporter expression in the MOE of Lipsi-lacZ transgenics, but very infrequent and low level reporter expression in the MOE of Lipsi (O/E-like mutant)-lacZ mice (Figure 15F, 15G).

Because in vivo footprinting, zebrafish reporter assays, and mouse transgenic experiments suggest a critical role of the novel O/E-like consensus sequence in OSN transcription that is not redundant to the canonical O/E site, we sought to identify the protein that may recognize this sequence. There are four Ebf family members in the mouse all of which are highly expressed in the MOE (S. Wang, 2002a). However, of the

four Ebf genes, only Ebf4 expression is restricted to the MOE, while the other three are also highly expressed in the vomeronasal organ (VNO) and other tissues (Song S Wang, Betz, & Reed, 2002). Given that the enhancer transgenes reported here (H, Sfaktiria, Kefallonia, and Lipsi) drive reporter expression only in the MOE and not the VNO, and taking into account that in zebrafish reporter assays the O/E-like motif drives expression in OSNs with remarkable specificity, we reasoned that Ebf4 recognizes this novel motif. To test this we performed ChIP-qPCR assays on whole MOE chromatin preparations using an antibody against Ebf4 (Figure 15H). In agreement with our hypothesis, ChIP-qPCR analysis reveals Ebf4 binding on enhancers containing a protected O/E-like motif, but not on sequences adjacent to these enhancers. Because antibodies available for the other Ebf family members were not suitable for ChIP assays (data not shown) we cannot compare the relative enrichment of Ebf1-3 to Ebf4 on enhancers with the O/E-like motif.

In addition to the O/E-like motif revealed by our in vivo footprinting assay, of great interest is the consensus for BPTF (Jordan-Sciutto, 1999), which is shared among several OR-proximal enhancers (Figure 16A) and is protected from DNase digestion at individual sites and in aggregate (Figure 6B,C). BPTF is the histone-binding component of the nucleosome remodeling complex NURF (Ruthenburg et al., 2011) and was previously shown to facilitate expression of Hox genes during development via chromatin remodeling at *cis*-regulatory sequences of Hox genes. (Wysocka et al., 2006) BPTF contains a bromodomain which binds acetylated lysines on one histone of a nucleosome, and a PHD domain which binds trimethylated lysines on another histone (Ruthenburg et al., 2011). We hypothesized that BPTF may be able to bind these

residues on histones from different nucleosomes, a possible mechanism by which it can mediate long-range interactions between different chromatin fibers.

To examine the functional significance of BPTF binding on OR-proximal enhancers we conditionally deleted BPTF in the MOE by crossing a floxed BPTF allele to a Foxg1-Cre driver that is expressed before the onset of OR expression (Hébert & McConnell, 2000; J. W. Landry et al., 2011). We refer to the Foxg1-cre, BPTF flox/flox mouse as the BPTF KO. These mice die perinatally, thus our analysis is restricted to E18.5 embryos. IF for MOR28 in sections from BPTF KO and control MOEs showed complete loss of MOR28 expression in the BPTF KO (Figure 16D). RNA *in situ* hybridization (RNA ISH) with a complex OR probe detecting several hundred ORs (see supplemental methods) also showed a dramatic reduction in OR expression (Figure 16E, 16J). GAP43 and Ncam1, markers of immature OSNs that are synchronous to OR expression (Iwema and Schwob, 2003; Krolewski et al., 2013, Lyons et al., 2013), are still expressed in BPTF KO MOEs (Fig 16 F,G), suggesting that the loss of OR expression is not caused by the loss of the OSN lineage. Finally, a general deficit in OR expression is corroborated by the loss of mature OSN markers, such as Adcy3 and Vglut2, in the BPTF KO MOEs (Figure 16H, 16I), because the terminal differentiation of OSNs is dependent upon stable OR expression (Lyons et al., 2013, Dalton et al., 2013). Taken together, biochemical data showing BPTF binding on OR-proximal enhancers and genetic experiments showing general loss of OR expression in the BPTF KO support a direct role of BPTF in OR gene regulation.

To test whether BPTF is required for the initial choice of OR or to maintain stable OR transcription, we crossed the conditional BPTF allele to Cre drivers that are expressed

after OR choice. Knocking out BPTF in cells expressing an OR does not result in down-regulation of OR expression. Specifically, we knocked out BPTF in MOR28 positive cells and assayed Cre reporter expression in the olfactory bulb (Figure 17A). IN this mouse we see proper targeting of Cre positive neurons to the olfactory bulb, indicating stable expression of MOR28. Further, we deleted BPTF in all mature olfactory neurons by crossing the conditional allele to an OMP-Cre driver and we assayed OR expression. Again, we see MOR28 expressing cells in the OMP-Cre knockout mouse (Figure 17B), and by RT-PCR for multiple OR cDNAs and markers of mature neurons we do not see OR expression affected in knockout mice relative to controls (Figure 17C). Taken together, these experiments suggest that BPTF is required for the initial choice of OR rather than maintenance of singular OR expression.

To test whether Bptf participates in the establishment or maintenance of interchromosomal associations between potential OR enhancers, we performed two-color DNA FISH in sections of control and Bptf KO MOEs. This analysis revealed a significant decrease in the frequency of interactions between H-Lipsi and H-Sfaktiria in Bptf KO OSNs (Figure 18A, p=0.0005, p=0.002, respectively, chi-square test n=356 nuclei), supporting a role in long-range genomic interactions. Importantly, the overall chromatin architecture of Bptf KO OSNs appears to remain intact and the aggregation of OR foci is not impaired (Figure 18B), suggesting a rather specific role of this protein in enhancer interactions. Of course we cannot distinguish between cause and effect with this genetic manipulation, and we cannot exclude indirect effects. Taking these caveats into account, the fact that deletion of Bptf results in reduced frequency of enhancer

interactions, concomitant with significant reduction of OR expression, is consistent with a role for these interactions in OR gene activation.

**Discussion**

To identify the genetic elements that determine the potency of these enhancers we modified a powerful, high throughput in vivo footprinting assay (Neph et al., 2012) to the needs of an extremely heterogeneous tissue, such as the MOE. Using sequence-capture technology we enriched our DHS-seq library for the 35 potential OR enhancers. With this approach we obtained unprecedented sequencing coverage on the sequences of interest, such that footprints of TFs that bind on these enhancers were revealed. Four in vivo occupied motifs are enriched in enhancers that support expression in zebrafish olfactory neurons and are predicted binding sites for Atf5, Evx/Lhx, Ebf, and Hdx transcription factors. The identification of Atf5 as a potential TF that binds on these enhancers is quite intriguing. We previously showed that Atf5 is translated transiently in the MOE and only in response to OR expression (Dalton et al., 2013). Atf5 translation plays a critical role in the stabilization of OR expression by activating Adcy3 expression and downregulating Lsd1. The fact that Atf5 may also bind to OR-proximal enhancers suggests that its role is not limited to mediating the OR-elicited negative feedback (Serizawa et al., 2003b, Lewcock and Reed 2004) but is also part of a positive feedback loop (Shykind et al., 2004b) that may enhance OR transcription.

Beyond providing a direct link between Atf5 and OR expression, our approach also revealed several additional motifs distributed across all 35 elements, thereby identifying novel TFs candidates that may bind to OR-proximal enhancers and regulate OR transcription. Here, we analyzed BPTF because a conditional KO allele was available (J.

Landry et al., 2008). Our genetic analysis shows that although OR expression is abolished if BPTF is deleted prior to OR activation, OR transcription is maintained when it is conditionally deleted in mature OSNs (data not shown). The genetic requirement of BPTF for only the initiation of OR transcription is intriguing because the Bromo- and PHD finger domains of BPTF allows it to bind to acetylated and methylated histone tails from different nucleosomes (Ruthenburg et al., 2011). Given that transcriptionally active OR alleles are marked by H3K4me3 (Magklara et al., 2011b), the modification recognized by the PHD-finger domain of BPTF, it is possible that BPTF binding to the OR enhancers facilitates their long-range interactions with active OR promoters. In this model, enhancer-promoter interactions initiated by BPTF may be stabilized by other factors, such as cohesin complexes, explaining the dispensable role of this protein in mature OSNs. Obviously, the global reduction of OR expression in the BPTF KO could be also caused by indirect effects or additional functions of BPTF in OR transcription.

Our zebrafish enhancer screen revealed specific GFP expression in OSNs, a result that was also observed in the mouse, where four different enhancer transgenics showed beta galactosidase expression only in the MOE. Identifying TFs that restrict the activity of OR-proximal enhancers to the MOE becomes critical, especially in light of recent computational predictions suggesting that OR and VR promoters share common TF binding motifs (Michaloski et al., 2011). An exciting possibility is that distant enhancers are the elements that restrict the expression of these chemoreceptor families to the proper sensory organ.

A candidate consensus sequence that may be restricting the activity of these enhancers to the MOE is a novel O/E-like motif that was identified by our in vivo footprinting assay.

As additional testament to the resolving power of *in vivo* DNase footprinting methods, our analysis revealed a 2bp difference from the canonical O/E site (M. M. Wang & Reed, 1993), which suffices for high and restricted expression in OSNs. Because previous studies have shown that Ebf4 is the only Ebf family member that has restricted expression in the MOE (S. Wang, 2002b) we hypothesized that this is the TF that binds to OR-proximal enhancers, a result supported by ChIP-qPCR experiments. Further analysis of the other highly protected motifs revealed by our in vivo footprinting experiments will provide more information on the regulatory restrictions imposed by these enhancers and perhaps some insight into the spatial expression pattern of OR genes in the MOE (Ressler, Sullivan, & Buck, 1993a). For example, class I ORs comprise 10% of the OR repertoire and are expressed in the most dorsal region in the MOE. Our ChIP-seq analysis uncovered a potential enhancer sequence in the class I OR cluster (Milos), which has also been identified by transgene mutagenesis experiments (Iwata et al., 2013), and may provide information on the spatial specification of OR expression.

**Methods**

Mice Strains: Mice were treated in compliance with the rules and regulations of IACUC. The Lipsi enhancer knockout mouse was generated by homologous recombination in ES cells (see Extended Experimental Methods). Other mouse strains used are: BPTF conditional (J. Landry et al., 2008), Foxg1-Cre (Hébert & McConnell, 2000), MOR28-IRES-GFP (Shykind et al., 2004b), and Ngn1-GFP BAC transgenic reporter mouse from GENSAT (Heintz, 2004). Strains were maintained on a mixed genetic background.

IF experiments were performed on pre-fixed MOE as previously described (Clowney et al., 2012) using anti-betagalactosidase (abcam, ab9361), anti-GFP (abcam, ab290), anti-vglut2 (Millipore, AB2251), and anti-adcy3 (santa cruz, sc-558). Olfr1507 antibody was designed by Gilad Barnea (Shykind et al., 2004a).

DHS-seq. Nuclei from the olfactory epithelium of five 6-8 week old wildtype mice were isolated as described (Magklara et al., 2011b) and resuspended in nuclease digestion buffer (0.32M sucrose, 50mM Tris-HCL pH 7.5, 4mM $MgCl_2$, 1mM $CaCl_2$, 0.1mM PMSF). 3 million nuclei were brought to 250mL nuclease digestion buffer, pre-warmed for one minute at 37C, and incubated with 20 U Dnase I (Ambion) for 2 minutes. Reactions were stopped with 30mL 0.5 M EDTA and placed on ice. Reactions were then incubated with 200 μg proteinase K (Ambion) at 55°C overnight, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, incubated with 10 μg RNase A (Roche) at 37°C for 30 minutes, extracted using phenol:chloroform:isoamyl (Invitrogen), and ethanol precipitated. Samples were resuspended in TE, and one half of each reaction was run on a 1% agarose/1x TAE gel. Regions 100-500 bp were excised and purified using Qiagen Gel Extraction Kit. Illumina sequencing libraries were then prepared using standard protocols, and amplified for 15 cycles with Phusion DNA Polymerase (NEB). DNA polymerase (NEB) and Illumina TruSeq primers.

DHS-seq sequence capture. DHS-seq library was further amplified for 10 PCR cycles and 1 mg was hybridized to Seqcap EZ choice probes (NImblegen, Roche) designed to

target and tile OR enhancer regions. Repetitive sequences were blocked using Mouse Hybloc (Applied Genetics Laboratories) and adapter sequences were blocked using custom oligos. A double capture protocol was followed as per the manufacturer's instructions and probes were pulled down using Streptavidin M270 beads. After elution, the sequence captured library was amplified for 18 cycles using Illumina TruSeq pimers targeting adapter sequences and Phusion DNA Polymerase (NEB).

DNase hypersenstivity (DHS) footprints were identified as in (Neph et al., 2012). DNase RPKM values were computed for each base within Seqcap-targeted regions. To compute the footprint occupancy score (FOS) for a given region, the following formula was used:

$$FOS = 0.5\left(\frac{C+1}{L} + \frac{C+1}{R}\right)$$

where *C* is the average RPKM within a central 6 bp region and *L* and *R* are the average RPKM within the 10 bp flanking regions to the left and right respectively. To exclude potential footprints in Seqcap areas with low DHS, regions in which average RPKM in the flanking windows was less than the average RPKM across the entire Seqcap region were filtered out. Footprints were reported as those regions with a FOS < 0.5.

Motif analysis. TF Footprints were extended by 2bp on each side and footprints from verified OR enhancers and H and P enhancers were inputted into MEME (Bailey et al., 2009). Motifs outputted by MEME were compared to known TF motifs in the JASPAR Vertebrate and UniProbe Mouse databases using TOMTOM (Gupta, Stamatoyannopoulos, Bailey, & Noble, 2007). Motifs were individually compared to

TRANSFAC database (Matys et al., 2006). *De novo* motif logos were generated by Weblogo (Crooks, Hon, Chandonia, & Brenner, 2004).

Lhx2 ChIP-qPCR: Whole main olfactory epithelium was dissected from 3 week old wildtype mice. Dissected tissue was thoroughly minced on ice using razor blades, then fixed for 5 minutes at room temperature with methanol-free formaldehyde (Pierce 28906) diluted to 1% in PBS. Fixation was quenched by adding 1/10$^{th}$ volume of 1.25M glycine. Fixed tissue was collected by centrifugation at 200g for 5 minutes at 4°C, then washed twice with cold PBS. Washed tissue was flash-frozen with liquid nitrogen then cryofractured using a Covaris CryoPrep Impactor on power setting 6. Nuclei were extracted by incubating fixed tissue in ChIP Lysis Buffer (50mM Tris-HCl pH 7.5, 150nM NaCl, 0.5% NP-40, 0.25% Sodium Deoxychoalate, 0.1% SDS) on a rotisserie rotator for 20 minutes at 4°C. Nuclei were collected by centrifugation (1750x g, 5 minutes, 4°C), resuspended in shearing buffer (0.5% Sodium Deoxychoalate, 1% NP40, 0.5% SDS in 1x PBS), then sheared on a Covaris S2 sonicator (11 minutes, 2% Duty Cycle, Intensity 3, 200 cycles per burst, frequency sweeping). Sheared chromatin was centrifuged (10,000g for 10 minutes at 4°C) to remove insoluble material. 5uL of sheared chromatin was reverse-crosslinked and DNA was column purified (Zymo D4013) to monitor shearing and quantify DNA yield. Sheared chromatin was diluted approximately 5 to 10-fold with ChIP Dilution Buffer (CDB: 16.7mM Tris-HCl pH8.1, 167mM NaCl, 1.2mM EDTA, 1.1% Triton X-100, 0.01% SDS), then pre-cleared with protein G dynabeads for two hours at 4°C. A 10% input control was set aside and the remaining cleared chromatin was used immediately for ChIP reactions.

Each ChIP was set up with 1ug of anti-Lhx2 and approximately 8ug of cleared chromatin and then incubated overnight at 4°C. Pre-blocked Protein G dynabeads (overnight with 2mg/mL yeast tRNA in CDB) were added to antibody bound chromatin and rotated for 2-3 hours at 4°C. Bound beads were washed 5 times with LiCl Wash Buffer (100mM Tris-HCl pH 7.5, 500mM LiCl, 1% NP-40, 1% Sodium Deoxycholate) and once with TE (pH7.5). DNA was eluted from beads by incubating at 65oC for 30 minutes with 100uL ChIP Elution Buffer (1% SDS, 0.1M Sodium Bicarbonate). This elution was repeated and the combined elution fractions were incubated overnight at 65oC. ChIP DNA was purified by Phenol-Chloroform extraction and ethanol precipitation. QPCR was run with input controls using the following primers targeting OR enhancer sequences and control sequences at distance of thousands of base pairs

ChIP-qPCR primers

| | Forward primer | Reverse primer |
|---|---|---|
| H +3kB | ggctctttgaccctcaacaac | gggaaacacagaggagtgga |
| H | agggtgcctctagtggttca | gggtccctgaggaattcagt |
| Sfaktiria +3kB | ttgcctgctgtttgctttta | cctgccaaatcttcttcgag |
| Sfaktiria | actatctccttgcggggttc | actcctggtcccctgagaat |
| Rhodes -3kB | gcaacgggaagagtgtctgt | atttttcccttcccatcagc |
| Rhodes | cccttggcccagtcatctat | cctctgctaatttgctggagac |
| Lipsi -2kB | attggggaatgtgacttt | ccactctcccctgtctcagta |
| Lipsi | tctcagaggctttccctcttc | Ctgaagcaggctactcac |

DNA FISH probes: BAC probes were prepared by standard Nick Translation. PanOR probe was generated as described (Clowney et al., 2012) from genomic DNA eluted from a custom array tiling ~40Mb of class II OR clusters. Amplified library was labeled with

High Prime (Roche) for Dig or Biotin and nicked to 200-500bp by Nick Translation (Roche). Probes were purified by G50 column, precipitated in the presence of 20x CotI, and resuspended in CamBio hybridization buffer. The following BACs were used to generate enhancer probes:

| | BAC clone |
|---|---|
| olfr1507 - H | RP24-290L16 |
| Lipsi | RP24-353M9 |
| Sfaktiria | RP23-65G23 |
| Crete | RP24-156B9 |
| Ikaria | RP24-212K15 |
| Symi | RP24-323I2 |
| Evia | RP23-97N5 |
| Ebf3 enhancer | RP23-200O13 |
| Ios | RP23-336P21 |
| Nimos | RP24-319L11 |

DNA FISH: Whole heads of P4 mice were directly embedded in OCT (Sakura) and frozen. 5 um cryosections were cut, air dried for 30 minutes, and fixed in cold 4% PFA for 5 minutes. Sections were permeabilized with PBS-0.1% Triton (PT), DNA was fragmented with 0.1M HCl for 5 minutes at RT, digested with 3mg/ml RNAse A in PT for one hour at 37C, and dehydrated at RT in ethanol series (70, 95, 100%), and slides were baked at 45C for ten minutes. Sections were denatured at 85C for minutes in 2xSSC, 75% formamide (Invitrogen), and immediately dehydrated in an ice-cold ethanol series (70, 95. 100%) and baked again at 45C for ten minutes. Probes were applied at 5 ng/uL concentration (BACs) or 25 ng/uL (complex panOR probe) under 8mm circular coverslips, sealed with rubber cement and incubated overnight at 37C. Slides were washed in 2xSSC, 55% formamide, 0.1% NP-40, three times for 15 minutes at 42C, rinsed with PT, blocked in TNB (Promega TSA kit), incubated 2 hours at room

temperature with anti-dig or anti-biotin conjugated to DyLight fluors (Jackson Immunoresearch). Then the slides were washed in PT+8% formamide and mounted.

Image analysis: Confocal images were collected on a Zeiss LSM700. All images are confocal slices. For FISH counts, stacks were collected with images 1uM apart and cells with OSN-typical centromeric focus were analyzed. Nuclei were counted if FISH signal was detectable for all BAC probes. BAC signals were colocalized if there was pixel overlap. BAC signals were colocalized with panOR probe if they were contained within discrete panOR foci in the Z-slice in which they were brightest.

In Situ Hybridization (ISH): To make the PanOR RNA in situ probe we isolated RNA from the MOE using Trizol (Invitrogen) and generated a cDNA library using Superscript III (Invitrogen). PCR was performed using degenerate primers against conserved OR transmembrane domains TMIII and TMVII (Buck & Axel, 1991; Malnic, Hirono, Sato, & Buck, 1999) and PCR products were Topo cloned into pCRII vector (Invitrogen) to generate antisense riboprobes. ISH was performed as previously described (Lyons et al., 2013). ISH experiments were quantified by counting cells over consecutive sections of the MOE.

IF experiments were performed on pre-fixed MOE as previously described (Clowney et al., 2012) using olfr1507 antibody designed by Gilad Barnea (Shykind et al., 2004a).

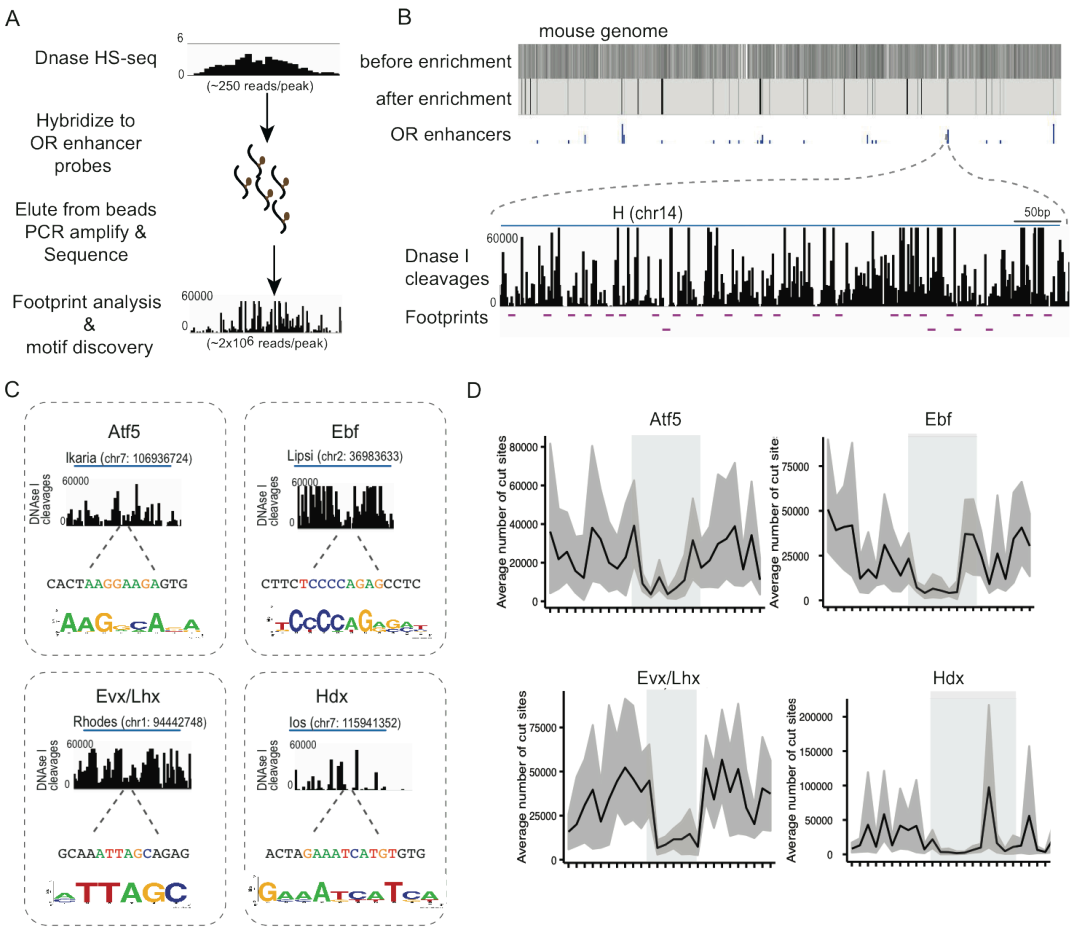

A
Dnase HS-seq
(~250 reads/peak)
Hybridize to OR enhancer probes
Elute from beads PCR amplify & Sequence
Footprint analysis & motif discovery
(~2x10^6 reads/peak)
B
mouse genome
before enrichment
after enrichment
OR enhancers
H (chr14)
50bp
Dnase I cleavages
Footprints
C
Atf5
Ikaria (chr7: 106936724)
CACTAAGGAAGAGTG
Ebf
Lipsi (chr2: 36983633)
CTTCTCCCCAGAGCCTC
Evx/Lhx
Rhodes (chr1: 94442748)
GCAAATTAGCAGAG
Hdx
Ios (chr7: 115941352)
ACTAGAAATCATGTGTG
DNAse I cleavages
D
Atf5
Ebf
Evx/Lhx
Hdx
Average number of cut sites

Figure 12: DNAse-seq uncovers protected TF footprints on OR enhancers

**(A)** Schematic of sequence capture-based method of DNase I sequencing library construction. (**B**) Heatmap of read density for DNase I sequencing reads across the mouse genome, before and after sequence capture. Blue bars indicate the chromosomal location of OR enhancer candidates. Below, per nucleotide DNAse I cleavages are mapped over the H enhancer locus, and pink bars indicate DNAse I footprints (footprint occupancy score (FOS) <0.5). **(C**) Four *de novo* consensus motifs that match TF motifs annotated within the Transfac, JASPAR, or UniPROBE databases. Example footprints are shown over instances of the motifs in different verified OR enhancers. (**D**) Average per nucleotide DNAse I cleavages at all instances of above consensus motifs in OR enhancers. X axis is position relative to consensus motif in center (shaded box). Error is bootstrapped 95% confidence intervals.

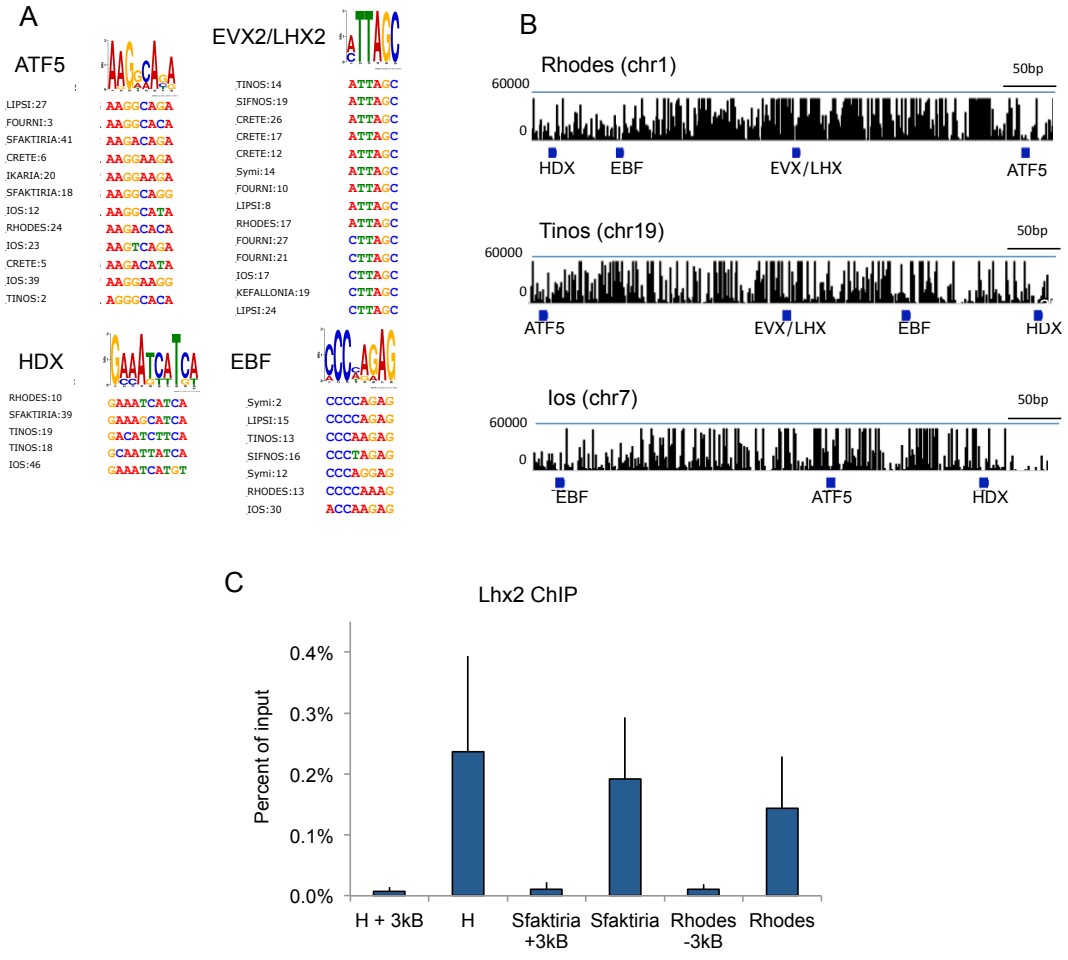

A
ATF5
LIPSI:27 AAGGCAGA
FOURNI:3 AAGGCACA
SFAKTIRIA:41 AAGACAGA
CRETE:6 AAGGAAGA
IKARIA:20 AAGGAAGA
SFAKTIRIA:18 AAGGCAGG
IOS:12 AAGGCATA
RHODES:24 AAGACACA
IOS:23 AAGTCAGA
CRETE:5 AAGACATA
IOS:39 AAGGAAGG
TINOS:2 AGGGCACA
EVX2/LHX2
TINOS:14 ATTAGC
SIFNOS:19 ATTAGC
CRETE:26 ATTAGC
CRETE:17 ATTAGC
CRETE:12 ATTAGC
Symi:14 ATTAGC
FOURNI:10 ATTAGC
LIPSI:8 ATTAGC
RHODES:17 ATTAGC
FOURNI:27 CTTAGC
FOURNI:21 CTTAGC
IOS:17 CTTAGC
KEFALLONIA:19 CTTAGC
LIPSI:24 CTTAGC
HDX
RHODES:10 GAAATCATCA
SFAKTIRIA:39 GAAAGCATCA
TINOS:19 GACATCTTCA
TINOS:18 GCAATTATCA
IOS:46 GAAATCATGT
EBF
Symi:2 CCCCAGAG
LIPSI:15 CCCCAGAG
TINOS:13 CCCAAGAG
SIFNOS:16 CCCTAGAG
Symi:12 CCCAGGAG
RHODES:13 CCCCAAAG
IOS:30 ACCAAGAG
B
Rhodes (chr1)
50bp
60000
0
HDX
EBF
EVX/LHX
ATF5
Tinos (chr19)
50bp
60000
0
ATF5
EVX/LHX
EBF
HDX
Ios (chr7)
50bp
60000
0
EBF
ATF5
HDX
C
Lhx2 ChIP
Percent of input
0.4%
0.3%
0.2%
0.1%
0.0%
H + 3kB
H
Sfaktiria +3kB
Sfaktiria
Rhodes -3kB
Rhodes

Figure 13: TF footprints on zebrafish-positive OR enhancers

**(A)** Consensus sequence for ATF5, EVX/LHX, HDX and O/E like motifs and alignment of footprinted sequences in OR enhancers. (**B)** DNase I cleavages mapped over verified OR enhancers Rhodes, Tinos, Ios and Lipsi. Blue bars indicate footprints containing ATF5, HDX, EVX/LHX, and O/E-like motifs. **(C)** Lhx2 ChIP on whole MOE, qPCR normalized to input. Primers target OR enhancers, and control primers amplify sequences 2-3 kB distance away from enhancers. Error bars are SEM from triplicate experiments.

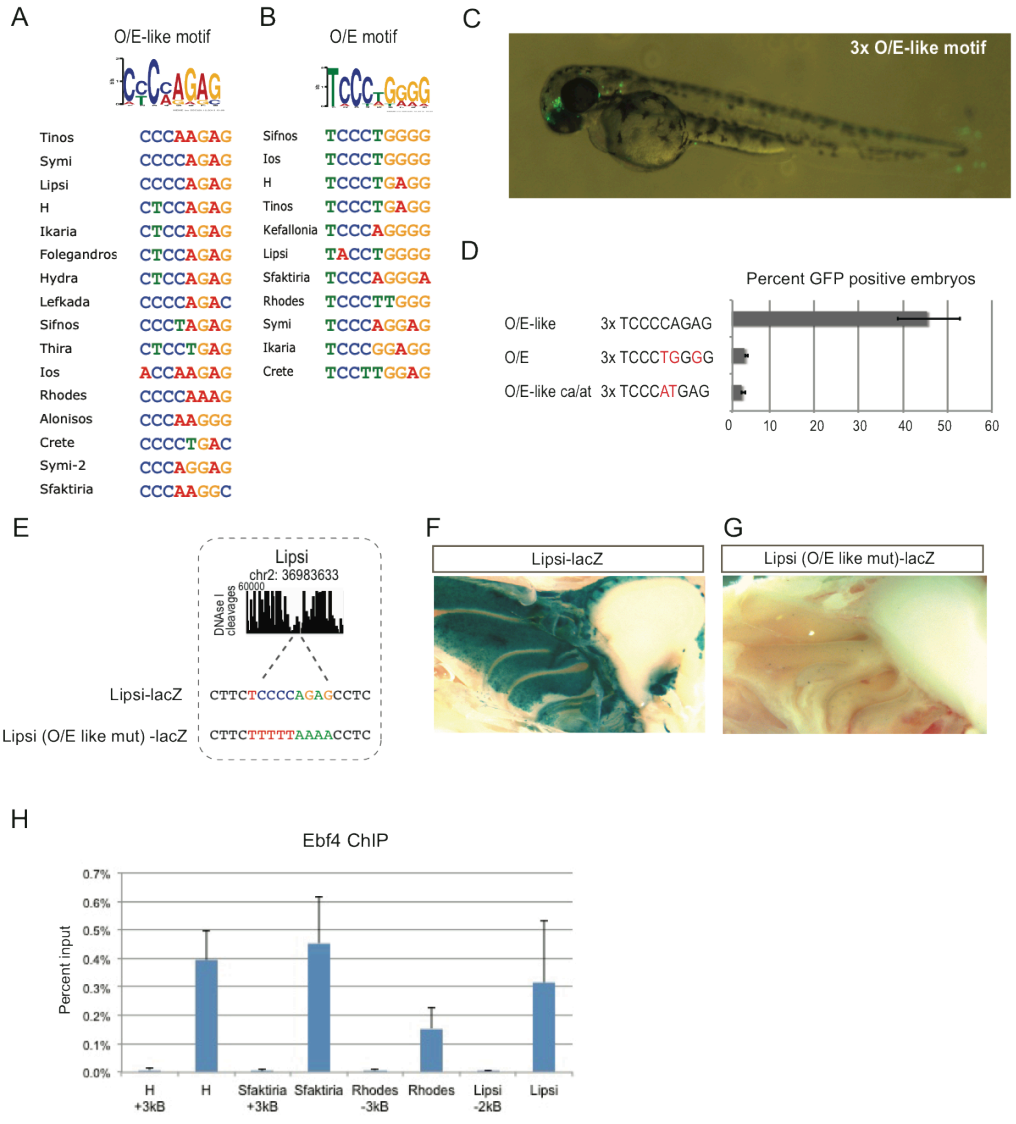

A
O/E-like motif
Tinos CCCAAGAG
Symi CCCCAGAG
Lipsi CCCCAGAG
H CTCCAGAG
Ikaria CTCCAGAG
Folegandros CTCCAGAG
Hydra CTCCAGAG
Lefkada CCCCAGAC
Sifnos CCCTAGAG
Thira CTCCTGAG
Ios ACCAAGAG
Rhodes CCCCAAAG
Alonisos CCCAAGGG
Crete CCCCTGAC
Symi-2 CCCAGGAG
Sfaktiria CCCAAGGC
B
O/E motif
Sifnos TCCCTGGGG
Ios TCCCTGGGG
H TCCCTGAGG
Tinos TCCCTGAGG
Kefallonia TCCCAGGGG
Lipsi TACCTGGGG
Sfaktiria TCCCAGGGA
Rhodes TCCCTTGGG
Symi TCCCAGGAG
Ikaria TCCCGGAGG
Crete TCCTTGGAG
C
3x O/E-like motif
D
Percent GFP positive embryos
O/E-like 3x TCCCCAGAG
O/E 3x TCCCTGGGG
O/E-like ca/at 3x TCCCATGAG
0
10
20
30
40
50
60
E
Lipsi
chr2: 36983633
60000
DNAse I cleavages
Lipsi-lacZ CTTCTCCCCAGAGCCTC
Lipsi (O/E like mut) -lacZ CTTCTTTTTAAAACCTC
F
Lipsi-lacZ
G
Lipsi (O/E like mut)-lacZ
H
Ebf4 ChIP
Percent input
0.7%
0.6%
0.5%
0.4%
0.3%
0.2%
0.1%
0.0%
H +3kB
H
Sfaktiria +3kB
Sfaktiria
Rhodes -3kB
Rhodes
Lipsi -2kB
Lipsi

Figure 14: An O/E-like motif is necessary and sufficient for enhancer activity

**(A)** Alignment of the *de novo* derived O/E-like motif sequences in OR enhancer candidates. **(B)** Alignment of the O/E motif sequences in OR enhancer candidates **(C)** Representative image of zebrafish embryo at 48hpf from oocyte injected with GFP reporter construct containing three tandem copies of O/E-like motif upstream of a minimal e1b promoter. **(D)** Results of zebrafish motif screen. Percent of injected oocytes with GFP expression in olfactory epithelium, error bars are standard error of the mean (SEM) over multiple injections. **(E)** Footprinted O/E-like motif in Lipsi. Motif coordinate (mm9) is indicated. Dnase cleavages are mapped in y-axis. Below, wildtype and mutated motifs in the Lipsi-lacZ and Lipsi (O/E-like)-lacZ transgene reporter constructs. **(F,G)** Whole mount x-gal staining of MOE in Lipsi-lacZ mouse and Lipsi (O/E-like mut) mouse. **(H)** Ebf4 ChIP-qPCR from whole MOE. Values shown are means from triplicate ChIP experiments, error bars represent the SEM.

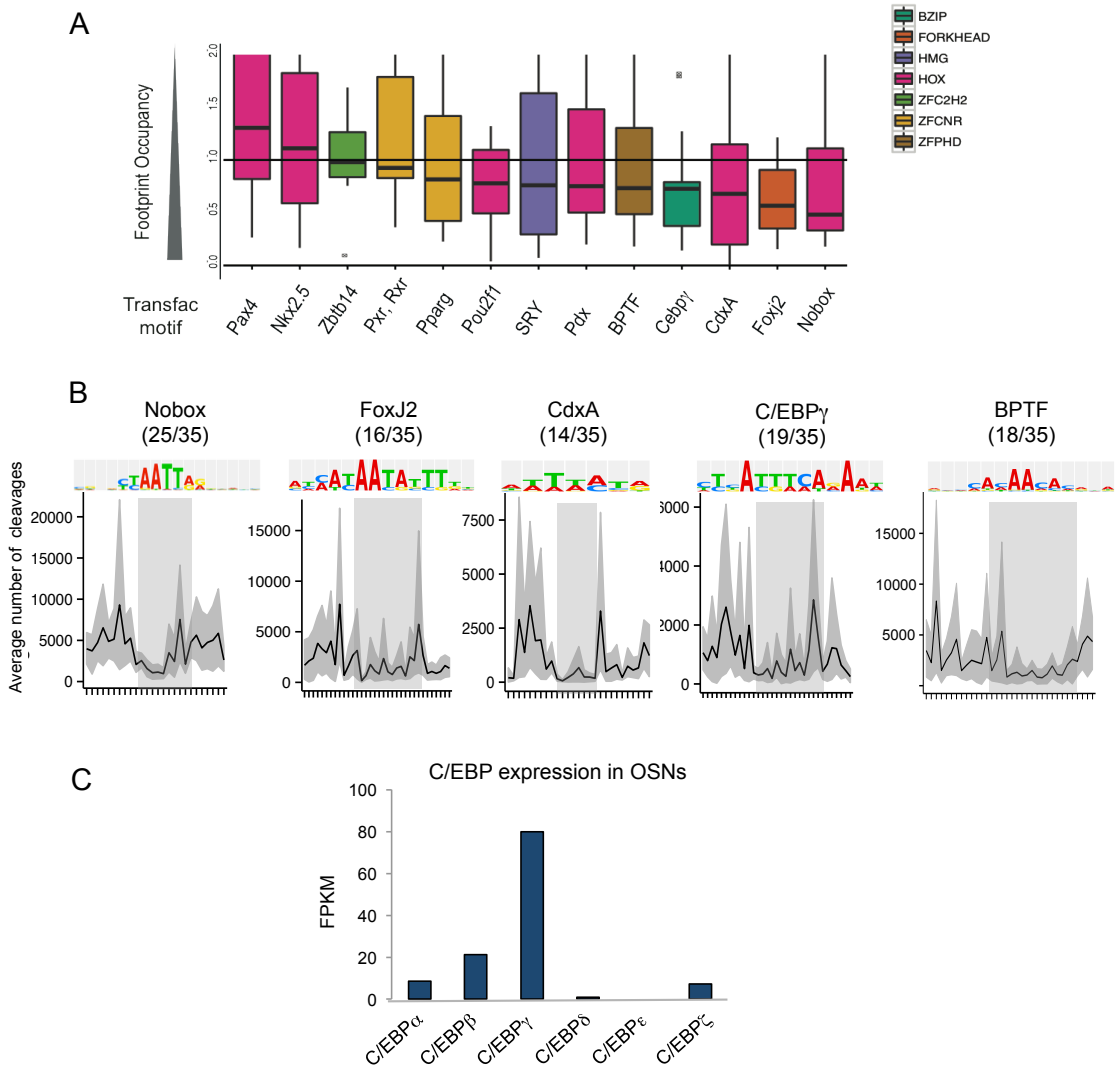

A
Footprint Occupancy
Transfac motif
Pax4
Nkx2.5
Zbtb14
Pxr, Rxr
Pparg
Pou2f1
SRY
Pdx
BPTF
Cebpγ
CdxA
Foxj2
Nobox
BZIP
FORKHEAD
HMG
HOX
ZFC2H2
ZFCNR
ZFPHD
B
Nobox (25/35)
FoxJ2 (16/35)
CdxA (14/35)
C/EBPγ (19/35)
BPTF (18/35)
Average number of cleavages
C
C/EBP expression in OSNs
FPKM
C/EBPα
C/EBPβ
C/EBPγ
C/EBPδ
C/EBPε
C/EBPζ

Figure 15: OR enhancers share multiple TF footprints

**(A)** Transcription factor (TF) motifs from the TRANSFAC database that are enriched on OR enhancer candidates and average footprint score for each TF motif. Low footprint occupancy scores (FOS) indicate greater footprint occupancy. TF motifs are color coded according to TF family. **(B)** Average DNase I cleavages for TRANSFAC motifs on 35 candidate OR enhancers. The X-axis is centered at the consensus sequence (shaded box). Fraction of OR enhancers containing each TF motif is indicated in parentheses. Error is bootstrapped 95% confidence intervals. **(C)** RNA-seq data measuring transcript levels of C/EBP family transcription factors in OSNs. RNA is from FAC-sorted cells in OMP-GFP mice. Y-axis is FPKM.

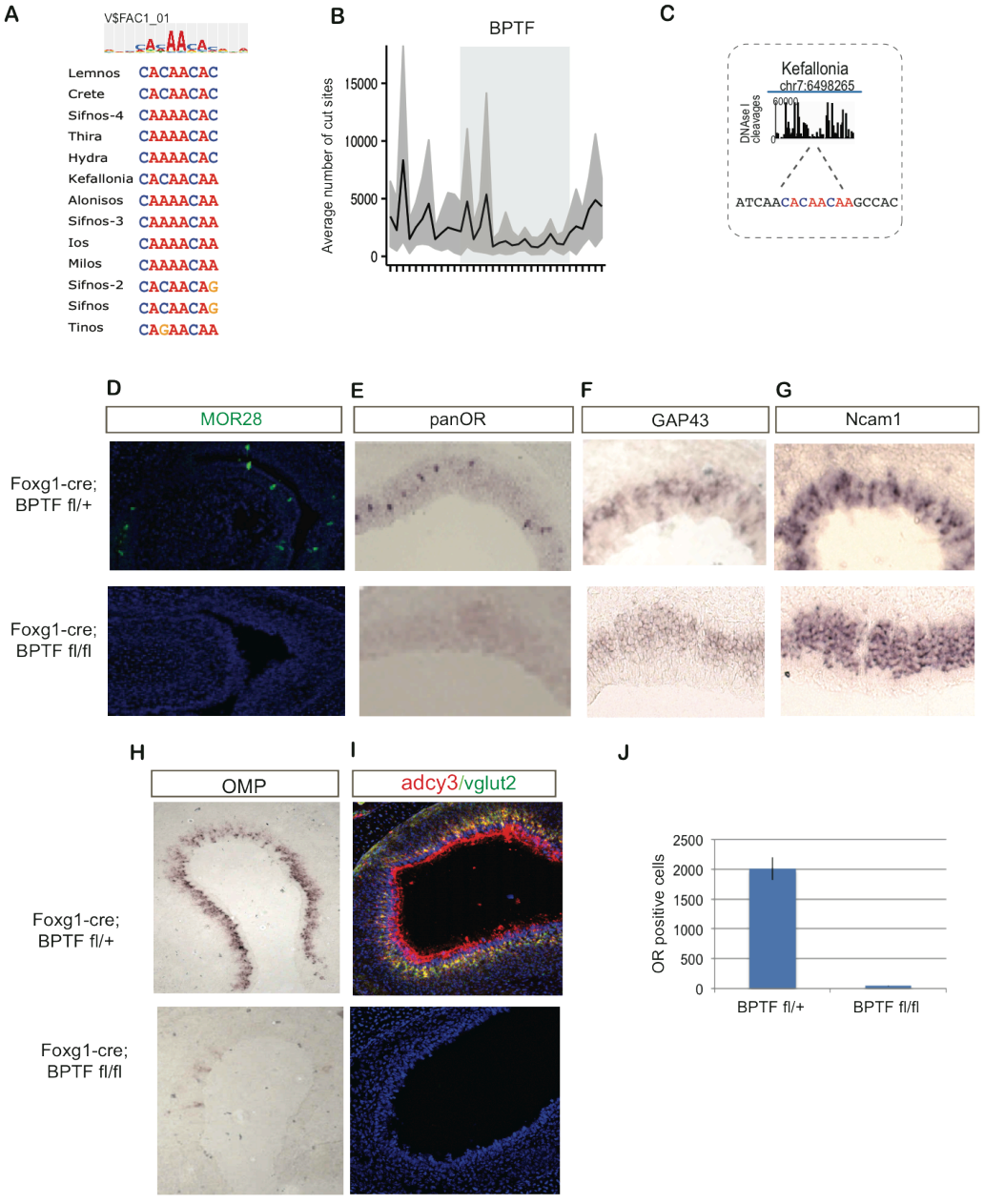
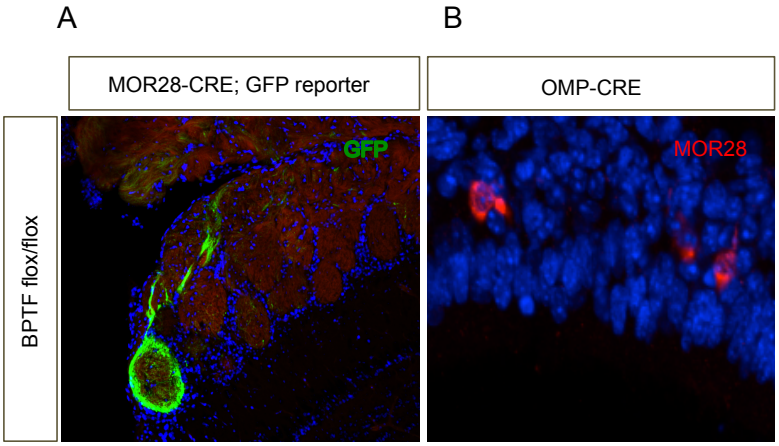

A
V$FAC1_01
Lemnos CACAACAC
Crete CACAACAC
Sifnos-4 CAAAACAC
Thira CAAAACAC
Hydra CAAAACAC
Kefallonia CACAACAA
Alonisos CAAAACAA
Sifnos-3 CAAAACAA
Ios CAAAACAA
Milos CAAAACAA
Sifnos-2 CACAACAG
Sifnos CACAACAG
Tinos CAGAACAA
B
BPTF
Average number of cut sites
15000
10000
5000
0
C
Kefallonia
chr7:6498265
60000
0
DNAse I cleavages
ATCAACACAACAAGCCAC
D
MOR28
E
panOR
F
GAP43
G
Ncam1
Foxg1-cre; BPTF fl/+
Foxg1-cre; BPTF fl/fl
H
OMP
I
adcy3/vglut2
Foxg1-cre; BPTF fl/+
Foxg1-cre; BPTF fl/fl
J
OR positive cells
2500
2000
1500
1000
500
0
BPTF fl/+
BPTF fl/fl

Fig 16: BPTF is required for OR gene expression.

**(A)** Alignment of BPTF motifs in OR enhancer candidates. **(B)** Average per nucleotide DNase I cleavages on BPTF motifs across OR enhancers. X-axis is position relative to BPTF consensus sequence in center (shaded box). Error is bootstrapped 95% confidence intervals. **(C)** Example footprint over the BPTF motif in Kefallonia. Motif coordinate is mm9. Y-axis is DNAse cleavages. **(D)** MOR28 IF in Foxg1-Cre; BPTF flox/+ and Foxg1-Cre; BPTF flox/flox MOE at e18.5. DAPI is nuclear stain. **(E)** Degenerate OR ISH in BPTF heterozygote and KO. **(F-H).** ISH for developmental markers in BPTF heterozygote (top) and KO (bottom). **(I)** Adcy3 IF (red) and vglut2 IF (green) in BPTF heterozygote (top) and KO (bottom). DAPI (blue) is nuclear stain. **(J)**Quantification of OR ISH experiment (panel E) in BPTF heterozygote and KO. Error bars represent the variance over duplicate experiments.


A
B
MOR28-CRE; GFP reporter
OMP-CRE
BPTF flox/flox
GFP
MOR28

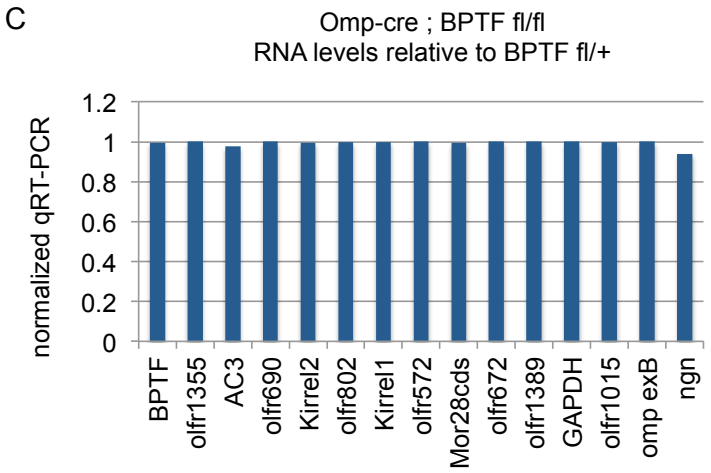

C
Omp-cre ; BPTF fl/fl
RNA levels relative to BPTF fl/+
normalized qRT-PCR
1.2
1
0.8
0.6
0.4
0.2
0
BPTF
olfr1355
AC3
olfr690
Kirrel2
olfr802
Kirrel1
olfr572
Mor28cds
olfr672
olfr1389
GAPDH
olfr1015
omp exB
ngn

Figure 17: Conditional deletion of BPTF after OR choice does not affect OR expression **(A).** MOR28-Cre; flox stop flox GFP reporter ; BPTF flox/flox mouse. IF of ofactory bulb shows targeting of GFP labeled axons to a single glomerulus. **(B)** OMP-Cre; BPTF flox/flox MOE. IF for olfr1507 (MOR28). (**C**) qRT-PCR for ORs and several mature neuronal markers in OMP-Cre; BPTF flox/flox MOE normalized to OMP-Cre; BPTF flox/+ RNA. Y axis is normalized levels of cDNA.

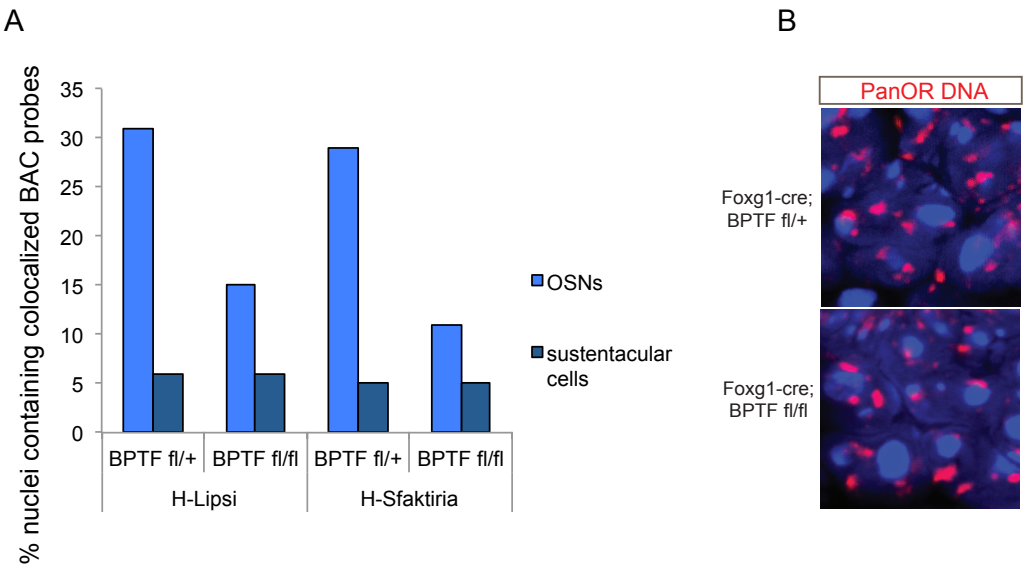

A
B
% nuclei containing colocalized BAC probes
35
30
25
20
15
10
5
0
BPTF fl/+
BPTF fl/fl
BPTF fl/+
BPTF fl/fl
H-Lipsi
H-Sfaktiria
OSNs
sustentacular cells
PanOR DNA
Foxg1-cre; BPTF fl/+
Foxg1-cre; BPTF fl/fl

Figure 18: BPTF facilitates interchromosomal enhancer-enhancer interactions.

**(A)** Quantification of DNA FISH co-localizations between H and Sfaktiria BACs and between H and Lipsi BAC probes in BPTF heterozygote and KO OSNs and sustentacular cells. The Y-axis is percent OSN nuclei containing co-localized probes. **(B)** DNA FISH with complex pan olfactory (PanOR) receptor probe (red) in BPTF heterozygote and KO OSNs. DAPI is nuclear stain (blue)

# Chapter 5: General Discussion

**Novel mechanism of *cis*-regulation**

In recent years there has been an explosion in our understanding of enhancers and gene regulation. Thousands of cell-type specific enhancers have been identified, and new types of enhancers are being discovered every year (Visel et al., 2009). Shadow enhancers are partially redundant enhancers that have been described for patterning genes in *Drosophila* (Mike Levine, 2010). Shadow enhancers may function to proffer robustness to patterning gene expression, as well as provide a second layer of patterning control. More recently super-enhancers were described, large transcriptionally active sequences which nucleate master transcription factors and drive gene expression programs in different cell types (Hnisz et al., 2013; Whyte et al., 2013).

In this thesis I describe a potentially new type of enhancer, the "weak" olfactory receptor enhancer that exhibits a unique epigenetic signature (Chapter 2), but more importantly forms frequent inter-chromosomal interactions in the nucleus with other OR enhancers to drive singular OR expression (Chapter 3). The "weak" OR enhancers sequences are bound by many common transcription factors which may mediate their long-range interactions in the nucleus (Chapter 4). Taken together, these observations support a model where a "weak" OR enhancer depends on the synergistic action of other weak enhancers on different gene clusters to drive olfactory receptor expression.

A similar type of synergistic action between weak enhancers has been described recently for the HoxD enhancer cluster (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011) where multiple enhancers interact in *cis* to drive robust expression of HoxD genes – each enhancer making an additive contribution to transcription and the developing digits. The situation is similar to the olfactory complex, except that the interacting enhancers are located on

different chromosomes. In both cases multiple "weak" enhancers converge within the nucleus to generate a single complex that may be dubbed a "super enhancer" of sorts.

In the case of HoxD expression in digits, the location of the participating enhancers in a *cis*-proximal gene desert is important because of the temporal control that can be exerted on these enhancers via the local chromatin. The enhancers only interact only upon removal of repressive histone methylation on the loci containing the enhancer sequences, a switch that occurs after expression of genes on the opposite end of the Hox cluster (Montavon, Soshnikova, Mascrez, Joye, Thevenet, Splinter, de Laat, et al., 2011; Noordermeer et al., 2014).

Temporal control is less important in the case of the olfactory receptor enhancer complex because there is a feedback system in place in which transcription of an OR leads to the terminal differentiation of the neuron and stabilization of the choice (Dalton et al., 2013; Serizawa et al., 2000). The OR enhancers may thus be exempt from such tight temporal control via heterochromatin silencing, which may not even be possible given their scattered genomic coordinates.

As discussed in Chapter 4, the low probability event of multiple enhancers interacting in *trans* may explain the singularity of expression of an OR within a neuron. The probability of an X number of enhancers interacting within the nucleus is low enough to be a single event within the nucleus. Moreover, the probability of a *particular* combination of enhancers interacting within the nucleus will be even lower, which could explain the sparse expression of particular ORs within the tissue. A question that remains unanswered is how a single choice is made between the OR genes that are located in *cis* to the interacting enhancers. It is possible that a second layer of regulation is in is in place to ensure that only one OR gene promoter interacts with the

enhancer complex (a looping that occurs at random, or perhaps the de-repression of the promoter of one OR promoter by Lsd1 (Lyons et al., 2013)).

Assembly of a multi-chromosomal multi-enhancer complex may emerge as a common motif in gene regulation. The modularity and redundancy built into the system makes it compatible with the evolution of gene families like the olfactory receptors (Clowney et al., 2011a). Furthermore, the random integration of multiple enhancers into an enhancer complex may confer stochastic gene expression patterns. Stochastic gene expression generates diversity of cell types, as has been mentioned in the stochastic choice of protocadherin promoters in neurons. It remains to be seen whether gene regulatory paradigms in developing nervous system rely on similar mechanisms to generate stochastic gene expression and neuronal diversity.

**Implications for olfactory receptor gene regulation**

The identification and characterization of the regulatory architecture of olfactory receptor enhancers has shed new light on the regulation of ORs. The olfactory receptor enhancer described in this work are active in a transgenic setting in a large number of olfactory sensory neurons in the zebrafish and mouse olfactory epithelium, as has been previously observed for the H enhancer (Nishizumi et al., 2007; Serizawa et al., 2003c). An OR enhancer, Lipsi on chromosome 2, is required for the choice of the proximal olfactory receptors in the gene cluster. A similar effect on proximal olfactory receptors has been observed for the H and P knockout mice (Khan et al., 2011; Nishizumi et al., 2007). As of yet, it is unclear what determines which *cis*-proximal ORs are regulated by an enhancer. OR genes within a cluster usually do not span topological association domains (data not shown), making it unlikely that an enhancer is unable to loop and activate other OR genes within the cluster. A comparative analysis of promoters of

ORs that are affected by enhancer deletion with unaffected ORs may show differential enrichment of TF motifs on the promoter sequences.

The model we propose may explain the stochastic choice of ORs, but it does not explain the stereotyped zonal expression of an olfactory receptor within the olfactory epithelium (Qasba & Reed, 1998; Ressler, Sullivan, & Buck, 1993b). The zonal of OR expression argues that there are some aspects of OR choice that are deterministic. One hypothesis that is compatible with the enhancer-complex model we propose is that within a zone of the tissue only a subset of enhancers are capable of forming long-range interactions. In other words, perhaps a subset of enhancers are de-commissioned within a particular topological area of the epithelium.

What might be the factors that control zonal enhancer activity? It is known that the zones of OR expression are not discrete, but rather are overlapping within the epithelium (Miyamichi, Serizawa, Kimura, & Sakano, 2005). This immediately suggests that the zonality could be the consequence of transcription factor gradients, and that enhancers integrate information from the transcription factor milieu of the zone of expression, as has been described in the proto-map enhancers that specify cortical patterning (Pattabiraman et al., 2014). Interestingly, the non-zonal expression of OR enhancer transgenes (Figure 4) indicates that the chromatin context of the *endogenous* enhancer may play an important role in specifying the zones of enhancer activity. In this case, it may be interesting to look for the zonal expression of chromatin remodeling factors that may insulate or protect an enhancer from the repressive heterochromatin milieu.

This analysis of the transcription factor landscape of OR enhancers uncovered some novel candidate regulators of ORs: the Hdx, Nobox, C/EBPgamma, and FoxJ2 motifs on OR enhancers are protected from DNAse digestion and are interesting candidate transcription factors that may

regulate OR expression. The binding of ATF5 on the OR enhancers is also puzzling, and further work will be required to understand the role of ATF5 in stabilizing OR choice via its binding at the OR enhancer complexes.

By DNAse footprinting (Figure 15) we uncovered a conserved, protected 9 base pair sequence which we dubbed the "O/E-like" motif. This sequence was necessary and sufficient for enhancer activity in mouse and zebrafish olfactory neurons, respectively. Chromatin immunoprecipitation experiments demonstrated that the transcription factor that binds these enhancers is Ebf4. Ebf4 has been shown by yeast one-hybrid to bind olfactory receptor promoters (Hoppe, Breer, & Strotmann, 2006) Ebf1 has been shown to bind a sequence similar to the O/E-like motif (M. M. Wang & Reed, 1993), but knocking out Ebf1 does not affect OR gene expression (S S Wang et al., 1997). A triple knockout of Ebf1, Ebf2, and Ebf3 also does not affect OR expression, which stresses the redundancy of these factors in the OE. The function of Ebf4 binding on OR enhancers in OR choice remains to be tested.

The analysis we performed here paints a rich picture of the transcription factor landscape of the OR enhancers. Each enhancer contains more than twenty identifiable transcription factor footprints – are the enhancers simultaneously occupied by twenty different transcription factors? Probably not. Future work could use the DNAse footprinting method I developed here to dissect the transcription factor footprints as they occur during different stages of the development of the OE or in different OSN cell populations. Such experiments would shed light onto the function of these transcription factors, and of the enhancers they bind, for initiating or stabilizing olfactory receptor choice.

# Chapter 6: References

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., & Thanos, D. (2000). Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell*, *103*(4), 667–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11106736

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202–8. doi:10.1093/nar/gkp335

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., … Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–37. doi:10.1016/j.cell.2007.05.009

Birnbaum, R. Y., Clowney, E. J., Agamy, O., Kim, M. J., Zhao, J., Yamanaka, T., … Ahituv, N. (2012). Coding exons function as tissue-specific enhancers of nearby genes. *Genome Research*, *22*(6), 1059–68. doi:10.1101/gr.133546.111

Bozza, T., Vassalli, A., Fuss, S., Zhang, J.-J., Weiland, B., Pacifico, R., … Mombaerts, P. (2009). Mapping of class I and class II odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. *Neuron*, *61*(2), 220–33. doi:10.1016/j.neuron.2008.11.010

Buck, L., & Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=1840504

Chan, Y. F., Marks, M. E., Jones, F. C., Jr, G. V., Shapiro, M. D., Brady, S. D., … Kingsley, D. M. (2010). Adaptive Evolution of Pelvic Reduction of a Pitx1 Enhancer, *327*(January), 302–305.

Chess, a, Simon, I., Cedar, H., & Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell*, *78*(5), 823–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8087849

Chess, A. (2005). Monoallelic expression of protocadherin genes. *Nature Genetics*, *37*(2), 120–1. doi:10.1038/ng0205-120

Clowney, E. J., LeGros, M. a, Mosley, C. P., Clowney, F. G., Markenskoff-Papadimitriou, E. C., Myllys, M., … Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their monogenic expression. *Cell*, *151*(4), 724–37. doi:10.1016/j.cell.2012.09.043

Clowney, E. J., Magklara, A., Colquitt, B. M., Pathak, N., Lane, R. P., & Lomvardas, S. (2011a). High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Research*, *21*(8), 1249–59. doi:10.1101/gr.120162.110

Clowney, E. J., Magklara, A., Colquitt, B. M., Pathak, N., Lane, R. P., & Lomvardas, S. (2011b). High-throughput mapping of the promoters of the mouse olfactory receptor genes reveals a new type of mammalian promoter and provides insight into olfactory receptor gene regulation. *Genome Research*, *21*(8), 1249–59. doi:10.1101/gr.120162.110

Crooks, G. E., Hon, G., Chandonia, J., & Brenner, S. E. (2004). WebLogo : A Sequence Logo Generator, 1188–1190. doi:10.1101/gr.849004.1

Dalton, R. P., Lyons, D. B., & Lomvardas, S. (2013). Co-opting the unfolded protein response to elicit olfactory receptor feedback. *Cell*, *155*(2), 321–32. doi:10.1016/j.cell.2013.09.033

Deato, M. D. E., Marr, M. T., Sottero, T., Inouye, C., Hu, P., & Tjian, R. (2008). MyoD targets TAF3/TRF3 to activate myogenin transcription. *Molecular Cell*, *32*(1), 96–105. doi:10.1016/j.molcel.2008.09.009

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–80. doi:10.1038/nature11082

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., … Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons, *428*(March), 0–5.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., … Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43–9. doi:10.1038/nature09906

Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A., & Mombaerts, P. (2004). Axon guidance of mouse olfactory sensory neurons by odorant receptors and the beta2 adrenergic receptor. *Cell*, *117*(6), 833–46. doi:10.1016/j.cell.2004.05.013

Ferreira, T., Wilson, S. R., Choi, Y. G., Risso, D., Dudoit, S., Speed, T. P., & Ngai, J. (2014). Silencing of Odorant Receptor Genes by G Protein βγ Signaling Ensures the Expression of One Odorant Receptor per Olfactory Sensory Neuron. *Neuron*, *81*(4), 847–859. doi:10.1016/j.neuron.2014.01.001

Fuss, S. H., Omura, M., & Mombaerts, P. (2007a). Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*. doi:10.1016/j.cell.2007.06.023

Fuss, S. H., Omura, M., & Mombaerts, P. (2007b). Matters Arising Local and cis Effects of the H Element on Expression of Odorant Receptor Genes in Mouse, 373–384. doi:10.1016/j.cell.2007.06.023

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., & Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science (New York, N.Y.)*, *318*(5853), 1136–40. doi:10.1126/science.1148910

Glusman, G., Sosinsky, a, Ben-Asher, E., Avidan, N., Sonkin, D., Bahar, a, … Lancet, D. (2000). Sequence, structure, and evolution of a complete human olfactory receptor gene cluster. *Genomics*, *63*(2), 227–45. doi:10.1006/geno.1999.6030

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L., & Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biology*, *8*(2), R24. doi:10.1186/gb-2007-8-2-r24

Hébert, J. M., & McConnell, S. K. (2000). Targeting of cre to the Foxg1 (BF-1) locus mediates loxP recombination in the telencephalon and other developing head structures. *Developmental Biology*, *222*(2), 296–306. doi:10.1006/dbio.2000.9732

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). *Nature Neuroscience*, *7*(5), 483. doi:10.1038/nn0504-483

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–12. doi:10.1038/nature07829

Heintzman, N. D., & Ren, B. (2009). Finding distal regulatory elements in the human genome. *Current Opinion in Genetics & Development*, *19*(6), 541–9. doi:10.1016/j.gde.2009.09.006

Hirota, J., & Mombaerts, P. (2004a). The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.0400940101

Hirota, J., & Mombaerts, P. (2004b). The LIM-homeodomain protein Lhx2 is required for complete development of mouse olfactory sensory neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 8751–8755. doi:10.1073/pnas.0400940101

Hirota, J., Omura, M., & Mombaerts, P. (2007a). Differential impact of Lhx2 deficiency on expression of class I and class II odorant receptor genes in mouse. *Molecular and Cellular Neurosciences*. doi:10.1016/j.mcn.2007.01.014

Hirota, J., Omura, M., & Mombaerts, P. (2007b). Differential impact of Lhx2 deficiency on expression of class I and class II odorant receptor genes in mouse. *Molecular and Cellular Neurosciences*, *34*(4), 679–88. doi:10.1016/j.mcn.2007.01.014

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., … Young, R. A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, *155*(4), 934–947.

Hoppe, R., Breer, H., & Strotmann, J. (2006). Promoter motifs of olfactory receptor genes expressed in distinct topographic patterns. *Genomics*, *87*(6), 711–23. doi:10.1016/j.ygeno.2006.02.005

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., … Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*. doi:10.1038/ng.2871

Imakaev, M., Fudenberg, G., Mccord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., … Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization, *9*(10). doi:10.1038/NMETH.2148

Iwata, T., Kaneko, S., Shiwa, Y., Enomoto, T., Yoshikawa, H., & Hirota, J. (2013). Bacillus subtilis genome vector-based complete manipulation and reconstruction of genomic DNA for mouse transgenesis. *BMC Genomics*, *14*(1), 300. doi:10.1186/1471-2164-14-300

Iwema, C. L., & Schwob, J. E. (2003). Odorant receptor expression as a function of neuronal maturity in the adult rodent olfactory system. *The Journal of Comparative Neurology*, *459*(3), 209–22. doi:10.1002/cne.10583

Jacquet, B. V, Muthusamy, N., Sommerville, L. J., Xiao, G., Liang, H., Zhang, Y., … Ghashghaei, H. T. (2011). Specification of a Foxj1-dependent lineage in the forebrain is required for embryonic-to-postnatal transition of neurogenesis in

the olfactory bulb. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(25), 9368–82. doi:10.1523/JNEUROSCI.0171-11.2011

Jordan-Sciutto, K. L. (1999). Fetal Alz-50 Clone 1, a Novel Zinc Finger Protein, Binds a Specific DNA Sequence and Acts as a Transcriptional Regulator. *Journal of Biological Chemistry*, *274*(49), 35262–35268. doi:10.1074/jbc.274.49.35262

Kehayova, P., Monahan, K., Chen, W., & Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(41), 17195–200. doi:10.1073/pnas.1114357108

Khan, M., Vaes, E., & Mombaerts, P. (2011). Regulation of the probability of mouse odorant receptor gene choice. *Cell*, *147*(4), 907–21. doi:10.1016/j.cell.2011.09.049

Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., … Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–7. doi:10.1038/nature09033

Kolterman, B. E., Iossifov, I., & Koulakov, A. A. (n.d.). A race model for singular olfactory receptor expression.

Krolewski, R. C., Packard, A., & Schwob, J. E. (2013). Global expression profiling of globose basal cells and neurogenic progression within the olfactory epithelium. *The Journal of Comparative Neurology*, *521*(4), 833–59. doi:10.1002/cne.23204

Kudrycki, K., Stein-izsak, C., Behn, C., Grillo, M., Akeson, R., & Margolis, F. L. (1993). Olf-l-Binding Site : Characterization of an Olfactory Neuron-Specific Promoter Motif, *13*(5), 3002–3014.

Landry, J., Sharov, A. a, Piao, Y., Sharova, L. V, Xiao, H., Southon, E., … Wu, C. (2008). Essential role of chromatin remodeling protein Bptf in early mouse embryos and embryonic stem cells. *PLoS Genetics*, *4*(10), e1000241. doi:10.1371/journal.pgen.1000241

Landry, J. W., Banerjee, S., Taylor, B., Aplan, P. D., Singer, A., & Wu, C. (2011). Chromatin remodeling complex NURF regulates thymocyte maturation. *Genes & Development*, *25*(3), 275–86. doi:10.1101/gad.2007311

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9. doi:10.1038/nmeth.1923

Lee, W., Cheng, T., & Gong, Q. (2008). Olfactory Sensory Neuron-Specific and Sexually Dimorphic Expression of, *1086*(August 2007), 1076–1086. doi:10.1002/cne

Lefebvre, J. L., Kostadinov, D., Chen, W. V, Maniatis, T., & Sanes, J. R. (2012). Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature*, *488*(7412), 517–21. doi:10.1038/nature11305

Lettice, L. a. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. doi:10.1093/hmg/ddg180

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. *Current Biology : CB*, *20*(17), R754–63. doi:10.1016/j.cub.2010.06.070

Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell*, *157*(1), 13–25. doi:10.1016/j.cell.2014.02.009

Levine, M., & Tjian, R. (2003). Transcription regulation and animal diversity. *Nature*, *424*(6945), 147–51. doi:10.1038/nature01763

Lewcock, J. W., & Reed, R. R. (2004). A feedback mechanism regulates monoallelic odorant receptor expression. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(4), 1069–74. doi:10.1073/pnas.0307986100

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal Interactions and Olfactory Receptor Choice. *Cell*, *126*(2), 403–413. Retrieved from http://www.sciencedirect.com/science/article/pii/S0092867406008555

Lomvardas, S., & Thanos, D. (2002). Modifying gene expression programs by altering core promoter chromatin architecture. *Cell*, *110*(2), 261–71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12150933

Lyons, D. B., Allen, W. E., Goh, T., Tsai, L., Barnea, G., & Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. *Cell*, *154*(2), 325–36. doi:10.1016/j.cell.2013.06.039

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., … Lomvardas, S. (2011a). An epigenetic signature for monoallelic olfactory receptor expression. *Cell*, *145*(4), 555–70. doi:10.1016/j.cell.2011.03.040

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., … Lomvardas, S. (2011b). An epigenetic signature for monoallelic olfactory receptor expression. *Cell*, *145*(4), 555–70. doi:10.1016/j.cell.2011.03.040

Malnic, B., Hirono, J., Sato, T., & Buck, L. B. (1999). Combinatorial receptor codes for odors. *Cell*, *96*(5), 713–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10089886

Mariani, F. V, & Martin, G. R. (2003). clues from the limb, *423*(May), 319–325.

Matys, V., Kel-Margoulis, O. V, Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., … Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research*, *34*(Database issue), D108–10. doi:10.1093/nar/gkj143

McIntyre, J. C., Bose, S. C., Stromberg, A. J., & McClintock, T. S. (2008). Emx2 stimulates odorant receptor gene expression. *Chemical Senses*, *33*(9), 825–37. doi:10.1093/chemse/bjn061

Michaloski, J. S., Galante, P. A. F., & Malnic, B. (2006). Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. *Genome Research*. doi:10.1101/gr.5185406

Michaloski, J. S., Galante, P. A. F., Nagai, M. H., Armelin-Correa, L., Chien, M.-S., Matsunami, H., & Malnic, B. (2011). Common promoter elements in odorant and vomeronasal receptor genes. *PloS One*, *6*(12), e29065. doi:10.1371/journal.pone.0029065

Miyamichi, K., Serizawa, S., Kimura, H. M., & Sakano, H. (2005). Continuous and overlapping expression domains of odorant receptor genes in the olfactory epithelium determine the dorsal/ventral positioning of glomeruli in the olfactory bulb. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*. doi:10.1523/JNEUROSCI.0324-05.2005

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A Regulatory Archipelago Controls Hox Genes Transcription in Digits. *Cell*, *147*(5), 1132–1145. Retrieved from http://www.sciencedirect.com/science/article/pii/S0092867411012736

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell*, *147*(5), 1132–45. doi:10.1016/j.cell.2011.10.023

Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., … Stamatoyannopoulos, J. a. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, *489*(7414), 83–90. doi:10.1038/nature11212

Niimura, Y., & Nei, M. (2005). Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(17), 6039–44. doi:10.1073/pnas.0501922102

Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A., & Sakano, H. (2007). Deletion of the core-H region in mice abolishes the expression of three proximal odorant receptor genes in cis. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 20067–72. doi:10.1073/pnas.0706544105

Noordermeer, D., Leleu, M., Schorderet, P., Joye, E., Chabaud, F., & Duboule, D. (2014). Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *eLife*, *3*, e02557. doi:10.7554/eLife.02557

Pattabiraman, K., Golonzhka, O., Lindtner, S., Nord, A. S., Taher, L., Hoch, R., … Rubenstein, J. L. R. (2014). Transcriptional regulation of enhancers active in protodomains of the developing cerebral cortex. *Neuron*, *82*(5), 989–1003. doi:10.1016/j.neuron.2014.04.014

Pennacchio, L. a, Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. a, Shoukry, M., … Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499–502. doi:10.1038/nature05295

Porter, S. C., North, a K., Wedel, a B., & Kustu, S. (1993). Oligomerization of NTRC at the glnA enhancer is required for transcriptional activation. *Genes & Development*, *7*(11), 2258–2273. doi:10.1101/gad.7.11.2258

Qasba, P., & Reed, R. R. (1998). Tissue and zonal-specific expression of an olfactory receptor transgene. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *18*(1), 227–36. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9412503

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. a, Flynn, R. a, & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279–83. doi:10.1038/nature09692

Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., … Hayashizaki, Y. (2010). An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *Cell*, *141*(2), 369. doi:10.1016/j.cell.2010.04.004

Ressler, K. J., Sullivan, S. L., & Buck, L. B. (1993a). A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=7683976

Ressler, K. J., Sullivan, S. L., & Buck, L. B. (1993b). A zonal organization of odorant receptor gene expression in the olfactory epithelium. *Cell*, *73*(3), 597–609. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7683976

Ribich, S., Tasic, B., & Maniatis, T. (2006). Identification of long-range regulatory elements in the protocadherin-alpha gene cluster. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19719–24. doi:10.1073/pnas.0609445104

Ruthenburg, A. J., Li, H., Milne, T. a, Dewell, S., McGinty, R. K., Yuen, M., … Allis, C. D. (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell*, *145*(5), 692–706. doi:10.1016/j.cell.2011.03.053

Sato, Y., Miyasaka, N., & Yoshihara, Y. (2007). Hierarchical regulation of odorant receptor gene choice and subsequent axonal projection of olfactory sensory neurons in zebrafish. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(7), 1606–15. doi:10.1523/JNEUROSCI.4218-06.2007

Serizawa, S., Ishii, T., Nakatani, H., Tsuboi, A., Nagawa, F., Asano, M., … Iwakura, Y. (2000). Mutually exclusive expression of odorant receptor transgenes. *Nature Reviews Neuroscience*. doi:10.1038/76641

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., & Sakano, H. (2003a). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science (New York, NY)*. doi:10.1126/science.1089122

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., & Sakano, H. (2003b). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science (New York, N.Y.)*, *302*(5653), 2088–94. doi:10.1126/science.1089122

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., & Sakano, H. (2003c). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science (New York, N.Y.)*, *302*(5653), 2088–94. doi:10.1126/science.1089122

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., … Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–20. doi:10.1038/nature11243

Shin, H., Liu, T., Manrai, A. K., & Liu, X. S. (2009). CEAS: cis-regulatory element annotation system. *Bioinformatics (Oxford, England)*, *25*(19), 2605–6. doi:10.1093/bioinformatics/btp479

Shykind, B. M., Rohani, S. C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., … Barnea, G. (2004a). Gene switching and the stability of odorant receptor gene choice. *Cell*. doi:10.1016/j.cell.2004.05.015

Shykind, B. M., Rohani, S. C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., … Barnea, G. (2004b). Gene switching and the stability of odorant receptor gene choice. *Cell*, *117*(6), 801–15. doi:10.1016/j.cell.2004.05.015

Soshnikova, N., & Duboule, D. (2009). Epigenetic temporal control of mouse Hox genes in vivo. *Science (New York, N.Y.)*, *324*(5932), 1320–3. doi:10.1126/science.1171468

Sosinsky, a, Glusman, G., & Lancet, D. (2000). The genomic structure of human olfactory receptor genes. *Genomics*, *70*(1), 49–61. doi:10.1006/geno.2000.6363

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 307. doi:10.1186/1471-2105-9-307

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323–348.

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. doi:10.1038/nature11232

Vassalli, A., Feinstein, P., & Mombaerts, P. (2011). Homeodomain binding motifs modulate the probability of odorant receptor gene choice in transgenic mice. *Molecular and Cellular Neurosciences*, *46*(2), 381–96. doi:10.1016/j.mcn.2010.11.001

Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., & Mombaerts, P. (2002a). Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=12194868

Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., & Mombaerts, P. (2002b). Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. *Neuron*, *35*(4), 681–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12194868

Vassar, R., Ngai, J., & Axel, R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. *Cell*, *74*(2), 309–18. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8343958

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., … Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–858. doi:10.1038/nature07730

Visel, A., Taher, L., Girgis, H., May, D., Golonzhka, O., Hoch, R. V, … Rubenstein, J. L. R. (2013). A high-resolution enhancer atlas of the developing telencephalon. *Cell*, *152*(4), 895–908. doi:10.1016/j.cell.2012.12.041

Wang, M. M., & Reed, R. R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature*. doi:10.1038/364121a0

Wang, S. (2002a). Cloning of a Novel Olf-1/EBF-like Gene, O/E-4, by Degenerate Oligo-based Direct Selection. *Molecular and Cellular Neuroscience*, *20*(3), 404–414. doi:10.1006/mcne.2002.1138

Wang, S. (2002b). Cloning of a Novel Olf-1/EBF-like Gene, O/E-4, by Degenerate Oligo-based Direct Selection. *Molecular and Cellular Neuroscience*, *20*(3), 404–414. doi:10.1006/mcne.2002.1138

Wang, S. S., Betz, A. G., & Reed, R. R. (2002). Cloning of a novel Olf-1/EBF-like gene, O/E-4, by degenerate oligo-based direct selection. *Molecular and Cellular Neurosciences*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=12139918

Wang, S. S., Tsai, R. Y., & Reed, R. R. (1997). The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=9151732

Wedel, a, Weiss, D. S., Popham, D., Dröge, P., & Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science (New York, N.Y.)*, *248*(4954), 486–90. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1970441

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., … Young, R. A. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell*, *153*(2), 307–319.

Wysocka, J., Swigut, T., Xiao, H., Milne, T. a, Kwon, S. Y., Landry, J., … Allis, C. D. (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, *442*(7098), 86–90. doi:10.1038/nature04815

Young, J. M., Luche, R. M., & Trask, B. J. (2011). Rigorous and thorough bioinformatic analyses of olfactory receptor promoters confirm enrichment of O/E and homeodomain binding sites but reveal no new common motifs. *BMC Genomics*, *12*(1), 561. doi:10.1186/1471-2164-12-561

Zang, C., Schones, D. E., Zeng, C., Cui, K., Zhao, K., & Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics (Oxford, England)*, *25*(15), 1952–8. doi:10.1093/bioinformatics/btp340

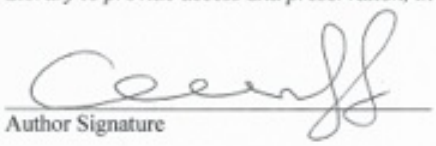
Zhang, X., & Firestein, S. (2002). The olfactory receptor gene superfamily of the mouse. *Nature Neuroscience*, *5*(2), 124–33. doi:10.1038/nn800

**Publishing Agreement**

*It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.*

***Please sign the following statement:***

*I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.*



Author Signature

9/10/2014

Date