Cooperative interactions enable singular olfactory receptor expression

Monahan K[1], Schieren I[1]., Cheung J[2]., Mumbey-Wafula A[1]., Monuki E.S[3],and Lomvardas S[1,4,5*]

[1]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032

[2]New York Structural Biology Center, New York, NY 10027

[3]Department of Pathology and Laboratory Medicine, School of Medicine, University of California Irvine, Irvine, CA 92697

[4]Department of Neuroscience, Columbia University, New York, NY 10032

[5]Zuckerman Mind Brain and Behavior Institute, Columbia University, New York, NY 10027

[*] Corresponding Author sl682@columbia.edu

Abstract

The monogenic and monoallelic expression of only one out of >1000 olfactory receptor (ORs) genes requires the formation of large heterochromatic chromatin domains that sequester the OR gene clusters. Within these domains, intergenic transcriptional enhancers evade heterochromatic silencing and converge into interchromosomal hubs that assemble over the transcriptionally active OR. The significance of this nuclear organization in OR choice remains elusive. Here, we show that transcription factors Lhx2 and Ebf specify OR enhancers by binding in a functionally cooperative fashion to stereotypically spaced motifs that defy heterochromatin. Specific displacement of Lhx2 and Ebf from OR enhancers resulted in pervasive, long-range, and *trans* downregulation of OR transcription, whereas pre-assembly of a multi-enhancer hub increased the frequency of OR choice in *cis*. Our data provide genetic support for the requirement and sufficiency of interchromosomal interactions in singular OR choice and generate general regulatory principles for stochastic, mutually exclusive gene expression programs.

Introduction

The mammalian main olfactory epithelium (MOE) provides an extreme example of cellular diversity orchestrated by the seemingly stochastic, monogenic, and monoallelic expression of a single olfactory receptor (OR) gene. Each mature olfactory sensory neuron (mOSN) in the MOE expresses only one OR that is chosen from a pool of more than two thousand alleles (Buck and Axel, 1991; Chess et al., 1994). The basis of the regulation of OR gene expression is chromatin-mediated transcriptional silencing followed by the stochastic de-repression and, thereby, transcriptional activation of a single OR allele that prevents the de-repression of additional OR genes (Dalton and Lomvardas, 2015; Monahan and Lomvardas, 2015). OR gene clusters are assembled into constitutive heterochromatin at early stages of OSN differentiation (Magklara et al., 2011), a process that represses OR transcription and preserves the monogenic and stochastic nature of OR expression (Lyons et al., 2014). Heterochromatic silencing is reinforced by the interchromosomal convergence of OR loci to OSN-specific, highly compacted nuclear bodies that assure complete transcriptional silencing of ORs in mOSNs (Clowney et al., 2012). Consequently, OR gene activation requires de-silencing by lysine demethylase Lsd1 (Lyons et al., 2013) and spatial segregation of the single chosen OR allele towards euchromatic nuclear territories (Armelin-Correa et al., 2014; Clowney et al., 2012). Translation of the newly transcribed OR mRNA activates a co-opted arm of the unfolded protein response (Dalton et al., 2013) and induces a feedback signal (Lewcock and Reed, 2004; Serizawa et al., 2005; Shykind et al., 2004) that turns off Lsd1, preventing the de-silencing and activation of additional OR genes (Lyons et al., 2013).

In the context of this repressive chromatin environment, OR gene choice requires the action of intergenic enhancers that escape heterochromatic silencing and activate the transcription of their proximal ORs (Khan et al., 2011; Markenscoff-Papadimitriou et al., 2014; Serizawa et al., 2003). These euchromatic enhancer “islands”, which we named after Greek Islands, engage in interchromosomal interactions with each other, and with the transcriptionally active OR allele, forming a multi-enhancer hub for OR transcription outside of the repressive OR foci (Clowney et al., 2012; Lomvardas et al., 2006; Markenscoff-Papadimitriou et al., 2014). The convergence of multiple Greek Islands to the chosen OR allele suggests that strong, feedback-eliciting OR gene transcription may be achieved only in the context of a multi-enhancer hub (Markenscoff-Papadimitriou et al., 2014). Yet, the molecular mechanisms that specify Greek Islands in the context of OR heterochromatin and, thus, enable their elaborate interactions during OSN differentiation remain unknown.

Here, we present a detailed molecular characterization of the Greek Islands, which revealed a common genetic signature and occupancy by shared sequence-specific transcription factors, allowing us, for the first time, to incapacitate them as a whole. ChIP-seq studies of FAC-sorted mOSNs revealed that most of the previously characterized Greek Islands, and several newly identified islands, are bound by two transcription factors: Lhx2 and Ebf. Computational analysis of the co-bound ChIP-seq peaks from Greek Islands revealed stereotypically positioned Lhx2 and Ebf binding sites that together constitute a "composite" binding motif that affords cooperative binding *in vivo.* This motif is highly enriched in Greek Islands relative to OR promoters and Lhx2/Ebf co-bound sites genome-wide. Considering the prevalence and specificity of this composite motif in Greek Islands, we designed a synthetic "fusion" protein that binds to this consensus sequence and not to individual Lhx2 or Ebf motifs *in vitro*. We found that overexpression of this fusion protein in mOSNs eliminated chromatin accessibility at most Greek Islands, and resulted in strong transcriptional downregulation of every OR, regardless of their genomic distance, or even their chromosomal linkage to a Greek Island. Finally, partial pre-assembly of a Greek Island hub in *cis,* by insertion of an array of 5 Greek Islands next to the Greek Island Rhodes, significantly increased the frequency of expression of Rhodes-linked OR genes. These manipulations provide genetic support for the requirement of *trans* enhancement in OR gene expression, and are consistent with the sufficiency of a multi-enhancer hub formation for OR gene choice.

Results

Greek Islands are co-bound by Lhx2 and Ebf

Greek Islands share a characteristic chromatin modification signature and *in vivo* footprints for transcription factors Lhx2 and Ebf (Markenscoff-Papadimitriou et al., 2014). To test the predicted binding of Lhx2 and Ebf, we performed ChIP-seq experiments using crosslinked chromatin prepared from FAC-sorted mOSNs, the neuronal population that stably expresses ORs in a singular fashion. To isolate mOSNs we FAC-sorted GFP^+ cells from the MOEs of OMP-IRES-GFP mice, as previously described (Magklara et al., 2011). The Ebf antibody we used for these experiments cross-reacts with all 4 Ebf proteins, Ebf1-4, (data not shown), which are all highly expressed in the MOE. Because of the genetic redundancy of the Ebf genes in the MOE (Wang et al., 2004), and because the 4 Ebf members form homo- and hetero-dimers with identical sequence specificity (Wang et al., 1997), we did not attempt to further distinguish between the 4 paralogues. For Lhx2

ChIP-seq studies we used a custom-made antibody (Roberson et al., 2001). The specificity of these antibodies is supported by motif analysis of the Lhx2 and Ebf ChIP-seq experiments, which revealed that the Lhx2 and Ebf binding sites are the most highly enriched motifs respectively (Figure 1A). Genome-wide, we identified 9,024 peaks for Ebf and 16,311 Lhx2 peaks, with 4,792 peaks being co-bound by both proteins (Figure 1B). Despite the *in vivo* recognition of an essentially identical motif in pro/pre-B cells(Gyory et al., 2012; Kong et al., 2016; Treiber et al., 2010b), where Ebf acts as master regulator of B-cell differentiation (Mandel and Grosschedl, 2010), there is little overlap between the genome-wide binding of Ebf in mOSNs and B-cell progenitors (data not shown). Genes proximal to Lhx2 and Ebf co-bound sites in mOSNs are statistically enriched for functions related to olfactory transduction and axonogenesis (Supplemental Figure S1A), consistent with a combinatorial role of these transcription factors in OSN differentiation and function (Hirota and Mombaerts, 2004; Wang et al., 1993; Wang et al., 2004; Wang et al., 1997).

The apparent coordinated binding of Lhx2 and Ebf to genomic DNA is exaggerated within the boundaries of heterochromatic OR clusters where individually bound peaks are rare and have low signal. Specifically, there are 63 peaks that are co-bound by both Lhx2 and Ebf, 2 Ebf-only, and 51 Lhx2-only peaks (Figure 1C) in the ~36MB of OR clusters, a significantly higher rate of overlap than the rate observed genome-wide ($p=1.5e^{-15}$ and $p=5.7e^{-9}$, respectively, Binomial test). Notably, most Ebf and Lhx2 co-bound sites in OR clusters have much stronger ChIP signal than singly bound sites (Supplemental Figure S1B). Several of these co-bound sites within OR clusters are among the regions of highest ChIP-seq signal in the genome, suggesting that they are bound in a large fraction of mOSNs (Supplemental Figure S1C, D), whereas individually bound peaks barely pass our peak-calling threshold (Supplemental Figure S1B). Co-bound sites within OR clusters coincide with 21 of the 35 previously characterized Greek Islands (table 1). For example, visual inspection of three Greek Islands, Crete, Sfaktiria and Lipsi, revealed strong Lhx2 and Ebf binding despite the high levels of flanking H3K9me3 on these OR clusters (Figure 1D). ATAC-seq analysis in the same cellular population revealed increased chromatin accessibility at the exact genomic location of the Lhx2 and Ebf ChIP-seq peaks, but very little accessibility across the rest of the OR cluster (Figure 1D). Each of these sites also exhibits a reduction of the heterochromatic modifications, H3K9me3 and H3K79me3, over the body of the element, and locally increased levels of the active enhancer mark H3K7ac (Supplemental Figure S1E). Overall, this chromatin signature is shared by the full set of Ebf and Lhx2 co-bound sites within OR gene clusters (Figure

1E and Supplemental Figure S1F). Thus, Lhx2/Ebf co-bound sites that do not correspond to the original Greek Islands (Supplemental Table 1) likely represent additional, less frequently active Islands that were only detected here due to the increased sensitivity of our mOSN-specific analysis (Anafi in Figure 1D and Supplemental Figure S1G for comparison between old and new Islands). In contrast, Greek Islands from the original set that lack Ebf and Lhx2 binding in mOSNs also deviate from the characteristic "epigenetic" signature obtained from whole MOE experiments (Supplemental Table1). Thus, these sites are likely to be functionally distinct or active in a different population of cells within the MOE, and are not included within our revised set of Greek Islands.

OR gene promoters are also significantly enriched for predicted Lhx2 and Ebf binding sites (Clowney et al., 2011; Michaloski et al., 2006; Plessy et al., 2012; Young et al., 2011), and mutations of individual Ebf and Lhx2 sites have been shown to reduce OR expression *in vivo* (Rothman et al., 2005). However, as a whole, OR gene promoters are inaccessible and not bound by these transcription factors in mOSNs (Figure 1F). Specifically, only 10 OR promoters show significant binding of Ebf and Lhx2 within 500bp of the TSS. Interestingly, these 10 ORs are expressed at levels similar to the median of OR expression (Supplemental Figure S1H and Supplemental Table 1). Thus, detection of Lhx2 and Ebf binding on these peaks is not explained by the unusually frequent transcriptional activation of their proximal ORs.

OR identity does not affect Greek Island accessibility

Based on the observation that most OR promoters display a complete lack of chromatin accessibility and Lhx2/Ebf binding, we asked if these promoters are accessible to transcription factors only in the OSNs that transcribe them. We FAC-sorted OSNs that express the same OR allele, by isolating GFP^{+} cells from Olfr17-IRES-GFP(Gogos et al., 2000), Olfr151-IRES-tauGFP(Bozza et al., 2002) , and Olfr1507-IRES-GFP(Shykind et al., 2004) knock-in mice (Figure 2A, B), and performed ATAC-seq (Buenrostro et al., 2013). As expected, the promoters Olfr1507, Olfr17 and Olfr151, are highly accessible when these genes are transcriptionally active (Figure 2C), consistent with local chromatin de-compaction being a prerequisite for OR gene transcription (Magklara et al., 2011). We also detect an increase in transposase accessibility at the 3'UTR of transcriptionally active OR alleles, an unusual feature that is not characteristic of most transcriptionally active genes in OSNs (Figure 2C, Supplemental Figure S2A-C).

In contrast to the differences between active and silent OR promoters, the overall pattern of accessibility of the Greek Islands is very similar in OSN populations that have chosen different ORs (Figure 2D). Very few Greek Islands display significantly different accessibility in the three OSN populations when compared to mOSNs (Figure 2E), and most fluctuations represent small but uniform shifts in Greek Island accessibility. For example, the H enhancer, which is proximal to Olfr1507 and is required for Olfr1507 expression, has a relatively strong ATAC-seq signal in all four cell populations and is not significantly stronger in Olfr1507+ cells than in mOSNs (Figure 2D,E, Supplemental Figure S2E). However, we do note some evidence for differential activity of Greek Islands. In particular, Kimolos, the Greek Island proximal to Olfr151, has relatively weak ATAC-seq signal in mOSN and in Olfr17+ and Olfr1507+ OSNs, but exhibits a nearly 10-fold increase in signal in Olfr151 expressing cells (Figure 2D,E, Supplemental Figure S2D). Thus, it appears that a large number of Greek Islands are broadly accessible in most OSNs, irrespective of the identity of the chose OR allele, whereas OR promoters are accessible only in the OSNs in which they are active.

Proximity of Lhx2/Ebf motifs correlates with binding on Greek Islands

What mechanism allows binding of Lhx2 and Ebf on Greek Islands but not OR promoters in most OSNs? We hypothesized that additional factors may bind specifically on Greek Islands but not on OR promoters, providing the functional distinction between the two types of regulatory elements. Motif analysis of the Lhx2 and Ebf ChIP-seq peaks using HOMER (Heinz et al., 2010) did not reveal additional known DNA binding sites that are shared by a significant portion of Greek Islands, other than Lhx2 and Ebf. *De novo* motif analysis, however, uncovered a novel, “composite” motif that corresponds to Lhx2 and Ebf sites positioned next to each other (Figure 3A). This composite Lhx2/Ebf motif is structurally very similar to the numerous heterodimeric motifs identified by an *in vitro* screen for sequences that are co-bound by a variety of transcription factor combinations (Jolma et al., 2015). A stringent Lhx2/Ebf composite motif, with score over 10 (see material and methods), is found in 35 of the 63 Greek Islands (Supplemental Figure S3A, Supplemental Table 2). This motif is significantly enriched in Greek Islands in comparison with OR promoters and with Lhx2/Ebf co-bound peaks outside of OR clusters (Figure 3B). In aggregate, the 43 strong composite motifs found in Greek Islands reside exactly at a local depletion of the ATAC-seq signal from mOSNs, consistent with *in vivo* occupancy of these sequences by transcription factors (Figure 3C) as previously described (Buenrostro et al., 2015).

Visual inspection of the aligned composite motifs revealed that the Ebf site is less constrained to stretches of C and G bases than solitary Ebf motifs, and instead tolerates stretches of pyrimidines and purines that retain a highly stereotypic spacing from the Lxh2 site (Figure 3D, E, top panel). Recent observations suggested that the relative positioning of DNA binding motifs compensates for the fluctuation of individual nucleotides *in vivo* (Farley et al., 2016). Similarly, the positioning of transcription factors on the face of the DNA double helix, as determined by the spacing between transcription factor binding sites, is more important than the relative strength of individual binding sites for the assembly of the IFN beta enhanceosome (Merika et al., 1998; Thanos and Maniatis, 1995). Thus, we asked if composite motifs with lower scores, which, predominantly, have degenerate Ebf motifs (Supplemental Figure S3B), still meet these stereotypic constraints. Indeed, despite increased fluctuation in the nucleotide level, the stereotypic distribution between purines and pyrimidines is retained in composites with score above 5 (Figure 3D, E bottom panel), with a new total of 55 out of 63 Greek Islands having a composite motif under this less stringent cutoff. Moreover, of the 28 Greek Islands that lack a strong composite, 20 have an Ebf site that is juxtaposed to an Lhx2 site. The distance between Ebf and Lhx2 sites in these Greek Islands is significantly shorter than the distance between Ebf and Lhx2 sites in OR promoters and in co-bound peaks outside of OR gene clusters (Figure 3F). In total, 61/63 islands contain a composite motif and/or very proximal Lhx2 and Ebf binding sites (Supplemental Table 2). Thus, although Lhx2 and Ebf frequently bind at the same genomic targets genome-wide, their binding on Greek Islands is restricted to stereotypically proximal Lhx2 and Ebf motifs.

Lhx2 is essential for Ebf binding on Greek Islands

An immediate prediction of our computational analyses is that Lhx2 and Ebf bind cooperatively to composite DNA binding motifs. In addition, Lhx2 and Ebf binding to these stereotypically spaced motifs may result in synergistic recruitment of coactivators that cannot be recruited by the individually bound proteins. In either case of functional cooperativity, deletion of either Lhx2 or Ebf should abolish the binding of the other transcription factor on Greek Islands. To test this we deleted Lhx2 from mOSNs, using a conditional Lhx2 allele (Mangale et al., 2008) that we crossed to OMP-IRES-Cre mice. Deletion of Lhx2 with OMP-IRES-Cre, results in loss of Lhx2 immunofluorescence (IF) signal from mOSNs, while Lhx2 protein levels are unaffected in progenitor and immature OSNs (Figure 4A). To enrich for Lhx2 KO mOSNs in our analyses, we introduced the

Cre-inducible fluorescent reporter tdTomato (Madisen et al., 2010) to our genetic strategy and we FAC-sorted Tomato^{+} Lhx2-/- mOSNs. RNA-seq of the FAC-sorted cells verifies the deletion of the floxed exons in mOSNs and the generation of a mutant Lhx2 mRNA that does not encode for Lhx2 protein (supplemental figure S4A). Lhx2 gene deletion results in significant downregulation of OR gene expression (Figure 4B), a result consistent with the partial deletion of a different floxed Lhx2 allele from mOSNs (Zhang et al., 2016). Furthermore, upon Lhx2 deletion the Lhx2 ChIP-seq signal is depleted genome-wide and from the Greek Islands (Figure 4C, D). Importantly, deletion of Lhx2 in mOSNs, results in loss of Ebf binding from Lipsi (Figure 4C) and from nearly all other Greek Islands (Figure 4E). ATAC-seq on the Lhx2 KO OSNs also shows strong reduction of ATAC-peaks from Greek Islands (Figure 4F), suggesting that Lhx2 and Ebf co-binding on Greek Islands is essential for their sustained accessibility in this heterochromatic environment. Consistent with the role of composite motifs on cooperative Lhx2 and Ebf binding, the effects of Lhx2 deletion on Ebf binding are weaker at co-bound sites outside the OR clusters compared to Greek Islands (Figure 4G, Supplemental Figure 4B). Interestingly, the general downregulation of OR gene transcription upon Lhx2 deletion extends to ORs that do not have Lhx2 motifs on their promoters (Supplemental Figure S4C), suggesting that Lhx2 activates OR transcription predominantly through the Greek Islands.

Inhibition of Greek Islands inhibits OR transcription

Our data suggest that composite motifs are an ideal target for genetic manipulations that could inhibit the function of Greek Islands as a whole. We reasoned that if we could fuse Lhx2 and Ebf DNA binding domains (DBD) at a proper distance, we could generate a DNA binding peptide that has high affinity for the composite but not for individual motifs. Because the DNA binding specificity of homeobox genes is low and is influenced by their partners (Chan et al., 1994; Passner et al., 1999), the Lhx2 DBD could be easily incorporated in this design. Ebf, however, has high affinity and specificity for its cognate palindromic motif, where it binds as a dimer (Hagman et al., 1993; Hagman et al., 1995; Travis et al., 1993; Wang and Reed, 1993; Wang et al., 1997). Crystal structure of an Ebf1 homodimer bound to DNA revealed that each DBD monomer contacts both halves of the palindromic motif and forms a clamp-like structure that likely stabilizes DNA binding (Treiber et al., 2010a). Thus, in order to reduce Ebf affinity for DNA without affecting its sequence specificity, we fused only one Ebf DBD to the Lhx2 DBD with various flexible linkers. Fusion of the two DNA binding domains with a

20aa protein linker generated a protein with affinity for the composite motif but not for individual Lhx2 and Ebf sites *in vitro* (Figure 5A). Competition experiments demonstrate that only unlabeled oligos containing the composite, and not individual Lhx2 or Ebf motifs, can compete off the binding of the fusion protein to the composite motif at up to 100x molar excess (Figure 5B,C). Remarkably, insertion of only 2 DNA bases between the Lhx2 and the Ebf binding sites on the composite motif impairs its ability to compete with the wild type composite (Figure 5C). Further increase of the distance between the two sites essentially eliminates any competitive advantage the composite motif had over the individual Lhx2 and Ebf sites (Figure 5C). Thus, the fusion of the Lhx2 DBD to a single Ebf DBD creates a novel DNA binding protein that recognizes the composite motif with sensitivity to the stereotypical distance of the two individual DNA binding sites.

To express the fusion protein in the MOE, we generated a transgenic construct under the control of the tetO promoter. This transgene includes a bi-cistronic mCherry reporter using the 2A peptide(Kim et al., 2011) (Supplemental Figure S5A), which allows isolation of the transgene-expressing OSNs by FACS. We analyzed two independent founders, which we crossed to OMP-IRES-tTA knock-in mice (Gogos et al., 2000), to obtain expression of the fusion protein specifically in mOSNs (Supplemental Figure S5B). We hypothesized that the fusion protein will compete with endogenous Lhx2 and Ebf for binding on composite motifs, acting as a repressor of the Greek Islands (Figure 5D). Indeed, ATAC-seq analysis shows strong reduction of ATAC-seq signal from the Greek Islands upon expression of the fusion protein in mOSNs (Figure 5E, F), suggesting the displacement of the heterochromatin-resisting transcription factors from OR enhancers. Unfortunately, both the Lhx2 and the Ebf antibodies we used in our ChIP-seq experiments cross-react with the DBD domains of the fusion protein (data not shown), thus we could not confirm by ChIP-seq their displacement from the Greek Islands. However, RNA-seq analysis of the FAC-sorted mCherry+ cells revealed significant reduction of OR transcription as a whole (Figure 5E, G). Although the repressing effect of the fusion protein does not extend to non-OR genes residing outside of OR clusters (Figure 5E), fusion protein expression has a ubiquitous repressive effect on OR transcription (Figure 5G). In fact, of the 500 most significantly downregulated genes 482 are ORs ($p<1e-313$, hypergeometric test). Consistent with this, genome-wide analysis shows that while ORs are homogeneously repressed by the fusion protein, genes containing Ebf-, Lhx2-, or Ebf and Lhx2-bound promoters are, on average, transcriptionally unaffected (Figure 5H, I). Similar to the effects of the Lhx2 deletion, the repressive effects of the fusion protein on OR transcription does not depend on the presence of

Ebf and Lhx2 motifs on OR promoters (Supplemental Figure 5C), supporting the Greek Island-mediated repressive effects of the fusion protein.

Multi-enhancer hubs activate OR transcription

The widespread downregulation of OR gene expression detected in Lhx2 KO and fusion protein expressing mOSNs suggests that the effects of Greek Island inhibition extend over large genomic distances, or even across chromosomes. Visual inspection of an isolated OR cluster on chromosome 16, which does not contain a Greek Island and is over 15MB away from the closest OR cluster with a Greek Island, supports the strong downregulation of ORs in *trans* (Supplemental Figure S6A). Genomewide, for both Lhx2 KO and fusion protein expressing mOSNs, there is a uniform reduction in OR expression regardless of the presence of a Greek Island in a cluster (Figure 6A, B). There is also a uniform reduction of OR expression independently of the distance between the OR and the closest Greek Island, and this reduction occurs irrespective of the motif content of OR promoters (Figure 6C, D). Moreover, comparable downregulation was observed for the ORs with a Greek Island in the promoter region (distance=1) and for ORs that lack a Greek Island in *cis* (distance set to 1e+08) (Figure 6C, D). Thus, functional incapacitation of Greek Islands by two distinct genetic manipulations results in specific but pervasive disruption of OR expression irrespective of OR promoter sequence, OR distance from a Greek Island, presence of a Greek Island within the OR cluster, or even presence of a Greek island within the same chromosome.

If *trans* interactions between Greek Islands are essential for OR transcription and the formation of a multi-enhancer hub over a stochastically chosen OR allele is the low probability event responsible for singular OR choice, then increasing the number of Greek Islands in an OR cluster should increase the expression frequency of the ORs in that cluster. To test this prediction, we introduced, by homologous recombination, an array of 5 Greek islands (Lipsi, Sfaktiria, Crete, H and Rhodes, hereafter termed LSCHR) next to the endogenous Rhodes, a Greek Island from chromosome 1 (Figure 6E). This array comprised the ATAC-seq accessible core of each Greek Island (392-497bp) together with 50bp of endogenous flanking sequence (supplemental table 6). We chose Rhodes for this manipulation for two reasons: First, the ATAC-seq and ChIP-seq signals on Rhodes are among the strongest between the 63 Greek Islands, which combined with the almost complete H3K9me3 local depletion suggest that it is accessible and bound by Lhx2 and Ebf in the

majority of mOSNs. Thus, any transcriptional changes observed by this manipulation would not be attributed to increased Lhx2 and Ebf binding on this locus. Second, there are no additional Greek Islands within a genomic distance of over 80MB on chromosome 1, thus formation of a Greek Island hub over this cluster requires recruitment of unlinked OR enhancers. We, therefore, reasoned that Rhodes-proximal ORs would be more responsive to the insertion of additional enhancers next to their local Greek Island, than ORs residing on chromosomes with multiple Greek Islands.

q-PCR analysis of cDNA prepared from the whole MOE of LSCHR knock-in mice and wild type littermates, shows strong transcriptional upregulation of the ORs in the Rhodes cluster that is almost doubled in homozygote knock-in mice in comparison to heterozygote littermates (Figure 6F). ORs from different clusters and non-OR genes are not strongly upregulated by this manipulation; however, four of the ORs in the Rhodes cluster are upregulated by more than 8 fold in the homozygote knock-in mice (Figure 6F). In fact, Olfr1412, which is the most upregulated OR in the Rhodes cluster approaches mRNA levels comparable to Olfr1507, the most highly expressed OR in the MOE (Figure 6F). RNA FISH experiments demonstrate that this transcriptional upregulation represents an increase in frequency of choice, rather than an increase of transcription rates in each cell (Figure 6G, H). ORs from different clusters do not appear significantly affected by this genetic manipulation, a result that is not surprising since the *trans* effects of this enhancer array would be distributed to more than a 1000 OR genes.

Discussion

In most cell types interchromosomal interactions are rare and thus far appear to represent technical or biological noise (Nagano et al., 2015), rather than to provide a reliable mechanism for gene regulation. Various studies suggest that the majority of genomic interactions are restricted within topologically associated domains (TADs) that show little variation between different tissues (Dixon et al., 2012). Specific genomic interactions between TADs are infrequent, and interactions between different chromosomes are even less prominent (Lieberman-Aiden et al., 2009; Rao et al., 2014). However, in certain biological contexts, specific interchromosomal interactions are readily detected by imaging and genomic approaches, or have been inferred genetically. For example, during X chromosome inactivation, there is a “chromosome kissing” step that occurs just before one of the two chromosomes is inactivated(Bacher et al., 2006; Masui et al., 2011; Xu et al., 2006). During T and B cell differentiation interchromosomal interactions regulate antigen receptor choice and cellular differentiation (Hewitt et al., 2008; Spilianakis et al., 2005). The stochastic induction of the human IFN beta gene by virus infection requires the formation of interchromosomal interactions between the IFN beta enhancer and NF-kappa B-bound Alu repeats (Apostolou and Thanos, 2008). Finally, stochastic photoreceptor choice in drosophila omatidia is determined by DNA elements that, genetically, appear to communicate in *trans* (Johnston and Desplan, 2014). Thus, although interchromosomal interactions may not be involved in gene regulation in most cell types, their stochastic and infrequent nature may be ideal for the execution of non-deterministic, and mutually exclusive regulatory processes like OR gene choice (Dekker and Mirny, 2016).

The involvement of interchromosomal interactions in OR gene choice was first postulated by the demonstration that the prototypical OR enhancer, the H enhancer (Serizawa et al., 2003), interacts in *trans* with transcriptionally active ORs (Lomvardas et al., 2006). The significance of these interactions was questioned as deletion of the H enhancer affected the expression of only three proximal ORs (Fuss et al., 2007; Nishizumi et al., 2007). Subsequently, however, additional OR enhancers, the Greek Islands, were discovered to a current total of 63 elements. The striking similarities between these elements in regards of the transcription factors that bind to them, combined with the demonstration that Greek Islands form a complex network of interchromosomal interactions (Markenscoff-Papadimitriou et al., 2014), suggested that extensive functional redundancy may mask the effects of single or even double (Khan et al., 2011) enhancer deletions in *trans*. The non-redundant role of Greek Islands for the expression of certain ORs in *cis* may be attributed to the

inability of some OR promoters to recruit enhancers from other chromosomes, making them completely dependent on the presence of a proximal enhancer for this function. In other words, even if *trans* enhancement is required for the activation of every OR gene, a fraction of them may depend on the assistance of a local Island for the recruitment of *trans* enhancers. Such qualitative promoter differences are consistent with the observation that enhancer deletions affect only some ORs in a cluster, and by the fact that certain ORs can be expressed as transgenic minigenes (Vassalli et al., 2002), while others can be expressed as transgenes only in the presence of an enhancer in *cis* (Serizawa et al., 2003). The proposed redundant function of Greek Islands as *trans* enhancers may have facilitated the rapid evolution of this gene family, which expanded dramatically during the transition from aquatic to terrestrial life (Niimura and Nei, 2007). Activation of OR transcription by Greek Islands in *trans* allows the functional expression of newly evolved OR alleles in mOSNs, without a requirement for physical linkage to an enhancer- a property fully compatible with gene expansion through retrotransposition, segmental duplication, and chromosomal translocation. Thus, OR gene activation through non-deterministic genomic interactions in *trans* may provide a mechanism that “shuffles the deck” and assures that a newly evolved OR allele will be expressed at a frequency similar to that of the existing OR repertoire.

Global and *trans* action of OR enhancers

A correlation between the formation of interchromosomal Greek Island hubs and OR transcription was previously established by ectopic expression of Lamin b receptor (Lbr) in mOSNs, and by conditional deletion of transcriptional co-activator Bptf, either of which caused reduction of Greek Island interactions in *trans* and pervasive OR downregulation (Clowney et al., 2012; Markenscoff-Papadimitriou et al., 2014). However, these manipulations have more general consequences that extend beyond the regulation of Greek Island interaction. For example, ectopic Lbr expression in mOSNs caused a general rearrangement of nuclear topology and disrupted the aggregation of OR clusters, making difficult to distinguish between the effects on interchromosomal OR clustering and interchromosomal Greek Island interactions. Deletion of Bptf on the other hand, although it only disrupted interchromosomal associations between Greek Islands, it also caused a developmental arrest in the OSNs that may, or may not, be related to the failure to activate OR expression.

To minimize indirect effects that may confound the interpretation of these manipulations, we targeted a common and highly specific genetic signature among Greek Islands, the composite motif. This DNA sequence

359 constitutes a remarkable example of highly constrained and stereotypically distributed transcription factor
360 binding motifs that is shared between most Greek Islands, and is highly enriched relative to OR promoters and
361 co-bound sites genomewide. Overexpression of a "synthetic" fusion protein that specifically recognizes the
362 composite motif eliminated ATAC-seq signal from Greek Islands in mOSNs, suggesting that it displaced the
363 endogenous Lhx2 and Ebf proteins on most OR enhancers. Similar observations were made for the conditional
364 Lhx2 deletion, which also reduced the chromatin accessibility of Greek Islands and abolished Ebf binding from
365 these elements. The strong and specific downregulation of the OR transcriptome in both Lhx2 knock out and in
366 fusion protein expressing mOSNs, clearly reveals the critical and ubiquitous role of the Greek Islands as key
367 regulators of OR expression. The fact that these transcriptional effects extend to ORs that have neither a
368 Greek Island in *cis* nor Lhx2/Ebf motifs on their promoters, is consistent with the role of Greek Islands as *trans*
369 OR gene enhancers. Although indirect effects are possible, the fact that three distinct genetic manipulation that
370 target the Greek Islands cause widespread downregulation of OR expression, provides strong genetic support
371 for the requirement of interchromosomal interactions in OR gene choice.
372
373 **Same transcription factors different chromatin states**
374 The experimental demonstration that every Greek Island is co-bound by Lhx2 and Ebf, the same
375 transcription factors predicted to bind on most OR promoters, is unexpected because of the fundamentally
376 different chromatin states of the two types of regulatory elements in mOSNs. OR promoters are inaccessible in
377 the mixed mOSN population, and only upon FAC-sorting cells that express the same OR, could we obtain
378 evidence for OR promoter accessibility. In contrast, the enhancers of OR genes appear accessible and bound
379 by Lhx2 and Ebf in a large fraction of mOSNs. The stereotypically proximal positioning of Lxh2 and Ebf motifs
380 on OR enhancers emerged as the key determinant for these differences, since the functionally cooperative
381 binding of Lhx2 and Ebf on proximal motifs *in vivo* appears to counteract the propagating properties of the
382 surrounding heterochromatin. Interestingly, the composite Lhx2/Ebf motif that we identified on Greek Islands is
383 structurally very similar to the numerous heterodimeric motifs identified by an *in vitro* screen for sequences that
384 are co-bound by a variety of transcription factors (Jolma et al., 2015). Thus, the solution that was adopted by
385 intergenic OR enhancers to generate heterochromatin-resistant binding sites, may be generally utilized by
386 other transcription factors in a variety of genomic contexts and regulatory needs. In support of this, the striking

stereotypy of Lhx2 and Ebf motifs in Greek Islands, also known as “rigid motif grammar” (Long et al., 2016), is reminiscent of the constraint spacing of transcription factor binding sites in the IFNbeta enhanceosome (Panne et al., 2007; Thanos and Maniatis, 1995).

A Multi-Enhancer Hub for Robust and Singular OR expression

The concept that Greek Islands may have stronger affinity for Lhx2 and Ebf than OR promoters, immediately provides a molecular solution for the need of a multi-enhancer hub for stable and robust OR transcription. In the event that an OR promoter becomes de-silenced and occupied by Lhx2 and Ebf, singular or weak binding by these transcription factors will be unstable, due to the competing forces of flanking heterochromatin. However, if an OR promoter is surrounded by multiple strong sites of cooperative binding, like the ones we detect in high frequency on the Greek Islands, then every time Lhx2 and Ebf fall off an OR promoter they will be sequestered by local, high affinity sites, which may also act as a replenishing source for these transcription factors. In other words, interchromosomal Greek island hubs may create local regions of high Lhx2 and Ebf concentration that is essential for continuous binding on the low affinity sites of a chosen OR promoter and high transcription rates.

Thus, we propose a model whereby the deployment of multiple, individually weak components that function in a coordinated and hierarchical fashion to activate OR transcription. According to this model, first, cooperative interactions between Lhx2 and Ebf result in stable binding to Greek Islands, which prevents flanking heterochromatin from spreading and silencing these intergenic elements. Because composite motifs are specifically enriched on Greek Islands similar cooperative interactions between Lhx2 and Ebf cannot protect OR promoters from heterochromatic silencing (Figure 7A). Second, cooperative interactions between Greek Islands assemble numerous Lhx2 and Ebf elements into a multi-chromosomal enhancer hub (Figure 7B). When this hub forms stable interactions with a stochastically chosen OR allele in *trans*, then heterochromatin is displaced, and cooperative enhancer-promoter interactions mediate stable Lhx2 and Ebf binding on the promoter, and therefore, transcriptional activation (Figure 7C). These cooperative interactions may be direct, homotypic interactions between Lhx2 and Ebf or facilitated by coactivator or mediator proteins that are recruited by these transcription factors. In either scenario, the same fundamental principles of cooperativity and synergy that govern the genetic switch between lysis and lysogeny in the lambda

bacteriophage (Ptashne, 2009), and promote the formation and function of the human IFNbeta enhanceosome (Thanos and Maniatis, 1995), may also regulate the formation of a 3-dimensional enhanceosome responsible for OR gene choice.

A multi-enhancer hub model explains why the few OR promoters that are bound by Lhx2 and Ebf in a large fraction of mOSNs are not transcribed at higher frequencies than most ORs. It also may explain why 63 OR genes, one for each Greek Island, are not simultaneously expressed in each mOSN: if numerous enhancers must cooperate for OR transcription, individual promoters, and even individual enhancer-promoter combinations, are not sufficient for OR transcription. But what prevents the formation of numerous multi-enhancer hubs, which could then activate more than one OR allele at a time? The answer to this critical question may be found in the transcriptional phenotype of the Rhodes knock-in mice, whereby 6 Greek Islands reside in tandem. In these mice, we detect a significant increase in the frequency of OR choice, suggesting that pre-assembly of an enhancer hub biases OR choice towards local ORs. However, this result also shows that despite an assumed increase in the potency of a Greek Island, the local ORs remain silent in the vast majority of the mOSNs. This implies the existence of a strong “thresholding” mechanism in the ability of Greek Islands to activate OR transcription, such that even 6 Islands acting together are inadequate to drive ubiquitous expression in most mOSNs. Thus, even if multiple enhancer hubs were to form in an mOSN nucleus, only the ones that surpass a critical number of interacting Greek Islands would lead to the activation of OR transcription. Such thresholding mechanism may be less strict in immature OSNs and progenitors, where low level OR co-expression is detected by single cell RNA-seq (Hanchate et al., 2015; Saraiva et al., 2015; Tan et al., 2015). Similar low level co-expression is detected in Lbr-expressing mOSNs, where the nuclear aggregation of OR clusters is prevented and the chromatin accessibility of OR genes is increased (Clowney et al., 2012). Thus, it is possible that the differentiation dependent silencing and aggregation of heterochromatic OR clusters into condensed nuclear foci contribute to this “all or none” transcriptional paradigm. In other words, the extreme silencing forces imposed by mOSNs to OR genes may result in extraordinary requirements for OR transcription, which can only be met by an activating multi-enhancer assembly of unprecedented complexity. Thus, even if more than one multi-enhancer hub could form in a nucleus, the number of transcription-competent hubs would be extremely limited if not singular. Combined with the kinetic restrictions imposed by

the OR-elicited feedback signal, and a recently reported post choice refinement process(Abdus-Saboor et al., 2016), our model provides a mechanistic solution for the singular choice of one out of >2000 OR alleles.

Acknowledgments

We would like to thank Drs. Richard Axel, Tom Maniatis, Richard Mann, and members of the Lomvardas lab for input, suggestions, and discussions and for critical reading of the manuscript. We are grateful to Dr. Mark Roberson for the anti-Lhx2 antibody and to Dr. Abbas Rizvi for assistance with statistical calculation. KM was funded by F32 post-doctoral fellowship GM108474 (NIH). This project was funded by R01DC013560 and R01DC015451 (NIH), and the HHMI Faculty Scholar Award.

Materials and Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Stavros Lomvardas (sl682@columbia.edu)

Experimental Model and Subject Details

Mice

Mice were treated in compliance with the rules and regulations of IACUC under a protocol number AC-AAAI1108. All experiments were performed on dissected whole main olfactory epithelium (MOE) or on freshly isolated, FAC sorted primary cells collected from whole main olfactory epithelium.

Mature olfactory sensory neurons (mOSNs) were sorted from OMP-IRES-GFP mice, which were previously described (Shykind et al., 2004). Olfr17+ cells were sorted from Olfr17-IRES-GFP mice (Shykind et al., 2004). Olfr151+ cells were sorted from Olfr151-IRES-tauGFP mice (Olfr151^{tm26Mom})(Bozza et al., 2002). Olfr1507+ cells were sorted from Olfr1507-IRES-GFP mice (Olfr1507^{tm2Rax})(Shykind et al., 2004).

Conditional deletion of Lhx2 in mOSNs was achieved by crossing Lhx2 conditional allele mice (Lhx2fl/fl: Lhx2$^{tm1.1Monu}$) (Mangale et al., 2008) with OMP-IRES-Cre mice (Omp$^{tm1(cre)Jae}$) (Eggan et al., 2004). In order to sort Lhx2 knock out mOSNs, a Cre-inducible tdtomato allele (Gt(ROSA)26Sor$^{tm14(CAG-tdTomato)Hze/J}$) (Madisen et al., 2010) was also included in this cross.

Transgenic mice bearing the TetO-Fusion-2A-mCherry construct were generated at the Columbia University Transgenic Mouse facility at the Irving Cancer Research Center. Fusion protein expression in mOSNs was achieved by crossing these mice with OMP-IRES-tTA mice (Omp$^{tm1(tTA)Gogo}$)(Yu et al., 2004).

Rhodes knock-in mice were generated by Biocytogen.

For ATAC-seq and ChIP-seq experiments, cells were sorted from adult male and female mice ranging in age from 7-16 weeks of age. For RNA-seq, cells were sorted from male and female mice ranging in age from 6-10 weeks.

Biological replicate samples are processed and collected separately from different mice.

Method Details

Fluorescence Activated Cell Sorting

Mice were sacrificed using CO_2 followed by cervical dislocation. The main olfactory epithelium (MOE) was dissected and transferred to ice-cold EBSS (Worthington Biochemical). The MOE was cut in to small pieces with a razor blade, and then dissociated with a papain dissociation system (Worthington Biochemical). Diced tissue was added to papain-EBSS, with at most 2 MOEs/mL, and incubated for 40 minutes at 37°C on a rocking platform. After 40 minutes, tissue was triturated 30 times, the supernatant containing dissociated cells was transferred to a new tube, and the cells were pelleted (300 rcf, 5 minutes, room temperature). Remaining papain was inhibited by resuspending the cell pellet with Ovomucoid protease inhibitor solution diluted 1:10 in EBSS, and the dissociated cells were pelleted (300 rcf, 5 minutes, room temperature).

For live cell sorts, dissociated cells were washed once with sort media (PBS with 2% Fetal Bovine Serum), and then resuspended in sort media supplemented with 100 U/mL DNase I (Worthington Biochemical), 4mM $MgCl_2$, and 500ng/mL DAPI (Invitrogen). These cells were passed through a 40uM cell strainer, and then FAC sorted. Live cells were selected by gating out DAPI positive cells.

For formaldehyde-fixed cell sorts, dissociated cells were resuspended in PBS + 1% methanol-free formaldehyde (Pierce). Cells were fixed at room temperature for 5 minutes, and then fixation was quenched by adding 1/10th volume of 1.25M glycine. Fixed cells were pelleted (500 rcf, 5 minutes, room temperature), washed once with sort media, resuspended in sort media, passed through a 40uM cell strainer, and then FAC sorted.

Preparation of Cross-linked Chromatin

Sorted fixed cells were pelleted (800 rcf,10 minutes, 4°C). Cell pellets were resuspended in ChIP Lysis Buffer (50mM Tris-HCl pH 7.5, 150nM NaCl, 0.5% NP-40, 0.25% Sodium Deoxycholate, 0.1% SDS, 1x protease inhibitors (Sigma, 05056489001)) and incubated on ice for 10 minutes. Nuclei were collected by centrifugation (1,000 rcf, 5 minutes, 4°C). The nuclei pellet was resuspended in shearing buffer (10mM Tris-HCl pH 7.5, 1mM EDTA, 0.25% SDS, 1x protease inhibitors) and then sheared to a size range of 200-500bp on a Covaris S220 Focused-ultrasonicator (16 minutes, 2% Duty Cycle, Peak Power 105W, 200 cycles per burst, 6°C). Sheared chromatin was centrifuged (10,000 rcf, 10 minutes, 4°C) to remove insoluble material. The DNA concentration

of the sheared chromatin was determined by fluorescent quantification (ThermoFisher, P7589). Shearing was assessed by agarose gel electrophoresis after DNA clean-up: chromatin was incubated for 30 minutes with RNase A, coss-links were reversed overnight at 65 C, and then DNA was column purified (Zymo Research, D4014).

Chromatin Immunoprecipitation of Cross-linked Chromatin

Formaldehyde cross-linked chromatin was used for ChIP of Ebf (Aviva, ARP32960_P050) and Lhx2 (Roberson et al., 2001). Approximately 2 ug of sheared chromatin was diluted to 500uL with ChIP Buffer (16.7mM Tris-HCl pH 8.1, 167mM NaCl, 1.2mM EDTA, 1.1% Triton X-100, 0.01% SDS, 1x protease inhibitors), and then pre-cleared with Protein G Dynabeads (Life Technologies) for one hour at 4°C. After preclearing, the supernatant containing cleared chromatin was transferred to a new tube, and approximately 100ng of chromatin was set aside as an input control. Input control chromatin was stored at 4°C until the elution step. The remaining pre-cleared chromatin was incubated overnight at 4°C with 1ug of Ebf antibody or 1uL of Lhx2 antibody. Protein G beads were blocked overnight with 2 mg/ml yeast tRNA (Life Technologies) in ChIP Buffer. The next day, the blocked beads were washed once with ChIP Buffer, then resuspended in antibody bound chromatin. Chromatin was incubated with beads for 1-2 hours at 4°C with rotation. Chromatin 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 pH 7.5. DNA was eluted from beads by incubating at 65°C for 30 minutes with 100uL ChIP Elution Buffer (1% SDS, 0.1M $NaHCO_3$, 4mM DTT) in a thermomixer set to 900 rpm. This elution was repeated and the elution fractions were pooled. The eluted DNA was incubated overnight at 65°C. Input chromatin was brought up to 200uL with elution buffer and also incubated at 65°C overnight. ChIP DNA and input DNA were column purified using Zymo ChIP DNA columns (Zymo Research, D5205) and eluted in 20uL of 10mM Tris-HCl pH 8.

Micrococcal Nuclease Digestion

Live sorted cells were pelleted (800 rcf, 15 minutes, 4°C) and then resuspended in Buffer 1 (0.3M Sucrose, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EGTA, 15mM Tris-HCl pH 7.5, 5mM Sodium Butyrate, 0.1mM PMSF, 0.5mM DTT, 1x protease inhibitors). Cells were lysed by adding an equal volume of Buffer 2 (0.4% Igepal CA-630, 0.3M Sucrose, 60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EGTA, 15mM Tris-HCl pH 7.5, 5mM Sodium Butyrate, 0.1mM PMSF, 0.5mM DTT, 1x Protease Inhibitor Cocktail). After addition of Buffer 2, cells were incubated on ice for 10 minutes, and then nuclei were pelleted (1,000 rcf, 10 minutes, 4°C). Nuclei were resuspended in MNase buffer (0.32M Sucrose, 4mM $MgCl_2$, 1mM $CaCl_2$, 50mM Tris-HCl pH 7.5, 5mM Sodium Butyrate, 0.1mM PMSF, 1x protease inhibitors). Nuclei were digested for 1 minute and 40 seconds with 0.2U of Micrococcal Nuclease (Sigma) per 1 million cells. Digestion was stopped by adding EDTA to a final concentration of 20mM. Undigested material was pelleted (10,000 rcf, 10 minutes, 4°C), and the supernatant (S1 fraction) was retained and stored at 4°C. The pelleted material was resuspended in Dialysis Buffer (1mM Tris-HCl pH 7.5, 0.2mM EDTA, 5mM Sodium Butyrate, 0.1mM PMSF, 1x protease inhibitors), rotated overnight at 4°C. Following dialysis, the insoluble material was pelleted (10,000 rcf, 10 minutes, 4°C) and the supernatant (S2 fraction) was retained. MNase digestion was assessed by agarose gel electrophoresis. The MNase treatment was optimized to yield a nucleosomal ladder comprising mostly mono and di-nucleosome sized fragments in the S1 fraction and di-nucleosome and larger sized fragments in the S2 fraction. The concentration of nucleic acid in the S1 and S2 fractions was determined by fluorescent quantification (ThermoFisher, P7589). Prior to ChIP equal volumes of S1 and S2 fractions were combined, and the total quantity of nucleic acid in the pooled fractions calculated to normalize between experiments.

Native Chromatin Immunoprecipitation

MNase digested native chromatin was used for ChIP with antibodies for H3K9me3 (Abcam, ab8898), H3K79me3 (Abcam ab2621), and H3K27ac (Active Motif, AM39133). Approximately 1ug of MNase digested chromatin was used per IP, with approximately 100ng reserved as an input control. Chromatin was diluted to 500uL in Wash Buffer 1 (50mM Tris-HCl pH 7.5, 10mM EDTA, 125mM NaCl, 0.1% Tween-20, 5mM Sodium Butyrate, 1x protease inhibitors), 1ug of antibody was added, and the binding reaction was rotated overnight at 4°C. For each IP, 10uL of Protein A Dynabeads (Life Technologies) and 10uL of Protein G Dynabeads were blocked overnight with 2 mg/ml yeast tRNA and 2mg/mL BSA in Wash Buffer 1. Blocked beads were added to antibody bound chromatin and rotated for 1-2 hours at 4°C. Bound beads were washed 4 times with Wash Buffer 1, 3 times with Wash Buffer 2 (50mM Tris-HCl pH 7.5, 10mM EDTA, 175mM NaCl, 0.1% Igepal CA-630,

5mM Sodium Butyrate, 1x protease inhibitors), and once with TE pH7.5. DNA was eluted from beads by incubating at 37°C for 15 minutes with 100uL Native ChIP Elution Buffer (10mM Tris-HCl pH7.5, 1mM EDTA, 1% SDS, 0.1 M $NaHCO_3$) in a thermomixer set to 900 rpm. This elution was repeated and the elution fractions were pooled. Input chromatin was brought up to 200uL with Native ChIP Elution Buffer. ChIP DNA and input DNA were column purified using Zymo ChIP DNA columns (Zymo Research, D5205) and eluted in 20uL of 10mM Tris-HCl pH 8.

ChIP-seq Library Preparation

ChIP-seq libraries were prepared with Nugen Ovation Ultralow v2 kits.

ATAC-seq

ATAC-seq libraries were prepared from live sorted cells using the protocol developed by Buenrostro et al (Buenrostro et al., 2015). Cells were pelleted (500 rcf, 5 minutes, 4°C) and then resuspended in lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630). Nuclei were immediately pelleted (1000 rcf, 10 minutes, 4°C). Pelleted nuclei were resuspended in transposition reaction mix prepared from Illumina Nextera reagents (for 50uL: 22.5uL water, 25uL 2xTD buffer, 2.5uL Tn5 Transposase). The volume of the Tn5 transposition reaction was scaled to the number of cells collected: 1uL mix per 1,000 cells. If fewer than 10,000 cells were collected by FACS, 10uL scale reactions were performed. See supplemental table 4 for a summary of ATAC-seq experiments. Transposed DNA column purified using a Qiagen MinElute PCR cleanup kit (Qiagen). The transposed DNA was then amplified using barcoded primers and NEBNext High Fidelity 2x PCR Master Mix (NEB). Amplified libraries were purified using Ampure XP beads (Beckman Coulter) at a ratio of 1.6ul of beads per 1uL of library and eluted in 30uL of elution buffer (10mM Tris-HCl pH 8, 0.1mM EDTA).

qRT-PCR

MOEs from 3 week old mice were dissected, cut in to small pieces with a razor blade, and then added to 1mL of Trizol. Samples were vortexed for 15 seconds, and then incubated for 5 minutes at room temperature. Total RNA was extracted by adding 200uL chloroform, vortexing for 15 seconds, incubating at room temperature for 2 minutes, then centrifugation at 12,000 rcf for 15 minutes at 4°C. The aqueous phase was collected and RNA was precipitated with isopropyl alcohol with 10ug/mL linear acrylamide (ThermoFisher) added as a carrier. The RNA pellet was washed twice with 75% ethanol, dried, then resuspended in nuclease free water. 3ug of RNA was DNase treated using the TURBO DNA-free Kit (ThermoFisher) according to manufacturer's instructions. cDNA was prepared from 800ng of RNA using SuperScriptIII (ThermoFisher) and used for qPCR with gene specific primers (supplemental table 8). Fold change was calculated using the ddCT approach, using Adcy3 as a reference gene to normalize between samples and expressing fold change relative to a wild type littermate control.

RNA-seq

Live sorted cells were pelleted (15 minutes, 800 rcf, 4°C), the supernatant was aspirated until 250uL of media remained, and then the cell pellet was resuspended in 750uL Trizol LS (ThermoFisher). Total RNA was extracted by adding 200uL chloroform, vortexing for 15 seconds, incubating at room temperature for 2 minutes, then centrifugation at 12,000 rcf for 15 minutes at 4°C. The aqueous phase was collected and RNA was precipitated with isopropyl alcohol with 10ug/mL linear acrylamide (ThermoFisher) added as a carrier. The RNA pellet was washed twice with 75% ethanol, dried, then resuspended in nuclease free water. 1ug of RNA was DNase treated using the TURBO DNA-free Kit (ThermoFisher) according to manufacturer's instructions. RNA-seq libraries were prepared from DNase-treated RNA using a TruSeq Stranded Total RNA with Ribo-Zero Gold Set B kit (Illumina RS-122-2302).

Deep Sequencing

Sequencing libraries were profiled on Bioanalyzer 2100 using a high sensitivity DNA kit (Agilent). Library concentration was determined by KAPA assay (KAPA Biosystems). Libraries were multiplexed and sequenced on an Illumina HiSeq with 50bp single-end or paired-end reads. See supplemental table 5 for a summary of sequencing data.

Recombinant DNA

Three versions of the fusion protein were designed with either 1, 2, or 4 repeats of a 5 amino acid linker sequence between the DNA binding domain of Ebf and the DNA binding domain of Lhx2 (supplemental table 6). Gene blocks encoding these proteins were synthesized by Integrated DNA Technologies. Gene blocks were TOPO cloned into a pcDNA3.1/V5-His expression vector (ThermoFisher). For in vivo expression, the fusion protein was subcloned into a pTRE2 vector that was modified to include a sequence encoding t2A-mCherry.

The 5-enhancer hub that was inserted into the Rhodes locus was generated using Gibson assembly of gene blocks synthesized by Integrated DNA Technologies (supplemental table 6).

Electrophoretic Mobility Shift Assay

Probe oligonucleotides (supplemental table 7) were annealed, gel purified, and end labeled with ^{32}P using T4 Polynucleotide kinase. Labeled probes were purified on a Sephadex G-50 column (GE Healthcare 27-5330-01).
Fusion protein, Lhx2, and Ebf were in vitro translated from pcDNA3 expression vectors bearing the T7 promoter (Promega, L1170). In vitro binding reactions were setup with 1uL of in vitro translation product, 0.5ug Poly(dI-dC), 1ug BSA, and 1xProtease Inhibitor cocktail in EMSA Binding Buffer (10mM HEPES pH 7.5, 40mM KCl, 5% Glycerol, 0.5% Igepal CA-630, 1mM DTT). For competition conditions, 2pmol (20-fold excess) or 10pmol (100-fold excess) of unlabeled, annealed oligonucleotides was added to the binding reactions. The binding reactions were incubated for 20 minutes on ice, and then 100fmol of radiolabeled probe was added. Following probe addition, binding reactions were incubated at room temperature for 10 minutes. Binding reactions were loaded on a native TBE polyacrylamide gel (6% acrylamide, 49:1 bis-acrylamide:acrylamide), and electrophoresed at 180 V for approximately 3 hours. After running, gels were transferred to filter paper, dried, and exposed to a phosphoimager screen overnight. Phosphoimager screens were scanned on a Typhoon FLA7000 or FLA9500.

RNA In Situ Hybridization

Regions of Olfr12 and Olfr1410 were cloned and verified by Sanger sequencing (Table S6). DNA for in vitro transcription was generated by PCR from these templates using an antisense primer bearing the T7 promoter. RNA probe was generated by in vitro transcription of 1ug of PCR product with T7 polymerase and Fluorescein RNA Labeling Mix (Sigma). Probe RNA was ethanol precipitated and resuspended in 50uL of hybridization buffer (50% formamide, 5x SSC, 5x Denhart's, 250ug/mL Yeast tRNA, 500ug/mL Salmon Sperm DNA). Prior to hybridization, probe was diluted 40x in hybridization buffer and denatured at 85 C for 5 minutes.

For RNA Fluorescent In Situ Hybridization (Figure 6G,H), MOE was dissected, embedded in OCT (ThermoFisher), and then frozen. Coronal cryosections were taken at a thickness of 10 um and then air dried for 10 minutes. Slides were fixed with 4% PFA in PBS for 10 minutes. After fixation, slides were rinsed three times with PBS, and then washed with PBST (PBS with 0.1% Triton X-100) for 10 minutes. Slides were then rinsed once with PBS and then incubated for 15 minutes in Acetylation Buffer (0.021 N HCl, 1.2% Triethanolamine (v/v), 0.25% Acetic Anhydride (v/v)). After acetylation, slides were rinsed 3 times with PBS, then probe was added and hybridized overnight at 65 C in a humid chamber. Following hybridization, slides were washed twice for 15 minutes with 0.2% SSC at 65 C, rinsed three times with PBS, and then blocked for 1 hour with TNB (0.1M Tris pH 7.5, 0.15M NaCl, 0.05% Blocking Reagent (Perkin Elmer)). After blocking, slides were stained overnight at 4 C with anti-fluoroscein POD antibody (Roche) diluted 1:100 in TNB. The next day, slides were rinsed twice with TNT buffer (0.1M Tris pH7.5, 0.15M NaCl, 0.1% Tween 20) and then washed in TNT buffer for 30 minutes. Slides were then treated with TSA amplification with Fluorescein labeling for 4 minutes, and then washed 6 times for 5-10 minutes with TNT buffer. DAPI was included in the final TNT wash at a concentration of 2.5ug/mL. Slides were then mounted with Vectashield (Vector Laboratories) and imaged.

Immunofluorescence

For imaging GFP and mCherry, MOE was dissected and fixed in 4% PFA for 30 minutes on ice prior to being embedded in OCT. Coronal cryosections were taken at a thickness of 12 to 14 um and then air dried for 10 minutes. Slides were fixed with 4% PFA for 10 minutes. After fixation, slides were washed with PBST (PBS with 0.1% Triton X-100), blocked in PBST-DS (PBST + 4% donkey serum), stained with DAPI (2.5ug/mL) in PBST-DS, washed with PBST, and then mounted with Vectashield and imaged.

For immunofluorescence (Figure 4A), MOE was dissected from 3 week old mice, embedded in OCT (ThermoFisher), and then frozen. Cryosections were taken and slides were fixed and washed as described above. Slides were stained with primary antibody (a-Lhx2, diluted 1:1000) in PBST-DS overnight at 4°C. Slides were then washed, stained with DAPI (2.5ug/mL) and secondary antibody (donkey a-rabbit conjugated to Alexa-488, diluted 1:1000, ThermoFisher) in PBST-DS for 1 hour, washed, and then mounted with Vectashield and imaged.

Microscopy

Confocal images were collected with a Zeiss LSM 700. Image processing was carried out with ImageJ (NIH).

Quantification and Statistical Analysis

ChIP-seq and ATAC-seq Sequencing Data Processing & Analysis

Adapter sequences were removed from raw sequencing data with CutAdapt. ChIP-seq and ATAC-seq reads were aligned to the mouse genome (mm10) using Bowtie2(Langmead and Salzberg, 2012). Default settings were used, except a maximum insert size of 1000 (-X 1000) was allowed for ATAC-seq and native ChIP-seq data since these data sets contained some large fragments. PCR duplicate reads were identified with Picard and removed with Samtools (Li et al., 2009). Samtools was used to select uniquely aligning reads by removing reads with alignment quality alignments below 30 (-q 30).

For Lhx2 and Ebf ChIP-seq, HOMER (Heinz et al., 2010) was used to call peaks of ChIP-seq signal using the “factor” mode and an input control. Consensus peak sets were generated by selecting peaks that overlapped between biological replicates and extending them to their combined size. For signal tracks, replicate experiments were merged, and HOMER was used to generate 1bp resolution signal tracks normalized to a library size of 10,000,000 reads.

For H3K9me3, H3K79me3, and H3K27ac ChIP-seq replicate experiments were merged, and HOMER was used to generate 1bp resolution signal tracks normalized to a library size of 10,000,000 reads. Regions enriched for H3K9me3 were identified by running HOMER peak calling in region mode, with the following settings: -L 0 -F 2.5 -size 2000 -minDist 4000. A consensus set of H3K9me3 enriched regions was generated by selecting regions that were enriched in both biological replicates.

For ATAC-seq, regions of open chromatin were identified by running HOMER peak calling in “region” mode, with a fragment size of 150bp and a peak size of 300bp. For ATAC-seq signal tracks, replicate experiments were merged, and HOMER was used to generate 1bp resolution signal tracks normalized to a library size of 10,000,000 reads. Reads were shifted 4bp upstream in order to more accurately map the Tn5 insertion site. Reads were extended to the full fragment length, as determined by paired-end sequencing, except for signal tracks of ATAC-seq fragment ends (Figure3C), which were generated by using a fragment size of 1bp.

A small number of failed ChIP-seq and ATAC-seq experiments were excluded from analysis. Failed experiments were identified based upon the presence of very few enriched peaks and low fold enrichment of reads in the identified peaks.

Classification of Greek Islands

The set of Greek Islands with function in Zebrafish or Mouse transgene assays (Markenscoff-Papadimitriou et al., 2014) was examined using data from OMP-GFP sorted mOSNs. With the exception of P, all were bound by Ebf and Lhx2, and were present within a region enriched for H3K9me3. Using these criteria, we used Bedtools2 (Quinlan and Hall, 2010) to generate a new, comprehensive list of sites with these properties in mOSNs. Specifically, the Greek Islands are defined as sites with overlapping peaks of Ebf and Lhx2 binding, within a region enriched for H3K9me3, and within an OR cluster (supplemental table 1).

OR Annotation

Annotation of OR gene transcripts was take from Ibarra-Soria et a l(Ibarra-Soria et al., 2014). ORs absent from this annotation but present in the UCSC mm10 annotation were added. For OR gene heatmaps, transcripts were merged by OR gene and the most 5’ annotated TSS and most 3’ annotate TES were used.

Lhx2 and Ebf co-binding

The background rate of overlap between Ebf and Lhx2 peaks was calculated for the whole genome excluding peaks within H3K9me3 positive regions of OR clusters. For Ebf, 4729 peaks overlapped Lhx2 peaks, whereas 4230 did not. In H3K9me3 positive regions of OR clusters, 63 out of 65 Ebf peaks overlapped Lhx2 peaks. In R, this overlap rate (63 out of 65), was compared to the genome-wide rate (4729/8850) using a Binomial test, with the alternative hypothesis that the overlap rate is greater in OR clusters, yielding a p-value of p = 2.557e-16. For Lhx2, 4729 peaks overlapped Ebf peaks, whereas 11468 did not. In H3K9me3 positive regions of OR clusters, 63 out of 114 Lhx2 peaks overlapped Ebf peaks. In R, this overlap rate (63 out of 114), was compared to the genome-wide rate (4729/16197) using a Binomial test, with the alternative hypothesis that the overlap rate is greater in OR clusters, yielding a p-value of p = 5.702e-09.

Heatmaps and Signal Plots

Heatmaps and signal plots were generated with Deeptools2(Ramirez et al., 2016). Unless otherwise specified, heatmaps were sorted by mean signal over the interval shown.

GO analysis

GREAT(McLean et al., 2010) was used for gene ontology analysis of sites bound by Lhx2 and Ebf in mOSN.

Motif Analysis

Motif analysis was performed with HOMER. Motif searches were run on Ebf and Lhx2 consensus peak sets for the 200bp region around the center of the peak, with repeat masking. The top de novo identified motif for Lhx2 and Ebf ChIP-seq was converted to TRANSFAC format, and plotted using Weblogo v3.5 (Crooks et al., 2004), together with motifs derived from published Ebf (GEO: GSE21978, Lin et al., 2010) and Lhx2 (GSE48068,(Folgueras et al., 2013)) ChIP-seq data.

HOMER was used to search Greek Island sequences for motifs with a variety of settings. A search for long motifs that allowed up to 4 mismatches (-len 18,20 -mis 4) recovered a sequence motif that was highly enriched relative to random background sequences (p=1e-59). HOMER was then used to optimize this motif against a background set of all mOSN Ebf and Lhx2 co-bound sites. This optimized motif is reported as the composite motif in Figure 3A. The HOMER derived composite motif was converted to TRANSFAC format and plotted using Weblogo v3.5.

For Figure 3B, HOMER was used to analyze sets of sequences and identify the highest scoring match to the composite motif in each sequence. The cumulative distribution of scores for each set were plotted in R. The binomial distribution was used to calculate the statistical significance the enrichment of composite motif sequence in Greek Island sequences relative to Ebf and Lhx2 Co-bound sites genome-wide.

To analyze specific instances of the composite motif for Figures 3C-F, HOMER was used to identify all instances of the composite motif scoring above a given threshold within Greek Islands. For figures 3D and E, the DNA sequence of Greek Island composite motifs, together with 20bp of sequence on either side, was converted to a matrix and visualized with deeptools2.

Composite motif multiple alignments (FigS3A,B) were generated with Jalview (Waterhouse et al., 2009).

Motif Proximity Analysis

HOMER was used to identify all instance of the de novo, ChIP-seq derived Ebf and Lhx2 motifs genome-wide. Bedtools was then used to identify all instances of Ebf motifs that occur within Greek Islands with a composite motif score of 10 or above. Greek Islands without an Ebf motif were excluded from further analysis. For the remaining islands, Bedtools2 was used to identify the closest Lhx2 motif to each of Ebf motif. For Greek Islands with multiple Ebf motifs, only the closest pair was retained. Additional sets of sequences were analyzed in the same manner, and the distribution of motif distances was plotted in R. An identical analysis centered on Lhx2 motifs rather than Ebf motifs yielded similar results.

Quantitative analysis of normalized ChIP-seq and ATAC-seq data

Normalized ATAC-seq and Ebf ChIP-seq data were generated in R using the Diffbind package (Ross-Innes et al., 2012). Diffbind was used to generate a read count for each peak for each data set. Count data was normalized using the “DBA_SCORE_TMM_READS_EFFECTIVE” scoring system, which normalizes using edgeR and the effective library size. After normalization, counts for biological replicates were averaged, and then counts were log transformed for plotting.

RNA-seq data processing and Analysis

Adapter sequences were removed from raw sequencing data with CutAdapt. RNA-seq reads were aligned to the mouse genome (mm10) using STAR (Dobin et al., 2013). Samtools was used to select uniquely aligning reads by remove reads with alignment quality alignments below 30 (-q 30). Signal tracks were generated with RSeQC (Wang et al., 2012), either retaining strand information (-d '+-,-+') (Supplementary Figure S4A), or without strand information (Figure 5F and Supplementary Figure S6A). RNA-seq signal plots are normalized to a library size of 1,000,000 reads. The Sashimi plot in Supplementary Figure S4A was generated using IGV(Robinson et al., 2011).

RNA-seq data analysis was performed in R with the DESeq2 package (Love et al., 2014). Genes with no counts in any condition were excluded. For Supplementary Figure 1H, DEseq2 was used to calculate FPKM values, and these values were plotted for subsets of OR genes. For all other plots, differential gene expression analysis was run comparing control mOSNs and Lhx2KO or Fusion protein expressing mOSNs. The base mean and log2fold change values from these analyses were used for plots. For MA-plots, significantly changed genes were identified with an adjusted p-value cutoff of 0.05.

For density and scatter plots of log2 fold change in OR transcript levels (FigS4C, FigS5C, Fig6C, Fig6D), ORs with low levels of expression (Normalized Base Mean < 5) were excluded.

Data and Software Availability

Data Resources

The data reported in this paper will be available through GEO.

Figure Legends

Figure 1. Greek Islands represent Lhx2 and Ebf co-bound regions residing in heterochromatic OR clusters

(A) The top sequence motif identified for mOSN ChIP-seq peaks is shown above sequence motifs generated from previously reported Lhx2 (Folgueras et al., 2013) and Ebf (Lin et al., 2010) ChIP-seq data sets. mOSN ChIPseq peaks were identified using HOMER and motif analysis was run on peaks present in both biological replicates.

(B) Overlap between mOSN Lhx2 and Ebf bound sites genome-wide.

(C) Overlap between mOSN Lhx2 and Ebf bound sites within OR clusters. For each factor, co-bound sites are significantly more frequent within OR clusters than in the rest of the genome ($p=5.702e^{-9}$ for Lhx2, $p=1.6e^{-15}$ for Ebf, Binomial test).

(D) mOSN ATAC-seq and ChIP-seq signal tracks for three representative OR gene clusters. Values are reads per 10 million. Below the signal tracks, OR genes are depicted in red and non-OR genes are depicted in blue. Greek Island locations are marked. Anafi is a newly identified Greek Island, located in a small OR cluster upstream of the Sfaktiria cluster. See also figure S1 and supplemental table 1. For ATACseq, pooled data is shown from 4 biological replicates, for ChIPseq, pooled data is shown from 2 biological replicates. For H3K9me3 ChIPseq, input control signal is subtracted from ChIP signal prior to plotting.

(E) mOSN ATAC-seq or ChIP-seq signal across 63 Greek Islands. Each row of the heatmap shows an 8kb region centered on a Greek Island. Regions of high signal are shaded red. Mean signal across all elements is plotted above the heatmap, values are reads per 10 million. All heatmaps are sorted in the same order, based upon ATAC-seq signal. See also figure S1 and supplemental table 1. For ATACseq, pooled data is shown from 4 biological replicates, for ChIPseq, pooled data is shown from 2 biological replicates.

(F) mOSN ATAC-seq and ChIP-seq signal tracks on OR genes. Each row of the heatmap shows an OR gene scaled to 4kb as well as the 2kb regions upstream and downstream. Plots and heatmap are scaled the same as in Figure 1E.

Figure 2. Greek island accessibility is independent of OR promoter choice

(A) GFP fluorescence (green) in MOE tissue sections from adult mice bearing Olfr17-IRES-GFP, Olfr151-IRES-tauGFP, or Olfr1507-IRES-GFP alleles. Nuclei are stained with DAPI (blue).

(B) Representative FACS data for Olfr-IRES-GFP mice. Data is shown from Olfr151-IRES-GFP mice. Viable (DAPI negative), GFP+ cells were collected for ATAC-seq.

(C) ATAC-seq signal tracks from GFP+ cells sorted from Olfr17-IRES-GFP (red), Olfr151-IRES-GFP (blue), or Olfr1507-IRES-GFP (green) mice. Values are reads per 10 million. The region spanning each targeted OR is shown for all three lines. See also figure S2. Pooled data is shown for 2 biological replicates.

(D) ATAC-seq signal over Greek Islands is shown for mOSNs and each Olfr-IRES-GFP line. All samples are sorted by signal in mOSNs. A blue arrow marks the H Enhancer, which is the Greek Island proximal to Olfr1507. A blue asterisk marks Kimolos, the Greek Island proximal to Olfr151, which has the strongest change in signal relative to mOSNs. See also figure S2. Pooled data is shown for 4 biological replicates for mOSNs, and 2 biological replicates for each Olfr-IRES-GFP sorted population.

(E) MA-plots showing fold change in ATAC-seq signal for each sorted Olfr-IRES-GFP population compared to mOSNs. Peak strength (normalized reads in peak) and fold change are shown for all ATAC-seq peaks; peaks that are not significantly changed are black and peaks that are significantly changed (FDR<0.001) are gold. Greek Islands are plotted as larger dots and are shown in red if significantly changed. Kimolos is marked with an asterisk in Olfr151 expressing cells, and H is marked with an arrow in Olfr1507 expressing cells. See also figure S2.

Figure 3. Greek Islands have stereotypically proximal Lhx2 and Ebf motifs

(A) Sequence logo of the Greek Island composite motif (center). The mOSN ChIP-seq derived Lhx2 and Ebf motifs logos are positioned above and below the corresponding regions of the composite motif. See also figure S3.

(B) Cumulative distribution plot of the score of the best composite motif site found in each of the 63 Greek Islands. Also plotted are cumulative distributions for co-bound sites outside of OR clusters and OR gene promoters. A score of 10 was selected as a stringent threshold for motif identification, and a score of 5 was selected for permissive motif identification. This motif is significantly enriched in Greek Islands relative to co-bound sites outside of OR clusters at both of these score cut-offs (Binomial test). See also supplemental table 2.

(C) Plot of the density of ATAC-seq fragment ends in the vicinity of Greek Island composite motifs sites scoring over 10. Plot shows mean signal and standard error in 5bp windows centered on 43 site composite motif sites (yellow).

(D) Multiple alignment of composite motif sequences from Greek Islands together with 20bp of flanking sequence. Each base is shaded by nucleotide identity: A= green, C=blue, G=yellow, T=red. Top panel depicts composite with score over 10 and bottom panel depicts composites with score between 5 and 10, together with a sequence logo of the motif present in those sequences.

(E) As in (D), except purines are shaded red and pyrimidines are shaded blue.

(F) For each site, the distance (in base pairs) between the closest Ebf-Lhx2 motif pair was determined. For each set of sites, the distribution of distances is shown as a boxplot. Sets of sites comprising Greek Islands with a strong composite motif, Greek Islands without a strong composite motif, Ebf and Lhx2 co-bound sites genome-wide, and OR gene promoters are compared. Sites without an Ebf motif are excluded. The distribution of distances between Ebf and Lhx2 motifs was significantly smaller for Greek Islands without a composite motif than for Ebf and Lhx2 bound sites genome-wide (two-sample, one-sided Kolmogorov–Smirnov test) See also supplemental table 2. n=25 for Greek Islands with Composite Score greater than 10; n=21 for Greek Islands with Composite Score less than 10; n=3805 for Co-bound sites genome wide; n=521 for OR promoters.

Figure 4. Lhx2 is required for Ebf binding predominantly on Greek Islands

(A) Lhx2 immunofluorescence (IF) (green) in MOE sections from 3 week old control (Lhx2fl/fl) and Lhx2 KO (OMP-IRES-Cre; Lhx2fl/fl) mice. Nuclei are stained with DAPI (blue). The Lhx2 immunoreactive cells on the basal layers of the MOE represent immature OSNs and progenitors that have not yet turned on OMP (and thus Cre) expression. See also figure S4 for demonstration of the Cre induced deletion at the mRNA level.

(B) MA-plot of OR transcript levels in FAC-sorted Lhx2KO mOSNs (OMP-IRES-Cre; Lhx2fl/fl; tdTomato) compared to FAC-sorted control mOSNs (OMP-IRES-GFP). Red dots correspond to OR genes with statistically significant transcriptional changes (adjusted p-value < 0.05). 3 biological replicates were included for control mOSNs and 2 biological replicates were included for Lhx2KO mOSNs.

(C) ChIP-seq and ATAC-seq signal tracks from FAC-sorted control mOSNs (OMP-IRES-GFP) and Lhx2KO mOSNs (OMP-IRES-Cre; Lhx2fl/fl; tdTomato) for the OR cluster containing the Greek Island Lipsi. Values are reads per 10 million. For ATACseq, pooled data from 4 biological replicates for control mOSNs are compared to data from 2 biological replicates for Lhx2 KO mOSNs. For ChIP, 2 biological replicates from mOSNs are compared to data from a ChIP from Lhx2 KO mOSNs.

(D-F) Heatmaps depicting Lhx2 and Ebf ChIP-seq and ATAC-seq signal across Greek Islands for FAC-sorted control and Lhx2KO mOSNs for the samples described in C.

(G) Log2 fold change in normalized Ebf ChIP-seq signal in Lhx2 KO mOSNs relative to control mOSNs for Greek Islands (red), compared to sites genome-wide that are bound by Ebf-only or both Ebf and Lhx2. See also figure S4B for MA-plot showing data for all peaks in each set. Mean counts for 2 biological replicates from mOSNs are compared to data from a ChIP from Lhx2 KO mOSNs.

Figure 5. Displacement of Lhx2 and Ebf from Greek Islands shuts off OR transcription

(A) Electrophoretic Mobility Shift Assay (EMSA) for binding of *in vitro* translated protein to DNA probes containing either an Ebf site, an Lhx2 site, or a composite site. Binding of three versions of the Fusion

protein with either 5, 10, or 20 amino acid linker peptides were compared to full length Lhx2 or full length Ebf1.

(B) EMSA for sequence selectivity of in vitro translated proteins. Binding of Fusion protein (20aa linker), Ebf1, and Lhx2 to composite motif probe was competed with a 20-fold molar excess of unlabeled oligo containing either an Lhx2 site, Ebf site, or composite site.

(C) EMSA for motif-spacing selectivity of in vitro translated proteins. Binding of Fusion protein (20aa linker) was competed with 100-fold molar excess of unlabeled oligo containing either wild type composite sequence or mutant composite generated by the insertion of 2-14 base pairs in two base pair increments. In the last two lanes the competitors are either a single Lhx2 or a single Ebf site.

(D) Schematic illustrating the proposed dominant-negative activity of the fusion protein for composite motif sites. See also figure S5 for depiction of the genetic strategy for mOSN overexpression.

(E) ATAC-seq and RNA-seq signal tracks from FAC-sorted control mOSNs and Fusion protein-expressing mOSNs for the OR cluster containing the Greek Island Lipsi. ATAC-seq values are reads per 10 million. RNA-seq values are reads per million. For ATACseq, pooled data from 4 biological replicates for control mOSNs are compared to data pooled from 2 independent founders of the Fusion Protein transgene. For RNAseq, representative tracks are shown for one of three biological replicates for control mOSNs and for one of 2 independent founders for the Fusion Protein transgene.

(F) Heatmap showing ATAC-seq signal across the Greek Islands for control mOSNs and Fusion protein-expressing mOSNs. Pooled data from 4 biological replicates for control mOSNs are compared to data pooled from 2 independent founders of the Fusion Protein transgene

(G) MA-plot (Dudoit and Fridlyand, 2002) of OR transcript levels in FAC-sorted mOSNs expressing fusion protein (OMP-IRES-tTA; tetO-Fusion-2a-mcherry) compared to FAC-sorted control mOSNs (OMP-IRES-GFP). Red dots correspond to OR genes with statistical significant transcriptional changes (adjusted p-value < 0.05). 3 biological replicates were included for control mOSNs and data from 2 independent founders were included for the Fusion Protein transgene.

(H) Violin plot of Log2 fold change in transcript levels of ORs (red) in mOSNs expressing fusion protein compared to control mOSN. ORs are compared to additional sets of genes: genes with Ebf and Lhx2 bound within 1kb of the TSS, genes with Lhx2-only bound within 1kb of the TSS, genes with Ebf-only bound within 1kb of the TSS, and non-OR genes without Ebf or Lhx2 binding.

(I) As in (H), with Log2 fold change in transcript levels shown as a heatmap for each set of genes.

Figure 6. Multi-enhancer hubs activate OR transcription

(A) MA-plot of OR transcript levels in FAC-sorted Lhx2KO (OMP-IRES-Cre; Lhx2fl/fl; tdTomato) mOSNs compared to FAC-sorted control mOSNs (OMP-IRES-GFP). Gold dots correspond to OR genes with statistical significant transcriptional changes. ORs in clusters without a Greek Island are shown as large dots, with significantly changed ORs in red. 3 biological replicates were included for control mOSNs and 2 biological replicates were included for Lhx2KO mOSNs.

(B) MA-plot of OR transcript levels in FAC-sorted Fusion protein expressing (OMP-IRES-tTA; tetO-Fusion-2a-mcherry) mOSNs compared to FAC-sorted control mOSNs (OMP-IRES-GFP). Gold dots correspond to OR genes with statistical significant transcriptional changes. ORs in clusters without a Greek Island are shown as large dots, with significantly changed ORs in red. 3 biological replicates were included for control mOSNs and data from 2 independent founders were included for the Fusion Protein transgene.

(C) Plot of OR distance from a Greek Island compared to Log2 Fold change in Lhx2KO mOSNs. ORs overlapping a Greek Island have distance set to 1. ORs on a chromosome without a Greek Island have distance set to 1e+08.

(D) Plot of OR distance from a Greek Island compared to Log2 Fold change in Fusion Protein expressing mOSNs. ORs overlapping a Greek Island have distance set to 1. ORs on a chromosome without a Greek Island have distance set to 1e+08.

(E) Targeted insertion of 5 Greek Islands (LSCHR) adjacent to Rhodes. Coordinates are mm10.

(F) RT-qPCR of OR transcript levels in MOEs of 3-week old LSCHR mice and wild-type littermate controls. Transcript levels are expressed as quantity relative to Adcy3, error bars are SEM. ORs are grouped by

presence inside or outside the OR cluster containing Rhodes, and within each group ORs are ordered by level of expression in wild-type mice. *p<0.05, **p<0.01, two-tailed student's t-test. For wild-type mice n=3, for LSCHR heterozygous and homozygous mice n=4.

(G) RNA in situ hybridization with probe for Olfr12 (green) in LSCHR homozygous and wild-type littermate control MOE at 2-weeks of age. Nuclei are labeled with DAPI (blue).

(H) RNA in situ hybridization with probe for Olfr1410 (green) in LSCHR homozygous and wild-type littermate control MOE at 2-weeks of age. Nuclei are labeled with DAPI (blue).

Figure 7. A Hierarchical Model for OR gene choice

(A) Lhx2 and Ebf bind in a functionally cooperative fashion on the composite motifs of the Greek Islands. Because these motifs are not juxtaposed in most OR promoters, Lhx2 and Ebf cannot overcome the heterochromatic silencing of OR promoters, thus their binding is restricted to the OR enhancers.

(B) Lhx2/Ebf bound OR enhancers are not strong enough to activate proximal OR alleles on their own and to facilitate stable transcription factor binding on their promoters.

(C) Lhx2/Ebf bound Greek Islands form an interchromosomal, multi-enhancer hub that recruits coactivators essential for the de-silencing of OR promoters and robust transcriptional activation of the OR allele that would be recruited to this hub.

Supplemental Figure Legends

Figure S1. Greek Islands represent Lhx2 and Ebf co-bound regions residing in heterochromatic OR clusters

(A) Top Gene Ontology terms from the Biological Process and MSigDB Pathway categories associated with genes proximal to sites bound by both Ebf and Lhx2.

(B) mOSN Lhx2 and Ebf ChIP-seq signal on OR Cluster Ebf+Lhx2 peaks (Greek Islands) compared to OR-cluster singly bound (Ebf or Lhx2) sites. Mean signal for each group is plotted above the heatmap, values are reads per 10 million. Both heatmaps are sorted in the same order, based upon ATAC-seq signal. Pooled data is shown from 2 biological replicates.

(C) Density plot of the distribution of peaks over Lhx2 ChIP-seq peak strength (normalized number of reads in each peak) for different categories of peaks. ChIP signal is calculated by averaging normalized peak counts from two biological replicates.

(D) Density plot of the distribution of peaks over Ebf ChIP-seq peak strength (normalized number of reads in each peak) for different categories of peaks. OR-cluster Ebf-only peaks are not included because there are only two peaks in this category. ChIP signal is calculated by averaging normalized peak counts from two biological replicates.

(E) ATAC-seq and ChIP-seq signal tracks for three Greek Islands, Sfaktiria, Crete and Lipsi. Greek Island position is highlighted in yellow. For heterochromatin modifications (H3K9me3 and H3K79me3), input control signal is subtracted from ChIP signal. Pooled data is shown from 4 biological replicates for ATACseq, 2 biological replicates for Lhx2, Ebf, H3K9me3, and H3K27ac, and one replicate for HeK79me3.

(F) ChIP-seq signal for histone modifications associated with heterochromatin and active enhancers in the vicinity of Greek Islands. Pooled data is shown from 2 biological replicates for H3K9me3 and H3K27ac, and one replicate for HeK79me3.

(G) Mean ATAC-seq or ChIP-seq signal for previously identified Greek Islands (Markenscoff-Papadimitriou et al., 2014)(blue shaded) that are bound by Ebf and Lhx2 compared to newly identified Ebf and Lhx2 bound islands (green shaded). Pooled data is shown from 4 biological replicates for ATACseq, 2 biological replicates for Lhx2, Ebf, H3K9me3, and H3K27ac, and one replicate for HeK79me3.

(H) Level of expression (FPKM) for OR genes in mOSNs determined by RNA-seq. ORs with a Greek Island within 500bp of the annotated TSS are plotted separately and in red. FPKM is the mean of three biological replicates.

Figure S2. Greek island accessibility is independent of OR promoter choice

Signal plots are from pooled data from 4 biological replicates for mOSNs and 2 replicates each for Olfr17-IRES-GFP+, Olfr151-IRES-GFP+, and Olfr1507-IRES-GFP+ cells.

(A) Profile of mean mOSN ATAC-seq signal over all genes. Genes are grouped into quartiles by level of expression in mOSNs.

(B) Profile of ATAC-seq signal over Olfr17 in all mOSNs and Olfr17-IRES-GFP expressing OSNs.

(C) Profile of ATAC-seq signal over Olfr1507 in all mOSNs and Olfr1507-IRES-GFP expressing OSNs.

(D) ATAC-seq signal in the vicinity of Olfr151 for each Olfr-IRES-GFP population. A blue asterisk marks Kimolos, the Greek Island with greatly increased signal in Olfr151-IRES-GFP expressing cells relative to mOSNs.

(E) ATAC-seq signal in the vicinity of Olfr1507 for each Olfr-IRES-GFP population. A blue arrow marks the location of H.

Figure S3. Greek Islands have stereotypically proximal Lhx2 and Ebf motifs

(A) Multiple alignment of composite motif sequences found in Greek Islands using a stringent cutoff (motif score > 10). Positions with at least 50% identity are shaded by nucleotide. A motif logo of the included sequences is shown below the alignment.

(B) Multiple alignment of weak composite motif sequences found in Greek Islands using a loose cutoff (10 > motif score >5). Positions with at least 50% identity are shaded by nucleotide. A motif logo of the included sequences is shown below the alignment.

Figure S4. Lhx2 is required for Ebf binding predominantly on Greek Islands

(A) Sashimi plot (Katz et al., 2010) of Lhx2 RNA-seq signal and splicing junctions in control and Lhx2KO mOSNs. A schematic of Lhx2 and the region affected by the conditional knockout is shown at the top. Representative data is shown for one replicate from each condition.

(B) MA-plots showing fold change in Ebf ChIP-seq signal for Lhx2KO mOSNs compared to control mOSNs. Peak strength (normalized reads in peak) and fold change are shown for all mOSN Ebf ChIP-seq peaks. Peaks are color coded by type; peaks that do not overlap a control mOSN Lhx2 peak are black, peaks that overlap an Lhx2 peak are blue, and Greek Islands are red. For ChIP, 2 pooled biological replicates from mOSNs are compared to data from a ChIP from Lhx2 KO mOSNs.

(C) Density plot of Log2 fold change in OR transcript levels in Lhx2KO mOSNs compared to control mOSNs, with ORs grouped based upon the motifs present in the promoter region (-500bp to the TSS). ORs with a very low level of expression (OR transcript level < 5 in Figure 4B) are not included. 3 biological replicates were included for control mOSNs and 2 biological replicates were included for Lhx2KO mOSNs.

Figure S5. Displacement of Lhx2 and Ebf from Greek Islands shuts off OR transcription

(A) Schematic of OMP-IRES-tTA driven expression of Fusion protein and mCherry in mOSNs

(B) mCherry fluorescence (red) in MOE tissue sections from animals bearing an OMP-IRES-tTA; tetO-Fusion-2A-mCherry transgene. Nuclei are stained with DAPI (blue).

(C) Density plots of Log2 fold change in OR transcript levels in Fusion protein expressing mOSNs compared to control mOSNs, with ORs grouped based upon the motifs present in the promoter region (-500bp to the TSS). ORs with a very low level of expression (OR transcript level < 5 in Figure 5G) are not included. 3 biological replicates were included for control mOSNs and data from 2 independent founders were included for the Fusion Protein transgene.

Figure S6. Multi-enhancer hubs activate OR transcription

(A) mOSN ATAC-seq and ChIP-seq signal tracks for an OR gene cluster without a Greek Island, scaled as in Figure 1A. Below the annotation, RNA-seq tracks show signal for control mOSNs, Lhx2KO mOSNs, and mOSNs expressing fusion protein. RNA-seq values are reads per million. An OR without Ebf or Lhx2 motifs in its promoter is circled. For ATACseq, pooled data from 4 biological replicates of control mOSNs is shown. For ChIPseq, pooled data from 2 biological replicates of control mOSNs is shown. For RNAseq, representative tracks are shown for one biological replicate from each condition.

References

Abdus-Saboor, I., Al Nufal, M.J., Agha, M.V., Ruinart de Brimont, M., Fleischmann, A., and Shykind, B.M. (2016). An Expression Refinement Process Ensures Singular Odorant Receptor Gene Choice. Curr Biol *26*, 1083-1090.

Apostolou, E., and Thanos, D. (2008). Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. Cell *134*, 85-96.

Armelin-Correa, L.M., Gutiyama, L.M., Brandt, D.Y., and Malnic, B. (2014). Nuclear compartmentalization of odorant receptor genes. Proceedings of the National Academy of Sciences of the United States of America *111*, 2782-2787.

Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. Nat Cell Biol *8*, 293-299.

Bozza, T., Feinstein, P., Zheng, C., and Mombaerts, P. (2002). Odorant receptor expression defines functional units in the mouse olfactory system. J Neurosci *22*, 3033-3043.

Buck, L., and Axel, R. (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell *65*, 175-187.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods *10*, 1213-1218.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-29.

Chan, S.K., Jaffe, L., Capovilla, M., Botas, J., and Mann, R.S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. Cell *78*, 603-615.

Chess, A., Simon, I., Cedar, H., and Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. Cell *78*, 823-834.

Clowney, E.J., LeGros, M.A., Mosley, C.P., Clowney, F.G., Markenskoff-Papadimitriou, E.C., Myllys, M., Barnea, G., Larabell, C.A., and Lomvardas, S. (2012). Nuclear aggregation of olfactory receptor genes governs their monogenic expression. Cell *151*, 724-737.

Clowney, E.J., Magklara, A., Colquitt, B.M., Pathak, N., Lane, R.P., and Lomvardas, S. (2011). 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*, 1249-1259.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome research *14*, 1188-1190.

Dalton, R.P., and Lomvardas, S. (2015). Chemosensory receptor specificity and regulation. Annual review of neuroscience *38*, 331-349.

Dalton, R.P., Lyons, D.B., and Lomvardas, S. (2013). Co-opting the unfolded protein response to elicit olfactory receptor feedback. Cell *155*, 321-332.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110-1121.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

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

Dudoit, S., and Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol *3*, RESEARCH0036.

Eggan, K., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004). Mice cloned from olfactory sensory neurons. Nature *428*, 44-49.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proceedings of the National Academy of Sciences of the United States of America *113*, 6508-6513.

Folgueras, A.R., Guo, X., Pasolli, H.A., Stokes, N., Polak, L., Zheng, D., and Fuchs, E. (2013). Architectural niche organization by LHX2 is linked to hair follicle stem cell function. Cell Stem Cell *13*, 314-327.

Fuss, S.H., Omura, M., and Mombaerts, P. (2007). Local and cis effects of the H element on expression of odorant receptor genes in mouse. Cell *130*, 373-384.

Gogos, J.A., Osborne, J., Nemes, A., Mendelsohn, M., and Axel, R. (2000). Genetic ablation and restoration of the olfactory topographic map. Cell *103*, 609-620.

Gyory, I., Boller, S., Nechanitzky, R., Mandel, E., Pott, S., Liu, E., and Grosschedl, R. (2012). Transcription factor Ebf1 regulates differentiation stage-specific signaling, proliferation, and survival of B cells. Genes & development *26*, 668-682.

Hagman, J., Belanger, C., Travis, A., Turck, C.W., and Grosschedl, R. (1993). Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes & development *7*, 760-773.

Hagman, J., Gutch, M.J., Lin, H., and Grosschedl, R. (1995). EBF contains a novel zinc coordination motif and multiple dimerization and transcriptional activation domains. The EMBO journal *14*, 2907-2916.

Hanchate, N.K., Kondoh, K., Lu, Z., Kuang, D., Ye, X., Qiu, X., Pachter, L., Trapnell, C., and Buck, L.B. (2015). Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. Science (New York, NY.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell *38*, 576-589.

Hewitt, S.L., Farmer, D., Marszalek, K., Cadera, E., Liang, H.E., Xu, Y., Schlissel, M.S., and Skok, J.A. (2008). Association between the Igk and Igh immunoglobulin loci mediated by the 3' Igk enhancer induces 'decontraction' of the Igh locus in pre-B cells. Nature immunology *9*, 396-404.

Hirota, J., and Mombaerts, P. (2004). 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.

Ibarra-Soria, X., Levitin, M.O., Saraiva, L.R., and Logan, D.W. (2014). The olfactory transcriptomes of mice. PLoS genetics *10*, e1004593.

Johnston, R.J., Jr., and Desplan, C. (2014). Interchromosomal communication coordinates intrinsically stochastic expression between alleles. Science (New York, NY *343*, 661-665.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384-388.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods *7*, 1009-1015.

Khan, M., Vaes, E., and Mombaerts, P. (2011). Regulation of the probability of mouse odorant receptor gene choice. Cell *147*, 907-921.

Kim, J.H., Lee, S.R., Li, L.H., Park, H.J., Park, J.H., Lee, K.Y., Kim, M.K., Shin, B.A., and Choi, S.Y. (2011). High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS ONE *6*, e18556.

Kong, N.R., Davis, M., Chai, L., Winoto, A., and Tjian, R. (2016). MEF2C and EBF1 Co-regulate B Cell-Specific Transcription. PLoS genetics *12*, e1005845.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lewcock, J.W., and 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*, 1069-1074.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O*., et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science (New York, NY *326*, 289-293.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J*., et al.* (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nature immunology *11*, 635-643.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403-413.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170-1187.

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

Lyons, D.B., Allen, W.E., Goh, T., Tsai, L., Barnea, G., and Lomvardas, S. (2013). An epigenetic trap stabilizes singular olfactory receptor expression. Cell *154*, 325-336.

Lyons, D.B., Magklara, A., Goh, T., Sampath, S.C., Schaefer, A., Schotta, G., and Lomvardas, S. (2014). Heterochromatin-mediated gene silencing facilitates the diversification of olfactory neurons. Cell reports *9*, 884-892.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R.*, et al.* (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nature neuroscience *13*, 133-140.

Magklara, A., Yen, A., Colquitt, B.M., Clowney, E.J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z.A., Kheradpour, P., Mountoufaris, G., Carey, C.*, et al.* (2011). An epigenetic signature for monoallelic olfactory receptor expression. Cell *145*, 555-570.

Mandel, E.M., and Grosschedl, R. (2010). Transcription control of early B cell differentiation. Curr Opin Immunol *22*, 161-167.

Mangale, V.S., Hirokawa, K.E., Satyaki, P.R., Gokulchandran, N., Chikbire, S., Subramanian, L., Shetty, A.S., Martynoga, B., Paul, J., Mai, M.V.*, et al.* (2008). Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. Science (New York, NY *319*, 304-309.

Markenscoff-Papadimitriou, E., Allen, W.E., Colquitt, B.M., Goh, T., Murphy, K.K., Monahan, K., Mosley, C.P., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. Cell *159*, 543-557.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupe, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. Cell *145*, 447-458.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol *28*, 495-501.

Merika, M., Williams, A.J., Chen, G., Collins, T., and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. Molecular cell *1*, 277-287.

Michaloski, J.S., Galante, P.A., and Malnic, B. (2006). Identification of potential regulatory motifs in odorant receptor genes by analysis of promoter sequences. Genome research *16*, 1091-1098.

Monahan, K., and Lomvardas, S. (2015). Monoallelic expression of olfactory receptors. Annu Rev Cell Dev Biol *31*, 721-740.

Nagano, T., Varnai, C., Schoenfelder, S., Javierre, B.M., Wingett, S.W., and Fraser, P. (2015). Comparison of Hi-C results using in-solution versus in-nucleus ligation. Genome Biol *16*, 175.

Niimura, Y., and Nei, M. (2007). Extensive gains and losses of olfactory receptor genes in mammalian evolution. PLoS ONE *2*, e708.

Nishizumi, H., Kumasaka, K., Inoue, N., Nakashima, A., and 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*, 20067-20072.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. Cell *129*, 1111-1123.

Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., and Aggarwal, A.K. (1999). Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature *397*, 714-719.

Plessy, C., Pascarella, G., Bertin, N., Akalin, A., Carrieri, C., Vassalli, A., Lazarevic, D., Severin, J., Vlachouli, C., Simone, R.*, et al.* (2012). Promoter architecture of mouse olfactory receptor genes. Genome research *22*, 486-497.

Ptashne, M. (2009). Binding reactions: epigenetic switches, signal transduction and cancer. Curr Biol *19*, R234-241.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic acids research *44*, W160-165.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S.*, et al.* (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665-1680.

Roberson, M.S., Meermann, S., Morasso, M.I., Mulvaney-Musa, J.M., and Zhang, T. (2001). A role for the homeobox protein Distal-less 3 in the activation of the glycoprotein hormone alpha subunit gene in choriocarcinoma cells. The Journal of biological chemistry *276*, 10016-10024.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., *et al.* (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389-393.

Rothman, A., Feinstein, P., Hirota, J., and Mombaerts, P. (2005). The promoter of the mouse odorant receptor gene M71. Molecular and cellular neurosciences *28*, 535-546.

Saraiva, L.R., Ibarra-Soria, X., Khan, M., Omura, M., Scialdone, A., Mombaerts, P., Marioni, J.C., and Logan, D.W. (2015). Hierarchical deconstruction of mouse olfactory sensory neurons: from whole mucosa to single-cell RNA-seq. Scientific reports *5*, 18178.

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., and Sakano, H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. Science (New York, NY *302*, 2088-2094.

Serizawa, S., Miyamichi, K., and Sakano, H. (2005). Negative feedback regulation ensures the one neuron-one receptor rule in the mouse olfactory system. Chemical senses *30 Suppl 1*, i99-100.

Shykind, B.M., Rohani, S.C., O'Donnell, S., Nemes, A., Mendelsohn, M., Sun, Y., Axel, R., and Barnea, G. (2004). Gene switching and the stability of odorant receptor gene choice. Cell *117*, 801-815.

Spilianakis, C.G., Lalioti, M.D., Town, T., Lee, G.R., and Flavell, R.A. (2005). Interchromosomal associations between alternatively expressed loci. Nature *435*, 637-645.

Tan, L., Li, Q., and Xie, X.S. (2015). Olfactory sensory neurons transiently express multiple olfactory receptors during development. Molecular systems biology *11*, 844.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091-1100.

Travis, A., Hagman, J., Hwang, L., and Grosschedl, R. (1993). Purification of early-B-cell factor and characterization of its DNA-binding specificity. Molecular and cellular biology *13*, 3392-3400.

Treiber, N., Treiber, T., Zocher, G., and Grosschedl, R. (2010a). Structure of an Ebf1:DNA complex reveals unusual DNA recognition and structural homology with Rel proteins. Genes & development *24*, 2270-2275.

Treiber, T., Mandel, E.M., Pott, S., Gyory, I., Firner, S., Liu, E.T., and Grosschedl, R. (2010b). Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin. Immunity *32*, 714-725.

Vassalli, A., Rothman, A., Feinstein, P., Zapotocky, M., and Mombaerts, P. (2002). Minigenes impart odorant receptor-specific axon guidance in the olfactory bulb. Neuron *35*, 681-696.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184-2185.

Wang, M.M., and Reed, R.R. (1993). Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. Nature *364*, 121-126.

Wang, M.M., Tsai, R.Y., Schrader, K.A., and Reed, R.R. (1993). Genes encoding components of the olfactory signal transduction cascade contain a DNA binding site that may direct neuronal expression. Molecular and cellular biology *13*, 5805-5813.

Wang, S.S., Lewcock, J.W., Feinstein, P., Mombaerts, P., and Reed, R.R. (2004). Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. Development (Cambridge, England) *131*, 1377-1388.

Wang, S.S., Tsai, R.Y., and Reed, R.R. (1997). The characterization of the Olf-1/EBF-like HLH transcription factor family: implications in olfactory gene regulation and neuronal development. J Neurosci *17*, 4149-4158.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Xu, N., Tsai, C.L., and Lee, J.T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. Science (New York, NY *311*, 1149-1152.

Young, J.M., Luche, R.M., and 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*, 561.

Yu, C.R., Power, J., Barnea, G., O'Donnell, S., Brown, H.E., Osborne, J., Axel, R., and Gogos, J.A. (2004). Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. Neuron *42*, 553-566.

Zhang, G., Titlow, W.B., Biecker, S.M., Stromberg, A.J., and McClintock, T.S. (2016). Lhx2 Determines Odorant Receptor Expression Frequency in Mature Olfactory Sensory Neurons. eNeuro *3*.

FIGURE 1

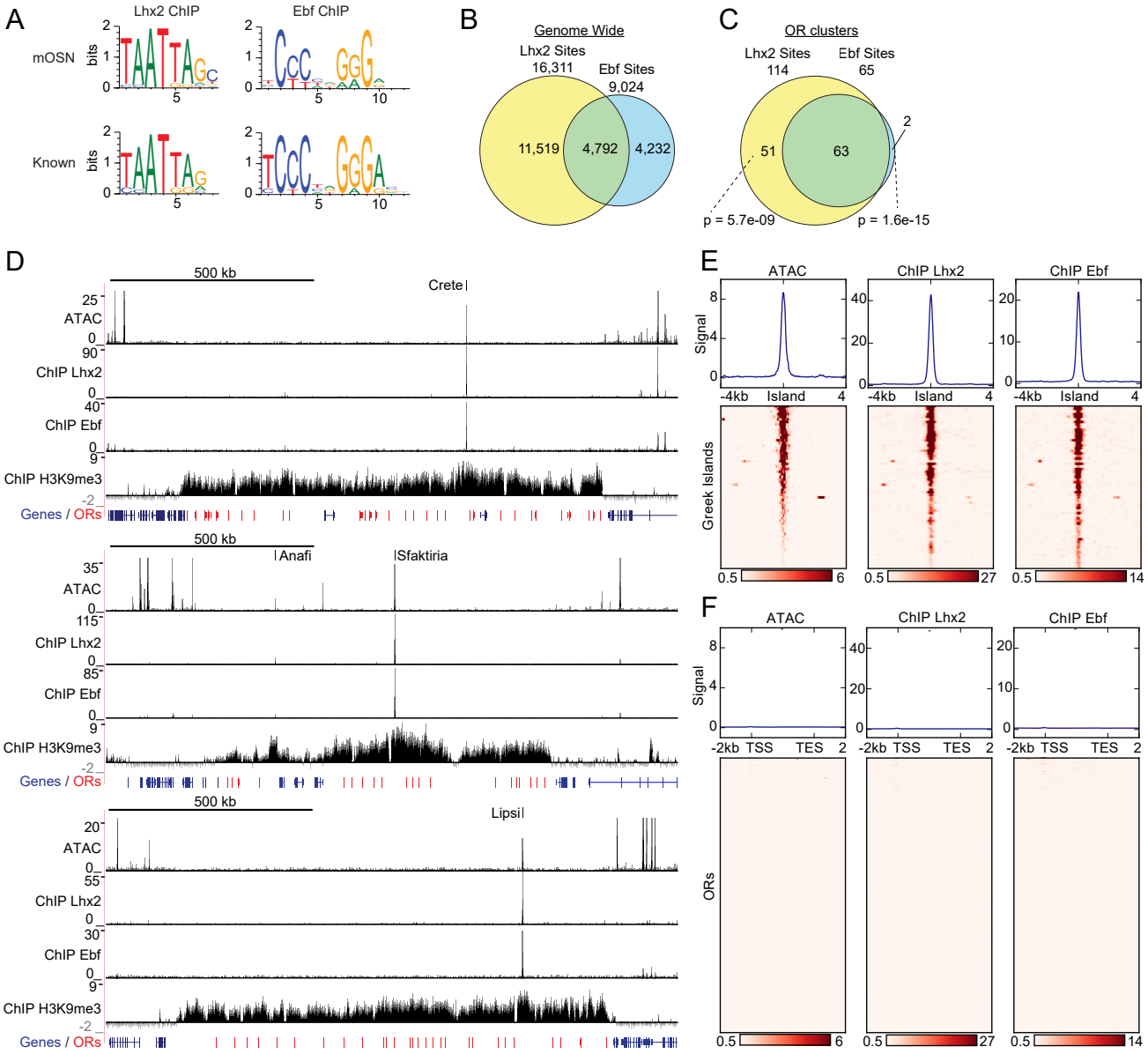


FIGURE 2

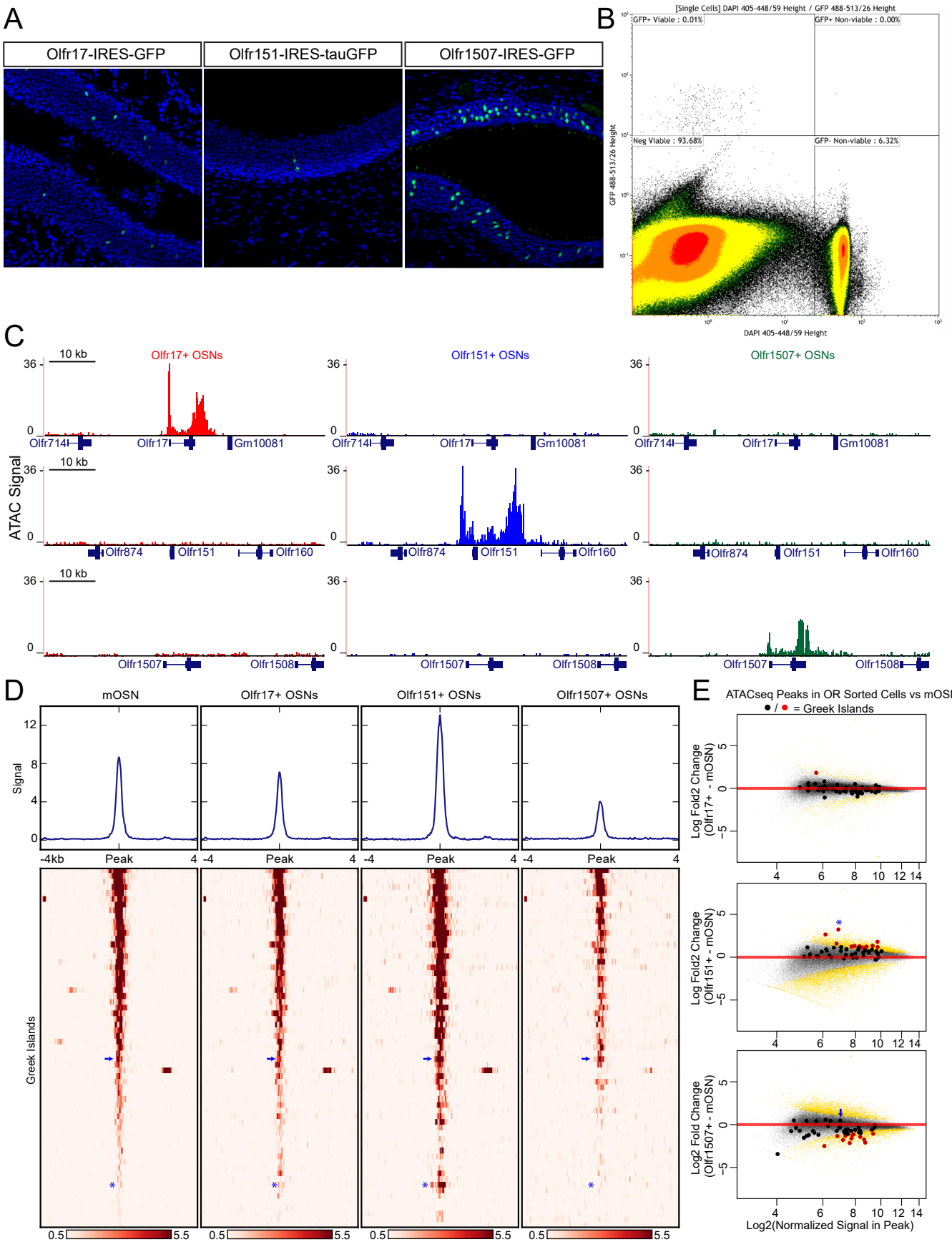


FIGURE 3

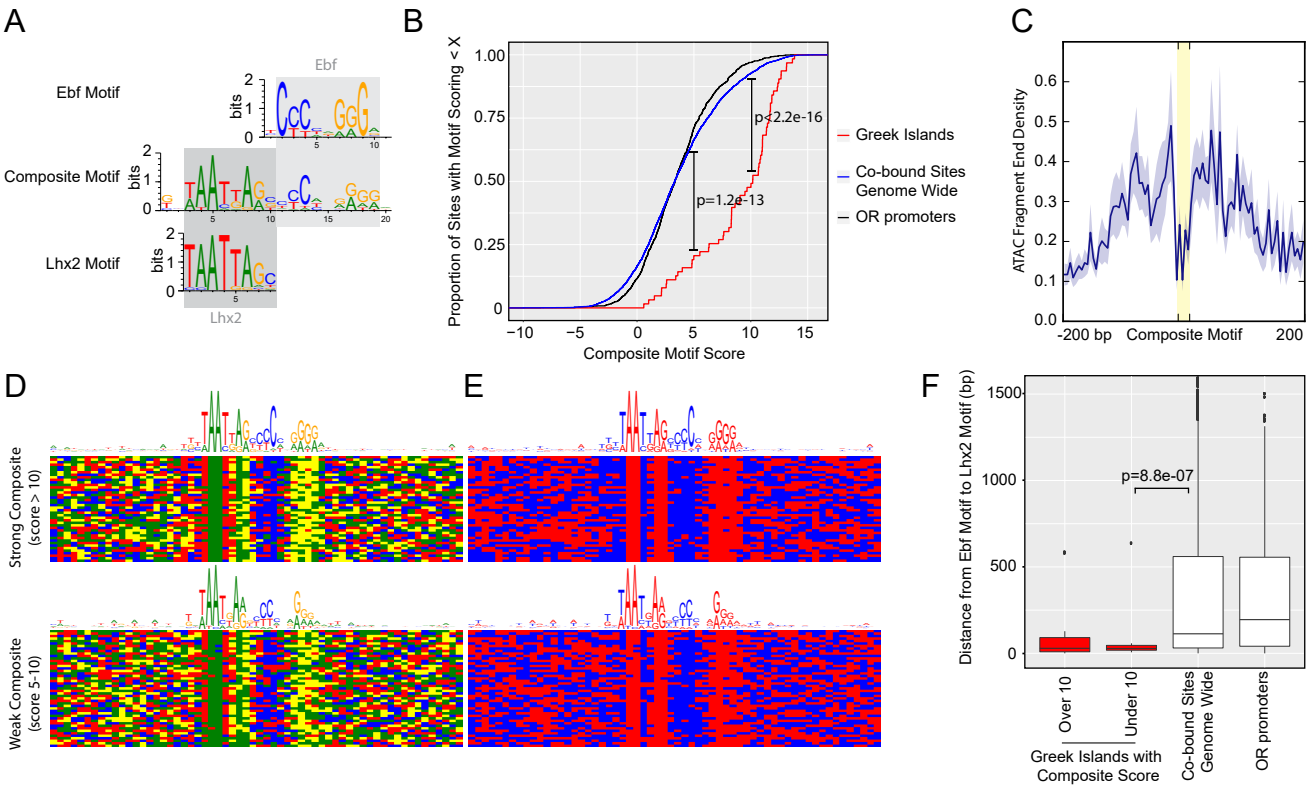


FIGURE 4

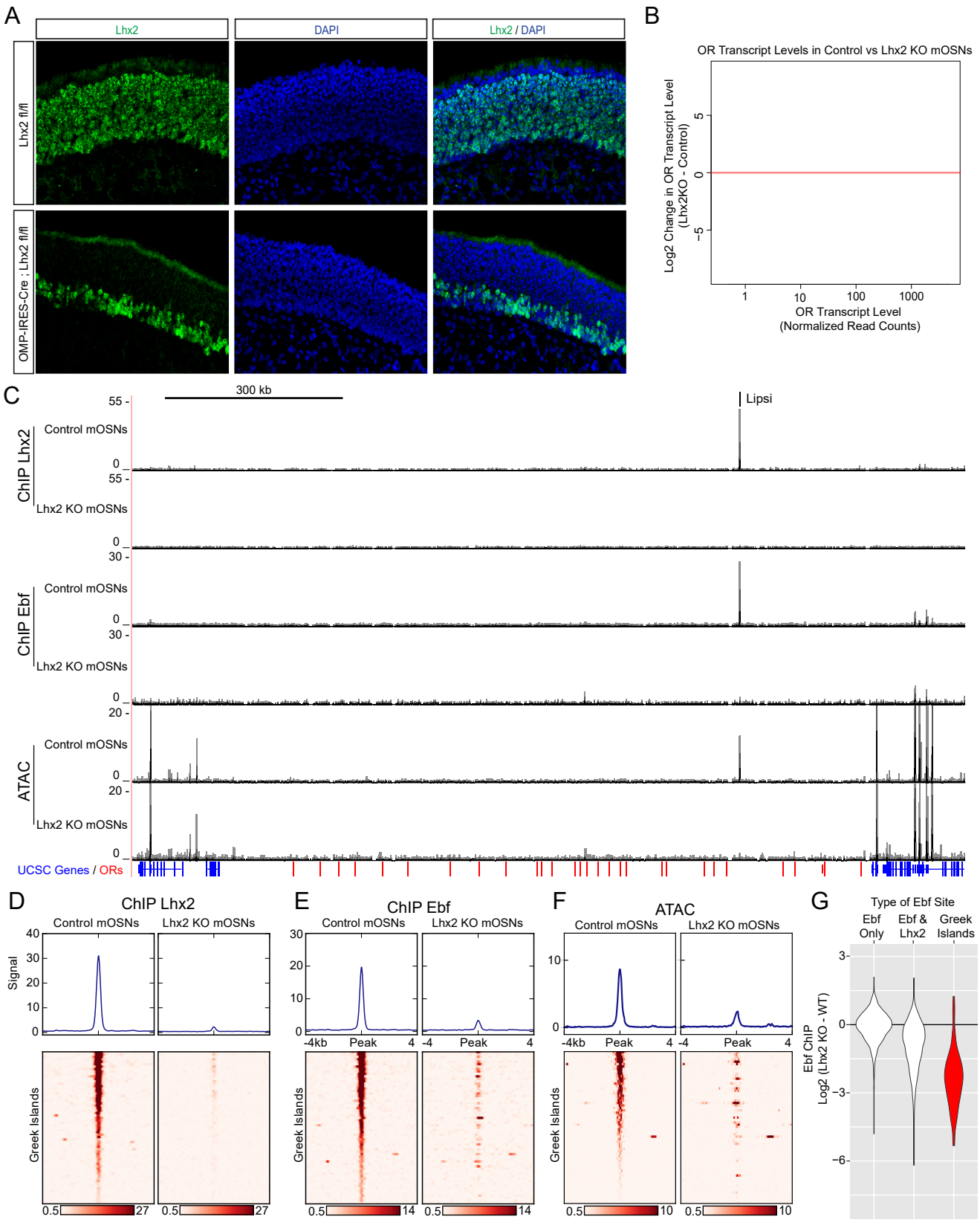


FIGURE 5

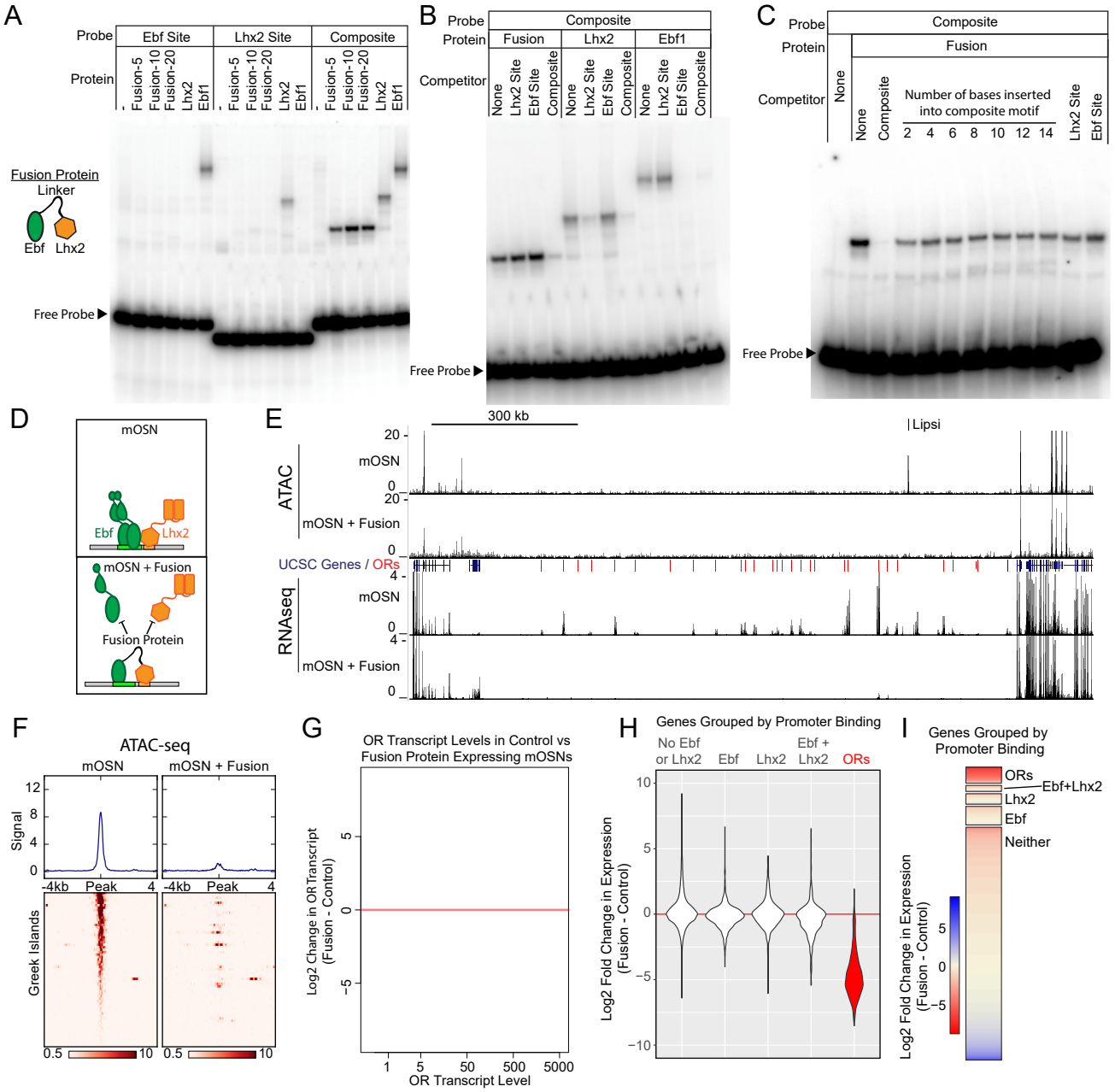


FIGURE 6

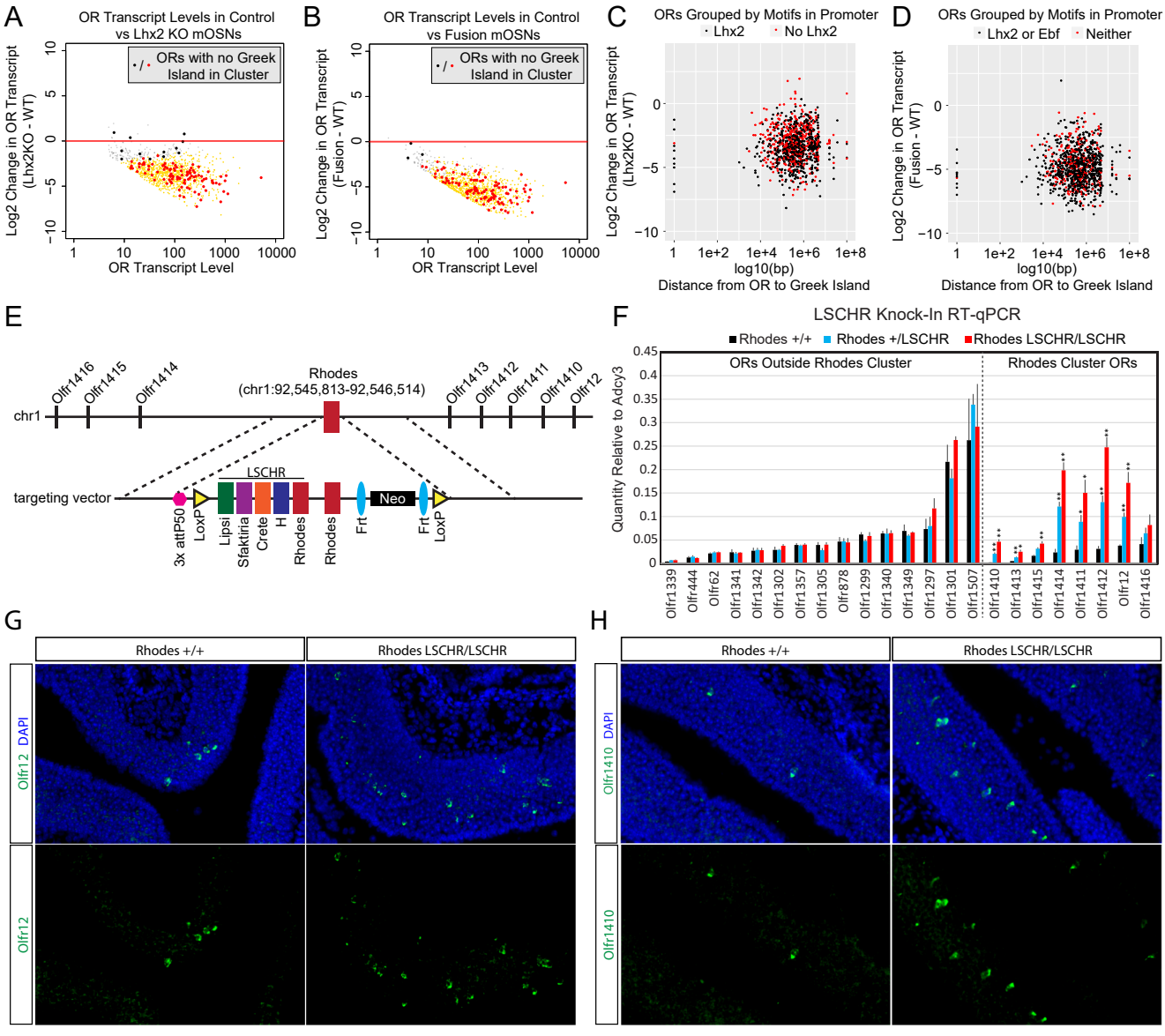


FIGURE 7

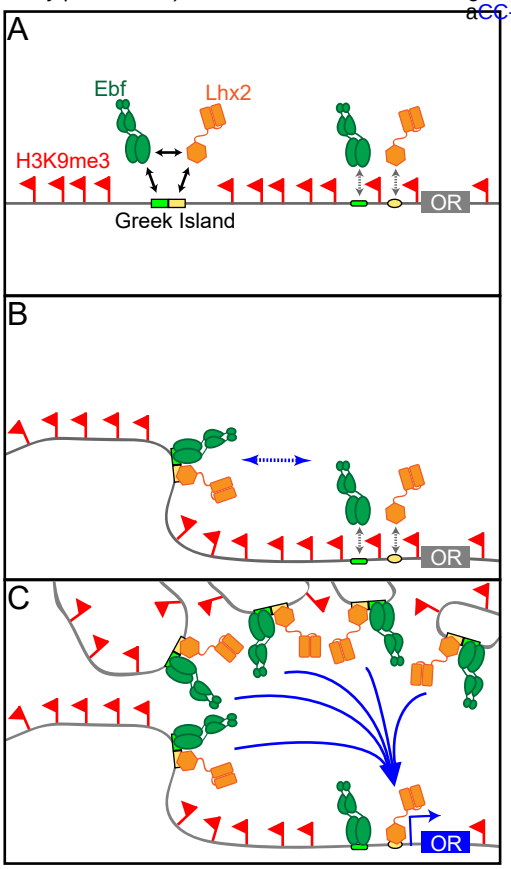


FIGURE S1

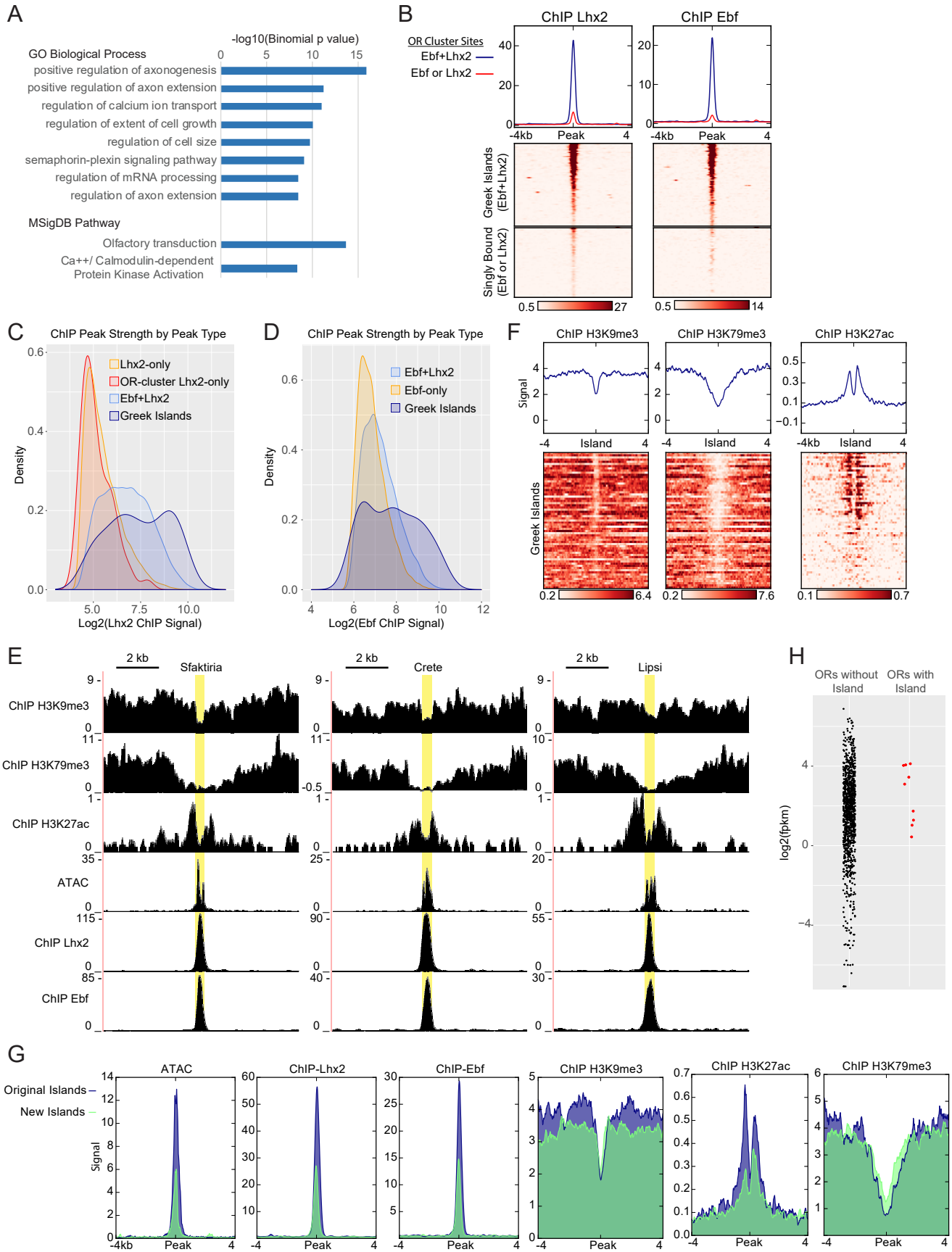


FIGURE S2

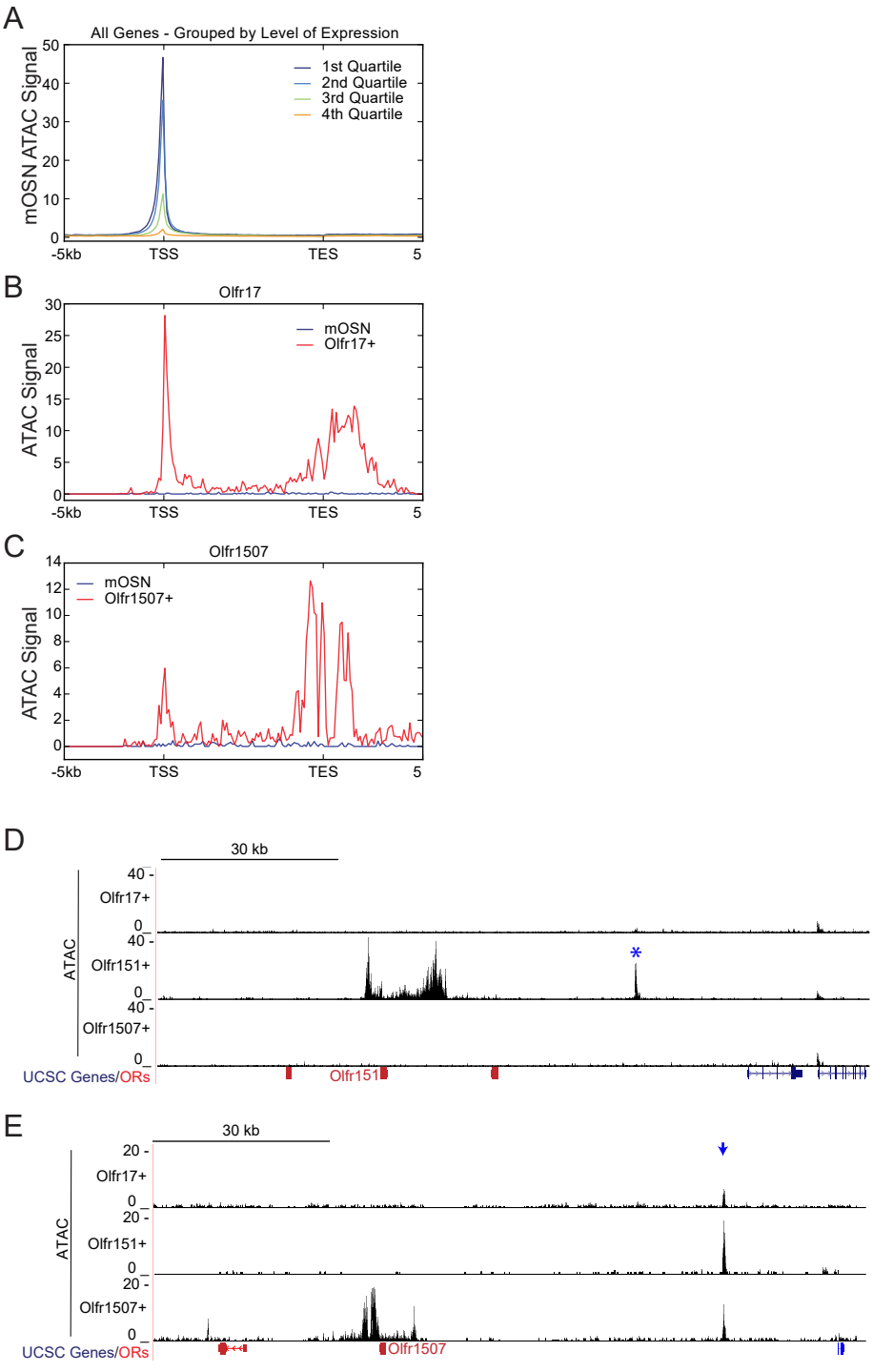


FIGURE S3

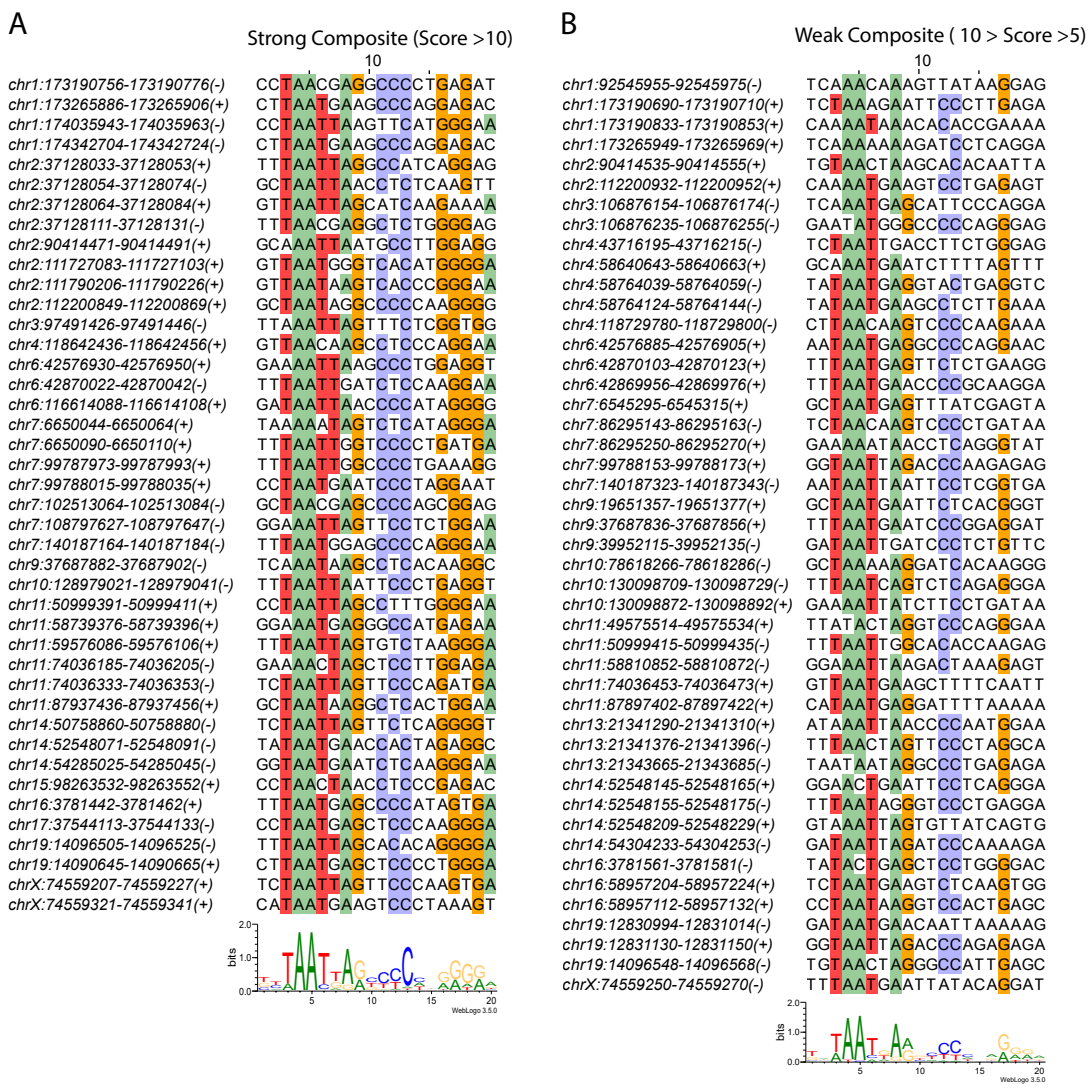
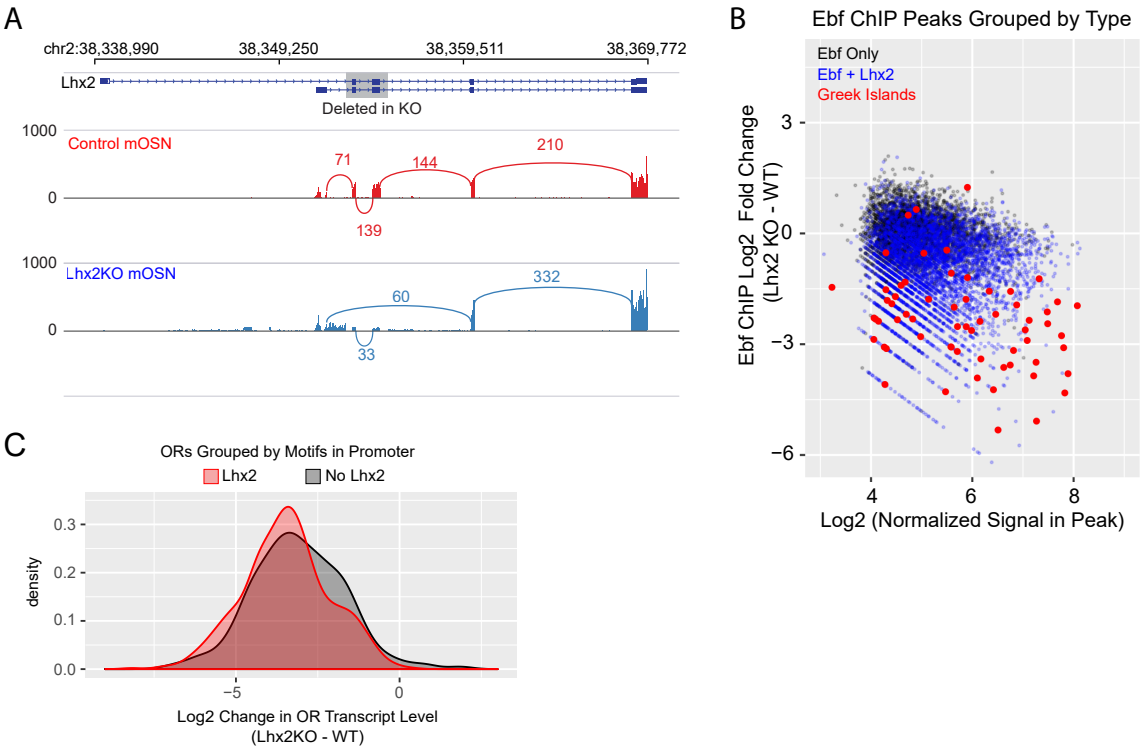
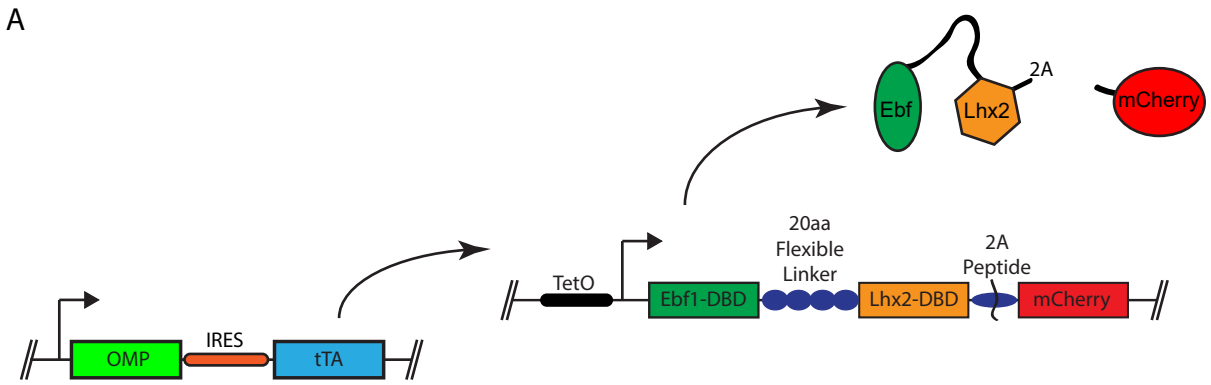
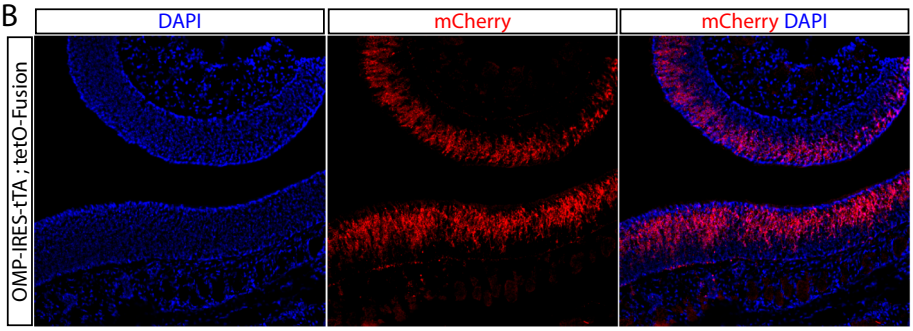


FIGURE S4







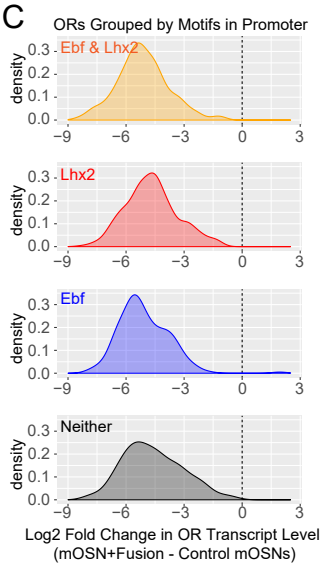


FIGURE S6