An Epigenetic Signature for Monoallelic Olfactory Receptor Expression

Angeliki Magklara[1], Bradley M. Colquitt[2$], E. Josephine Clowney[3$], Angela Yen[4,5], William Allen[6], Eirene-Markenscoff-Papadimitriou[2], Pouya Kheradpour[4,5], George Mountoufaris[1], Catriona Carey[1], Gilad Barnea[6], Manolis Kellis[4,5] and Stavros Lomvardas [1,2,3*]

[1]Department of Anatomy, University of California, San Francisco, CA 94158, USA.
[2]Program in Neurosciences, University of California, San Francisco, CA 94158, USA.
[3]Program in Biomedical Sciences, University of California, San Francisco, CA 94158, USA.
[4]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
[5]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[6]Department of Neuroscience, Brown University, Providence, RI 02912, USA.

*Corresponding Author. Email: stavros.lomvardas@ucsf.edu: T (415)514-4811; F (415)476-1974
$ These authors contributed equally to this work.

SUMMARY
Constitutive heterochromatin is traditionally viewed as the static form of heterochromatin that silences pericentromeric and telomeric repeats in a cell cycle and differentiation independent manner. Here, we show that in the mouse olfactory epithelium, olfactory receptor (OR) genes are marked, in a highly dynamic fashion, with the molecular landmarks of constitutive heterochromatin. The cell-type and differentiation dependent deposition of H3K9me3 and H4K20me3 along the OR clusters is, most likely, reversed during the process of OR choice. In contrast to the current view of OR choice, our data suggest that OR silencing takes place before OR expression, indicating that it is not the product of an OR-elicited feedback signal. This implies that chromatin mediated silencing provides the molecular prerequisite upon which a mechanism of singular and stochastic OR selection can be applied.

INTRODUCTION
In mammals, olfactory perception is accomplished by the detection of volatile chemicals in the olfactory epithelium and transmission of the odorant information to the brain, where it is processed. Unlike other sensory systems, olfaction relies on a large family of ~1000 OR genes that are expressed in a mutually exclusive and monoallelic fashion in olfactory sensory neurons (OSNs) (Buck and Axel, 1991; Chess et al., 1994). These neurons detect environmental chemicals, which are perceived as odors, and project their axons to the olfactory bulb, the first relay station of the olfactory circuit. OSNs that express the same receptor converge to the same neuropil structure called glomerulus (Mombaerts et al., 1996; Ressler et al., 1994; Vassar et al., 1994). It is well established that ORs participate both in odor detection and in guiding the axons to the proper glomeruli, ascribing this way the functional identity of each neuron (Barnea et al., 2004; Feinstein et al., 2004; Serizawa et al., 2006; Wang et al., 1998). The dual role of ORs in the wiring and physiology of the olfactory system emphasizes the importance of their proper expression. Each neuron faces the challenging task of expressing one OR allele at high levels, while keeping the rest of the repertoire completely silent. The effective repression of the non-chosen alleles is crucial for this system; due to the exceptionally high number of gene family members, even the lowest levels of basal transcription from the non-chosen ORs would result into thousands of inappropriately expressed OR molecules. Although, each individual receptor would have insignificant representation, altogether they would generate OR activity comparable to the one from the chosen allele, resulting, most likely, to wiring perturbations and subsequent sensory confusion.

In the mouse, OR genes are expressed in the main olfactory epithelium (MOE) in a spatial and temporal fashion that is almost certainly organized by positional cues (Ressler et al., 1993; Rodriguez-Gil et al., 2010; Vassar et al., 1993). Within a zone of expression, however, each neuron expresses only one out of several hundred alleles that have the potential to be transcribed in that particular region, and OR choice is seemingly stochastic (Fuss and Ray, 2009; Shykind, 2005). The molecular mechanisms regulating the “one receptor per neuron” rule remain elusive. Genetic experiments suggest that the production of OR protein elicits a feedback signal that stabilizes the expression of the encoding allele and prevents the activation of additional

OR genes (Lewcock and Reed, 2004; Serizawa et al., 2003; Shykind et al., 2004). Moreover, there is evidence that the OR coding sequence represses heterologous promoters suggesting that it contains important information for the regulation of OR expression (Merriam and Chess, 2007; Nguyen et al., 2007). Regulatory information is also included in proximal promoter sequences of OR genes as well as in distant enhancer elements (Rothman et al., 2005; Serizawa et al., 2003). One of them, the H enhancer, interacts with active OR alleles in *cis* or *trans,* which led us to propose that this interaction might be instructive for OR expression (Lomvardas et al., 2006). However, genetic ablation of H disrupts the expression of only three proximal ORs disputing a model in which H is a singular enhancer that orchestrates OR choice (Fuss et al., 2007; Nishizumi et al., 2007). Thus, despite immense efforts the molecular mechanisms regulating the activation of one OR allele and the stable transcriptional repression of the rest remain unknown.

Chromatin mediated silencing constitutes an effective form of transcriptional repression. There are distinct forms of transcriptionally inactive chromatin known as heterochromatin. Recent high-throughput approaches begin to reveal the regulatory principles of the heterochromatic genome. Facultative heterochromatin, a term assigned to the chromatin structure of silenced genes, is generally hypoacetylated and has di- and tri- methyl groups on lysine 27 and/or dimethyl groups on lysine 9 of histone H3 (Trojer and Reinberg, 2007). This type of heterochromatin is dynamic and appears to be developmentally regulated (Bernstein et al., 2006; Ezhkova et al., 2009). In contrast, constitutive heterochromatin, which is mostly found on pericentromeric and telomeric repeats, intergenic regions and LINE elements, remains condensed during the cell cycle and stable during differentiation. Moreover, it has distinct molecular features, such as the trimethylation of lysine 9 and lysine 20 of histones H3 and H4 respectively (Fodor et al., 2010); (Schotta et al., 2004).

Here, we test the hypothesis that chromatin mediated silencing prevents the inappropriate expression of multiple OR genes in each sensory neuron (McClintock, 2010). Our data show that in the olfactory epithelium OR genes are subject to an unusual form of heterochromatic silencing that combines characteristics of both constitutive and facultative heterochromatin. Our ChIP-on-chip experiments detect two histone modifications, H3K9me3 and H4K20me3, on OR loci. Although these modifications are characteristic of constitutive heterochromatin, the dynamic nature by which they mark OR genes is consistent with properties of facultative heterochromatin. The cell-type and differentiation dependent deposition of the two trimethyl-marks on OR clusters forms highly compacted and inaccessible heterochromatic macrodomains. Surprisingly, heterochromatic silencing of OR clusters occurs before OR transcription and does not require OR expression, suggesting that it is not the product of an OR-elicited feedback signal. The enrichment for these silent marks is significantly reduced on an active OR allele, which is marked instead with H3K4me3, a modification found on active or poised promoters (Guenther et al., 2007), revealing the “epigenetic” underpinnings of non-deterministic monoallelic gene expression. Finally, we show that insertion of the OMP-promoter transgene within OR heterochromatin results to an OR-like, sporadic and zonal expression of this transgene, indicating that stochastic escape from heterochromatic silencing could be the basis of monogenic and monoallelic gene expression.

RESULTS

Whole genome analysis of H3K9me3 and H4K20me3 in the MOE

To explore the possibility that OR genes are repressed via heterochromatic silencing, we performed chromatin immunoprecipitation (ChIP) experiments with antibodies against all methylated forms of H3K9, H3K27 and H4K20 using native chromatin preparations from the MOE and liver. Preliminary screening experiments showed that the tested OR sequences were enriched for H3K9me2, H3K9me3 and H4K20me3 in the MOE (data not shown), whereas, in the liver, significant enrichment was obtained only for H3K9me2, as previously described (Wen et al., 2009). In contrast, the ORs tested were not enriched for H3K27me3 in any cell type examined. Further investigation, using preparations from hippocampus and mouse embryonic fibroblasts, failed to detect H3K9me3 or H4K20me3 on ORs, suggesting that these marks are deposited on these genes in a tissue-dependent manner (data not shown).

To examine the genome-wide distribution of the two heterochromatic marks on ORs, we performed ChIP-on-chip experiments, using Nimblegen whole-genome tiling arrays, with immunoprecipitated DNA from MOE and liver (see Experimental Procedures). We found that most ORs are hypermethylated on H3K9 and H4K20 in the

MOE but not in the liver. The distribution of the two modifications along chromosomes 2, 7 and 9, which contain large OR genomic clusters, is depicted as heatmaps in Figure 1A, where genes are ordered by chromosomal position. Most genes, independently of their transcription status (Sammeta et al., 2007), appear to be devoid of both modifications in both tissues (green in heatmaps). However, in the MOE, there is significant enrichment for H3K9me3 and H4K20me3 on OR genes. The high enrichment signals for the two modifications, in combination with the tandem chromosomal organization of these genes, highlights the position of each OR cluster in these three chromosomes (Figure 1A). Vomeronasal receptor (VR) genes, which encode chemoreceptors that detect pheromones, are enriched for H3K9me3 and H4K20me3 in the MOE (Figure 1A Chromosome 7). In contrast to the high enrichment values in the MOE, ORs and VRs are hypomethylated in the liver, in agreement published ChIP-on-chip experiments that do not detect significant enrichment for H3K9me3 or H4K20me3 on OR genes in several different cell lines (Celniker et al., 2009).

A Heterochromatic Signature for Chemoreceptors

To obtain a global view of our ChIP-on-chip data, we ranked the mouse genes based on the average signal intensity of H3K9me3 and H4K20me3 over a region of 2Kb (Figure 1B,C). Since most OR promoters are not annotated, we used the translational start site (TSS) as a coordinate for the alignment of the mouse genes. Also, to present these data in a visually comprehensive manner we chose to include in the presentation of these heatmaps only every 15^{th} mouse gene, although the analysis was performed for all mouse genes (see Extended Experimental Procedures). In Figure 1B, 1,000 randomly selected genes are ranked in descending order by their enrichment values for the two modifications. OR genes, which are depicted by blue lines at the side of the heatmap, are clustered on the very top, indicating that they are the most enriched genes for H3K9me3 and H4K20me3 in the MOE. Indeed, as seen in Figure 1C, which provides a zoomed-in view of the top 1,000 genes in this ranking system, OR genes constitute the majority of genes with significant enrichment for both trimethyl-marks ($p<10^{-7}$). Sequential ChIPs with chromatin from the MOE confirmed the conclusion that these modifications are present on the same OR loci at the same time (Figure S1A).

Interestingly, most of the non-OR genes that are enriched for H3K9me3 and H4K20me3, represented by red lines in Figure 1B,C are chemoreceptors, mostly VRs and Formyl-Peptide receptors (FPRs) that are clustered in extremely AT-rich isochores (Figures S1B, C) . These genes are also expressed in a mutually exclusive fashion in the Vomeronasal organ (VNO), where they detect pheromones (VRs) or degrading bacterial proteins (FPRs) (Dulac and Axel, 1995; Liberles et al., 2009; Riviere et al., 2009). VRs, which are also expressed monoallelicaly, are subject to the same receptor-mediated feedback signaling in the VNO, consistent with the idea that similar molecular principles apply to the regulation of these gene families (Capello et al., 2009; Fleischmann et al., 2008; Rodriguez, 2007). Although very little is known for the regulatory mechanisms of FPRs, it is noteworthy that only the members of the FPR gene family that are expressed in the VNO are enriched for these repressive marks (Figure S1D). The three family members that are dedicated to immunological functions are H3K9me3 and H4K20me3 free. Therefore, H3K9me3 and H4K20me3 mark monogenically expressed chemoreceptors in a tissue dependent manner. Although VRs and FPRs are not expressed in the MOE, this tissue has several common molecular and physiological properties with the VNO, which may account for the common histone modifications between these chemoreceptor families. As seen in our ChIP-on-chip analysis though, and verified by ChIP-qPCR (data not shown), ORs are more highly enriched for H3K9me3 and H4K20me3 than VRs and FPRs in the MOE. The only exception from this trend comes from the type I OR genes that are all clustered on chromosome 7 (Figure 1A). These genes are evolutionarily older than the rest of the mammalian ORs (type II ORs) and have a dorsally restricted expression pattern in neurons that are believed to belong to a different developmental lineage than the neurons that express type II ORs (Bozza et al., 2009; Tsuboi et al., 2006). Type I ORs have lower enrichment levels of H3K9me3 and H4K20me3, comparable to the levels of VRs and FPRs. Future experiments will test whether the highest deposition of these repressive marks occurs in the neuronal lineage that is committed to the expression of a specific type of chemoreceptors.

Another family of genes that appear highly enriched for H3K9me3 and H4K20me4 are the KRAB-ZFP (Figure S1D), a subgroup of the ZFP family of transcription factors. Unlike the aforementioned chemoreceptor genes, the heterochromatinization of the KRAB-ZFPs seems to be ubiquitous, as it is also observed in the liver (data not shown) as well as in other cell types examined (Barski et al., 2007; O'Geen et al., 2007). Interestingly, none of the genes that were recently reported as being subject to sexually dimorphic imprinting in the mouse

brain (Gregg et al., 2010a; Gregg et al., 2010b), were positive for the two trimethyl-marks in the MOE, suggesting that either imprinting is under different regulation if it occurs in this tissue or that this heterochromatin-mediated silencing occurs only in the tissue of monoallelism.

Defined Heterochromatic Macrodomains cover the OR clusters in the MOE

In order to identify significant sites of enrichment within OR clusters, we used the Ma2C algorithm that was developed for peak detection in ChIP-on-chip experiments with Nimblegen tiling arrays (Song et al., 2007). By using a sliding window of 0.5 Kb with FDR≤5%, we observed that, in the MOE, the peaks for the two histone modifications were contained in broadly enriched genomic regions spreading throughout the OR clusters in an almost continuous arrangement, rather than forming focal sites of enrichment (Figure 2A). Modification of the Ma2C parameters (see Extended Experimental Procedures), in such a way that local signal fluctuations would be averaged over a larger sliding window of 10Kb, confirmed that H3K9me3 and H4K20me3 form tissue-specific, heterochromatic macrodomains that, often, cover several megabases of clustered OR genes in the MOE (an example for OR cluster 2.2 is shown in Figure 2A). Using this analysis, we found that 1376 ORs fall in H4K20me3 blocks and 1109 ORs fall in H3K9me3 blocks, out of a total of 1441 annotated OR genes (see Figure S1A) ($p<10^{-7}$). As we show in the double ChIP experiment (Figure S1A), H3K9me3 and H4K20me3 coexist in OR chromatin, therefore, the lower number of ORs embedded in H3K9me3 blocks reflects, most likely, qualitative differences between the antibodies used and not a biologically significant variation. In addition, ChIP-qPCR analysis for several ORs that appear positive only for H4K20me3 in the ChIP-on-chip analysis revealed enrichment also for H3K9me3 (data not shown). In contrast to our findings in the MOE, there are much fewer and smaller H3K9me3 and H4K20me3 peaks and blocks on the ORs in the liver (Figure 2A) that do not overlap; 22 ORs fall in H3K9me3 blocks and 25 different ORs in H4K20me3 blocks (out of 913 ORs, in total, located on chromosomes 1 to 9 that were examined in the liver). The presence of heterochromatic macrodomains over OR clusters was also confirmed by the application of an independent algorithm that was developed and used for the detection of long stretches of H3K9me2 in the liver, brain and ES cells (Figure S2) (Wen et al., 2009).

We extensively validated our ChIP-on-chip results by ChIP-qPCR for multiple OR gene clusters in both tissues. Representative data demonstrating the gene- and tissue- specific marking of OR genes are presented in Figure 2. In Figure 2A-C, randomly selected ORs (highlighted in black rectangles in Figure 2A) from cluster 2.2 were assayed and their MOE-specific enrichment was confirmed by ChIP-qPCR (Figures 2B-C). An interesting observation is that the external borders of the heterochromatic blocks, coating the OR clusters, coincide with the borders of the OR genomic loci. For example, the *Ptprj* (protein tyrosine phosphatase receptor j) gene, located at the border of cluster 2.2 (Figure 2A) is partially methylated for H3K9me3 and H4K20me3 only at the most OR-proximal sequences, as confirmed by ChIP-qPCR (Figure 2A-C). Similar observations were made for the borders of all OR clusters examined. The reported binding of CTCF outside of OR clusters (Kim et al., 2007), or other insulating elements (Dickson et al., 2010), might contribute to the confinement of OR heterochromatin within the OR clusters.

In few instances, the embedding of transcriptionally active non-OR genes (eg. *Mgat1* and *Mapk9*, highlighted by green rectangles in Figure 2D), in an OR cluster, interrupts the heterochromatin blocks, until the presence of another OR gene reconstitutes them. In contrast, in the same cluster, the *Btnl9* (butyrophilin-like protein 9) and *Flt4* (FMS-like tyrosine kinase 4) genes (highlighted by red rectangles in Figures 2D-E), which are not transcribed in the MOE, appear to be partially covered by heterochromatin marks, suggesting that in the absence of a competing transcriptional state, or an insulating activity, the OR heterochromatin can extend to the body of non-OR genes that are inserted in an OR cluster. These results were also confirmed by ChIP-qPCR as shown in Figures 2F-G.

Characterization of the OR heterochromatin

The data presented above demonstrate that all OR clusters in the MOE fall in long-range silenced domains decorated with markers of constitutive heterochromatin (H3K9me3 and H4K20me3), whereas, in the liver, they are enriched, only, for H3K9me2 (our data and (Wen et al., 2009)). In order to determine whether there are functional differences stemming from these distinct OR chromatin modifications, we treated nuclei from olfactory epithelium and liver with DNase I and examined the accessibility of different DNA loci by qPCR

analysis of the recovered DNA. As seen in Figure 3A, there is significant difference in DNase I sensitivity of OR loci between the two tissues. In MOE, the OR genes get hardly digested, even after an hour of incubation with DNase I, suggesting that they possess a chromatin structure that renders them inaccessible to the enzyme. Notably, there is a significant increase of the OR qPCR signal at the first timepoints of DNase treatment. This is probably due to the fact that OR chromatin is less soluble when it is not digested, which results to significant loss of OR chromatin fragments during extraction and purification. In sharp contrast to the prolonged DNase I resistance of ORs, non-OR genes are rapidly digested independently of their transcription status in the MOE. On the other hand, in the liver, the OR loci are indistinguishable from the other genes tested regarding their DNase I digestion pattern. These data argue for a less accessible chromatin structure for OR genes in the MOE, which correlates with the tissue dependent addition of H3K9me3 and H4K20me3 to the ubiquitously present H3K9me2. Thus, there is a regulatory system in place that most likely renders OR genes inaccessible to the transcription machinery in the cell type that has the potential to express these genes.

To test directly the idea that OR heterochromatin is insensitive to DNase I digestion due to a more compacted chromatin arrangement in the olfactory epithelium, we took advantage of the fact that chromatin condensation affects the buoyancy of the chromatin fiber (Ghirlando et al., 2004; Gilbert et al., 2004). We performed limited MNase digestion of MOE and liver chromatin followed by ultracentrifugation in sucrose gradient (4-60%) (Figure 3B). It is well established, that unlike DNase I, MNase digests native chromatin without being affected by the compaction levels or the transcriptional state of each locus (Weintraub and Groudine, 1976). Thus, digestion with MNase produces chromatin fragments of similar length distribution in the two tissues (input lanes in Figure 3C). Fractions were collected, from top to bottom, and upon deproteination, they were analyzed by gel electrophoresis and southern blot. As a probe we used a ~600bp RT-PCR product, generated by a degenerate primer pair, which contains several hundred OR sequences (Buck and Axel, 1991). As seen in Figure 3C the distribution of OR fragments is dramatically different across fractions from the two tissues, despite the fact that the chromatin was equally digested. In the liver preparation the strongest OR signal appears in the 2nd and 3rd fraction. In contrast, in the MOE, there is significant amount of OR DNA molecules in the first fraction, but also a substantial signal in the bottom 5 fractions. Importantly, the signal in these fractions, which correspond to the highest sucrose concentration, originates from OR DNA molecules of lower molecular weight (pointed by black arrows), a migration pattern consistent with highly compacted chromatin. Selected fractions from the same preparation analyzed by a probe specific for ribosomal genes, which represent the only other multigenic family of similar size with the ORs, did not reveal similar differences between the two tissues or such lower molecular weight fragments in the bottom fractions (Figure 3D).

In summary, these experiments provide evidence that the addition of H3K9me3 and H4K20me3 to OR loci, in the MOE, coincides with the establishment of macrodomains that have characteristics of heterochromatin: decreased DNase I accessibility, and reduced buoyancy in sucrose gradients. However, unlike other forms of heterochromatin that is marked with H3K9me3 and H4K20me3, OR heterochromatin is tissue dependent and is not restricted to intergenic regions. In contrast, it is spread throughout the largest family of mammalian genes.

OR silencing is independent of OR expression

The MOE is a heterogeneous tissue comprised of mature and immature olfactory sensory neurons (OSNs), stem-like progenitor cells (Horizontal Basal Cells, HBCs and Globose Basal Cells, GBCs), and supporting epithelial cells called sustentacular (Leung et al., 2007; Duggan and Ngai, 2007) (Figure S3A). Although OSNs comprise the majority of cells in our dissections, we sought to confirm that our findings describe the chromatin state of OR genes in the OR-expressing neurons. For this reason, we performed a series of fluorescence-activated cell sorting (FACS) experiments followed by ChIP-qPCR analysis.

We isolated mature OSNs from OMP-IRES-GFP mice (Figure 4A) and, as seen in Figure 4B**,** the OR genes tested have high levels of enrichment for both H3K9me3 and H4K20me3 in this cell population. Although these are the neurons that express ORs, each OR gene is expressed in 0.1% of the cells, therefore it is not surprising that we can detect these silencing marks on them. The finding that OR genes have heterochromatic modifications in the cells that can express them is counterintuitive but not unexpected; this silencing system probably evolved to prevent multigenic OR transcription in the cellular microenvironment that has the potential to express ORs. Moreover, the confirmation that ORs are heterochromatinized in OSNs

raises the question whether this silencing is induced by OR expression as a consequence of the previously reported feedback signal.

To address this question, we examined the chromatin modification status of ORs in progenitor cells, starting with the most multipotent cells of the MOE, the HBCs, (Heintz, 2004; Leung et al., 2007) by sorting ICAM-1 positive cells (Carter et al., 2004) (Figure 4C). As seen in Figure 4D, there is no enrichment for H3K9m3 and H4K20me3 on OR genes, whereas the pericentromeric repeats (major satellite repeats) and the KRAB-ZFP genes (Zfp560) are hypermethylated even in this less differentiated stage. This result is consistent with the idea that distinct tissue- and gene-specific mechanisms are employed for the heterochromatinization of ORs. Interestingly, though, we detect high signal for H3K9me2 on ORs in the HBCs (Figure S3B) suggesting that even in this multipotent state OR genes are already repressed. It is possible that H3K9me2 constitutes a sufficient layer of repression in cells that do not have the machinery responsible for OR transcription.

Although this analysis demonstrates that chromatin-mediated repression of OR genes occurs before OR expression, the observation that ORs are marked with H3K9me3 and H4K20me3 in mature OSNs but not HBCs, poses the question of when this additional lysine methylation happens. To examine this, we sorted an additional population from the MOE that does not express OR genes, the sustentacular cells. Sustentacular cells line the apical surface of the epithelium (Figure 4E), have common developmental origin with the olfactory neurons but perform non-neuronal functions and they do not express OR genes (Figure 4H). These cells were sorted with the use of SUS4 antibody (Chen et al., 2004). As seen in Figure 4F, in sustentacular cells, the levels of H3K9me3 and H4K20me3 on ORs are also comparable to the levels of these modifications in the mature OSNs. This result is consistent with the marking of OR genes with H3K9me3 and H4K20me3 in the absence of OR expression. In this scenario however, it is possible that trimethylation of lysine 9 and lysine 20 occurs before the selection of one OR allele for activation.

To test this, we examined the chromatin state of OR genes in populations from the MOE that do not express OMP, ICAM-1 or iLR (immature laminin receptor) (another marker for multipotent basal cells) (Jang et al., 2007) and they are also negative for SUS4. This quadruple negative population consists mostly of progenitor cells and immature sensory neurons that do not express ORs at levels that are detectable by RT-PCR (Figure 4H). As seen in Figure 4G, in this population the enrichment levels of H3K9me3 and H4K20me3 on the tested ORs are as high as in the OMP positive cells. This observation corroborates the hypothesis that trimethylation of ORs occurs in the absence of an OR- elicited feedback signal and it is consistent with a model of OR silencing preceding OR choice.

To perform ChIP analysis in a more specific immature cell population from the MOE, we obtained a Neurogenin1-IRES-GFP BAC transgenic mouse from GENSAT (Heintz, 2004). As seen in Figure 5A, sections of the olfactory epithelium of this strain show GFP-labeled cells in the basal layer of the epithelium, consistent with the reported expression of Neurogenin-1 in globose basal cells (Cau et al., 2002). FACS sorting and native ChIP-qPCR analysis of the GFP-positive cells revealed high levels of enrichment for H3K9me3 and H3K20me3 on the tested ORs, suggesting a similar heterochromatic signature with the mature OSNs (Figure 5B). RT-PCR analysis on these cells showed that they represent a mixed population of progenitor cells and immature sensory neurons (Figure S3C). In order to obtain a global view of the levels of OR transcription in the Neurogenin1 positive cells, we performed a deep sequencing analysis (RNAseq) using cDNA from Neurogenin-1 and OMP positive cells (see Experimental Procedures). We detected transcripts for 1185 OR genes in the mature OSNs that have, on average, 8 fold higher mRNA levels than in the Neurogenin-1 positive cells (Figure 5C). Most likely, the extremely low levels of OR transcripts that are found in the Neurogenin-1 positive cells reflect the existence of a small percentage of contaminating mature OSNs that are detected due to the extreme sensitivity of RNAseq. In agreement with this, 25 genes that are expressed only in mature OSNs are also detected in low levels in the Neurogenin-1 sample, at a similar 7-fold average reduction (see Extended Experimental Procedures). Had only the few OR-expressing cells contributed the H3K9me3 and H4K20me3 signal on OR genes, the trimethylation signal would have also been 8-fold lower in the Neurogenin-1 positive cells. However, the enrichment levels of H3K9me3 and H4K20me3 on the tested OR genes in the Neurogenin-1 positive cells are identical to the levels of these marks in OMP positive cells. Therefore the ChIP-qPCR data from the quadruple negative cells and the Neurogenin-1 positive cells are consistent with H3K9me3 and H4K20me3 being deposited on OR genes before OR expression.

An “epigenetic” switch accompanies OR choice

If OR repression precedes OR choice, then, for an OR allele to be transcribed, these repressive modifications must be removed at a later stage. Alternatively, it is possible that these modifications are not incompatible with transcription and that an OR allele is activated by removal of the repressive proteins that these methyl- marks recruit, resulting in subsequent opening of the local chromatin architecture. To test whether the active OR allele is free of these repressive modifications we isolated, by FACS, neurons that express the olfactory receptor P2 (Olfr17) that is genetically tagged with GFP from P2-IRES-GFP knocked-in mice (Figure 6A). Preliminary experiments had shown that, in the MOE of these mice, the GFP sequence is also marked with H3K9me3 and H4K20me3, following the histone methylation pattern of the adjacent P2 gene (data not shown). Consequently, the GFP sequence can be used as a target for allele-specific qPCR, by allowing the distinction between the two P2 alleles in heterozygote P2-IRES-GFP mice (see Figure 6B). We isolated ~40,000 GFP-positive and GFP-negative neurons from P2-IRES-GFP heterozygote mice and performed ChIP-qPCR for H3K9me3 and H4K20me3. We used a primer pair specific for GFP (marked GFP) to monitor the active P2 allele, along with primers that cannot distinguish between the two alleles (p2CDS or p2prom) and primers specific for the wild type allele (p2WT) (Figure 6B). As seen in Figures 6C and 6D the enrichment for H3K9me3 and H4K20me3 is significantly reduced at the active OR allele, compared to the enrichment of the inactive allele or the enrichment of the same sequence in the GFP negative population. The fact that there still appears to be some H3K9me3 on the active allele is, probably, due to the high background of the assay caused by the limited number of cells and the particular antibody used (compare to the H4K20me3 results). This idea is supported by an independent experiment, which showed that the levels of this modification on the active P2 allele are equal to the levels on Omp, which serves as our negative control (Figure 6E). Olfr177, which is normally expressed in a different zone than P2, is also highly enriched for H3K9me3 and H4K20me3. These results indicate that OR silencing surpasses zonal information and more importantly reject the scenario that the repressive modifications described here are used only for the prevention of OR expression outside of the appropriate zone.

H3K4me3 histone marks on the active OR allele

To further explore the idea of a different chromatin state between two OR alleles in such a pure neuronal population, we performed the same ChIP experiment in P2-IRES-GFP sorted neurons using an antibody against H3K4me3. This modification is found on active or poised promoters(Guenther et al., 2007) and has a mutually exclusive distribution with H3K9me3 and H4K20me3 (Regha et al., 2007). In agreement with the proposed incompatibility of these methyl-marks, H3K4me3 cannot be detected on OR promoters using chromatin preparations from the whole MOE (data not shown), because each gene is active in only ~0.1% of the cells and silenced in the rest. Since we already established that the inactive P2 allele is enriched for the two heterochromatic marks, we performed our analysis on homozygote P2-IRES-GFP mice to increase the yield of GFP positive neurons. As seen in Figure 6F, in the GFP positive population, where one of the P2 alleles is active, there is high enrichment for H3K4me3 on the P2 promoter and CDS, whereas there is no enrichment for these sequences in the GFP negative population. This supports the findings presented in Figure 6C-E indicating that activation of the P2 allele correlates with the reduction of H3K9me3 and H4K20me3. Interestingly, although H3K4me3 is very abundant on the active P2 allele, throughout the transcribed body of the gene, it is missing from the neighboring P3 and P4 genes (Figures 6F and S4A-B), despite the high sequence similarity between these genes and their expression in the same zone. This finding, along with the results from the H3K9me3 and H4K20me3 ChIPs in the GFP positive cells, suggest that OR choice is an extremely precise process albeit being stochastic. Only the chosen allele is relieved of or protected from heterochromatic silencing, whereas the non-chosen allele and the almost identical neighboring genes remain repressed. Moreover, our observations support the hypothesis that monoallelic OR expression is reflected in distinct post-translational histone modifications between the active and inactive allele. By default, these differences are not directed by the primary DNA sequence of the two alleles, thus they constitute a *bona fide* “epigenetic” signature of monoallelic gene expression.

Heterochromatic Silencing correlates with OR-like expression

Our data suggest that heterochromatic silencing might be the basis for the monogenic OR expression. To test whether this heterochromatic structure can influence gene expression, we examined a transgenic mouse, whereby an OMP-promoter driven LacZ transgene had been inserted proximal to a singular OR gene (*Olfr459*)

on chromosome 6 (Pyrski et al., 2001). Unlike numerous OMP-LacZ or OMP-GFP independent transgenes that are expressed in the majority of olfactory neurons (Nguyen et al., 2007; Walters et al., 1996), this transgene manifests a sporadic and mostly zonal expression pattern reminiscent of that of the neighboring *Olfr459*, which led to the suggestion that it is under the transcriptional control of this OR locus (Pyrski et al., 2001).

We reasoned that this transgene had been inserted within the heterochromatic block flanking *Olfr459*. Unlike other non-OR genes located in OR clusters, and which, through selective evolutionary processes, developed insulating mechanisms that prevented heterochromatic silencing (for example *Mgat1*, Figures 2D-G), transgenes are vulnerable to the influence of the local chromatin architecture (Henikoff, 1990). Mapping the exact insertion site of this transgene revealed that it resides ~55kbs from *Olfr459* (Figure 7A)(see Experimental Procedures). ChIP-qPCR experiments showed that the insertion site is heterochromatinized in wild type mice and remained so after the transgenic insertion. (Figure 7B, 7C). Importantly, as shown in Figure 7C ChIP-qPCR analysis using chromatin prepared from transgenic mice confirms that this transgene has similar methyl- marks with the surrounding OR heterochromatin, unlike the endogenous OMP promoter, which is H3K9me3 and H4K20me3 free. Notably, these modifications are specific for the olfactory epithelium, as in the liver the OMP-LacZ transgene and *Olfr459* are H3K9me3 and H4K20me3 free (Figure 7C). This observation suggests that the chromatin state of this transgene is influenced by that of the endogenous locus, rather than being the product of a transgenesis-induced heterochromatic silencing (Hiragami-Hamada et al., 2009).

To examine whether the insertion of the OMP transgene within OR-heterochromatin also results to monoallelic expression we compared the number of beta galactosidase positive cells between homozygous and heterozygous transgenic mice. As seen in Figure 7D, OMP-LacZ homozygotes have in average 1.8 fold more beta galactosidase positive cells, consistent with a monoallelic expression pattern.

DISCUSSION

In the mouse, olfactory receptor choice is a seemingly stochastic regulatory process that culminates in the expression of one out of the ~2800 annotated OR alleles. The molecular mechanisms that regulate the activation of one allele and prevent the expression of the rest remain, largely, unknown. Here, we show that the landmarks of constitutive heterochromatin, H3K9me3 and H4K20me3, are deposited on OR loci in a tissue- and differentiation-dependent manner. The extensive marking of OR genes by these methyl-groups results in the generation of compacted heterochromatic macrodomains that are likely incompatible with transcription. Although our experiments cannot yet determine whether these methyl- marks cause OR silencing or they are the consequence of it, they most certainly depict a “footprint” of regulatory events that result in the silencing of the non-chosen ORs and allow the expression of a single allele. The fact that the insertion of an OMP-LacZ transgene within OR heterochromatin results in an OR-like, sporadic and likely monoallelic expression pattern is consistent with an instructive role of this chromatin structure in gene expression.

The whole genome ChIP-on-chip analysis presented here provides a unique example for the involvement of H3K9 and H4K20 trimethylation in transcriptional choices in the mouse. Although it is well established that the methyl-marks we studied are markers of transcriptional repression, they are predominantly found on silenced pericentromeric and telomeric repeats or LINE elements, rather than being implicated in precise developmental and transcriptional decisions, such as OR choice (Fodor et al., 2010; Pauler et al., 2009; Peters et al., 2001). Consistent with a role in OR choice, H3K9me3 and H4K20me3 mark the OR genes in a differentiation regulated fashion. They are added on the -already repressed via H3K9me2- OR chromatin during the transition from Horizontal basal cells to immature sensory neurons. Subsequently, they are possibly removed from one allele during OR selection, as suggested by the reduced levels of these trimethyl-marks on the active P2 allele. Therefore, our data describe a novel form of silencing that combines characteristics of both constitutive and facultative heterochromatin; OR heterochromatin has the same molecular and biochemical characteristics as the pericentromeric and telomeric heterochromatin, but it is dynamic and it depends on the identity of the cell and its differentiation state.

OR silencing precedes OR choice

Our data suggest a model for OR choice that incorporates our biochemical findings and previously reported genetic observations consistent with the existence of a feedback signal (Lewcock and Reed, 2004; Serizawa et

al., 2005; Shykind et al., 2004). According to this, all the OR alleles become silenced before the neuron reaches a differentiation state that supports OR transcription. At this stage, a limited enzymatic activity removes H3K9me3 and H4K20me3, from a stochastically chosen allele, allowing its transcriptional activation. The following synthesis of an intact OR elicits a feedback signal that prevents this enzyme/selector from activating another allele, and probably through a different mechanism, stabilizes the transcription of the chosen allele. In other words, the feedback signal does not silence the non-chosen OR alleles, but, instead, prevents their de-silencing. As a result, an OR-generated feedback is not responsible for creating the singularity in OR choice, but, instead, for preserving it for the life of the neuron.

Alternatively, it is possible that, at the moment of OR choice, OR genes are repressed only via H3K9me2, which is already deposited on these genes at an earlier differentiation state. In this scenario the expression of an OR gene could trigger the subsequent trimethylation of H3K9 and H4K20, resulting in permanent OR silencing. However, our data are consistent with the former model. First, we detect high levels of H3K9me3 and H4K20me3 on sustentacular cells (Figure 4F) that do not express OR genes, and therefore do not "experience" such a feedback signal. Second, in a quadruple negative (OMP, ICAM-1, iLR and SUS4 negative) cell population from the MOE, where we cannot detect OR mRNA either (Figure 4H), we find high enrichment levels for the two trimethyl-marks on OR genes (Figure 4G). Finally, in a Neurogenin-1 positive population, which, by RNAseq, has very low levels of OR mRNA (Figure 5C), the enrichment for H3K9me3 and H4K20me3 on OR genes is as high as in the mature OSNs (Figure 5B). Although only the identification of the enzymes involved in OR silencing and their genetic manipulation thereof can provide a definite distinction of the two models, the molecular logic of OR choice is conceptually the same: in both models, chromatin mediated silencing precedes OR choice.

Potential role for lysine 9 and lysine 20 trimethylation

If, however, H3K9me2 is sufficient for the retention of OR genes in a repressed state throughout the organism, what is the significance of adding one extra methyl-group on H3K9 and of trimethylating H4K20? A simple interpretation is that the additional lysine methylation events result in a more compact chromatin structure, as shown in Figure 3, which prevents the binding of OR-activating and OSN-specific transcription factors on multiple OR alleles. In other words, this transition in methyl-marks provides better protection from the activating transcription factor milleux and, therefore, more efficient repression, in accordance with the regulatory needs of olfactory neurons.

It is possible though, that trimethylation of H3K9 and H4K20 offers additional regulatory benefits. For example, since H3K9me3 and H4K20me3 decorate almost exclusively OR genes in the MOE (Figures 1C,D), the de-methylases responsible for their removal can be dedicated to the transcriptional needs of OR choice. These enzymes could be sequestered in a unique nuclear location, under the control of H-like enhancer elements, providing the necessary singularity for OR choice. Following OR expression these enzymes can be inactivated in response to the feedback signal, without affecting expression of the rest of the genes in the MOE.

However, a model whereby heterochromatic silencing of OR genes occurs before the activation of one allele poses a problem: How can the activator "read" the primary DNA sequence of the OR promoters while they are embedded in this inaccessible chromatin configuration (Ptashne, 2007)? A simple solution to this problem would be that the same heterochromatic histone modifications, which are used to induce chromatin compaction, function also as recruiters for the activation machinery. It is well established that pericentromeric repeats from different chromosomes converge to large heterochromatic chromo-centers in the eukaryotic nucleus (Ferrai et al., 2010). If H3K9me3 and H4K20me3 play an instructive role in this process, similar organization might exist for OR genes. We have previously reported that the H element interacts preferentially with OR promoters and repetitive DNA elements, which are both enriched for these heterochromatic marks (Lomvardas et al., 2006). Thus, it is possible that the deposition of these marks on OR genes extends beyond transcriptional silencing; it could provide instructive cues for the organization of OR genes in the nucleus in a way that allows for the monoallelic activation of a single OR gene.

Nucleation of OR heterochromatin

Our data raise questions as to how the silencing machinery is recruited to the OR genes. The only region with extensive sequence identity between ORs resides within their coding sequences, providing a possible

candidate target for the recruitment of the silencing activity. In agreement with this hypothesis, the CDS of OR genes has been previously implicated in transcriptional silencing (Nguyen et al., 2007). In this scenario, given the tandem organization of OR genes, there would be multiple heterochromatin nucleation sites within each cluster. In addition to the CDS, the LINE elements that are highly enriched in OR and VR loci (Pauler et al., 2009) might contribute to this heterochromatic silencing as has been suggested for X inactivation (Chow et al., 2010). Finally, our analysis revealed an intriguing correlation between the high AT content of chemoreceptor clusters and enrichment for H3K9me3 and H4K20me3 (Figures S1B, S1C), providing an additional mechanism by which genomic properties could instruct epigenetic changes. Thus, a combinatorial recruitment of the OR silencing machinery, through the OR CDS, LINE elements and a high AT-content, is possible.

In conclusion, our experiments provide a molecular glimpse into the monoallelic expression of olfactory receptors in the mouse. In the other well-characterized stochastic regulatory process in mammals, X inactivation, the choice is made between two transcriptionally active X chromosomes; whereas one will remain on, the other will be silenced as a consequence of the choice (Royce-Tolland and Panning, 2008). In OR choice, the logic is different. Silencing occurs before a choice is made and the choice itself is, most probably, mediated by de-repression. Similar logic applies to *var* gene choice in *Plasmodium falciparum,* where heterochromatic silencing prevents multigenic expression and relief of the silencing probably coincides with var choice (Freitas-Junior et al., 2005). A common theme between OR choice and the choice of var genes is the high number of available alleles. If OR genes were maintained in a poised and accessible state, then the concomitant selection of multiple alleles would be unavoidable. Therefore, the high number of OR alleles, together with the need for a strictly monogenic and monoallelic OR expression gave rise to an unusual regulatory circuit. It remains to be seen whether the regulatory principles proposed here apply to other neuronal systems whereby neurons commit permanently to differentiation processes regulated by irreversible transcriptional choices.

FIGURE LEGENDS

Figure 1. **Genome-wide mapping of H3K9me3 and H4K20me3 reveal a tissue-dependent heterochromatinization of the ORs in the MOE.** ChIP-on-chip experiments with antibodies against H3K9me3 and H4K20me3 using native chromatin preparations from the MOE and liver. The log2 ratio of IP/input was calculated and used for the construction of the heatmaps presented here.
(A) Positional heatmaps of chromosomes 2, 7 and 9 are shown. Each row represents one gene in 1kb windows from -5kb to +5kb of the translation start site. Four states are shown as adjacent columns: liver-H3K9me3, OE-H3K9me3, liver-H4K20me3 and OE-H4K20me3. Arrows indicate the OR clusters found on these chromosomes, which appear to be hypermethylated (red) only in the OE but not in the liver. V1R and V2R clusters as well as the only type I OR cluster are shown in the heatmap from chromosome 7.
(B) Ranked heatmap illustrating 1,000 randomly selected genes (approx. 1 in 15 genes). Each row represents one gene in 200 bp windows from -1kb to +1kb of the translation start site. The ORs (blue lines) make up the vast majority of the genes that are positive for both modifications and are placed at the top of the heatmap. The red lines represent VRs and FPRs (see text) that are also enriched but, as shown here, have lower methylation levels.
(C) Ranked heatmap, constructed as the previous one but showing the top 1,000 genes positive for H3K9me3 and H4K20me3 in the MOE. Most of these genes are ORs that, also, rank the highest, as depicted by the blue lines next to the heatmap

Figure 2. OR clusters in the MOE are coated by tissue-specific heterochromatic blocks of H3K9me3 and H4K20me3. Ma2C analysis of our ChIP-on-chip data viewed on the UCSC genome browser.
(A) Part of the biggest OR cluster located on chromosome 2, which contains ~240 genes and spans a 5-MB region. The thin blue (H3K9me3) or red (H4K20me3) bars represent significant peaks (FDR ≤5%) identified in the MOE by MA2C using standard parameters (window=0.5 kb, min number of probes= 5, max gap=0.25 kb); the thick blue or red bars represent the blocks identified with modified parameters (window=10 kb, min number of probes= 20, max gap= 1kb). In the liver, there are only a few, sporadic H3K9me3 peaks and blocks (purple).
(B) Results from H3K9me3 ChIP-qPCR analysis using native chromatin preparations from MOE and liver. The Ptprj gene (marked by red rectangle in 2A) stands at the border of the OR cluster which coincides with the border of the heterochromatic block. Its most proximal -to the OR cluster- intron is enriched for H3K9me3 and H4K20me3 and its most distal intron, located 43kb downstream, is free of these modifications. Zfp560 serves as positive control.
(C) Same as B but for H4K20me3.
(D) Part of an OR cluster on chromosome 11 is interrupted by a small group of non-OR. Genes that are marked by green rectangle are transcriptionally active in the MOE and genes marked by red rectangles do not have detectable transcripts.
(E) Zoomed-in picture of the cluster, which shows that genes Btnl9 and Flt4, which are transcriptionally inactive, are partly methylated. Two sets of primers for each of these genes, one at the beginning (most proximal to the neighboring OR gene) and one at the end of the gene (most distal from the neighboring OR) were used in ChIP-qPCR.
(F) Results from H3K9me3 ChIP-qPCR analysis using native chromatin preparations from the MOE. OR genes tested, as well as *Zfp354c*, and part of the *Btnl9* and *Flt4* genes were enriched. In contrast, the active genes *Mgat1* and *Mapk9*, *Zfp879* and the distal part of *Btnl9* and *Flt4* were devoid of modifications. Experiments were performed in two biological replicates with similar results. Values are the mean of triplicate qPCR. Error bars indicate the S.E.M.
(G) Same as F but for H4K20me3.
Figure 3. The ORs acquire a highly compacted chromatin structure in the MOE.
(A) To examine the structure of OR chromatin, we performed a DNase I accessibility assay with nuclei from both MOE and liver. Nuclei were treated with DNase I, DNA was isolated at various time points (2 to 60 min) and equal amounts were used for qPCR. The amount of DNA measured at each interval was expressed as a fraction of the DNA present at 2 min of enzyme treatment and was plotted over time. We assayed several ORs as well as genes that are active or inactive in the MOE or liver, and their mean is shown here (see Suppl.Fig3A

for detailed analysis of all genes). Experiments were performed in 3 biological replicates with similar results. Representative data from one experiment are shown here.
(B) MNase digested chromatin was submitted to ultracentrifugation through a sucrose gradient. The largest or most compacted chromatin fragments are collected in the fractions with the highest sucrose concentration.
(C) Fractions from the MOE and liver were analyzed by agarose gel electrophoresis and Southern blot analysis with a degenerate OR probe that is used to monitor the distribution of OR sequences in different sucrose fractions. Black arrows mark the low molecular weight OR sequence that appear in the bottom fractions of the chromatin from the MOE. Equal amounts of MOE and liver chromatin were loaded on the gradient as demonstrated by the input lanes that represent DNA extracted from chromatin that was not loaded in the gradient.
(D) Same fractions form the MOE and liver analyzed with the use of a probe specific for ribosomal genes.

Figure 4. ChIP-qPCR assays for H3K9me3 and H4K20me3 in sorted cell populations from the MOE.
(A) Section of the MOE from an adult OMP-IRES-GFP mouse. Mature OSNs are labeled green, as they specifically express GFP under the transcriptional control of the *OMP* gene.
(B) GFP^+ cells (mature neurons) were isolated with FACS from OMP-IRES-GFP mice and were used for ChIP-qPCR experiments. *Golf*, *Tbp* and *Omp* are active genes in the mature neurons and are used as negative controls. *Zfp560* and major satellite repeats are used as positive controls. *Olfr690* is a type I OR.
(C) Staining of MOE section with an antibody against ICAM (PE-ICAM) labels the HBCs that specifically express this protein.
(D) ChIP-qPCR experiments with isolated HBCs.
(E) Staining of MOE section with SUS4 antibody that specifically labels the sustentacular cells.
(F) ChIP-qPCR experiments with isolated sustentacular cells.
Experiments were performed in biological replicates and values are the mean of technical triplicates of these experiments. Error bars represent the SEM.
(G) Immature neurons and progenitors from the MOE were isolated by collecting OMP-, ICAM-, iLR- and Sus4-negative cells (thus called quadruple negative). All ORs tested were enriched for the two modifications. Values are the mean of triplicate qPCR. Error bars represent SEM.
(H) RNA was isolated from OMP-GFP+ cells, sustentacular and basal cells combined, and quadruple negative cells and was used in qRT-PCR reactions with primers amplifying different ORs. Actin was used as an endogenous control. Error bars represent SEM.

Figure 5. **ChIP-qPCR and RNAseq analysis of Neurogenin-1 positive cells.**
(A) Section of the MOE from an adult Ngn1-iIRES-GFP mouse. GBCs that specifically express Neurogenin-1 also express GFP and are labeled green.
(B) GFP^+ cells (GBCs) were isolated with FACS from Ngn1-IRES-GFP mice and were used for ChIP experiments for H3K9me3 and H4K20me3.
(C) Expression levels of OR transcripts, as determined by Illumina mRNA-seq, are quantified by normalized RPKM (reads per kilobase of exon model per million mapped reads – see Extended Experimental Procedures). RPKM increases with radius from the center of the figure, clamped at a maximum of 1. Each radial bar represents the level of expression of a single OR. For each OR gene, red indicates the expression level in Neurog1-positive neurons and blue indicates the expression level in OMP-positive neurons. ORs are sorted first by chromosome, indicated by the number or letter exterior to each wedge of the figure, and then by increasing gene start position within each chromosome. The two classes of ORs (Class I and II) are demarcated by background shading

Figure 6. The active OR allele is not enriched for H3K9me3 or H4K20me3, but it is marked with H3K4me3. Heterozygote P2-IRES-GFP mice were used to isolate GFP^+ cells, whereby the targeted P2 (Olfr17) gene is active, and GFP- cells by FACS. ChIP experiments were performed in these cells with antibodies against H3K9me3 and H4K20me3.
(A) Confocal image of P2-IRES-GFP+ cells in sections of the MOE depicting the extremely low number of GFP positive cells (~0.1% of the total cells).

(B) The location of the primers used in this experiment. Primers for the GFP sequence were used to specifically monitor the active allele, while the p2WT primers specifically amplified the inactive allele.
(C-D) As shown here, GFP is hypomethylated on H3K9 in the GFP^+ cells, where it is transcribed, but not in the GFP- cells, where this P2 allele is inactive. The transcribed region of the P2 gene appears to be hypermethylated in both sorted populations, probably because the primers cannot distinguish the two alleles. The inactive allele, amplified specifically by the p2WT primers, shows high enrichment for H3K9me3 and H4K20me3 in both GFP+ and GFP- populations. Due to extremely limited starting material we were able to perform very few qPCR reactions in each experiment, for this reason the positive (major satellite repeats) and negative (OMP gene) controls are represented from separate experiments (C and D)
(E) Same as C but for H4K20me3 (similar results obtained from biological duplicates).
(F) We repeated the same ChIP-qPCR experiment with an antibody against H3K4me3. There is significant enrichment for H3K4me3 throughout the P2 gene, but not on the neighboring P3 gene or a distant OR (Olfr177) in the GFP^+ cells. As expected, there was no H3K4me3 on the P2 gene, or any other OR gene, in the GFP^- cells. Values are the mean of triplicate qPCR. Error bars represent SEM.

Figure 7. Tissue-specific OR modifications are associated with OR-like transgene expression. (**A**) Graphic representation of the *Olfr459* locus and the OMP-LacZ insertion site located 55 kb away. Positions marked A, B, and C indicate assayed regions in the qPCR analysis below.
(**B**) ChIP-qPCRs with chromatin from the MOE of wild type mouse show that the *Olfr459* is enriched for H3K9me3 and H4K20me3 and both modifications appear to extend to the insertion site.
(**C**) ChIP-qPCR analysis of the MOE and liver from OMP-LacZ positive animals. Both H3K9me3 and H4K20me3 show MOE-specific deposition on *Olfr459*, the OMP-LacZ transgene, and the regions proximal to these loci. Experiments were performed on two biological replicates with similar results. Values are the mean of triplicate qPCR. Error bars are SEM.
(**D**) X-gal stains of lateral whole mounts of the nasal cavities from hemizygote and homozygote OMP-LacZ animals. Quantification as described in Experimental Procedures. *N,* number of biological replicates. $*p < 10^{-4}$, Student's t-test.

Experimental procedures

Mice were treated in compliance with the rules and regulations of IACUC under a protocol approval number AN084169-01.

ChIP-qPCRs and ChIP-on-chip experiments

ChIP-qPCRs assays, sequential ChIPs and ChIP-on-chip experiments were performed according to standard protocols and are described in detail in Extended Experimental Procedures.

Data analysis

Quality control of the ChIP-on chip data was performed both by NimbleGen (according to their protocols) and by our group. The log2 (ChIP/input) ratio was normalized in a 'weighted global' manner. For peak analysis we used the Model-based Analysis of 2-Color arrays (MA2C) (Song et al., 2007), which is a variation of the general sliding window approach. We also confirmed our results by a different algorithm (Wen et al., 2009). See Extended Experimental Procedures for more details.

DNase I accessibility assay

Nuclei were isolated from MOE and liver and digested with 20U of DNase I (Ambion) for 2-60 min at 37^{o}C; reactions were terminated with 0.5M EDTA, pH 8.0. Following proteinase K digestion, DNA was purified by phenol-chloroform extraction and ethanol precipitation, treated with RNase A and further purified with the PCR-column purification kit (Qiagen), before being quantified using the Nanodrop. Equal amounts of DNA were used in qPCR. As a control, we also digested naked DNA (nuclei were treated with proteinase K, DNA was purified and digested with DNase I) from the two tissues and there were no differences between them (data not shown). The assay was performed 3 times with similar results.

Chromatin Fractionation

The MOE and liver from 6 to 8-week-old C57BL/6 mice were dissected and native chromatin was prepared as described in Extended Experimental Procedures. Briefly, nuclei were extracted and digested with diluted MNase to yield DNA fragments with an average size larger than 20Kbs. The fractionation was performed as described before (Gilbert et al., 2004) but the gradient was generated with the freeze-thaw method using a 30% starting sucrose solution. Since lower molecular weight OR fragments could be detected with regular gel electrophoresis (0.8% agarose gel run at 6v/cm for 20h) we did not use PFGE.

Fluorescence activated cell sorting of MOE cell populations

For FACS experiments the olfactory epithelium of 6 to 10-week-old OMP-IRES-GFP, Ngn1-IRES-GFP or P2-IRES-GFP mice was dissected and cells were dissociated using a papain dissociation kit (Worthington Biochemical, Freehold, NJ) following the manufacturer's instructions. Cell sorting was performed on an AriaII (BD Biosciences) cell sorter. Sorted cells were used for ChIPs-qPCR and expression analysis. See Extended Experimental Procedures for more details.

Immunofluorescence of sections of the MOE

Immunofluorescence for ICAM-1 and SUS-4 and subsequent confocal microscopy was performed as described (Lomvardas et al., 2006)

Transgene mapping and X-galactosidase staining

Mapping of the OMP-LacZ transgene and staining of neurons with X-galactosidase were performed as per standard protocols. See Extended Experimental Procedures for more details.

SUPPLEMENTAL DATA

Supplemental data include 4 figures, 1 table (Excel file), Extended Experimental Procedures and References.

ACKNOWLEDGMENTS

We are grateful to Dr. Frank Margolis for the OMP-LacZ mouse strain, Dr. James E. Schwob for the SUS4 antibody, Drs. Andrew Feinberg and Bo Wen for the primer sequences for the positive controls for the H3K9me2 ChIPs and Dr. Rafael Irizarry for his suggestions for performing our ChIP-on-chip experiments. We would also like to thank Drs. Hiten Madhani, Barbara Panning, Nirao Shah, Dimitris Thanos and Keith Yamamoto for critical reading of the manuscript. B.M.C and E.J.C. are supported by fellowships from the National Science Foundation W.A. is supported by a Royce Fellowship. G.B. is a Pew Scholar in the Biomedical Sciences and S.L. is funded by the McKnight Endowment Fund for Neuroscience ,Rett Syndrome Research Trust, Hellman Family Foundation, Sandler Program for Biomedical Sciences, NIDCD (R03 DC010273) and Director's New Innovator Award Program (1DP2 OD006667).

REFERENCES

Barnea, G., O'Donnell, S., Mancia, F., Sun, X., Nemes, A., Mendelsohn, M., and Axel, R. (2004). Odorant receptors on axon termini in the brain. Science (New York, NY *304*, 1468.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bozza, T., Vassalli, A., Fuss, S., Zhang, J.J., Weiland, B., Pacifico, R., Feinstein, P., and Mombaerts, P. (2009). Mapping of class I and class II odorant receptors to glomerular domains by two distinct types of olfactory sensory neurons in the mouse. Neuron *61*, 220-233.

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

Capello, L., Roppolo, D., Jungo, V.P., Feinstein, P., and Rodriguez, I. (2009). A common gene exclusion mechanism used by two chemosensory systems. The European journal of neuroscience *29*, 671-678.

Carter, L.A., MacDonald, J.L., and Roskams, A.J. (2004). Olfactory horizontal basal cells demonstrate a conserved multipotent progenitor phenotype. J Neurosci *24*, 5670-5683.

Cau, E., Casarosa, S., and Guillemot, F. (2002). Mash1 and Ngn1 control distinct steps of determination and differentiation in the olfactory sensory neuron lineage. Development (Cambridge, England) *129*, 1871-1880.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M.*, et al.* (2009). Unlocking the secrets of the genome. Nature *459*, 927-930.

Chen, X., Fang, H., and Schwob, J.E. (2004). Multipotency of purified, transplanted globose basal cells in olfactory epithelium. The Journal of comparative neurology *469*, 457-474.

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

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A.G. (2010). VEZF1 elements mediate protection from DNA methylation. PLoS genetics *6*, e1000804.

Dulac, C., and Axel, R. (1995). A novel family of genes encoding putative pheromone receptors in mammals. Cell *83*, 195-206.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.H., Hannon, G., Tarakhovsky, A., and Fuchs, E. (2009). Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. Cell *136*, 1122-1135.

Feinstein, P., Bozza, T., Rodriguez, I., Vassalli, A., and Mombaerts, P. (2004). Axon guidance of mouse olfactory sensory neurons by odorant receptors and the beta2 adrenergic receptor. Cell *117*, 833-846.

Ferrai, C., de Castro, I.J., Lavitas, L., Chotalia, M., and Pombo, A. (2010). Gene positioning. Cold Spring Harb Perspect Biol *2*, a000588.

Fleischmann, A., Shykind, B.M., Sosulski, D.L., Franks, K.M., Glinka, M.E., Mei, D.F., Sun, Y., Kirkland, J., Mendelsohn, M., Albers, M.W.*, et al.* (2008). Mice with a "monoclonal nose": perturbations in an olfactory map impair odor discrimination. Neuron *60*, 1068-1081.

Fodor, B.D., Shukeir, N., Reuter, G., and Jenuwein, T. (2010). Mammalian Su(var) Genes in Chromatin Control. Annu Rev Cell Dev Biol.

Freitas-Junior, L.H., Hernandez-Rivas, R., Ralph, S.A., Montiel-Condado, D., Ruvalcaba-Salazar, O.K., Rojas-Meza, A.P., Mancio-Silva, L., Leal-Silvestre, R.J., Gontijo, A.M., Shorte, S.*, et al.* (2005). Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell *121*, 25-36.

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.

Fuss, S.H., and Ray, A. (2009). Mechanisms of odorant receptor gene choice in Drosophila and vertebrates. Molecular and cellular neurosciences *41*, 101-112.

Ghirlando, R., Litt, M.D., Prioleau, M.N., Recillas-Targa, F., and Felsenfeld, G. (2004). Physical properties of a genomic condensed chromatin fragment. J Mol Biol *336*, 597-605.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555-566.
Gregg, C., Zhang, J., Butler, J.E., Haig, D., and Dulac, C. (2010a). Sex-specific parent-of-origin allelic expression in the mouse brain. Science (New York, NY *329*, 682-685.
Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science (New York, NY *329*, 643-648.
Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.
Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nature neuroscience *7*, 483.
Henikoff, S. (1990). Position-effect variegation after 60 years. Trends Genet *6*, 422-426.
Hiragami-Hamada, K., Xie, S.Q., Saveliev, A., Uribe-Lewis, S., Pombo, A., and Festenstein, R. (2009). The molecular basis for stability of heterochromatin-mediated silencing in mammals. Epigenetics Chromatin *2*, 14.
Jang, W., Kim, K.P., and Schwob, J.E. (2007). Nonintegrin laminin receptor precursor protein is expressed on olfactory stem and progenitor cells. J Comp Neurol *502*, 367-381.
Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231-1245.
Leung, C.T., Coulombe, P.A., and Reed, R.R. (2007). Contribution of olfactory neural stem cells to tissue maintenance and regeneration. Nature neuroscience *10*, 720-726.
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.
Liberles, S.D., Horowitz, L.F., Kuang, D., Contos, J.J., Wilson, K.L., Siltberg-Liberles, J., Liberles, D.A., and Buck, L.B. (2009). Formyl peptide receptors are candidate chemosensory receptors in the vomeronasal organ. Proceedings of the National Academy of Sciences of the United States of America *106*, 9842-9847.
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.
McClintock, T.S. (2010). Achieving Singularity in Mammalian Odorant Receptor Gene Choice. Chemical senses *35*, 447-457.
Merriam, L.C., and Chess, A. (2007). cis-Regulatory elements within the odorant receptor coding region. Cell *131*, 844-846.
Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J., and Axel, R. (1996). Visualizing an olfactory sensory map. Cell *87*, 675-686.
Nguyen, M.Q., Zhou, Z., Marks, C.A., Ryba, N.J., and Belluscio, L. (2007). Prominent roles for odorant receptor coding sequences in allelic exclusion. Cell *131*, 1009-1017.
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.
O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. PLoS genetics *3*, e89.
Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D.P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome research *19*, 221-233.
Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A.*, et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323-337.
Ptashne, M. (2007). On the use of the word 'epigenetic'. Curr Biol *17*, R233-236.
Pyrski, M., Xu, Z., Walters, E., Gilbert, D.J., Jenkins, N.A., Copeland, N.G., and Margolis, F.L. (2001). The OMP-lacZ transgene mimics the unusual expression pattern of OR-Z6, a new odorant receptor gene on mouse chromosome 6: implication for locus-dependent gene expression. J Neurosci *21*, 4637-4648.

Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H., Schotta, G., Jenuwein, T.*, et al.* (2007). Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. Molecular cell *27*, 353-366.
Ressler, K.J., Sullivan, S.L., and Buck, L.B. (1993). A zonal organization of odorant receptor gene expression in the olfactory epithelium. Cell *73*, 597-609.
Ressler, K.J., Sullivan, S.L., and Buck, L.B. (1994). Information coding in the olfactory system: evidence for a stereotyped and highly organized epitope map in the olfactory bulb. Cell *79*, 1245-1255.
Riviere, S., Challet, L., Fluegge, D., Spehr, M., and Rodriguez, I. (2009). Formyl peptide receptor-like proteins are a novel family of vomeronasal chemosensors. Nature *459*, 574-577.
Rodriguez, I. (2007). Odorant and pheromone receptor gene regulation in vertebrates. Current opinion in genetics & development *17*, 465-470.
Rodriguez-Gil, D.J., Treloar, H.B., Zhang, X., Miller, A.M., Two, A., Iwema, C., Firestein, S.J., and Greer, C.A. (2010). Chromosomal location-dependent nonstochastic onset of odor receptor expression. J Neurosci *30*, 10067-10075.
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.
Royce-Tolland, M., and Panning, B. (2008). X-inactivation: it takes two to count. Curr Biol *18*, R255-256.
Sammeta, N., Yu, T.T., Bose, S.C., and McClintock, T.S. (2007). Mouse olfactory sensory neurons express 10,000 genes. The Journal of comparative neurology *502*, 1138-1156.
Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes & development *18*, 1251-1262.
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.
Serizawa, S., Miyamichi, K., Takeuchi, H., Yamagishi, Y., Suzuki, M., and Sakano, H. (2006). A neuronal identity code for the odorant receptor-specific and activity-dependent axon sorting. Cell *127*, 1057-1069.
Shykind, B.M. (2005). Regulation of odorant receptors: one allele at a time. Human molecular genetics *14 Spec No 1*, R33-39.
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.
Song, J.S., Johnson, W.E., Zhu, X., Zhang, X., Li, W., Manrai, A.K., Liu, J.S., Chen, R., and Liu, X.S. (2007). Model-based analysis of two-color arrays (MA2C). Genome Biology *8*, R178.
Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Molecular cell *28*, 1-13.
Tsuboi, A., Miyazaki, T., Imai, T., and Sakano, H. (2006). Olfactory sensory neurons expressing class I odorant receptors converge their axons on an antero-dorsal domain of the olfactory bulb in the mouse. The European journal of neuroscience *23*, 1436-1444.
Vassar, R., Chao, S.K., Sitcheran, R., Nunez, J.M., Vosshall, L.B., and Axel, R. (1994). Topographic organization of sensory projections to the olfactory bulb. Cell *79*, 981-991.
Vassar, R., Ngai, J., and Axel, R. (1993). Spatial segregation of odorant receptor expression in the mammalian olfactory epithelium. Cell *74*, 309-318.
Walters, E., Grillo, M., Oestreicher, A.B., and Margolis, F.L. (1996). LacZ and OMP are co-expressed during ontogeny and regeneration in olfactory receptor neurons of OMP promoter-lacZ transgenic mice. Int J Dev Neurosci *14*, 813-822.
Wang, F., Nemes, A., Mendelsohn, M., and Axel, R. (1998). Odorant receptors govern the formation of a precise topographic map. Cell *93*, 47-60.
Weintraub, H., and Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science (New York, NY *193*, 848-856.
Wen, B., Wu, H., Shinkai, Y., Irizarry, R.A., and Feinberg, A.P. (2009). Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. Nature genetics *41*, 246-250.

Figure 1

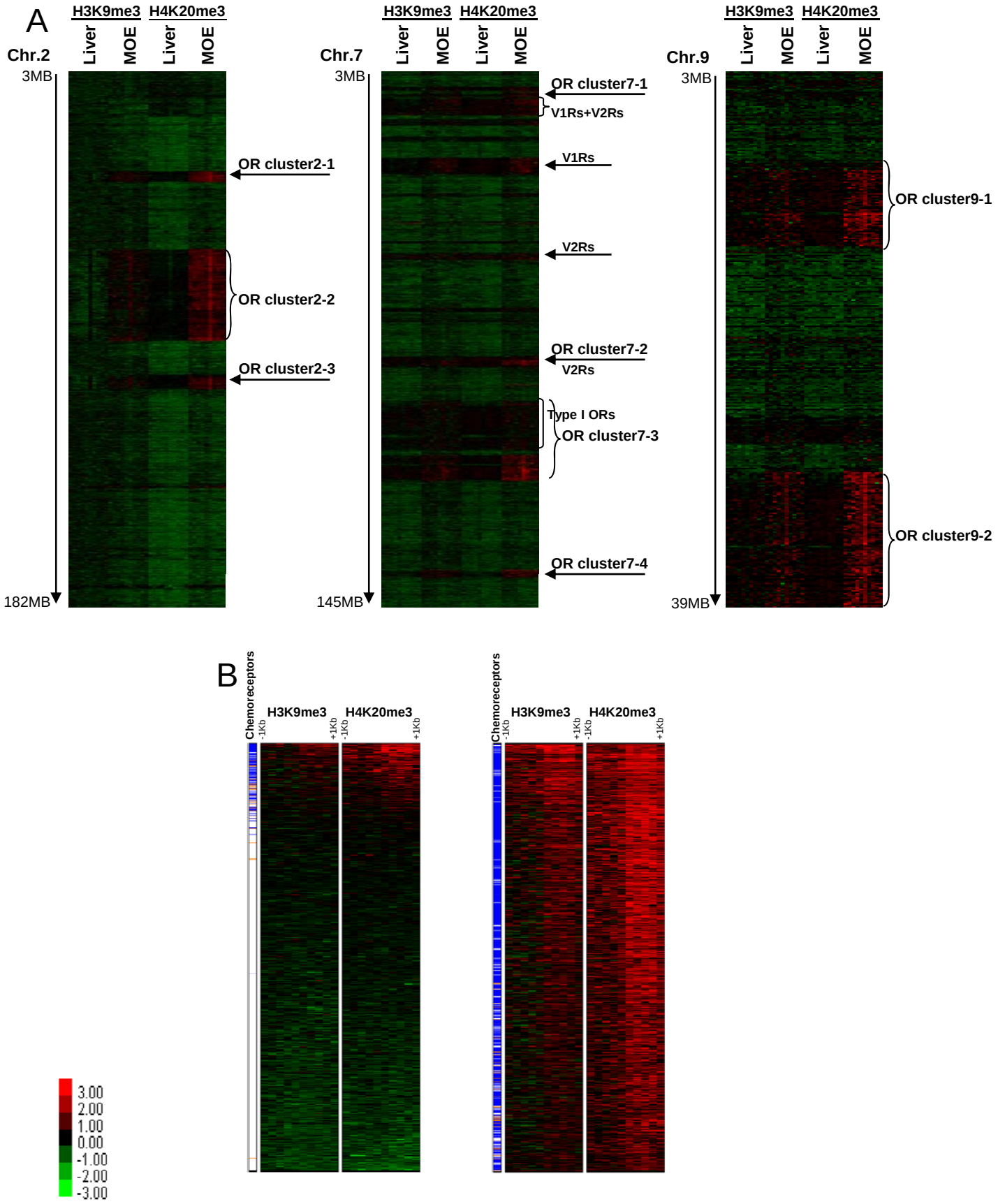
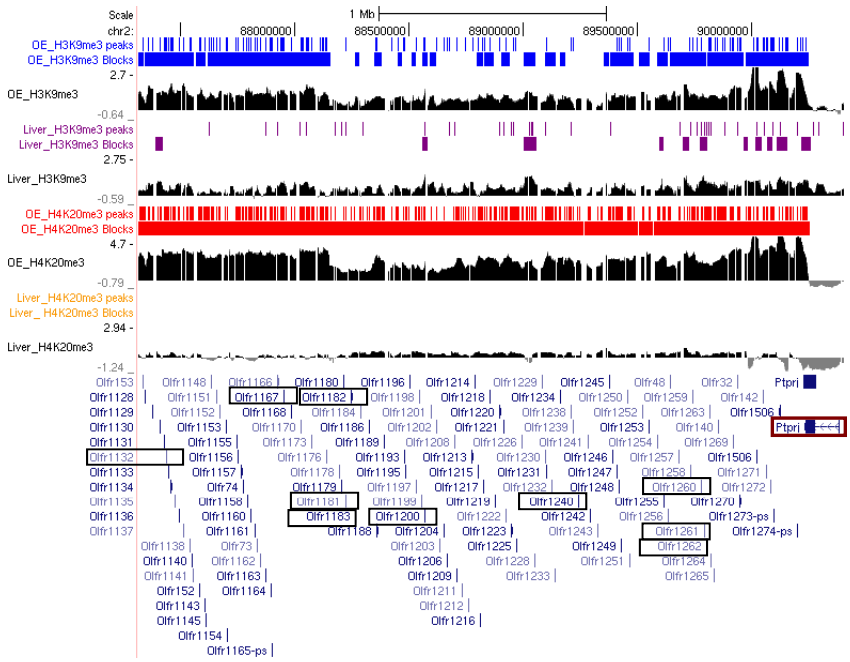
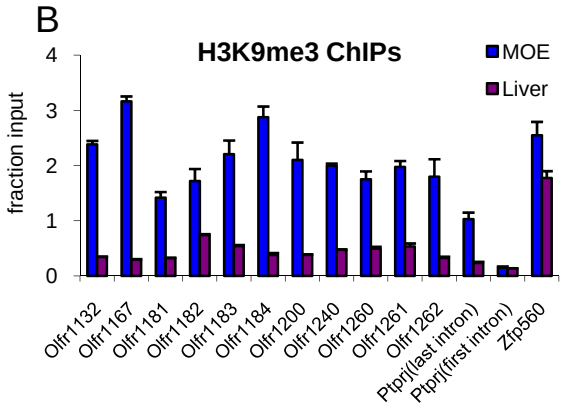
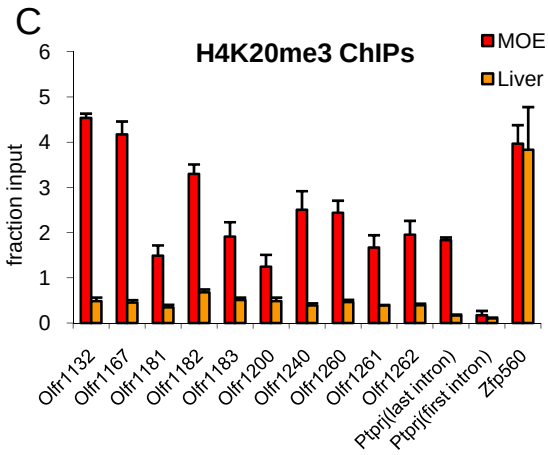
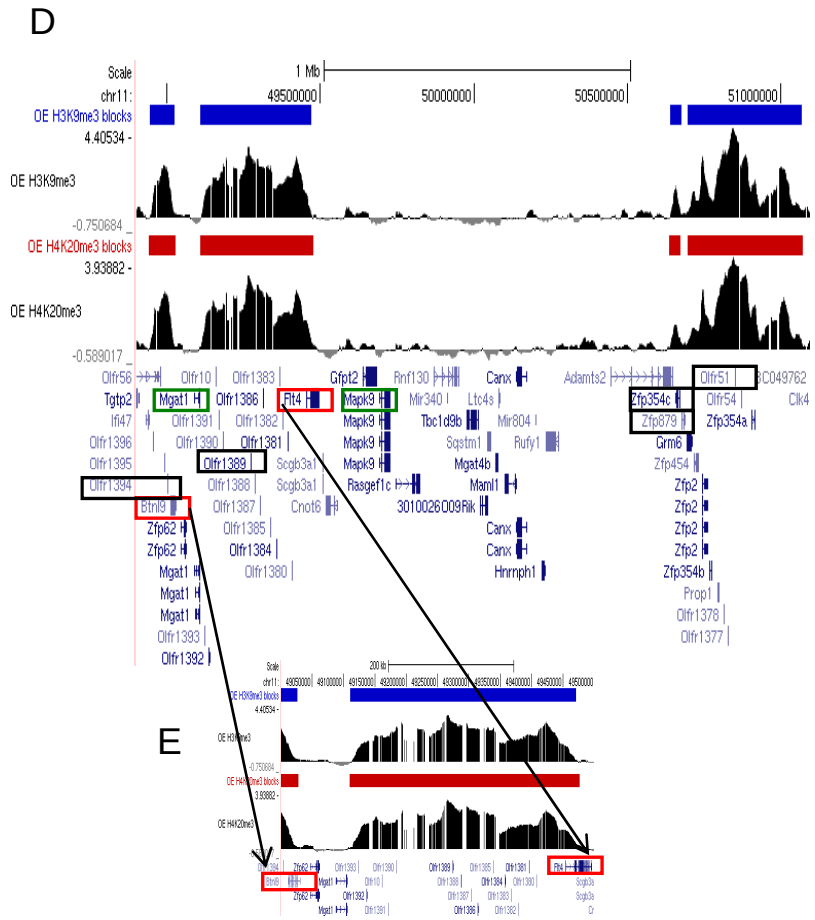

A
H3K9me3
H4K20me3
Liver
MOE
Liver
MOE
Chr.2
3MB
182MB
OR cluster2-1
OR cluster2-2
OR cluster2-3
Chr.7
3MB
145MB
OR cluster7-1
V1Rs+V2Rs
V1Rs
V2Rs
OR cluster7-2
V2Rs
Type I ORs
OR cluster7-3
OR cluster7-4
Chr.9
3MB
39MB
OR cluster9-1
OR cluster9-2
B
Chemoreceptors
H3K9me3
H4K20me3
-1Kb
+1Kb
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00

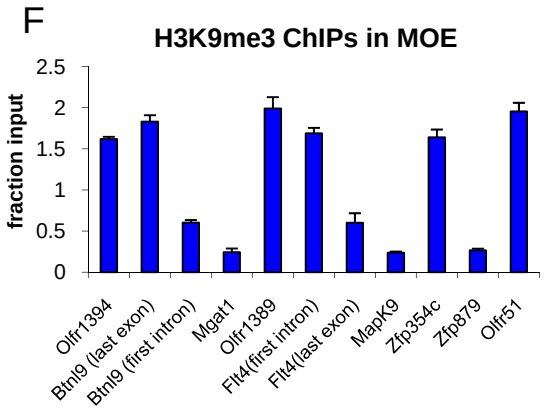
Figure 2


A
Scale
chr2:
1 Mb
88000000
88500000
89000000
89500000
90000000
OE_H3K9me3 peaks
OE_H3K9me3 Blocks
OE_H3K9me3
Liver_H3K9me3 peaks
Liver_H3K9me3 Blocks
Liver_H3K9me3
OE_H4K20me3 peaks
OE_H4K20me3 Blocks
OE_H4K20me3
Liver_H4K20me3 peaks
Liver_H4K20me3 Blocks
Liver_H4K20me3
Olfr1167
Olfr1182
Olfr1183
Olfr1200
Olfr1132
Ptpri


B
H3K9me3 ChIPs
MOE
Liver
fraction input
Olfr1132
Olfr1167
Olfr1181
Olfr1182
Olfr1183
Olfr1184
Olfr1200
Olfr1240
Olfr1260
Olfr1261
Olfr1262
Ptprj(last intron)
Ptprj(first intron)
Zfp560


C
H4K20me3 ChIPs
MOE
Liver
fraction input
Olfr1132
Olfr1167
Olfr1181
Olfr1182
Olfr1183
Olfr1200
Olfr1240
Olfr1260
Olfr1261
Olfr1262
Ptprj(last intron)
Ptprj(first intron)
Zfp560


D
Scale
chr11:
1 Mb
49500000
50000000
50500000
51000000
OE H3K9me3 blocks
OE H3K9me3
OE H4K20me3 blocks
OE H4K20me3
Olfr1389
Btnl9
Flt4
Mgat1
Mapk9
Zfp354c
Zfp879
Olfr51
E


F
H3K9me3 ChIPs in MOE
fraction input
Olfr1394
Btnl9 (last exon)
Btnl9 (first intron)
Mgat1
Olfr1389
Flt4(first intron)
Flt4(last exon)
MapK9
Zfp354c
Zfp879
Olfr51

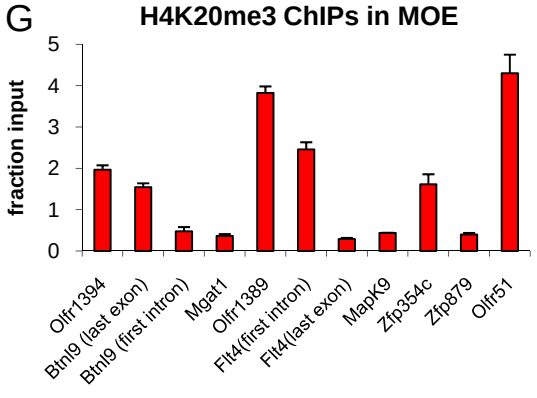
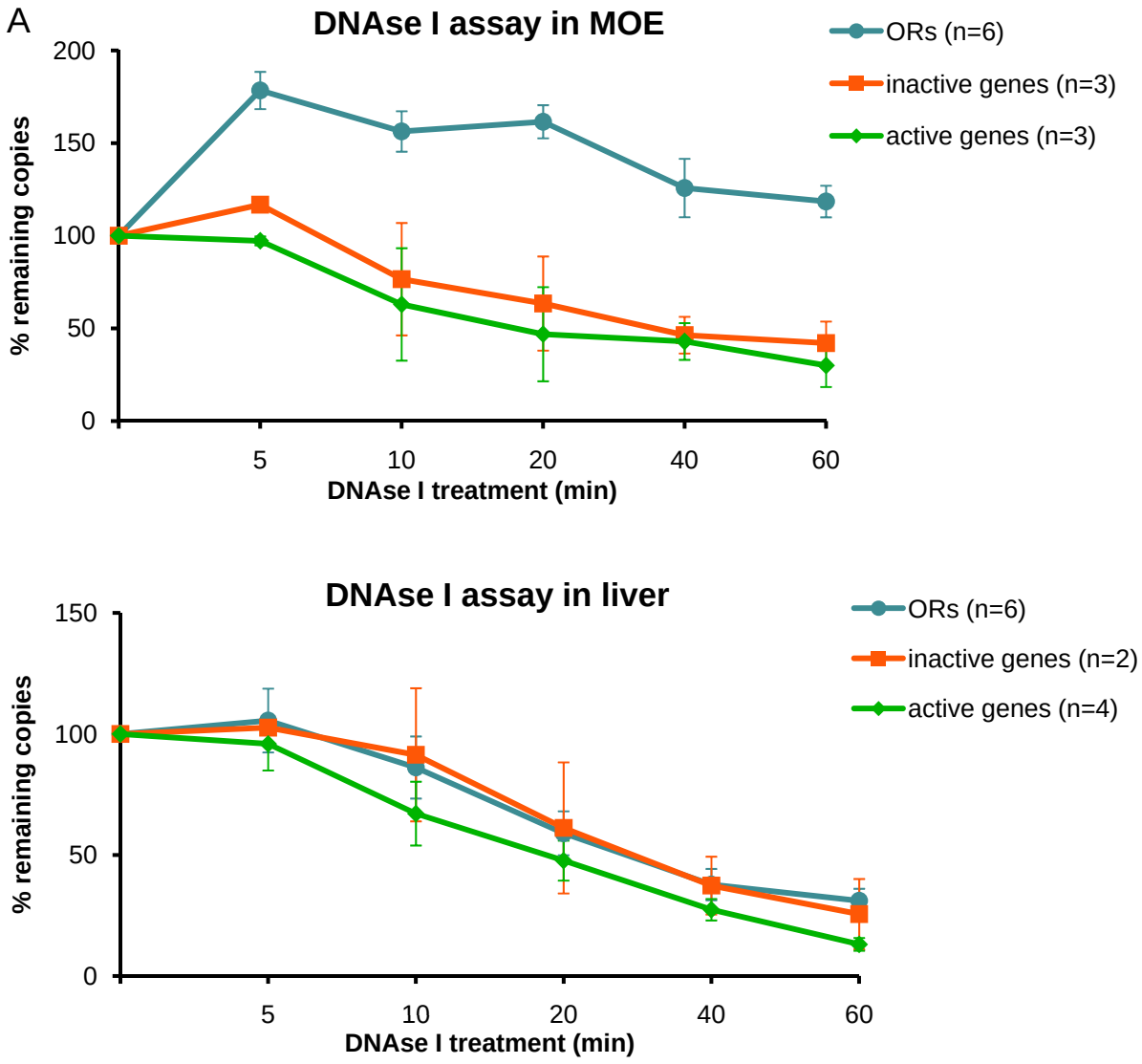
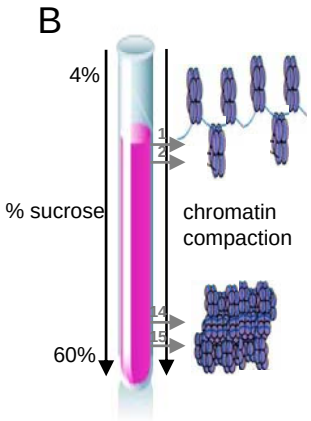

G
H4K20me3 ChIPs in MOE
fraction input
Olfr1394
Btnl9 (last exon)
Btnl9 (first intron)
Mgat1
Olfr1389
Flt4(first intron)
Flt4(last exon)
MapK9
Zfp354c
Zfp879
Olfr51

Figure 3


A
DNAse I assay in MOE
ORs (n=6)
inactive genes (n=3)
active genes (n=3)
% remaining copies
200
150
100
50
0
5
10
20
40
60
DNAse I treatment (min)
DNAse I assay in liver
ORs (n=6)
inactive genes (n=2)
active genes (n=4)
% remaining copies
150
100
50
0
5
10
20
40
60
DNAse I treatment (min)


B
4%
% sucrose
60%
chromatin compaction

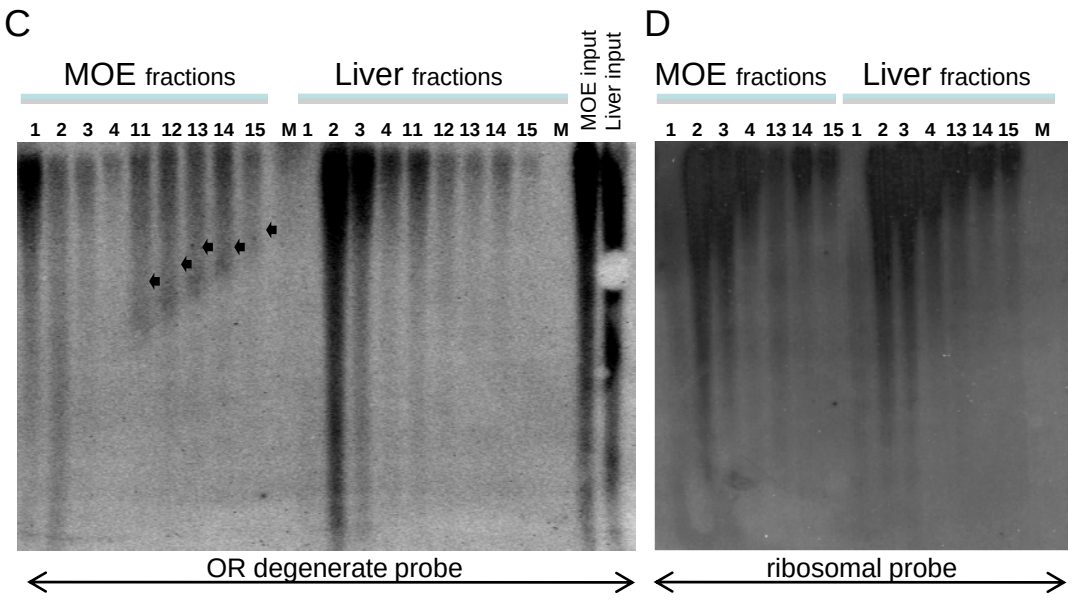

C
MOE fractions
Liver fractions
MOE input
Liver input
1 2 3 4 11 12 13 14 15 M 1 2 3 4 11 12 13 14 15 M
OR degenerate probe
D
MOE fractions
Liver fractions
1 2 3 4 13 14 15 1 2 3 4 13 14 15 M
ribosomal probe

Figure 4

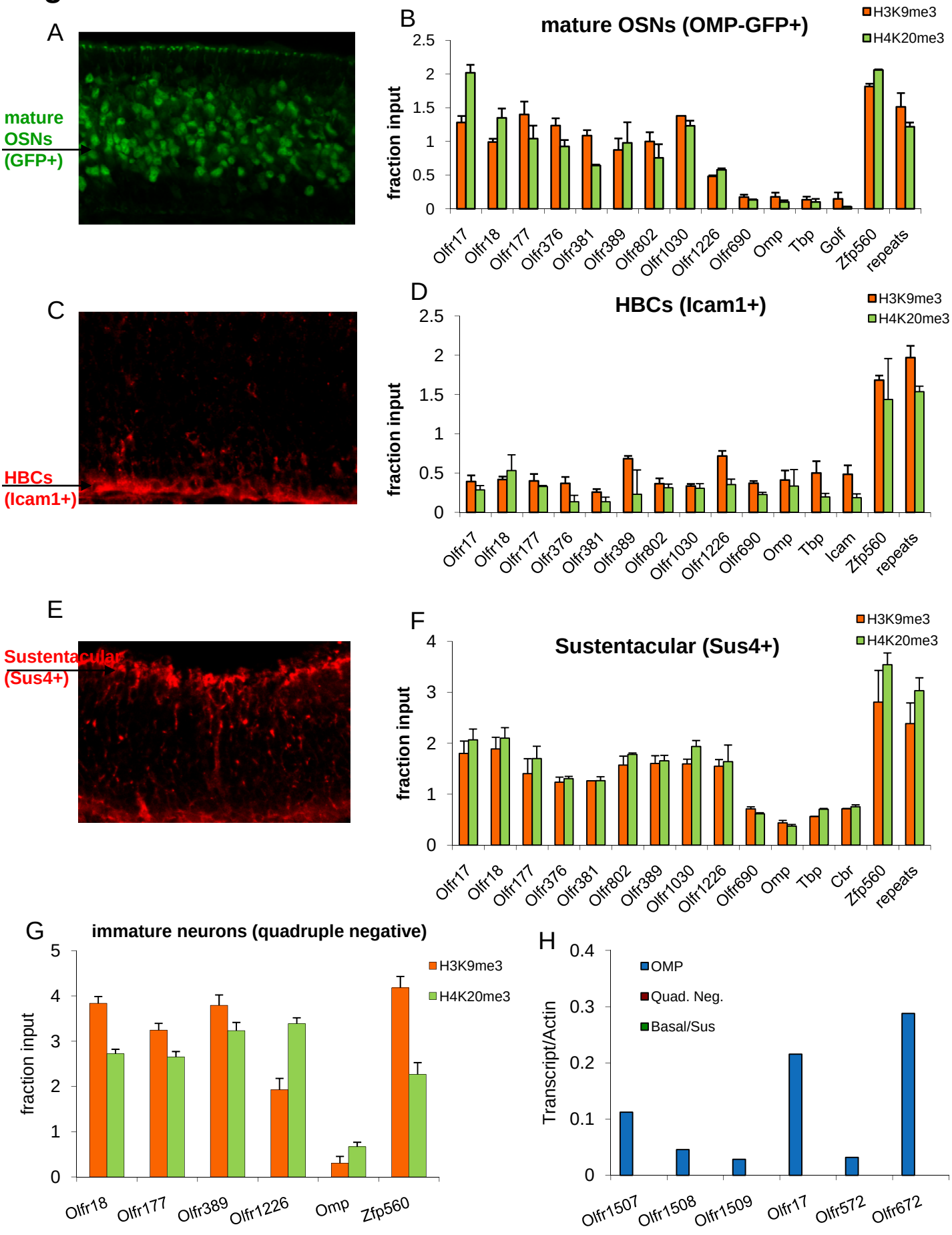
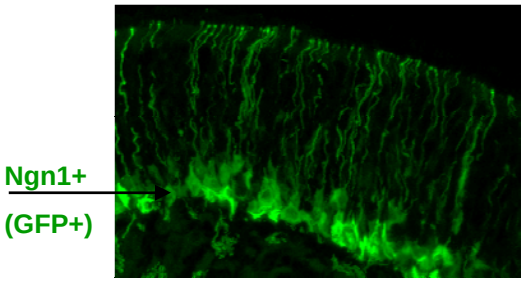
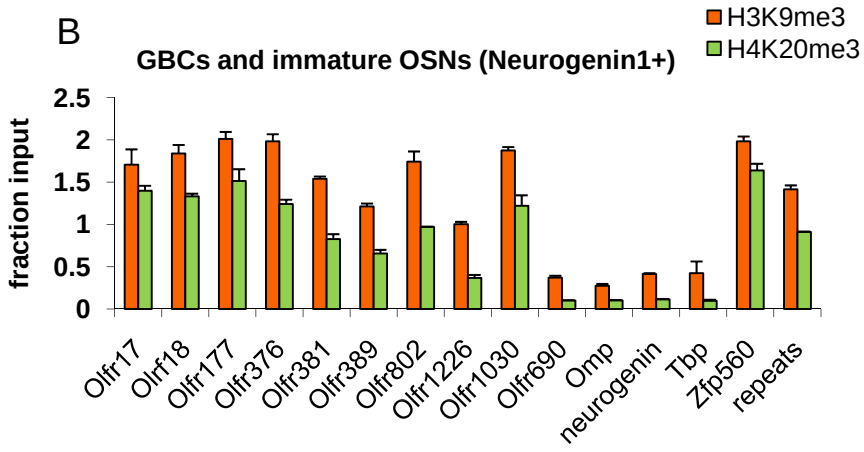

A
mature
OSNs
(GFP+)
B
mature OSNs (OMP-GFP+)
H3K9me3
H4K20me3
fraction input
2.5
2
1.5
1
0.5
0
Olfr17
Olfr18
Olfr177
Olfr376
Olfr381
Olfr389
Olfr802
Olfr1030
Olfr1226
Olfr690
Omp
Tbp
Golf
Zfp560
repeats
C
HBCs
(Icam1+)
D
HBCs (Icam1+)
H3K9me3
H4K20me3
fraction input
2.5
2
1.5
1
0.5
0
Olfr17
Olfr18
Olfr177
Olfr376
Olfr381
Olfr389
Olfr802
Olfr1030
Olfr1226
Olfr690
Omp
Tbp
Icam
Zfp560
repeats
E
Sustentacular
(Sus4+)
F
Sustentacular (Sus4+)
H3K9me3
H4K20me3
fraction input
4
3
2
1
0
Olfr17
Olfr18
Olfr177
Olfr376
Olfr381
Olfr802
Olfr389
Olfr1030
Olfr1226
Olfr690
Omp
Tbp
Cbr
Zfp560
repeats
G
immature neurons (quadruple negative)
H3K9me3
H4K20me3
fraction input
5
4
3
2
1
0
Olfr18
Olfr177
Olfr389
Olfr1226
Omp
Zfp560
H
OMP
Quad. Neg.
Basal/Sus
Transcript/Actin
0.4
0.3
0.2
0.1
0
Olfr1507
Olfr1508
Olfr1509
Olfr17
Olfr572
Olfr672

Figure 5

A


Ngn1+
(GFP+)

B


GBCs and immature OSNs (Neurogenin1+)
H3K9me3
H4K20me3
fraction input
2.5
2
1.5
1
0.5
0
Olfr17
Olfr18
Olfr177
Olfr376
Olfr381
Olfr389
Olfr802
Olfr1226
Olfr1030
Olfr690
Omp
neurogenin
Tbp
Zfp560
repeats

C

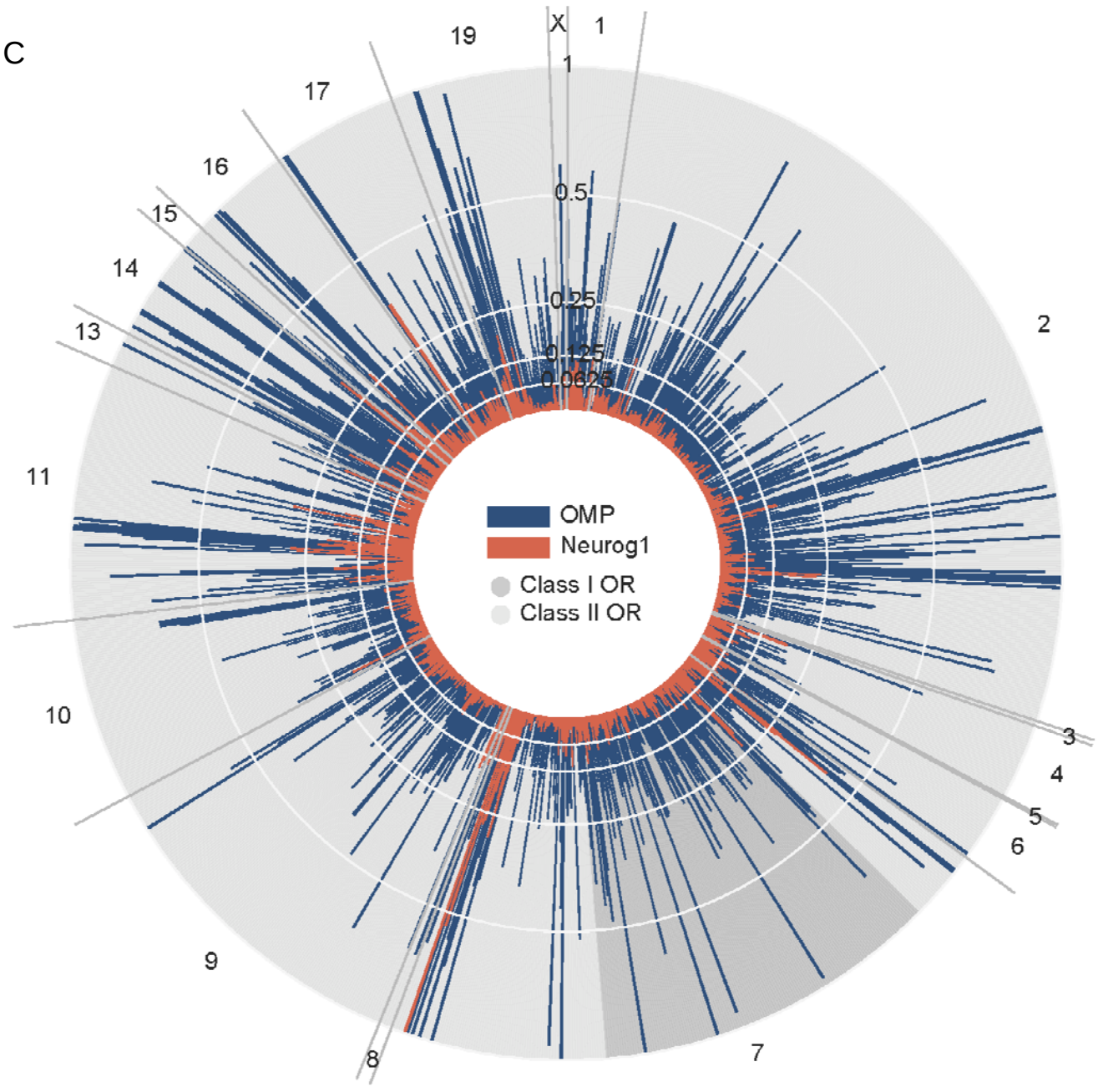
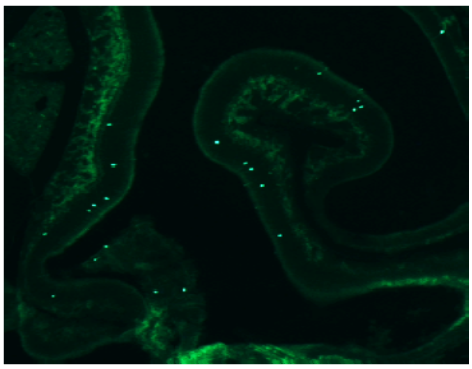
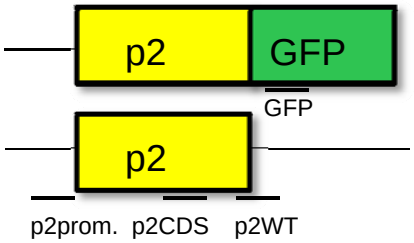

X
1
2
3
4
5
6
7
8
9
10
11
13
14
15
16
17
19
0.5
0.25
0.125
0.0625
OMP
Neurog1
Class I OR
Class II OR

Figure 6

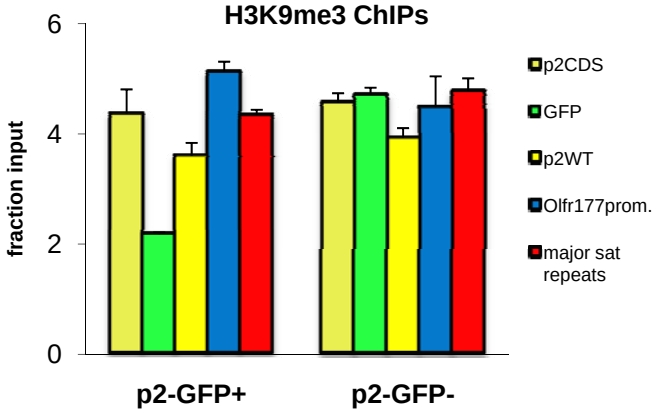
A



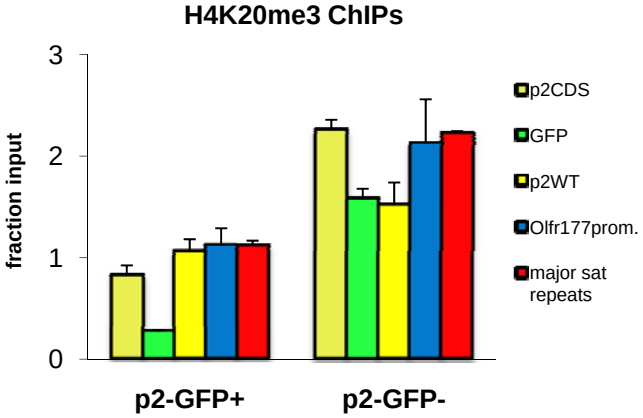
B



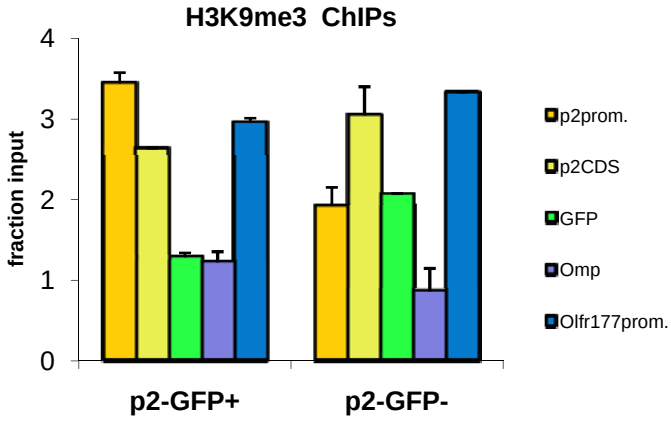
C



D



E



F

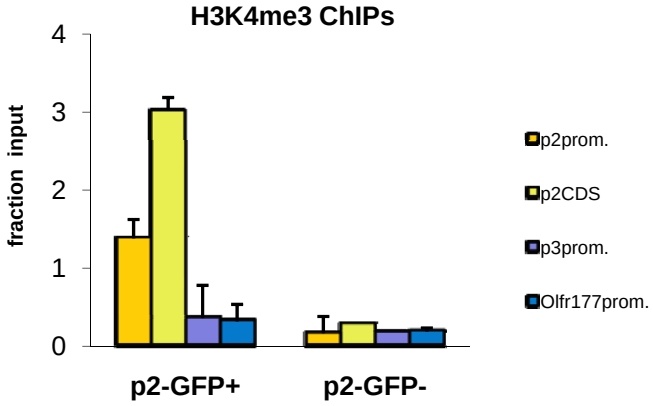
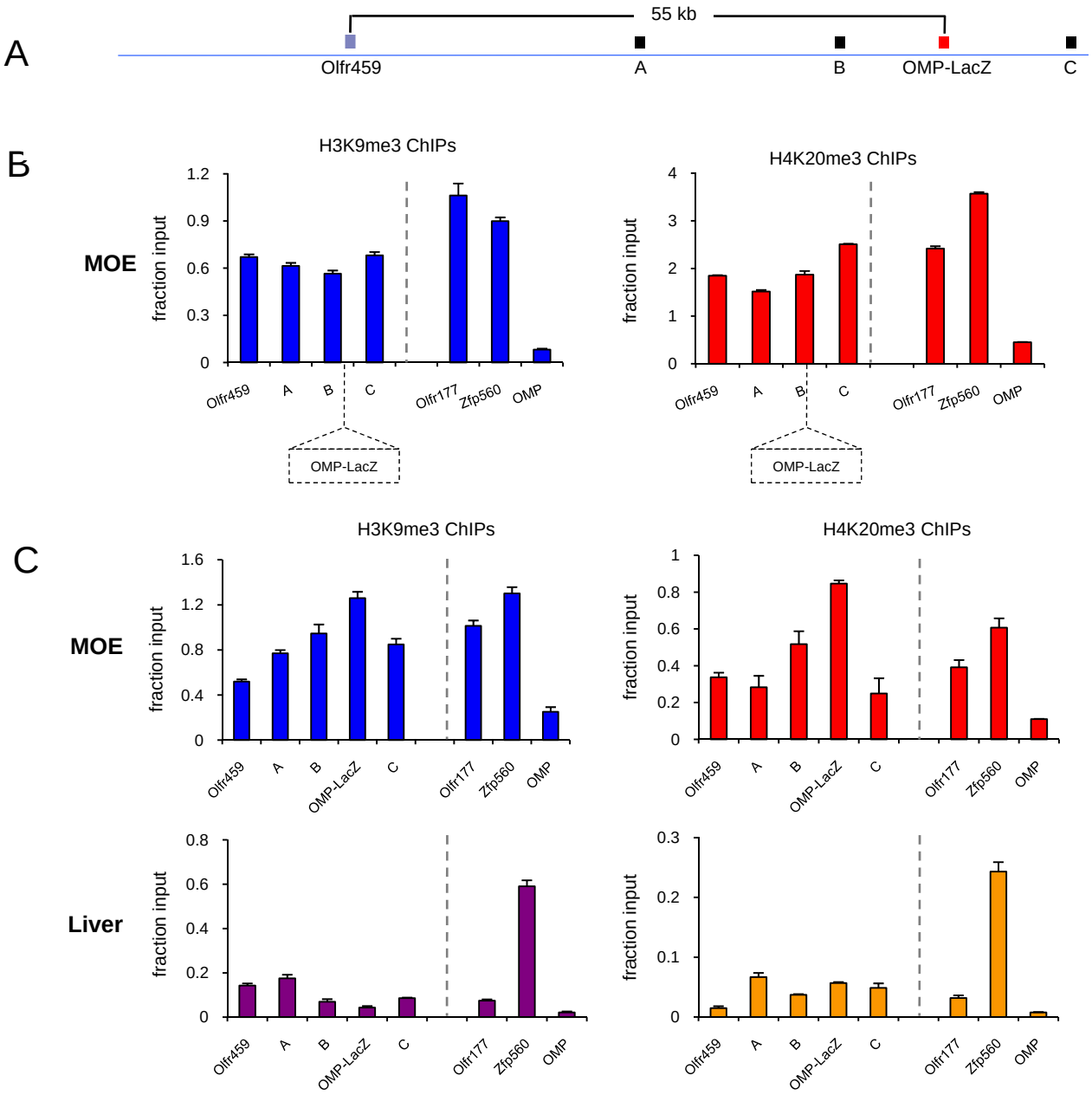
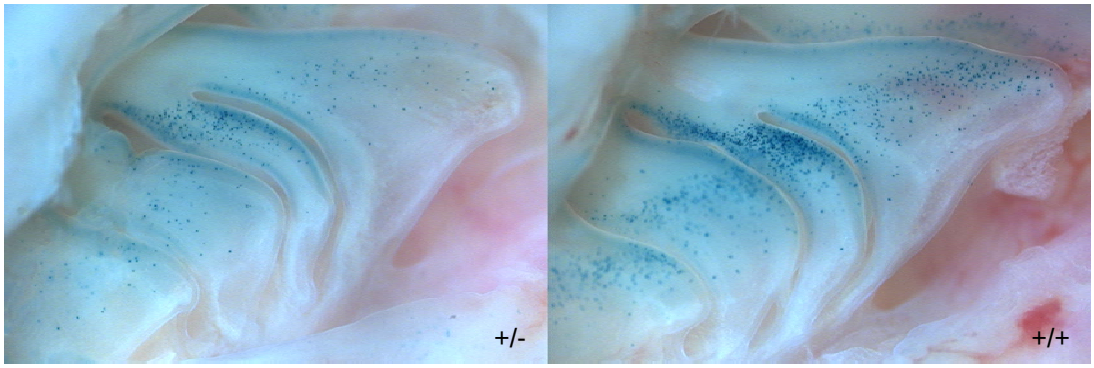


Figure 7



D



OMP-LacZ	LacZ+ cells / area		*N*
+/-	51±3	} 1.8 *	3
+/+	93±8		4

SUPPLEMENTAL INFORMATION

Supplemental Figure Legends

Figure S1. Genome-wide mapping of H3K9me3 and H4K20me3 in the MOE
(A) Sequential ChIPs confirm that H3K9me3 and H4K20me3 are present at the same time on the same loci. Chromatin immunoprecipitated with an antibody against H3K9me3 was eluted and submitted to a second round of immunoprecipitation with the same antibody (tK9/tK9, as a control) or an antibody against H4K20me3 (tK9/tK20). Real-time PCR ratios reflect the relative enrichment of indicated loci over *Omp*, which is used as a negative control. Representative results of two experiments are shown.
(**B**) Ranked heatmap illustrating 1,000 randomly selected genes (approx. 1 in 15 genes) in descending order for H3K9me3 and H4K20me3 enrichment. An additional heatmap on the side demonstrates the average AT-content of each represented gene (30-70%).
(**C**) As above but the ranked heatmap illustrates the top 1,000 genes positive for H3K9me3 and H4K20me3 in descending order of enrichment.
(D) Table presenting the genes that are enriched for both H3K9me3 and H4K20me3 based on the MA2C analysis with a sliding window of 10Kb. Blocks in mm8 built were intersected with murine genes using RefGene mm9 records lifted over to mm8. Gene families with an asterisk were completed by manual annotation; others were simply summarized by RefGene name2 prefix. Genes were called "positive", if they were at least 50% covered by both H3K9me3 and H4K20me3. Included in the table are families with at least four positive members and with p value for enrichment of the family, relative to the whole positive set, less than .05 (chi square test). The column on the right depicts whether the ChIP-on-chip results were confirmed by ChIP-qPCR or not.

Figure S2. Analysis of the Chip-on-chip data with two independent, computer algorithms yield similar results.
(A-B) Two examples, from chromosomes 2 **(A)** and 9 **(B)**, are shown here. The blue (H3K9me3) and red (H4K20me3) bars represent the blocks identified by MA2C (Song et al., 2007), while the green ones represent large domains (called LOCKs by the authors) identified by a method described in (Wen et al., 2009) (see Suppl. Experimental Procedures). Note the presence of blocks, apart from the OR clusters, only in a small group of Zfps (*) and in gene-poor areas (**) with high AT-content. Images generated on UCSC browser.

Figure S3. ChIP-qPCRs and expression analysis of sorted cells from the MOE.
(A) Cartoon depicting the cell lineage relationships in the MOE. Basal stem cells (HBCs and GBCs) give rise to OSNs and non-neuronal supporting cells (sustentacular) (adapted from Duggan and Ngai, 2007).
(B) ChIP-qPCR experiments with DNA immunoprecipitated by an antibody against H3K9me2 in HBCs. All OR loci tested with qPCR were enriched for this modification. Primers for Wnt2 and intergenic region (on chromosome 7) were used as positive control (Wen et al., 2009). *Omp* and *Icam* are devoid of H3K9me2. Interestingly Zfp560, which is enriched for H3K9me3 in these cells, is negative for H3K9me2. Values are the mean of triplicate qPCR. Error bars represent SEM.
(C) Expression profiling of the sorted populations confirms that each isolated population expresses predominantly the respective cell type marker.

Figure S4. ChIP-qPCR and expression analysis of P2-GFP+ neurons.
(A) Graphic representation of the P2 locus on chromosome 7 with the P3 (Olfr713), P4 (Olfr714) and P2 (Olfr17) genes. The arrows mark the location where the indicated primers bind; their distance from the translation start site is also shown. Picture not to scale.
(**B**) Immunoprecipitated DNA with an antibody against H3K4me3, as well as input, from P2-GFP+ cells, was amplified using the WGA4 kit (Sigma). Equal amount of IP DNA and input were used for qPCR with primers located in the promoter of P2, P3 and P4 genes, as well as that of a distant OR (Olfr177 on chr.16). H3K4me3 is detected only in the promoter and the CDS of the

active p2 gene, while, just, 500bp upstream of the promoter (p2upstr. primers) there is no detectable signal. *Tbp* and *Omp* are active genes in these cells and were used as positive controls.
(C) RNA was isolated from P2-GFP+ cells and was used in RT-PCR. As expected, only the P2 gene is expressed in these cells confirming the significant enrichment for P2 expressing neurons of this isolated population. Actin was used as an endogenous control. Error bars represent SEM.

Extended Experimental Procedures

Chromatin immunoprecipitation assays, sequential ChIPs and real-time qPCR

The MOE and liver from 6 to 8-week-old C57BL/6 mice were dissected and native chromatin was prepared as described before (Umlauf, 2004). Briefly, nuclei were extracted and digested with micrococcal nuclease (MNase, Sigma) to yield a population of mono- to tetra- nucleosomes that was, subsequently, used in chromatin immunoprecipitation assays. The antibodies used were specific to H3 monomethyl lysine-9 (ab8896), H3 dimethyl lysine-9 (ab1220), H3 trimethyl lysine-9 (ab8898), H3 trimethyl lysine-4 (ab1012), H4 monomethyl lysine-20 (ab9051) and H4 trimethyl lysine-20 (ab9053) from Abcam, and H4 dimethyl lysine-20 (07-369), H3 monomethyl lysine-27 (07-448), H3 dimethyl lysine-27 (07-421) and H3 trimethyl lysine-27 (07-449) from Upstate. Immunoprecipitated DNA was isolated by phenol-chloroform extraction and ethanol precipitation and used in quantitative PCR reactions on an ABI 7900 Real-Time PCR machine. Primer sequences and locations are provided by request. Data were processed with the ABI software and the enrichment of ChIP sample over input was calculated by the $2^{-\Delta Ct}$ method after correcting for the no antibody control. For sequential ChIP experiments, we followed the same protocol except that chromatin was eluted with 10mM DTT, diluted 20-fold and after the addition of BSA, salmon sperm DNA and yeast tRNA, it was subjected to a second immunoprecipitation with a different antibody and then eluted with standard elution buffer (0.1M NaHCO3, 1% SDS). The relative fold-enrichments were determined by the $2^{-\Delta\Delta Ct}$ method, using *Omp* as the normalizer. All ChIP experiments were performed multiple times using different chromatin preparations.

ChIP-on-chip experiments and data analysis

ChIP and input DNA samples were amplified using the WGA2 kit (Sigma). Labeling of DNA, hybridization and scanning were conducted at NimbleGen in Iceland, according to their standard protocols, on 2.1 Whole-Genome Economy arrays. We performed duplicate or triplicate ChIP-on-chip experiments for the whole genome for H3K9me3 and H4K20me3 in the MOE, and singletons for chromosomes 1-9 in the liver.
Quality control of the data was performed both by NimbleGen (according to their protocols) and by us using the protocol advised by the Epigenome Network of Excellence (Straub T., 2009). The log2 (ChIP/input) ratio was normalized in a 'weighted global' manner. This is similar to the usual global normalization (Wit E. and McClure J, 2004), except that it weights the data for each of the states (H3K9me3 in OE, for example) equally, in spite of how many replicates there are for a given state. Specifically, each sample of data is subtracted by its median and divided by its mean absolute deviation (MAD) as usual. Then, the 'weighted global' median and 'weighted global' MAD is calculated by first finding the average median and average MAD within each state, and then averaging these values across all four states. Since we expect each state to have a similar distribution, this allows each state to weight the global distribution similarly, even if certain states have more replicates than others. Using the normalized log-ratios, we generated positional heatmap representations for the genes to visually summarize our findings. For each gene and state, we centered a 10Kb window at the translation start site. Each window consisted of 10 buckets of 1Kb each, and the normalized log ratio (IP/input) of each probe was added to the appropriate bucket. All values in a bucket were averaged, including data from replicate experiments. Since there were multiple 'states' of the data, the values for each of the four states – H3K9me3 in liver and OE tissue, and H4K20me3 in liver and OE tissue - were concatenated. Once each gene was represented by its values (all four states with 10 buckets per modification),

we created heatmaps of the data, where the rows were ordered by position in the genome. We also ranked the genes by histone modification enrichment in the OE data. More specifically, we averaged the values in all 20 buckets for each gene (10 buckets of 200 base pairs each for each of the two histone modifications), and then ranked all the genes in the mouse genome by these values in descending order. A corresponding heatmap was created for this ranking. To quantify the significance of the OR genes showing stronger signal in this ranking than random, we used the Mann-Whitney Rank-Sum test.
For peak analysis, two different pipelines of existing tools were used: one was for Large-Organized Chromatin K9 Modifications (LOCKs) (Wen et al., 2009) and the other one was the Model-based Analysis of 2-Color arrays (MA2C) (Song et al., 2007). Both pipelines are variations on the general sliding window approach, but we adjusted the parameters to, especially, search for large-scale enrichment. In the LOCKs method, averaging was performed across 500 base pair windows, while the minimum block size was 10,000. In the MA2C pipeline, we used the default parameters for peak finding (window of 500bp, minimum number of probes in a window 5 and maximum gap of 250 base pairs) and also the adjusted parameters for large domains (window of 10Kb, minimum number of probes 20, max gap of 1Kb). An FDR≤5% was used in both cases.

Fluorescence activated cell sorting of MOE cell populations and ChIP-qPCRs with sorted cells

Because of the expected low yields after sorting, we performed pilot experiments with dissociated cells to determine the lowest cell number, as well as the MNase digestion conditions that would allow us to, reliably, detect histone modifications by ChIP-qPCR. We found that we could use as few as 40,000 cells per IP by slightly modifying our native chromatin protocol. To test the specificity of the FACS-ChIP protocol we used H3K4me3 IPs to test if the appropriate cellular markers are enriched for this mark according to their expression pattern (data not shown).
For FACS experiments, the olfactory epithelium of 6 to 10-week-old OMP-IRES-GFP, Ngn1-IRES-GFP or P2-IRES-GFP mice was dissected and cells were dissociated using a papain dissociation kit (Worthington Biochemical, Freehold, NJ) following the manufacturer's instructions. Cell sorting was performed on an AriaII (BD Biosciences) cell sorter. For the Ngn1-IRES-GFP sorting we collected only the populations with the highest GFP signal, since our pilot experiments showed that the population with lower GFP signal included OR expressing cells. To isolate various cell populations from the MOE, dissociated cells were incubated with primary antibodies SUS4 for sustentacular cell labeling (a gift from Dr. JE Schwob) or PE-ICAM for HBCs labeling, and subsequently with secondary antibodies conjugated to R-phycoerythrin (Jackson laboratories) or Alexa 594 (Molecular Probes, Invitrogen). Data were analyzed using FLowJo. After FACS, the cells were visually inspected under a microscope to confirm the purity of the sorted populations. Sorted cells were collected by centrifugation and nuclei were extracted and digested with MNase under pre-optimized conditions. ChIP assays were performed as described above. All FACS-ChIPs experiments were performed 2-4 times with similar results.

Expression analysis of sorted cells

Sorted cells were spinned down and resuspended in Trizol. RNA was extracted according to manufacturer's instructions and used for reverse transcription (Superscript III by Invitrogen). cDNA was used for qPCR with primers for cell type-specific markers or various OR genes. Results were normalized to actin.

Immunohistochemistry and confocal microscopy

To visualize developmental markers, fresh frozen 12 μm MOE sections were prepared from wildtype, OMP-ires-GFP, and Ngn1-GFP mice. Immunostaining was performed by briefly post-fixing in 4% PFA and incubating with anti-ICAM-1 and SUS4 antibodies following standard procedures. Sections were imaged using a Nikon C1si confocal.

RNA-Seq and data analysis

RNA-seq libraries were prepared by the standard Illumina protocol from RNA extracted from FACS sorted OMP-GFP or Neurog1-GFP positive neurons. An Illumina Genome Analyzer IIx was used to generate 35 bp single-end reads. These reads were then quality filtered and aligned to the UCSC mm9 mouse reference genome with the short-read aligner Bowtie (Langmead et al., 2009). Reads that could not be aligned or that aligned to more than one position in the genome were discarded. A custom Python script was used to count the number of reads that overlapped with each exon in the Ensembl mouse reference annotation to compute a mapped exon read count for each annotated gene. A custom R script was used to normalize these data by RPKM (reads per kilobase of exon model per million mapped reads - (Mortazavi et al., 2008)), with a total read count that was upper-quartile normalized (Bullard et al., 2010). The figure was generated with a custom script using the visualization library Protovis. The mature OSN markers that were used to estimate the fraction of mature OSNs in the Neurogenin-1 sample were: Pde3c, Kcnc4, Clic3, Faim2, Pcdh10, Kcna6, Adcy3, Sema3e, Scn8a, Jakmip1, Pde7b, Mt3, Epha5, Eomes, Lect, S100a5, Nap1l5, Omp, Crocc, Stom, Cidea, Pde1c, Cnga2, Olfm1, Gnal.

Transgene mapping and X-galactosidase staining

Genomic DNA from OMP-LacZ positive animals was digested with *XbaI*, diluted, and self-ligated. Inverse PCR was performed using primers directed against the SV40 polyadenylation sequence in the transgene, and the resulting amplicon was cloned using the TOPO-TA system (Invitrogen). Multiple inserts were sequenced and mapped to the mm8 reference sequence. The exact location of the insertion is 3' end: chr6:41758578 (mm8). Lateral whole mounts of the nasal cavities were fixed for 30 minutes with 4% PFA and stained for 1 hour with 1 mg/ml X-gal in X-gal staining solution (35 mM potassium ferricyanide, 35 mM potassium ferrocyanide, 2 mM $MgCl_2$, 0.02% NP-40, 0.01% sodium deoxycholate). Counts of OMP-LacZ+ cells were collected in two regions of each whole mount and normalized to region area using ImageJ.

Supplemental References

Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics *11*, 94.

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology *10*, R25.

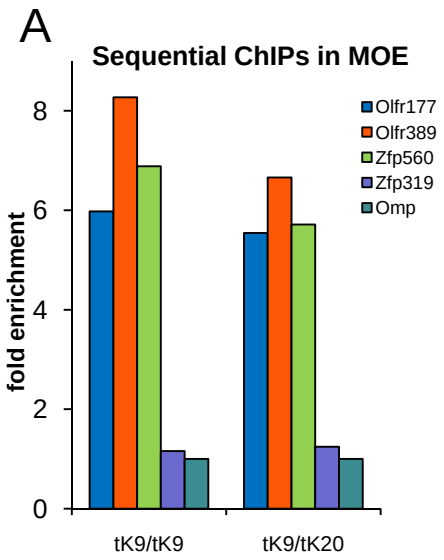
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

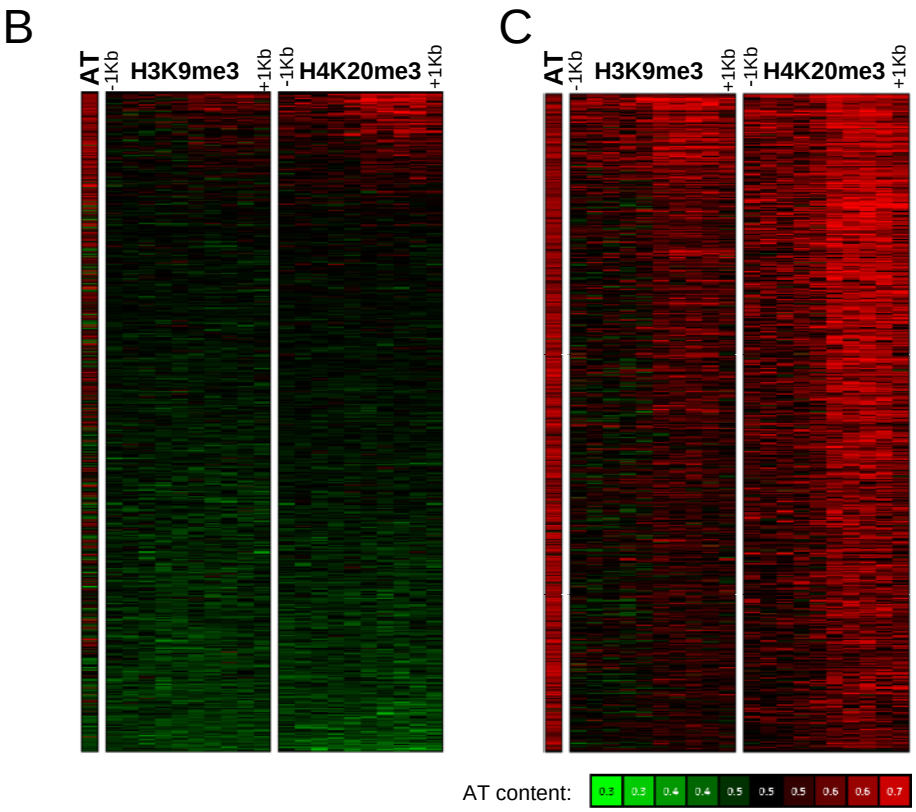
Straub, T. (2009). Epigenome NoE - protocol: Basic Analysis of NimbleGen ChIP-on-chip Data using Bioconductor/R. The Epigenome Network of Excellence. [Online]

Wit, E. and McClure, J. (2004) Statistics for Microarrays: Design, Analysis, and Inference. West Sussex : John Wiley & Sons, Ltd.

Umlauf, D., Goto, Y. and Feil, R. (2004). Site-specific analysis of histone methylation and acetylation. Methods Mol Biol 287, 99-120.

Figure S1



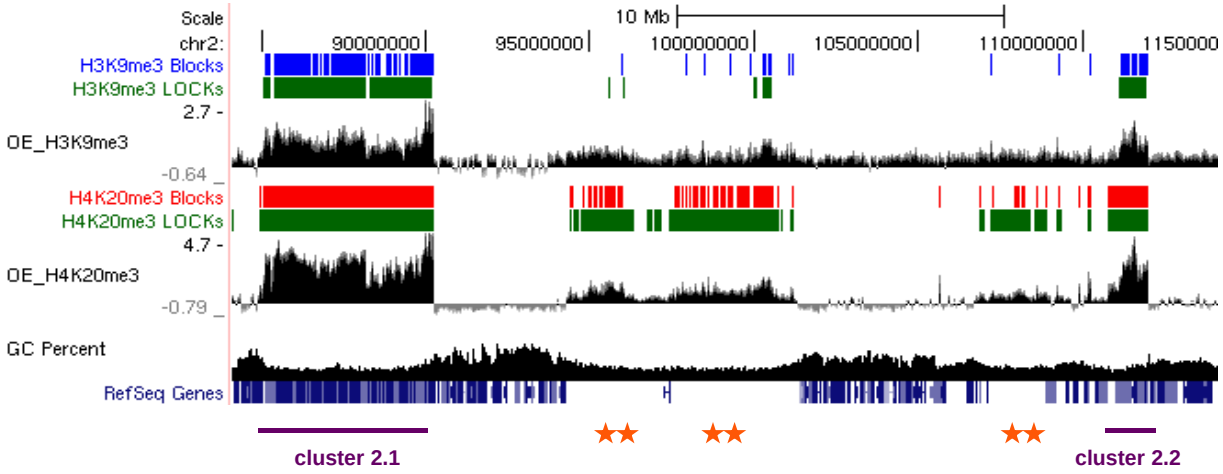


D

Gene Family	Positive	Total	P value	Function/ Description	qPCR
Olfr*	1106	1441	0.E+00	Olfactory GPCR	Confirmed
Zfp	82	325	2.E-24	Zinc Finger Transcription	Confirmed
V2r / Vmn2r	71	118	7.E-61	Vomeronasal GPCR	Low Levels
V1r / Vmn1r	67	212	1.E-30	Vomeronasal GPCR	Low Levels
Pcdha,b,g	35	59	2.E-41	Protocadherin Adhesion	Confirmed
Klra	16	21	6.E-27	NK Cell Immune Receptor	Not Confirmed
Defa	14	14	6.E-28	Defensin Immune Peptide	
Mup	10	13	1.E-17	Urine Secreted	Not Confirmed
Ssx	8	9	4.E-15	KRAB Transcription	
Zscan	7	13	2.E-08	Zinc Finger Transcription	
Esp	6	7	7.E-12	Exocrine Secreted	
Prame	6	11	1.E-07	Unknown	
Mrg	6	20	1.E-03	Nociception GPCR	
Tdpoz*	5	8	2.E-02	Unknown	Not Confirmed
Cts	5	23	4.E-02	Cathepsin Migration	
Amy	4	5	3.E-08	Amylase Digestion	
Obox	4	5	3.E-08	Homeobox Transcription	
Fpr	4	6	8.E-07	Vomeronasal GPCR	
Ms4	4	15	4.E-02	Unknown	
Misc / Unknown Family	456				
Total Positive Genes	1916				
Total Genes	20427				
Overall Proportion	0.094				

Figure S2

A


Scale
chr2:
10 Mb
90000000
95000000
100000000
105000000
110000000
1150000
H3K9me3 Blocks
H3K9me3 LOCKs
2.7 -
OE_H3K9me3
-0.64 _
H4K20me3 Blocks
H4K20me3 LOCKs
4.7 -
OE_H4K20me3
-0.79 _
GC Percent
RefSeq Genes
cluster 2.1
cluster 2.2

B

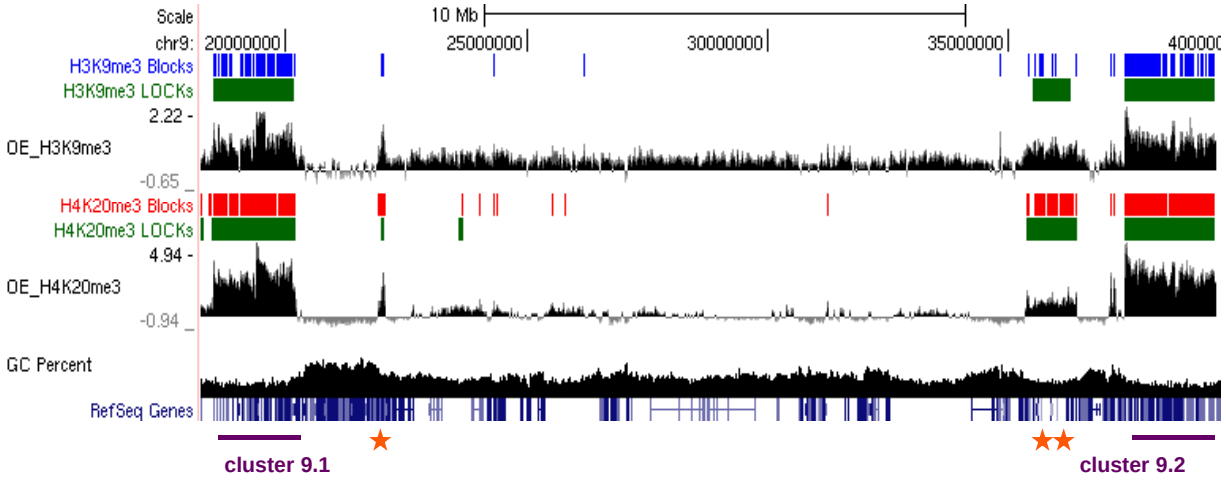
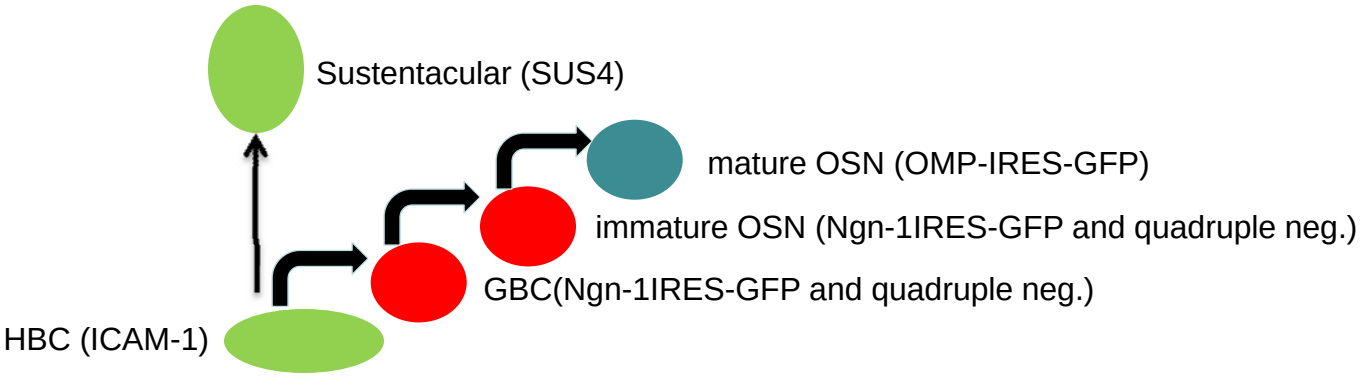
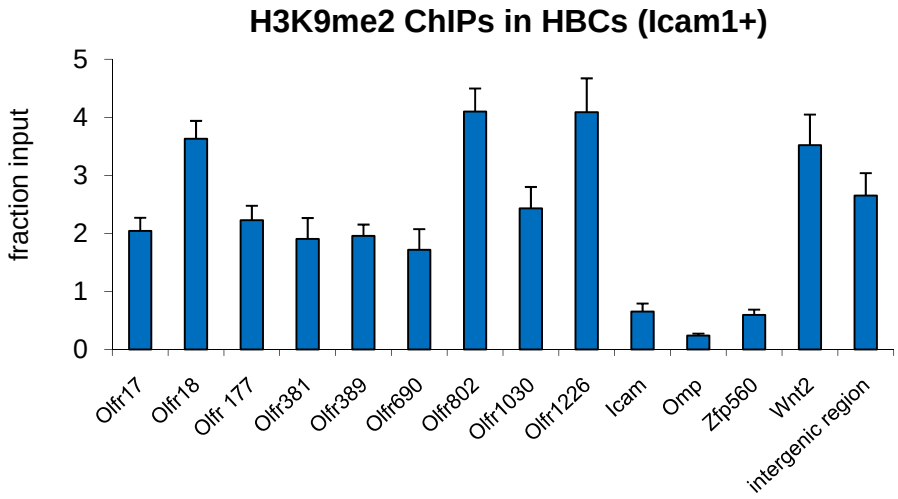

Scale
chr9:
10 Mb
20000000
25000000
30000000
35000000
400000
H3K9me3 Blocks
H3K9me3 LOCKs
2.22 -
OE_H3K9me3
-0.65 _
H4K20me3 Blocks
H4K20me3 LOCKs
4.94 -
OE_H4K20me3
-0.94 _
GC Percent
RefSeq Genes
cluster 9.1
cluster 9.2

Figure S3

A



B



C

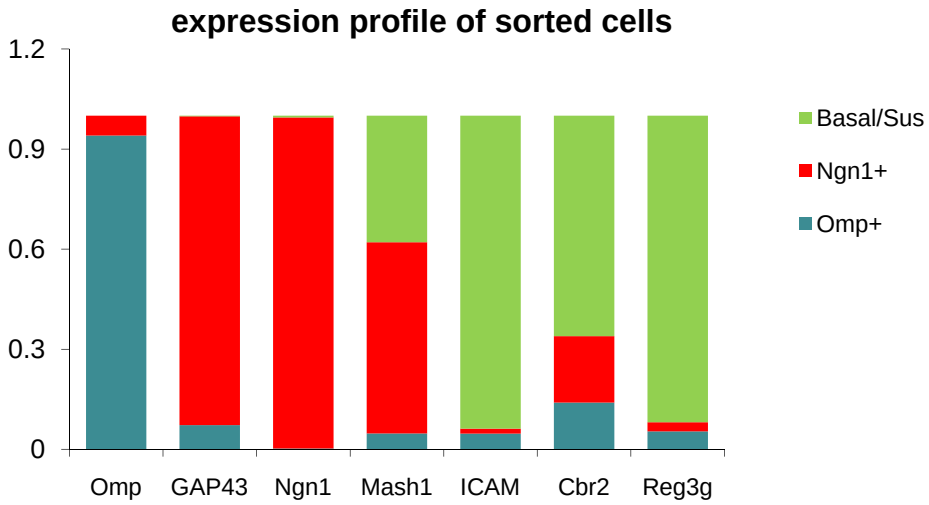
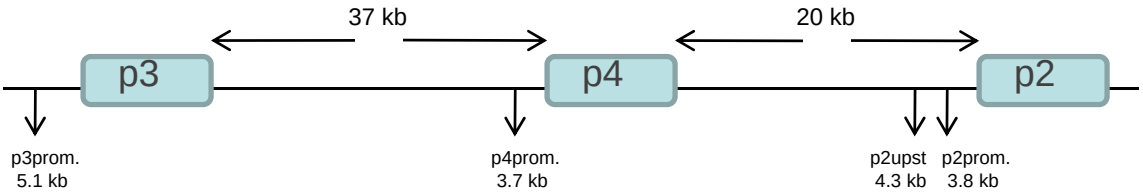
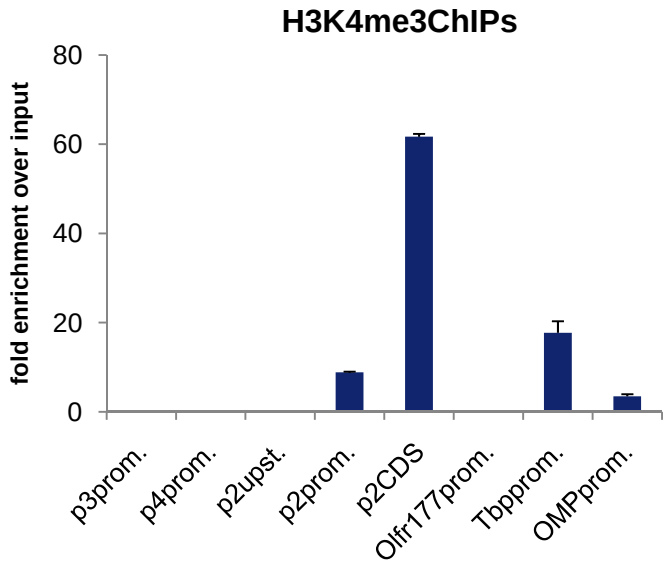


Figure S4

A



B



C