Mouse olfactory receptors (ORs) are encoded by a family of more than 1,000 genes \(^1\) that are organized in heterochromatic clusters \(^2\) distributed across chromosomes. Every mature olfactory sensory neuron (mOSN) expresses only one OR gene in a monoallelic and stochastic fashion \(^3,4\). Activation of OR genes requires removal of heterochromatic marks \(^2\) and the concerted action of 63 intergenic enhancers, the Greek islands, which are bound by the transcription factors LHX2 and EBF\(^6,7\). Expression of a single OR coincides with nuclear convergence of OR gene clusters \(^8,9\), which promotes interchromosomal interactions between Greek islands and the chosen OR gene \(^6\). The specificity by which Greek islands associate with the active OR allele, as well as the importance of their interchromosomal contacts in OR transcription, are uncertain. Although interchromosomal interactions occur in other systems \(^10–14\), unbiased approaches such as in situ Hi-C \(^15\) fail to detect robust \(\text{trans}\) contacts between non-repetitive regions \(^16,17\), raising questions about the frequency and biological roles of genomic interactions between chromosomes \(^17,18\). To obtain quantitative and functional insight into the regulation and function of multi-chromosomal interactions, we performed in situ Hi-C in 10 distinct wild-type and mutant cell types from the mouse main olfactory epithelium (MOE) (Extended Data Fig. 1a–d).

**Trans compartments form during OSN differentiation**

First, we analysed fluorescence-activated cell (FAC)-sorted mOSNs, which represent terminally differentiated, post-mitotic neurons that are heterogeneous with regards to OR identity. In situ Hi-C on mOSNs revealed extensive interchromosomal interactions corresponding to 35.6% of total Hi-C contacts (Extended Data Fig. 1e), consistent with previous imaging results \(^19\). Zoomed-in genomic views show strong, OR-specific \(\text{trans}\) contacts between OR gene clusters (Fig. 1a, b) with the median OR cluster having approximately 7.5% of all its Hi-C contacts mapping to OR clusters from different chromosomes (Extended Data Fig. 1f). Aggregate peak analysis (APA) and unbiased compartment prediction \(^15\) (Fig. 1b, Extended Data Fig. 1g, h) confirmed that most OR clusters participated in the assembly of OR-selective multi-chromosomal compartments. Notably, \(\text{trans}\) contacts between OR clusters represented only 0.25% of all interchromosomal contacts in mOSNs, but accounted for 50% of the 1,000 strongest \(\text{trans}\) Hi-C contacts (Extended Data Fig. 1i). In horizontal basal cells (HBCs), the quiescent stem cells of the MOE, \(\text{trans}\) OR contacts were almost absent, representing only 2% of the strongest 1,000 \(\text{trans}\) contacts genome-wide, whereas intercluster \(\text{cis}\) OR contacts were strong but less specific than in mOSNs (Extended Data Fig. 2a–c, g–j). In the more-differentiated immediate neuronal precursors (INPs) \(^15\), \(\text{trans}\) OR contacts were abundant but less frequent than in mOSNs (Extended Data Fig. 2d–f, j). Thus, OR gene compartments form in a hierarchical fashion, with \(\text{cis}\) contacts appearing first and \(\text{trans}\) interactions strengthening with differentiation (Extended Data Fig. 2j–o). In vitro bacterial artificial chromosome (BAC) Hi-C and in silico Hi-C assays showed that intrachromatid Hi-C fragments do not map to other OR clusters, excluding the possibility of homology-derived mapping artefacts (Extended Data Fig. 3).

Within OR compartments, the 63 euchromatic Greek islands represent Hi-C ‘hotspots’ of specific and frequent \(\text{cis}\) and \(\text{trans}\) contacts (Fig. 1d, e, Extended Data Fig. 4a, b). Similar to OR interactions, \(\text{trans}\) Greek island contacts are not detected in HBCs (Extended Data Fig. 4e, f), which do not express ORs. By contrast, in INPs, in which multiple OR genes are weakly transcribed in each cell \(^2,20–22\), Greek islands interact with each other but lack the focal contact distribution detected in mOSNs (Extended Data Fig. 4c, d). The differentiation-dependent enhancement and specification of \(\text{trans}\) interactions is a property of most Greek islands (Fig. 1f, g, Extended Data Fig. 4g, h). In total, 4.5% of Greek island Hi-C contacts in mOSNs are made with the other Greek islands, with half of these contacts being \(\text{trans}\) (Fig. 1f). Notably, this exceeds the mean and cumulative...
frequency of contacts that Greek islands make with LHX2 and EBF co-bound intergenic sequences present in cis (Fig. 1g, Extended Data Fig. 4i, j), consistent with the differentiation-dependent assembly of a multi-chromosomal enhancer hub composed exclusively of Greek islands.

**Greek islands promote compartmentalization**

To mechanistically dissect Greek island interactions, we explored the role of the core sequences of these enhancers. In situ Hi-C in mOSNs carrying homozygous deletions for islands H18 (2 kb), Lipsi6 (1 kb), and Sfaktiria (0.6 kb) (triplet enhancer deletions) showed strong reductions in trans interactions between genomic bins containing these deletions and the remaining Greek islands, an effect that extended over large genomic distances (Fig. 2a–c, Extended Data Fig. 5a, b). Notably, the reduction in cumulative trans Greek island contacts correlates with the transcriptional downregulation of OR genes observed in cells in which Greek islands have been deleted (Fig. 2c). If we exclude Greek island bins from this analysis, we also observe reduction in trans OR contacts (Fig. 2c, d, Extended Data Fig. 5c). Thus, DNA elements as small as 0.6 kb coordinate genomic contacts extending over hundreds of kilobases, similarly to ZIP elements that affect nuclear positioning in yeast or the Igk enhancer that affects the positioning of immunoglobulin loci in pre-B cells. The partial effects of the triple enhancer deletions on cluster-wide contacts suggest that additional sequences participate in interactions between OR clusters.

**Protein regulators of compartmentalization**

We next examined the role of Greek-island-bound transcription factors in compartmentalization of ORs. We deleted Lhx2 in HBCs, and then used methimazole to induce these cells to differentiate25, 26. Using TdTomato intensity as a marker, we identified two distinct cell populations, the dimmest of which consisted of HBC-derived INPs and mOSNs (Extended Data Fig. 5d, e). RNA sequencing (RNA-seq) of the FAC-sorted cells showed that early deletion of Lhx2 caused a developmental delay in the OSN lineage and an increase in INP-specific markers (Extended Data Fig. 5f). With differentiation deficits and possible cell-identity changes taken into account, trans OR and trans Greek island contacts were strongly reduced in these cells in comparison to mOSNs and even INPs (Fig. 3a–d, Extended Data Fig. 5g). The frequency of interchromosomal interactions remained high in the early Lhx2 knockout cells, but OR–OR contacts represented only 16% of the 1,000 strongest trans contacts (Extended Data Figs. 1e, 5h). Late deletion of Lhx2 diminished trans and long-range cis contacts between Greek islands (Fig. 3b, d, Extended Data Fig. 5j), consistent with widespread OR downregulation.

To understand how LHX2 stabilizes Greek island contacts, we investigated whether LHX2 (a LIM domain protein) recruits LIM-domain-binding proteins (LDB1 and LDB2), which are known mediators of long-range genomic interactions. Chromatin immunoprecipitation and sequencing (ChIP-seq) for LDB1 and LDB2, which is the
Deletion of Greek islands disrupts local recruitment of trans Greek islands and impairs OR compartmentalization.

A, In mOSNs in which three Greek islands (H, Lipsi and Sfaktiria) are homoeurgously deleted (triple KO), the 50-kb regions containing the deleted islands have reduced trans Greek island contacts, expressed as fraction of total Hi-C contacts. Interactions among the remaining islands are not significantly different (P = 0.80, two-sided, paired Wilcoxon signed-rank test, n = 56). B, Pairwise heat map of Greek island contacts reveals that the 50-kb regions containing the deleted Greek islands (arrowheads) exhibit reduced contacts, plotted as log2contacts in KO/contacts in control, across the full set of Greek islands. Greek islands are ordered by genomic position and colour bar indicates chromosome.

C, The OR gene cluster containing Lipsi makes fewer Hi-C contacts with trans Greek islands and OR gene clusters in triple knockout mOSNs than in control mOSNs. Count data for trans Greek island contacts and trans OR cluster contacts from two biological replicates were analysed to identify loci with a significant difference in contacts between conditions (see Methods). Significantly changed regions, corrected for multiple comparisons, are indicated with asterisks (Padj < 0.05, Wald test). Lower panel, RNA-seq analysis of the expression of OR genes in triple knockout mOSNs relative to control mOSNs. Significantly changed ORs are red (P < 0.01, Wald test, five biological replicates for control mOSNs and four for knockout mOSNs). D, OR gene clusters containing the deleted Greek islands (red) make fewer contacts with trans OR gene clusters in triple knockout mOSNs, plotted as fraction of the total Hi-C contacts. Contacts made by the non-targeted clusters are not significantly different (P = 0.79, two-sided, paired Wilcoxon signed-rank test, n = 64).

Fig. 2 | Deletion of Greek islands disrupts local recruitment of trans Greek islands and impairs OR compartmentalization.

a, b, c, d, Only family member expressed in mOSNs (Extended Data Fig. 6a, b), revealed close overlap with LHX2 peaks in mOSNs (Extended Data Fig. 6c–e). Consistent with this, every Greek island was bound by LDB1 in an LHX2-dependent fashion (Extended Data Fig. 6f). Greek islands represented some of the strongest LDB1 peaks in the genome, suggesting that LHX2 and EBF act synergistically in recruitment of LDB1 (Extended Data Fig. 6g, h). Greek islands and OR clusters are not bound by CTCF and the cohesin subunit RAD21 (Extended Data Fig. 6i, j), which is not surprising given the inhibitory role of cohesin complexes in the formation of genomic compartments35,36. Finally, there is very little LDB1 signal on OR gene promoters (Extended Data Fig. 6k), even for the active Olfr1507 promoter in OLFR1507+ OSNs (Extended Data Fig. 6l). Deletion of Ldb1 in mOSNs (Extended Data Fig. 7a, b) caused strong reduction in trans and long-range cis Greek island interactions (Fig. 4a, b, Extended Data Fig. 7c–f), a smaller decrease in trans contacts between OR clusters (Extended Data Fig. 7g, h) and even weaker genome-wide effects in trans (Extended Data Fig. 1e). Notably, RNA-seq showed that deletion of Ldb1 caused widespread transcriptional downregulation of ORs (Fig. 4c) that appeared to be highly restricted to the OR gene family (Fig. 4d, Extended Data Fig. 7i).

Fig. 3 | LHX2 is essential for the formation of OR compartments and the assembly and stability of Greek island hubs.

A, Pairwise views of Hi-C contacts between OR clusters located on different chromosomes in control (top), early Lhx2 knockout (middle) and late Lhx2 knockout (bottom) OSNs. A Hi-C hotspot between interacting Greek islands in control mOSNs (arrowhead) is absent in both early and late Lhx2 knockout cells. In addition, a strong reduction in the surrounding OR–OR contacts is observed in the early Lhx2 knockout cells. B, Pairwise heat maps of Greek island contacts reveal reduced Hi-C contacts across the full set of Greek islands in early and late Lhx2 knockout cells. C, Contacts made by each OR cluster (n = 67) to OR clusters located in trans, expressed as fraction of the total Hi-C contacts, in control mOSNs versus INPs, early Lhx2 knockout cells, and late Lhx2 knockout cells. Dashed lines are linear fits. D, As in c, but for trans contacts between Greek islands (n = 59). All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately.
Greek island hubs contact only the active OR gene

To test whether Greek island hubs regulate OR transcription by direct interaction with the chosen OR gene, we performed in situ Hi-C in OSNs expressing the OR genes Olfr16, Olfr17 and Olfr1507. In these OSN populations, the overall network of OR cluster and Greek island interactions was largely the same (Extended Data Fig. 8a–d) but OSN-type-specific variability was also observed (Extended Data Figs. 8e–m, 9a, b). However, in each OSN type, the transcriptionally active OR consistently formed frequent interactions with Greek islands. For example, in OLFR16+ OSNs the Olfr16 locus interacted strongly (5% of total Hi-C contacts mapped on Olfr16) with long-range \( \text{cis} \) and \( \text{trans} \) Greek islands (Fig. 5a, b, Extended Data Fig. 9c, d), whereas in OLFR17+ and OLFR1507+ OSNs it interacted primarily with nearby Greek islands (Fig. 5a, b). Notably, in OLFR16+ cells, Greek island contacts were enriched specifically over the Olfr16 locus (Fig. 5b) relative to the full OR repertoire (Fig. 5c). Thus, in situ Hi-C accurately identifies the position (5-kb resolution), between the Olfr16 locus and Greek islands in OLFR16-expressing cells. c, Manhattan plot of Greek island contacts with OR genes reveals that in OLFR16+ cells the Olfr16 locus is the OR gene most significantly enriched for Greek island contacts relative to heterogeneous mOSNs (see Methods). All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately.

Fig. 4 | LDB1 is essential for the stability of Greek island hubs and for OR transcription. a, Pairwise heat map of Greek island contacts reveals broad reductions in Hi-C contacts in \( \text{Ldb1} \) knockout mOSNs. b, Left, for each Greek island, the fraction of total Hi-C contacts made to other Greek islands located in \( \text{cis} \) at short range (<5 Mb apart, grey), long range (>5 Mb apart, blue) and in \( \text{trans} \) (red). Top, control mOSNs; bottom, \( \text{Ldb1} \) knockout cells. Right, the effect of \( \text{Ldb1} \) knockout on the mean fraction of Hi-C contacts across all Greek islands (two-sided, paired Wilcoxon signed-rank test, \( n = 59 \)). c, RNA-seq analysis of gene expression in \( \text{Ldb1} \) knockout cells relative to control mOSNs. Significantly changed genes are coloured red (\( P_{\text{adj}} < 0.05 \) for greater than 1.5-fold change, Wald test, \( 10^{-5} \log_2(KO/\text{control}) \)). d, Effect of \( \text{Ldb1} \) knockout on genes not associated with LDB1 ChIP–seq peaks (\( n = 9,548 \)), genes located closest to a non-promoter LDB1 ChIP–seq peak (\( n = 5,624 \)), genes with an LDB1 ChIP–seq peak within the promoter region (\( n = 1,640 \)) and ORs (\( n = 1,135 \)). The percentage of significantly changed genes in each category is shown (\( P_{\text{adj}} < 0.05 \) for greater than 1.5-fold change, Wald test, \( n = 5 \) for control mOSNs and \( n = 4 \) for \( \text{Ldb1} \) knockout cells). Box indicates median, upper, and lower quartiles; whiskers indicate 1.5 \( \times \) the interquartile range. a and b present pooled data from two independent biological replicates that yielded similar results when analysed separately.

Fig. 5 | Greek island hubs interact specifically with the transcriptionally active OR locus. a, Increased contacts between the active OR promoter and Greek islands located in short-range \( \text{cis} \) (<5 Mb, grey), long-range \( \text{cis} \) (>5 Mb, blue) and \( \text{trans} \) (red). Greek island interactions are expressed as the fraction of the total Hi-C contacts mapped to each promoter (5-kb resolution). b, Profile of the OR cluster containing Olfr16 reveals increased contacts, expressed as fraction of the total Hi-C contacts mapped to each position (5-kb resolution), between the Olfr16 locus and Greek islands in OLFR16-expressing cells. c, Manhattan plot of Greek island contacts with OR genes reveals that in OLFR16+ cells the Olfr16 locus is the OR gene most significantly enriched for Greek island contacts relative to heterogeneous mOSNs (see Methods). All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately.
transcriptionally active OR from a pool of more than 1,000 genes through its cumulative interactions with Greek islands (Fig. 5c, Extended Data Fig. 9e–h).

Discussion

Our results have revealed a type of genomic compartment with multichromosomal composition and very high exclusivity. Genomic compartments represent more complex assemblies than segregation products of transcriptionally active and inactive chromatin6,7,37. However, the demonstration that more than 1,000 genes from 18 chromosomes form exclusive compartments implies that this process is precisely regulated, comparable with the assembly of the nucleolus38. Unlike the nucleolus, however, OR compartments and Greek island hubs are regulated by proteins with widespread binding in the OSN genome. In the absence of an OR-specific factor that would explain the specificity of OR contacts, we propose that Greek islands bound by LHX2, EBF and LDB1 and OR heterochromatin create a unique molecular ‘barcode’ that assembles OR-specific compartments. These heterochromatin compartments may achieve efficient OR silencing through the phase separation properties of HP139,40, but they also confine in close proximity Greek islands from different chromosomes6,8, forcing them to interact. As proposed for super-enhancers41,42, this confinement may promote an adjacent euchromatin phase consisting of locally concentrated activators. Where the two phases are incompatible, the Greek island hub would insulate the active OR allele from the surrounding repressive environment, resulting in stable OR choice (Extended Data Fig. 10). Given that this multi-chromosomal super-enhancer interacts only with the single chosen OR gene, and its disruption perturbs OR transcription, interchromosomal interactions emerge as essential regulators of OR transcription6,7,43. This concept of trans enhancement was initially challenged by the cis-only effects of enhancer deletions14,44,45. However, the demonstration that Greek islands promote OR compartmentalization and recruit trans Greek islands towards proximal ORs explains why these elements are essential cis, but redundant trans, enhancers. As long-range genomic interactions have been implicated in transcriptional stochasticity12,46, cell-type-specific interchromosomal contacts may serve as an additional generator of multicellular diversity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0845-0.

Received: 9 April 2018; Accepted: 6 December 2018; Published online 9 January 2019.
52. Zhao, Y. et al. LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. *Proc. Natl Acad. Sci. USA* **104**, 13182–13186 (2007).

53. Mangale, V. S. et al. Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science* **319**, 304–309 (2008).

54. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).

55. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

56. Droettboom, M. matplotlib/matplotlib v2.2.2 https://zenodo.org/record/1202077 (2018).

57. Waskom, M. mwaskom/seaborn: v0.8.1 https://zenodo.org/record/883859 (2017).

58. McKinney, W. Data Structures for Statistical Computing in Python. *Proc. 9th Python Sci. Conf*. 1697900, 51–56 (2010).

59. Freese, N. H., Norris, D. C. & Loraine, A. E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* **32**, 2089–2095 (2016).

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

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

62. Klein, F. A. et al. FourCSeq: analysis of 4C sequencing data. *Bioinformatics* **32**, 841–842 (2010).

63. Pedregosa, F. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **28**, 2224–2230 (2017).

64. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

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

66. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

67. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

68. Benjamini, Y. & Speed, T. P. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72 (2012).

Acknowledgements We thank A. Mumbay-Wafula for assistance with mouse colony management, I. Schieren for assistance with fluorescence-activated cell sorting, P. Love for sharing the Ldb1 fl/fl mice and E. Monuki for sharing the Lhx2 fl/fl mice. A.H. was funded by F31 post-doctoral fellowship DC016785 (NIH) and K.M. was funded by F32 post-doctoral fellowship GM108474 (NIH). S.L. acknowledges support from the NIH Common Fund 4D Nucleome Program (grant 1U01DA040582). In addition, this project was funded by R01DC013560, R01DC015451 (NIH) and the HHMI Faculty Scholar Award. Research reported in this publication was also performed in the CTTI Flow Cytometry Core, supported in part by the Office of the Director, National Institutes of Health under award S10OD020056. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Author contributions K.M, A.H., and S.L. designed the study. K.M. performed in situ Hi-C in mOSNs, INPs, HBCs, and OLFR1507, OLFR16, and OLFR17-expressing mOSNs and methimazole-treated cells from the MOE. A.H. performed in situ Hi-C in wild-type and Lhx2 knockout mOSNs, and performed RNA-seq in Ldb1 knockout and control Ldb1 and Lhx2 knockout mice. A.H. was funded by F31 post-doctoral fellowship DC016785 (NIH) and K.M. was funded by F32 post-doctoral fellowship GM108474 (NIH). S.L. acknowledges support from the NIH Common Fund 4D Nucleome Program (grant 1U01DA040582). In addition, this project was funded by R01DC013560, R01DC015451 (NIH) and the HHMI Faculty Scholar Award. Research reported in this publication was also performed in the CTTI Flow Cytometry Core, supported in part by the Office of the Director, National Institutes of Health under award S10OD020056. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Competing interests The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0845-0.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0845-0.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.L.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS
No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Mice. Mice were treated in compliance with the rules and regulations of IACUC under protocol number AC-AAAT2450. Mice were killed using CO2, followed by cervical dislocation. Both male and female mice were used for experiments. All experiments were performed on dissected olfactory epithelium tissue or on dissociated cells prepared from whole olfactory epithelium tissue. Dissociated cells were prepared using papain (Worthington Biochemical) and FAC-sorted as previously described.

This study used several mouse lines to allow isolation of cells at specific stages of olfactory sensory neuron (OSN) development, OSNs that express one of three specific olfactory receptors, and cells with specific targeted mutations. Mature OSNs (mOSNs) were sorted from adult (8–12 weeks of age) Omp-ires-gfp mice. Neural progenitors (INPs) were isolated by sorting the brightest of two GFP populations from 1–2-week-old Ngn1-gfp mice. The dim population of Ngn1–GFP cells represents a more mature population of OSNs, as determined by RNA-seq (data not shown). Multipotent olfactory progenitors (horizontal basal cells) were isolated by injecting perinatal Krt5-CreERT2, B6N.129S6-G(Rosa)26Sor1(CAG-CAG-Tomato, EApG)(test) mice16 with tamoxifen 24 and 48 hours before sorting GFP-positive, tdTomato-negative cells. OLFRI17+ cells were sorted from adult Olf1r1507-ires-gfp mice. OLFRI6+ cells were sorted from adult Olf16-16-tausqf (Olf16tm2Ryoll) mice23. Triple enhancer knockout mice were generated by crossing mice bearing three individual Greek island deletions (H38, Lipsi5 and Sfaktiria) and Omp-ires-gf and sorting for GFP + mature OSNs from adult mice. The Sfaktiria deletion was generated by Biozycogen using TALENs to target the region chr6:42869802-42870400 (mm10).

Conditional deletion of Lhx2 early in mOSN differentiation was achieved by crossing Lhx2 conditional allele mice to mice bearing Krt5-CreERT2 and Cre-inducible tdTomato (ROSA26-tdtomato, Gt(Rosa)26Sor1(CAG-CAG-Tomato, EApG)(test)). At 6 weeks of age, deletion of the conditional allele in horizontal basal cells was induced by two intraperitoneal injections of tamoxifen 24 hours apart. One week later, differentiation of horizontal basal cells into olfactory cell types was induced by intraperitoneal injection with methimazole, which triggers afoiation of olfactory epithelium and regeneration of the tissue from horizontal basal cells. The olfactory epithelium was allowed to regenerate for eight weeks, producing bright tdTomato+ cells that localized to the basal (HBCs) and apical (sustentacular cells) layers of the MOE, and dim tdTomato+ cells that populated the neuronal cell layers of the MOE. Fluorescence-activated cell sorting (FACS) of the bright and dim populations separately, followed by RNA-seq, confirmed that the dim cell population consisted mostly of mOSNs and INPs (Extended Data Fig. 3e, f).

Conditional alleles were deleted specifically in mOSNs using Omp-ires-Cre22 mice. Conditional deletion of Lhx2 in mOSNs was achieved by crossing Lhx2 conditional allele mice35 (Lhx2-2–Fluc, Lhx2CreMm1Susa) with Cre-inducible tdTomato and Omp-ires-Cre. Similarly, conditional deletion of Ldb1 in mOSNs was achieved by crossing Ldb1 conditional allele mice45 (Ldb1-1–Fluc, Ldb1tm1Kpsnp) with Cre-inducible tdTomato and Omp-ires-Cre mice. Recombined cells were purified by selecting tdTomato-positive cells by FACS from adult mice.

Fluorescence-activated cell sorting. Cells were dissociated into a single-cell suspension by incubating freshly dissected MOE with papain for 40 min at 37 °C according to the Worthington Papain Dissociation System. Following dissociation and filtering three times through a 35-μm cell strainer, live cells were sorted by collecting fluorescent, DAPI-negative cells for RNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq). Alternatively, cells were fixed with 1% PFA in PBS for 5 min (ChIP) or 10 min (Hi-C) at room temperature. Fixed fluorescent cells were then sorted on a BD Aria II, BD Influx or Beckman Coulter MoFlo Astrios EQ cell sorter. Representative FACS plots for the cells used in this study are available at https://data.4dnucleome.org/search/lab/display_title=Stavros+Lomvardas%20COLUMBIA&protocol_type=Cell%20sorting%20protocol&type=Protocol.

In situ Hi-C. Depending on the genotype, between 20 thousand and 3 million cells were used for in situ Hi-C. Sorted cells were lysed and intact nuclei were processed through an in situ Hi-C protocol as previously described13 with a few modifications. In brief, cells were lysed with 50 mM Tris pH 7.5 0.5% Igepal, 0.25% sodium deoxycholate 0.1% SDS, 150 mM NaCl and protease inhibitors. Pelleted intact nuclei were then resuspended in 0.5% SDS and incubated for 20 min at 65 °C for nuclear permeabilization. After being quenched with 1.1% Triton-X for 10 min at 37 °C, nuclei were digested with 6 U/ml DpnII in 1 x DpnII buffer over-night at 37 °C. Following initial digestion, cells were pelleted (2.5g/50 ml), buffor replaced to 1 x DpnII containing the fresh DpnII was added at 30°C for an additional 2 h of digestion. Following digestion, the restriction enzyme was inactivated at 65 °C for 20 min. For the 1.5-h fill-in at 37 °C, biotinylated dGTP was used instead of dATP to increase ligation efficiency. Ligation was performed at 25 °C for 4 h with rotation. Nuclei were then pelleted and sonicated in 10 mM Tris pH 7.5, 1 mM EDTA, 0.25% SDS on a Covaris S220 for 16 min with 2% duty cycle, 105 intensity, 200 cycles per burst, 1.8–1.85 W, and maximum temperature of 6 °C. DNA was reverse cross-linked overnight at 65 °C with proteinase K and RNase A. Each experiment was performed with two biological replicates.

Hi-C library preparation and sequencing. Reverse cross-linked DNA was purified with 2× Ampure beads following the standard protocol and eluted in 300 μl water. Biotinylated fragments were enriched as previously described15 using Dynabeads MyOne Streptavidin T1 beads. The biotinylated DNA fragments were prepared for next-generation sequencing directly on the beads by using the Nugen Ovation Ultralow kit protocol with some modifications. Following end repair, magnetic beads were washed twice at 55 °C with 0.05% Tween, 1 M NaCl in Tris/EDTA pH 7.5, instead of heat-inactivating end-repair enzymes. Residual detergent was removed by washing beads twice in 10 mM Tris pH 7.5. End repair buffers were replenished to original concentrations, but the enzyme and enhancer were omitted before adaptor ligation. Following adaptor ligation, beads underwent five washes with 0.05% Tween, 1 M NaCl in Tris/EDTA pH 7.5 at 55 °C and two washes with 10 mM Tris pH 7.5 to remove ligation enzymes and buffers. DNA was amplified by 10 cycles of PCR. Beads were reclaimed and amplified unbiotinylated DNA fragments were purified with 0.8× Ampure beads. The quality and concentration of libraries were assessed using Agilent Bioanalyzer and KAPA Library Quantification Kit. Hi-C libraries were sequenced paired-end on NextSeq 500 (2 × 75 bp), or NovaSeq 6000 (2 × 300 bp).

A full protocol and gel electrophoresis of a typical Hi-C experiment is available at https://data.4dnucleome.org/search/lab/display_title=Stavros+Lomvardas%20COLUMBIA&protocol_type=Experimental+protocol&type=Protocol.

Hi-C data processing pipeline. Raw fastq files were processed using the Juicer Tools Version 1.76 pipeline with one modification. Reads were aligned to mm10 using BWA 0.7.17 mem algorithm and specifying the -5 option implemented specifically for Hi-C data. The -5 option always takes the leftmost alignment (5′) on a read as the primary read. This alignment gets its own alignment score independent of subsequent alignments. Following alignment, independently mapped reads are merged to generate chimeric reads. After reads are aligned, merged, and sorted, chimaeras are de-duplicated and finally Hi-C contact matrices are generated by binning at various resolutions and matrix balancing. Importantly, all reads mapping to multiple locations are discarded as ‘chimeric ambiguous reads.’ To remove multi-mapers, we used a stringent cutoff of MAPQ > 30. All data used in this paper, including data generated by other groups, were aligned in this way.

Hi-C data analysis. Hi-C matrices used in this paper were matrix-balanced using Juicer’s built-in Knight-Ruiz (KR) algorithm. Where noted, values were instead normalized to target counts/total Hi-C contacts for that bin at a specified resolution (for example, per cent OR contacts/total Hi-C contacts per bin). This accounts for terminal read alignment depth of a given bin. Matrices were graphed using pandas, seaborn and matplotlib57–59 packages for python, or R-Studio Server (R version 3.5.1).

Genome wide Hi-C maps were constructed from KR-normalized matrices at 1-Mb resolution and normalized to library size. The maximum value of the colour scale was set to 1,000 reads per billion Hi-C contacts per 1-Mb bin. Cumulative interchromosomal contacts at the resolutions noted in the text were constructed by calling Juicer Tools dump to extract genome wide un-normalized data from a .hic file. Subsequently, single-ended bins for regions of interest were selected for genome-wide interchromosomal counts. Counts pertaining to a particular bin were divided by the total Hi-C contacts sequenced for the respective bin. These normalized counts were then aggregated per genomic bin to construct a bedGraph and visualized using Integrated Genome Browser60. Alternatively, all bins contacted by a bin of interested were categorized by genomic location (for example, Greek islands overlapping, OR cluster overlapping, intergenic EBF/LHX2 peak overlapping) and then counts were aggregated by category. For 50-kb and 25-kb analyses, only the bin directly overlapping a feature (for example, a Greek island) was assigned to that category. For 5-kb resolution analyses the bin containing a feature and the two bins directly upstream and downstream were assigned to that feature category. Aggregate counts were converted to fraction of Hi-C contacts by dividing by the total number of Hi-C contacts made by the bin of interest. Mean counts per interaction were determined by dividing the aggregate counts for each category (for example, Greek island overlapping, OR cluster overlapping, and so on) by the number of bins that matched that category present in cis or in trans.

APA was done using Juicer Tools. Normalized APA matrices were graphed with the maximum scale set to five times the mean of the matrix. OR gene cluster contact matrices were constructed by extracting pairwise contact counts defined for each cluster (cluster 1 × size of cluster 2) of the respective pairwise OR gene cluster interaction. The logarithm of these values was then taken to account for the strength of cis interactions and plotted.
Specific OR gene cluster contacts were constructed through programmatic access to .hic files using straw for python. These matrix files can also be used to form 3D contour maps with the same software to better visualize the focal peaks in the contact matrix. KR-normalized matrix values were further normalized by dividing by Hi-C library size for directly comparing samples.

For box plots quantifying the strength of interchromosomal interactions, the box indicates median and upper and lower quartiles; the whiskers indicate 1.5 × the interquartile range. Outliers are not shown.

DESeq2 was used to detect differences between conditions for individual sites. A similar approach has previously been used to analyse count data from 4C-seq. The raw, un-normalized number of Hi-C contacts mapping to OR clusters located in trans or to Greek islands located in trans was determined for every region of the genome at a given resolution (25-kb bins). For each condition, counts from two biological replicates were analysed using DESeq2. Regions with zero counts in any condition were excluded. DESeq2 identifies regions where the observed change in counts between conditions is significantly greater than amount of change expected based upon an analysis of variance between replicates. For the analysis of triple enhancer knockout mOSNs compared to control mOSNs, we used for all of our datasets.

Following generation of in silico Hi-C fastqs, we aligned our data using the same pipeline we used for all of our datasets.

To address potential mapping issues by an orthogonal computational approach, we performed in silico Hi-C on BAC clone RP23-374F2, a 165-kb clone containing mostly OR sequences but also non-OR sequences. The Hi-C protocol is analogous to our experimental Hi-C. In brief, we digested the BAC clone with DpnII, filled-in overhangs with DNA Pol I Klown fragment, performed in vivo digestion with Tf4 ligases and purified DNA using Illumina Truseq Stranded RNA-seq Gold kits. All datasets were processed using 50 bp of single end; 75-bp reads were trimmed to 50 bp and only read 1 was used from paired-end data. Cutadapt was used to remove adapter sequences from raw sequencing data and then filtered reads were aligned to the mouse genome (mm10) using STAR. Peaks of ChIP-seq signal were identified using DiffBind and were used to calculate the ChiP–seq signal in each peak. For this analysis, DiffBind was used to normalize ChiP–seq scores across biological replicate experiments using the ‘DBA_SCORE_TMM_READS_EFFECTIVE’ scoring system, which normalizes using edgeR and the effective library size. The ChiP–seq signal for each peak was then calculated by averaging the normalized score across biological replicates.

ATAC-seq. ATAC-seq data were analysed as previously described. RNA-seq. See Supplementary Information 2 for a summary of RNA-seq sequencing data. RNA-seq experiments were conducted as previously described. In brief, RNA was extracted from FACS-purified cells using Trizol and libraries were prepared using Illumina TruSeq Stranded RNA-seq Gold kits. All datasets were processed using 50 bp of single end; 75-bp reads were trimmed to 50 bp and only read 1 was used from paired-end data. Cutadapt was used to remove adapter sequences from raw sequencing data and then filtered reads were aligned to the mouse genome (mm10) using STAR. RNA-seq data analysis was performed in R with the DESeq2 v1.20.0 package. Very low-abundance transcripts (genes with fewer than 10 counts combined across all samples) were excluded. DESeq2 was used to calculate normalized counts (regularized log transformed), FPKM values, fold change values, P values, and P values adjusted for multiple comparisons (P_adj).

Immunofluorescence. MOE was dissected from 6-week-old Ldb1 KO (Ldb1<sup>fl/fl</sup>; Omp-cre) mice and littermate controls. MOE tissue was embedded in OCT and then coronal cryosections were collected at a thickness of 12 µm. Tissue sections were prepared and stained as previously described. Tissue sections were stained with primary antibodies against LDB1 (1:1,000 dilution, Santa Cruz Biotechnology Cat. no. sc-11198, RRID:AB_2288368, RRID:AB_650839). DNA was labelled with DAPI (2.5 µg/ml, Thermo Fisher Scientific Cat. no. D3571).

Consensus peak sets were generated by selecting peaks that overlapped in at least two biological replicates and extending them to their combined size. Bedtools<sup>27</sup> v2.26.0 was used to compare peak sets.

For signal tracks, biological replicates were merged and HOMER was used to generate 1-bp resolution signal tracks normalized to a library size of 10,000,000 reads. Values in all ChiP–seq signal plots are counts per 10 million reads. Plots of ChiP–seq signal over individual loci were generated using the UCSC Genome Browser. Deeptools<sup>28</sup> v3.1.1 was used to generate ChiP–seq heat maps and mean signal plots. For heat maps, each row of the heat map is an 8-kb region centred on a Greek island or ChIP–seq peak for the factor shown. For the heat map in Extended Data Fig. 6j, all Greek islands are shown alongside 500 randomly selected ChiP–seq peaks for each factor. For Extended Data Fig. 6k, each row corresponds to an OR gene, showing 1 kb upstream of the transcriptional start site, 1 kb downstream of the transcriptional end site, and the gene body scaled to 2 kb. Signal plots present average data for all regions each set. Heat maps are sorted by mean signal.

DiffBind<sup>30</sup> v2.8.0 was used to calculate the ChiP–seq signal in each peak. For this analysis, DiffBind was used to normalize ChiP–seq scores across biological replicate experiments using the ‘DBA_SCORE_TMM_READS_EFFECTIVE’ scoring system, which normalizes using edgeR and the effective library size. The ChiP–seq signal for each peak was then calculated by averaging the normalized score across biological replicates.

ATAC-seq. ATAC-seq data were analysed as previously described<sup>27</sup>. RNA-seq. See Supplementary Information 2 for a summary of RNA-seq sequencing data. RNA-seq experiments were conducted as previously described. In brief, RNA was extracted from FACS-purified cells using Trizol and libraries were prepared using Illumina TruSeq Stranded RNA-seq Gold kits. All datasets were processed using 50 bp of single end; 75-bp reads were trimmed to 50 bp and only read 1 was used from paired-end data. Cutadapt was used to remove adapter sequences from raw sequencing data and then filtered reads were aligned to the mouse genome (mm10) using STAR. RNA-seq data analysis was performed in R with the DESeq2 v1.20.0 package. Very low-abundance transcripts (genes with fewer than 10 counts combined across all samples) were excluded. DESeq2 was used to calculate normalized counts (regularized log transformed), FPKM values, fold change values, P values, and P values adjusted for multiple comparisons (P_adj).

Immunofluorescence. MOE was dissected from 6-week-old Ldb1 KO (Ldb1<sup>fl/fl</sup>; Omp-cre) mice and littermate controls. MOE tissue was embedded in OCT and then coronal cryosections were collected at a thickness of 12 µm. Tissue sections were prepared and stained as previously described<sup>27</sup>. Tissue sections were stained with primary antibodies against LDB1 (1:1,000 dilution, Santa Cruz Biotechnology Cat. no. sc-11198, RRID:AB_2288368, RRID:AB_650839). DNA was labelled with DAPI (2.5 µg/ml, Thermo Fisher Scientific Cat. no. D3571).

Confocal images were collected with a Zeiss LSM 710 and image processing was carried out with Imaris (NIH).

Statistics. A sample size of two independent biological replicates was selected for high-throughput sequencing experiments. This size was selected because the large number of genes or loci measured in high-throughput sequencing datasets allows the analysis and modelling of dispersion and variance within and between replicates, thereby allowing the identification of genes or loci with significant differences between conditions using a limited number of replicates. When possible, additional biological replicates were included.

For ChiP–seq, statistically significant peaks were identified using HOMER on each replicate dataset. Each replicate peak set was selected by setting a read count threshold based upon an input control false discovery rate of 0.001, and then peaks were filtered based upon the following criteria: Poisson P value over input <1.00 × 10<sup>-4</sup> and Poisson P value over local region <1.00 × 10<sup>-4</sup>. Consensus peak

© 2019 Springer Nature Limited. All rights reserved.
sets were then generated by selecting peaks that overlapped in at least two biological replicates. A two-tailed Wilcoxon rank-sum test was used to determine whether there was a significant difference in the median ChIP–seq peak strength between sets of peaks. For RNA-seq, five biological replicates of control mOSNs, four biological replicates of triple enhancer knockout mOSNs, and four biological replicates of Ldb1 knockout mOSNs were analysed with DESeq2, which generates two-tailed Wald test $P$ values, and generates adjusted $P$ values using the Benjamini–Hochberg method. For Hi–C data, two independent biological replicates were generated for each condition and analysed separately. Individual biological replicates yielded similar results and were pooled for the analyses presented here. A paired, two-tailed Wilcoxon rank-sum test was used to determine whether the mean frequency of Hi–C contacts for the set of Greek islands was different between conditions.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

**Data availability**
All Figures include publicly available data. All ChIP–seq and RNA-seq data reported in this paper (see Supplementary Information 1, 2) are available from GEO under accession number GSE112153. Additional data (mOSN RNA-seq, mOSN LHX2 ChIP–seq, mOSN EBF ChIP–seq and OLFR1507+ ATAC-seq) were previously described and are available from GEO under accession number GSE93570. All Hi–C data generated in this study are publicly available at https://data.4dnucleome.org/ under the following accession numbers: 4DNESH4UTRNL, 4DNESNYBDSLX, 4DNES54YB6TQ, 4DNESRE7AK5U, 4DNES425UDGS and 4DNESEPDL6KY.

72. Yan, J. et al. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res.* **28**, 387 (2018).

73. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
Extended Data Fig. 1 | See next page for caption.
Extended Data Fig. 1  |  Hi-C on FAC-sorted primary cells from the MOE reveals extensive interchromosomal interactions between OR clusters. a, Table summarizing all Hi-C experiments in this manuscript separated by biological replicates. The total number of Hi-C contacts in each replicate and the total number of interchromosomal (trans) Hi-C contacts are shown. b–d, Hi-C contact curves for wild-type conditions (b), for wild-type and mutant MOE populations (c), and for cells sorted on the basis of expression of specific OR genes (d). All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately. e, Graphs showing the proportion of trans Hi-C contacts between replicates of each genotype and cell type. Pooled data from publicly available datasets is shown for ES cells12, B cells15 and cortical neurons73. f, As in e, but showing the median fraction of Hi-C contacts made to trans OR clusters for OR cluster regions divided into 50-kb bins. g, Machine-learning HMM score for a given number of compartments (see Methods). Nine compartments were used for further analysis. h, From the nine HMM-derived compartments, one includes predominantly OR clusters (magenta, bottom) and overlaps with OR compartments defined by biased analysis of trans OR contacts (black, top). OR gene clusters depicted in red. Scale on the biased analysis represents the percentage of Hi-C contacts mapped to trans OR clusters (pooled data from two biological replicates). Scale in the HMM-derived compartments represents the average value of a given locus in a given compartment. i, Circos plots depicting the strongest 1,000 interchromosomal interactions genome-wide at 1-Mb resolution in mOSNs. Red lines represent OR-to-OR contacts and black lines non-OR-to-non-OR contacts. Line thickness increases with contact frequency. Chromosome numbers given at the periphery of the circle.
Extended Data Fig. 2 | Extensive interchromosomal contacts form between OR gene clusters over OSN differentiation. a–i, In situ Hi-C contact matrices of chromosomes 2 and 9, APAs, and Circos plots depicting the strongest 1,000 interchromosomal interactions genome-wide for mOSNs (a–c), INPs (d–f) and HBCs (g–i). All three sets of analyses reveal an increase in trans OR cluster interactions over the course of differentiation. j, For OR gene clusters (divided into 50-kb bins, \( n = 768 \) bins) the frequency of cis short (<5 Mb distance, including self), cis long (>5 Mb) and trans contacts with OR clusters is shown, expressed as the fraction of total Hi-C contacts mapped to each bin. k, Number of Hi-C contacts, normalized to a library size of one billion Hi-C contacts genome-wide, observed for each OR cluster region (divided into 50-kb bins, \( n = 768 \) bins) in HBCs, INPs, mOSNs, ES cells, B cells and cortical neurons. l–n, For OR cluster regions (divided into 50-kb bins, \( n = 768 \) bins), the fraction of total Hi-C contacts that are made to ORs clusters located in short range cis (l), long range cis (m) and trans (n). o, The six most distinct HMM-derived compartments of chromosome 2 in HBCs (green, left), INPs (blue, middle) and mOSNs (magenta, right). OR clusters emerge as distinct compartment in INPs and strengthen in mOSNs. For all box plots, box indicates median, upper, and lower quartiles; whiskers indicate 1.5 × the interquartile range. All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately.
Extended Data Fig. 3 | See next page for caption.
Extended Data Fig. 3 | In vitro and in silico Hi-C experiments show that OR Hi-C contacts are generated by unique sequences that do not map to other OR clusters. **a**, Contact matrices from in vitro Hi-C (top) using a 165-kb BAC plasmid containing seven OR genes from an OR cluster from chromosome 1 and in situ Hi-C from mOSNs (bottom). Hi-C contacts in the BAC Hi-C are restricted to the coordinates of the BAC plasmid and do not extend to two OR genes from this cluster that are absent from the BAC. **b**, Virtual 4C from the 165-kb BAC region to chromosome 2, which contains the highest number of OR genes. Top, virtual 4C from the BAC in vitro Hi-C shows that no reads mapped to ORs from chromosome 2, whereas the same 165-kb regions makes abundant trans contacts with these ORs in mOSNs. **c**, Of all the BAC Hi-C contacts, 99.3% map within the BAC, whereas in mOSNs only 21.7% of the BAC region Hi-C contacts map within the BAC. **d**, In silico Hi-C analysis shows complete absence of mis-mapped reads corresponding to OR clusters under the mapping conditions used throughout the manuscript (removing mapq < 30). Each OR cluster was subjected to intracluster in silico Hi-C (**g**) and then the Hi-C contacts of the 69 OR clusters were mapped in aggregate to the whole genome. As seen in the contact matrix from chromosomes 2 and 9 (**d**), the in silico reads only map within clusters, with no mis-mapped reads that would erroneously be interpreted as intercluster cis or trans contacts. **e**, For reference, the corresponding in situ Hi-C from mOSNs. **f**, Aggregate analysis for all 69 OR gene clusters shows that our mapping protocol does not mis-map any Hi-C contacts to the wrong OR cluster. **g**, Brief description of the pipeline used for the in silico analysis.
Extended Data Fig. 4 | Greek islands make differentiation-dependent contacts with other Greek islands in trans that are stronger than cis contacts with LHX2 and EBF peaks. **a, b,** Heat maps and 3D projections of Hi-C contacts between a pair of OR gene clusters in cis (a) and trans (b) reveal a local maximum of in situ Hi-C interactions between Greek island loci (arrowheads) in mOSNs. **c–f,** As in a, b but for INPs and HBCs. **g,** For chromosome 2, fraction of all Hi-C contacts made to trans Greek islands in mOSNs (top), INPs (middle) and HBCs (bottom). **h,** For each Greek island, the distribution of Hi-C contacts—expressed as contacts per billion—made to individual Greek islands located in trans for HBCs, INPs and mOSNs. Boxes indicate median, upper, and lower quartiles; whiskers indicate 1.5 $\times$ the interquartile range. For each Greek island, the number of trans Greek islands is listed. **i,** Left, comparison of the total fraction of Hi-C contacts made by each Greek island to intergenic LHX2- and EBF-bound peaks present in cis versus Greek islands present in trans for HBCs, INPs and mOSNs. For each category, we compare roughly equal numbers of peaks (number of trans Greek islands for each island versus number of cis LHX2/EBF sites for each island, mean ± s.d.). Right, mean fraction of Hi-C contacts across all Greek islands (two-sided, paired Wilcoxon signed-rank test, $n = 59$). Contacts with trans Greek islands (red) constitute a higher fraction of Hi-C contacts than short-range cis (dark blue) or long-range cis (light blue) contacts with intergenic LHX2/EBF peaks. All panels present pooled data from two independent biological replicates that yielded similar results when analysed separately.
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5  | Greek islands and LHX2 are required for OR compartmentalization in developing OSNs. a, Pairwise Hi-C contacts between all pairs of Greek islands ordered by genomic position in control (left) and Greek island triple KO (right) mOSNs. The 50-kb regions containing the deleted Greek islands are marked with arrowheads. Plotting the log2(fold difference) in Hi-C contacts (right) reveals that consistent strong reductions are observed for the deleted islands. Colour bar depicts chromosome. b, c, The genomic regions exhibiting the most significant reductions in Hi-C contacts with trans OR Greek islands (b) or trans OR clusters (c), in triple KO mOSNs relative to control mOSNs, are mostly located within the three OR clusters containing the deleted Greek islands (two biological replicates per condition, see Methods). d, Genetic and experimental strategy for early Lhx2 deletion. Tamoxifen induction with Krt5-creER deletes Lhx2 in HBCs and then methimazole treatment ablates INPs and mOSNs, leading to regeneration from Lhx2-deleted HBCs. e, Fluorescent labelling of the HBC-derived cells upon methimazole induction reveals two major populations, bright and dim. f, RNA-seq shows that the dim population expresses markers of INPs and mOSNs; the bright population expresses markers of HBCs. Counts are normalized by row. g, Three-dimensional projection of Hi-C contacts between OR clusters located on different chromosomes in control mOSNs, INPs, early Lhx2 KO cells and late Lhx2 KO cells (from left to right). A Hi-C hotspot between interacting Greek islands is observed only in control mOSNs (arrowhead). In addition, a strong reduction in the surrounding OR–OR contacts relative to mOSNs or INPs is observed in the early Lhx2 KO cells. h, Circos plots depicting the strongest 1,000 interchromosomal interactions genome-wide at 1-Mb resolution in (left to right) mOSNs, INPs, early Lhx2 KO cells, and late Lhx2 KO cells. Red lines represent OR–OR contacts and black lines non-OR–non-OR contacts. Line thickness increases with contact frequency. Chromosome numbers depicted at the periphery of the circle. i, Genetic strategy for late Lhx2 deletion and fluorescent marking of Lhx2 KO mOSNs. j, Left, for each Greek island, the fraction of total Hi-C contacts made to other Greek islands located in cis at short range (<5 Mb apart, grey), long range (>5 Mb apart, blue), and in trans (red). Top, control mOSNs; bottom, late Lhx2 KO cells. Right, the effect of late Lhx2 KO on the mean fraction of Hi-C contacts across all Greek islands (two-sided, paired Wilcoxon signed-rank test, n = 59).
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | LDB1 expression and genomic distribution in mOSNs. a, Transcript level, expressed as fragments per kilobase per million mapped reads (FPKM), of the two LDB family members in mOSN RNA-seq datasets (n = 5 biological replicates). b, Sections of olfactory epithelium stained for LDB1 (green) and ADCY3 (magenta), a marker for mOSNs. Nuclei are labelled with DAPI (blue). Scale bar, 25 µm. Similar results were obtained from four independent experiments. c, LDB1 and LHX2 ChIP–seq signal in mOSNs across the OR gene cluster containing the Greek island Lipsi. OR genes are red and all other genes are blue. Plot shows pooled data from three biological replicates for LDB1 and two biological replicates for CTCF and RAD21. Values are counts per 10 million reads. Analysing each replicate separately yielded similar results. j, mOSN ChIP signal over Greek island and non-Greek island ChIP–seq peaks. For ChIP–seq peaks, the heat map shows 500 randomly selected peaks and the plot shows data from the full consensus set of peaks (n = 22,791 for LDB1, n = 24,883 for CTCF, and n = 9,882 for RAD21). Plots show pooled data; similar results were obtained with each replicate (n = 3 for LDB1 ChIP–seq and n = 2 for CTCF and RAD21 ChIP–seq). Units are counts per 10 million reads. k, As in j, but showing LDB1 ChIP signal over OR genes (n = 1,255) in mOSNs. l, LDB1 ChIP–seq from control mOSNs (top) and OLFR1507-expressing cells (middle). Strong signal is observed on the Greek island H, in both populations, but only a very weak signal on the Olfr1507 promoter when it is transcriptionally engaged. Pooled data from three biological replicates is shown for the mOSNs. One of two biological replicates is shown for OLFR1507 + OSNs; the other replicate yielded similar results but with lower enrichment in peaks. ATAC-seq from the OLFR1507-expressing cells (bottom) shows that the promoter of Olfr1507 has similar accessibility to the H element. ATAC-seq plot shows pooled data from two biological replicates that yielded similar results.

Extended Data Fig. 6 | LDB1 expression and genomic distribution in mOSNs. a, Transcript level, expressed as fragments per kilobase per million mapped reads (FPKM), of the two LDB family members in mOSN RNA-seq datasets (n = 5 biological replicates). b, Sections of olfactory epithelium stained for LDB1 (green) and ADCY3 (magenta), a marker for mOSNs. Nuclei are labelled with DAPI (blue). Scale bar, 25 µm. Similar results were obtained from four independent experiments. c, LDB1 and LHX2 ChIP–seq signal in mOSNs across the OR gene cluster containing the Greek island Lipsi. OR genes are red and all other genes are blue. Plot shows pooled data from three biological replicates for LDB1 and two biological replicates for CTCF and RAD21. Values are counts per 10 million reads. Analysing each replicate separately yielded similar results. j, mOSN ChIP signal over Greek island and non-Greek island ChIP–seq peaks. For ChIP–seq peaks, the heat map shows 500 randomly selected peaks and the plot shows data from the full consensus set of peaks (n = 22,791 for LDB1, n = 24,883 for CTCF, and n = 9,882 for RAD21). Plots show pooled data; similar results were obtained with each replicate (n = 3 for LDB1 ChIP–seq and n = 2 for CTCF and RAD21 ChIP–seq). Units are counts per 10 million reads. k, As in j, but showing LDB1 ChIP signal over OR genes (n = 1,255) in mOSNs. l, LDB1 ChIP–seq from control mOSNs (top) and OLFR1507-expressing cells (middle). Strong signal is observed on the Greek island H, in both populations, but only a very weak signal on the Olfr1507 promoter when it is transcriptionally engaged. Pooled data from three biological replicates is shown for the mOSNs. One of two biological replicates is shown for OLFR1507 + OSNs; the other replicate yielded similar results but with lower enrichment in peaks. ATAC-seq from the OLFR1507-expressing cells (bottom) shows that the promoter of Olfr1507 has similar accessibility to the H element. ATAC-seq plot shows pooled data from two biological replicates that yielded similar results.
Extended Data Fig. 7 | Effects of conditional Ldb1 deletion on Greek island interactions and OR expression. a, Schematic of the genetic strategy used to generate Ldb1 KO mOSNs that are fluorescently labelled. b, In Ldb1fl/fl;OMP-cre mice, LDB1 (green) is lost from mOSNs but retained in basal immature cells. Nuclei are stained with DAPI (magenta). Scale bar, 20 μm. Similar results were obtained from three independent experiments. c, Hi-C contacts between a pair of OR clusters located on different chromosomes in control (top) and Ldb1 KO (bottom) mOSNs. A Hi-C hotspot between interacting Greek islands in control mOSNs (arrowheads) is absent in Ldb1 KO OSNs. d, Three-dimensional projection of the same OR cluster pair in control and Ldb1 KO OSNs. e, Trans interactions of each Greek island (n = 59) with the other Greek islands as fraction of the total Hi-C contacts in mOSNs versus Ldb1 KO cells. Greek islands that changed more than twofold are red. f, For each Greek island, the mean number of cis long-range (left) and trans (right) Hi-C contacts per billion made to every non-OR sequence (at 50-kb resolution), intergenic LHX2- and EBF-bound peak (outside OR clusters), or Greek island. Boxes indicate median, upper and lower quartiles; whiskers indicate 1.5 × the interquartile range. g, As in e but for trans contacts between OR gene clusters (n = 67). Clusters that changed more than 1.5-fold are red. h, Circos plots depicting the strongest 1,000 interchromosomal interactions genome-wide at 1-Mb resolution in control mOSNs (left) and Ldb1 KO mOSNs (right). Red lines represent OR–OR contacts and black lines non-OR–non-OR contacts. Line thickness increases with contact frequency. Chromosome numbers depicted at the periphery of the circle. i, Transcript levels of Greek island-binding factors in RNA-seq data from control mOSNs and Ldb1 KO mOSNs. Transcript levels of Ebf3 are reduced approximately twofold (P = 0.031 for greater than 1.5-fold change, DESeq2 normalized Wald test with n = 5 for control mOSNs and n = 4 Ldb1 KO). The expression of other factors is not significantly different between conditions.
Extended Data Fig. 8 | Long-range interactions in homogeneous OSN subpopulations. a–c, Circos plots representing the 1,000 strongest trans contacts in OLFR16+ (a), OLFR17+ (b) and OLFR1507+ (c) OSNs. d, Left, comparison of the frequency of local cis (grey), long-range cis (blue) and trans (red) Greek island interactions in mixed mOSNs and OSNs expressing specific OR genes. Right, mean values for OLFR16+, OLFR17+ and OLFR1507+ cells are not significantly different from those for mixed mOSNs (P > 0.05 for all comparisons, two-tailed paired Wilcoxon signed-rank test). e, In situ Hi-C contact matrices from OLFR16+, OLFR17+ and OLFR1507+ cells focused on the Olfr16 gene locus. Arrowhead points to specific long-range contacts between Olfr16 and the Greek island Astypalea that occur only in OLFR16+ cells. Grey arrowheads mark Greek island–Greek island contacts that also differ between cell types. f, g, Similar analysis for the Olfr16 locus in OLFR17+ and OLFR1507+ cells. h–j, As in e–g, but for the Olfr17 locus. k–m, As in e–g, but for the Olfr1507 locus.
Extended Data Fig. 9 | See next page for caption.
Extended Data Fig. 9 | Long-range cis and trans contacts between Greek islands and the active OR gene. a, Hi-C contacts that span more than 80 Mb are observed between the Olfr16 locus and Greek islands in OLFR16+ cells. b, Close examination of the contacts (dashed box from a) reveals that Greek islands contact Olfr16 only in OLFR16+ cells (top, black arrowhead). Extremely long-range contacts between Greek islands (grey arrowheads), but not involving the Olfr16 locus, are also observed in OLFR17+ and OLFR1507+ cells (middle, bottom). c, Heat maps depicting interchromosomal contacts between Olfr16 (chromosome 1) and Greek islands from different chromosomes in in situ Hi-C from OLFR16+, OLFR17+ and OLF1507+ cells. d, Three-dimensional projection of APA between the Olfr16 locus and trans Greek islands in the three specific mOSN populations. e, Heat maps for contacts between Olfr16, Olfr17, or Olfr1507 and trans Greek islands reveals an accumulation of contacts centred around the active allele. f, APA for OR versus trans Greek islands shows the accumulation of contacts on the active allele at 10-kb resolution. The poor mappability of the Olfr17 locus and the lower sequencing depth perturbs the expected focal peak. For the Olfr1507 locus, the presence of the Greek island H, 50 kb from Olfr1507, results in Hi-C contacts spanning a broad area. g, h, Short, long and trans contacts with Greek islands across the OR gene clusters containing Olfr17 (g) and Olfr1507 (h) plotted as a fraction of the total Hi-C contacts mapped to each position (5-kb resolution). Top, contacts in cells in which Olfr17 and Olfr1507 are active; bottom, contacts in OLFR16+ cells in which Olfr17 and Olfr1507 are silent.
Extended Data Fig. 10 | A model for specific OR compartmentalization and the generation of mutually exclusive phases that regulate OR gene choice. a, b, Coincidence of LHX2 and LDB1 peaks with trimethylation of histone 3 lysine 9 (H3K9me3) enrichment may generate an OR-enriched molecular barcode that promotes specific interactions between OR gene clusters. c, In INPs, where OR compartments first form, Greek islands do not make specific contacts with each other. d, However, in mOSNs Greek islands specifically interact with each other through homotypic LDB1 interactions, forming a multi-enhancer hub that is segregated from the OR compartment. We hypothesize that OR compartments and Greek island hubs form incompatible liquid phases driven by HP1 proteins and the unstructured domains of LHX2 and LDB1, respectively. e, Upon deletion of Ldb1 (or Lhx2), the Greek island phase falls apart and the Greek islands become incorporated to the OR compartments, as in the INPs.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

| n/a | Confirmed |
|-----|-----------|
| ☑   | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☑   | An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☑   | The statistical test(s) used AND whether they are one- or two-sided |
| ☑   | Only common tests should be described solely by name; describe more complex techniques in the Methods section. |
| ☑   | A description of all covariates tested |
| ☑   | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☑   | A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☑   | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted |
| ☑   | Give P values as exact values whenever suitable. |
| ☑   | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☑   | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☑   | Estimates of effect sizes (e.g. Cohen's d, Pearson’s r), indicating how they were calculated |
| ☑   | Clearly defined error bars |
| ☑   | State explicitly what error bars represent (e.g. SD, SE, CI) |

Software and code

Policy information about availability of computer code

Data collection

Zeiss Zen2012 SP1 (v8.1.9.484) was used for capturing confocal images

Data analysis

Data was analyzed on linux workstations running Ubuntu 18.04.1 LTS or Red Hat Enterprise Linux 7.3. The following software was used:
Juicer Tools Version 1.76
BWA 0.7.17
Integrated Genome Browser 9.0.0
CutAdapt v1.17
Bowtie2 v2.3.2
Picard
Samtools v1.4.1
HOMER v4.10.3
Bedtools2 v2.26.0
Deeptools2 v2.1.1
STAR v2.5.3a
RSeQC v2.6.4
ImageJ 2.0.0
circos 0.69.5,
slurm 17.11.2

Our web collection on statistics for biologists may be useful.
python 2.7.3 with the following packages: seaborn 0.9.0, matplotlib 1.5.3, scikit-learn 0.20.0, hmmlearn 0.2.1, numPy 1.15.4, and pandas v0.23.4.

R-Studio Server (R version 3.5.1) with the following packages: tidyverse1.2.1,DiffBind v2.8.0, DESeq2 v1.20.0, VennDiagram v1.6.20, cowplot v0.9.9, reshape2 v1.4.3, RColorBrewer v1.1.2, BiocParallel v1.14.2, GenomicRanges v1.32.7, Rsamtools v1.32.3, pheatmap v1.0.10, ggplot v3.0.1.1, and SummarizedExperiment v1.10.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data
Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All figures include publicly available data. All ChIP-seq and RNA-seq data reported in this paper are available from GEO (GSE112153). Additional data (mOSN RNA-seq, mOSN Lhx2 ChIP-seq, mOSN Ebf ChIP-seq, and Olfr1507+ ATAC-seq) were previously described and are available from GEO (GSE93570). All HiC data generated in this study are publicly available at www.data.4Dnucleome.org under the following accession numbers: 4DNESH4UTRN, 4DNESNYBOSLY, 4DNES54YB6TQ, 4DNESRE7AK5U, 4DNES425UDGS, 4DNESEPDL6KY.

Field-specific reporting
Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design
All studies must disclose on these points even when the disclosure is negative.

Sample size
A sample size of two independent biological replicates was selected for high throughput sequencing experiments. This size was selected because multiple statistical approaches have been developed to allow identification of significantly changed loci or genes from two biological replicates of high throughput sequencing data (e.g. edgeR, DEseq2). These approaches work by using the large number of genes/loci measured to analyze and model the dispersion and variance within and between replicates, thereby allowing the identification of genes/loci with significant differences between conditions. Wherever possible, additional biological replicates were included.

Data exclusions
For ChIP-seq data, peaks within ENCODE black-listed regions (https://sites.google.com/site/anshulkundaje/projects/blacklists) were excluded from analysis. This exclusion was pre-established. No other data were excluded.

Replication
At least two independent biological replicates were performed for each experiment. Replicate experiments yielded the same results.

Randomization
No experiments were performed with live animals. For the purpose of purifying primary cells, animals of similar age were grouped by genotype and sorted together.

Blinding
Animals were used as a source of tissue and primary cells, so knowledge of genotype was required for proper handling and cell sorting. Blinding was not attempted for in vitro experiments performed on sorted cells.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Unique biological materials |
| ☒   | Antibodies |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☐   | ChIP-seq |
| ☐   | Flow cytometry |
| ☒   | MRI-based neuroimaging |
Antibodies

| Antibodies used | CTCF (Millipore Cat# 07-729, RRID:AB_441965, serum, diluted 1:500 for ChIP), Rad21 (Abcam Cat# ab992, RRID:AB_2176601, polyclonal, used at 0.2ug/mL for ChIP), or Ldb1 (Santa Cruz Biotechnology Cat# sc-11198, RRID:AB_2137017, polyclonal, used at 0.1ug/mL for immunofluorescence and 0.2ug/mL for ChIP) Adcy3 (Santa Cruz Biotechnology Cat# sc-588, RRID:AB_630839, polyclonal), anti-goat IgG conjugated to Alexa-488 (Thermo Fisher Scientific Cat# A-11055, RRID:AB_2534102, polyclonal, used at 2ug/mL), anti-rabbit IgG conjugated to Alexa-555 (Thermo Fisher Scientific Cat# A-31572, RRID:AB_162543, polyclonal, used at 2ug/mL). |

Validation

| Validation | CTCF - manufacturer states that antibody is reactive for mouse and validated for ChIP-seq. Antibody has been extensively used and CTCF depletion eliminates immunoreactivity (e.g. Nora et al, Cell 2017).
Rad21 - Manufacturer states that antibody reacts with mouse and is validated for ChIP-seq. Antibody has been used extensively and Rad21 depletion eliminates Rad21 ChIP-seq signal (e.g. Rao et al, Cell 2017).
Ldb1 - Manufacturer states that antibody reacts with mouse. Antibody has been extensively used and Ldb1 deletion eliminates reactivity by immunofluorescence (this study) and western blot (e.g. Song et al, Blood, 2010).
Adcy3 - Manufacturer states that antibody reacts with mouse and is validated for immunofluorescence.
anti-rabbit IgG conjugated to Alexa-555 - Manufacturer states that antibody is validated for immunofluorescence.
anti-goat IgG conjugated to Alexa-488 - Manufacturer states that antibody is validated for immunofluorescence. |

Animals and other organisms

| Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research |

Laboratory animals

This study used several mouse lines (mus musculus) on mixed C57BL/6J and 129 backgrounds. Experimental genotypes were:
Krt5Cre;R26R-tdtomato-GFP - Male and female adult mice were used
Ngn1-GFP - Male and female mice were used a 1-2 weeks of age.
OMP-ires-GFP - Male and female adult mice were used
H-/-;Lipari/-;Sfatkiri/-;OMP-ires-GFP - Male and female adult mice were used
Krt5Cre;Lhx2fl/fl;R26R-tdtomato - Male and female adult mice were used
OMP-ires-Cre;Lhx2fl/fl;R26R-tdtomato - Male and female adult mice were used
OMP-ires-Cre;Ldb1fl/fl;R26R-tdtomato - Male and female adult mice were used
Ofr16-ires-GFP - Male and female adult mice were used
Ofr17-ires-GFP - Male and female adult mice were used
Ofr1507-ires-GFP - Male and female adult mice were used

Wild animals

Study did not involve wild animals.

Field-collected samples

Study did not involve specimens collected from the field.

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as GEO.

Data access links

- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Files in database submission

- mOSN_RNAseq_rep4.fastq.gz
- mOSN_RNAseq_rep5.fastq.gz
- mOSN_Ldb1KO_RNAseq_rep1.fastq.gz
- mOSN_Ldb1KO_RNAseq_rep2.fastq.gz
- mOSN_Ldb1KO_RNAseq_rep3.fastq.gz
- mOSN_Ldb1KO_RNAseq_rep4.fastq.gz
- mOSN_3enhKO_RNAseq_rep1.fastq.gz
- mOSN_3enhKO_RNAseq_rep2.fastq.gz
- mOSN_3enhKO_RNAseq_rep3.fastq.gz
- mOSN_3enhKO_RNAseq_rep4.fastq.gz
- Krt5Cre_Lhx2Het_Dim_RNAseq_rep1.fastq.gz
- Krt5Cre_Lhx2Het_Bright_RNAseq_rep1.fastq.gz
- Krt5Cre_Lhx2KO_Dim_RNAseq_rep1.fastq.gz
- Krt5Cre_Lhx2KO_Bright_RNAseq_rep1.fastq.gz
- mOSN_CTCF_Chipseq_rep1.R1.fastq.gz
- mOSN_CTCF_Chipseq_rep1.R2.fastq.gz
- mOSN_Rad21_Chipseq_rep1.R1.fastq.gz
- mOSN_Rad21_Chipseq_rep1.R2.fastq.gz
- mOSN_Ldb1_Chipseq_rep1.R1.fastq.gz
- mOSN_Ldb1_Chipseq_rep1.R2.fastq.gz
The methodology section of the document includes the following details:

**Replicates**
Replicates closely agree when assessed by peak overlap or by clustering when analyzed together with input control or ChIP-seq data for other factors. Replicates are also enriched for similar motifs.

**Sequencing depth**
See Supplementary Information 1.

**Antibodies**
- CTCF (Millipore Cat# 07-729, RRID:AB_441965)
- Rad21 (Abcam Cat# ab992, RRID:AB_2176601)
- Ldb1 (Santa Cruz Biotechnology Cat# sc-11198, RRID:AB_2137017)

**Peak calling parameters**
All data sets were processed using 50bp of single end data; 75bp reads were trimmed to 50bp and only read 1 was used from paired end data. Adapter sequences were removed from raw ChIP-seq data using CutAdapt (RRID:SCR_011841) and filtered reads were aligned to the mouse genome (mm10) using Bowtie2 v2.3.2 (RRID:SCR_006646) with default settings. Picard (RRID:SCR_006525) was used to identify duplicate reads, which were then removed with Samtools (RRID:SCR_002105). Samtools was used to select uniquely aligning reads by removing reads with alignment quality alignments below 30 (~30). Peaks of ChIP-seq signal were identified using HOMER (RRID:SCR_010881) in “factor” mode with an input control.

**Data quality**
Peaks were called using an input control. Peaks were called at an FDR of 0.1% and required 4-fold enrichment over the input control and a 4-fold enrichment over the local (10 Kb) background. Number of peaks called with these settings is listed in Supplementary Information 1. Motif analysis was run on called peaks for each data set and returned the expected motif as

---

Genome browser session
(e.g. UCSC)
https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=kevingmonahan&hgS_otherUserSessionName=monahan_et_al
the top hit in each case.

Software

ChiP-seq data was analyzed using Bedtools2, Deeptools2, and DiffBind. Diffbind was used to normalize ChIP-seq scores across biological replicate experiments using the “DBA_SCORE_TMM_READS_EFFECTIVE” scoring system, which normalizes using edgeR and the effective library size. The ChIP-seq signal for each peak was then calculated by averaging the normalized score across biological replicates.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were dissociated into a single-cell suspension by incubating freshly dissected main olfactory epithelium with papain for 40 min at 37°C according to the Worthington Papain Dissociation System. Following dissociation and filtering three times through a 35μm cell strainer, live cells were sorted by collecting fluorescent, DAPI-negative cells for RNA-seq and ATAC-seq. Alternatively, cells were fixed with 1% PFA in PBS for 5 minutes (ChIP) or 10 minutes (HiC) at room temperature. Fixed fluorescent cells were then sorted on a BD Aria II, Influx, or MoFlo cell sorter.

Instrument

BD Aria II, BD Influx, or Beckman Coulter MoFlo Astrios EQ

Software

No post-sort analysis of flow data was performed

Cell population abundance

Post-sorted cells were periodically checked for presence of the selected fluorescent label. Purities of greater than 95% were routinely observed.

Gating strategy

Unlabeled negative control population were analyzed to establish gate positions.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.