Dietary palmitic acid promotes a prometastatic memory based on Schwann cell activation

Gloria Pascual[1*#], Diana Domínguez[1*], Marc Elosúa-Bayes[3], Felipe Beckedorff [5], Carmelo Laudanna[1], Claudia Bigas[1], Delphine Douillet[4], Carolina Greco[6], Aikaterini Symeonidi[1], Inmaculada Hernández[1,3], Sara Ruiz Gil[3], Neus Prats[1], Coro Bescós[6], Ramin Shiekhattar[5], Moran Amit[7], Holger Heyn[3], Ali Shilatifard[4#] & Salvador Aznar Benitah[1,2#].

[1]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [2]ICREA, Catalan Institution for Research and Advanced Studies, 08010 Barcelona, Spain. [3]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [4]Department of Biochemistry and Molecular Genetics and Simpson Querrey Center for Epigenetics, Northwestern University Feinberg School of Medicine. [5]Silvester Cancer Center, University of Miami, USA. [6]Vall D´Hebron Hospital, Barcelona, Department of Oral and Maxillofacial Surgery. Universitat Autònoma de Barcelona, Spain. [7]Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

* Equal contribution

Correspondence to: Gloria.pascual@irbbarcelona.org; ASH@northwestern.edu and Salvador.aznar-benitah@irbbarcelona.org

Fatty acid (FA) uptake and altered metabolism constitute hallmarks of metastasis [1-2] yet it is unclear the biology behind it, or whether all dietary FAs are prometastatic. Here we show that dietary palmitic acid (PA), but not oleic acid or linoleic acid, promotes metastasis in oral carcinomas and melanoma. Unexpectedly, tumours from mice fed a short-term palm oil (PA)–rich diet, or tumour cells briefly exposed to PA *in vitro*, remain highly metastatic even when serially transplanted (without further exposure to high levels of PA). This PA–induced prometastatic memory requires the fatty acid transporter CD36 and is associated with the stable deposition of histone H3 lysine 4 trimethylation by the methyltransferase Set1A/COMPASS. Bulk, single-cell and positional RNA sequencing indicate that genes with this prometastatic memory predominantly relate to a neural signature that stimulates intratumour Schwann cells and innervation, two parameters strongly correlated with metastasis but etiologically poorly understood[3-4]. Mechanistically, tumour-associated Schwann cells secrete a specialized pro-regenerative extracellular matrix, whose ablation inhibits metastasis initiation. Both the PA-induced memory of this proneural signature and its long-term boost in metastasis require the transcription factor EGR2 and the glial cell–stimulating peptide galanin. In sum, we provide evidence that a dietary metabolite provokes stable transcriptional and chromatin changes that lead to a long-term stimulation of metastasis, and that this is related to a pro-regenerative state of tumour-activated Schwann cells.

Metastatic growth from primary tumours is multifactorial and may require non-genetic factors, including lifestyle. For instance, high-fat diets promote tumourigenesis in pre-clinical models of cancer[1-2]; obesity is associated with a high aggressiveness of certain types of cancer[5]; and altered uptake and metabolism of fatty acids (FAs) drive cancer progression[6-9]. Mechanistically, FA–interacting proteins and transporters (including CD36) are relevant for tumourigenesis, chemotherapy resistance and metastasis[6-9]. However, we still do not know which dietary FAs can trigger metastasis, or the mechanisms behind this.

To address these questions, we stimulated human oral squamous cell carcinoma (OSCC) cells with palmitic acid (PA), oleic acid (OA) or linoleic acid (LA) for 4 days in culture prior to orthotopically inoculating cells into mice fed a standard diet (Extended Data Fig. 1a). No FA treatment affected primary tumour initiation (note that PA was used at 300 μM, within its physiological range in human serum[10-11], and OA and LA, at 50 μM to avoid lipotoxicity) (Extended Data Table 1a). Notably, however, PA strongly increased the penetrance and size of metastatic lesions and induced CD36 cell surface expression, while LA and OA had no significant effects (Extended Data Fig. 1a–c).

To test if the prometastatic effect of PA required constant high levels of the fatty acid, we stimulated OSCC cells with PA for 4 days, removed it from the cell medium and grew cells for a further 14 days in standard medium before orthotopically inoculating them into mice (referred to herein as 14D post-PA) (Fig. 1a). CD36 membrane expression returned to background levels at 14D post-PA (Extended Data Fig. 1d). However, strikingly these cells were much more competent in generating metastases than control cells (Fig. 1b, Extended Data Fig. 1e-h and Extended Data Table 1b, c). The prometastatic effect of a pre-treatment with PA was so strong that $CD36^{bright}$ cells, and even $CD36^{dim}$ cells (i.e., with low levels of CD36, previously described as non-metastatic[6]), retained it when serially transplanted (with no changes to primary tumour sizes) (Extended Data Fig. 1i,j and Extended Data Table 1d,e). The beneficial effects of OA did not counteract the prometastatic effect of PA (Fig. 1b, Extended Data Fig. 1g,h and Extended Data Table 1f-i). PA stimulated metastasis even at a lower concentration (of 50 μM, as used for the

other FAs), while another saturated FA such as stearic acid did not (Fig. 1b, Extended Data Fig. 1g, h and Extended Data Table 1f-i). Thus, short-term exposure to high levels of PA, rather than to dietary FAs in general, induces a prometastatic memory in OSCC.

To test whether a PA-induced prometastatic memory could also come from a palm oil–rich diet, we orthotopically inoculated OSCC cells into mice; once primary tumours appeared, mice were fed short-term (ten days) with a diet rich in palm oil, or olive oil, or a standard one; mice were then fed a standard diet until being sacrificed at the endpoint. We then collected primary tumours (all of which contained a similar number of $CD36^{bright}$ cells), purified OSCC cells and serially transplanted tumour cells into secondary mice fed a normal diet (Fig. 1c and Extended Data Fig. 2a, b). Corroborating our *in vitro* results for PA and OA, tumour cells from palm oil–fed primary recipient mice were much more metastatic, and those from olive oil–fed mice, less metastatic, in secondary recipient mice (Fig. 1d, Extended Data Fig. 2c, d and (Extended Data Table 1j). The PA–induced stable boost in metastasis also occurred after serial orthotopic inoculation of melanoma cells under the same dietary conditions (Extended Data Fig. 2h, i). Removing CD36 from the primary tumour (by constitutive CD36 knockdown or treatment with a CD36-neutralizing antibody) led to a loss of the prometastatic phenotype but did not affect primary tumour competency (Extended Data Figs. 2e-i, 3a and Extended Data Table 1j-n). CD36 depletion using an inducible shRNA in the secondary recipient mice also inhibited the prometastatic memory of PA in the tumour cells (Extended Data Fig. 3b, c and Extended Data Table 1o, p).

To investigate whether the epigenomes of tumour cells are affected by PA regarding this metastatic memory, we treated OSCC cells *in vitro* for 4 days with PA or OA, followed by 14D post-PA. We then studied by chromatin immunoprecipitation followed by sequencing (ChIP-seq) the genome-wide changes in histone marks indicative of promoters (histone H3K4me3), enhancers (histone H3K4me1), active transcription (H3K27ac), Polycomb–mediated repression (histone H3K27me3) and non-Polycomb repression (H3K9me3) (Fig. 2a-c, Extended Data Figs. 3d-h, 4a, 5a and Extended Data Tables 2a, b, 3a-d, 4a, k-q). Intriguingly, we observed stable changes in H3K4me3 in the promoters of specific genes after either PA or OA, but with a

significantly higher number of genes affected by PA than by OA exposure (Extended Data Fig. 4b, e and Extended Data Table 5c, f, g, o-q). A stably altered H3K4me3 profile was also evident *in vivo* by ChIP-seq of secondary tumour transplants that had retained the prometastatic memory of a palm oil–rich diet described in Fig. 1 (Extended Data Fig. 4c-d and Extended Data Table 2e). Importantly, the overlap of H3K4me3-marked genes after PA *in vitro* or palm oil *in vivo* was statistically significant, underscoring the physiological relevance of our *in vitro* studies (Extended Data Fig. 4e, f).

The distribution of H3K4me1, H3K27me3 and H3K27ac was also altered after a 4-day PA treatment (Extended Data Figs. 4a, 5a and Extended Data Tables 3a, 4d, o, q), but the majority of these changes had faded at the 14-days post-PA timepoint (Extended Data Figs. 3d, 5a and Extended Data Tables 3b-e, 4d, p, 5h). Indeed, no stable changes in H3K4me1 and H3K27ac were observed in the enhancers of the genes with long-term differences in H3K4me3 induced by PA, as determined by ChIP-seq and nascent transcription at enhancers through PRO-seq analysis (Extended Data Fig.3 g, h and Extended Data Table 3f-h). Furthermore, promoters with the PA-induced stable changes in H3K4me3 were not bivalent, since they did not show stable changes in H3K27me3 (Extended Data Fig. 3e and Extended Data Table 5d, e). Of note, the H3K9me3 levels were significantly reduced in several genes in a stable manner after PA treatment (e.g., even after PA withdrawal); however, there was less than a 2% overlap between these genes with reduced H3K9me3 and those that gained H3K4me3 in their promoters (Extended Data Fig. 3i, j and Extended DataTable 4a-c). Altogether, these results indicate that while different chromatin marks respond to PA stimulation, these changes are stably maintained mainly by i) genes with newly-marked H3K4me3 promoters, and ii) genes with specific regions that lose the repressive H3K9me3 mark.

To distinguish whether PA treatment can induce the CD36+ prometastatic state *de novo* or only enhances the potential of CD36+ cells already present within a tumour, we analyzed OSCC cells stimulated with PA for 4 days in culture using t-SNE–based cluster and pseudo-time analyses of single-cell RNA-sequencing (scRNA-seq). Strikingly, almost all cells responded to PA and

acquired the CD36+ signature of metastasis-initiating cells, including genes such as *CD36* and *IFIT3* (Extended Data Fig. 5b, c and Extended Data Table 4s, t). Furthermore, cells slowed down their cell cycle during PA treatment but were proliferating normally at the 14-days post-PA timepoint, with no associated apoptotic death at any point, ruling out the selection or out-competition of CD36+ cells over any other cell state (Extended Data Fig. 5d and Extended Data Table 4e). These results indicate that PA induces, rather than selects, a stable prometastatic memory in OSCC cells.

Interestingly, gene ontology (GO) analyses of the genes with the PA-induced stable changes in H3K4me3 revealed many functions related to neurogenesis and neural remodeling (Extended Data Tables 2c, d, f and 5a, b, i, j, l). Conversely, the levels of H3K4me3 were downregulated in genes pertaining to the same neural categories after oleic acid stimulation (Extended Data Table 5m, n). We were particularly intrigued by these categories, as perineural invasion (PNI) and tumour innervation are among the strongest predictors of poor prognosis in oral, prostate, breast and pancreatic carcinomas, among other cancer types[3-4, 12-16.] However, the mechanisms underlying their contribution to tumour aggressiveness, remain mostly unknown.

There was a significant correlation between the stable changes in H3K4me3 and differentially-expressed genes (DEGs) in OSCC cells stimulated with PA *in vitro* (Extended Data Fig. 5e and Extended Data Table 4f, g and 5r, s). Many of these DEGs are related to neural categories, among other biological functions (Extended Data Tables 4h-j, and 5t). Importantly, the same neural categories appeared (in a CD36-dependent manner) in the transcriptome of $CD36^{bright}$ cells purified from oral tumours and melanomas harbouring the *in vivo* prometastatic memory of the palm oil diet but not in olive oil–diet-derived secondary tumours (Fig. 3a, Extended Data Fig. 6a–h and Extended Data Tables 6, 7). These proneural transcriptional changes were functionally relevant, since chemical sympathectomy through intraperitoneal administration of the noradrenaline-selective neurotoxic drug 6-hydroxydopamine (6-OHDA) strongly reduced prometastatic memory of PA without affecting primary tumour competency (Extended Data Fig. 7a-c and Extended Data Table 8a-d).

Transcription factor–binding analysis in the promoters of the in vivo palm diet- neural-related DEGs of OSCC cells showed a strong enrichment for EGR2 binding sites (Extended Data Fig. 7d and Extended Data Table 9). In addition, a search for the *in vivo* neural signature induced by PA that was common to all the tumours we had tested resulted in a list of genes that included the glial-inducing peptide galanin (GAL) as one of the most upregulated ones[17-18] (Extended Data Fig. 7e and Extended Data Table 9). The genomic localization of EGR2 responded to PA stimulation of OSCC cells, and a significant proportion of the genes that gained EGR2 binding, including galanin, were neural-related in a CD36-dependent manner (Fig. 3b, c, Extended Data Fig. 7f-h and Extended Data Table 10 a, i, j, m, o). Furthermore, knockdown of EGR2 altered the distribution of H3K4me3 (Extended Data Fig. 8a-b and Extended Data Table 10 b-h, k,l,n), which were enriched for neural-related genes (Extended Data Fig. 8b and Extended Data Table 5a). Importantly, depletion of either EGR2 or galanin prevented $CD36^{bright}$ tumour cells from acquiring a stable neural signature (Extended Data Fig. 8c, d and Extended Data Table 9) and strongly reduced the *in vivo* prometastatic memory of OSCC established by the palm oil–rich diet (Fig. 3d, Extended Data Fig. 8e and Extended Data Table 9). This antimetastatic effect was also seen upon systemic inhibition of galanin signaling through the intraperitoneal administration of the pan-galanin receptor inhibitor galantide (M15) (Extended Data Fig. 9a-c and Extended Data Table 11a-c). Interestingly, expression of *EGR2* and *galanin* significantly correlated with poor prognosis in patients with head and neck squamous cell carcinoma as well as other types of solid tumours (Extended Data Table 12).

H3K4 trimethylation is deposited by methyltransferases of the MLL/COMPASS family, including MLL1, MLL2, Set1A and Set1B[19-20]. In OSCC cells at the 14-days post-PA timepoint, no changes were observed in either the protein expression of these four COMPASS family members (Extended Data Fig. 10a, b) or in the localization of MLL1 and MLL2 on chromatin (as determined by ChIP-seq) (Extended Data Fig. 10c-e, and Extended Data Table 13a, d ,e). Further, inhibition of MLL1 did not affect the long-term PA-induced metastatic potential of OSCC cells after PA stimulation (Extended Data Table 13b, c). Conversely, OSCC cells depleted of Set1A

did not acquire the stable PA-induced H3K4me3 neural signature (Extended Data Fig. 10f, g and Extended Data Table 14a, b, f, g), and significantly reduced their ability to retain a PA-induced prometastatic memory *in vitro* and *in vivo* (Extended Data Fig. 10h). Although the size of the primary tumours was diminished in Set1A-depleted cells, their penetrance was unchanged (Extended Data Table 14c-e).

We next studied whether the predominant neural signature of tumours retaining the prometastatic memory of PA influenced the composition and function of the tumour stroma, specifically relating to its neural component. Strikingly, the bulk transcriptome of the stroma from these OSCC tumours was distinct from that of tumours derived from all other dietary conditions in a CD36-dependent manner (Extended Data Fig. 11a-c and Extended Data Table 15). The DEGs of the palm oil–derived tumour stroma related predominantly to extracellular matrix (ECM) organization, neurogenesis, innervation and gliogenesis, which were not observed in OSCC cells depleted of CD36, EGR2 or galanin (Extended Data Fig. 11d-g and Extended Data Table 15). Node pathway interactome predictions further highlighted neuron projection and Schwann cell development as functions significantly enhanced in the stroma of palm oil–derived tumours (Extended Data Fig. 12a and Extended Data Table 15).

We next performed 10X single cell (sc)-RNA-seq to specifically characterize the cellular composition of the bulk tumour stroma that we had purified. UMAP analysis identified five clusters annotated to erythrocytes, vascular and lymphatic endothelium, smooth muscle cells and tumour-associated Schwann cells (Fig. 4a, Extended Data Fig. 12b, c and Extended Data Table 16). Within these five cell types, tumour-associated Schwann cells and lymphatic endothelial cells displayed the highest proportional changes in cell content between both conditions PA-induced memory as compared to control tumours, in a CD36-dependent manner (Fig. 4a). However, the DEGs relating to ECM organization, neurogenesis and gliogenesis/Schwann cells, which we identified as the most predominant GO analysis of the bulk transcriptome data, corresponded primarily to the tumour-associated Schwann cell population (Fig. 4b, Extended Data Fig, 12d and Extended Data Table 16).

We were particularly interested in the expression changes in ECM components specifically associated with the tumour-associated Schwann cell cluster. Several of these ECM proteins are components of perineuronal nets, a specialized matrix secreted by glial cells during nerve injury and essential for nerve regeneration[21-23] (Fig. 4b, Extended Data Figs. 13a-c and Extended Data Table 16). Using the TCGA database, we observed a significantly positive correlation between their expression and the presence of the PA-induced CD36+ metastatic signature in human head and neck SCCs and melanomas (Extended Data Fig. 13d). Immunofluorescence analyses further showed their increased expression in the stroma of the tumours with PA-induced prometastatic memory in very close proximity to tumour-associated Schwann cells, as compared to their control diet counterparts (Extended Data Fig. 13e, f). Positional RNA-seq allowed us to clearly define the bulk and invasive front of the tumours surrounded by the healthy mouse tissue of the tongue (Extended Data Fig. 14a, b). Lesions that retained the PA prometastatic memory showed a strong increase in the number of Schwann cells infiltrating the tumour compared to control tumours in a CD36-dependent manner (Extended Data Figure 14b, c). This was confirmed by immunofluorescence the Schwann cell-specific marker S100 (Extended Data Fig. 15a-f). The positional RNA-seq also confirmed high levels of ECM components related to tumour-associated Schwann cells (Extended Data Fig.14b).

We next expressed a bacterial Chondroitinase ABC (modified to be expressed in eukaryotic cells) in OSCC cells (Extended Data Fig. 16a), which specifically digests the Schwann cell–derived ECM components, allowing us to directly test their potential role in metastasis[24]. Strikingly, chondroitinase ABC expression prevented the stable increase in metastatic competency induced by PA in OSCC cells, without affecting the penetrance of primary tumours (Fig. 4c and Extended Data Table 17a-d). Digestion of Schwann-cell specific ECM components, but not of collagen 5A1 (not produced by Schwann cells) was confirmed by IF staining (Extended Data Fig. 16b, c).

Metastatic progression of tumours requires driver mutations, but is also associated with stable (*i.e.* epigenetic) alterations in chromatin. For instance, pancreatic adenocarcinoma cells modify large euchromatic regions decorated by H3K9me3 downstream of the pentose phosphate pathway

to acquire metastatic competency[25]. In addition, metastasizing melanomas show global increases in histone H3K27me3 and H3K9me3, and depend on the histone methyltransferases SETDB1 and EZH2 to efficiently metastasize[26]. Our results indicate that dietary palmitic acid not only stimulates metastasis, but does so in a long-term stable manner through a transcriptional state that stimulates intratumour Schwann cells. Interestingly, this prometastatic memory is only retained by tumour cells but not the tumour stroma (Extended Data Table 18).

Schwann cells are essential for the protection of peripheral nerves as well as for their recovery after damage[27-28]. The presence of intratumour Schwann cells is associated with metastasis of numerous cancers, although the factors responsible for their stimulation and the mechanism behind how they promote metastasis are ill-defined[3-4, 12-16]. Our data indicate that intratumour Schwann cells activated by metastatic cells downstream of dietary PA, adopt a pro-regenerative state related to the secretion of a specialized ECM, which occurs at least in part via communication with the neuropeptide galanin secreted by CD36+ metastatic cells. Work from us and others has shown that the presence of the CD36+ signature correlates with poor survival in several tumour types[29-30]. We now show that there is a significant positive correlation between the PA-induced CD36+ metastatic signature and the expression of perineuronal net components in many different tumour types (Extended Data Table 19).

Our data suggest that this proneural/prometastatic state is established in part via the transcription factor EGR2 and stabilized by the methyltransferase Set1A/COMPASS downstream of CD36. Notably, CD36 is required for fat tasting and sensing[31], and Set1A deposits H3K4me3 at neural genes to epigenetically establish the neural lineage during development[32-33]. Therefore, one hypothesis is that metastatic cells aberrantly hijack the natural fat-sensing function of CD36 that is normally mediated by the interactions between specialized cells, such as oral keratinocytes and intestinal epithelial cells, and their associated nerves and glial cells, thereby establishing specific epigenetic states involved in fat metabolism and signalling[31,34]. Extensive future work will be necessary to test this exciting possibility.

Different tumour types seem to display specific fatty acid preferences related to metastasis. For example, OA inhibited the metastatic spread of oral carcinoma and melanoma in our orthotopic models, yet it stimulates metastasis of cervical and gastric carcinomas[35]. In addition, metastasizing melanoma cells that invade through the lymphatic system acquire a high metastatic competency by being exposed to high levels of oleic acid present in the lymph nodes[36]. Different tumours also fuel their growth and progression by converting palmitate into sapienate[37]. However, irrespective of the fatty acid, inhibition fatty acid metabolism has potential for treating numerous tumours with an unmet clinical need[30]. In sum, our findings not only underscore the long-term health risks associated with a diet rich in PA regarding metastatic progression, but also provide mechanistic insights to identify novel epigenetic- and neural/glial-related therapeutic strategies to attenuate and prevent distant dissemination.

References

1. Peck, B. & Schulze, A. Lipid Metabolism at the Nexus of Diet and Tumor Microenvironment. *Trends in Cancer* **5,** 693–703 (2019).

2. Pascual, G., Domínguez, D. & Benitah, S. A. The contributions of cancer cell metabolism to metastasis. *Dis. Model. Mech.* **11,** dmm032920 (2018).

3. Boilly, B., Faulkner, S., Jobling, P. & Hondermarck, H. Nerve Dependence: From Regeneration to Cancer. *Cancer Cell* **31,** 342–354 (2017).

4. Zahalka, A. H. & Frenette, P. S. Nerves in cancer. *Nat. Rev. Cancer* **20,** 143–157 (2020).

5. Solans, M., Chan, D. S. M., Mitrou, P., Norat, T. & Romaguera, D. A systematic review and meta-analysis of the 2007 WCRF/AICR score in relation to cancer-related health outcomes. *Ann. Oncol.* **31,** 352–368 (2020).

6. Pascual, G. *et al.* Targeting metastasis-initiating cells through the fatty acid receptor CD36. *Nature* **541,** 41–45 (2017).

7. Lee, C. *et al.* Tumor metastasis to lymph nodes requires YAP-dependent metabolic adaptation. *Science* **363,** 644 LP – 649 (2019).

8. Rambow, F. *et al.* Toward Minimal Residual Disease-Directed Therapy in Melanoma. *Cell* **174,** 843–855.e19 (2018).

9. Haobin, Ye et al. Leukemic stem cells evade chemotherapy by metabolic adaptation to an adipose tissue niche. Cell Stem Cell **19,** 23-37.

10. Abdelmagid, S. A. *et al.* Comprehensive Profiling of Plasma Fatty Acid Concentrations in Young Healthy Canadian Adults. *PLoS ONE* **10,** e0116195 (2015).

11. Feng, R. *et al.* Free fatty acids profile among lean, overweight and obese non-alcoholic fatty liver disease patients: a case – control study. *Lipids Health Dis* **16,** 1–9 (2017).

12. Nyanbol Kuol *et al.* Role of the nervous system in cancer metastasis. *Journal of Experimental & Clinical Cancer Research* **37,** 1-12 (2018).

13. Wang, W. *et al*. Nerves in the tumour microenvironment: origin and effects. *Front. Cell Dev. Biol.* **8**, 1630 (2020).

14. Venkatesh, H. S. *et al.* Electrical and synaptic integration of glioma into neural circuits. *Nature* **573,** 539–545 (2019).

15. Amit, M. *et al.* Loss of p53 drives neuron reprogramming in head and neck cancer. *Nature* **578,** 449–454 (2020).

16. Venkataramani, V. *et al.* Glutamatergic synaptic input to glioma cells drives brain tumour progression. *Nature* **573,** 532–538 (2019).

17. Ubink, R., Calza, L. & Hökfelt, T. 'Neuro'-peptides in glia: Focus on NPY and galanin. *Trends in Neurosciences* **26,** 604–609 (2003).

18. Gresle, M. M. *et al.* Galanin is an autocrine myelin and oligodendrocyte trophic signal induced by leukemia inhibitory factor. *Glia* **63,** 1005–1020 (2015).

19. Avgustinova, A. & Benitah, S. A. The epigenetics of tumour initiation: Cancer stem cells and their chromatin. *Current Opinion in Genetics and Development* **36,** 8–15 (2016).

20. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171,** 34–57 (2017).

21. Fawcett, J. W., Oohashi, T. & Pizzorusso, T. The roles of perineuronal nets and the perinodal extracellular matrix in neuronal function. *Nat Rev Neurosci* **20,** 451–465 (2019)

22. Quraishe, S., Forbes, L. H. & Andrews, M. R. The Extracellular Environment of the CNS: Influence on Plasticity, Sprouting, and Axonal Regeneration after Spinal Cord Injury. *Neural Plasticity* **2018,** (2018).

23. Dzyubenko, E., Gottschling, C. & Faissner, A. Neuron-Glia Interactions in Neural Plasticity: Contributions of Neural Extracellular Matrix and Perineuronal Nets. *Neural Plasticity* **2016,** (2016).

24. Bartus, K. *et al.* Large-scale chondroitin sulfate proteoglycan digestion with chondroitinase gene therapy leads to reduced pathology and modulates macrophage phenotype following spinal cord contusion injury. *Journal of Neuroscience* **34,** (2014).

25. McDonald, O. G. *et al.* Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49,** 367–376 (2017).

26. Shi, X. *et al.* The abundance of metabolites related to protein methylation correlates with the metastatic capacity of human melanoma xenografts. *Sci. Adv.* **3,** (2017).

27. Clements. M. P. *et al.* The Wound Microenvironment Reprograms Schwann Cells to Invasive Mesenchymal-like Cells to Drive Peripheral Nerve Regeneration. *Neuron* **96,** 98–114.e7 (2017).

28. Johnston, A. *et al.* Dedifferentiated Schwann Cell Precursors Secreting Paracrine Factors Are Required for Regeneration of the Mammalian Digit Tip. *Stem Cell* **19,** 433–448 (2016).

29. Ana-Maria Enciu *et al.* Targeting CD36 as biomarker for metastasis prognostic: how far from translation into clinical practice? *BioMed Research International* **2018**, 7801202 (2018).

30. Gabriele Bergers & Sarah-Maria Fendt. The metabolism of cancer cells during metastasis. *Nature Reviews Cancer* **21,** 162–180 (2021).

31. Sundaresan, S. & Abumrad, N. A. Dietary Lipids Inform the Gut and Brain about Meal Arrival via CD36-Mediated Signal Transduction. *J. Nutr.* **145,** 2195–2200 (2015).

32. Wang, L. *et al.* A cytoplasmic COMPASS is necessary for cell survival and triple-negative breast cancer pathogenesis by regulating metabolism. *Genes Dev.* **31,** 2056–2066 (2017).

33. Bledau, A. S. *et al.* The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Dev.* **141,** 1022–1035 (2014).

34. Liu, D., Archer, N., Duesing, K., Hannan, G. & Keast, R. Mechanism of fat taste perception: Association with diet and obesity. *Progress in Lipid Research* **63,** 41–49 (2016).

35. Yang, P. *et al.* Dietary oleic acid-induced CD36 promotes cervical cancer cell growth and metastasis via up-regulation Src/ERK pathway. *Cancer Lett.* **438,** 76-85 (2018).

36. Ubellacker, J. M. *et al.* Lymph protects metastasizing melanoma cells from ferroptosis. *Nature* **585,** 113–118 (2020).

37. Vriens, K. *et al.* Evidence for an alternative fatty acid desaturation pathway increasing cancer plasticity. *Nature* **566,** 403-406 (2019).

Acknowledgements

Research in the S.A.B. lab is supported partially by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 787041), the Government of Cataluña (SGR grant), the Government of Spain (MINECO), the La Marató/TV3 Foundation, the Foundation Lilliane Bettencourt, the Spanish Association for Cancer Research (AECC) and The Worldwide Cancer Research Foundation (WCRF). D.G. was supported by a La Caixa International Fellowship for Doctoral studies, and C.B., by an FPI fellowship (MINECO); I.H., by the EU H2020 Marie Skłodowska-Curie award (GA No. 754510). H.H. has received funding from the Ministerio de Ciencia, Innovación y Universidades (SAF2017-89109-P; AEI/FEDER, UE). Studies in the Shilatifard laboratory related to COMPASS are supported by NCI's Outstanding Investigator Award R35CA197569; other research, by funding from University of Miami Miller School of Medicine, Sylvester Comprehensive Cancer Center, grant R01 GM078455, and DP1 CA228041 from the National Institute of Health (to R.S.) and the National Cancer Institute of the National Institutes of Health (Award Number P30CA240139) (note that the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health). The IRB Barcelona is a Severo Ochoa Center of Excellence (MINECO award SEV-2015-0505). We would like to thank the histology and genomics facilities of the IRB Barcelona for their assistance in this work. We thank Veronica Raker for manuscript editing.

Author contributions

G.P., D.D. and S.A.B. designed the study. C.G. helped to design some of the experiments. G.P. performed all *in vivo* palm oil and olive oil diet studies and analyzed the innervation phenotype, including all *in vivo* transcriptome analyses of tumour and neural compartments (bulk arrays, 10X sc RNA-seq and positional RNA-seq). D.D. performed all *in vitro* and *in vivo* epigenetic studies and the mechanistic studies for the Set1A, MLL1 and MLL2 proteins. G.P. performed

the experiments in melanoma. C.L. analysed the scRNA-seq and gene expression data. M.E. analysed the 10X scRNA-seq and the positional RNA-seq data. C.B. performed the correlative analyses of the CD36+ signature and perineuronal nets in human tumours. M.A. helped to perform the sympathectomy experiments. I.H. and A.S. analyzed ChIP-seq data. D.D. performed the *in vitro* scRNA-sequencing experiment. S.R.G., I.H. and H.H. contributed to and analyzed scRNA-seq data. N.P. characterised histology samples. F.B and R.S. performed some transcriptome analyses. C.B. provided the tumour samples to establish the patient-derived VDH15 oral cancer cell line. S.A.B. wrote the manuscript with the input of G.P and D.D.

Competing Interests

S.A.B. is a co-founder and scientific advisor of *ONA Therapeutics*. The remaining authors declare no competing interests. Primary contacts: S.A.B., salvador.aznar-benitah@irbbarcelona.org; G.P., gloria.pascual@irbbarcelona.org; A.S., ASH@northwestern.edu

Main Figure Legends

Figure 1. OSCC metastatic cells are selectively sensitive to PA and display a metastatic memory. a, Schematic diagram of an in vitro 4-day fatty acid (FA) treatment of OSCCs cells followed by FA withdrawal and orthotopic injection into NSG mice 14 days after *in vitro* cell culture. **b,** Frequency of developed lung metastasis (Lung Met) of animals injected as schematized in a, expressed as percentages. n = 15 mice per group; lung metastases: $***P = 0.0007$, $*P = 0.04$; two-tailed Fisher's exact test. PA (palmitic acid), SA (stearic acid), LA (linoleic acid), OA (oleic acid), PA/OA denotes a 4-days treatment with PA followed by a 4-days treatment with OA and 14-days in culture without treatment. **c,** Schematic diagram of the *in vivo* diet experiments

whereby OSCCs were exposed to a high-fat diet in primary recipient mice and injected into secondary recipients with a control diet. **d,** Frequency of developed lymph node metastasis (LN Met) and lung metastasis (Lung Met) of animals injected as schematized in c) expressed as percentages. For LN Met, **P* = 0.003; ***P* = 0.02; for Lung Met, **P* = 0.01, ****P* = 0.0004; two-tailed Fisher's exact test.

Figure 2. Palmitic acid induces stable H3K4me3 chromatin changes in OSCC cells. a, H3K4me3 ChIP-seq analysis of CTRL pLKO.1 SCC-25 cells upon *in vitro* PA treatment. TSS-centered heatmaps of H3K4me3 memory (differential) and non-memory (non-differential) peaks for 14D Untreated and 14D post-PA conditions. **b,** TSS-centered density plot showing H3K4me3 signal of memory peaks between 14D post-PA and control conditions. **c,** H3K4me3 representative PA memory peaks (CHRDL2, OLIG1 and NEFM genes) for all tested conditions (4/14D Untreated/PA-treated cells) shown at the same scale (Diff. Bind FDR≤0.1).

Figure 3. **EGR2 and galanin regulate the palm oil diet pro-metastatic memory. a,** Top biological processes Gene Ontology (GO) terms up-regulated in [$CD36^{bright}CD44^{bright}$]–sorted populations from secondary (2ary) primary tumours (PTs) palm oil diet–exposed in primary recipients. **b,** Representative peaks within the galanin (GAL) locus in 4D/14D post-PA SCC-25 pLKO.1 samples. **c,** Top biological processes GO terms significantly up-regulated in 4D PA-treated EGR2 ChIP-seq samples compared to 4D Untreated samples for SCC-25 pLKO.1 cells (Neural/lipid-related terms are highlighted in purple and blue respectively). **d,** Frequency of LN Mets from 2ary recipients after orthotopic injection of primary recipient–PT cells, derived from control (pLKO.1), shRNA-knockdown of EGR2 (shEGR2 #38_9 and shEGR2 #40_9) or shRNA-knockdown of GAL (shGAL #74_4 shRNAs) SCC-25 cells; primary recipients were fed a normal (CTRL) or palm oil–rich (PALM) diet. (pLKO.1/control diet, $n = 22$; pLKO.1/PALM, $n = 21$; shEGR2/control diet and shEGR2/PALM, $n = 20$) (n reflects 20 per shRNA used);

shGAL/control diet and shGAL/PALM, $n = 10$ (each). *$P = 0.03$, two-tailed Fisher's exact test; *n.s.*, not significant.

Figure 4. **Tumour cells with a PA-induced metastatic memory activate tumour-associated Schwann cells. a,** 10X scRNA-seq data showing the odds ratio (OR) due to a palm memory within the annotated tumour stroma clusters. OR and *P*-values are shown for control (pLKO.1) or shCD36 groups. ** denotes OR > 10. **b,** 10X scRNA-seq UMAP plot showing the cluster distribution and expression levels of selected gene markers. **c,** Frequency of developed metastases of secondary recipient mice injected with wild-type VDH-15 (WT) or LV-chondroitinase ABC enzyme (chABC)–derived cells from primary recipient mice fed a normal diet or a palm oil–rich diet. For WT–VDH-15, $n = 12$ mice per group; for chABC–VDH-15, $n = 10$ mice per group. (For lymph node met frequency, *$P = 0.03$, **$P = 0.008$; for lung metastasis, **$P = 0.002$, **$P = 0.001$; two-tailed Fisher's exact test; frequency is expressed as the percentages).

Extended Data Figure Legends

Extended Data Figure 1. a, Flow cytometry analysis of *in vitro* cultured SCC-25 cells after 4 days of fatty acid stimulation. The percentage of CD36 membrane expression and cell viability (as measured by DAPI incorporation) are shown. **b,** BLI quantification of lymph node (LN) metastases, showing number and size, from mice inoculated with untreated SCC-25 cells ($n = 17$) or *in vitro* treated with 300 μM PA ($n = 17$), 50 μM OA ($n = 10$) or 50 μM LA ($n = 10$). BLI signals are expressed as the relative normalized photon flux. $P = 0.05$, two-tailed *t*-test. **c,** Frequency of developed LN metastases from animals in b. *$P = 0.01$, **$P = 0.001$; two-tailed Fisher's exact test. **d,** Flow cytometry analysis of *in vitro* cultured SCC-25 cells immediately after 4 days (4D) PA stimulation or at 14D after PA withdrawal (wdl). The numbers indicate [$CD44^{bright}CD36^{bright}$] and [$CD44^{bright}CD36^{dim}$] populations in the represented gate, expressed as percentage of total DAPI-negative cells from *in vitro* cultured cells (samples are representative of $n = 5$ independent experiments). **e,** BLI quantification of primary tumours generated from mice injected with SCC-25 cells *in vitro* 14D after PA withdrawal. The BLI signal is expressed as the

relative normalized photon flux. Data are given as the mean and s.e.m. (untreated, $n = 14$; 14D after PA wdl, $n = 14$, $^{*}P = 0.05$, two-tailed *t*-test). **f,** Frequency of developed LN metastases from animals in e ($^{*}P = 0.02$, two-tailed Fisher's exact test). **g,** *ex vivo* BLI lung metastasis quantification of mice injected with VDH-15 cells at 14D fatty acid withdrawal (PA, palmitic acid; SA, stearic acid; OA, oleic acid; LA, linoleic acid). BLI signals are expressed as the relative normalized photon flux. Images are representative of $n = 15$ mice per group. **h,** Frequency of developed lung metastasis of mice injected with SCC-25 after FA withdrawal expressed as percentages (n = 10 mice per group, lung metastases: $^{*}P = 0.05$; two-tailed Fisher's exact test). PA/OA denotes a 4-day treatment with palmitic acid followed by a 4-days treatment with oleic acid, followed by 14-day with no fatty acid. **i,** BLI quantification of FACS-sorted and serially-transplanted PT populations (as indicated) from SCC-25 primary recipients. BLI signals are shown as the normalized photon flux ($n = 10$ mice per group. $^{*}P = 0.03$, $^{**}P = 0.05$, $^{***}P = 0.01$; two-tailed t-test). **j,** Frequency of developed LN metastases of mice in i ($^{*}P = 0.05$, two-tailed Fisher's exact test).

Extended Data Figure 2. a, Schematic diagram representing the *in vivo* diet experiments in which OSCC that were exposed to a high-fat diet in primary NSG mice recipients were injected into secondary recipients that were maintained on a control diet. At the final point primary tumour (PT) cells were purified by FACS-SORT for molecular characterization. **b,** Flow cytometry analysis of PTs from VDH-15– and SCC-25–injected primary recipients fed a high-fat or control diet, as schematized in a. Numbers indicate [$CD44^{bright}CD36^{bright}$] and [$CD44^{bright}CD36^{dim}$] populations in the represented gate, expressed as percentages of total DAPI–GFP+Lin– cells. Samples are representative of $n = 3$ independent experiments. **c, d,** BLI tumour monitoring of secondary recipients injected with VDH-15–derived cells from primary recipient mice fed with palm oil–rich, olive oil–rich diet and normal diet (as control). BLI signals are shown as the normalized photon flux ($n = 20$ mice per group, two independent experiments. LN metastasis, $P = 0.04$; lung metastasis, $^{*}P = 0.04$, $^{**}P = 0.007$; two-tailed *t*-test). **e,** Frequency of developed LN metastases from animals in c expressed as percentage (for LN metastasis, $^{*}P = 0.003$;

P = 0.02; for lung metastasis: *P = 0.01, ***P = 0.0004, two-tailed Fisher's exact test). **e, f, BLI metastasis quantification of secondary recipient mice injected with control pLKO.1 or shCD36 VDH-15- (f) and SCC-25- (g) derived cells from primary recipient mice fed with a high-fat palm oil or a control diet. BLI signals are expressed as the normalized photon flux (for VDH-15, n = 10 mice per group; for lymph node and lung met, *P = 0.04; *n.s.*, not significant; two-tailed *t*-test, data are mean ± s.e.m.; for SCC-25, n = 20 mice per group, data from two independent experiments; *P = 0.03, ****P < 0.0001; two-tailed *t*-test, data are mean ± s.e.m.). **g,** Frequency of developed metastases from animals in g expressed as percentage (for lymph node met, ***P = 0.003, **P = 0.0004; for lung met, ****P < 0.0001, two-tailed Fisher's exact test). **h,** *Ex vivo* BLI lung metastasis monitoring from secondary recipient mice injected with control pLKO.1 or shCD36 melanoma-derived cells from primary recipient mice fed with a high-fat palm oil or a control diet. Pictures are representative of n = 5 mice per group. **i,** BLI metastasis quantification of animals in i. BLI signals are expressed as the normalized photon flux. (***P = 0.0009, ****P <0.0001, two-tailed *t*-test, data are mean ± s.e.m.).

Extended Data Figure 3. **a,** Frequency of developed lung metastasis from secondary mice injected with SCC-25–derived from primary recipient mice fed a normal or palm oil–rich diet; Primary recipients were treated with an anti-CD36 neutralizing antibody (JC63.1) or a control isotype (IgA). Data is expressed as the percentages. (n = 13 mice per group; *P = 0.01, *P = 0.03, two-tailed Fisher's exact test). **b, c,** *ex vivo* BLI imaging and quantification of lung metastasis from VDH-15 secondary recipients injected with a doxycycline-inducible shCD36 (shCD36 #23 or #76) VDH-15–derived tumour cells from primary recipient mice fed a normal or palm oil–rich diet. The secondary mice were untreated or continuously doxycycline-treated. (Data are mean ± s.e.m. Images are representative of n = 10 mice per group; *P = 0.01; **P = 0.001; ****P < 0.0001; two-tailed *t*-test). **d, e,** TSS-centered heatmaps of d) H3K4me1 and e) H3K27me3 non-memory peaks for SCC-25 pLKO.1 14D Untreated and 14D post-PA conditions. **f,** TSS-centered density plot showing the signal of three histone marks of interest (H3K4me3, H3K4me1 and H3K27me3) in the genomic areas where H3K4me3 memory peaks were identified. **g,** ChIP-seq

signal TSS-centered distribution plots of H3K4me1 and H3K27ac histone marks used to map enhancer regions in SCC-25 pLKO.1 cells, both at 4D (upper panel) and 14D (bottom panel) time points for Untreated and PA-treated conditions. A total of 2,183 enhancers were mapped. **h,** Volcano plots showing the enhancer regions displaying differential transcription of eRNAs (FC>+/-1.6, P<0.05) at 4D PA (upper panel) or 14D post-PA (bottom panel) according to our Pol II traveling ratio analysis (PRO-seq). Up-/down-regulated eRNAs are colored in red and blue respectively. **i,** PCA plots of 4D (left) and 14D (right) Untreated/PA-treated H3K9me3 ChIP-seq sSCC-25 pLKO.1 samples. **j,** Images showing representative H3K9me3 neural-related peaks (NRXN3 and GABRG3 genes) lost in 4D PA and 14D post-PA SCC-25 pLKO.1 cells, as compared to control samples.

Extended Data Figure 4. a, PCA plots of 4D/14D Untreated/PA-treated H3K4me3 (left), H3K4me1 (center) and H3K27me3 (right) ChIP-seq samples of CTRL pLKO.1 SCC-25 cells. On the side, PCA plot of 14D Untreated/PA-treated H3K4me1 ChIP-seq samples of CTRL pLKO.1 VDH-15 cells. **b,** PCA plots of 14D Untreated/OA -treated H3K4me3 ChIP-seq samples of CTRL pLKO.1 SCC-25 (left)/VDH-15 (right) cells. The bottom figures show representative H3K4me3 peaks (NEFM and CHRDL2 genes) at 14D post-OA for CTRL pLKO.1 SCC-25/VDH-15 cells. **c,** H3K4me3 ChIP-seq from secondary (2ary) primary tumors (PTs) of CTRL pLKO.1 VDH-15 cells upon *in vivo* exposure to Control (CTRL)/Palm oil (PALM)-enriched diets. On the bottom, a table displaying the number of total, differentially up-/down-regulated H3K4me3 peaks identified when comparing CTRL and PALM diet samples. **d,** Left, TSS-centered heatmaps of H3K4me3 memory and non-memory representative peaks for 2ary CTRL/PALM PTs; middle plot, TSS-centered density plot showing the H3K4me3 signal in differential peaks for both conditions assessed**;** right, representative *in vivo* memory peaks (FABP3, DRGX and CNTFR genes) for both CTRL/PALM tumors (Diff.Bind FDR≤0.1). **e,** Venn diagrams showing the overlap between genes harboring total (non-differential) H3K4me3 ChIP-seq peaks for VDH-15 pLKO.1 (left) Untreated 14D and CTRL diet 2ary recipient *in vivo* samples or (right) 14D post-PA and PALM diet 2ary recipient *in vivo* conditions. **f,** Venn

diagrams showing the overlap between differential H3K4me3 ChIP-seq peaks-bearing genes for VDH-15 pLKO.1 14D post-PA and PALM diet 2ary recipient *in vivo* conditions. (a,b: *a representation factor >/<1 indicates a higher/lower enrichment than expected by chance in gene overlap; P-Values were calculated using a hypergeometric test*). **g, h,** Bar plots showing the top biological processes GO terms built by the g) common or h) unique genes indicated in f). Unique refers to the PALM diet 2ary recipient *in vivo* condition. Neural-related terms are highlighted in purple.

Extended Data Figure 5. a, Left, PCA plot of 4D/14D Untreated/PA-treated H3K27ac ChIP-seq samples of SCC-25 pLKO.1 cells; middle, TSS-centered density plot showing the H3K27ac signal of all differential peaks in 4D/14D PA samples as compared to their control counterparts; right, Image showing representative H3K27ac peaks (ANGPTL4 gene) gained upon 4D PA treatment in SCC-25 pLKO.1 cells, as compared to the corresponding 14D samples. **b,** Left, plot displaying the three main clusters (0, 1 and 2) detected upon scRNA-seq data tSNE analysis of CTRL pLKO.1 SCC-25 cells after 4D PA exposure. On the side, trajectory plot displaying the predicted Cluster distribution of 4D Untreated/PA-treated individual cells; right, trajectory plot showing the distribution of 4D Untreated (blue)/PA-treated (red) cells and their corresponding clusters. **c,** tSNE plot showing distribution of cells enriched in the 4D PA transcriptional signature; on the side, trajectory plot displaying the predicted distribution of the PA response score of each cell analyzed; right, bar plot showing the quantification of the Cluster distribution of 4D Untreated/PA-treated cells shown as the proportion of total cells per condition. **d,** Cell cycle Analysis using Propidium Iodide; PI staining. Representative FACS plots displaying the PI cell cycle profiles of both 4/14D Untreated/ PA-treated cells. **e,** Box plots showing the mRNA expression detected by RNA-seq in Untreated and PA-treated SCC-25 pLKO.1 cells at 4D /14D in those regions displaying H3K4me3 PA-driven changes (UP/DOWN in 4D/14D PA box plots *****P<0.0001 of a two-tailed t-test).*

Extended Data Figure 6. a, Principal component analysis (PCA) from microarray data of secondary (2ary) recipient-primary tumours (PTs)-sorted [$CD36^{bright}CD44^{bright}$] or [$CD36^{dim}CD44^{bright}$]. Diets are indicated as (B) primary recipient mice fed a normal (red), olive oil–rich (green) or palm oil–rich (blue) diet, and secondary mice fed with a normal diet and (A) primary mice fed with a normal diet and secondary mice fed with a palm oil–rich diet (fuchsia). The axis shows the percent variability covered by each of the represented components. **b,** Heatmap displaying the DGE levels of 2ary VDH-15–derived PTs with a palm oil memory. CD36b, [$CD36^{bright}CD44^{bright}$] and CD36d, [$CD36^{dim}CD44^{bright}$]. **c,** Gene ontology (GO) and Gene Set Enrichment (GSEA) analysis showing top categories from biological processes that are upregulated in palm oil diet–memory from [$CD36^{bright}CD44^{bright}$]-sorted populations from secondary PTs. For GO analysis, $FC > 1.5$, $P < 0.05$). **d,** Principal component analysis (PCA) from microarray data of secondary recipient-PTs control pLKO.1– or shCD36 sorted [$CD36^{bright}CD44^{bright}$] or [$CD36^{dim}CD44^{bright}$], ±palm oil memory. **e,** Gene Set Enrichment (GSEA) analysis showing top categories from biological processes that were upregulated in palm oil diet–memory from [$CD36^{bright}CD44^{bright}$]-sorted populations from secondary PTs. **f, g,** Principal component analysis (PCA) and heatmap plot from the microarray data of cells FACS-sorted from secondary PTs from control (pLKO.1) or shCD36 melanoma ± palm oil memory. In f, the axis shows the percent variability covered by each of the represented components. The heatmap plot (g) shows the differentially expressed genes in control (pLKO.1)–palm oil memory PTs as compared to control–normal diet and their correspondence with the gene expression levels from the shCD36–palm oil. **h,** GO and GSEA analysis showing the top biological process categories that were upregulated by a palm oil memory of [GFP+]-sorted populations from secondary recipient, melanoma-derived PTs, as analysed in RNA microarrays. For GO analysis, $FC > 1.5$, $P < 0.05$.

Extended Data Figure 7. a, Diagram of the experimental *setup*. Secondary (2ary) recipient mice were injected on day 0 with OSCC from primary tumours (PTs) ± palm oil memory. 2ary recipients were treated with the neurotoxin 6-hydroxydopamine (6-OHDA; to induce the

apoptosis of dopaminergic neurons) or vehicle on days –6, –3 and +3 relative to the day of injection. **b,** Frequency of developed lung metastasis from animals treated as shown in a, expressed as the percentages (*P = 0.03, **P = 0.001; two-tailed Fisher's exact test). **c,** Tyrosine hydroxylase (TH) immunofluorescence analysis of VDH-15 primary tumours (PTs) from secondary recipients ± palm oil memory after vehicle or 6-OHDA treatment. Nuclei are stained with DAPI. KT-14, cytokeratin 14. Note that 6-OHDA-lesioned-tumours display a marked reduction in the expression level of TH. Images are representative of n = 3 biological replicates per group. **d,** Top common predicted binding site motifs in promoter regions of the co-regulated neural-related genes in SCC-25–or VDH-15–primary tumours with a palm memory. Z-scores and P values are shown for each cell line. **e,** Integrative gene set enrichment analysis (GSEA) from PTs of secondary recipients SCC-25–, VDH-15– or melanoma tumour–derived cells ± palm oil memory. The graph shows the biological processes enrichment in palm oil memory compared to control diet. On the right side, detail of the neuropeptide signalling pathway–associated gene-set assayed by integrative GSEA showing galanin (*GAL*) as the top significantly represented gene in palm oil memory tumours. **f, g,** PCA plots of differentially bound signal (DBS) regions detected for 4D Untreated/PA-treated EGR2 ChIP-seq samples in f) SCC-25 pLKO.1 and g) SCC-25 CD36 KD cells. **h,** Heatmaps showing the EGR2 DBS regions (differential peaks) detected for 4D SCC-25 pLKO.1 (left) and SCC-25 CD36 KD (right) shown in f) and g) respectively (FDR<0.05).

Extended Data Figure 8. a, PCA plots of differentially bound signal (DBS) regions detected for 4D (left) and 14D (right) Untreated/PA-treated H3K4me3 ChIP-seq samples in SCC-25 EGR2 KD cells. **b,** Bar plots showing the top biological processes GO terms uniquely down-regulated in 4D PA (upper panel) and 14D post-PA (bottom panel) H3K4me3 ChIP-seq SCC-25 EGR2 KD samples. Neural-related terms are highlighted in purple**. c,** Venn diagrams showing the overlap between *in vivo* proneural-induced gene signatures in secondary OSCC primary tumours with a palm oil memory and the knockdowns, as indicated. **d,** Gene set enrichment analysis (GSEA)

showing the negative enrichment of biological processes in secondary PTs derived from SCC-25–shEGR2/PALM vs pLKO.1/control diet or shGAL/PALM vs pLKO.1/control diet. **e,** BLI metastasis quantification of lymph node (LN) and lung of secondary recipient mice injected with primary recipient SCC-25 cells, derived from control (pLKO.1), shRNA-knockdown of EGR2 (shEGR2 #38_9 and shEGR2 #40_9) or shRNA-knockdown of GAL (shGAL #74_4 shRNAs) ±palm oil memory (pLKO.1/control diet, $n = 22$; pLKO.1/PALM, $n = 21$; shEGR2/control diet and shEGR2/PALM, $n = 20$; shGAL/control diet and shGAL/PALM, $n = 10$; in the knockdowns, n reflects number of mice per shRNA used; $^*P = 0.01$, two-tailed t-test. Data are the mean ±s.e.m.).

Extended Data Figure 9. a, Experimental setup of the experiment for galanin receptor inhibition. **b, c,** *Ex vivo* BLI lung metastasis (b) and frequency of lung metastasis development (c), expressed as the percentages, from secondary (2ary) recipient mice injected with VDH-15 cells derived from primary recipient mice fed a palm oil–rich diet or a control diet. 2ary recipients were treated with the galanin receptor antagonist galantide (M15) or vehicle as explained in (a) (vehicle-treated, $n = 12$ mice; M15-treated, $n = 15$ mice. $^*P = 0.03$; $^*P = 0.04$; two-tailed Fisher's exact test).

Extended Data Figure 10. a, b, Western blot analysis of the COMPASS family of methyltransferases (Set1A/B and MLL1/2) in VDH-15 or SCC-25 pLKO.1 cells that were Untreated, a) 4D after PA treatment or b) 14D after removal of PA. Cells were infected with a non-targeting (nt)-shRNA (pLKO.1). Hsp90 was used as a loading control. **c, d,** PCA plots displaying ChIP-seq for c) MLL1or d) MLL2 from untreated or 14D post-PA VDH-15 pLKO.1 cells. **e,** Representative peaks from MLL1/MLL2 ChIP-seq of MLL1/MLL2-regulated genes (*HOX* gene cluster) and non-regulated genes (neural-related *CHRDL2* gene) for all conditions tested (U ntreated/4D PA and 14D post-PA pLKO.1 VDH-15 cells). **f,** PCA plot showing 14D Untreated/PA-treated H3K4me3 ChIP-seq samples of CTRL pLKO.1 and Set1A #42 KD VDH-15 cells. **g,** GO analysis of PA-treated Set1A KD VDH-15 H3K4me3 ChIP-seq samples. The plot

shows the top biological processes GO terms significantly down-regulated in Set1A KD #42 samples 14D post-PA exposure when compared to the 14D Untreated samples (Diff. Bind FDR≤0.05). On the side, H3K4me3 peaks of Set1A-regulated neural genes (CHRDL1 and GRIP2 genes). H3K4me3 representative examples are shown at 4/14D time points for all assessed conditions (CTRL pLKO.1 and Set1A #42 KD VDH-15 cells). **h,** Left, secondary (2ary) recipient orthotopic injections of CTRL pLKO.1 and Set1A #42 KD VDH-15 cells upon *in vivo* exposure to Control/Palm-oil enriched diets in primary recipient mice. The frequency of developed LNmets (at the end-point of the experiment is shown for all conditions (n= 10 mice/group; *LNmets: CTRL pLKO.1 CTRL Diet vs CTRL pLKO.1 PALM Diet n.s= 0.065, CTRL pLKO.1 PALM Diet vs Set1A KD PALM Diet n.s= 0.091; n.s.* is not significant Fisher's exact test). Right, frequency of developed LNmets at the end-point of the experiment in 1ary recipient orthotopic injections of CTRL pLKO.1/Set1A KD #07 VDH-15 cells 14D post-PA *in vitro* exposure (*n= 20 mice/group; LNmets: CTRL pLKO.1 UNTR vs CTRL pLKO.1 14D post-PA *P=0.038, CTRL pLKO.1 14D post-PA vs Set1A KD 14D post-PA *P=0.031 of Fisher's exact test).*

Extended Data Figure 11. a, Flow cytometry analysis of SCC-25-primary tumours (PTs) from secondary (2ary) recipients ± palm oil memory. [GFP–/CD31–/CD45–] tumour stroma–selected cells were purified and processed to perform RNA microarrays. Data are representative of $n = 3$ independent experiments. Numbers indicate the population in the represented gate, expressed as the percentages. **b,** Principal component analysis (PCA) of the microarray data of VDH-15–associated bulk stroma purified from secondary (2ary) recipients. Diet samples are indicated as (B) primary recipient mice fed a normal (red), olive oil–rich (green) or palm oil–rich (blue) diet, and 2ary mice fed with a normal diet and (A) primary mice fed with a normal diet and 2ary mice fed with a palm oil–rich diet (fuchsia). The axis shows the percent variability covered by each of the represented components. **c,** Principal component analysis (PCA) from microarray data of 2ary recipient tumour–associated stroma derived from injection in primary recipients of control (pLKO.1)– or shCD36–VDH-15 or SCC-25 cells, as indicated by the colour-code. The axis shows the percent variability covered by each of the represented components. **d,** Volcano plots for the

differential gene expression microarray analysis and gene ontology (GO) analysis from a purified tumour stroma from 2ary recipient mice injected with OSCC from PTs ± palm oil memory. Volcano plots show the up- or downregulated genes in control–palm oil memory tumours compared to control–normal diet tumours (top) and the expression of these genes in the shCD36–palm oil (lower plot). Data from $n = 2$ biological replicates; fold-change, $FC > 2$, $P < 0.01$. The GO analysis shows the top biological processes categories upregulated in the palm oil memory tumour stroma, $FC > 2$, $P < 0.05$. (Integrated analysis, derived from VDH-15 or SCC-25 cells). **e,** Neural–mouse stromal signature induced in 2ary recipients after orthotopic injection of primary recipient–PT cells, derived from control (pLKO.1), shRNA-CD36, shRNA-EGR2 or shRNA-GAL SCC-25 cell. Each dot in the plot represents a neural gene. **f, g,** GSEA showing the negative enrichment of biological processes in neural–mouse stroma associated with 2ary PTs from shEGR2/PALM or shGAL/PALM vs pLKO.1/control diet conditions.

Extended Data Figure 12. a, Functional network interaction between tumour–mouse stromal palm oil (PALM)–controlled genes and the human OSCC primary tumours with a palm memory. The network graph shows the co-regulated functional nodes of interaction between the two compartments (fold-change, $FC > 1.5$; $P\ value < 0.05$). **b, c,** 10X single-cell (sc) RNA-seq clustering analysis of the tumour-associated stroma purified from secondary (2ary) oral SCC-25 PTs ±palm oil memory. The principal component UMAP plot is shown in which the specific cell types have been annotated to each respective cluster. **d,** Overlap analysis represented by Venn diagrams showing the intersection between bulk-stroma palm memory-controlled genes and 10X single-cell (sc) RNA-seq clusters from SCC-25–derived stromal cells in 2ary recipients. Representation factor (RF) and *P values* are shown for the overlap. Hypergeometric test; estimated number of protein-encoding genes = 25,000.

Extended Data Figure 13. a, b, c, Integrated UMAP cluster visualization, annotated for cell types, of 10X single-cell (sc) RNA-seq data of the tumour-associated stroma purified from secondary (2ary) oral SCC-25 primary tumours (PTs) ±palm oil memory. The UMAP plot shows

the expression level and cluster distribution of selected gene markers relative to glial cells and progenitor glia a), specialized extracellular matrix constituent (ECM) b) and nerve injury/ nerve regenerative processes c). **d,** Regression analysis showing the correlation between the long-lasting palm oil–related signature of the OSCC tumours (derived from SCC-25 or VDH-15 cells) or melanoma (501-mel), and the expression of markers related with perineuronal nets (GO:0072534). R-squared coefficients (R) and *P values* are shown for each analysis. **e,** Immunofluorescence analysis of PTs from 2ary recipient mice injected with VDH-15–derived cells from primary recipient ± palm oil memory. Cytokeratin-14 (KT-14) is shown as epithelial marker of OSCCs, and the specialized ECM markers Hyaluronidase 1 (Has1), Tenascin R (TNR) and the glial/Schwann marker s100 are shown as tumour stroma markers. Yellow arrows indicate areas of Has1 positive cells. Dashed lines delimitate the interface between the tumour and the tumour-associated stroma. **f,** Magnification from b) (palm diet memory condition). Note the double labelling of TNR and S100 in the close proximity of the tumour front. Images are representative of $n = 4$ biological replicates.

Extended Data Figure 14. a, Images showing Haematoxylin-Eosin (H&E) staining of primary tumours (PTs) from secondary (2ary) recipients injected with control (pLKO.1) and shCD36 SCC-25 cells derived from primary recipient mice fed a normal diet or a palm oil–rich diet. **e,** Spatial transcriptomics analysis of tumours in a) showing the proportion content and spatial distribution of mouse and human transcriptome per spot, the analysis stratification of the tissue as healthy/tumour invasive front/tumour and the proportion content and spatial distribution of the tumour-associated Schwann cells. **c,** Quantitative analysis of the proportional content of tumour-associated Schwann cells within each compartment (healthy/tumour front/tumour) of the PTs from 2ary recipients injected with control (pLKO.1) and shCD36 SCC-25 cells derived from primary recipient mice fed a normal diet or a palm oil–rich diet.

Extended Data Figure 15. a, b. e, Immunofluorescence analysis and quantification of primary tumours (PTs) from secondary (2ary) recipient mice injected with VDH-15– or SCC-25-derived

cells, as indicated, derived from primary recipient mice fed a normal diet, a palm oil–rich or an olive oil–rich diet. Cytokeratin-14 (KT-14) is shown as an epithelial marker of OSCC, and the glial/Schwann markers s100 or GAP43 as a tumour stroma marker. Nuclei are stained with DAPI. Images are representative of $n = 3$ independent experiments. **c,** Graphs showing the values of integrated density of (a, b). ($n = 3$ biological replicates per group; $^{*}P = 0.05$, $^{****}P < 0.0001$; two-tailed *t*-test). **f,** Immunofluorescence analysis of PTs from M15-treated or vehicle-treated mice ± palm oil memory. The expression of cytokeratin-14 (KT-14) is shown as an epithelial marker of OSCC, and the glial Schwann cell marker s100 is shown in the tumour-associated stroma. Nuclei are stained with DAPI. Note that the increased expression of s100 in the stroma of palm memory tumours is slightly reduced in the case of M15-treated palm condition.

Extended Data Figure 16. a, Western blot analysis of total protein extracted from *in vitro* cultured VDH-15 cells infected with the lentiviral LV-chondroitinase ABC (ch-ABC) vector, showing the overexpression of the chondroitinase ABC enzyme in the infected cells ($n = 2$ replicates per group). **b, c,** Immunofluorescence analysis of primary tumours from secondary recipient mice injected with wild-type VDH-15 (WT) or LV-chondroitinase ABC enzyme (chABC)–derived cells from primary recipient mice fed a normal diet or a palm oil–rich diet. Cytokeratin-14 (KT-14) is shown as epithelial marker of OSCCs, and the specialized ECM markers Versican (Vcan), Tenascin R (TNR) and the glial/Schwann marker s100 as tumour stroma markers d) or collagen 5A1 (Col5A1) as ECM marker e). Nuclei are stained with DAPI. Images are representative of $n = 3$ biological samples.

Extended Data Table 1. a, Frequency of developed primary tumours (PTs) from mice inoculated with untreated SCC-25 cells ($n = 17$) or *in vitro* treated with 300 μM PA ($n = 17$), 50 μM OA ($n = 10$) or 50 μM LA ($n = 10$). **b,** BLI quantification of PTs generated from mice injected with SCC-25 cells *in vitro* untreated or palmitic acid (PA)-treated, 14 days after PA withdrawal. The BLI signal is expressed as the relative normalized photon flux (untreated, $n = 14$; 14D after PA

withdrawal, *n* = 14). **c,** Frequency of developed PTs from mice in (b). **d,** Flow cytometry analysis of primary tumours (PTs) from the control empty vector (pLKO.1) or shCD36 SCC-25; prior to injection, SCC-25 cells were either treated *in vitro* with palmitic acid (PA) for 4 days followed by 14 days PA withdrawal or not treated, as indicated. Numbers indicate [$CD44^{bright}CD36^{bright}$] and [$CD44^{bright}CD36^{dim}$] expressed as percentages of total DAPI-GFP+Lin– cells (samples are representative of *n* = 3 independent experiments). **e,** BLI quantification of FACS-sorted and serially-transplanted tumour populations from SCC-25 primary recipients. The BLI signal is expressed as the normalized photon flux (*n* = 10 mice per group). **f,** Frequency of developed PTs and lung metastasis of mice injected with VDH-15 cells, FA (fatty acid)-treated or untreated, at 14 days after FA withdrawal (*n* = 15 mice per group; PA (palmitic acid), SA (stearic acid), LA (linoleic acid), OA (oleic acid), PA/OA denotes a 4-day treatment with palmitic acid followed by a 4-days treatment with oleic acid, followed by 14-day with no fatty acid. **g,** Lung BLI quantification of mice in (f) BLI signals are expressed as the relative normalized photon flux. **h,** Frequency of developed PTs of mice injected with VDH-15 cells, FA (fatty acid)-treated or untreated, at 14 days after FA withdrawal (*n* = 15 mice per group; PA (palmitic acid), SA (stearic acid), LA (linoleic acid), OA (oleic acid), PA/OA denotes a 4-day treatment with palmitic acid followed by a 4-days treatment with oleic acid, followed by 14-day with no fatty acid; BLI signals are expressed as the relative normalized photon flux). **i,** *ex vivo* BLI lung metastasis quantification of mice in (h). **j,** BLI primary tumour quantification of secondary (2ary) recipient mice injected with SCC-25 control (pLKO.1) or shCD36–derived tumour cells from primary recipient mice ±palm oil memory. BLI signals are expressed as the normalized photon flux (*n* = 20 mice per group, data from two independent experiments). **k,** Frequency of developed PTs from 2ry mice in (j). **l,** BLI primary tumour quantification of 2ary recipient mice injected with SCC-25-derived cells from primary recipient mice Primary recipients were treated with an anti-CD36 neutralizing antibody (JC63.1) or a control isotype (IgA). BLI signals are expressed as the normalized photon flux. **m,** *Ex vivo* BLI quantification of lung metastasis from the 2ary mice in (l). BLI signals are expressed as the normalized photon flux. **n,** Frequency of developed PTs and lung metastasis from the 2ary mice in (l). (For IgA treatment, *n* = 17 mice per group; for anti-CD36 JC63.1

treatment, *n* = 13 mice per group). **o,** Flow cytometry analysis of *in vitro* VDH-15 cells infected with a control vector (a doxycycline-inducible scrambled siRNA) or a doxycycline-inducible shCD36 #23 or shCD36 #76 vector. The cells were *in vitro*-treated with doxycycline (2 μg ml^{-1}) and collected after 24 h to analyze the expression of RFP and CD36. Numbers indicate the population in the represented gate, expressed as the percentages. **p,** BLI primary tumour quantification of 2ary recipient mice injected with VDH-15 with a doxycycline-inducible shCD36 (using shCD36 #23 or #76)–derived from primary recipient mice fed with a normal or palm oil–rich diet. The secondary recipient mice were untreated or continuously doxycycline-treated. BLI signals are expressed as the normalized photon flux (*n* = 10 mice per group).

Extended Data Table 2. a, b, H3K4me3 ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for a) SCC-25 and b) VDH-15 pLKO.1 cells 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.1). **c, d,** Biological processes Gene Ontology (GO) terms built using the differential peaks and annotated genes shown in a) and b) respectively. **e,** H3K4me3 ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for VDH-15 pLKO.1 secondary (2ary) primary tumours (PTs) exposed to an *in vivo* palm oil-enriched diet as compared to 2ary control diet PTs (FDR<0.1). **f,** Biological processes GO terms build using the differential peaks and annotated genes shown in e).

Extended Data Table 3. a, b, Gene overlap analysis between genes harboring H3K4me1 (left) and H3K27me3 (right) peaks for both a) 4D and b) 14D Untreated/ post-PA SCC-25 pLKO.1 cells (Peak calling FDR<0.05; *a representation factor >/<1 indicates a higher/lower enrichment than expected by chance in gene overlap; P-Values were calculated using a hypergeometric test*). **c, d,** H3K4me1 and **e,** H3K27me3 ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for c), e) SCC-25 and d) VDH-15 pLKO.1 cells 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.1). **f, g,** PRO-Seq-derived genomic coordinates and quantitative information (*Log2 Fold Change and adjusted p-value and RPKM*

values) of the SCC-25 pLKO.1 eRNAs found statistically significant differentially expressed at f) 4D and g) 14D post-PA as compared to control cells. **h,** Non-filtered PRO-Seq-derived list of eRNAs identified for SCC-25 pLKO.1 cells as described in the Methods section.

Extended Data Table 4. a, Number of H3K9me3 ChIP-Seq. differential bound signal (DBS) regions (up-/down-regulated) detected for SCC-25 pLKO.1 4D/14D samples. **b, c,** Overlap analysis amongst genes harboring differentially up-regulated H3K4me3 and down-regulated H3K9me3 peaks in our b) 4D and c) 14D *in vitro* SCC-25 pLKO.1 ChIP-seq samples (*a representation factor >/<1 indicates a higher/lower enrichment than expected by chance in gene overlap; P-Values were calculated using a hypergeometric test*). **d,** Number of H3K27ac DBS regions (up-/down-regulated) detected for SCC-25 pLKO.1 4D/14D samples. **e,** Table displaying SCC-25 pLKO.1 cell cycle phase quantification for both 4/14D Untreated/ PA-treated cells, as shown in Extended Data Figure 5d. **f, g,** Overlap analysis amongst differentially expressed genes (DEGs), as detected by RNA-seq and genes harboring differential H3K4me3 peaks in our f) 4D and g) 14D *in vitro* SCC-25 pLKO.1 ChIP-seq samples. **h,** Top biological processes GO terms built using up-regulated DEGs in 14D post-PA RNA-seq SCC.25 pLKO.1 samples and **i,** Neural-related and **j,** Epigenetics-related-filtered GO terms. **k, l,** H3K9me3 ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for SCC-25 pLKO.1 cells at k) 4D and l) 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.05). **m, n,** Biological processes GO terms built using the differential peaks and annotated genes shown in k) and l). **o, p,** H3K27ac ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for SCC-25 pLKO.1 cells at o) 4D and p) 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.05). **q, r,** Biological processes GO terms built using the differential peaks and annotated genes shown in o) and p). **s, t,** scRNA-Seq. data displaying the s) DEGs identified for 4D PA for SCC-25 pLKO.1 cells as compared to its control condition and t) the gene markers defining the tSNE Cluster 0 at 4D time point,

Extended Data Table 5. a, b, Top Biological processes GO terms significantly up-regulated in a) 14D post-PA H3K4me3 ChIP-seq SCC-25 pLKO.1 cells and in b) VDH-15 pLKO.1 tumors from secondary (2ary) recipient mice exposed to a palm oil-enriched diet in the primary recipient, as compared to their control counterparts (Neural-related terms are highlighted in purple). **c, d, e,** Overlap analysis between genes harboring c) H3K4me3 differential up-regulated peaks for 14D post-PA *vs* 14D post-OA conditions and differential up-regulated d) H3K4me3 and H3K4me1 peaks for 14D post-PA conditions for both CTRL pLKO.1 SCC-25 and VDH-15 cells; and e) H3K4me3 and H3K27me3 peaks for 14D post-PA conditions for CTRL pLKO.1 SCC-25 (*a representation factor >/<1 indicates a higher/lower enrichment than expected by chance in gene overlap; P-Values were calculated using a hypergeometric test*; percentages refer to either the PA 14D condition or the H3K4me3 histone mark). **f, g, h,** Number of differential peaks obtained upon differential binding analysis of f), g) H3K4me3 14D Untreated/post-PA or 14D Untreated/post-OA conditions; and h) H3K4me1 14D Untreated/post-PA for both CTRL pLKO.1 SCC-25 and VDH-15 cells (Diff. Bind FDR≤0.1). **i, j,** Top biological processes GO terms significantly down-regulated in i) 14D post-PA H3K4me3 ChIP-seq SCC-25 pLKO.1 cells, and j) in VDH-15 pLKO.1 tumors from 2ary recipient mice exposed to a palm oil-enriched diet in the primary recipient, as compared to their control counterparts. **k, l, m, n,** Top biological processes GO tables displaying the top k), m) up-regulated and l), n) down-regulated terms for k), l) H3K4me3 VDH-15 pLKO.1 cells 14D post-PA or m), n) SCC-25 pLKO.1 cells 14D post-OA, as compared to their Untreated counterparts. **o, p,** H3K4me3 ChIP-Seq. differentially bound (up-/down-regulated) peaks and corresponding annotated genes for o) SCC-25 and p) VDH-15 pLKO.1 cells 14D post-OA treatment *in vitro* as compared to control cells (FDR<0.1). **q,** Biological processes GO terms built using the differential peaks and annotated genes shown in o). **r, s,** SCC-25 pLKO.1 RNA-Seq. Differentially expressed genes (DEGs) detected at r) 4D and s) 14D post-PA treatment when compared to control cells. **t,** Biological processes GO terms built using the DEGs and annotated genes shown in s).

Extended Data Table 6. Contains microarray data, gene ontology (GO) and gene set enrichment (GSEA), and integrative analysis from VDH-15, SCC-25 and melanoma primary tumours from secondary recipient ±palm oil memory.

Extended Data Table 7. Contains microarray data, gene ontology (GO) and gene set enrichment (GSEA) analysis from melanoma (501mel) primary tumours from secondary recipient ±palm oil memory.

Extended Data Table 8. a, BLI quantification of primary tumours (PTs) from secondary recipient mice injected with VDH-15-derived cells ±palm oil memory and treated with the neurotoxin 6-hydroxidopamine (6-OHDA), to induce the apoptosis of dopaminergic neurons, or vehicle respectively **b, c,** BLI quantification of lymph node (LN) (b) and lung (c) metastasis from animals in a. BLI signals are expressed as the normalized photon flux. **d,** Frequency of developed PTs from animals in a, b, c, (n = 10 mice per group; for 6-OHDA-treated 2ry mice, n = 8 mice per group).

Extended Data Table 9. Contains data of the *transcription factor binding site* (TFBS) analysis in the pro-neural signature of VDH-15 and SCC-25 primary tumours (PTs) with a palm memory and microarray data and gene set enrichment (GSEA) of SCC-25 shEGR2 and SCC-25 shGAL PTs ±palm oil memory.

Extended Data Table 10. a, b, Number of a) EGR2 and b) H3K4me3 differentially bound signal (DBS) regions (up-/down-regulated) detected for a) 4D PA SCC-25 pLKO.1 and CD36 KD and b) SCC-25 EGR2 KD samples 4D/14D post-PA treatment as compared to control ones. **c, d, e, f,** Overlap analysis amongst genes harboring c), e) up- and d), f) down-regulated H3K4me3 differential peaks in SCC-25 pLKO.1 and SCC-25 EGR2 KD cells both at c), d) 4D and e), f) 14D time points (*a representation factor >/<1 indicates a higher/lower enrichment than expected by chance in gene overlap; P-Values were calculated using a hypergeometric test*). **g, h,** Top

biological processes GO terms uniquely g) up- and uniquely h) down-regulated in g) 4D PA and h) 14D post-PA H3K4me3 ChIP-seq SCC-25 EGR2 KD samples as compared to pLKO.1 ones (Neural-related terms are highlighted in purple). **i,** EGR2 ChIP-seq differentially bound (up-/down-regulated) peaks (peaks called with narrow and broad peak calling settings displayed) and corresponding annotated genes for SCC-25 pLKO.1 cells at 4D PA treatment *in vitro* as compared to control cells (FDR<0.05). **j,** Biological processes GO terms built using the differential annotated genes shown in i). **k, l,** H3K4me3 ChIP-seq differentially bound (up-/down-regulated) peaks and corresponding annotated genes for SCC-25 EGR2 KD cells at k) 4D and l) 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.05). **m, n,** Biological processes GO terms built using the differential annotated genes shown in k) and l) respectively. **o,** EGR2 ChIP-seq differentially bound (up-/down-regulated) peaks and corresponding annotated genes for SCC-25 CD36 KD cells at 4D PA treatment *in vitro* as compared to control cells (FDR<0.05).

Extended Data Table 11. a, BLI quantification of primary tumours (PTs) from secondary recipient mice injected with VDH-15 cells derived from primary recipient mice fed with a palm oil–rich or a control diet. Secondary recipients were treated with the galanin receptor antagonist galantide (M15) or vehicle. BLI signals are expressed as the normalized photon flux. **b,** *Ex vivo* BLI lung metastasis quantification of mice from a. **c,** Frequency of developed PTs and lung metastases of secondary recipients in a, b,. (Vehicle-treated, n = 12 mice; M15-treated, n = 15 mice).

Extended Data Table 12. Contains data of the correlation of EGR2 or GAL expression and overall survival in different types of tumours. Numbers at risk for expression higher than median (red) and expression lower than the median (blue) are shown. *P* values, hazard ratios, mean times and sample size are shown for each type of tumour.

Extended Data Table 13. a, Number of differential MLL1/MLL2 ChIP-seq peaks identified upon differential binding analysis of SCC-25 pLKO.1 cells (FDR ≤ 0.1; FC = +/–1; up/down refer

to the 14D post-PA condition). **b,** Relative quantification values of MLL1 mRNA levels from VDH-15 cells infected with nt-shRNA (pLKO.1) or MLL1 shRNA #54 or #56. **c,** Lymph node metastasis (LNmet) bioluminescent intensity (BLI) quantification for all conditions assessed upon orthotopic injection of VDH-15 pLKO.1 and MLL1 #54 and #56 KD cells (*n = 10 mice/group; n.s., not significant; two-tailed t-test*). **d,** MLL1 and **e,** MLL2 ChIP-seq differentially bound (up-/down-regulated) peaks and corresponding annotated genes for VDH.15 pLKO.1 cells at 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.1).

Extended Data Table 14. a, Top biological processes GO terms significantly up-regulated in Set1A #42 KD H3K4me3 ChIP-seq samples 14D post-PA exposure when compared to Set1A KD 14D Untreated ones (Diff. Bind FDR≤0.05). **b,** Relative quantification values of Set1A mRNA levels for the nt-shRNA pLKO.1 and Set1A shRNAs #42 and #07. **c,** Bioluminescent intensity (BLI) and **d,** frequency of individual PTs developed in primary recipient animals orthotopically inoculated with control pLKO.1 or Set1A KD #07 VDH-15 cells after 14D post-PA *in vitro* exposure PA (*n= 20 mice/group).* **e,** Secondary (2ary) recipient orthotopic injections of control pLKO.1 and Set1A KD #42 VDH-15 cells upon *in vivo* exposure to Control/Palm-oil enriched diets in primary recipient mice. Frequency of developed PTs at the end-point of the experiment shown for all conditions. (n= 10 mice/group; PTs: *CTRL PLKO.1 CTRL Diet vs Set1A KD CTRL Diet *P= 0.0433* of a Fisher's exact test). **f,** H3K4me3 ChIP-seq differentially bound (up-/down-regulated) peaks and corresponding annotated genes for VDH-15 Set1A KD #42 cells 14D post-PA treatment *in vitro* as compared to control cells (FDR<0.05). **g,** Biological processes GO terms built using the differential annotated genes identified comparing control pLKO.1 and Set1A KD #42 VDH-15 conditions.

Extended Data Table 15. Contains microarray data, gene ontology (GO) and gene set enrichment analysis (GSEA) of bulk tumour stroma from VDH-15 and SCC-25 secondary recipients ±palm oil memory.

Extended Data Table 16. Contains data relative to 10X single cell (sc) analysis of bulk tumour stroma from pLKO.1 and shCD36 SCC-25 secondary recipients ±palm oil memory, overlap analysis showing the intersection between bulk-stroma palm memory-controlled genes and 10X sc-RNA-seq clusters from SCC-25–derived stromal cells in secondary recipients. Representation factor (RF) and *P* values are shown for the overlap. Hypergeometric test; estimated number of protein-encoding genes = 25,000 and palm diet–responsive genes in the tumour-associated stroma that are functionally related with specialized ECM constituent (*FC* > 1.5; *P* value < 0.05).

Extended Data Table 17. a, b, c, BLI quantification of primary tumours (PTs) (b) lymph node (LN) (c) and lung metastasis (d) from secondary recipient mice that had been injected with wild-type (WT) VDH15 or chondroitinase ABC (chABC)-expressing VDH-15 cells derived from primary recipients fed a normal diet or a palm oil–rich diet. BLI signals are expressed as the normalized photon flux. **d,** Frequency of developed PTs of animals (WT VDH-15, *n* = 12 mice per group; for chABC VDH-15, *n* = 10 mice per group).

Extended Data Table 18. Contains data from *ex vivo* BLI quantification lung metastasis in secondary recipient mice injected with VDH-15–derived cells from primary recipients ±palm oil diet, or with a mixture of VDH-15–derived cells and mouse tumour stroma cells from primary recipients ±palm oil diet; the combinations are indicated. The BLI signal is expressed as the relative normalized photon flux and the frequency of developed primary tumours (*n* = 10 mice per group).

Extended Data Table 19. Pearson correlation analysis showing that the CD36+ cells metastatic transcriptome significantly correlates with enhanced expression of perineuronal net genes in several tumor types. GEPIA: Gene Expression Profiling Interactive Analysis.

Methods

Animal studies. NOD *scid* gamma (NSG) (NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ) mice were purchased from Charles River and crossed in-house. All mice (of an approximate 50:50 mix of male:female) were housed under a regimen of 12 h light/12 h dark cycles and SPF conditions, and all procedures were evaluated and approved by the CEEA (Ethical Committee for Animal Experimentation) from the Government of Catalunya. SCC intra-tongue injection was performed as previously described[1,2]. Briefly, mice were anesthetized by intraperitoneal injection with a mixture of ketamin (50 mg per kg) and medetomidin (0.5 mg per kg), and SCC cells were resuspended in 30 µl PBS and injected into each mouse tongue with a BD ultra-fine 6-mm needle. For melanoma orthotopic cell transplant, mice were anesthetized, and a 0.8-cm incision was made on the dorsal back skin to implant a silicon chamber (SH, MEREFSA, S.L.), as previously described[3]. A mixture with 100,000 melanoma cells and 100,000 primary human keratinocytes were introduced into the silicon chamber with a p200 pipet. Mice were housed individually and anesthetized after 8 days, and the chamber was removed to bring the grafted area in direct contact with air. Mice were monitored for the luciferase bioluminescent signal immediately after injection (T0) and then once weekly with a Xenogen IVIS Imaging System-100 (Caliper Life Sciences). Briefly, animals were injected by retro-orbital injection with 50 µl of D-luciferin (Promega) diluted in 1× PBS at 5 mg ml^{-1}. Continuous administration of isofluorane gas was provided to ensure that animals were anesthetized during imaging. Data were quantified with the Living Image software version 4.4 (Caliper Life Sciences). Quantifications were calculated with unsaturated pixels. Colour scale minimum and maximum values are shown in pictures.

High-fat diet experiments entailed feeding mice for 10 days with 42% Kcal fat-modified Western diet supplemented with either palm oil (TD 150067, Envigo) or olive oil (TD 09820, Envigo). Normal chow diet was used for control groups.

For histological analysis, animal tissue was collected, fixed with 4% paraformaldehyde (PFA) for 2 h at room temperature, dehydrated and then embedded in paraffin.

To induce expression of inducible shCD36, doxycycline hydrochloride (D3447 SIGMA) was diluted at dose of 2 mg ml^{-1} in drinking water with 5% sucrose and continuously administered to

mice in secondary recipients.

To treat mice *in vivo* with neutralizing anti-CD36 antibodies, mice were injected intraperitoneally with 100 μl of physiological serum containing 10 μg of neutralizing monoclonal anti-CD36 JC63.1 (CAYMAN, CAY-10009893-500) or 10 μg of the corresponding IgA (mouse IgA, kappa [S107], Abcam, ab37322). All antibodies were azide-free with no added preservative compound. For sympathectomy experiments, secondary recipient NSG mice were injected intraperitoneally (i.p.), at day –6 and day –3 (e.g., days before the OSCC orthotopic injection); at day 3 after the OSCC injection, mice were injected again with a solution of 0.1% ascorbic acid and 0.9% saline, and either with 6-hydroxydopamine hydrochloride (6-OHDA, H4381 SIGMA) or without (as the vehicle), at doses of 250 mg kg^{-1} (day –6), 100m /kg^{-1} (day –3) or 100 mg kg^{-1} day 3)[4].

For blocking experiments with M15 (galantide/galanin receptor antagonist, G1278 SIGMA), secondary recipient NSG mice were injected intraperitoneally (i.p.) with either vehicle alone or M15 prepared in DMSO, injected at dose of 40 nmol kg^{-1} every 4 days after OSCC orthotopic injection[5], starting the day of the OSCC injection and finishing one day before the final point of the experiment.

Clinical material. Biological samples were obtained from patients of the Hospital Vall d'Hebron (Barcelona, Spain) with informed consent and approval of the Bank of Tumor Committees of the hospital according to Spanish ethical regulations. The study followed the guidelines of the Declaration of Helsinki, and patient identity and pathological specimens remained anonymous (in the study and elsewise).

Plasmids. The MSCV-IRES-luciferase-GFP retrovirus was kindly provided by Dr. Johannes Zuber (Research Institute of Molecular pathology, Vienna)[6]. CD36, EGR2, galanin, MLL1/2 and Set1A knockdown experiments used lentiviral shRNAs to target the selected gene (Sigma Aldrich and Dharmacon, respectively).

Inducible lentiviral shCD36 TRIPZ vectors were provided by HORIZON (https://horizondiscovery.com/en/products/gene-modulation/knockdown-reagents/shrna/PIFs/TRIPZ-Inducible-Lentiviral-shRNA?nodeid=entrezgene-948).

The empty pLKO.1-TRC lentiviral vector was used as a control (Addgene, plasmid #10879). (Supplementary Table S1). The LV-ChonABC lentivirus was kindly provided by Drs. Joost Verhaagen and Fred de Winter (Netherlands Institute for Neuroscience, Amsterdam)[7].

Cell culture studies. Cells were cultured in a humidified incubator at 37˚C with 5% $C0_2$. The human tongue squamous cell carcinoma cell line SCC-25 (ATCC^{R} CRL-168^{TM}) and patient-derived cells (VDH-15) were grown in keratinocyte serum-free media (KSFM; Gibco) supplemented with 5 µg ml^{-1} penicillin/streptomycin, 0.025 mg ml^{-1} bovine pituitary extract and 0.2 µg ml^{-1} hEGF. The 501mel human cell line (kindly provided by Dr. Claudia Wellbrock, Manchester Cancer Research Centre, The University of Manchester, UK) was grown in DMEM supplemented with 5 µg ml^{-1} penicillin-streptomycin and 10% FBS (Gibco). Primary human keratinocytes were grown as previously described[8]. PhoenixA and 293T cells grown in DMEM/10% FBS were used for retrovirus and lentivirus production, respectively, after transfection with the $CaCl_2$ method. For selection, 2.5 µg ml^{-1} puromycin or 0.3 mg ml^{-1} G418 was added to the media. All cell lines tested negative for mycoplasma contamination.

For palmitic acid, oleic acid and linoleic acid *in vitro* OSCC treatment, sodium palmitate (P9767, SIGMA), sodium oleate (O7501, SIGMA), sodium stearate (S3381, SIGMA) and linoleic acid sodium salt (L8134, SIGMA), respectively, were prepared as a 2.5 mM stock solution by dissolving in 1 ml solution of 0.1 M NaOH and warming at 80 ºC until clear. The fatty acid solution was complexed with fatty acid–free BSA (A7030, Sigma) in a molar ratio fatty acid/BSA of 5:1; briefly, 0.325 g BSA was dissolved in 8 ml of 0.9% NaCl, and the mixture was warmed to 45 ºC. The clear solution of palmitate, oleate, stearate and linoleate was added drop-by-drop by pipet with agitation. The final stock solution was filtered at 0.45 uM, aliquoted and stored at –20 ºC. SCC cells growing in serum–free media were treated *in vitro* with 300 uM or 50uM palmitate, 50 uM stearate, 50 uM oleate, 50 uM linoleic acid in KSFM media for 96 h (4 days).

For intra-tongue injection of SCC cells and backskin melanoma transplant, cultured cells were flask-detached with 0.25% trypsin-EDTA (Gibco), diluted in PBS/Trypan Blue, counted in a Neubauer chamber and resuspended in 1× PBS. For SCC-25 and VDH-15 cells, 100,000 cells per mouse were injected in the mice primary recipients. For SCC-25 and VDH-15 secondary recipient injection, 10,000 SCC DAPI-GFP+ purified sorted cells were injected per mouse. Cells were resuspended in cold KSFM/30% Matrigel (BD Matrigel, growth factor reduced, ref. 366231) before being injected into secondary recipients. For melanoma transplant, a mixture of cells containing 100,000 melanoma cells plus primary human keratinocytes (1:1) were introduced into the silicon chamber of primary mice recipients. For secondary recipients, 10,000 melanoma cells *plus* primary human keratinocytes (1:1) were transplanted.

For doxycycline *in vitro* treatment, SCC cells were incubated at 37˚C with 5% $C0_2$ with 2 µg ml^{-1} doxycycline hydrochloride (D3447 SIGMA), diluted in KSFM medium, and were collected 24 h to 48 h later for flow cytometry analysis.

Tumour disaggregation from xenografted mice. Tumours were isolated from mouse tongues or melanoma grafts, and connective tissue was removed to the greatest extent possible. Tissue was chopped in 0.5% trypsin 1-300 (MP Biomedical) in KSFM media (Gibco) in a McIlwain Tissue Chopper. After complete homogenization, samples were incubated at 37 ºC for 90 min with shaking. Homogenates were filtered sequentially in 100 µm, 70 µm and 40 µm BD strainers and then centrifuged at 1000 rpm for 10 min at 4 ºC. Supernatant was discarded, and each pellet was resuspended in 1× PBS/4% calcium chelated FBS[5] for antibody staining and subsequent FACS analysis.

FACS analysis. For xenograft analysis, cell samples were incubated for 45 min at room temperature with anti-human CD36-APC (clone 5-271, ref. 336208, BioLegend) and anti-human CD44-PeCy7 (clone GG-26, ref. 560533, BD Pharmigen) at 1:100 dilution; to exclude contaminant mouse cells, a lineage-negative cocktail conjugated to biotin composed of anti-CD31 clone MEC13.3 (ref. 553371, BD Pharmigen, 1:200 dilution), anti-CD45 clone 30-F11 (ref. 13-045-81, eBioscience, 1:200 dilution), anti-H2Kd (ref. 553564, eBioscience, 1:200 dilution) and

the mouse Lin Cocktail (ref. 120-003-582, MACS Miltinyi Biotech, 1:20 dilution) was used. After the first incubation, samples were washed in 1× PBS/calcium–chelated FBS[5] and centrifuged for the second incubation with Brilliant Violet (BV) 605 streptavidin (ref. 405229, Biolegend, 1:50 dilution) for 30 min at room temperature and then resuspended in 1× PBS/4% calcium–chelated FBS[5] with DAPI at a 1:1000 dilution for FACS analysis. For neural progenitor cells purification, samples were incubated with an APC mixture (*referred as Neural Mix*) of antibodies to enrich mouse neural and glial fractions (Supplementary Table 2), excluding endothelial cells (BV-605 anti-mouse CD31, clone 390 (RUO), ref. 740356, BD Biosciences, 1:200 dilution), macrophages (Pecy7 anti-mouse CD45, clone 30-F11, ref. 103113, BioLegend, 1:200 dilution) and tumoral-human GFP+ cells. After the first incubation, samples were washed in 1× PBS/calcium-chelated FBS[5] and centrifuged for the second incubation with APC streptavidin (ref. 405207, BioLegend, 1:50 dilution). Unstained, single-color and FMO (*Fluorescence Minus One*) controls were used in each case. Samples were analysed in a BD FACS ARIA FUSION. For FACS sorting, cells were selected on the basis of their Forward and Side scatter excluding cellular debris. Doublets and dead cells were eliminated by DAPI. GFP-positive cells were gated excluding lineage-BV–positive cells. This population was selected for further CD36 human expression analysis. Tumour-associated stromal cells were gated excluding endothelial (CD31+), macrophages (CD45+) and human (GFP+) cells. This population was selected for further mouse stromal gene expression analysis.

10× single-cell RNA sequencing (scRNA-seq).

Sample Processing. Xenografted tumours were processed and FACS-sorted as previously described (*see section Tumour disaggregation from xenografted mice*) to purify a tumour stroma-associated fraction enriched in glial and neural progenitors. Cells were collected in PBS + 0.5% bovine serum albumin (BSA) at 4 ºC in LoBind[R] tubes (Eppendorf) and processed immediately.

Sequencing. Cell concentration and viability of the single-cell suspension were verified by counting with a TC20™ Automated Cell Counter (Biorad). Cells were partitioned into Gel Bead-

In-Emulsions (GEMs) using the Chromium Controller system (10× Genomics) with the aim of a target cell recovery of 5,000 cells. Single-cell Gene Expression (GEX) libraries from PBMC samples were prepared using Chromium™ Single Cell 3′ Library & Gel Bead Kit v2 (10× Genomics, Cat. N. 120237), while CLL and cultured PBMC samples were prepared with Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3 (10× Genomics, Cat. N. 1000075), following manufacturer's instructions. Briefly after GEM-RT clean up, cDNA was amplified using 11 cycles. cDNA QC and quantification were performed on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). Libraries were indexed by PCR using the PN-220103 Chromiumi7 Sample Index Plate (10× Genomics). Size distribution and concentration of 3′ GEX libraries were verified on an Agilent Bioanalyzer High Sensitivity chip (Agilent Technologies). Sequencing was carried out on an Illumina Hiseq4000 sequencer to obtain approximately 40,000 reads/cell.

Primary processing. Sequencing reads were processed with CellRanger v4.0.0 data, using a mouse genome assembly as a reference genome (provided by 10× Genomics; https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest). The --chemistry and --expect-cells flags of *cellranger count* were set to “SC3Pv3” and “5000”, respectively.

Quality control and normalization. Quality control (QC) and normalization were performed independently for each dataset. The distributions of three QC metrics were assessed both individually and jointly: library size (total UMI), library complexity (number of detected genes) and mitochondrial expression. Note that analyzing these metrics jointly ensured that cells with high mitochondrial expression were not metabolically active. Cells barcodes with an aberrantly low number of UMI and genes, or with an abnormally high mitochondrial percentage, were classified as poor quality. Doublets were classified as barcodes with an aberrantly large library size and complexity, and by using the Scrublet doublet detection method [https://doi.org/10.1016/j.cels.2018.11.005]. Thresholds for the

aforementioned parameters were set individually for each dataset. Genes that were detected in less than 8 barcodes were rules out. The *Seurat v4.0.0* [https://doi.org/10.1101/2020.10.12.335331] was used to normalize and scale UMI counts, using the function *SCTransform* [https://doi.org/10.1186/s13059-019-1874-1].

Data integration. The four datasets were first merged without correcting for any technical artifacts, which revealed marked dataset-dependent clustering. These datasets were then integrated using *Harmony* [https://doi.org/10.1038/s41592-019-0619-0]. The raw count matrices were integrated using the top 30 principal components obtained from PCA of the merged dataset. Post-integration clustering and non-linear dimensionality reduction were run using the functions *FindNeighbors*, *FindClusters* and *RunUMAP* on the top 30 components obtained by Harmony with default parameters. The resolution parameter in FindCluster set to 0.01, 0.05, 0.1, 0.25, 0.5, 0.75 or 1, to assess varying degrees of granularity.

Downstream analysis. To annotate the cell identity, cells were clustered with the *FindClusters* function. The resolution parameter was set to 0.05 for a high-level annotation. Cluster 5 was then further split into 2 subclusters based on the cluster labels obtained at resolution 0.5. Marker genes were computed for each cell identity, and a Wilcoxon signed-rank test was performed to test for differential expression of each gene; note that this test is among the best performing for scRNA-seq differential-expression analysis[21]. Finally, a combination of differentially expressed and well-known marker genes was used to annotate each cluster to its specific cell type. Lastly, the differential abundance analysis between the control and palm diet in the different conditions was carried out using a binomial logistic regression. The control diet group was set as the reference and obtained the odds ratio for each type.

Immunofluorescence and histological analysis. De-paraffinized antigen retrieved sections (10 min in boiling 10 mM Trizma base, 1 mM EDTA, pH 9.0, or autoclaved in 10 mM Tri-sodium citrate, pH 6.0) of 8 μm were permeabilized for 25 min in 0.25% Triton X-100/PBS and blocked for 90 min in 0.25% gelatin/PBS. Primary antibodies were incubated overnight at 4 °C, and

secondary antibodies were incubated 2 h at room temperature in 0.25% gelatin/PBS. Nuclei were stained with DAPI (1:5000, Roche). The primary antibodies used were keratin-14 polyclonal chicken (ref. 906001, BioLegend) at 1:500, S100 polyclonal rabbit (ref. ab868, Abcam) at 1:200, S100 mouse (4C4.9) (ref. ab4066, abcam) at 1:200, neurofilament light (NEF-L) polyclonal rabbit (ref. AB9568, Chemicon) at 1:200 and Gap43 polyclonal rabbit (ref. ab16053, Abcam) at 1:200, Has1 polyclonal rabbit (ref. bs-2946R Bioss Inc) at 1:100, Versican (Vcan) polyclonal rabbit (ref. ABIN2443952 Antibodies-Online) at 1:100, Tenascin R (TnR) polyclonal rabbit (ref. PA5-31143 THERMO). Tyrosine Hydroxylase (TH) polyclonal rabbit (ref. ab112 ABCAM). Secondary antibodies were Alexa Fluor conjugated (R37118, A-11041, Molecular Probes). Hematoxilin and eosin (H&E) staining was done according to the standard protocol. Images were acquired using a Nikon E600+Olympus DP72 and a Super resolution Elyra PS1+AiryScan confocal microscope. Image processing and quantification analysis was performed with ImageJ software (https://imagej.nih.gov/ij/index.html).

Spatial Transcriptomics.

Visium samples Origin. Xenografted tumours were dissected from mice, embedded in Tissue-Tek O.C.T. compound (Ref. 4583 Sakura) within biopsy 10×10×5 mm cryomolds (Ref 62534-10 Sakura) and quickly snap-frozen in an isopentane bath chilled with liquid nitrogen (LN2).

Samples were stored at –80 ºC until further processing.

Sequencing. Frozen tissue samples were tested for RNA quality with RIN > 7.0 (RNA pico, Agilent), and a tissue optimization experiment (10× Genomics, Visium Spatial Tissue Optimization, Rev A) was performed with imaging of fluorescence footprint on upright epifluorescence (Nikon Eclipse E800), which identified 10 min as an optimum permeabilization time.

Samples were then processed for full ST experiment as per manufacturer's instructions (10× Genomics, Visium Spatial, Rev D), cut in a pre-cooled cryostat at 10 µM thickness onto four 6.5 mm × 6.5 mm capture areas with 5000 oligo-barcoded spots. Slides then were fixed and

stained by H&E with immediate imaging on NanoZoomer S60 (Hamamatsu). Tissues were permeabilized with a proprietary enzyme (10 min), and reverse-transcription and second-strand synthesis were performed on the slide, with cDNA quantification via qRT-PCR using KAPA SYBR FAST-qPCR kit (KAPA Biosystems), and analyzed on the LightCycler®480 Real-Time PCR System (Roche). qRT-PCR results (Cq value at 25% of peak fluorescence) revealed cDNA amplification. The Visium slide serial number was V19S23-040.

Libraries were constructed as per manufacturer's instructions; ST libraries were then quantified by High-Sensitivity DNA Assay 2100 Agilent. Libraries were sequenced on a NovaSeq6000 platform (Illumina, US) using 100 bp paired-end dual-indexed set up. Around 200 million paired-end reads were obtained for each tissue section. Read 1 was sequenced with 28 cycles; i7 index, 10 cycles; i5 index, 10 cycles; and read 2, 82 cycles.

Primary processing. Reads were mapped against a mixed human (GRCh38) - mouse reference dataset downloaded from 10× Genomics (https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest) using 10× Genomics Space Ranger 1.1.0 with default parameters. Images were processed with Inkscape to divide into individual capture areas and rotated them to align the printed fiducial spot pattern.

Quality control and normalization. Quality control was carried out by looking at the number of UMIs and genes and mitochondrial and ribosomal percentage. For sample shCD36/palm diet, spots that did not overlap with the tissue sections were removed. Expression matrix counts were normalized using Seurat's NormalizeData, which uses log1p to natural-log transform the data and then multiplies it by a scale factor of 10,000. A subset of the 2,000 most highly variable genes was identified using the *FindVariableFeatures* function with default parameters; this subset was further scaled to mean 0 and variance 1 using the function *ScaleData*.

Downstream analysis. Principal component analysis (PCA) was used to reduce the dimensionality using *RunPCA*, clustering and non-linear dimensionality reduction were run using the functions *FindNeighbors*, *FindClusters*, and *RunUMAP* respectively on the top 10 PC. The number of principal components was determined by looking at the elbow plot of the variance % explained by each principal component.

The percentage of mouse vs human for each spot was set by determining the fraction of UMIs mapped to mouse genes over human genes at each capture location. Tissue region was stratified by clusters overlapping pure mouse, pure human and intermixing regions as being healthy, tumour and tumour-front, respectively.

To map the single-cell reference dataset, some granularity was added to the basal cells by dividing it into 5 subclusters and removing erythrocytes. *SPOTlight* (https://doi.org/10.1093/nar/gkab043) was used to map the cell types found in the reference dataset to the spatial transcriptomics data. To train the model, up to 100 cells per cell type were used, with the geneset being the union between the marker genes of the cell types and the 3000 most-highly variable genes. Marker genes for each cell type were obtained with the Seurat's function *FindAllMarkers*, whereby only markers with positive log2FC were considered. Cell types contributing <4% to the spot's predicted composition were considered to be fitting noise and were set to 0. To quantify the changes in tumour-associated Schwann cells across the healthy, tumour-front and tumour, a pairwise Wilcoxon test was used between the regions; the Kruskal-Wallis test was used to compare the three distributions simultaneously.

***Data and code Availability*.** R version 4.0.1 and Python 3.6.0 were used. All the code used to analyze the single cell and positional RNAseq data can be found in the GitHub repository https://github.com/MarcElosua/10X-EPID. Data can be found in [TBD which data is available and where].

RNA sequencing from *in vitro* cultured OSCC cells.

Sample Processing. *In vitro* SCC-25 untreated/PA-treated cells at 4 days or 14 days post-PA were collected for FACS-sorting. 500,000 CD44+ cells were collected per condition and 10ug of total RNA was purified using the RNeasy Mini Kit (Cat. 74104 Qiagen) and further processed for mRNA-seq, applying the Illumina sequencing technology (https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html).

Downstream Analysis. After adapter-cleaning and quality correction using Trimmomatic (v0.36), paired-end reads were aligned to the hg19 genome (Ensembl GRCh37.75) using the TopHat Aligner (v2.1.1).

For differential expression analysis, differentially expressed genes (DEGs) were determined according to read count values, as performed using featureCounts (v1.6.0). DEGs were obtained using the DESeq2 software (v1.26.0) and using an adjusted p-value < 0.05 and fold change (FC) >1.5.

Chromatin immunoprecipitation-sequencing (ChIP-seq) assays. For the 4-day chromatin time point, OSCC cells were either stimulated with palmitic acid (300 μM) or oleic acid (50 μM) for 4 days *in vitro*, and chromatin was collected on the 4th day. For the 14-day time point, the corresponding fatty acid was withdrawn from the media after the 4-day stimulation, and cells were passaged normally for 2 weeks until collection time for fixation. All ChIP-seq experiments targeting histone marks were performed using the ChIP-IT High Sensitivity kit® from Active Motif (ref. 53040) according to the manufacturer's instructions. Sonication was performed for 30 min in a Bioruptor Pico (Diagenode). The optimal conditions for each of the histone marks and antibodies used are listed below:

Target Histone Mark	Chromatin Amount/IP	Antibody Used
H3K4me1	**10 µg**	**ref. ab8895, Abcam**
H3K4me3	**10 µg**	**ref. C15410003, Diagenode**
H3K27me3	**15 µg**	**ref. ab6002, Abcam**
H3K27ac	**15 ug**	ref. **07-360**, Millipore
H3K9me3	**20 ug**	ref. **ab8898**, Abcam

Optimal chromatin fixation for MLL1 and MLL2 ChIPs consisted of double crosslinking using ChIP-IT High Sensitivity kit® fixation reagents and the ChIP Cross-link Gold reagent from Diagenode (ref. C01019027). Samples were sonicated for 15 min in a Bioruptor Pico

(Diagenode). About 60 µg chromatin was used for both MLL1 and MLL2 IPs following the ChIP-IT High Sensitivity kit® instructions; MLL1 and MLL2 antibodies were produced in-house in the A. Shilatifard laboratory (Northwestern Medicine).

EGR2 ChIPs were performed following the ChIP-IT High Sensitivity kit® recommendations. Chromatin was sonicated for 15 min in a Bioruptor Pico (Diagenode), and 45 ug of chromatin and 8 ug of EGR2 antibody (ref. Ab43020) were used in the immunoprecipitation step.

H3K4me3 ChIPs-seq from sorted human tumour cells was performed using the True MicroChIP kit from Diagenode (ref. C01010130). In brief, 30,000 cells per condition were FACS-sorted, fixed and sonicated together. Optimal sonication was achieved after 20 min in a Bioruptor Pico (Diagenode). Samples were then split in three, and each IP was performed from 10,000 cells using the ChIP-seq–grade H3K4me3 positive control antibody provided. All libraries for sequencing were prepared using NEBNext Ultra DNA Library Prep Kit from Illumina (E7370L) following the manufacturer's instructions.

ChIP-seq data processing and differential binding analysis. Trimmomatic (v0.36) software[9] was used to remove possible contaminations from single-end ChIP-seq data. Reads were aligned to the human genome build hg19 using BWA (v0.7.12) software[10] and BowTie2 (v2.3.3.1) software for H3K9me3, H3K4me3 (EGR2 KD cells) and EGR2 ChIP-seq datasets. Aligned bam files were coordinate-sorted, and duplicates were removed with samtools (v1.3.1) software[11]. Quality control was assessed on every dataset before running any other bioinformatics analysis with FastQC (v0.11.3) software (https://www.bioinformatics.babraham.ac.uk/). Samples were subsampled to the lowest coverage for each dataset.

Peak calling to determine enrichment over the background was done with MACS (v2.1.2) callpeak[12]. The broad flag method was used for H3K9me3, H3K4me1, H3K4me3 and Set1A with q-value narrow and broad cut-off 0.05, and for H3K27me3, MLL1 and MLL2, with q-value narrow cut-off 0.05 and q-value broad cut-off 0.0001. Signal tracks on ChIP-seq peak enrichment

levels over whole-genome were generated with MACS (v2.1.2) bdgcmp on the bedGraph files generated by MACS (v2.1.2) callpeak. For H3K27ac, peaks were called using GeneRich (version 0.6) and a false discovery rate (FDR) ≤ 0.05. GeneRich was also used to remove duplicate reads. Differential peaks were determined using the DiffBind package (version 2.4.8) in R (version 3.54.24) and an FDR ≤ 0.05.

The distribution of the differential regions within genomic features was determined using annotatePeaks.pl script of HOMER suit of tools (v4.10) [13].

R software (v3.4.4)[14] and Bioconductor R package DiffBind[15] were used for differential binding analysis of H3K4me1, H3K4me3, H3K27me3, MLL1 and MLL2. The binding matrix was calculated with affinity scores based on TMM normalization (EdgeR), using read counts minus control read counts and effective library size. PCA plots were generated using affinity data for all sites. Differential analysis using DESeq2 was used for H3K27ac, H3K4me3 (SCC-25 EGR2 KD cells) and EGR2 ChIP-seq datasets. The FDR threshold was fixed as either 0.1 or 0.05 (as specified in each figure legend).

Heatmaps and average profiles of ChIP binding to TSS regions were plotted with ChIPseeker R Bioconductor package [16].

The Enrichr (https://amp.pharm.mssm.edu/Enrichr/)[17] online platform was used for ChIP-seq gene ontology (GO) analysis. Integrative Genome Visualizer IGV (v2.4.14) (http://software.broadinstitute.org/software/igv/)[18] was used for visualization of peaks.

ChIP-seq data and RNA-seq data combined analysis. To assess transcription in the genomic loci bearing PA-driven H3K4me3 changes at 4D/14D time points, the already-defined set of H3K4me3 peaks for each condition was used (as detailed in the ChIP-seq data processing section; see below). The H3K4me3 peaks of all replicates were merged for each condition using BEDTools Intersect (v2.28.0) fuction. The BEDTools Window (v2.28.0) function was then applied to overlap the TSS of all expressed genes in each experimental condition, as determined by the RNA-seq data (see RNA-seq and data analysis section), with all H3K4me3 peaks in that condition in a window of ± 300 bp. Regions containing differentially expressed genes, as defined

by RNA-seq using a fold-change (FC) cutoff > 2 and *P* value < 0.05, were selected. The RNA-seq TMM of all the differentially expressed genes in each time point (4 days or 4+14 days) was used as the input for the corresponding box plots.

Precision run-on sequencing (PRO-seq) studies. The precision nuclear run-on sequencing (PRO-seq) assays were performed at Ali Shilatifard's laboratory following a previously detailed protocol[19]. In brief, nuclei from OSCC cells were isolated using a Dounce homogenizer with a loose pestle, and the nuclear run-on reaction was performed at 30 ºC for 3 min using 25 μM biotin-11-ATP/GTP/CTP/UTP (Perkin Elmer). Total RNAs were extracted and hydrolyzed using a 0.2 M NaOH solution for 10 min on ice. Biotinylated nascent RNAs were purified using Dynabeads M-280 streptavidin (Invitrogen), followed by 3′-VRA3 and 5′-VRA5 adapter ligation and two additional rounds of purification with Dynabeads. cDNA was then generated by reverse transcription with the SuperScript III (Invitrogen), and a 10-cycle amplification step of the cDNA libraries was performed. The indexed DNA libraries were size selected (140–350 bp) and sequenced in NextSeq 500 system (Illumina) with single-read runs. Two biological replicates per experimental condition were sequenced for the SCC-25 cell line.

PRO-seq data processing. For eRNA mapping, the BEDTools Window (v2.28.0) function was used to overlap H3K27ac and H3K4me1 peaks for each time point (4 days [4D] or 4+14 days [14D]) and each condition (PA or untreated) in a window of ±500 bp, following the merging of all common peaks using the BEDTools merge (v2.28.0) function. Homer annotatePeaks v4.9.1-5 was used for peak annotation and only intergenic peaks were selected, as described[13]. The H3K27ac and H3K4me1 intergenic peaks were overlapped with BEDTools Window (v2.28.0) function in a window of ± 250 bp, and a set of intergenic enhancers was defined. After identifying intergenic regions containing H3K27ac/H3K4me1 peaks, the BEDtools Intersect function was used with the option –v to discard peaks overlapping with any gene from the hg19 reference genome (ENSENBL v87), including an additional threshold of 2 kb upstream of every TSS. In order to determine which intergenic enhancers presented transcriptional activity, we separated the

PRO-seq mapped reads were separated by strands and individually-called peaks using MACS2.1.2 with the parameters –nomodel --extsize 200 --broad-cutoff 0.1 and –broad to identify broad peaks. The common peaks were merged from each sample (e.g., 4D or 14D) for each stranded peak; all peaks from common plus strand were then merged with the common minus-strand peaks. BEDtool intersects between the intergenic enhancer peaks and the common peaks from PRO-seq were used to define enhancers that were actively transcribing eRNAs. All transcripts shorter than 800 nt were excluded to minimize any bias in the Pol II Traveling Ratio (TR) analysis (as these eRNAs are short), to precisely separate promoter and gene body regions. Lastly, the TSS coordinates were defined for each eRNA based on the position of the highest signal occurrence in the PRO-seq peak window, and those with a count per million read value (CPM) > 0.5 were discarded. For eRNA TR calculations, the promoter region was defined as –100 nt from PRO-Seq summit to + 150 nt, and the gene body, as +151 nt from PRO-Seq summit to TES plus 1 Kb, for all transcripts used in the analysis (n= 2,182). The BEDTools closest (v2.28.0) function was used to identify the closest expressed upstream and downstream genes.

Single cell RNA-sequencing (scRNA-seq) assays – SMART-seq. Full-length single-cell RNA-seq libraries were prepared using the Smart-seq2 protocol[20] with minor modifications. Briefly, freshly harvested single cells were sorted into 96-well plates containing the lysis buffer (0.2% Triton-100, 1U/ul RNase inhibitor). Reverse transcription was performed using SuperScript II (ThermoFisher Scientific) in the presence of 1 μM oligo-dT30VN (IDT), 1 μM template-switching oligonucleotides (QIAGEN), and 1 M betaine. cDNA was amplified using the KAPA Hifi Hotstart ReadyMix (Kapa Biosystems) and IS PCR primer, with 23 cycles of amplification. Following purification with Agencourt Ampure XP beads (Beckmann Coulter), product size distribution and quantity were assessed on a Bioanalyzer using a High Sensitivity DNA Kit (Agilent Technologies). A total of 120 pg of the amplified cDNA was fragmented using Nextera XT (Illumina) and amplified with double indexed Nextera PCR primers (IDT). Products of each well of the 96-well plate were pooled and purified twice with Agencourt Ampure XP beads. Final libraries were quantified and checked for fragment size distribution using a Bioanalyzer High

Sensitivity DNA Kit. Pooled sequencing of Nextera libraries was carried out using a HiSeq2500 (Illumina) to an average sequencing depth of 500,000 million reads per cell. Sequencing was carried out as paired-end (PE75) reads with library indexes corresponding to cell barcodes.

scRNA-seq data analysis – SMART-seq. Single-cell RNA-seq data analysis was performed using R package Seurat (v2.3.0)[21]. Gene expression count matrices were filtered both to exclude cells with mitochondrial percentage higher than 15, percentage of mapped reads lower than 85 and small library size (corresponding to the first quartile from total sum of counts across all features) and to exclude features that were not expressed in at least 5 cells and had minimum 90 counts. Gene expression levels for each cell were normalized by total expression, multiplied by a scale factor of 10,000 and log-transformed. Before clustering, highly variable genes, which represent 10 % of total gene features, were identified based on their average expression and dispersion and data was scaled regressing out batch effects. The number of significant principal components was assigned by the randomization approach "jack straw", and clustering was performed on the top 10 principal components with a resolution of 0.5 by a shared nearest neighbor (SNN) modularity optimization-based clustering algorithm. T-distributed stochastic neighbor embedding (t-SNE) was applied to display the data to cell loadings of 10 selected principal components. Cells were scored by palmitic acid responder genes and projected on the t-SNE. Gene expression markers were found for all the clusters by Wilcoxon rank-sum test with a return threshold of 0.01. To specifically find genes that were different between each time point and condition, Wilcoxon rank-sum test was applied between untreated cells at 4 days and palmitic acid treated cells at 4 days; only genes that were detected in a minimum fraction of 0.1 in either of the two populations were tested.

Trajectory analysis was performed using Monocle (v2.10.1)[22] with the highly variable genes obtained in previous steps for each time point separately. DDTree method was used for dimensionality reduction and batch effects were subtracted from the data to avoid their contribution to the trajectory. Finally, clusters and different treatment conditions were projected on the inferred trajectory.

Western Blotting.

After a wash with cold PBS, cells were lysed for 30 min at 4 ºC using buffer containing 1× EDTA, 150 mM NaCl, 20 mM HEPES, supplemented with protease and phosphatase inhibitors.

All samples were pipetted several times to ensure proper lysis. Cell lysates were centrifuged at 4 ºC for 15 min at 12,000 rpm, and the supernatant was kept as the protein extract. All samples were diluted in Laemmli buffer and boiled at 95ºC for 5 min. Equal amounts of protein for all samples were loaded into 6% polyacrylamide gels for protein separation, followed by transferring to LF-PVDF membranes, 90 min at 4ºC. Membrane blocking incubation was done with 5% milk-supplemented TBS for 30 min at room temperature (RT). In-house MLL1, MLL2, Set1A or Set1B or loading control Hsp90 (ref. Ab13495, Abcam) antibodies were incubated at 4ºC over-night on a rocking platform. Next, membranes were incubated with a secondary anti-rabbit HRP conjugated antibody at 1:5,000 dilution for 1 h at RT and washed. Blots were visualized using SuperSignal™ West Pico Chemiluminescent Substrate (ThermoFisher) in a ChemiDoc Gel Digital Imaging System (Bio Rad).

For chABC western blot, 5 µg total protein lysate from *in vitro*-VDH-15 infected with LV-chABC was loaded into 12% polyacrylamide gel for protein separation and then transferred to LF-PVDF membrane, for 90 min at 4ºC. Membrane blocking incubation was done with 5% milk-supplemented TBS for 30 min at RT followed by incubation with anti-chABC (1E10) (NBP 96141AF594 Novus Biologicals). Membranes were then incubated with a secondary anti-rabbit HRP conjugated antibody at 1:5,000 dilution for 1 h at RT and washed. Blots were visualized using SuperSignal™ West Pico Chemiluminescent Substrate (ThermoFisher) in a ChemiDoc Gel Digital Imaging System (Bio Rad).). Western blot analysis was performed according to standard protocols.

Gene expression analysis. RNA isolation and cDNA amplification was performed as previously described[19]. In brief, RNA was isolated using magnetic beads (RNAClean XP beads, Agencourt). RNA was reverse transcribed and amplified using Whole Transcriptome Amplification chemistry

(WTA2, Sigma Aldrich). For monitoring amplification, SYBR Green was added to the reaction; it was stopped when the SYBR Green signal reached a plateau. cDNA were purified using PureLink Quick PCR Purification Kit (Invitrogen) and quantified using Nanodrop ND-1000 spectrophotometer (Thermo-FisherScientific).

Subsequently, 8 µg cDNA was fragmented and labelled using GeneChip Mapping 250K Nsp Assay Kit (Affymetrix; catalog #900766) according to manufacturer's instructions. Mouse MG-430 PM Array were hybridized, washed, stained and scanned according to the protocol described in GeneAtlas™ Hybridization, Wash, and Stain Kit for 3′ IVT Arrays User Manual.

Samples ready to hybridize were denatured at 96 °C for 10 min prior to incubation on Mouse MG-430 PM Array Strip. Hybridization was performed for 16 h at 45 °C at GeneAtlas Hybridization Oven (Affymetrix). Washing and stain processing were performed at GeneAtlas Fluidics Station (Affymetrix), following the specific script for Mouse MG-430 PM Arrays. Subsequently arrays were scanned with GeneAtlas Scanner (Affymetrix). Normalized expression signals were calculated from Affymetrix CEL files using RMA algorithm[20].

Microarrays data are available in the Gene Expression Omnibus (GEO) repository under accession code GSE148321.

Gene overlap analysis. For all gene overlap analyses (performed on expression and/or ChIP-seq data, a hypergeometric test was used, considering a background of 25,000 genes in total. Hypergeometric tests yielded a representation factor and *P* value for each comparison, indicative of the gene overlap enrichment and randomness respectively.

Gene expression analysis. RNA isolation and cDNA amplification were performed as previously described[23]. In brief, RNA was isolated using magnetic beads (RNAClean XP beads, Agencourt). RNA was reverse-transcribed and amplified using Whole Transcriptome Amplification chemistry (WTA2, Sigma Aldrich). For monitoring amplification, SYBR Green was added to the reaction, and reactions were stopped when the SYBR Green signal reached a plateau. cDNA were purified

using PureLink Quick PCR Purification Kit (Invitrogen) and quantified using Nanodrop ND-1000 spectrophotometer(Thermo-FisherScientific).

Subsequently, 8 µg of cDNA were fragmented and labelled using GeneChip Mapping 250K Nsp Assay Kit (Affymetrix; catalog #900766) according to manufacturer's instructions. Mouse MG-430 PM arrays were hybridized, washed, stained and scanned according to the protocol described in GeneAtlas™ Hybridization, Wash, and Stain Kit for 3' IVT Arrays User Manual. Samples ready to hybridize were denatured at 96°C for 10 min prior to incubation on Mouse MG-430 PM Array Strip. Hybridization was performed for 16 h at 45 °C in a GeneAtlas Hybridization Oven (Affymetrix). Washing and stain processing were performed at GeneAtlas Fluidics Station (Affymetrix), following the specific script for Mouse MG-430 PM Arrays. Subsequently arrays were scanned with GeneAtlas Scanner (Affymetrix). Normalized expression signals were calculated from Affymetrix CEL files using RMA algorithm[24].

Real-time PCR. Real-time PCR using TaqMan gene expression probes (Applied Biosystems) (Supplementary Table 3) was performed and analysed using a 7900-HT Fast Real-Time PCR Instrument (Applied Biosystems). Relative expression levels were determined by normalization to beta-2-microglobulin (B2M) and/or beta-actin (ACTB) using the $\Delta\Delta C_t$ method.

Bioinformatics analysis. ToppGene Suite (http://toppgene.cchmc.org)[25], Gene Set Enrichment Analysis (GSEA) (www.broadinstitute.org/gsea)[26,27] and the Specific Expression Analysis tool (genetics.wustl.edu/jdlab)[28,29], were used for gene ontology and gene set enrichment analysis. Cytoskape software platform (http://cytoskape.org)[30] was used to visualize and analyse the functional networks. PSCAN platform (http://www.beaconlab.it/pscan)[31] was used for transcription factor motif analysis in promoter of co-regulated genes. The analysis of survival biomarkers was performed using the gene dataset libraries from the Kaplan Meier Plotter website (https://kmplot.com/analysis)[32].

Gene expression analysis in the microarray data was performed using R[33] and Bioconductor[34]. Microarrays were processed using RMA normalization as implemented in the Bioconductor R

package affy[35]. Data were annotation using probeset information provided by Affymetrix in its product support web (http://www.affymetrix.com/support/ downloaded 10/05/2017). Standard quality controls were performed in order to identify abnormal samples[36] regarding: a) spatial artifacts in the hybridization process (scan images and pseudo-images from probe level models); b) intensity dependences of differences between chips (MvA plots); c) RNA quality (RNA digest plot); d) global intensity levels (boxplot of perfect match log-intensity distributions before and after normalization and RLE plots); e) anomalous intensity profile compared to the rest of the samples (NUSE plots, Principal Component Analyses); f) impact of quality metrics[37] on expression measures. For the CD36-bright / CD36-dim and the neural stromal fraction datasets, expression values from samples sharing the same hybridization batch were a-priori corrected genewise by scanning batch using a linear model, in which sample group was included as covariate. Additionally, and in these datasets, hybridization batch was also a-priori corrected using the same approach. Expression for each specimen was summarized gene-wise using the mean across its technical replicates.

Differential expression analysis was carried out at the probeset level using moderated t-statistics by empirical Bayes shrinkage as implemented in the limma R package[38], including hybridization batch (CD36-bright / CD36-dim and neural stromal fraction datasets) or scanning batch (shEGR2 / shGAL datasets) as covariate in the model. In the shEGR2 and shGAL human dataset, an additional control by RMA.IQR metric[37] was required and thus included in the model. The Benjamini and Hochberg correction method[39] was applied for multiple contrast adjustment.

Functional enrichment analyses in the microarray data were performed using a modification of ROAST[40], a rotation-based approach implemented in the R package *limma*[38] that is especially suitable for small size experiments. Such modifications were implemented to accommodate the re-standardized maxmean statistic proposed in *Bradley Efron et al.*[41], in the ROAST algorithm, in order to enable it for competitive testing[42]. When no biological replicates were available, functional enrichment was carried out with the pre-ranked version of the Geneset Enrichment Analysis (GSEA)[23], for which genes were ranked according to the base 2 logarithm of the fold-changes obtained in the differential expression analysis. In both cases, data were collapsed to the

gene level by selecting the probeset showing the highest standard deviation within each gene. Genesets for these analyses were derived from the Gene Ontology (GO)[43], as collected in the R packages org.Hs.eg.db[44] and org.Mm.eg.db[45].

Statistical analysis. For all experiments, an adequate sample size was determined based on results of pilot studies. No statistical methods were used to determine sample size. All animals that fulfilled proper experimental conditions during the experimental procedures were included in the analysis. Based on the results of pilot studies, homogeneous groups of males and females between 8 and 10 weeks, and their control littermates, were used for the experimental studies. No statistically differences were found with respect to the sex of mice, and no randomization method previously described was used. Data are generally shown as the mean ± s.e.m. Statistical significance was analysed using Prism 6 software (GraphPad) with a two-tailed *t*-test and Fisher's exact test. Significance was considered at $P \leq 0.05$.

Supplementary Table 1

Target Gene	5′-3′ target DNA sequence	Library Clone ID	User ID
MLL1	GCACTGTTAAACATTCCACTT	TRCN0000005954	#54
	CGCCTAAAGCAGCTCTCATTT	TRCN0000005956	#56
Set1a	CAAAGGACAACAACGAATGAA	TRCN0000152242	#42
	GCTGAAACTCAACCAGCTCAA	TRCN0000154707	#07
CD36	GAAGTTACATATTAGGCCAT	TRCN0000056998	-
	CCGACGTTAATCTGAAAGGA	TRCN0000056999	-
GAL	CCTGAAGTCAAACCTTAAGAT	TRCN0000083173	-
	CAGGTCATTCAGCGACAAGAA	TRCN0000083174	-
EGR2	CGTAGCAATATCTGCTCCTTT	TRCN0000013838	-
	CCCAGACTATCCTGGATTCTT	TRCN0000013839	-
	CTCTCTACAATCCGTAACTTT	TRCN0000013840	-
CD36-i	TAAGCTAATATTATTGAAG	V2THS_260223	#23
	TGATTTTGATAGATATGGG	V3THS_343076	#76

Supplementary Table 2

Antibody	Working dilution	Cat. reference	Reactivity

APC Rat CD184; Clone 2B11/CXCR4 (RUO)	1:100	558644, BD Biosciences	Neuroepithelial/Glioblast
CD271 (LNGFR)-Biotin	1:100	130-110-076, Miltenyi Biotec	Neuroepithelial/Neuroblast
APC CD15 (SSEA-1)	1:100	FAB2155A-100, RD Systems	Neuroepithelial
DyLight 650 CD133	1:100	NB120-16518C, Novus Biologicals	Neuroepithelial/Neuroblast
APC CD29 Clone HMβ1-1	1:100	102215, BioLegend	Neuroepithelial/Neuroblast
APC- CD171 (L1CAM)	1:100	130-102-221, Miltenyi Biotec	Intermediate neural progenitors
APC CD24, Clone M1/69 (RUO)	1:100	562349, BD Biosciences	Intermediate neural progenitors
APC GLAST (ACSA-1, astrocyte *cell* surface antigen)	1:100	130-095-814, Miltenyi Biotec	Glial
APC A2B5	1:100	130-093-582, 130-093-582	Glial
APC NCAM-1/CD56	1:100	FAB7820A, RD Systems	Glial/Neural

Supplementary Table 3

GENE	Assay ID
B2M	Hs00187842_m1
HPRT1	Hs01003267_m1
ACTB	Hs01060665_g1
GAL	Hs00544355_m1
EGR2	Hs00166165_m1
CD36	Hs01567185_m1
MLL1	Hs00610538_m1
Set1a	Hs00322315_m1

Supplementary References

1 Myers, J. N., Holsinger, F. C., Jasser, S. A., Bekele, B. N. & Fidler, I. J. An orthotopic nude mouse model of oral tongue squamous cell carcinoma. *Clin Cancer Res* **8**, 293-298 (2002).

2 Benaich, N. *et al.* Rewiring of an epithelial differentiation factor, miR-203, to inhibit human squamous cell carcinoma metastasis. *Cell Rep* **9**, 104-117, doi:10.1016/j.celrep.2014.08.062 (2014).

3 Schneider, TE *et al.* M*easuring stem cell frequency* in *epidermis*: a *quantitative* in *vivo functional assay for long-term repopulating cells. Proc Natl Acad Sci USA **100**, 11412-11417, doi:*10.1073/pnas.2034935100 (2003).

4 Shwartz, Y. et al. Cell Types Promoting Goosebumps Form a Niche to Regulate Hair Follicle Stem Cells. *Cell **182**,* 578-593, (2021).

5 Bhandari, M. *et al.* Galanin receptor antagonist M35 but not M40 or C7 ameliorates cerulein-induced acute pancreatitis in mice. *Pancreatology* **10,** (2011).

6 Zuber, J. *et al.* Mouse models of human AML accurately predict chemotherapy response. *Genes Dev* **23**, 877-889, doi:10.1101/gad.1771409 (2009).

7 Bartus, K. *et al.* Large-scale chondroitin sulfate proteoglycan digestion with chondroitinase gene therapy leads to reduced pathology and modulates macrophage phenotype following spinal cord contusion injury. *Journal of Neuroscience* **34,** (2014).

8 Nowak, J. A. & Fuchs, E. Isolation and culture of epithelial stem cells. *Methods Mol Biol* **482**, 215-232, doi:10.1007/978-1-59745-060-7_14 (2009).

9 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

10 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

11 Li, H. *et al.* The Sequence Alignment / Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. *Bioinformatics* **25,** 1–2 (2009).

12 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** (2008).

13 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).

14 R Core Team & R Foundation for Statistical Computing. *R: a Language and Environment for Statistical Computing. http://www.R-project.org/* (2018).

15 Stark, R. & Brown, G. DiffBind iffBind : differential binding analChIP-Seq peak data. *Bioconductor* 1–27 (2011).

16 Yu, G., Wang, L. G. & He, Q. Y. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31,** 2382–2383 (2015).

17 Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** (2013).

18 Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnology* **29,** 24–26 (2011).

19 Douillet, D. *et al.* Uncoupling histone H3K4 trimethylation from developmental gene expression via an equilibrium of COMPASS, Polycomb and DNA methylation. *Nat. Genet.* (2020). doi:10.1038/s41588-020-0618-1

20 Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10,** 1096–1100 (2013).

21 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36,** 411–420 (2018).

22 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32,** 381–386 (2014).

23 Gonzalez-Roca, E. *et al.* Accurate expression profiling of very small cell populations. *PLoS One* **5**, e14418, doi:10.1371/journal.pone.0014418 (2010).

24 Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* **4**, 249-264, doi:10.1093/biostatistics/4.2.249 (2003).

25 Chenj, J. *et al.* ToppGene Suite for gene list enrichment analysis and candidate genes for priorizitation. *Nucleic Acid Res*, **37**, W305-W311 doi:10.1093/nar/gkp427 (2009).

26 Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**, 2502-2504 (2003).

27 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).

28 Dougherty, J.D. et al. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res, **38** 4218-4230, doi: 10.1093/nar/gkq130 (2010).

29 Xu, X. et al. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. J. Neurosci, 34 1420-31, doi: 10.1523/JNEUROSCI.4488-13.2014 (2014).

30 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genom Res,* **13,** 2498-504, doi: 10.1101/gr1239303 (2003).

31 Zambelli, F. *et al.* Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res **37**, W247-W252, doi:10.1093/nar/gkp464l (2009).

32 Nagy, A. *et al.* Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep, **8**, 11515 (2018).

33 R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (2008).

34 Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80, doi:10.1186/gb-2004-5-10-r80 (2004).

35 Gautier, L., Cope, L., Bolstad, B. M., Bioinformatics, R. I.2004. affy—analysis of Affymetrix GeneChip data at the probe level. *academic.oup.com* doi:10.1093/bioinformatics/btg405

36 Bolstad, B. M. *et al.* in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 33–47 (Springer, New York, NY, 2005). doi:10.1007/0-387-29362-0_3

37 Eklund, A. C. & Szallasi, Z. Correction of technical bias in clinical microarray data improves concordance with known biological information. *Genome Biol* **9,** 1–8 (2008).

38 Smyth, G. Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor. *Springer, New York*, 397-420 (2005).

39 Benjamini, Y. H. Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

40 Wu D. et al. ROAST: rotation gene set tests for complex microarray experiments. Bioinformatics, **26**, 2176-82 (2010).

41 Bradley Efron & Robert Tibshirani. On Testing the Significance of Sets of Genes. The Annals of Applied Statistics. **1,** 107-129 (2007).

42 Goeman JJ, & Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics. **23**, 980-7 (2007).

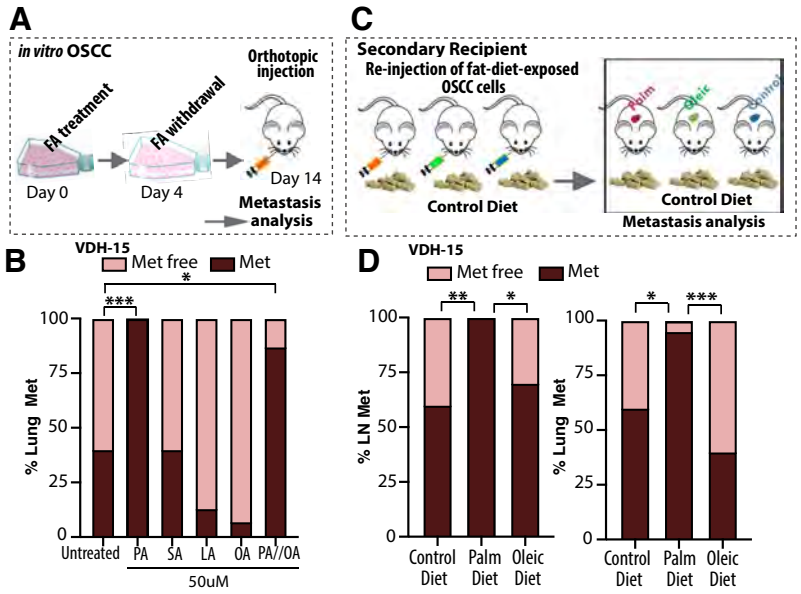
43 Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25,** 25–29 (2000).

44 Carlson M (2019). org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2.

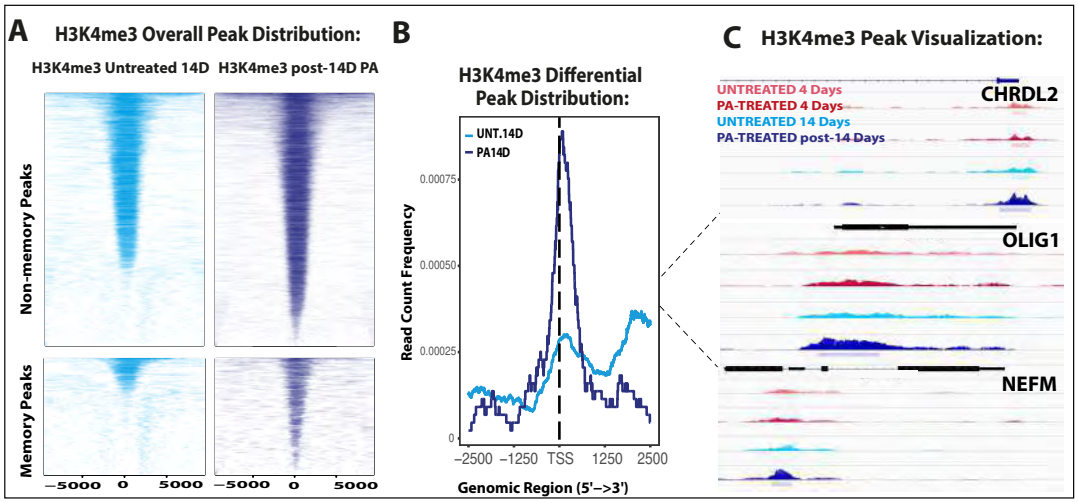
45 Carlson M (2019). org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.8.2.

46 Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat Meth* **15,** 255–261 (2018).

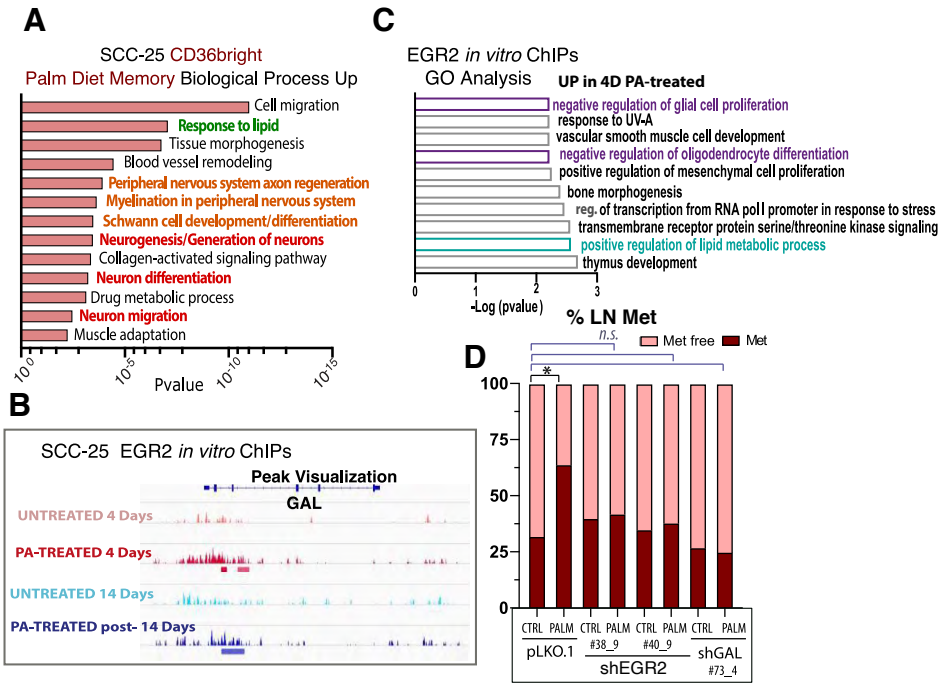
Main Figure 1



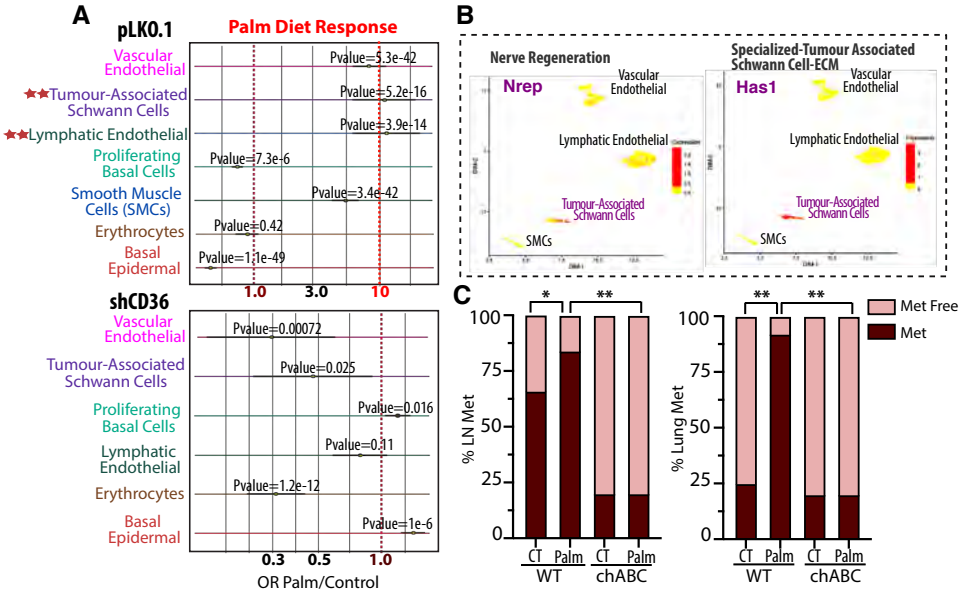
Main Figure 2



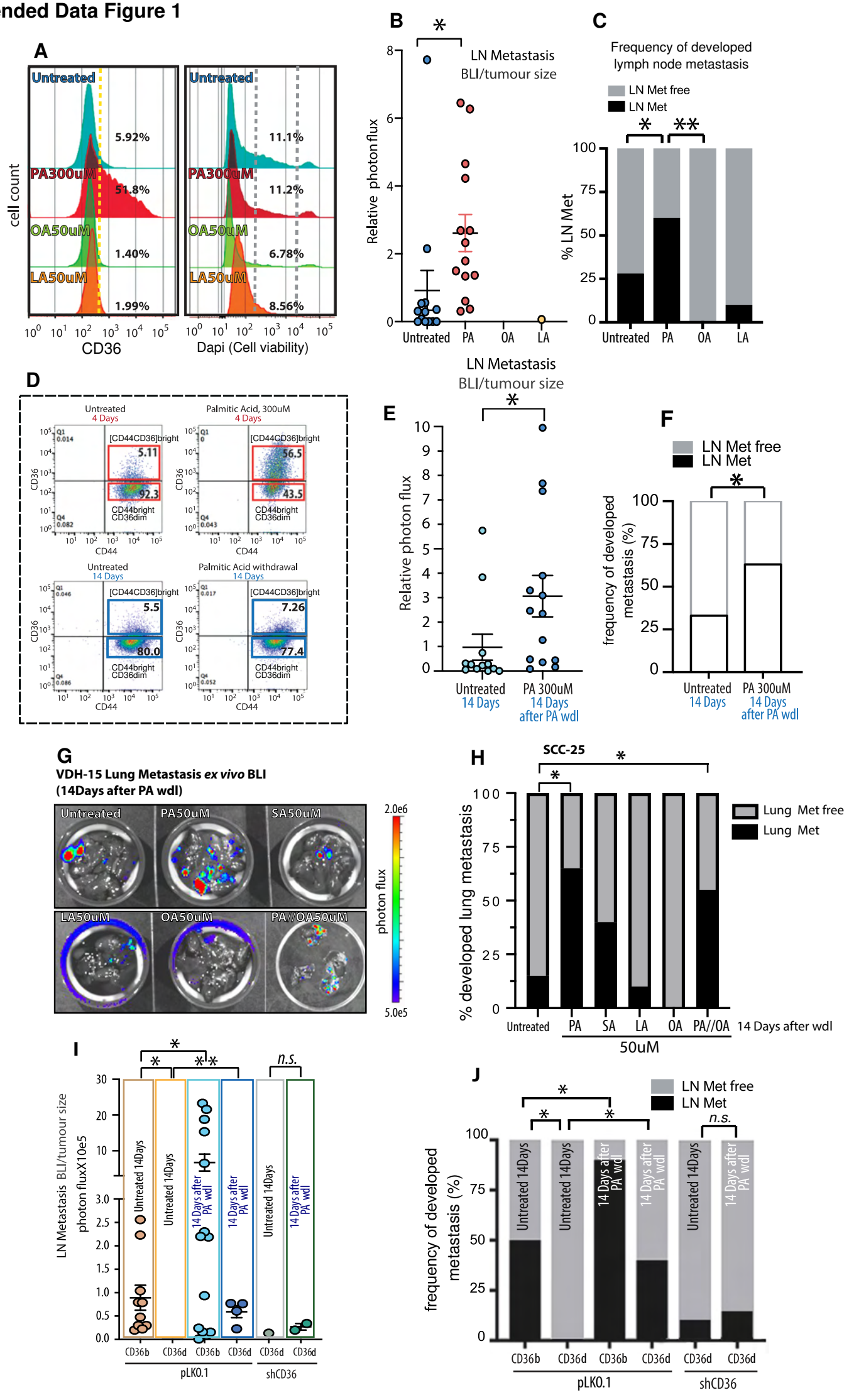
Main Figure 3



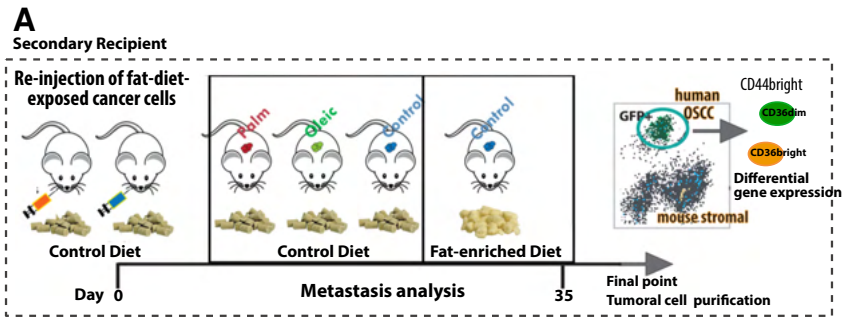
Main Figure 4

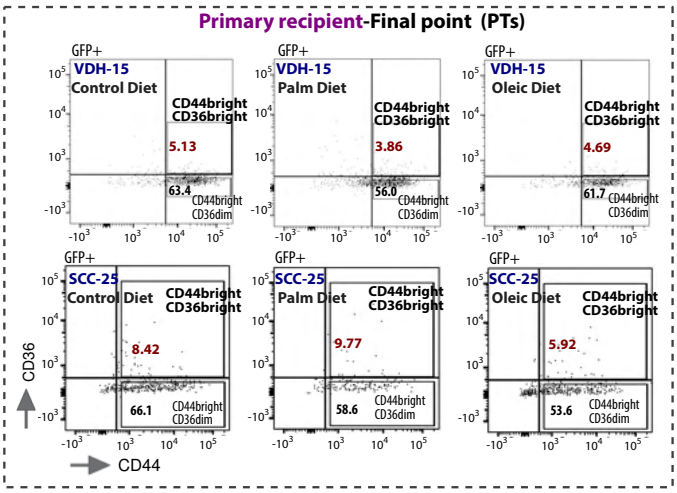


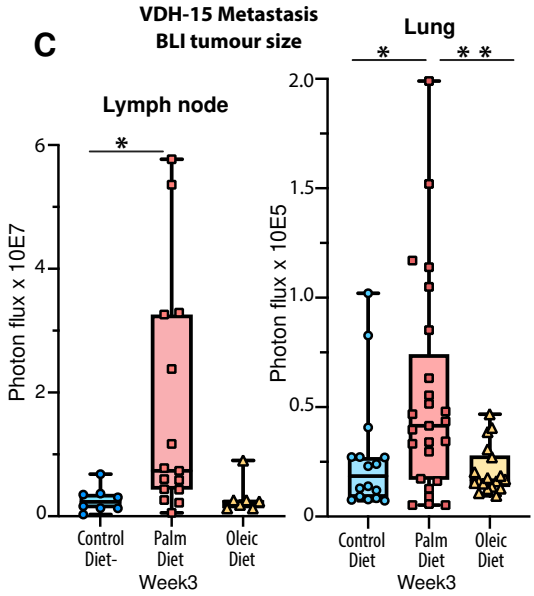
Extended Data Figure 1

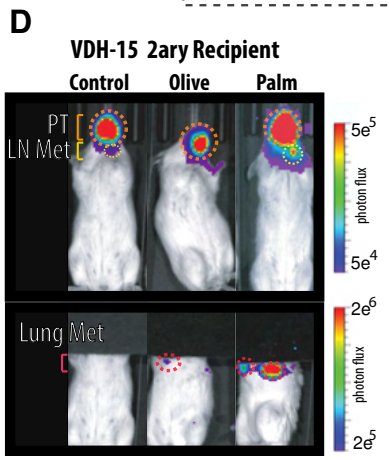


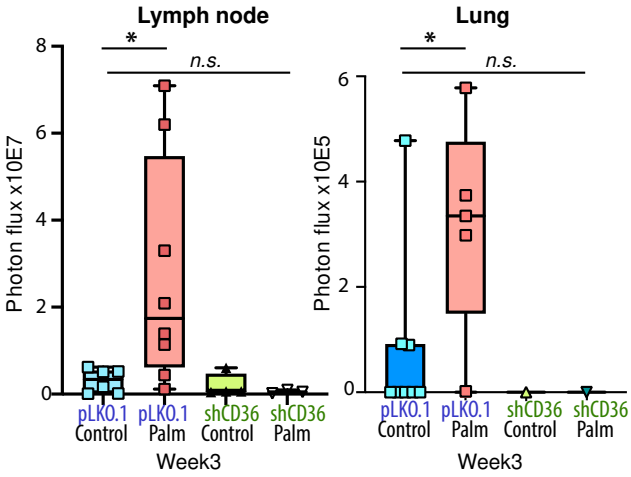
Extended Data Figure 2

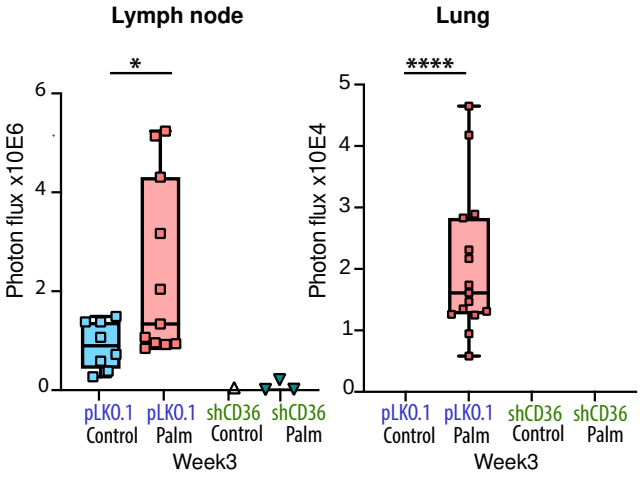


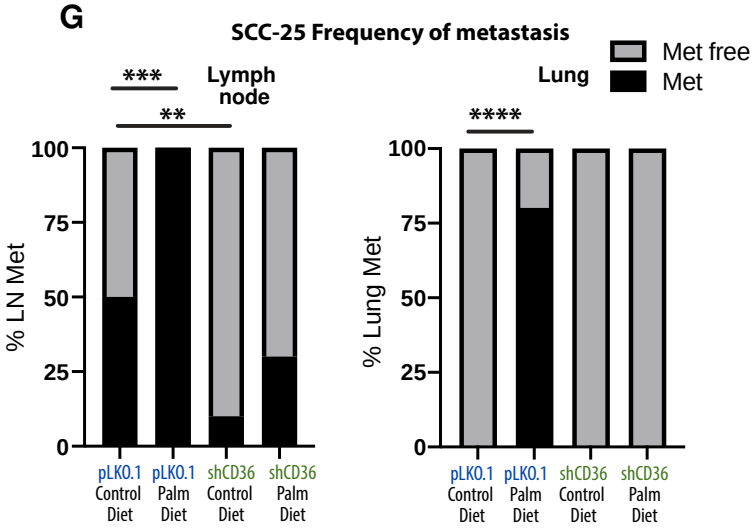


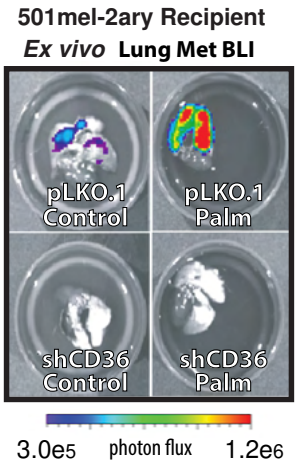


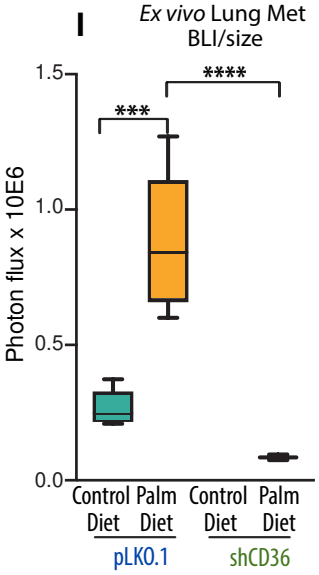




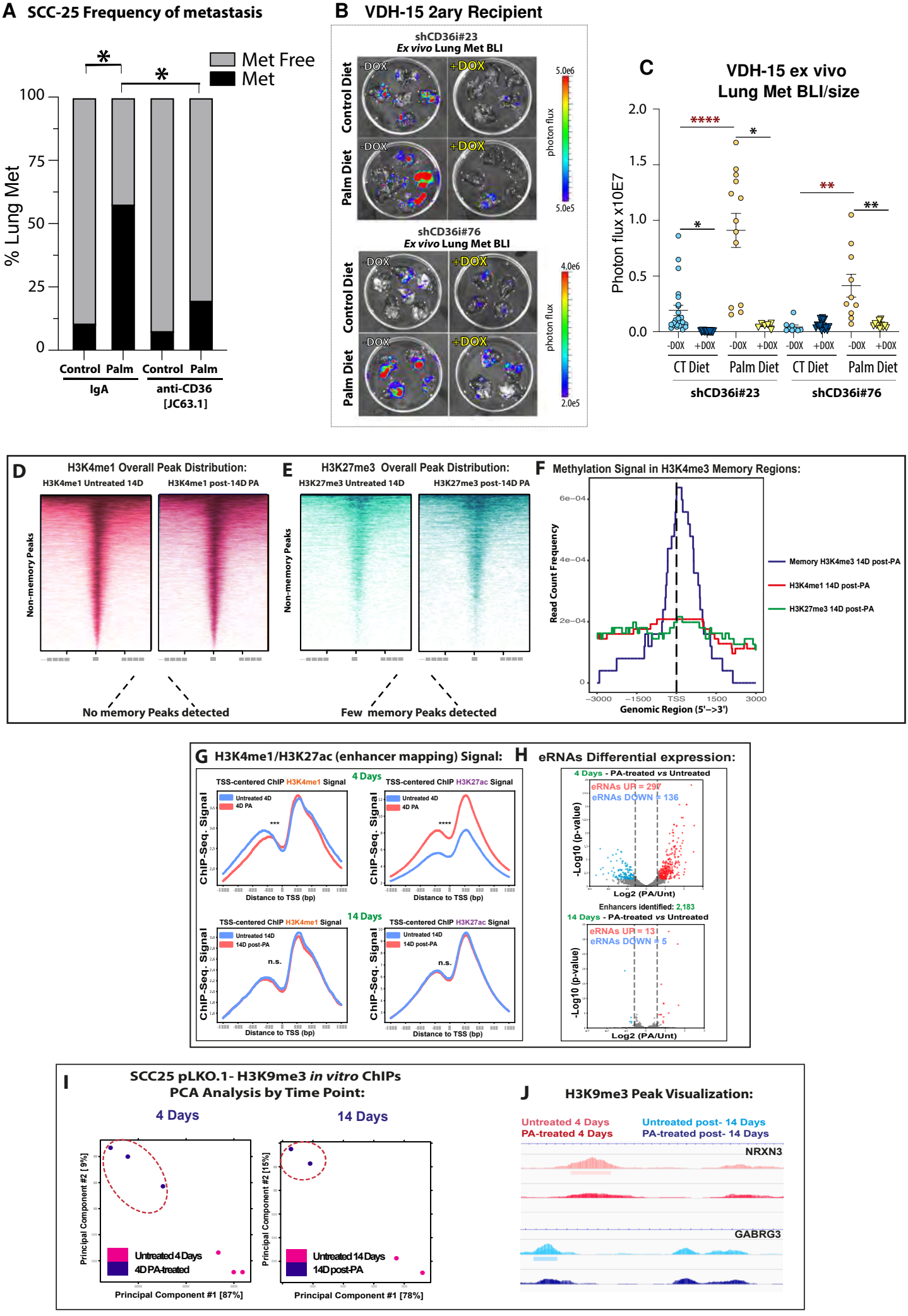




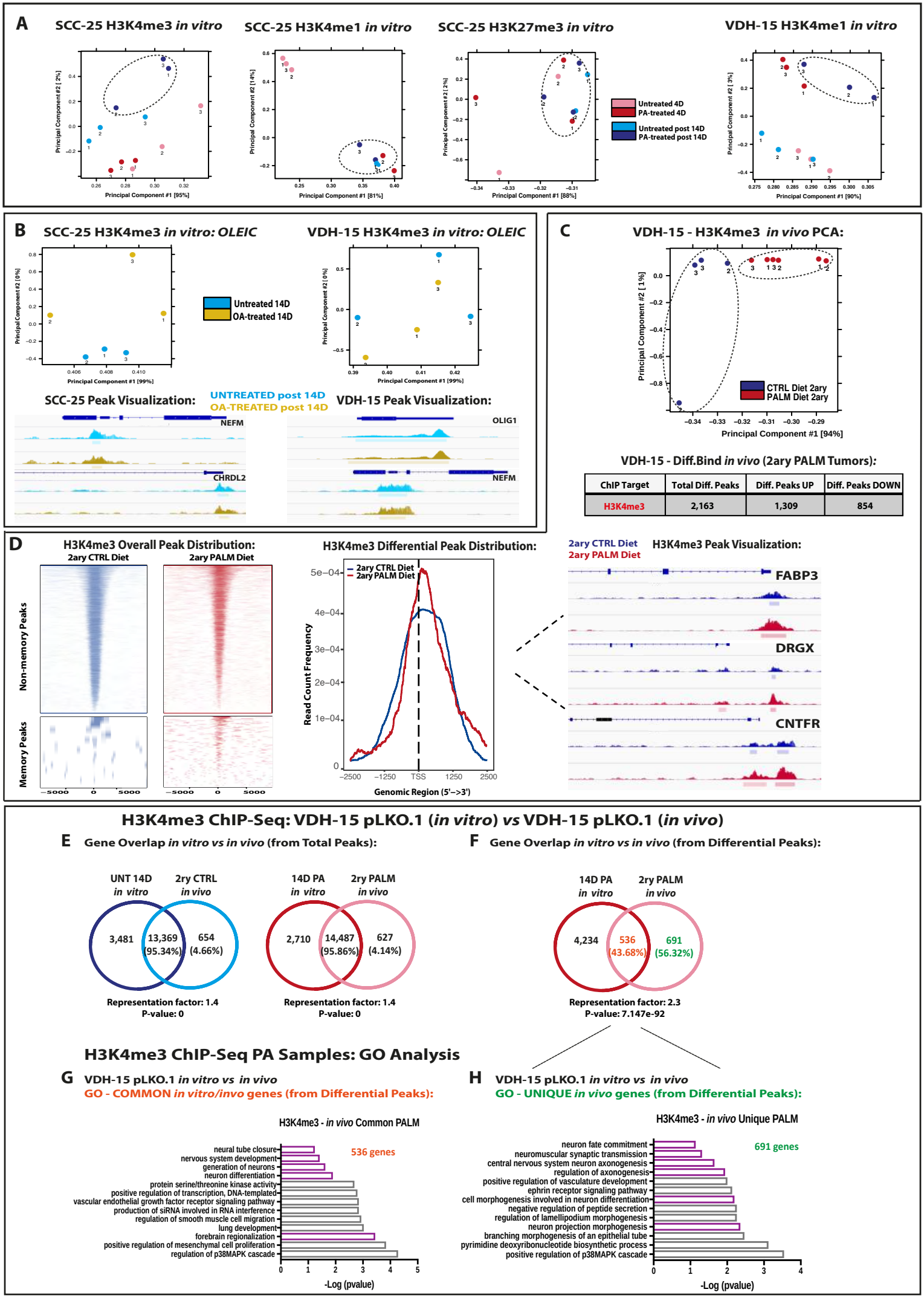




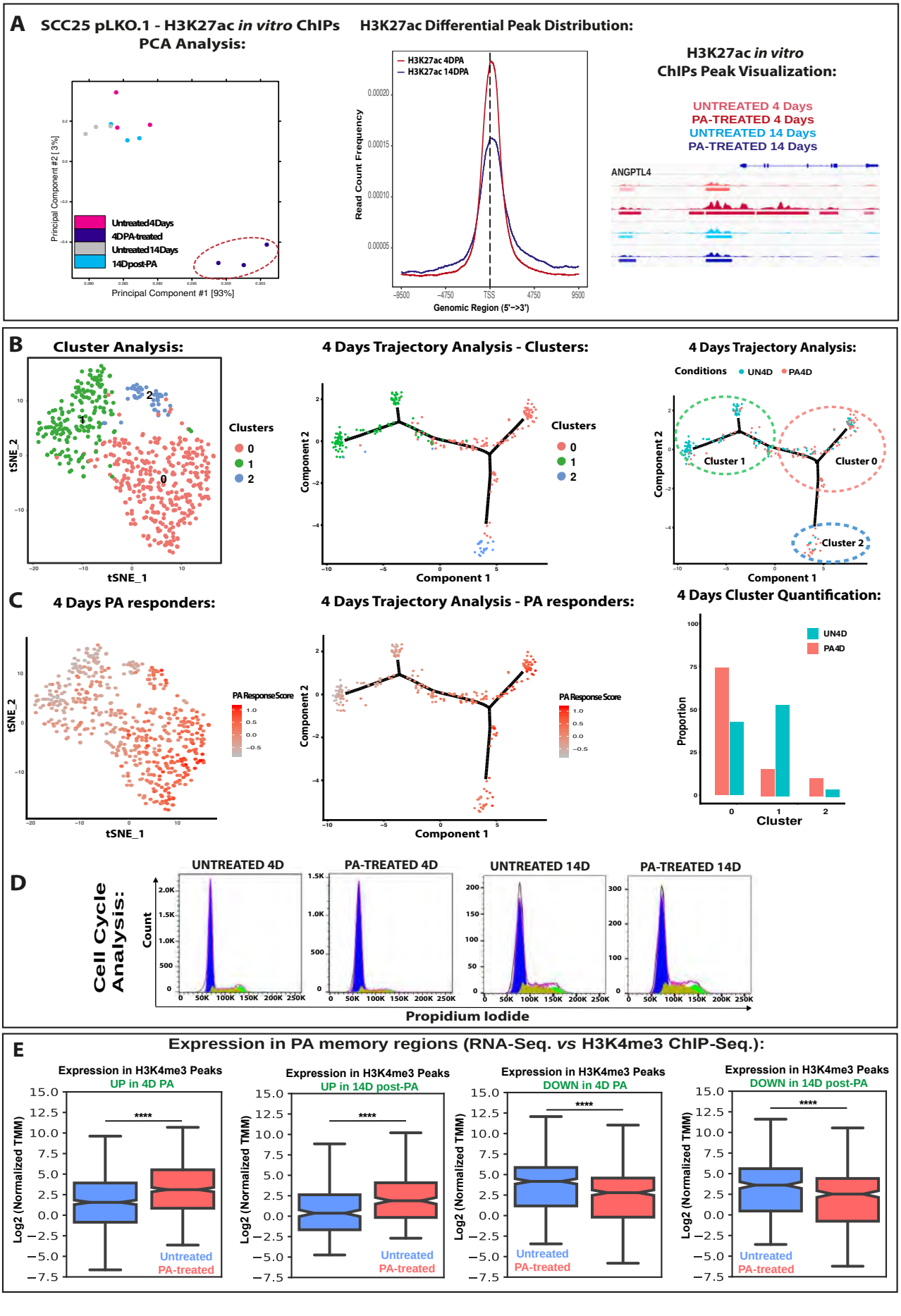
Extended Data Figure 3



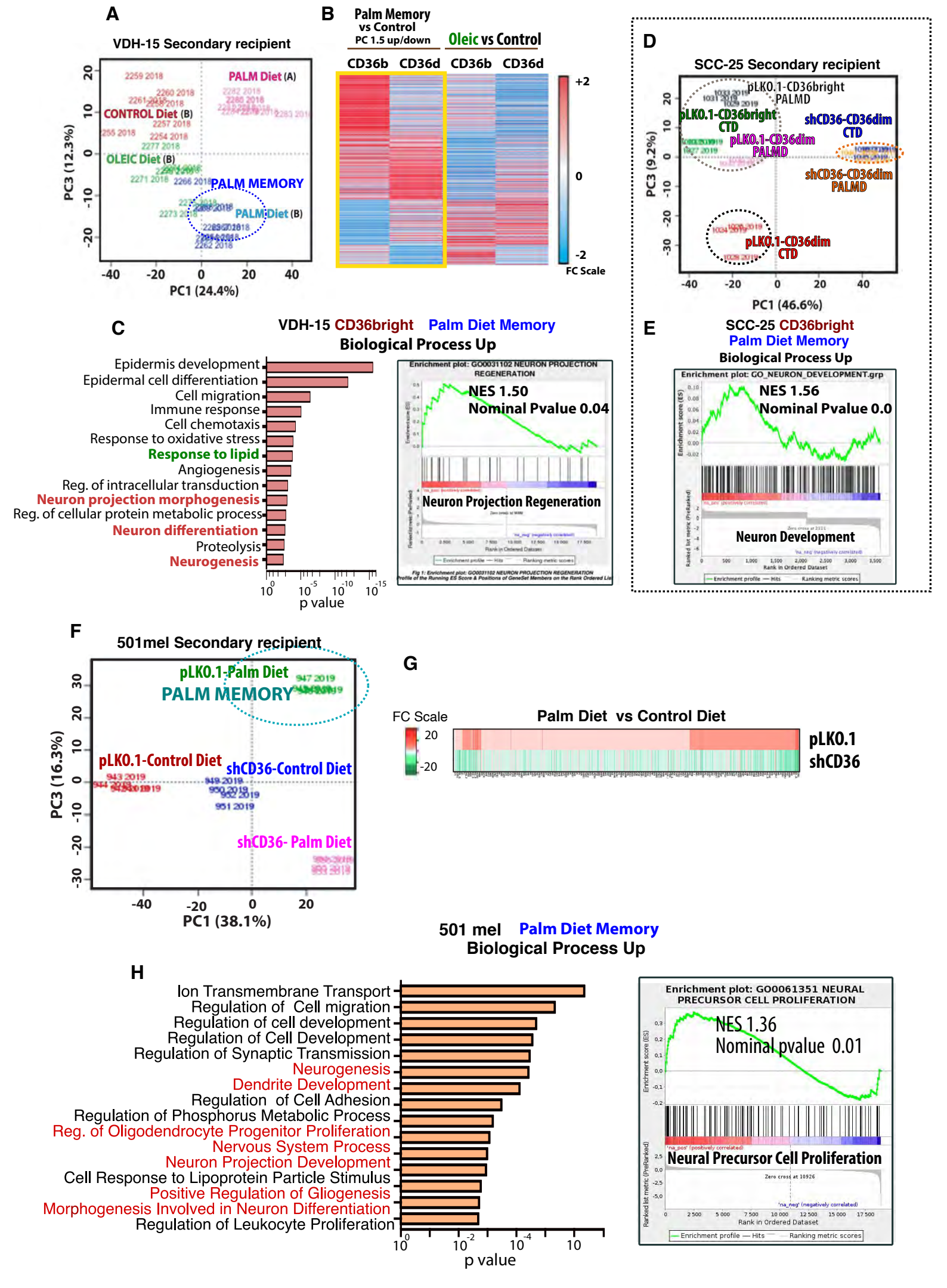
Extended Data Figure 4



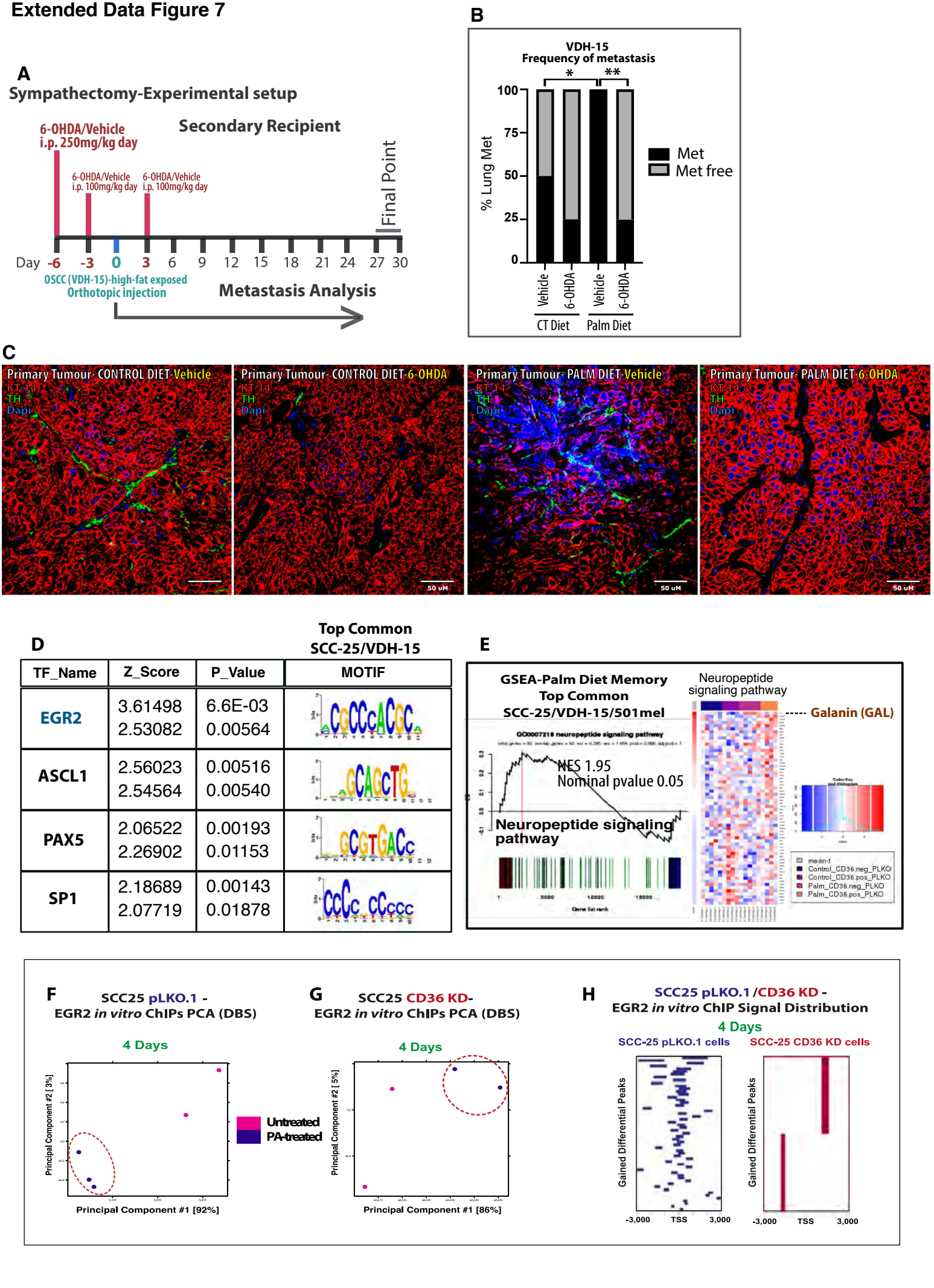
Extended Data Figure 5



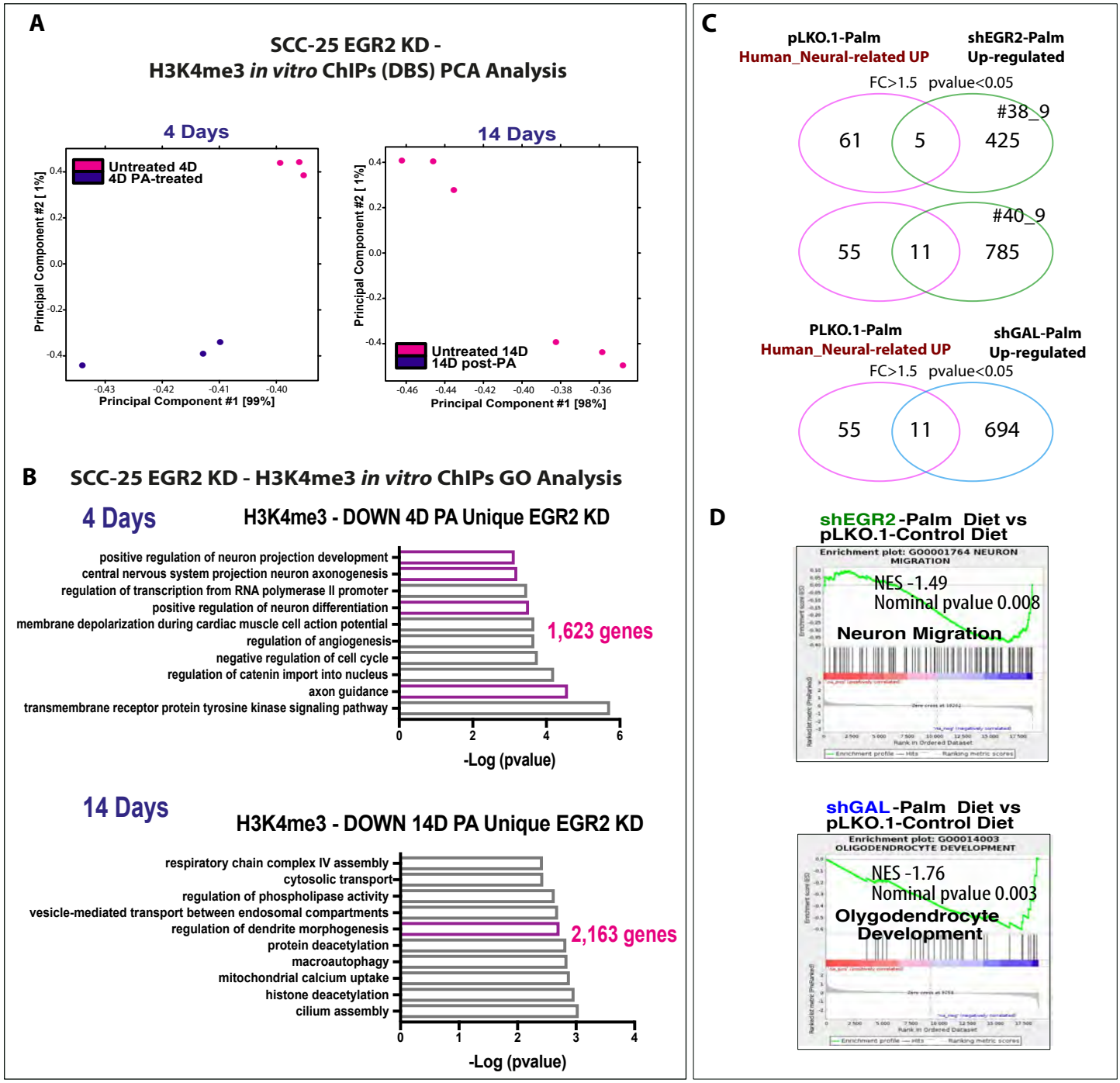
Extended Data Figure 6

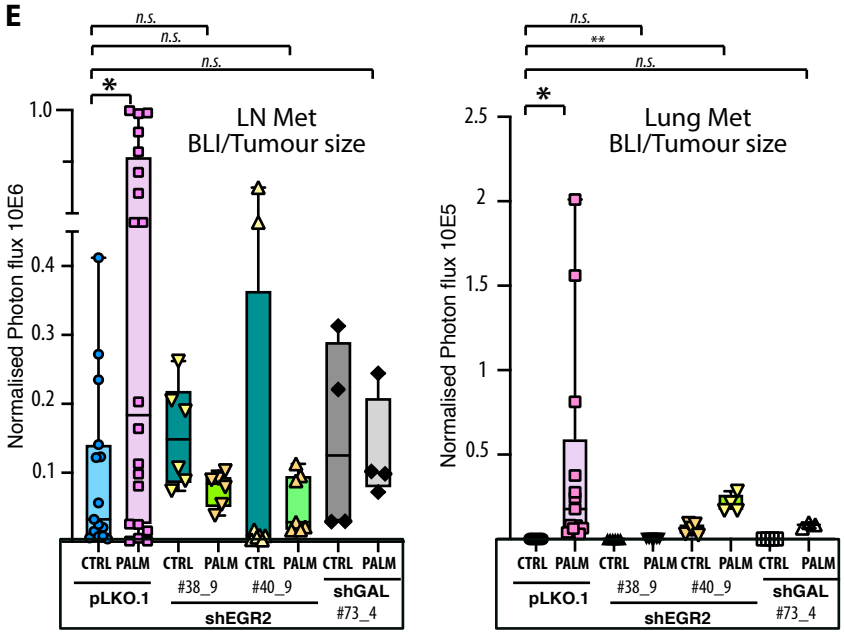


Extended Data Figure 7

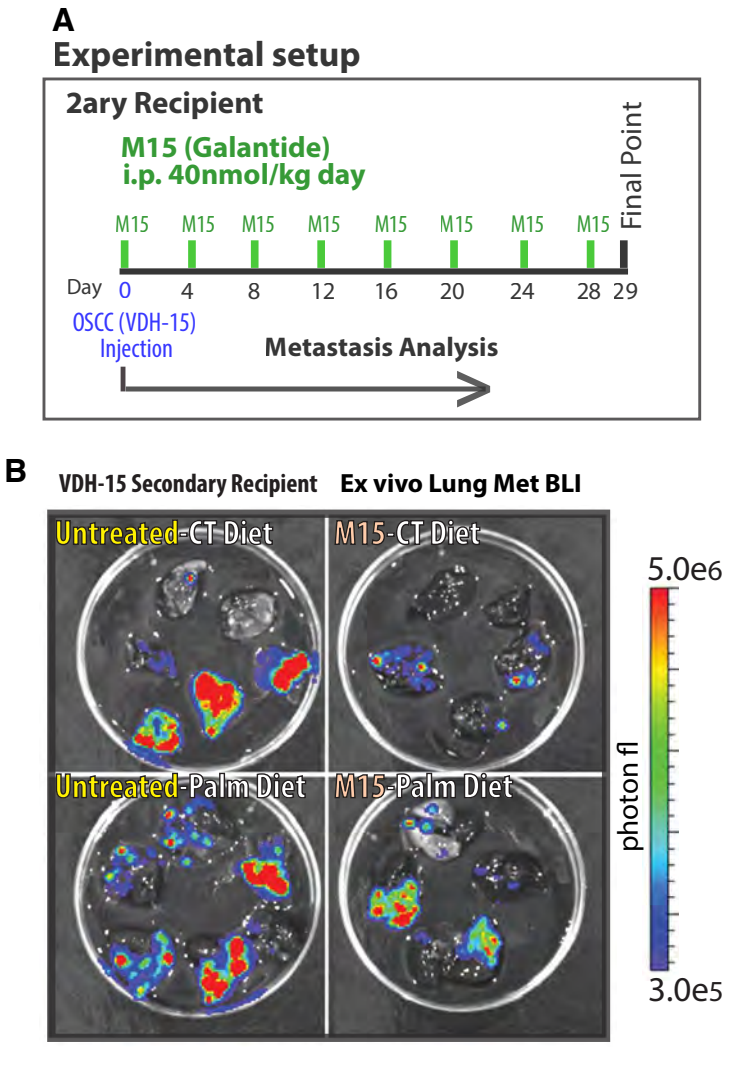


Extended Data Figure 8

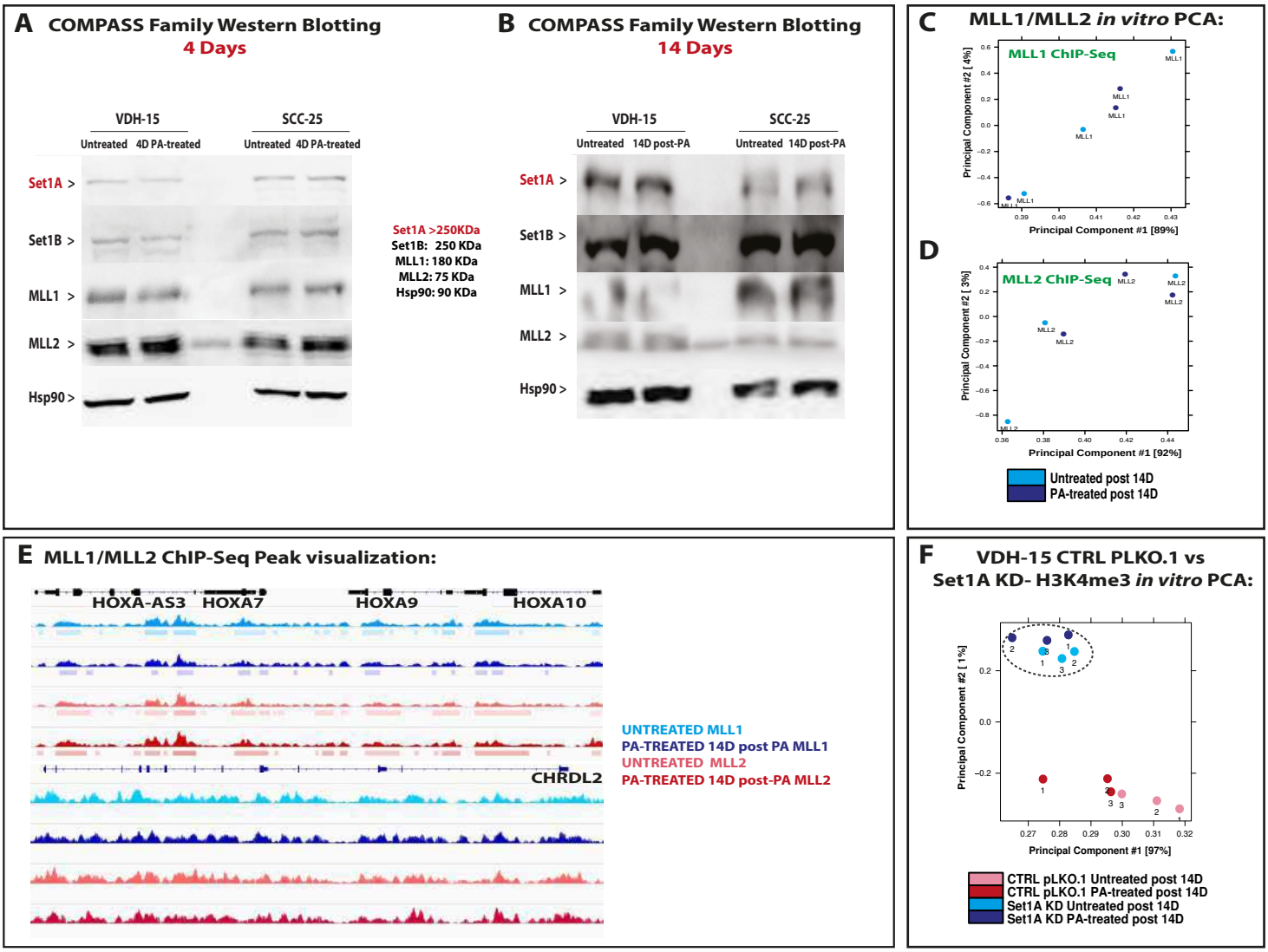


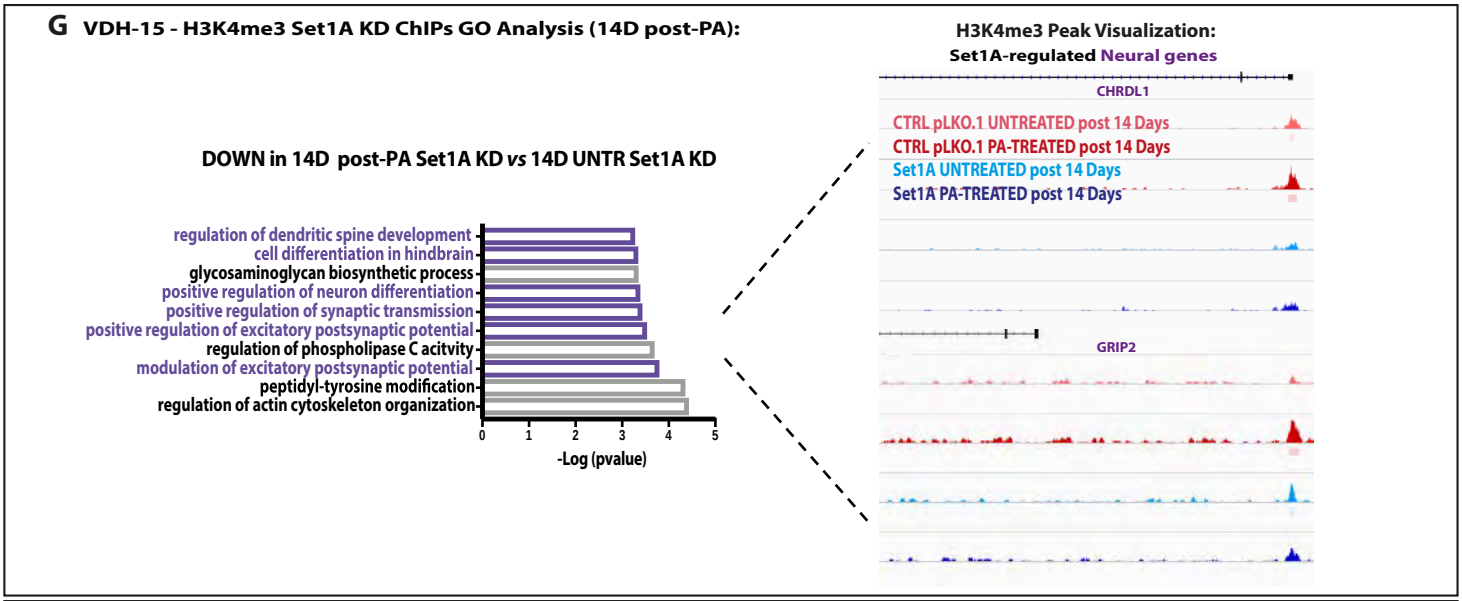


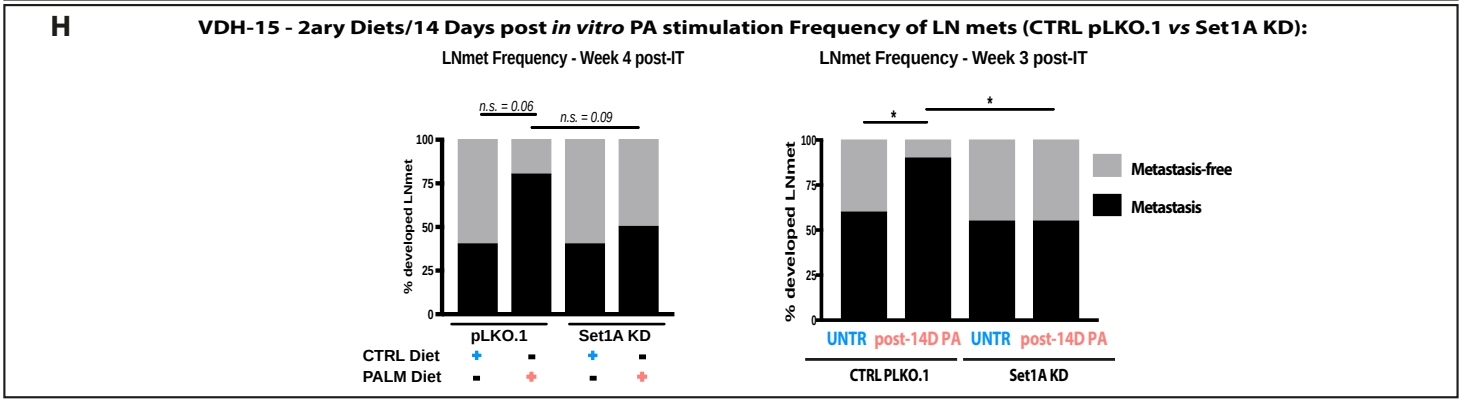
Extended Data Figure 9



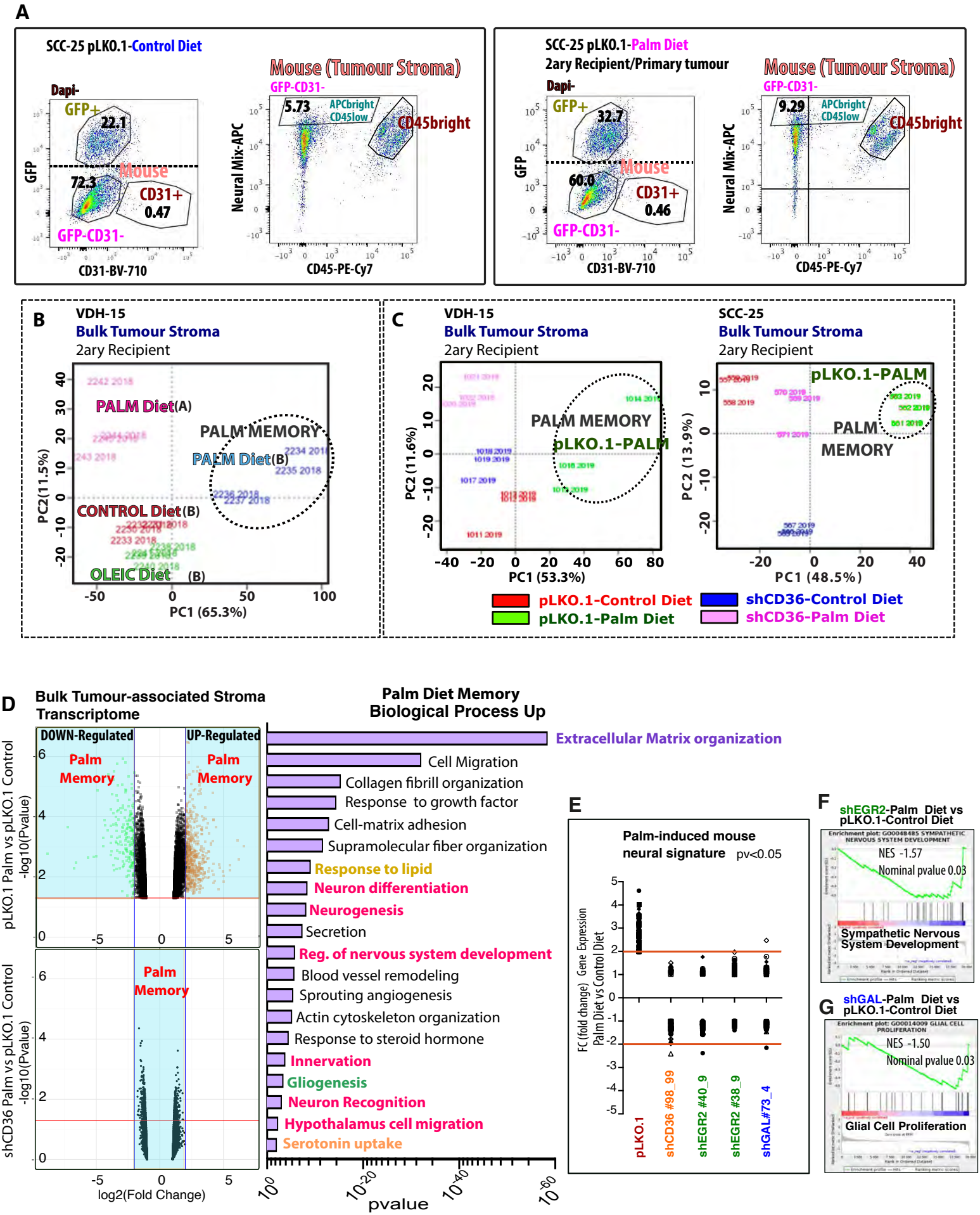
Extended Data Figure 10



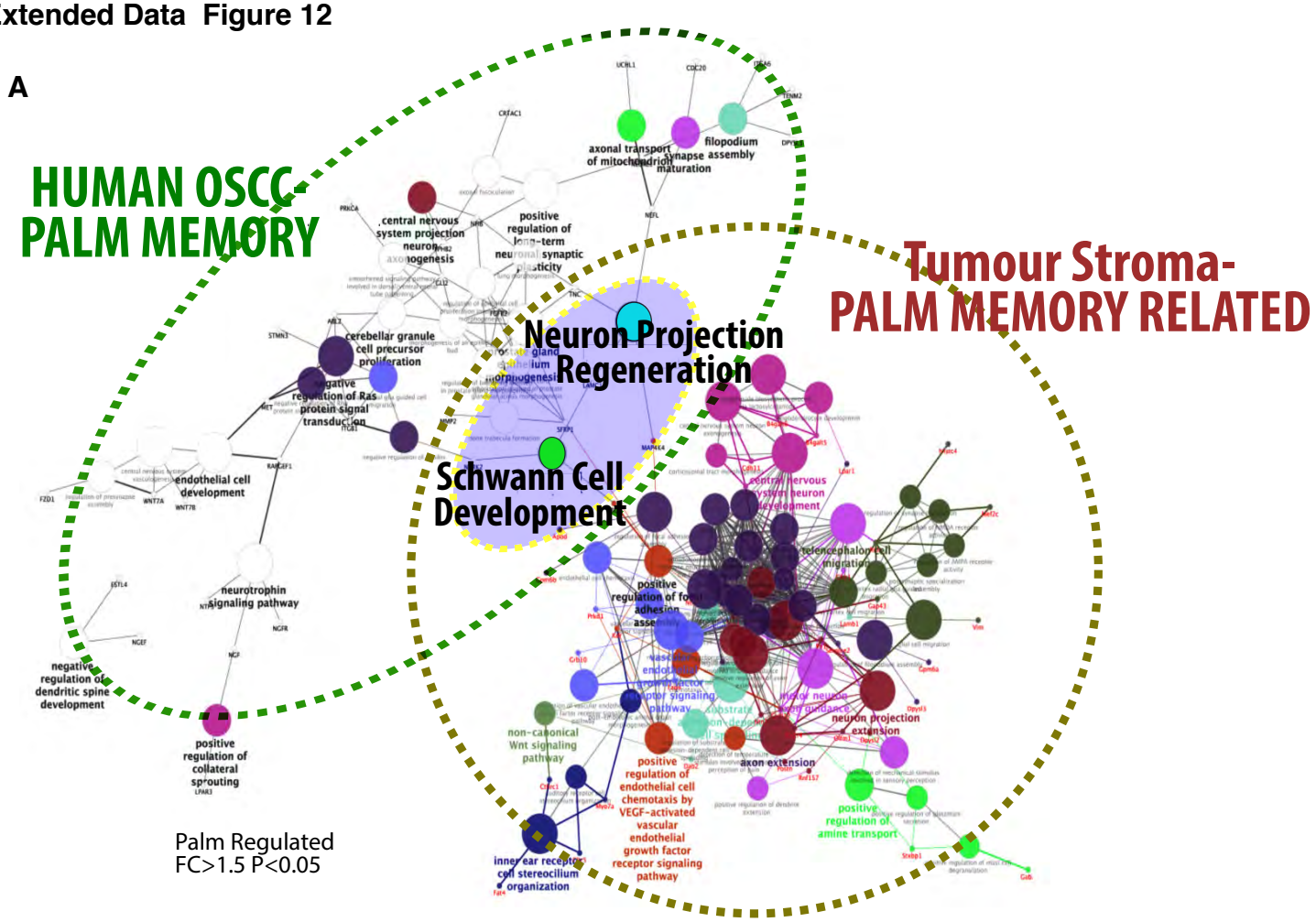


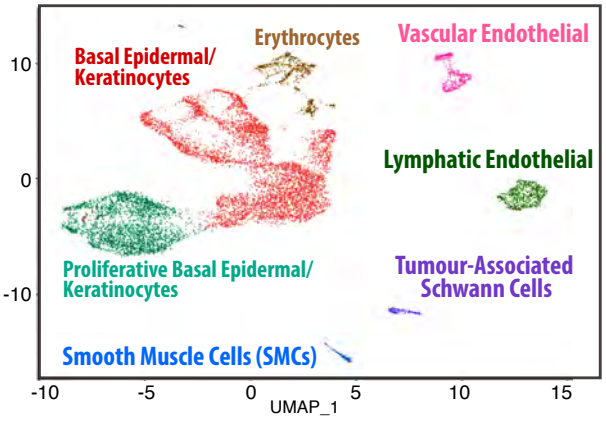


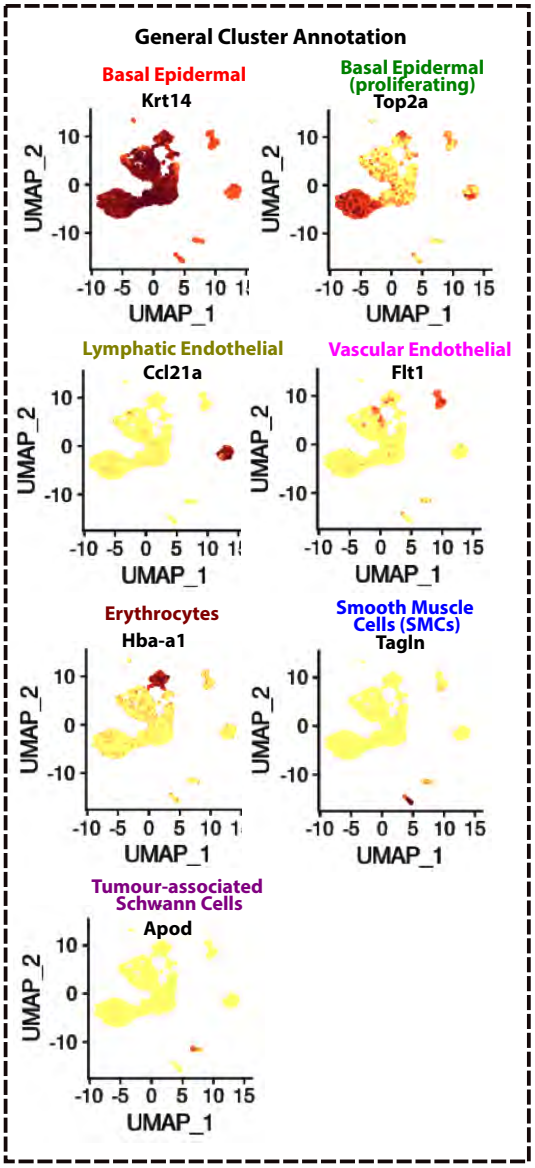
Extended Data Figure 11

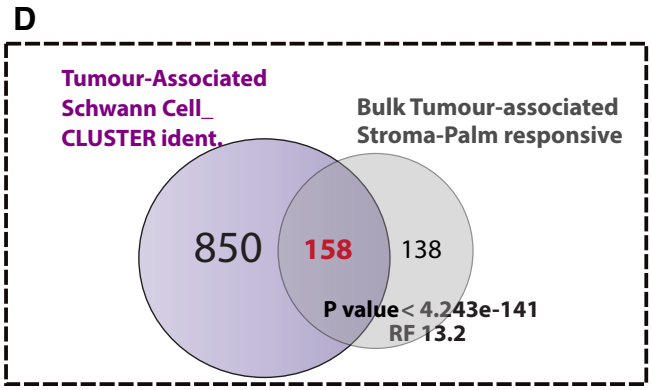


Extended Data Figure 12

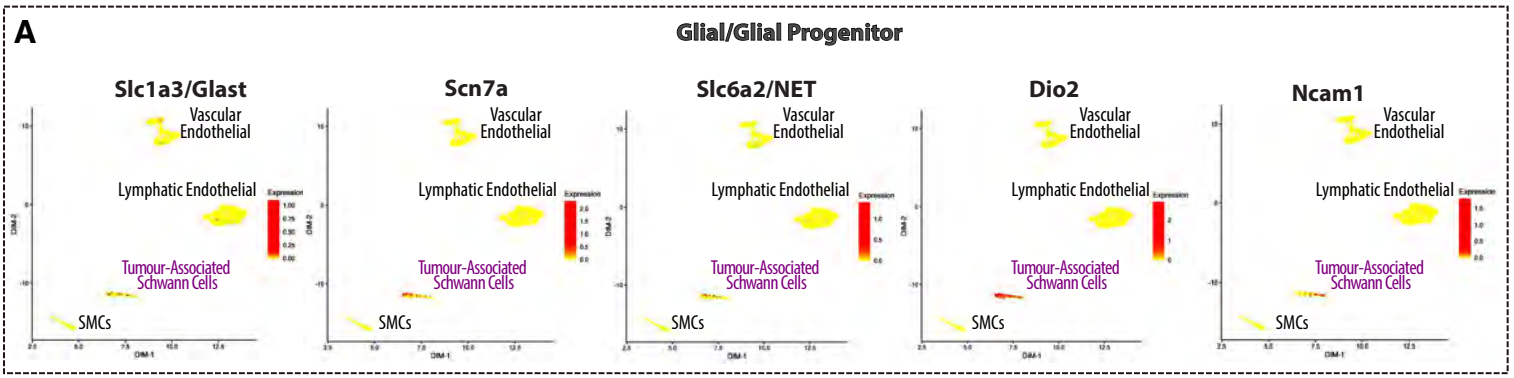


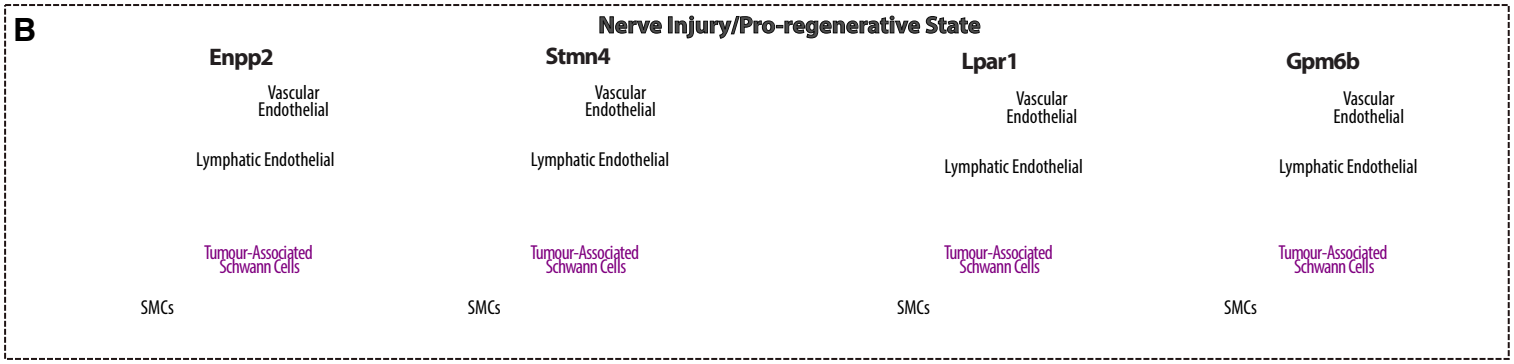


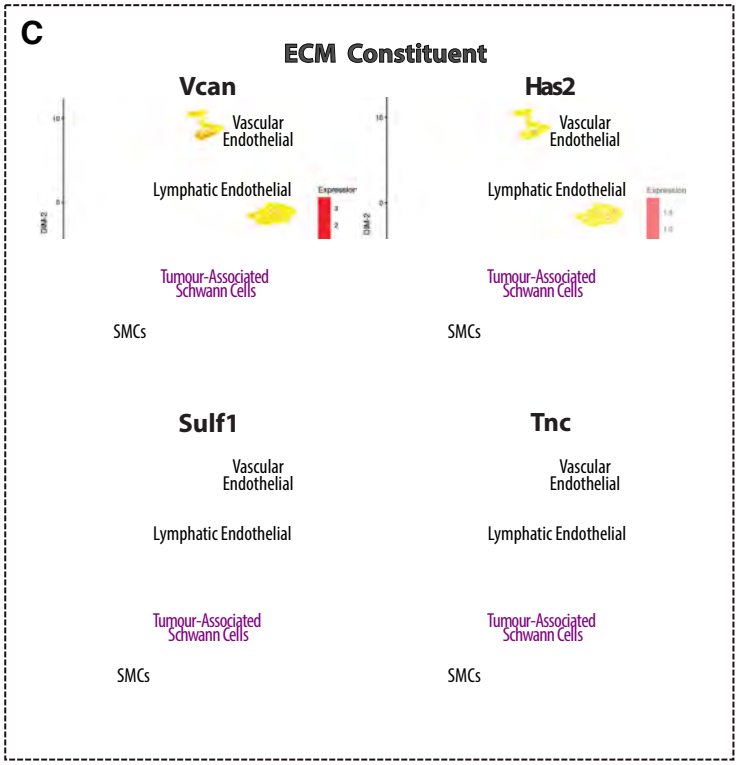


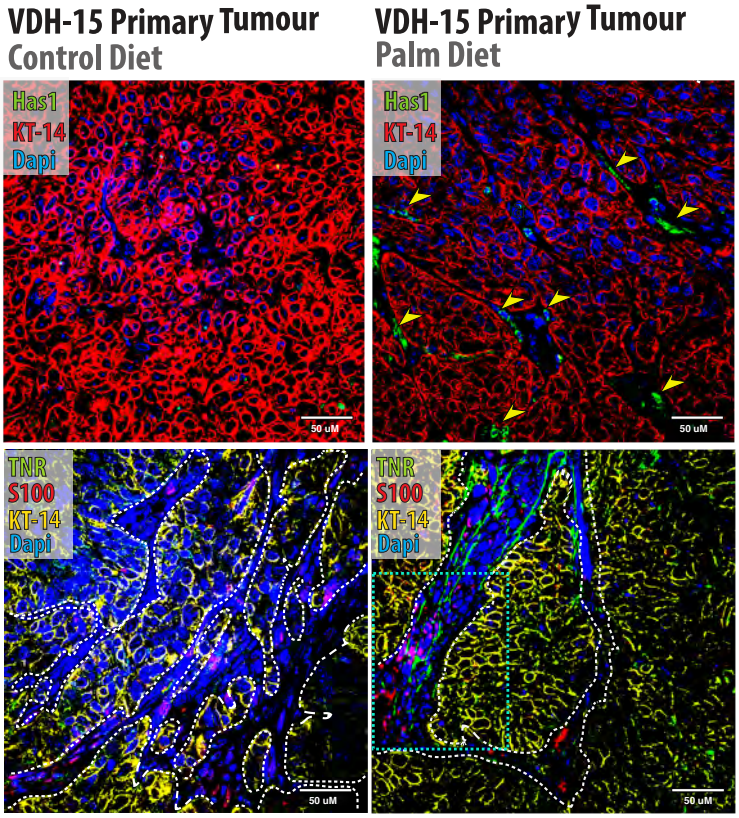


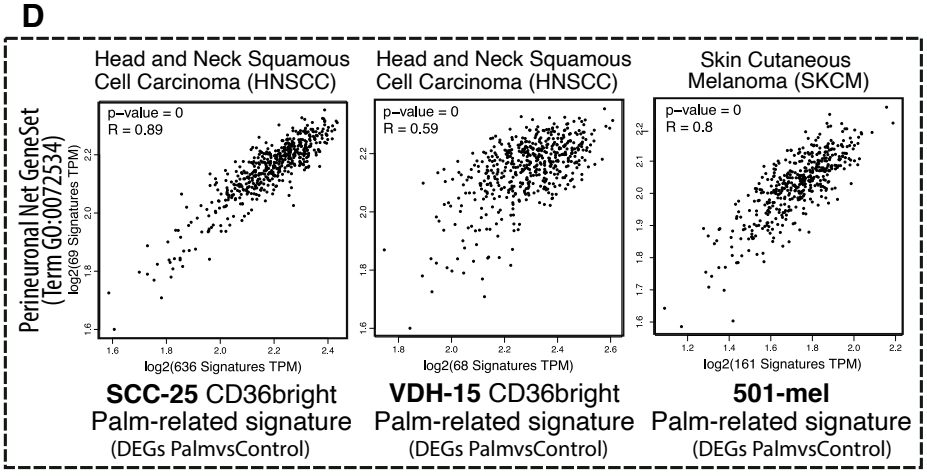
Extended Data Figure 13

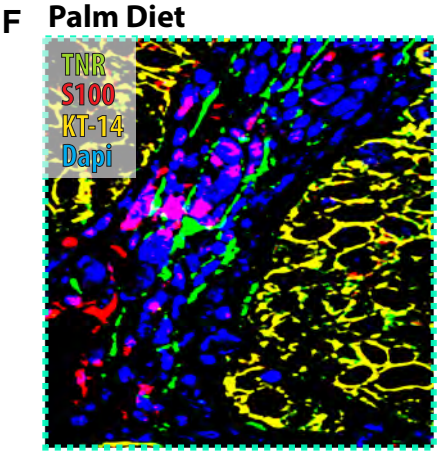




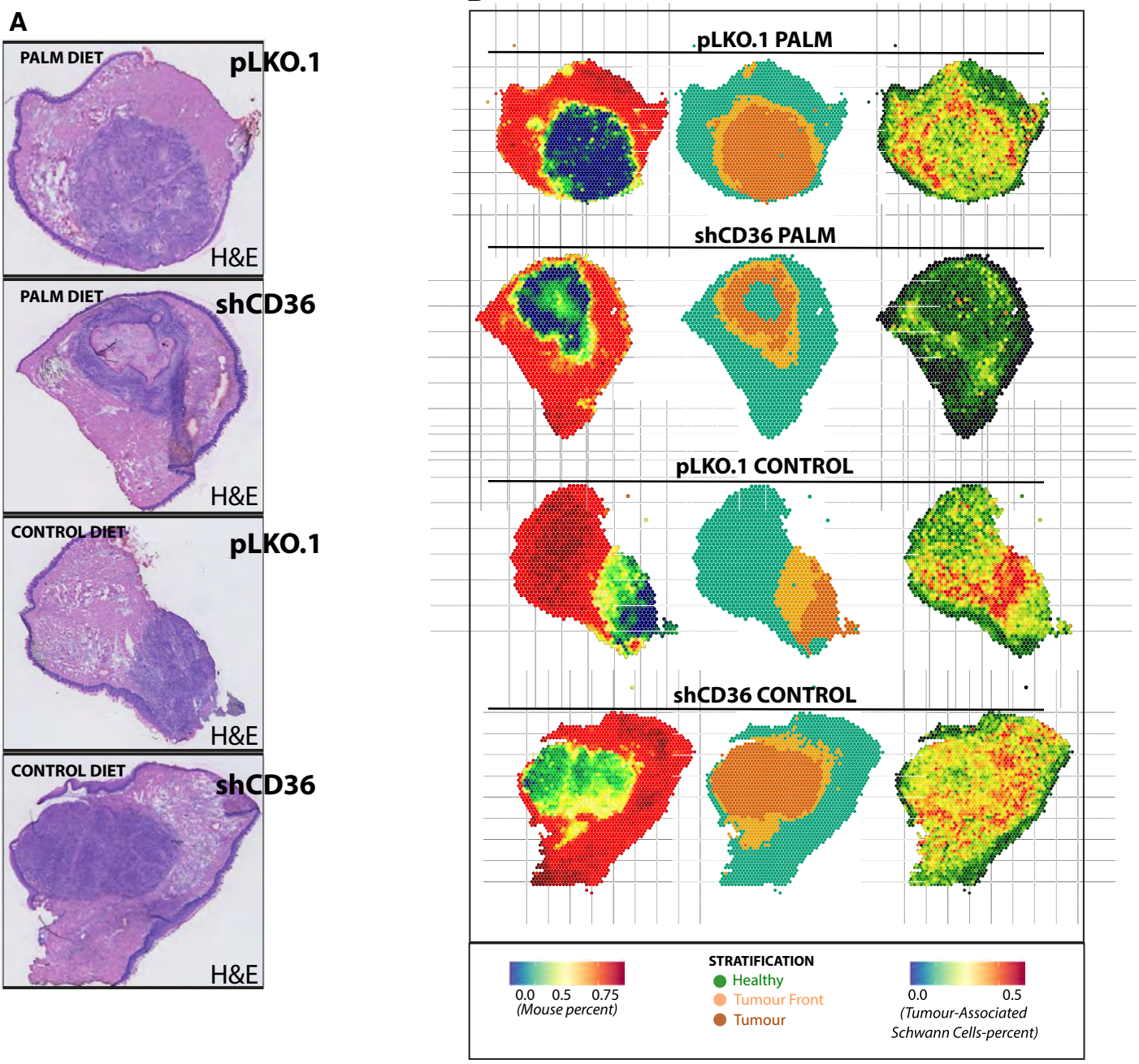


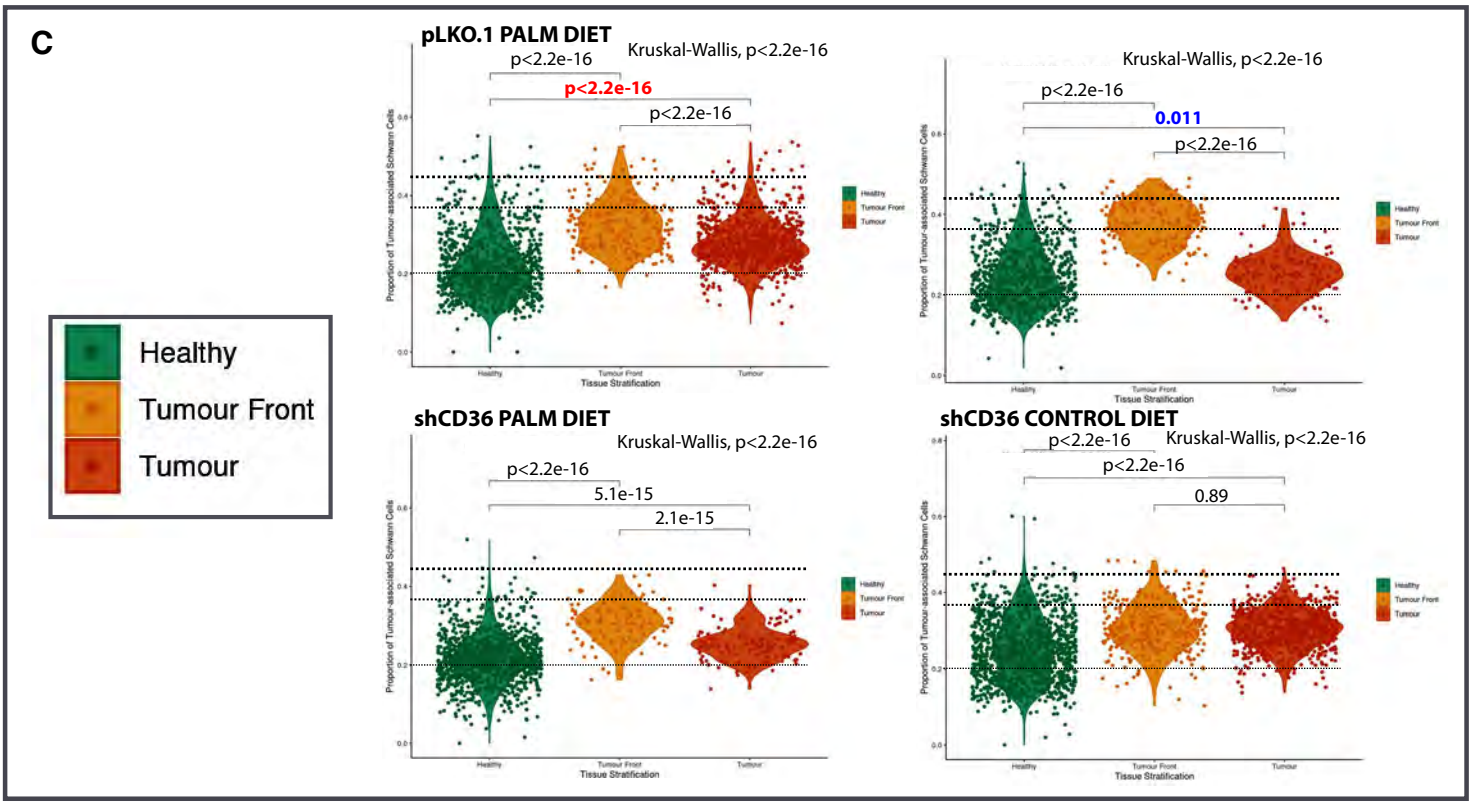




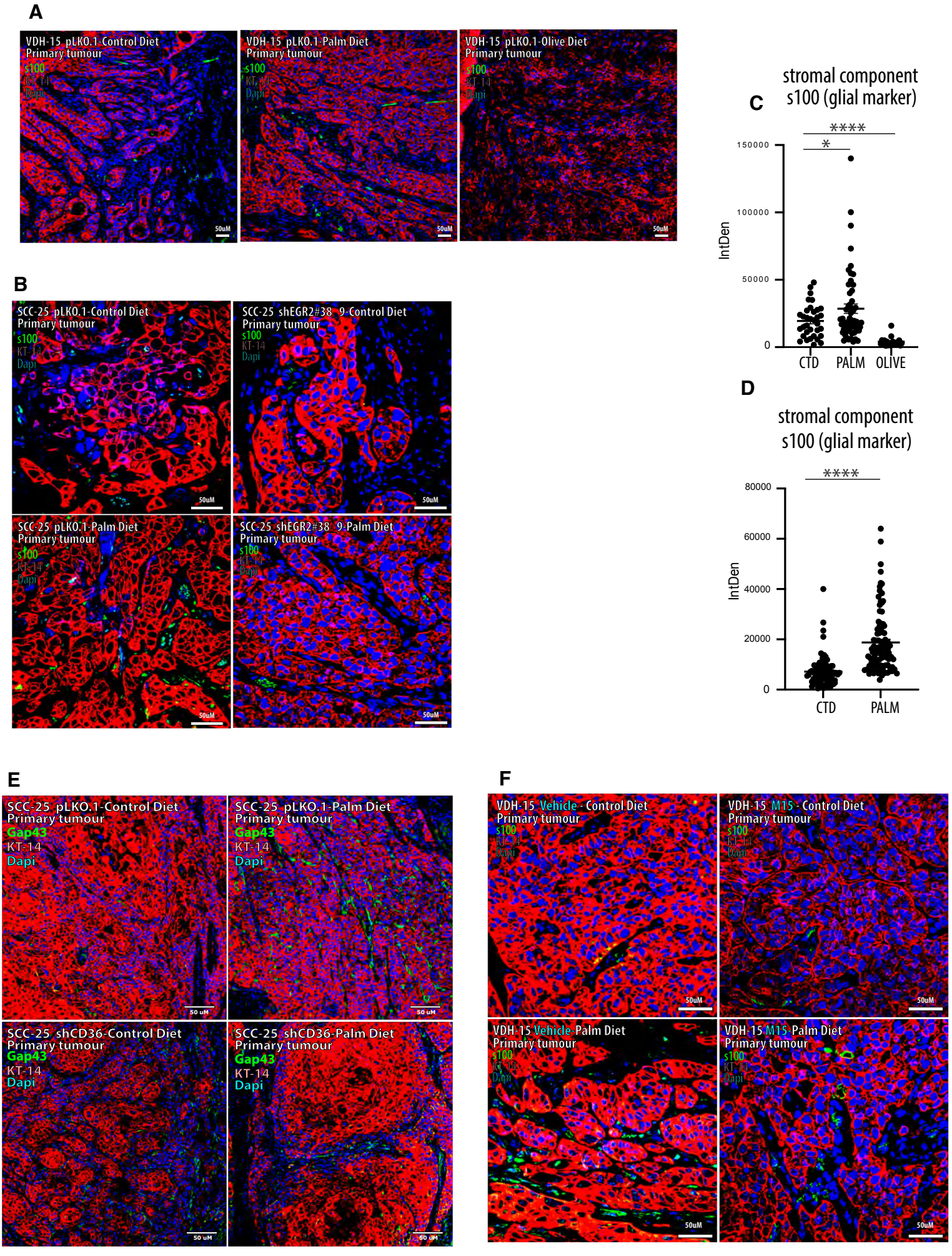


Extended Data Figure 14





Extended Data Figure 15



Extended Data Figure 16