HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2022 August 27.

Published in final edited form as:

Nature. 2022 July ; 607(7917): 169–175. doi:10.1038/s41586-022-04842-7.

OCA-T1 and OCA-T2 are coactivators of POU2F3 in the tuft cell lineage

Xiaoli S. Wu[1,2], **Xue-Yan He**[1], **Jonathan J. Ipsaro**[1,3], **Yu-Han Huang**[1], **Jonathan B. Preall**[1], **David Ng**[1], **Yan Ting Shue**[4,5], **Julien Sage**[4,5], **Mikala Egeblad**[1], **Leemor Joshua-Tor**[1,3], **Christopher R. Vakoc**[1,✉]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

[2]Genetics Program, Stony Brook University, Stony Brook, NY, USA.

[3]Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

[4]Department of Pediatrics, Stanford University, Stanford, CA, USA.

[5]Department of Genetics, Stanford University, Stanford, CA, USA.

Abstract

Tuft cells are a rare chemosensory lineage that coordinates immune and neural responses to foreign pathogens in mucosal tissues[1]. Recent studies have also revealed tuft-cell-like human tumours[2,3], particularly as a variant of small-cell lung cancer. Both normal and neoplastic tuft cells share a genetic requirement for the transcription factor POU2F3 (refs. [2,4]), although the transcriptional mechanisms that generate this cell type are poorly understood. Here we show that binding of POU2F3 to the uncharacterized proteins C11orf53 and COLCA2 (renamed here OCA-T1/*POU2AF2* and OCA-T2/*POU2AF3*, respectively) is critical in the tuft cell lineage. OCA-T1 and OCA-T2 are paralogues of the B-cell-specific coactivator OCA-B; all three proteins are encoded in a gene cluster and contain a conserved peptide that binds to class II POU transcription

✉**Correspondence and requests for materials** should be addressed to Christopher R. Vakoc. vakoc@cshl.edu.

Author contributions X.S.W. and C.R.V. conceived this project and wrote the manuscript with input from all of the authors. X.S.W. and C.R.V. designed the experiments. X.S.W. performed experiments with help from X.-Y.H., J.J.I., Y.-H.H., D.N. and Y.T.S.; J.J.I. performed size-exclusion column for C11orf53/OCA-T1 and POU2F3 protein purification, assisted in microscale thermophoresis data analysis, and performed and analysed all analytical gel-filtration experiments. Y.-H.H. set up the initial breeding cage for *C11orf53* WT and knockout mice, performed immunostaining of trachea, small intestine and thymus, and performed subcutaneous injection and tumour measurement of NCI-H526 xenografts with the assistance of X.-Y.H. and D.N.; X.-Y.H. performed all RNA-FISH, histology and immunostaining image analysis, assisted in or performed tissue collection for histology and immunostaining, and assisted in NCI-H1048 subcutaneous injection and tumour measurement and cell cycle arrest experiments. Y.T.S. sorted the YT330 cell line. J.B.P. assisted in scRNA-seq library preparation, initial mapping and designed the machine learning algorithm for cell type assignment. C.R.V., L.J.-T., M.E. and J.S. supervised the studies and acquired the fundings.

Competing interests C.R.V. has received consulting fees from Flare Therapeutics, Roivant Sciences and C4 Therapeutics; has served on the advisory boards of KSQ Therapeutics, Syros Pharmaceuticals and Treeline Biosciences; has received research funding from Boehringer-Ingelheim and Treeline Biosciences; and owns a stock option from Treeline Biosciences. M.E. is a member of the research advisory board for brensocatib for Insmed; a member of the scientific advisory board for Vividion Therapeutics; and a consultant for Protalix. J.S. has received research funding from Abbvie and Pfizer and licensed a patent (11046763 to Stanford University) to Forty Seven/Gilead on the use of CD47-blocking strategies in SCLC.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04842-7.

Peer review information *Nature* thanks Christoph Schneider, Dean Tantin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

factors and a DNA octamer motif in a bivalent manner. We demonstrate that binding between POU2F3 and OCA-T1 or OCA-T2 is essential in tuft-cell-like small-cell lung cancer. Moreover, we generated OCA-T1-deficient mice, which are viable but lack tuft cells in several mucosal tissues. These findings reveal that the POU2F3–OCA-T complex is the master regulator of tuft cell identity and a molecular vulnerability of tuft-cell-like small-cell lung cancer.

Tuft cells are solitary chemosensory cells that coordinate immune and neural functions within mucosal epithelial tissues[1]. This cell type possesses a unique transcriptome that includes taste receptors, ion channels, cytokines and neurotransmitters that together allow for paracrine regulation of type 2 immunity[5–7]. For example, intestinal tuft cells detect parasite-derived metabolites and respond by secreting IL-25 and leukotrienes to activate innate lymphoid cells[8–10]. Recent studies identified a tuft-cell-like variant of small-cell lung cancer (also known as SCLC-P), which has a similar transcriptome and genetic dependencies to normal tuft cells[2,11]. An important requirement for both normal and malignant tuft cell development is the class II POU domain transcription factor POU2F3 (also known as OCT11 and SKN-1a)[2,5,12,13]. Owing to the sparsity of tuft cells in epithelial tissues, the biochemical mechanisms used by POU2F3 to drive tuft cell development are largely unknown.

Using single-cell RNA-sequencing (scRNA-seq) data from human and mouse tissues, we identified a previously undescribed gene, *C11orf53* (long isoform; also known as *1810046K07Rik* in mouse), as being selectively expressed in tuft cells of the small intestine, trachea, thymus and colon in a pattern resembling *POU2F3* (refs. [9,14–20]) (Fig. 1a,b and Extended Data Fig. 1a–g). We confirmed this co-expression pattern in mouse tissues using RNA fluorescence in situ hybridization (RNA-FISH) (Extended Data Fig. 2). Across more than 1,000 human cancer cell lines, we found that *C11orf53* (long isoform) is selectively co-expressed with *POU2F3* in SCLC-P cells, a finding that we validated by western blotting[21] (Fig. 1c,d, Extended Data Fig. 3a,b and Supplementary Fig. 1a). Co-expression of *C11orf53* and *POU2F3* was observed in a subset of primary human SCLC-P tumours, but expression of both genes was low or undetectable in other molecular subtypes of SCLC[22] (Fig. 1e). Collectively, these observations suggest that *C11orf53* expression is a conserved marker of normal and malignant tuft cells.

The protein product of *C11orf53* is predicted to be intrinsically disordered and lacks any known structured domains (Extended Data Fig. 4a). However, we noticed that a 22 amino acid segment near the N terminus is conserved across species (Extended Data Fig. 4b). PSI-BLAST analysis revealed that this peptide bears homology to OCA-B (encoded by *POU2AF1*), which is a B-cell-specific coactivator of class II POU transcription factors POU2F1 (also known as OCT1) and POU2F2 (also known as OCT2)[23] (Fig. 2a). This analysis revealed another uncharacterized protein, COLCA2, that similarly harbours a conserved OCA-B-like peptide (Fig. 2a and Extended Data Fig. 4c,d). In the transcriptomic datasets described above, we found that *COLCA2* (long isoform; also known as *Gm684* in mouse) is expressed in human and mouse tuft cells, albeit at lower levels and with less specificity than *C11orf53* (refs.[9,14–21,24]) (Extended Data Fig. 1a–g). Exceptions to this pattern are human thymic tuft cells, which express *COLCA2* at comparable levels

to *C11orf53* (Extended Data Fig. 1d), and human bronchial epithelium, in which a high *COLCA2* mRNA level instead of *C11orf53* is detected in tuft cells[15,24] (Extended Data Fig. 1e). *COLCA2* is also co-expressed with *POU2F3* in a subset of human SCLC-P tumours and cell lines in a mutually exclusive pattern to *C11orf53* (refs. [15,24]) (Fig. 1c–e, Extended Data Figs. 1h and 3c and Supplementary Fig. 1a). Notably, the conserved amino acids present on C11orf53 and COLCA2 precisely match the sites of OCA-B that engage in direct physical contacts with POU2F1 or to the octamer DNA element in an existing crystal structure of this ternary complex[25–27] (Fig. 2a and Extended Data Fig. 4g). Moreover, the genes that encode C11orf53, COLCA2, and OCA-B exist in a cluster in both the mouse and human genomes (Fig. 2b and Extended Data Fig. 4e), suggesting that these three proteins have a common evolutionary origin. Considering that the OCA-B interaction surface of POU2F1 is conserved on POU2F3 (Extended Data Fig. 4f,g), we hypothesized that C11orf53 and COLCA2 are previously overlooked paralogues of OCA-B that support the function of POU2F3 in tuft cells.

Using immunoprecipitation and GST pull-down assays performed in nuclear extracts prepared from SCLC-P lines, we found that C11orf53 and COLCA2 associate with POU2F3 (Fig. 2c,d, Extended Data Fig. 5a and Supplementary Fig. 1b). Guided by the existing OCA-B crystal structure[25], we generated C11orf53 and COLCA2 mutations predicted to disrupt binding to POU2F3 (C11orf53(V22E) and COLCA2(V17E)) or to octamer-motif DNA (C11orf53(V16D) and COLCA2(V11D)) (Fig. 2a, Extended Data Fig. 4g and Supplementary Fig. 1b). In contrast to the wild-type (WT) proteins, each of these mutations was unable to bind to POU2F3 (Fig. 2c,d, Extended Data Fig. 5a and Supplementary Fig. 1b). DNA pull-down assays revealed that C11orf53 bound to POU2F3 in the presence of octamer motif A (ATGCAAAT), but was unable to do so with octamer motif T (ATGCTAAT) (Fig. 2e and Supplementary Fig. 1c). The requirement for adenine as the fifth base of the octamer motif is consistent with previous studies of OCA-B[26,28]. We next reconstituted a stable C11orf53–POU2F3–DNA complex using purified full-length proteins (Fig. 2f and Supplementary Fig. 1d). Using analytical gel filtration and microscale thermophoresis assays, we confirmed that C11orf53 and POU2F3 can form a stable complex only when DNA is present, and this binding interaction occurred with a binding constant (K_D) of 98 nM (Fig. 2g,h, Extended Data Fig. 5d,e, Supplementary Fig. 1d and Supplementary Table 1). By contrast, POU2F3 could bind to DNA independently of C11orf53 in both assays (Fig. 2g, Extended Data Fig. 5e,f and Supplementary Table 1). Consistent with a conserved mode of binding, we found that C11orf53, COLCA2 and OCA-B could each bind to any of the three class II POU transcription factors, but not to other POU transcription factors (Extended Data Fig. 5b,c,g–i and Supplementary Fig. 1e,f). Similar to OCA-B, the C-terminal regions of C11orf53 and COLCA2 have trans-activation functions when tethered to a heterologous promoter, and their trans-activation abilities were significantly more potent than the ability of the trans-activation domains of POU2F3 (ref. [29]) (Fig. 2i, Extended Data Fig. 5j, Supplementary Fig. 1g and Supplementary Table 1). Collectively, these experiments validate C11orf53 and COLCA2 as bona fide OCA-B paralogues that form a DNA-dependent coactivator complexes with POU2F3.

We next performed chromatin immunoprecipitation (ChIP) analysis of C11orf53, COLCA2 and POU2F3 in SCLC-P cell lines to compare the genome-wide binding sites of these

proteins. By inspecting individual tuft -cell-specific genes (such as *TRPM5*, *ASCL2* and *AVIL*) and by performing a genome-wide analysis, we observed a close correlation between POU2F3 and C11orf53 in three independent SCLC-P lines (Fig. 3a,b and Extended Data Fig. 6a). These shared binding sites were almost exclusively found at distal elements containing ATGCAAAT motifs and were enriched for active enhancer binding proteins p300, MED1 and BRD4 (Fig. 3a and Extended Data Fig. 6b,c). By expressing HA-tagged proteins in NCI-H211 cells, we found that WT C11orf53 occupied tuft-cell-specific enhancers, whereas the V16D and V22E mutations abolished this interaction (Fig. 3c, Extended Data Fig. 6d and Supplementary Fig. 1h). Moreover, our epigenomic analysis in NCI-H1048 cells—a SCLC-P cell line that expresses COLCA2 rather than C11orf53 (Fig. 1c,d, Extended Data Figs. 1h and 3c and Supplementary Fig. 1a)—revealed that COLCA2 genomic occupancy also overlapped with POU2F3 (Extended Data Fig. 6a). Results from sequential ChIP experiments also support that POU2F3 and C11orf53 or COLCA2 co-occupy the same chromatin fragments (Extended Data Fig. 6e).

To evaluate the functional link between C11orf53/COLCA2 and POU2F3, we performed early-time-point RNA-seq analysis after genetic knockout of each factor in SCLC-P lines. These experiments revealed a strong correlation between the transcriptional changes incurred in C11orf53-, COLCA2- and POU2F3-deficient cells, including down-regulation of tuft cell lineage markers (Fig. 3d, Extended Data Fig. 6f,g, Supplementary Fig. 1h and Supplementary Table 2). One caveat to this experiment is that C11orf53 and POU2F3 maintain each other's expression in an apparent positive-feedback loop, which may contribute to the observed correlation in these data (Fig. 3d, Extended Data Fig. 6f,g, Supplementary Fig. 1h and Supplementary Table 2). To overcome this issue, we performed a gain-of-function experiment in which we co-expressed *C11orf53* and *POU2F3* cDNAs in an undifferentiated mouse SCLC cell line[30] (Extended Data Fig. 6h and Supplementary Fig. 1i). Using RNA-seq, we found that co-expression of *C11orf53* and *POU2F3* led to a robust activation of tuft-cell-specific genes (for example, *Trpm5*, *Ascl2* and *Avil*), whereas no change in expression was observed after expressing each factor individually (Fig. 3e and Supplementary Table 2). Taken together, these transcriptomic analyses suggest that C11orf53 and COLCA2 can each cooperate with POU2F3 to activate the transcription of tuft-cell-specific genes.

Using data produced by the Dependency Map Project that profiled gene essentiality in more than 900 human cancer cell lines[21], we found that POU2F3, C11orf53 and COLCA2 are required for the growth of SCLC-P cell lines but are dispensable for the growth of all other cancer types (Extended Data Fig. 7a,b). Consistent with its expression pattern (Fig. 1c,d, Extended Data Figs. 1h and 3c and Supplementary Fig. 1a), the NCI-H1048 line required COLCA2, whereas the other SCLC-P lines required C11orf53 (Extended Data Fig. 7b). We used CRISPR–Cas9 to inactivate POU2F3, C11orf53 and COLCA2 in a diverse panel of SCLC cell lines and validated the selective requirement for all three factors in SCLC-P to support cell proliferation (Fig. 4a, Extended Data Fig. 7c,d, Supplementary Fig. 2 and Supplementary Table 3). Furthermore, we validated the dependency of SCLC-P lines on POU2F3, C11orf53 or COLCA2 in vivo by inactivating each gene in NCI-H211 and/or NCI-H1048 cell line xenografts (Fig. 4b, Extended Data Fig. 7e,f and Supplementary Table 4). Using a cDNA rescue assay of the knockout phenotype, we found that point mutations that

disrupt binding to POU2F3 or deletion of the trans-activation domain rendered C11orf53 or COLCA2 incapable of promoting SCLC-P proliferation (Fig. 4c, Extended Data Figs. 6d and 7g–i, Supplementary Fig. 1i,j and Supplementary Table 3). In these cDNA rescue experiments, we found that C11orf53 and COLCA2 can substitute for one another's essential function in SCLC-P (Extended Data Figs. 6d and 7h,i, Supplementary Fig. 1i,j and Supplementary Table 3). However, expression of OCA-B could not substitute for the essential function of C11orf53 in SCLC-P lines, which is due to intrinsic differences between the C-terminal trans-activation domains (Fig. 4c, Extended Data Figs. 6d and 7h,i, Supplementary Fig. 1i,j and Supplementary Table 3). These genetic experiments suggest that an interaction between POU2F3 and C11orf53 or COLCA2 is essential in SCLC-P.

To evaluate the role of C11orf53 in normal mammalian development, we used CRISPR–Cas9 to generate mice containing a four-nucleotide deletion of the start codon of *C11orf53* (Fig. 4d). Breeding experiments revealed that *C11orf53*$^{-/-}$ and *C11orf53*$^{+/-}$ mice were born at normal Mendelian ratios, and had indistinguishable weight, morphology, fertility and organ histology from *C11orf53*$^{+/+}$ mice (Fig. 4e, Extended Data Fig. 8 and Supplementary Table 5). However, immunofluorescence staining of DCLK1 and POU2F3 markers revealed no detectable tuft cells in the trachea, small intestine, urethra, gall bladder, tongue and nasal epithelium of *C11orf53*$^{-/-}$ mice, whereas a partial loss of tuft cells was observed in the colon (Fig. 4f–h, Extended Data Fig. 9, Supplementary Table 5 and Supplementary Videos 1 and 2). We further validated the absence of tuft cells and the preservation of other lineages in the small intestine of *C11orf53*$^{-/-}$ mice using scRNA-seq (Extended Data Fig. 10a,b). Moreover, injection of *C11orf53*$^{-/-}$ mice with recombinant IL-25, a cytokine that induces type 2 immune response and tuft cell expansion[5], was not able to rescue the tuft cell deficiency (Extended Data Fig. 10c–f and Supplementary Table 5). Gastric and thymic tuft cells remained present in *C11orf53*$^{-/-}$ mice, which could be explained by high expression of *COLCA2* in tuft cells from these tissues (Fig. 4h, Extended Data Figs. 1a, 2c and 9 and Supplementary Table 5). The selective absence of tuft cells in *C11orf53*$^{-/-}$ mice bears resemblance to *Pou2f3*$^{-/-}$ mice[5,12,13,31], suggesting that the function of C11orf53 is to support normal tuft cell development as a POU2F3 coactivator.

Since its discovery over 30 years ago as a lineage-specific transcriptional coactivator, OCA-B has been widely viewed as the only protein of its kind encoded in the human genome[23]. Here we have identified two previously overlooked OCA-B paralogues that are likely to have arisen during evolution through gene duplications. Like OCA-B, these proteins possess a conserved peptide for bivalent binding to class II POU domain-containing transcription factors and to octamer motif DNA. Moreover, all three paralogues possess a C-terminal trans-activation domain. Although these three proteins are biochemically similar, their distinct expression patterns and trans-activation domains are likely to confer specialized transcriptional outputs through mechanisms that remain unclear. On the basis of the paralogue relationship to OCA-B, we propose renaming the protein C11orf53 to OCA-T1 and COLCA2 to OCA-T2 (Oct co-activator from tuft cells 1 and 2). We propose renaming the gene *C11orf53* to *POU2AF2* (POU class 2 homeobox associating factor 2) and *COLCA2* to *POU2AF3*. Although additional experiments will be needed to clarify the function of OCA-T2 in vivo, we hypothesize that binding of POU2F3 to OCA-T1 versus OCA-T2 leads to distinct transcriptional effects, which may contribute to tuft cell

heterogeneity in various tissue contexts. As POU2F1 is ubiquitously expressed and also has the ability to bind to OCA-T proteins, this transcriptional complex might also have a role in normal and malignant tuft cell biology. Several of the findings in our study are in agreement with a preprint article[32].

Cellular reprogramming experiments have shown that POU domain transcription factors are necessary but insufficient to carry out lineage-specifying transcriptional functions[33,34]. In the case of POU5F1/OCT4 in embryonic stem cells, cooperativity with SOX2 at composite octamer–sox elements is critical for enhancer-mediated gene activation and pluripotency[35]. Our studies suggest that class II POU domain transcription factors uniquely rely on OCA coactivators to achieve their lineage-defining functions in the B cell and tuft cell lineages. Despite having two activation domains and a high-affinity DNA-binding domain, our functional experiments indicate that POU2F3 cannot specify tuft cell identity without binding to one of the OCA-T proteins. It is probable that OCA-T binding endows POU2F3 with a critical trans-activation domain, which may allow for increased binding affinity for the general transcriptional machinery to support enhancer-mediated gene activation[35–38]. In summary, our study defines a specialized transcription factor–coactivator interaction that is required for the development of normal and malignant tuft cells. The mechanisms defined here may have therapeutic significance, as pharmacological blockade of the POU2F3 interaction with OCA-T1 and OCA-T2 would be expected to selectively suppress SCLC-P tumours without having a significant effect on normal tissue homeostasis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04842-7.

Methods

Cell lines and cell culture

NCI-H211, NCI-H526, COR-L311, NCI-H69, DMS-79, NCI-H82, HTB-184, NCI-H524, DMS 114 and YT330 cell lines (SCLC) were cultured in RPMI supplemented with 10% FBS. NCI-H446 (SCLC) were cultured in RPMI supplemented with 20% FBS. HEK293T cells were cultured in DMEM with 10% FBS. NCI-H1048, NCI-H209, NCI-H1436, NCI-H1836 and replicates of YT330 reprogrammed (SCLC) cells were cultured in HITES medium, which is composed of DMEM:F12 supplemented with 0.005 mg ml^{-1} insulin, 0.01 mg ml^{-1} transferrin, 30 nM sodium selenite, 10 nM hydrocortisone, 10 nM β-estradiol, 4.5 mM L-glutamine and 5% FBS. Penicillin–streptomycin was added to all media. All cell lines were cultured at 37 °C with 5% CO_2 and were periodically tested mycoplasma negative. COR-L311 was obtained from Sigma-Aldrich, and other human SCLC cell lines were purchased from ATCC. For NCI-H211, NCI-H526, COR-L311 and NCI-H1048, cells were further validated by STR profiling at external facility (Genetics core, University of Arizona) at the end of the study to ensure cell identity.

Sorting of the YT330 cell line

The mouse neuroendocrine YT330 cell line used for the reprogramming experiment was sorted out from Rb/p53/p130 triple KO (TKO) *Hes1*$^{GFP/+}$ mice. In brief, tumours from the lungs of TKO *Hes1*$^{GFP/+}$ mice (6 months after tumour induction) were isolated, pooled and finely chopped with a razor blade. They were then digested in 6 ml of PBS with 120 μl of 100 mg ml^{-1} collagenase/dispase (Roche) for 45 min with shaking at 37 °C then cooled on ice before adding 15 μl of 1 mg ml^{-1} DNase (Sigma-Aldrich) for 5 min. The digested mixture was filtered through a 40 μm filter, pelleted and resuspended in 1 ml of red blood cell lysis buffer (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA) for 90 s. Cells were washed once in DMEM and resuspended in FACS buffer (10% BGS in PBS, 1 million cells per 100 μl). The final single-cell suspension was stained with FACS antibodies. For negative lineage selection antibodies against CD45-PE-Cy7 (eBioscience 25–0451-82, 30-F11, 1:100), CD31-PE-Cy7 (eBioscience 25–0311-82, 390, 1:100), TER-119-PE-Cy7 (eBioscience 25–5921-82, TER-119, 1:100) were used. For positive lineage selection, antibodies against CD24-APC (eBioscience, 17–0242-82, M1/69, 1:200) were used. DAPI staining was used to identify dead cells. Cells were kept in RPMI medium for culturing.

Analysis of publicly available expression datasets

For all public scRNA-seq analysis, the gene expression matrices along with sample and cluster annotations were either obtained from the indicated studies or processed using Scanpy (v.1.7.2)[39]. Samples were concatenated together using anndata (v.0.7.6), processed and plotted with Scanpy (v.1.7.2) in Python (v.3.8.8)[39]. The mRNA level was normalized by subtracting the minimum and then dividing by its maximum to calculate the mean expression in the group. The mouse thymus and human thymus datasets were obtained from ref. [15] (6,811 mouse stromal cells with 308 tuft cells detected, 255,901 human thymus cells with 44 tuft cells in dataset), and the human small intestine dataset was obtained from ref. [14] (22,502 cells with 33 tuft cells detected), the human colon mucosa dataset was obtained from ref. [20] (34,772 cells with 228 tuft cells), the human bronchi dataset was obtained from ref. [24] (36,248 cells with 92 tuft-like cells detected). For Extended Data Fig. 1f, the dataset was obtained from the Tabular Muris Consortium, in which a scRNA-seq analysis of 20 mouse organs was performed. The data were processed using scanpy.pl.dotplot with the cell type annotation provided by each study to calculate the fraction of cells expressed in each group.

For the cancer cell line expression and dependency analysis, the corresponding matrix was downloaded from Cancer Dependency Map (DepMap) portal with version 21Q2 (refs. [21,40]). All the data were plotted with ggplot2 (v.3.3.3) in R (v.4.0.5). For two-class comparison analysis of cancer dependency in tuft cells versus other cancer cell lines, ashr with empirical Bayes moderated *t*-statistics was used.

For the SCLC patient transcriptome analysis, paired-end raw sequencing results were obtained from ref. [22]. The samples were then pseudoaligned to the reference human transcriptome (Gencode v.35), followed by TPM calculation for genes using Salmon (v.1.0.0) with the default settings[41].

For isoform expression of mouse *C11orf53* and *Colca2*, raw data were obtained from ref. [9] and pseudoaligned to the reference mouse transcriptome (GRCm38.p6) for transcript quantification using Salmon (v.1.0.0) with the default settings[41].

Western blot

Cell pellets (1 million cells) were washed three times with 1× PBS, resuspended in 200 µl Laemmli sample buffer (Bio-Rad) containing β-mercaptoethanol and boiled at 98 °C for 20 min, followed by centrifugation at 13,000*g* for 5 min. These whole-cell extracts were separated by SDS–PAGE (NuPAGE 4–12% Bis-Tris Protein gels, Thermo Fisher Scientific), followed by transfer to nitrocellulose membranes and immunoblotting. Antibodies used in this study included C11orf53/OCA-T1 (in-house generated, 1:200 (1 mg ml^{-1})), POU2F3 (in-house generated; and Sigma-Aldrich, HPA019652, 1:500 or 1:1,000), H3 (Abcam, ab18521, 1:10,000 or 1:50,000), GAL4-DBD (Santa Cruz, SC-510, 1:1,000), HRP-conjugated secondary antibodies (rabbit Cytiva/Amersham, NA934; mouse, Agilent/Dako, P026002–2, 1:10,000), HRP-conjugated β-actin (Sigma-Aldrich, A3854, 1:10,000), HA (3F10, Roche, 12013819001, 1:1,000), and FLAG (Sigma-Aldrich, A8592, 1:1,000–1:5,000). For the GAL4 luciferase construct, HEK293T cells that were transfected with the plasmid of interest were collected 48 h after infection and washed twice with 1× PBS. For the sgRNA knockout experiments, NCI-H211 cells were collected 4 days after infection. For the overexpression experiments in YT330 cells, samples were collected 18 days after infection.

Antibody generation

Peptides GDPAHFLFRDSWEQTLPD and EADTGSLHDPSPWVKEDGS of C11orf53/OCA-T1 were purchased from GenScript with KLH conjugation. A mixture of both peptides (4 mg each) was used for rabbit immunizations for two rabbits with a 73 day rabbit antibody production protocol from Pocono Rabbit Farm & Laboratory. The whole-blood serum of exsanguination was used to validate the efficiency of antibodies for western blot and immunoprecipitation with HEK293T HA–OCA-T1 overexpression cell lysates. The whole-blood serum was then aliquoted in 10 ml and stored at −80 °C for long-term storage. An aliquot of 10 ml blood serum was then purified with columns containing the mixture of both antigens using the Thermo Fisher Scientific SulfoLink Immobilization Kit for Peptides and eluted 2 ml with IgG elution buffer (Thermo Fisher Scientific, 44999), neutralized with 200 µl 1 M Tris (pH 7.5) and stored at −20 °C with 50% glycerol. This purified antibody was used for western blot and chromatin immunoprecipitation assays.

For POU2F3, the same procedures were followed, except a mixture of peptides NSRPSSPGSGLHASSPTC, ASQNNSKAAMNPSSAAFNC and SSGSWYRWNHPAYLHC was used as antigens for rabbit immunization.

Protein conservation, BLAST, similarity and disorder analysis

For protein conservation analysis, the protein sequence for C11orf53/OCA-T1 (NCBI: NM_198498.3 with NP_940900.2) or COLCA2/OCA-T2 (NCBI: NM_001271458.2 with NP_001258387.1) was submitted to the ConSurf server with using the HMMER search algorithm using the default parameters[42,43]. For protein homology

searching, the conserved peptide sequence of C11orf53 from the above analysis (KRVYQGVRVKHTVKDLLAEKRSG) was submitted for PSI-BLAST analysis on NCBI using the refseq protein database (NCBI Protein reference sequences) in *Homo sapiens* with the default parameters[44,45]. For protein disorder prediction, the protein sequences were submitted to the IUPred2A and PONDR web portals for analysis[44,46]. For protein alignment and similarity calculations, Clustal Omega from EMBL-EBI was used with the default parameters. All data were plotted with matplotlib (v.3.4.2) and the seaborn (v.0.11.1) package with Python (v.3.8.8).

Structural analysis

Protein structure prediction for C11orf53/OCA-T1, COLCA2/OCA-T2 and POU2F3 were performed independently using SWISS-MODEL[47]. Superpositions of the predicted C11orf53, COLCA2 and POU2F3 structure to OCA-B with POU2F1 (PDB: 1CQT) and molecular graphics preparation were performed with PyMOL (v.2.3.2). The overlayed structure was then used to infer residues involved in interaction with DNA and POU2F3 for Fig. 2a and Extended Data Fig. 4g.

Plasmid construction and sgRNA cloning

The sgRNA lentiviral expression vector with optimized sgRNA scaffold backbone (LRG2.1T, Addgene, 108098) and the lentiviral Cas9 vector (Addgene, 108100). *POU2F3*, *C11orf53*, *COLCA2* and OCA-B/*POU2AF1* cDNA were cloned into the lentiV_P2A_Neo (Addgene, 108101) vector using the In-Fusion cloning system (Takara). A 3×FLAG tag was added to POU2F3 at the N terminus, and a 3×HA tag was added to C11orf53, COLCA2 and OCA-B at their C termini with an SV40 nuclear localization signal. Construction of POU2F3, C11orf53 and COLCA2 CRISPR-resistant synonymous mutants and loss-of-function mutants were cloned using the Phusion Flash master mix (Thermo Fisher Scientific) and the In-Fusion cloning system. For POU3F4, POU4F3, POU5F1, POU2F2 and POU2F1, plasmids were cloned into the pcDNA3.1 vector with a 3× FLAG tag at the N terminus. For co-expression in YT330 cells, POU2F3 and C11orf53 were co-expressed with a P2A linker in between. For co-expression of POU2F3 and COLCA2 in NCI-H1048 cells, 3×HA-tagged COLCA2 (C terminus) was cloned into the lentiV_IRES_BSD vector (modified from lentiV_P2A_Neo) to enable co-selection of POU2F3- and COLCA2-positive cells.

Lentiviral transduction

Lentivirus was produced in HEK293T cells by transfecting plasmids with helper plasmids (VSVG and psPAX2 (Addgene, 12260)) using polyethylenimine (PEI 25000). In brief, for one 10 cm dish of HEK293T cells, 10 μg of plasmid DNA, 5 μg of VSVG and 7.5 μg of psPAX2 were mixed in 500 μl Opti-MEM (mix A). In a separate tube, 90 μl of 1 mg ml^{-1} PEI was added into 500 μl Opti-MEM, and then mixed with mix A by vortexing. After incubation for 15 min at room temperature, the DNA and PEI mixture was added dropwise onto HEK293T cells. Medium was exchanged 6–8 h after transfection. The lentivirus-containing supernatant was collected at 48 h and 96 h after transfection, pooled, then centrifuged at 1,200 rpm for 5 min. Virus-containing supernatants were then filtered through 0.45 μm SFCA filters (Corning). For lentivirus infection, target cells were

mixed with the virus and 4 μg ml^{-1} polybrene, then centrifuged at 1,700 rpm for 30 min in 6-well plates or 24-well plates. Medium was exchanged at 24 h after infection, and the corresponding antibiotics (1 μg ml^{-1} puromycin, 10 μg ml^{-1} blasticidin, 1 mg ml^{-1} G418 sulfate) were added at 48 h after infection if selection was needed.

Immunoprecipitation analysis

For immunoprecipitaiton (IP) analysis of SCLC-P nuclear extracts, 80 million cells stably expressing the desired construct were collected by centrifugation, washed three times with PBS and resuspended in 1 ml buffer A (20 mM HEPES (pH 8.0), 10 mM KCl, 300 mM sucrose, 0.1% NP-40, 10% glycerol), collected by centrifugation, then similarly washed in buffer A followed by 5 min of centrifugation at 600*g*. The nuclear pellet was then resuspended in A60 buffer (50 mM Tris (pH 7.5), 60 mM NaCl, 0.5% NP-40, 10% glycerol, 1.5 mM $MgCl_2$, protease inhibitors) followed by 30 min rotation at 4 °C. Samples were cleared by centrifugation at the maximum speed in a tabletop centrifuge; the resulting supernatant (fraction 1) was kept cold. Insoluble chromatin pellets were digested with 1 μl MNase in 200 μl A60 buffer with 1 mM $CaCl_2$ at 37 °C for 10 min and stopped with 2 mM EGTA and 5 mM EDTA followed by centrifugation at maximum speed in a tabletop centrifuge for 20 min. This supernatant was mixed with fraction 1 as nuclear extract. Anti-HA magnetic beads (Thermo Fisher Scientific) were equilibrated twice in PBS and once in A60 buffer. 2% input was taken for each sample. Extracts were incubated with 30 μl equilibrated anti-HA beads at 4 °C overnight with rotation. The next day, beads were washed three times with A60 buffer and proteins were eluted in 200 μl SDS Laemmli buffer with β-mercaptoethanol at 98 °C for 15 min followed by western blotting.

For IP analysis of HEK293T cells, one 10 cm dish of HEK293T cells was transfected 8 μg of each HA-tagged cDNA and FLAG-tagged cDNA with 64 μl of 1 mg ml^{-1} PEI 25000 in 1 ml Opti-MEM. For COLCA2, given that its expression is lower, 12 μg of HA-tagged cDNA was used. Then, 48 h later, HEK293T cells were collected and washed twice with PBS, followed by resuspension in 1 ml A150 (20 mM Tris (pH 8.0), 150 mM NaCl, 1% Triton X-100, 1.5 mM $MgCl_2$) and moved forward for IP as described above with A150 buffer.

DNA pull-down assay

Single-stranded DNA oligos were synthesized by Integrated DNA Technologies with 5′ biotin modification. Double-stranded DNA probes for the pull-down assays were generated by mixing complementary oligonucleotides (a biotinylated forward-strand/non-biotinylated reverse-strand pair) in a 1:1 ratio, then heating at 95 °C for 5 min before annealing by cooling to 25 °C at 5 °C min^{-1} in a thermocycler. For each reaction, 25 μl of Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific, 65602) were washed with the binding buffer (20 mM Tris (pH 7.5), 1 M NaCl, 1 mM EDTA) followed by two washes with PBS. DNA probe (5 pmol) was incubated with 25 μl washed MyOne T1 beads (Thermo Fisher Scientific, 65001) in 200 μl of binding buffer at 4 °C with rotation for 1 h, followed by two washes with binding buffer, one wash with PBS and two washes with NTN150 buffer (20 mM Tris (pH 7.5), 150 mM NaCl, 0.5% NP-40, protease inhibitors).

One 15 cm dish of HEK293T cells was transfected with 15 μg 3×HA-tagged C11orf53 (pcDNA 3.1) and/or 10 μg HA tagged POU2F3 (pcDNA 3.1) with polyethylenimine (PEI 25000). Then, 48 h after transfection, cells were collected and washed three times with 1× PBS and lysed with 1 ml NTN150 buffer for 30 min at 4 °C with rotation. The samples were then cleared by centrifugation at maximum speed in a tabletop centrifuge for 30 min at 4 °C. 2% input was taken. Supernatants were then aliquoted equally into three tubes and incubated with biotinylated DNA containing T1 beads at 4 °C with rotation. After 2 h incubation, protein-bound T1 beads were washed twice with the NTN150 buffer and once with PBS. Proteins were then eluted from beads by heating at 98 °C for 20 min in 200 μl Laemmli sample buffer with β-mercaptoethanol. After 5 min of centrifugation at maximum speed in a tabletop centrifuge, the samples were analysed using western blotting.

Purification of recombinant protein

For POU2F3 and POU2F1 full-length purification, His_6–GFP-tagged POU2F3 or POU2F1 constructs in pHIS-Parallel-GFP were transformed into BL21-CodonPlus (DE3)-RIPL (Agilent) for large-scale expression using standard methods. In brief, cultures were grown in Luria Broth supplemented with the appropriate antibiotic(s) at 37 °C to a culture density with an optical density at 600 nm (OD_{600}) of 1.0. The cultures were then cooled at 4 °C for 1 h followed by induction of protein expression with 1 mM IPTG. Induction proceeded overnight at 16 °C with shaking at 180 rpm. Cells were collected by centrifugation at 2,600*g* for 10 min at 4 °C. The supernatant was discarded, and the pellets were taken for protein purification. Cell pellets were resuspended in 20 ml of lysis buffer (50 mM sodium phosphate (pH 8.0), 500 mM NaCl and 10 mM imidazole, supplemented with protease inhibitors) per litre of culture. The suspension was lysed by sonication (2 min 30 s at 40% amplitude, 2 sec on, 2 sec off) then clarified by ultracentrifugation at 13,000*g* for 45 min at 4 °C. The soluble supernatant was next taken for affinity-column purification with 500 μl Ni-NTA resin (Qiagen), pre-equilibrated with lysis buffer. After loading, the column was washed with wash buffer (50 mM sodium phosphate (pH 8.0), 1 M NaCl and 30 mM imidazole, supplemented with protease inhibitors). The target protein was eluted in elution buffer (50 mM sodium phosphate (pH 8.0), 1 M NaCl and 200 mM imidazole). Immediately after elution, 1 mM dithiothreitol and 1 mM EDTA were added to limit aggregation and degradation, respectively. Further purification was performed by gel filtration using a Superdex 200 increase 10/300 column (Cytiva/GE Healthcare Life Sciences). The protein was chromatographed over ~30 ml at a flow rate of 0.6 ml min^{-1} in a running buffer of PBS. Peak fractions were assessed using SDS–PAGE. Fractions with highly purified protein were concentrated, then taken for binding assays. Typical yields were 1–2 mg of highly purified protein (>98% pure as assessed using SDS–PAGE) per litre of culture.

For full-length C11orf53/OCA-T1 or POU2AF1/OCA-B, $Strep_2$SUMO-tagged C11orf53, or OCA-B constructs were cloned into the vector pFL then integrated into bacmids using DH10MultiBac cells (Geneva Biotech). Isolated bacmids were then transfected into *Sf9* cells for baculoviral-driven expression in CCM3 medium (HyClone). After expression, cells were collected by centrifugation at 1,000*g* for 10 min at 4 °C, then lysed by sonication in lysis buffer (50 mM Tris (pH 8.0), 100 mM KCl and 1 mM dithiothreitol with protease

inhibitor). Cell lysates were clarified by centrifugation at 4 °C for 45 min at 13,000*g*. The supernatant was then applied to a Strep-Tactin (IBA) column equilibrated with lysis buffer, followed by one wash with 20 ml lysis buffer, one wash with 5 ml lysis buffer with 2 mM ATP, and two washes with lysis buffer. Protein was then eluted in lysis buffer supplemented with 50 mM D-desthiobiotin. Further purification was performed by gel filtration using a Superdex 200 increase 10/300 column (Cytiva/GE Healthcare Life Sciences). The protein was chromatographed over ~30 ml at a flow rate of 0.6 ml min^{-1} in a running buffer of PBS. Peak fractions were assessed by SDS–PAGE. Fractions with highly purified protein were concentrated, then taken for binding assays. Typical yields were 1–2 mg of highly purified protein (>98% pure as assessed by SDS–PAGE) per litre of culture.

For purification of GST, and GST-tagged COLCA2/OCA-T2(1–49) and C11orf53/OCA-T1(1–52) variants, plasmids were transformed into BL21-CodonPlus (DE3)-RIPL (Agilent) for large-scale expression using standard methods. In brief, cultures were grown in Luria Broth supplemented with the appropriate antibiotic(s) at 37 °C to a culture density with an OD_{600} of approximately 0.6–0.8. Cells were then induced with 1 mM IPTG at 30 °C for 4 h for protein expression. Cells were collected by centrifugation at 2,600*g* for 10 min at 4 °C. The supernatant was discarded, and the pellets were taken for protein purification. For 1 litre of each culture, the pellet was resuspended in 35 ml PBS with protease inhibitor and lysed by sonication (2 min 30 s at 40% amplitude) then clarified by ultracentrifugation at 13,000*g* for 45 min at 4 °C. The soluble supernatant was next taken for affinity column purification with 1 ml of 75% glutathione resin (GE, 17075601), pre-equilibrated with PBS. After loading, the column was washed with three washes of 1× PBS. The target protein was eluted in 10 mM reduced L-glutathione (Sigma-Aldrich, G4251–25G).

Analytical gel filtration

A motif-A-containing 22 bp DNA oligonucleotide (2 μg) was mixed at an equimolar ratio with $Strep_2$SUMO–C11orf53, His_6–GFP–POU2F3, or both (as indicated) at a final sample volume of 20 μl, then incubated for approximately 1 h at 4 °C. The samples were injected onto a Superdex200 3.2/300 increase column (Cytiva/GE Healthcare Life Sciences) and chromatographed over ~3.5 ml at a flow rate of 0.05 ml min^{-1} in 20 mM Tris (pH 7.4), 50 mM NaCl with absorbance monitoring at 260 nm and 280 nm. Data presented in the figures correspond to the normalized measurements at 260 nm. Assessment of other complexes (containing selected combinations of $Strep_2$SUMO–C11orf53, $Strep_2$SUMO–OCA-B, His_6–GFP–POU2F1, His_6–GFP–POU2F3 and DNA, as indicated), were prepared similarly, although the molar ratio for His_6–GFP–POU2F1-containing mixtures was increased to 3:1 as not all the recombinantly purified His_6–GFP–POU2F1 was competent for binding. Formation of ternary complexes was verified by collection of peak fractions followed by SDS–PAGE analysis.

Microscale thermophoresis analysis

Binding of purified POU2F3 and/or C11orf53 to DNA was measured using a Monolith NT.115 Pico running MO Control version 1.6 (NanoTemper Technologies). Assays were conducted in 50 mM Tris (pH 8.0), 50 mM NaCl, 0.02% Tween-20. DNA probes were prepared as in DNA pull-down assays except modified with Alexa-647N in place of biotin.

For GFP, POU2F3, and/or C11orf53 binding to binding to DNA, 500 pM Alexa647N-labelled DNA was mixed with serial dilutions of C11orf53 or POU2F3 and loaded into microscale thermophoresis (MST) premium coated capillaries (NanoTemper Technologies). For C11orf53 binding to GFP-POU2F3 + DNA or GFP + DNA, 500 pM of DNA with 50 nM of GFP–POU2F3 or GFP were mixed with serial dilutions of C11orf53. For C11orf53 + POU2F3, equal molar amounts of C11orf53 and POU2F3 were pre-mixed then serially diluted prior to incubation with 500 pM DNA for MST. MST measurements were recorded at room temperature using 30% excitation power and medium MST power. Measurements were performed in triplicate. Determination of the binding constant (K_D) was performed using MO Affinity Analysis v.2.3. For plotting, data are presented as fraction bound; each data point represents the mean ± s.e.m. with the fit corresponding to the average fit of the three replicates.

GST pull-down assays

GST, GST–COLCA2$^{1–49}$ or GST–C11orf53$^{1–52}$ (2 μM) variants were immobilized to 25 μl of glutathione beads by incubating the protein of interest with beads in PBS at 4 °C for 1 h followed by two washes in PBS and one wash with A60 buffer. The NCI-H211 nuclear extract was prepared as described in the ‘Immunoprecipitation analysis’ section. After taking input samples, the nuclear extract from 60 million cells was incubated with the beads containing immobilized protein of interest and rotated at 4 °C for 4 h. The beads were then washed four times with 1 ml A60 buffer and eluted in 200 μl 2× Laemmli buffer followed by western blot alongside input samples.

Luciferase reporter assays

Plasmids encoding the GAL4 DNA-binding domain (DBD) fusions (modified from pFN26A (BIND) hRlucneo Flexi Vector, E1380; Promega) were co-transfected with pGL4.35[luc2P/9XGAL4UAS/Hygro] Vector (E1370, Promega) into HEK293T cells for 48 h in a 96-well plate (Corning) with 100 ng of each plasmid. Luciferase activity was measured using the Dual Luciferase Reporter Assay System (E1910; Promega) according to the manufacturer’s instructions. All data shown represent Firefly luciferase activity normalized to the internal Renilla luciferase activity, the latter of which was expressed through a constitutive promoter on the pFN26A plasmid. For each biological replicate, three technical replicates were averaged.

ChIP–seq and sequential ChIP sample preparation

For each ChIP, 20 million cells were used. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature with agitation and quenched with 0.125 M glycine for 5 min at room temperature. After washing twice with PBS, cells were incubated in 1 ml cell lysis buffer (10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40 with protease inhibitor) for 15 min on ice. Nuclei were isolated by centrifugation at 600*g* for 30 s, resuspended in 1 ml nuclear lysis buffer (50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS with protease inhibitor) and sonicated using a Bioruptor Pico (Diagenode) (30s on/off, 10 cycles). For NCI-H1048 cells, 20 million cells were incubated in 500 μl nuclear lysis buffer and processed for 10 cycles of sonication. In all cases, the chromatin was centrifuged at maximum speed in a tabletop centrifuge for 15 min at 4 °C. The supernatant was mixed

with 7 ml IP dilution buffer (20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 1% Triton X-100), and incubated with the 2.5 μg of the indicated antibody for 2 h. Protein A/G magnetic beads (25 μl, Dynabeads, Thermo Fisher Scientific) were washed twice with PBS and added to the antibody–chromatin mixture at 4 °C overnight. The beads were washed once with IP wash 1 buffer (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% Triton X-100, 0.1% SDS), twice with high salt buffer (20 mM Tris-HCl (pH 8.0), 2 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.01% SDS), once with IP wash 2 buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate) and twice with TE (pH 8.0). Chromatin DNA was eluted, and cross-linking was reversed in 200 μl nuclear extraction buffer with 12 μl of 5 M NaCl and 1 μg ml^{-1} RNase A at 65 °C overnight. The beads were then discarded. The DNA-containing supernatant was treated with 4 μg ml^{-1} proteinase K at 56 °C for 20 min and purified using the QIAquick PCR purification kit (QIAGEN) in 60 μl water.

For sequential ChIP, 30 million cells stably expressing HA–C11orf53 or HA–COLCA2 and FLAG–POU2F3 were collected, and regular ChIP was performed for each reaction, except that, after the first IP, only nuclear lysis buffer was added to the beads followed by incubation at 65 °C for 30 min twice. After two elutions, the samples were diluted with IP dilution buffer then processed as a typical ChIP. The first empty control was beads only or 3 μg of the M2-FLAG antibody (Sigma-Aldrich, F1804). For the second IP, 4 μg of Cell Signaling Technology mAb#3724 was used.

ChIP–qPCR was performed using SYBR green (ABI) on the ABI 7900HT system. The amplification efficiency of various primer sets was determined using an input standard curve dilution series of the pre-immunoprecipitated genomic DNA; only primer sets that amplified linearly according to the DNA amount were used for qPCR analysis. The IP signal for each sample was normalized to the input.

For ChIP-seq of H3K27Ac (Abcam, ab4729, 2 μg in total), 5 million cells were used; for POU2F3 ChIP–seq (Santa Cruz, SC-330; and two in-house-generated antibodies, 2 μg per IP), 3 ChIP samples were pooled before IP wash one; for C11orf53/OCA-T1 (Sigma-Aldrich, 12CA5, 11583816001 for lentivirally overexpressed HA–C11orf53 and HA-Cherry-NLS; two in-house generated antibodies for endogenous C11orf53), p300 (Active Motif, 61401), MED1 (Bethyl, A300–793), BRD4 (Behtyl A301–985), COLCA2/OCA-T2 (Sigma-Aldrich, 12CA5; and Cell Signaling Technology, mAb#3724 for lentivirally expressed HA–COLCA2), 5 ChIP reactions were combined before IP wash one and 2 μg antibody was used for each ChIP. ChIP-seq libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (E7645) according to the manufacturer's protocol with AMPure XP beads (Beckman, A63881) with no size selection. One extra amplification cycle was added to final PCR enrichment. The final PCR product was cleaned using AMPure XP beads twice. Library quality was assessed via Bioanalyzer equipped with the high-sensitivity DNA chip (Agilent) and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854). ChIP-seq libraries were pooled and analysed by single-end sequencing 76-bp sequencing using the NextSeq (Illumina) system.

ChIP-seq data analysis

Sequencing reads were first trimmed and filtered with Trimmomatic (0.39-Java-11) with a minimal read length of 50 using the default settings[48]. Reads that passed the quality control criteria were then mapped to the hg38 genome assembly using bowtie2 v.2.4.2 with the very-sensitive setting and allowing 2 mismatches[49]. Bam files were then sorted using Samtools v.1.11 with the default settings[49,50]. PCR duplicates were removed using Picard v.2.21.6 with the default settings and indexed using Samtools[50]. Peaks were called with Macs2 v.2.2.7.1 with input as negative control with default settings and a 5% FDR as cut-off. Bedtools v.2.30.0 was used to intersect as well as merge peaks from different experiments to generate high-confidence peaks or all C11orf53–POU2F3 occupying sites[51,52]. Homer v.4.11 makeTagDirectory and makeUCSCfile were used to generate bigwig files for visualizing ChIP-Seq tracks on the UCSC genome browser[53,54]. The annotatePeak function was used to annotate peak features for C11orf53–POU2F3 occupying peaks as well as calculating tag counts on C11orf53/POU2F3 intensity on all C11orf53–POU2F3 binding sites with ±250 bp around the peak centre. MEME SUITE v.5.4.1 MEME was used to discover motifs enriched on defined peak regions[55,56]. Plots were generated using customized scripts in Python with matplotlib and seaborn[57,58]. For NCI-H1048 cells, POU2F3 ChIP-seq data were obtained from ref. [2].

RNA-seq library preparation

For knockout experiments, cells were collected 5 days after infection with sgRNA for NCI-H211, or 6 days after infection with sgRNA for NCI-H526, COR-L311 and NCI-H1048 cells. Three independent sgRNAs were used for C11orf53/OCA-T1 and COLCA2/OCA-T2, and two independent sgRNAs were used for POU2F3 and negative controls. For overexpression experiments in YT330 cells, cells were collected 21 days after infection with cDNAs for three independent replicates. Total RNA was extracted using TRIzol (Thermo Fisher Scientific) according to the manufacturer's protocol and resuspended in RNase-free water. Poly(A) RNA was selected and fragmented with NEBNext Poly(A) mRNA Magnetic isolation module (NEB, E7350) and the library was prepared using the NEBNext Ultra II RNA Library Prep kit for Illumina (NEB, E7770) with 2 μg RNA according to the manufacturer's protocol. RNA-seq libraries were pooled and analysed by single-end 76 bp sequencing using the NextSeq (Illumina) system.

RNA-seq data analysis

Sequencing reads were mapped onto reference human genome hg38 using Hisat2 v.2.1.0 with the default settings[59]. Differentially expressed genes were analysed using counts from HTSeq-count v.0.13.5. Genes with low expression (maximum read counts in all conditions are less than 10) were filtered out. Expressed genes were then analysed by DESeq2 with masking structural RNAs and non-coding RNA for calculating fold change[60,61]. TPM was quantified using Salmon v.1.0.0 with the default settings[41].

Competition-based cell proliferation assays

Cas9-expressing human cell lines were infected with sgRNA linked with GFP (LRG2.1T, Addgene, 108098). The percentage of GFP^+ cells was measured every 4 days from day

4 to day 22 after infection using the Guava easyCyte Flow Cytometer (Millipore). Except for NCI-H1048 and DMS 114, all SCLC cells were trypsinized at 37 °C for 10 min with repetitive pipetting to break cell clumps before flow cytometry. For each passage, the GFP percentage was normalized by dividing the initial GFP percentage to calculate the fold change. The average fold change in two independent sgRNAs was calculated for sgNEG and sgPOU2F3. Three independent sgRNAs were used for calculation for sgC11orf53/OCA-T1 and sgCOLCA2/OCA-T2 among three biological replicates.

Generation of *C11orf53*-knockout mice

The 4-week-old C57BL6 female mice were treated with 5 international units (IU) of PMSG, followed by 5 IU hCG 48 h later. These female mice were then paired with C57BL6 males one-on-one immediately after the hCG treatment. One-cell zygotes were collected from these donor females with M2 medium. The zygotes were maintained at 37 °C in 5% CO_2 until microinjection with the appropriate CRISPR mix. The CRISPR reagent mix was freshly made on the morning of the day of injection. The CRISPR reagent mix contained 10 ng μl^{-1} sgRNA and 70 ng μl^{-1} Cas9 2NLS Nuclease. The sgRNA and Cas9 were both obtained from Synthego. The CRISPR reagent mix was centrifuged in a microcentrifuge at maximum speed for at least 10 min, and the supernatant was then transferred into a fresh 0.5 ml tube and kept on ice until injection.

Mouse zygote CRISPR injections were conducted in droplets of M2 medium covered with mineral oil on a glass slide at room temperature. This was conducted on a Nikon Ti inverted microscope with a DIC optical objective. The CRISPR reagent mix was then back-filled into a microinjection needle with a sterile glass pipette. About 50 mouse zygotes were transferred into one of the micro droplets of M2 medium on the glass slide under the microscope and the zygotes aligned into a line within the droplet. CRISPR reagent was introduced into the pronuclei of the zygotes one-by-one using CELLectro technology. Injected zygotes were transferred back into droplets of KSOM medium after injection and kept in an incubator at 37 °C in 5% CO_2 until embryo transfer. Surviving zygotes were transferred into recipients on the same day of the injection. CD-1 females of 2 months of age were used as transfer recipients. Standard embryo transfer was conducted according to standard animal use protocol. Surviving zygotes were transferred into oviducts on both sides. GTSR usually transfers 10 to 15 zygotes per oviduct and 20 to 30 zygotes per recipient. Recipients are monitored and cared for according to standard animal use protocol during and after the embryo transfer. Once pups are born, tail samples are collected within 7 days after birth. These samples are boiled in 200 μl of freshly-prepared tail lysis buffer (25 mM NaOH with 0.2 mM EDTA) at 95 °C for 20 min then neutralized with 200 μl of neutralization buffer (40 mM Tris, pH 5.0). The supernatant was prepared for PCR with primers for 40 cycles and sent for Sanger sequencing for genotyping.

SCLC line xenograft experiments

For xenograft experiments, Cas9-expressing NCI-H526 or NCI-H1048 cells were transduced with lentiviral sgRNA in LRG-2.1T-blast vectors. Cells were preselected with blasticidin for 48 h before injection. Then, 10 million cells were resuspended in 100 μl of 50% Matrigel (Corning, 356231) in PBS for each condition before subcutaneous injection into NSG mice

(NOD.Cg-Prkdc scid Il2rg tm1Wjl /SzJ, female, 6 weeks, 005557, Jackson Laboratory). Mice were monitored daily, and tumour size was measured using callipers twice weekly until the tumour in any group reached a maximum diameter of 2 cm, which is the maximum tumour size allowed. Tumour volume was calculated as (length × width^2)/2, which did not exceed 4 × 10^3 mm^3, the maximum tumour volume allowed, throughout the course of study. Experimental protocols involving mice were approved by the institutional animal care and use committees (IACUC) at Cold Spring Harbor Laboratory.

Tissue preparation for histology and immunofluorescence staining

Freshly isolated tissues were fixed in 4% of paraformaldehyde (PFA) at 4 °C overnight. Notably, freshly isolated intestines (including small intestine, colon and caecum) were first flushed three times with ice-cold PBS using a blunted 18G needle and 10 cm^3 syringe, then cut longitudinally and fixed in 4% PFA for 30 min at room temperature. Pre-fixed intestines were then rolled as ‘Swiss rolls’ and transferred to a tissue cassette and further fixed in 4% of PFA at 4 °C overnight. Fixed tissues were then washed three times with PBS and transferred into a tissue cassette for processing and paraffin-embedding. Paraffin blocks were sectioned at 5 μm for further histological analyses and immunofluorescence staining. For H&E staining, in brief, paraffin sections were first deparaffinized and rehydrated, then stained with Gill’s Hematoxylin (REF6765008; Thermo Fisher Scientific) and eosin (HT110180; Sigma-Aldrich). For Alcian Blue staining, paraffin sections were first deparaffinized and rehydrated, then stained with Alcian Blue solution (Poly Scientific, S111A-16oz) for 30 min, followed by counterstaining with Nuclear Fast Red (StatLab, STNFRPT) for 10 min. Rinsed slides were dried for 1 h at 60 °C in an incubator before mounting. Mounted slides were scanned using the Aperio Light Field Slide Scanner (Leica Biosystems) and representative areas were taken and evaluated.

Immunofluorescence staining was performed as previously described[2]. In brief, antigen retrieval was performed by boiling the deparaffinized and rehydrated paraffin sections in Tris-EDTA buffer (10 mM Tris base, 1 mM EDTA, pH 9.0) for 6 min in a pressure cooker. The sections were blocked with Fc receptor blocker (Innovex Biosciences) for 30 min followed by blocking buffer (PBS containing 0.1% Triton X-100, 2.5% bovine serum albumin and 5% goat serum) for 1 h at room temperature. Sections were then incubated with primary rabbit anti-POU2F3 (1:100; Sigma-Aldrich, HPA019652) or anti-DCLK1 (1:300; Abcam, ab37994) antibodies in 0.5× blocking buffer overnight at 4 °C. After washing the slides twice with PBS, the sections were incubated with secondary antibody (goat anti-rabbit Alexa568; 1:300 dilution; Life Technologies) and counterstained with DAPI (1:1,000; Life Technologies) for 1 h at room temperature. Images were collected at ×40 magnification using the Leica TCS SP8 confocal microscope and were processed using Leica LAS X software. Quantification of POU2F3- and DCLK1-expressing cells in trachea was performed by manually counting POU2F3- or DCLK1-positive cells and dividing by the total DAPI-positive cell number of the epithelial layer in the whole trachea. Quantification of POU2F3 and DCLK1-expressing cells in the gastrointestinal tract (including gastric, small and large intestines) was performed by calculating POU2F3- and DCLK1-positive cells divided by all DAPI-positive cells using ImageJ.

RNA-FISH

RNA-FISH analysis of *Pou2f3*, *C11orf53* and *Colca2* was performed on fresh frozen tissue sections using an RNAscope Fluorescent Multiplex Assay Kit (ACD, 320850) with the following probes (ACD): *Pou2f3*, 539211-C1; *C11orf53* (*1810046K07Rik*), custom designed by ACD; and *Colca2* (*Gm684*), custom designed by ACD. The custom-designed probe was unfortunately listed as confidential by ACD and cannot be shared. The assay was performed according to the manufacturer's instructions. Slides were counterstained with DAPI. Images were collected at ×40 magnification plus ×3 zoom using a Leica TCS SP8 confocal microscope and were processed with Leica LAS X software.

Whole-mount immunostaining and tissue clearing of nasal cavity

To detect the presence or abundance of tuft cells in the nasal epithelium layer of nasal cavity, we performed whole-mount immunostaining and tissue clearing as previously reported of nasal cavity tissue[62]. In brief, nasal cavity tissues (from nose tip to eye) were collected from 8–10-day-old mice, and then fixed at 4 °C overnight in PBS with 4% PFA and 30% sucrose. After washing three times with PBS at room temperature, nasal tissues were permeabilized in methanol gradients in PBS for 30 min each (50% methanol, 80% methanol and 100% methanol). Next, nasal cavity was bleached with Dent's bleach (15% H_2O_2, 16.7% DMSO in MeOH) for 1 h at room temperature, and further rehydrated through descending methanol gradients in PBS (80% methanol, 50% methanol and PBS). The tissues were then blocked overnight at 4 °C with shaking in blocking buffer (PBS with 0.3% Triton X-100, 0.2% BSA, 5% DMSO and 5% donkey serum). Next, nasal cavities were stained with primary rabbit anti-POU2F3 (1:100; Sigma-Aldrich, HPA019652) or anti-DCLK1 (1:300; Abcam, ab37994) antibodies in blocking buffer for 5 days at 4 °C with shaking. After washing overnight in wash buffer (PBS with 0.2% Triton X-100 and 3% NaCl), the tissues were stained with donkey anti-rabbit-Alexa568 (1:300) secondary antibody and DAPI (1:1,000) for 3 days in blocking buffer without DMSO at 4 °C. After washing the tissues for 24 h in washing buffer, the tissues were then dehydrated in methanol gradients in distilled H_2O for 30 min in each step (50% methanol, 70% methanol, 90% methanol, and three times with 100% methanol). Finally, tissues were cleared in glass containers with 1:1 of methanol:BABB (benzyl alcohol:benzyl benzoate, 1:2) for 30 min and in 100% BABB for 1 h before being imaged in a Leica TCS SP8 confocal microscope. *z*-Stacked images were processed using Leica LAS X software. Videos were generated using LAS X 3D Visualization of Leica LAS X software.

IL-25 treatment experiment

Age- and sex-matched WT and *C11orf53* knockout mice ($n = 5$) were treated with 500 ng of IL-25 (R&D Systems) in 200 μl of PBS through intraperitoneal injection for a consecutive 7 days. At day 8, the small intestines were collected and processed as described in the 'Tissue preparation for histology and immunofluorescence staining' section. Paraffin sections of small intestines were used for H&E, Alcian Blue and IF staining for quantifying goblet and tuft cell frequency. Quantification of POU2F3- and DCLK1-expressing cells was performed by calculating POU2F3- and DCLK1-positive cells divided by all DAPI-positive cells using ImageJ. Immunofluorescence staining procedures were performed as described above.

Sample preparation for scRNA-seq

The age- and sex-matched *C11orf53*$^{+/+}$ and *C11orf53*$^{-/-}$ mice were euthanized (littermates). The small intestines were cut and washed with PBS in cold conditions multiple times until no clumps were observed in the PBS. The cells were detached from the small intestine by incubating with 7.5 mM EDTA in PBS on a 4 °C rotator for 45–60 min. After detachment, cells were digested with TrypleExpress containing 1–2 μM Y-Factor in a 37 °C water bath for 75 s followed by a repetitive pipetting at least 30 times to ensure full dissociation. DMEM/F12 medium with 10% FBS was then added to quench the digestion. The cells were then centrifuged and resuspended in sorting medium (1× PBS, 1 μM Y-factor and 0.5 mM EDTA) and filtered through a 40 μM filter. For the second replicate, cells were incubated with CD45 microbeads according to the manufacture's protocol (Miltenyi Biotec, 30–052-301) to deplete CD45^{high} cells before sorting.

scRNA-seq analysis

Cell suspensions were stained for viability using Sytox blue (Thermo Fisher Scientific, S34857) and sorted on a Sony SH800S (Sony Biotechnology) system using a 100 μm chip (LE-C3110) in ultra purity mode. Sorted cells were washed and resuspended in PBS containing 0.04% BSA. An aliquot was stained with ViaStain AOPI (Nexcelom, CS2–0106-5mL) and counted using the Countess FL II automated cell counter. Up to 12,000 cells were loaded per lane on 10x Chromium microfluidic chips. Single-cell capture, barcoding and library preparation were performed using the 10x Chromium chemistry, using the NextGEM Single-Cell 3′ Library Kit v.3.1 (1000121; 10x Genomics). cDNA and libraries were checked for quality using an Agilent Bioanalyzer, quantified by KAPA qPCR, and sequenced on a NextSeq 500 (Illumina) system to an average depth of approximately 24,000 reads per cell. The Cell Ranger pipeline (v.5.0.0, 10x Genomics) was used to align FASTQ reads to the mouse reference genome (gex-mm10–2020-A, 10x Genomics) and produce digital gene–cell count matrices. The resulting gene–barcode matrices were further analysed by Scanpy v.1.8.1 for secondary analysis. First, cells with more than 20% mitochondrial gene reads were filtered out and processed with Scanpy using the default settings. The raw read counts were then normalized and log-transformed after masking highly expressed genes with the function scanpy.pp.normalize_total(), using a target sum of 1×10^6 counts. Principal component analysis dimensionality reduction was performed on the top 4,000 highly variable genes (ignoring ribosomal and mitochondrial genes), followed by neighbour graph generation and *t*-SNE projection with the top 50 principal components using the default settings. Cells were clustered based on the Leiden algorithm with a resolution of 0.8. For cell type annotation, data from ref. [16] was used for building customized support vector machines using the scikit-learn.svm Python module. In brief, a simplified gene expression matrix was extracted from the top 50 differentially expressed genes from each cell type. A Gaussian svm kernel was trained on this matrix using a 70/30 train/test split, and the fitted model was used to transfer cell type labels to epithelial cells in our dataset with around 80% accuracy. Epithelial cells were subsetted by examining clusters expressing canonical markers (for example, *Epcam*$^{+}$, *Ptprc*$^{-}$ expression), and the predicted cell types were visualized with marker expression to assess accuracy.

Cell cycle arrest experiment and analysis

Cas9-expressing NCI-H211 or NCI-H1048 cell lines were infected with sgRNA linked with GFP (LRG2.1T, Addgene, 108098) at a high titre. Then, 5 days after infection, cells were labelled with 10 μM BrdU for 2 h at 37 °C. Cells were then collected and prepared according to the manufacturer's protocol (BD, APC BrdU Flow Kit, 552598). Cells were then co-stained with DAPI for measuring the total DNA content and run on an LSR II flow cytometer (BD Biosciences). Data were analysed using FlowJo (v.10, BD Biosciences) according to standard procedures.

Statistical analysis

All statistical analysis was performed using Python scipy v.1.5.0 or Graph Pad Prism 9. All statistical tests are indicated in the figure legends. In principle, when only two groups were considered, two-tailed Student's *t*-tests were used. For BrdU apoptosis analysis, two-way analysis of variance was used for calculating the difference of sgPOU2F3, sgC11orf53/OCA-T1 versus sgCOLCA2/OCA-T2 versus sgNEGs. If the variance among the groups was significantly different, Welch's correction was applied before performing the Student's *t*-test.

Reporting summary

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

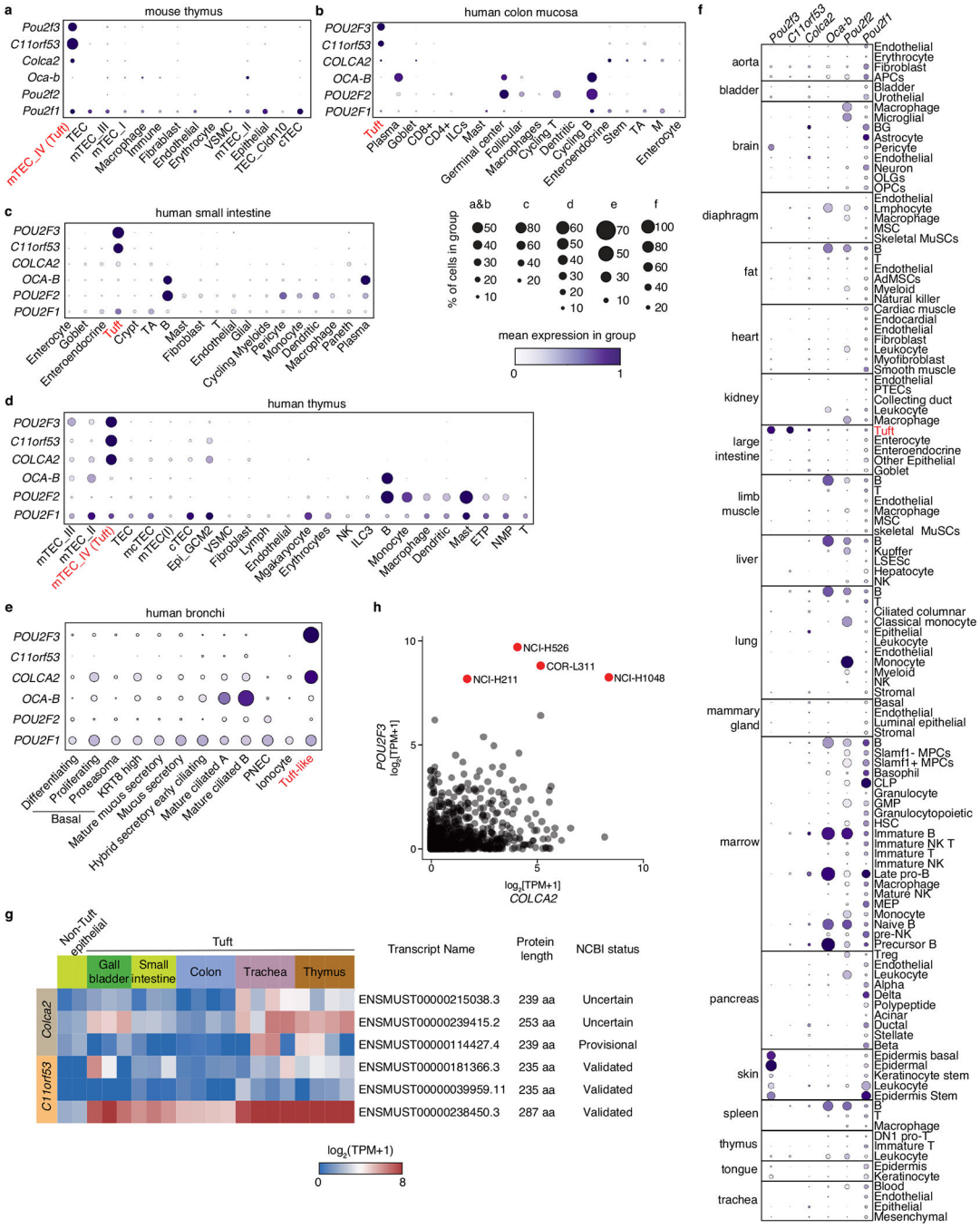
Data availability

All genomic datasets are available at the GEO database under accession code GSE186614. The thymic scRNA-seq dataset was obtained from ref. [15]; the human colon mucosa dataset was obtained from ref. [14]; and the human bronchi dataset was obtained from ref. [24]. The Tabula Muris Consortium dataset was obtained from Figshare (https://figshare.com/projects/Tabula_Muris_Transcriptomic_characterization_of_20_organs_and_tissues_from_Mus_musculus_at_single_cell_resolution/27733). The cancer dependency dataset as well as expression was obtained online (https://depmap.org/portal/download/, DepMap Public 21Q2). The transcriptome dataset for patients with SCLC was obtained from ref. [22]. The original dataset for analysing isoform level of C11orf53/OCA-T1 and COLCA2/OCA-T2 in mouse tuft cells is derived from ref. [9].

Code availability

Customized code used to analyse the data is available on GitHub (https://github.com/xlw1207/paper).

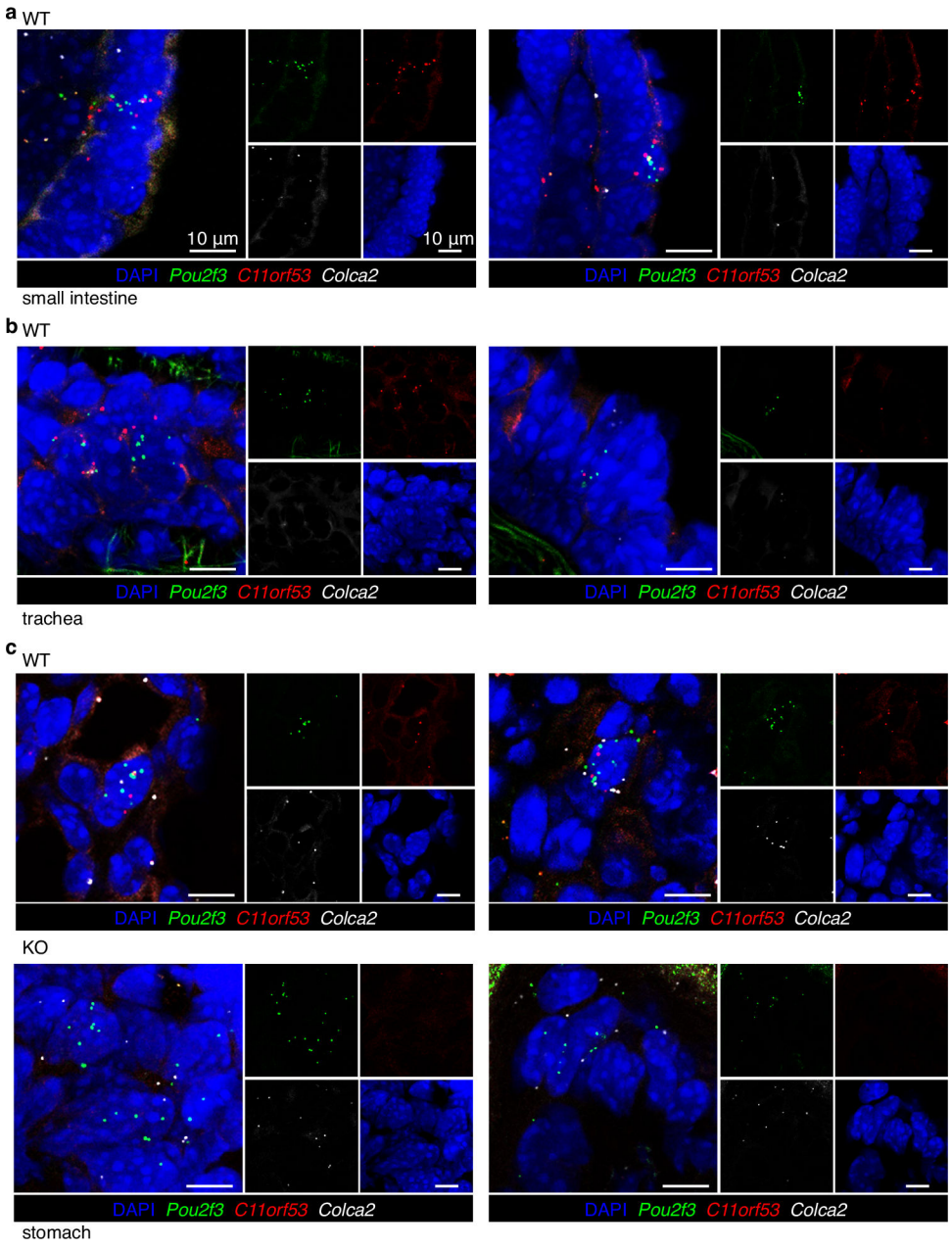
Extended Data



Extended Data Fig. 1 |. C11orf53, COLCA2, and POU2F3 are expressed in murine and human tuft cells.

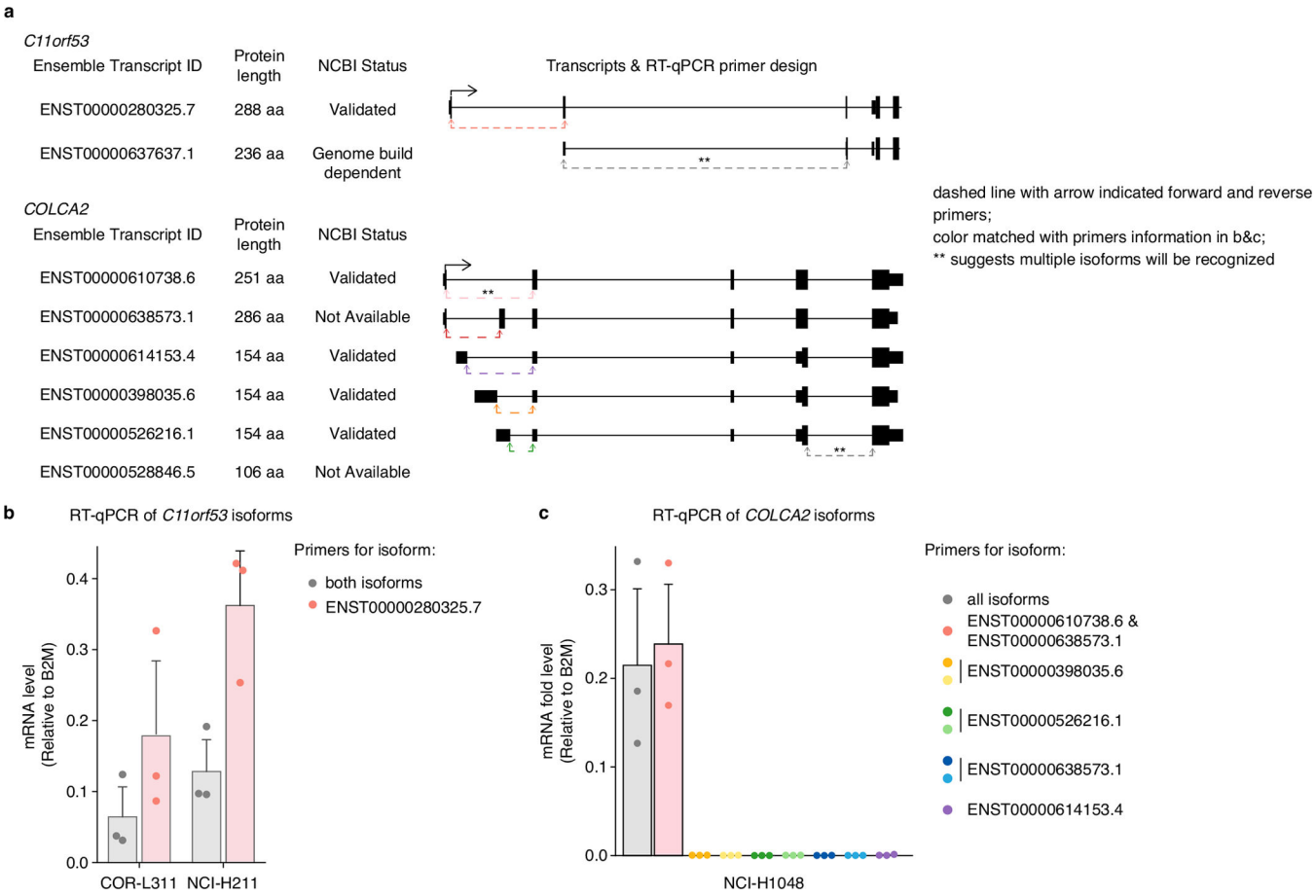
(**a**)-(**f**) Summary of scRNA-seq data. The plots depict the average expression of indicated genes in the indicated cell types. *COLCA2*: mouse gene name *Gm684*. (**a**) mTEC: medullary thymic epithelial cells; TEC: thymic epithelial cells; VSMC: vascular smooth muscle cell; cTEC: cortical thymic epithelial cells. (**b**) ILCs: innate lymphoid cells; TA: transit amplifying. (**d**) NK: natural killer; ETP: early thymic progenitor. NMP: neutrophil-

myeloid progenitor. (**e**) PNEC: pulmonary neuroendocrine cells. (**f**) APCS: antigen presenting cells; BG: Bergmann glial cell; OLGs: oligodendrocytes; OPCs: oligodendrocyte precursor cell; MSC: mesenchymal stem cell; MuSC: muscle satellite stem cell; AdMSC: adipose mesenchymal stem cell; PTECs: proximal tubular epithelial cells; LSESc: Liver sinusoidal endothelial cells; MPCs: multipotent progenitor cell; CLP: common lymphoid progenitor cells; GMP: granulocyte monocyte progenitor cell; HSC: hematopoietic precursor cell; MEP: megakaryocyte-erythroid progenitor cell; Treg: regulatory T cells; DN1 pro-T: double negative 1 progenitor T cells. (**g**) Summary of *C11orf53* and *Colca2* isoform expression in sorted murine tuft cells. (**h**) mRNA levels of *POU2F3* and *COLCA2* across 1,379 human cancer cell lines from CCLE database. Data are obtained from the Cancer Dependency Map (DepMap) portal (21Q2). Source of each single-cell or bulk RNA-seq dataset as well as plotting is indicated in method section.



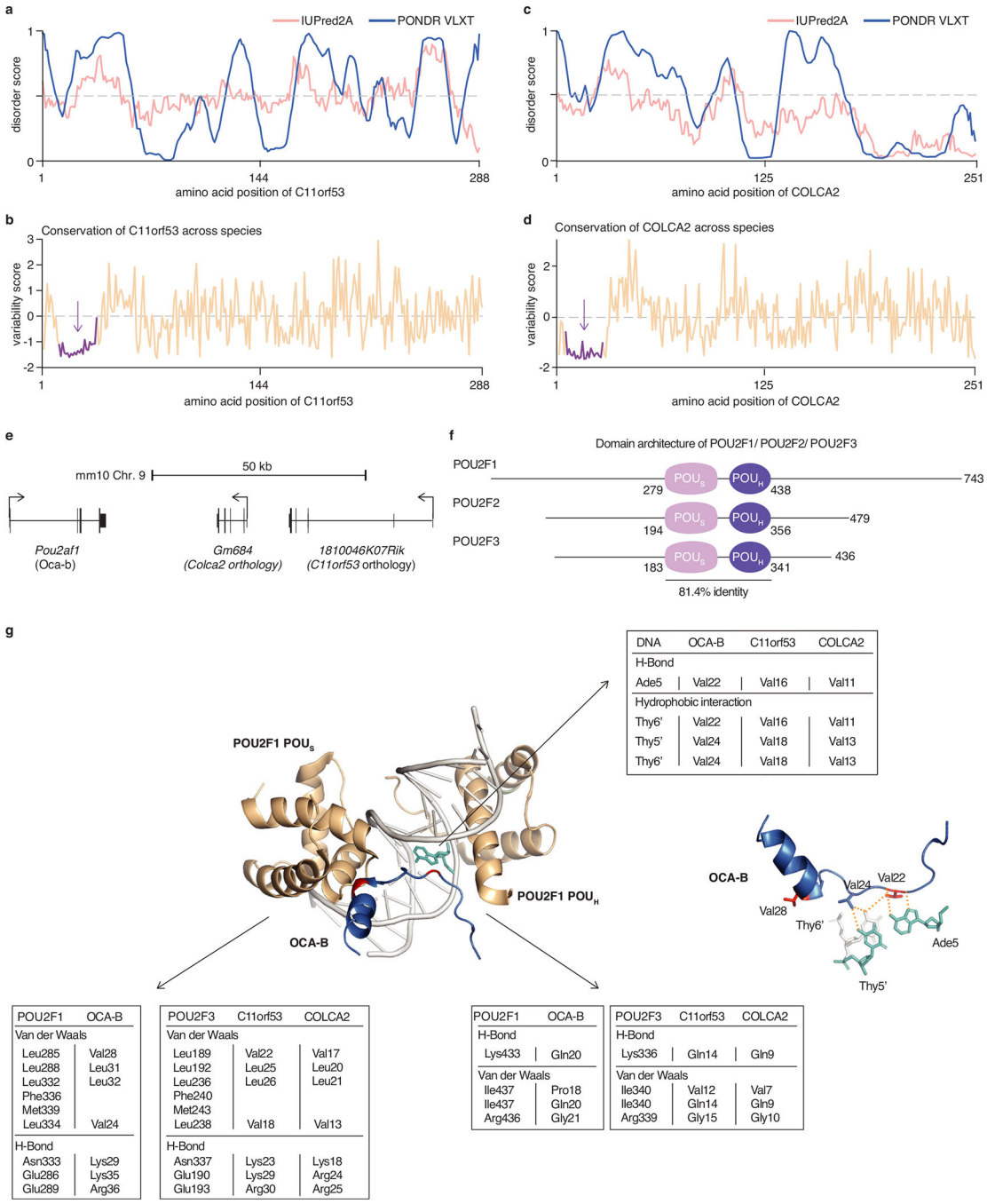
Extended Data Fig. 2 |. Evaluation of *Pou2f3*, *C11orf53*, and *Colca2* mRNA level through RNA-FISH in different tissues.

(**a**)-(**c**) Representative RNA-FISH analysis of *Pou2f3*, *C11orf53*, and *Colca2* expression in mouse small intestine (**a**), trachea (**b**), and stomach tissue of wild-type mice (**c, top**). (**c, bottom**) RNA-FISH analysis *Pou2f3*, *C11orf53*, and *Colca2* expression in the stomach of C11orf53$^{-/-}$ (KO) mice (one independent experiment from three different mice). All scale bar represents 10 μM.



Extended Data Fig. 3 |. Isoform analysis of human *C11orf53* and *COLCA2*.

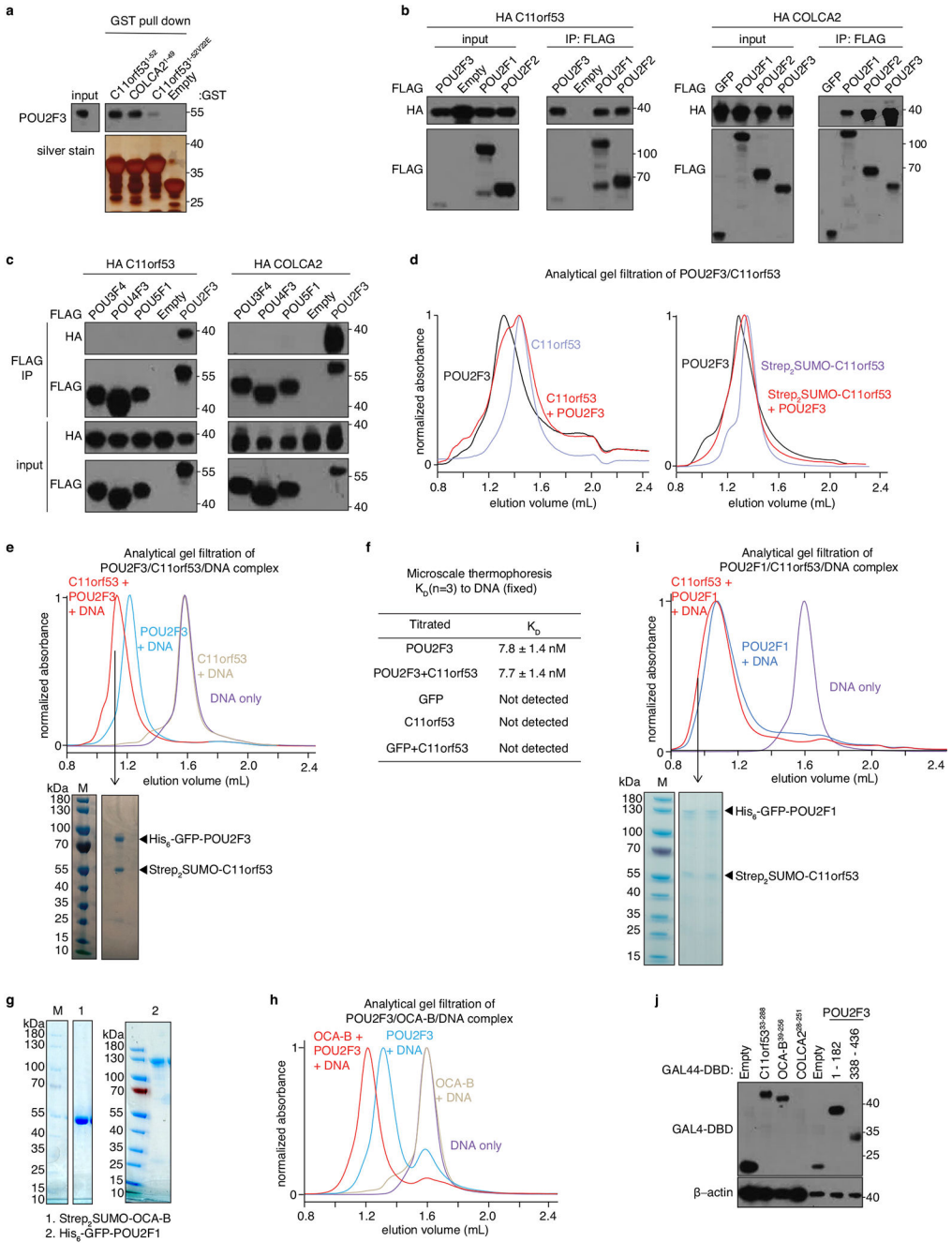
(**a**) Isoform information of *C11orf53* and *COLCA2* extracted from the Ensemble and NCBI databases. (**b**) RT-qPCR analysis of RNA from SCLC-P cell lines using the primers indicated in (**a**), that quantify either the short isoform of *C11orf53* or all isoforms. (**c**) RT-qPCR analysis of *COLCA2* isoforms in NCI-H1048 cells using primers indicated in (**a**). for (**b**) and (**c**), mRNA level (Ct value) of isoform indicated is normalized to the level of *B2M*. Bar graph represents the mean of normalized mRNA level from three biological replicates with each replicate depicted as individual dot. Primers are provided in Supplementary Table 6.



Extended Data Fig. 4 |. C11orf53, COLCA2, and OCA-B share a conserved peptide that corresponds the known binding site of OCA-B with POU2F1/OCT1.

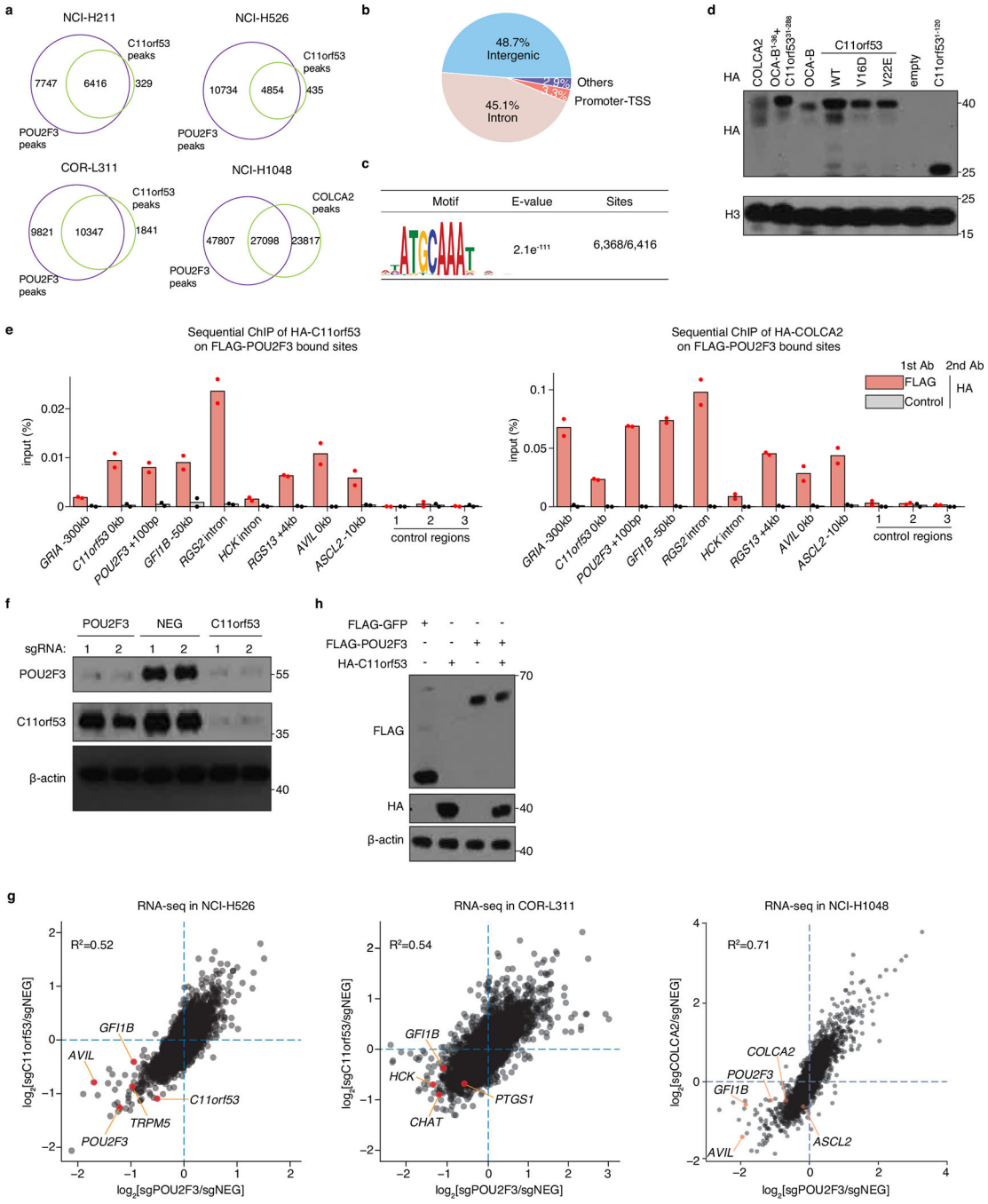
(**a**) and (**c**) Predicted intrinsic disorder for C11orf53 and COLCA2. PONDR (Predictor of Natural Disordered Regions) VLXT scores and IUPred2A (Intrinsically Unstructured Proteins Prediction) scores are shown on the y axis, and the amino acid positions are shown on the x axis. (**b**) and (**d**) The conservation analysis of C11orf53 and COLCA2 protein sequences across species with Bayesian method on ConSurf Server. The purple region is the conserved peptide (residues 10–32 for C11orf53 and 5–27 for COLCA2). (**e**)

C11orf53, *COLCA2*, and *Pou2af1*/OCA-B gene cluster in the mouse genome (mm10). (**f**) Architecture and conservation of POU homeodomains of POU2F1/OCT1, POU2F2/OCT2, and POU2F3/OCT11. (**g**) Summary of residues involved in protein-DNA and protein-protein close contacts (<4.0 Å) for OCA-B with POU2F1, C11orf53 with POU2F3, and COLCA2 with POU2F3. The crystal structure of OCA-B$^{1–44}$ and POU2F1^{DBD} (PDB: 1CQT^{25})^{25} was used as a model for predicting the structure of C11orf53 or COLCA2 with POU2F3^{DBD} on an octamer motif (ATGCAAAT) in PyMOL.



Extended Data Fig. 5 |. Additional biochemical evidence that C11orf53/OCA-T1 and COLCA2/OCA-T2 are paralogues of OCA-B.

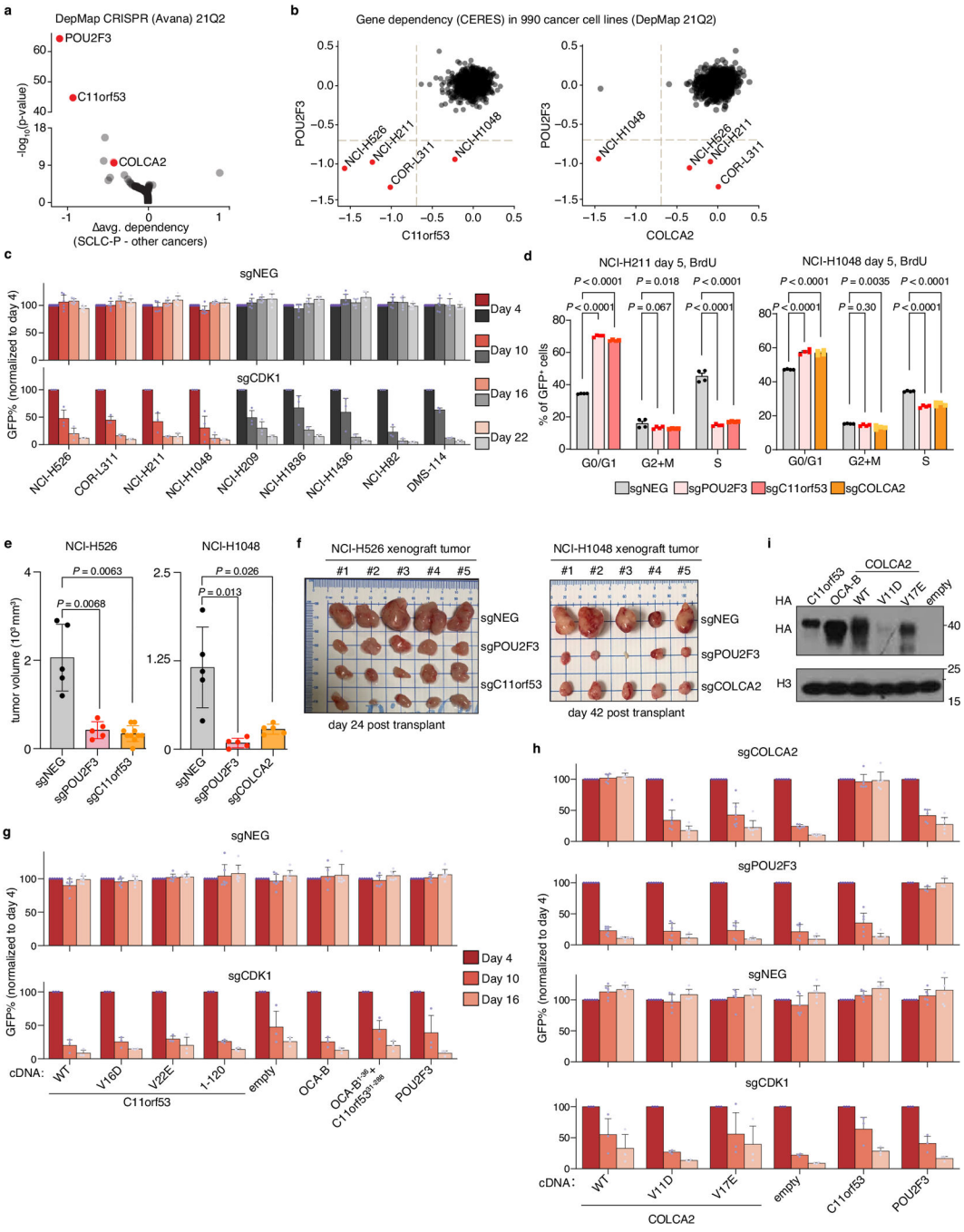
(a) GST-pulldown with western blot of endogenous POU2F3 from NCI-H211 nuclear extracts with indicated constructs (n = 1). **(b)** Co-IP testing the interaction of HA-C11orf53 (HEK293T whole cell lysates, n = 2) or HA-COLCA2 (NCI-H1048 nuclear extract, n = 1) with FLAG-POU2F1, FLAG-POU2F2, and FLAG-POU2F3. **(c)** Co-IP testing the interaction of HA-C11orf53 or HA-COLCA2 with FLAG-POU3F4/OCT9, FLAG-POU4F3/BRN3C, FLAG-POU5F1/OCT4, and POU2F3/OCT11 in HEK293T whole cell lysates. **(d)** Analytical gel filtration of His_6-GFP-POU2F3 with $Strep_2$SUMO-C11orf53 or untagged C11orf53, performed in the absence of any DNA. **(e)** Replicate of analytical gel filtration of His_6-GFP-POU2F3/$Strep_2$SUMO-C11orf53/octamer motif A alone or in combination, accompanied by Coomassie blue staining of the peak fraction of the ternary complex (red curve). Data are representative of two biological replicates (**c-e**). **(f)** Summary of microscale thermophoresis measurements of protein binding affinity for the octamer motif A DNA. Highest protein concentration tested is 1 μM for all experiments. **(g)** Purity assessment of recombinant OCA-B (expressed and purified from *Sf9* cells) and POU2F1/OCT1 (expressed and purified from *E. coli*) proteins by SDS-PAGE and Coomassie blue staining. **(h)** Analytical gel filtration of POU2F3, OCA-B, octamer DNA motif A assemblies. **(i)** Analytical gel filtration of POU2F1, C11orf53, and octamer DNA motif A assemblies. For (h) and (i), the maximum absorbance at 260 nm for each injection was normalized to 1.0 for the ease of comparison. As POU2F1 is larger in molecular weight than POU2F3, the shift in elution volume is less prominent for the ternary complex. Complex formation was validated by SDS-PAGE assessment of the peak fraction to confirm the presence of both proteins (red curve). Data are representative of two biological replicates (**i-j**). Source data for microscale thermophoresis assay is provided in Supplementary Table 1, uncropped gels are provided in Supplementary Fig. 1b, d–g.



Extended Data Fig. 6 |. Epigenomic evaluation of POU2F3, C11orf53/OCA-T1, and COLCA2/OCA-T2 in SCLC-P cell lines.

(**a**) Comparison of POU2F3, C11orf53, and HA-COLCA2 ChIP-seq peak overlap in the indicated cell lines. For each protein, peaks represent the ones that are consistently identified from 2–3 replicate of experiments. Detailed information is in method. (**b**) Annotation of POU2F3 and C11orf53 overlapping peaks NCI-H211 cells. (**c**) Position weight matrix of discovered motif enriched on POU2F3/C11orf53 co-binding sites by MEME from NCI-H211 cell line. (**d**) Western blot analysis of lentivirally overexpressed constructs used for

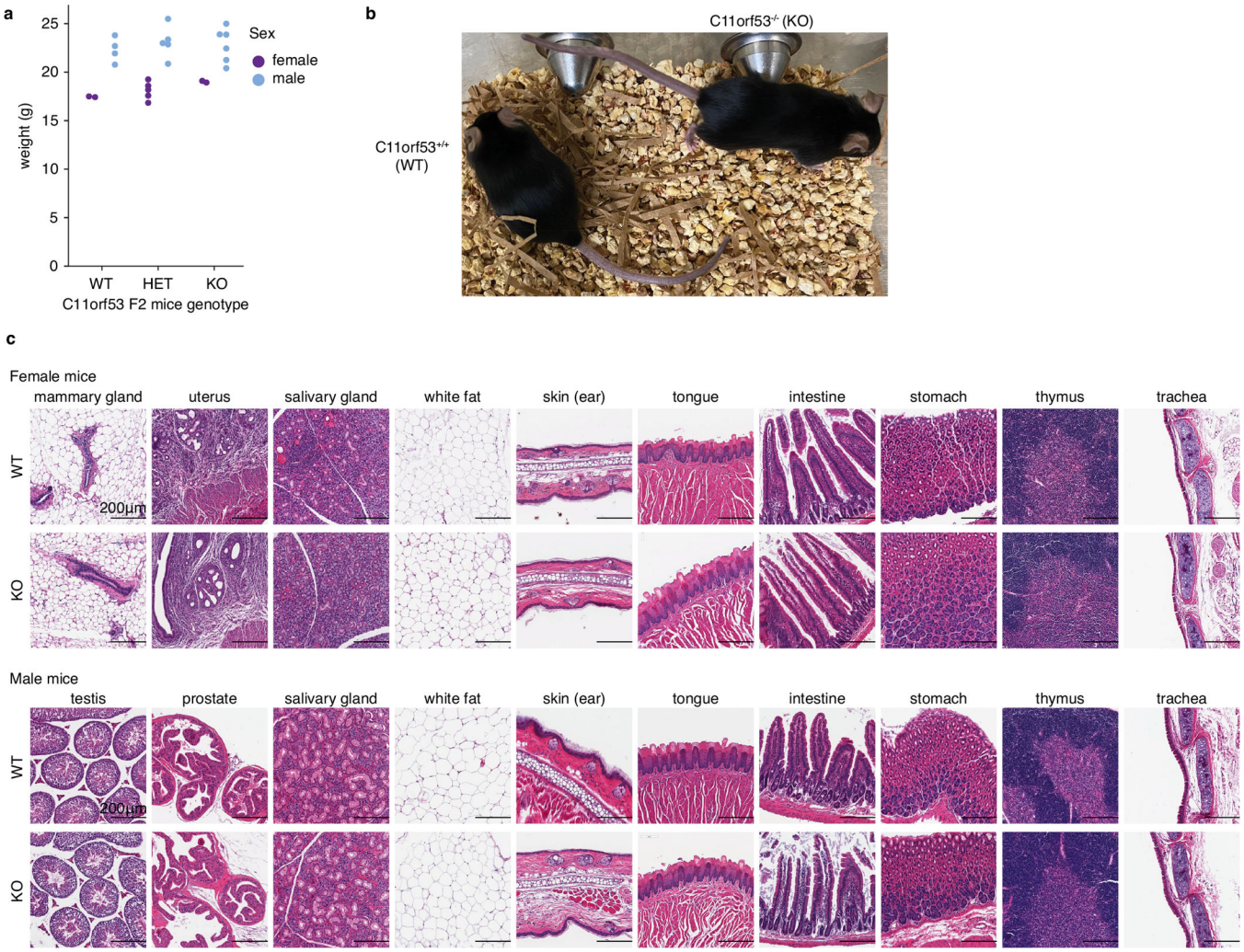
ChIP-qPCR and gene complementation assay in NCI-H211 cells (n = 2). (**e**) Sequential ChIP-qPCR analysis of overexpressed HA-C11orf53 (NCI-H211) or HA-COLCA2 (NCI-H1048) co-expressed with FLAG-POU2F3 binding sites in NCI-H211 and NCI-H1048 cells respectively with anti-FLAG (1^{st} IP) or protein G beads (1^{st} IP) and anti-HA (2^{nd}) antibodies. The enrichment is adjusted to the input amount. Two biological replicates are performed and represented as individual dots; the bar value represents the average enrichment over input of two biological replicates. (**f**) Western blot analysis of POU2F3 or C11orf53 in NCI-H211 (Cas9) cells transduced with indicated sgRNAs. Samples were collected 4 days post infection (n = 1). (**g**) RNA-seq analysis comparing mRNA changes following C11orf53, COLCA2, or POU2F3 knockout compared to control in three SCLC-P cells. RNA was collected five- or six-days post sgRNA infection. Each dot represents the log_2fold-change of a single protein-coding genes (read outs > = 10). (**h**) Western blot analysis of FLAG-POU2F3, HA-C11orf53, and FLAG-GFP expression in murine YT330 SCLC cells. Data is representative of two biological replicates. Uncropped gel is provided in Supplementary Fig. 1i. Source data for gene expression changes upon knockout in different cells are provided in Supplementary Table 2. sgRNA and primer sequences are provided in Supplementary Table 6.



Extended Data Fig. 7 |. C11orf53/OCA-T1 and COLCA2/OCA-T2 are selective dependencies in SCLC-P lines.

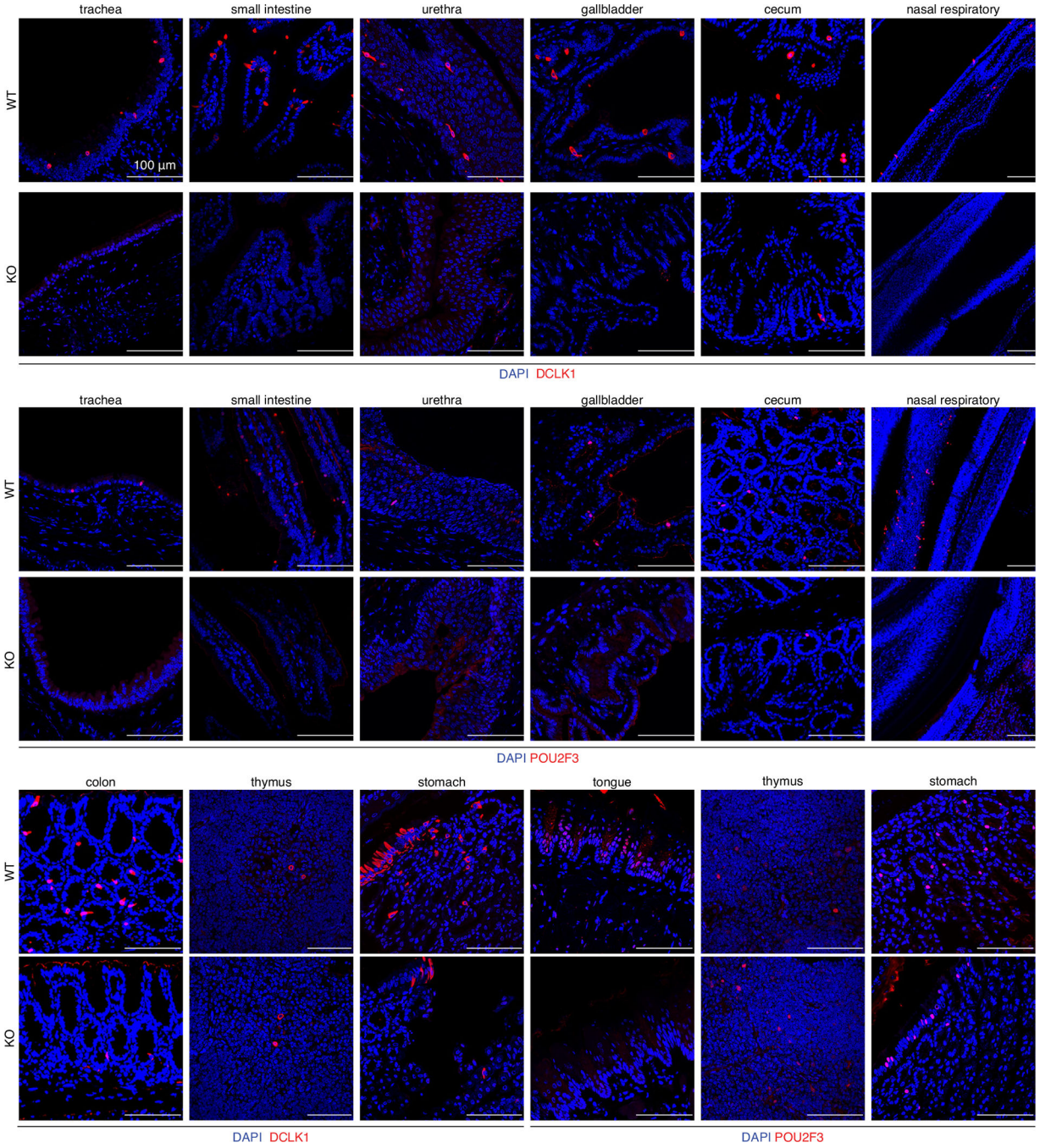
(a) Two-class comparison of gene dependencies of four SCLC-P versus 986 other cancer cell lines. The average essentiality of each gene from SCLC-P is subtracted to its mean in other cell lines. This difference is moderated with an empirical Bayes method using the adaptive shrinkage method described in CRNA package - ashr. (**b**) Scatter plot of *C11orf53*, *COLCA2*, and *POU2F3* dependency scores across a panel of human cancer cell lines. Data obtained from DepMap (21Q2). **(c)** Control sgRNAs (2 negative controls or one sgCDK1)

for the competition-based proliferation assays (related to Fig. 4a), (n = 3). Mean ± s.d. is plotted. (**d**) BrdU incorporation assays following CRISPR-based targeting of *POU2F3*, *C11orf53*, or *COLCA2* or negative control in the indicated $Cas9^+$ cell lines. Two technical replicates with two independent sgRNAs for each gene as biological replicates. Adjusted *P* value was calculated with two-way ANOVA with Tukey's multiple comparison tests. (**e-f**) Tumour weights and imaging at the terminal timepoint of the xenograft experiments shown in Fig. 4b. Mean ± s.e.m. is plotted for (**d-e**). P-values are derived from two-tailed unpaired student's t-test with Welch's correction. (**g**) Controls for NCI-H211 gene complementation assay shown in Fig. 4c. (n = 3). Mean ± s.d. is plotted. (**h**) Competition-based proliferation assays in NCI-H1048 cells cotransduced with indicated cDNAs and sgRNAs lentivirally to assess functionality of indicated mutants. cDNAs were engineered to be resistant to Cas9/sgRNA-mediated cutting. Mean ± s.d. of normalized GFP percentage is plotted (n = 3). (**i**) anti-HA western blot of the indicated cDNAs from (**h**). Data are representative of two biological replicates. Source data for all GFP depletion assays, BrdU assays and tumour weights are provided in Supplementary Table 3–4, gating strategy for BrdU assay is provided in Supplementary Figure 2, sgRNA sequences are provided in Supplementary Table 4, uncropped gels are provided in Supplementary Fig. 1j.



Extended Data Fig. 8 |. Similar weights, morphology, and organ histology of C11orf53$^{+/+}$ and C11orf53$^{-/-}$ mice. WT: C11orf53$^{+/+}$ mice, KO: C11orf53$^{-/-}$ mice.

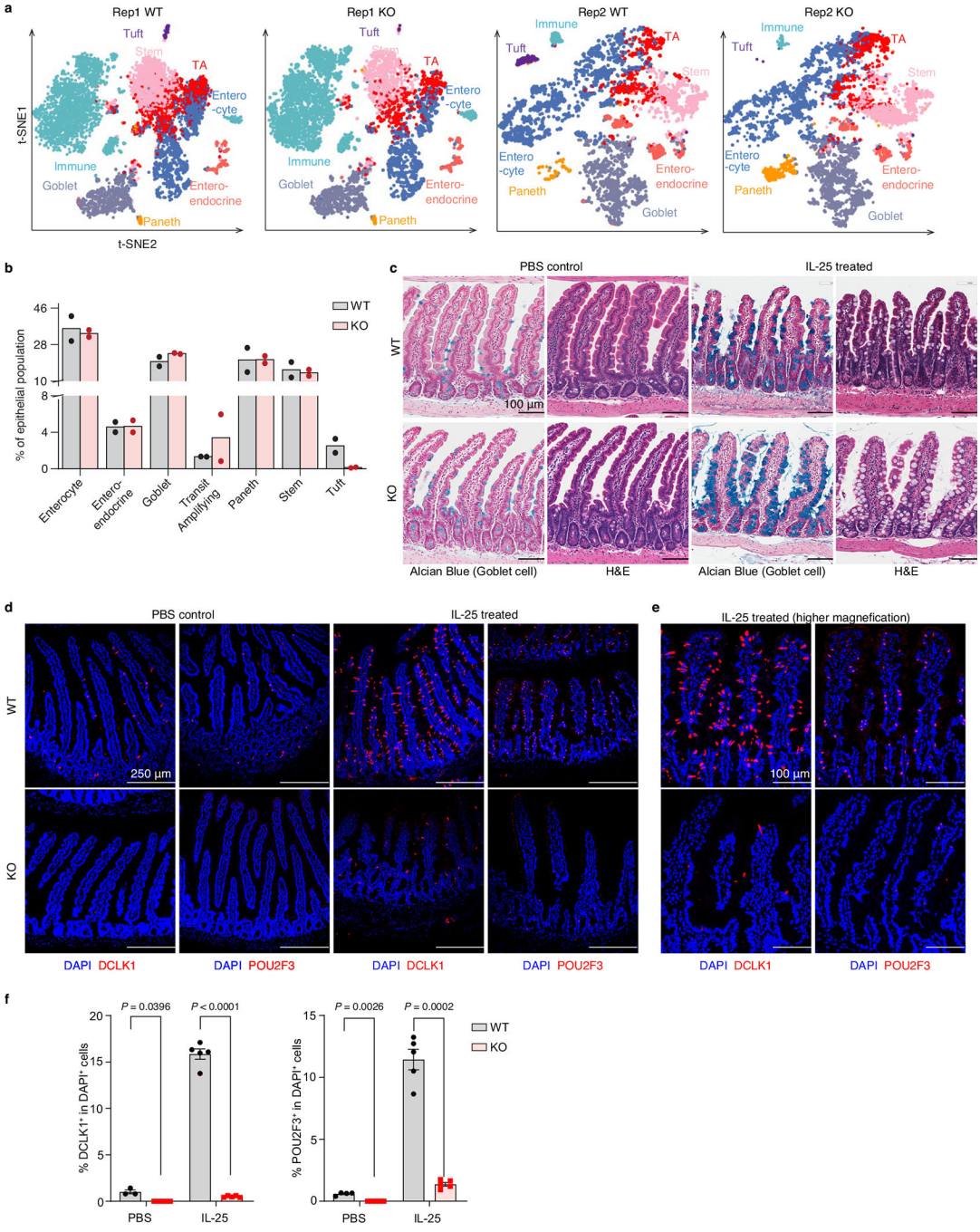
(**a**) Body weights of C11orf53 F2 mice of each genotype (2 female WT mice, 4 male WT mice; 3 female KO mice, 4 male KO mice; 5 HET male and female mice were included in the experiment with 6–10 weeks old littermate mice). (**b**) Representative images of C11orf53$^{+/+}$ and C11orf53$^{-/-}$ mice F2 mice (representative image of four mice). (**c**) H&E analysis of representative organs from age- and sex-matched C11orf53$^{+/+}$ and C11orf53$^{-/-}$ littermate mice (one female and one male mice for both genotypes were included in the experiment). All scale bars represent 200 μm. Source data for mice body weight is provided in Supplementary Table 5.



Extended Data Fig. 9 |. Representative immunofluorescence staining of POU2F3 and DCLK1 in tissues of C11orf53$^{+/+}$ and C11orf53$^{-/-}$ mice.

WT: C11orf53$^{+/+}$ mice, KO: C11orf53$^{-/-}$ mice. For trachea staining, 5 WT, 6 KO and 10 HET mice were included, except for POU2F3 staining which 9 HET mice were included in. For small intestine, 10 WT, 10 KO, 8 HET mice were included for POU2F3 staining and 9 WT, 10 KO, 5 HET mice for DCLK1 staining were included in. For urethra, 5 mice were included. For gallbladder, 6 WT and 5 KO mice were included. For caecum, colon, tongue and nasal respiratory, 3 WT and KO mice were included. For thymus, 6 WT and KO mice

were included. For stomach, 10 mice were included. Data shown are representative images of acquired for each tissue. All scale bars represent 100 μm.



Extended Data Fig. 10 |. RNA-FISH, sSingle-cell RNA-seq, and IL-25 treatment experiments confirming in C11orf53$^{-/-}$ mice.

(**a**) t-distributed stochastic neighbour embedding (t-SNE) of small intestine epithelial cells (points), coloured by cell type assignments. Two independent replicates of single-cell RNA-seq data from C11orf53 wildtype (WT) and knockout (KO) littermate female mice. For the second replicate, cells that expressed a high level of CD45 were depleted by using

CD45 microbeads before library preparation (Miltenyi Biotec, 30–052-301). TA: transient amplifying. (**b**) Quantification of different cell types in the intestinal epithelium from single-cell RNA-seq experiments in small intestine from two independent replicates. Bar graphs represent the mean. (**c**) H&E and Alcian Blue (Goblet cell) staining of intestines of C11orf53 WT and KO mice treated with PBS control or 500 ng of IL-25 for 8 days (for PBS control, four mice were included. For IL-25, five mice were included). Scale bars represent 100 μm. (**d**) Immunofluorescence staining of DCLK1 and POU2F3 markers in intestines of C11orf53 WT and KO mice treated with PBS control or IL-25, scale bars represent 250 μm. (**e**) Same as (**d**), but higher magnification (scale bars represent 100 μm). For DCLK1 staining, three WT mice were included in PBS control and five WT mice were included in IL-25 treatment. For POU2F3 staining, four and five WT mice were included in PBS and IL-25 treatment respectively, five KO mice were included in both PBS control and IL-25 treatment. (**f**) Quantification of (d). Mean ± s.e.m. is plotted. Two-tailed unpaired student's t-test with Welch's correction was used to evaluate significance. Source data for IF quantification is provided in Supplementary Table 3.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge C. Hammell, E. Luk, J. Sheltzer and K. Adelman for discussions and suggestions throughout the course of this study. This work was supported by Cold Spring Harbor Laboratory NCI Cancer Center Support grant CA045508. Additional funding was provided to C.R.V. by the Pershing Square Sohn Cancer Research Alliance, National Institutes of Health grants CA013106 and CA242919, Department of Defense grant W81XWH1910317, and the Cold Spring Harbor Laboratory and Northwell Health Affiliation. L.J.-T. is an investigator of the Howard Hughes Medical Institute. M.E. was supported by the Dr Marcia Kramer Mayer, William C. and Joyce C. O'Neil Charitable Trust, and the Pershing Square Foundation. X.-Y.H. was supported by the 2021 AACR-AstraZeneca Breast Cancer Research Fellowship (grant no. 21–40-12-HE). J.S. was supported by NIH grant CA231997. Y.T.S. was supported by Singapore National Science Scholarship (PhD) A*STAR.

References

1. Schneider C, O'Leary CE & Locksley RM Regulation of immune responses by tuft cells. Nat. Rev. Immunol. 19, 584–593 (2019). [PubMed: 31114038]
2. Huang Y-H et al. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. Genes Dev. 32, 915–928 (2018). [PubMed: 29945888]
3. Yamada Y et al. A tuft cell-like signature is highly prevalent in thymic squamous cell carcinoma and delineates new molecular subsets among the major lung cancer histotypes. J. Thorac. Oncol. 16, 1003–1016 (2021). [PubMed: 33609752]
4. Gerbe F, Legraverend C & Jay P The intestinal epithelium tuft cells: specification and function. Cell. Mol. Life Sci. 69, 2907–2917 (2012). [PubMed: 22527717]
5. Gerbe F et al. Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. Nature 529, 226–230 (2016). [PubMed: 26762460]
6. von Moltke J, Ji M, Liang H-E & Locksley RM Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit. Nature 529, 221–225 (2016). [PubMed: 26675736]
7. Howitt MR et al. The taste receptor TAS1R3 regulates small intestinal tuft cell homeostasis. Immunohorizons 4, 23–32 (2020). [PubMed: 31980480]
8. McGinty JW et al. Tuft-cell-derived leukotrienes drive rapid anti-helminth immunity in the small intestine but are dispensable for anti-protist immunity. Immunity 52, 528–541 (2020). [PubMed: 32160525]

9. Nadjsombati MS et al. Detection of succinate by intestinal tuft cells triggers a type 2 innate immune circuit. Immunity 49, 33–41 (2018). [PubMed: 30021144]
10. Lei W et al. Activation of intestinal tuft cell-expressed Sucnr1 triggers type 2 immunity in the mouse small intestine. Proc. Natl Acad. Sci. USA 115, 201720758 (2018).
11. Rudin CM et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat. Rev. Cancer 19, 289–297 (2019). [PubMed: 30926931]
12. Yamaguchi T et al. Skn-1a/Pou2f3 is required for the generation of Trpm5-expressing microvillous cells in the mouse main olfactory epithelium. BMC Neurosci. 15, 13 (2014). [PubMed: 24428937]
13. Matsumoto I, Ohmoto M, Narukawa M, Yoshihara Y & Abe K Skn-1a (Pou2f3) specifies taste receptor cell lineage. Nat. Neurosci. 14, 685–687 (2011). [PubMed: 21572433]
14. Elmentaite R et al. Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. Dev. Cell 55, 771–783 (2020). [PubMed: 33290721]
15. Park J-E et al. A cell atlas of human thymic development defines T cell repertoire formation. Science 367, eaay3224 (2020). [PubMed: 32079746]
16. Haber AL et al. A single-cell survey of the small intestinal epithelium. Nature 551, 333–339 (2017). [PubMed: 29144463]
17. Plasschaert LW et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature 560, 377–381 (2018). [PubMed: 30069046]
18. Bornstein C et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature 559, 622–626 (2018). [PubMed: 30022162]
19. Schaum N et al. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. Nature 562, 367–372 (2018). [PubMed: 30283141]
20. Smillie CS et al. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. Cell 178, 714–730 (2019). [PubMed: 31348891]
21. Behan FM et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. Nature 568, 511–516 (2019). [PubMed: 30971826]
22. George J et al. Comprehensive genomic profiles of small cell lung cancer. Nature 524, 47–53 (2015). [PubMed: 26168399]
23. Luo Y, Fujii H, Gerster T & Roeder RG A novel B cell-derived coactivator potentiates the activation of immunoglobulin promoters by octamer-binding transcription factors. Cell 71, 231–241 (1992). [PubMed: 1423591]
24. Duclos GE et al. Characterizing smoking-induced transcriptional heterogeneity in the human bronchial epithelium at single-cell resolution. Sci. Adv. 5, eaaw3413 (2019). [PubMed: 31844660]
25. Chasman D, Cepek K, Sharp PA & Pabo CO Crystal structure of an OCA-B peptide bound to an Oct-1 POU domain/octamer DNA complex: specific recognition of a protein-DNA interface. Genes Dev. 13, 2650–2657 (1999). [PubMed: 10541551]
26. Gstaiger M, Georgiev O, Leeuwen H, van, Vliet, Pvander & Schaffner, W. The B cell coactivator Bob1 shows DNA sequence-dependent complex formation with Oct-1/Oct-2 factors, leading to differential promoter activation. EMBO J. 15, 2781–2790 (1996). [PubMed: 8654375]
27. Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F & Matthias P B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. Nature 383, 538–542 (1996). [PubMed: 8849727]
28. Cepek KL, Chasman DI & Sharp PA Sequence-specific DNA binding of the B-cell-specific coactivator OCA-B. Genes Dev. 10, 2079–2088 (1996). [PubMed: 8769650]
29. Luo Y & Roeder RG Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. Mol. Cell. Biol. 15, 4115–4124 (1995). [PubMed: 7623806]
30. Lim JS et al. Intratumoural heterogeneity generated by Notch signalling promotes small-cell lung cancer. Nature 545, 360–364 (2017). [PubMed: 28489825]
31. Andersen B et al. Functions of the POU domain genes Skn-1a/i and Tst-1/Oct-6/SCIP in epidermal differentiation. Genes Dev. 11, 1873–1884 (1997). [PubMed: 9242494]

32. Szczepanski AP, Tsuboyama N, Zhao Z & Wang L POU2AF2/C11orf53 functions as a co-activator of POU2F3 by maintaining chromatin accessibility and enhancer activity. Preprint at bioRxiv 10.1101/2022.03.17.484753 (2022).
33. Dailey L & Basilico C Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. J. Cell. Physiol. 186, 315–328 (2001). [PubMed: 11169970]
34. Malik V, Zimmer D & Jauch R Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming. Cell. Mol. Life Sci. 75, 1587–1612 (2018). [PubMed: 29335749]
35. Reményi A et al. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. Genes Dev. 17, 2048–2059 (2003). [PubMed: 12923055]
36. Chu C-S et al. Unique immune cell coactivators specify locus control region function and cell stage. Mol. Cell 80, 845–861 (2020). [PubMed: 33232656]
37. Ren X, Siegel R, Kim U & Roeder RG Direct interactions of OCA-B and TFII-I regulate immunoglobulin heavy-chain gene transcription by facilitating enhancer-promoter communication. Mol. Cell 42, 342–355 (2011). [PubMed: 21549311]
38. Montoro DT et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature 560, 319–324 (2018). [PubMed: 30069044]
39. Wolf FA, Angerer P & Theis FJ SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018). [PubMed: 29409532]
40. Stephens M False discovery rates: a new deal. Biostatistics 10.1093/biostatistics/kxw041 (2016).
41. Patro R, Duggal G, Love MI, Irizarry RA & Kingsford C Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419 (2017). [PubMed: 28263959]
42. Ashkenazy H et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–W350 (2016). [PubMed: 27166375]
43. Celniker G et al. ConSurf: using evolutionary data to raise testable hypotheses about protein function. Israel J. Chem. 53, 199–206 (2013).
44. Mészáros B, Erdős G & Dosztányi Z IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46, W329–W337 (2018). [PubMed: 29860432]
45. Romero P, Obradovic Z, Kissinger C, Villafranca JE & Dunker AK Identifying disordered regions in proteins from amino acid sequence. In Proc. International Conference on Neural Networks ICNN´97 Vol. 1, 90–95 (IEEE, 1997).
46. Romero P et al. Sequence complexity of disordered protein. Proteins Struct. Funct. Bioinform. 42, 38–48 (2001).
47. Waterhouse A et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303 (2018). [PubMed: 29788355]
48. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
49. Langmead B & Salzberg SL Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
50. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]
51. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008). [PubMed: 18798982]
52. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010). [PubMed: 20110278]
53. 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). [PubMed: 20513432]

54. Kent WJ et al. The Human Genome Browser at UCSC. Genome Res. 12, 996–1006 (2002). [PubMed: 12045153]

55. Bailey TL & Elkan C Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proc. International Conference on Intelligent Systems for Molecular Biology Vol. 2, 28–36 (CAAAAI Press, 1994).

56. Ramírez F et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016). [PubMed: 27079975]

57. Hunter JD Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9, 90–95 (2007).

58. Waskom M seaborn: statistical data visualization. J. Open Source Softw. 6, 3021 (2021).

59. Kim D, Langmead B & Salzberg SL HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015). [PubMed: 25751142]

60. Roe J-S et al. Enhancer reprogramming promotes pancreatic cancer metastasis. Cell 170, 875–888 (2017). [PubMed: 28757253]

61. Love MI, Huber W & Anders S Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014). [PubMed: 25516281]

62. Adrover JM et al. Programmed ‘disarming’ of the neutrophil proteome reduces the magnitude of inflammation. Nat. Immunol. 21, 135–144 (2020). [PubMed: 31932813]

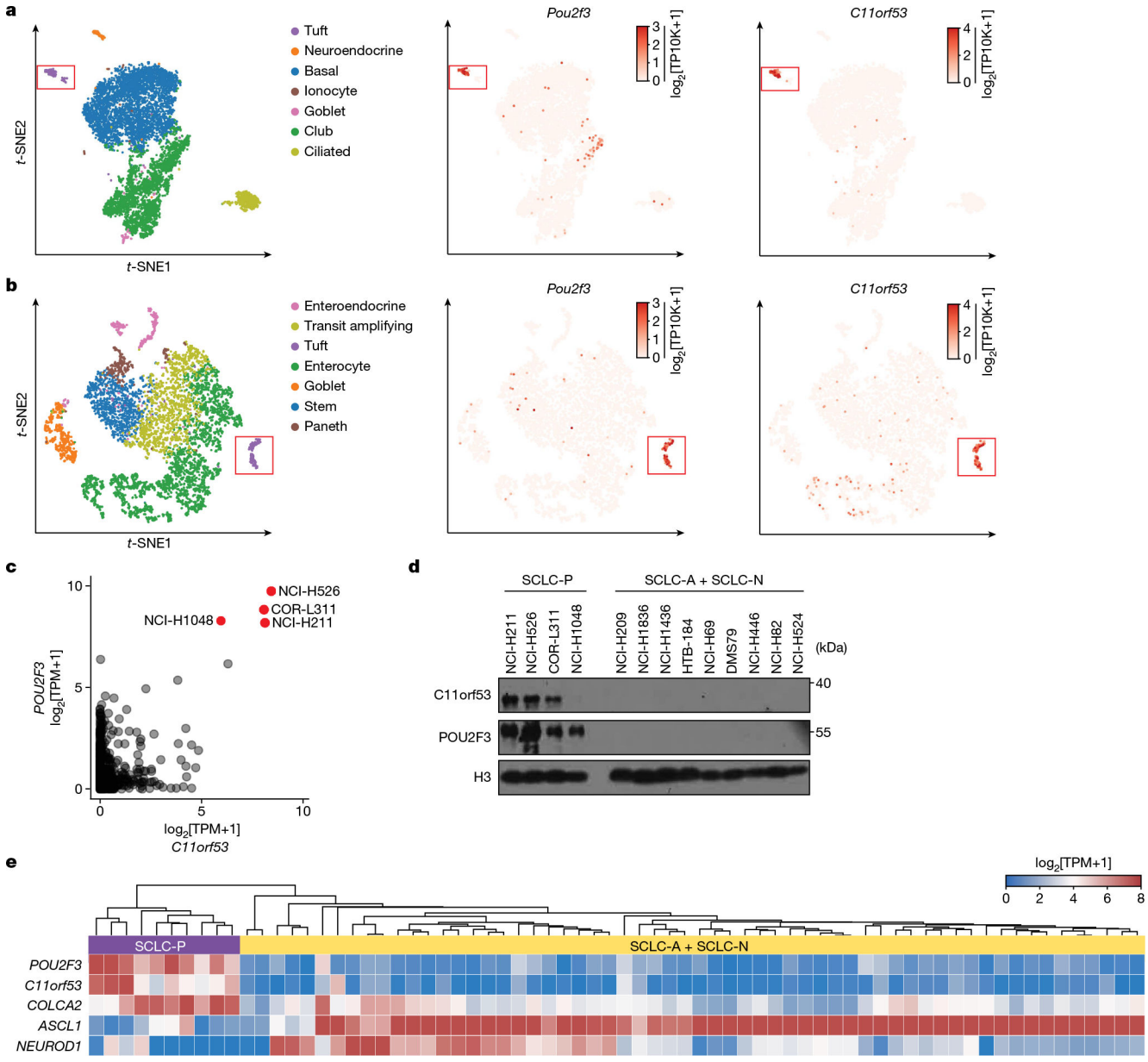


Fig. 1 |. *C11orf53* is selectively expressed in normal and malignant tuft cells.

a,**b**, scRNA-seq analysis of the mRNA levels of *Pou2f3* and *C11orf53* (mouse gene name: *1810046K07Rik*) in mouse trachea epithelium[38] (**a**) and small intestine epithelium[16] (**b**). *t*-Distributed stochastic neighbour embedding (*t*-SNE) analysis of 7,193 trachea epithelial cells (points) or 7,216 small intestine epithelial cells (points), coloured by cell type assignments or by the mRNA level ($\log_2$[transcripts per 10,000 reads (TP10K) + 1] of *Pou2f3* and *C11orf53*). **c**, mRNA level of *POU2F3* and *C11orf53* across 1,379 human cancer cell lines from the CCLE database. Data were obtained from Cancer Dependency Map (DepMap) portal (21Q2). TPM, transcripts per million. **d**, Western blot analysis of C11orf53 and POU2F3 in a panel of SCLC cell lines. SCLC-P, SCLC-POU2F3^{high}; SCLC-A, SCLC-ASCL1^{high}; SCLC-N, SCLC-NEUROD1^{high}. Data are representative of two independent experiments. **e**, mRNA levels of indicated genes from 72 patients with SCLC[22] that are

grouped on the basis of unsupervised clustering. Source data for western blots are provided in Supplementary Fig. 1a.

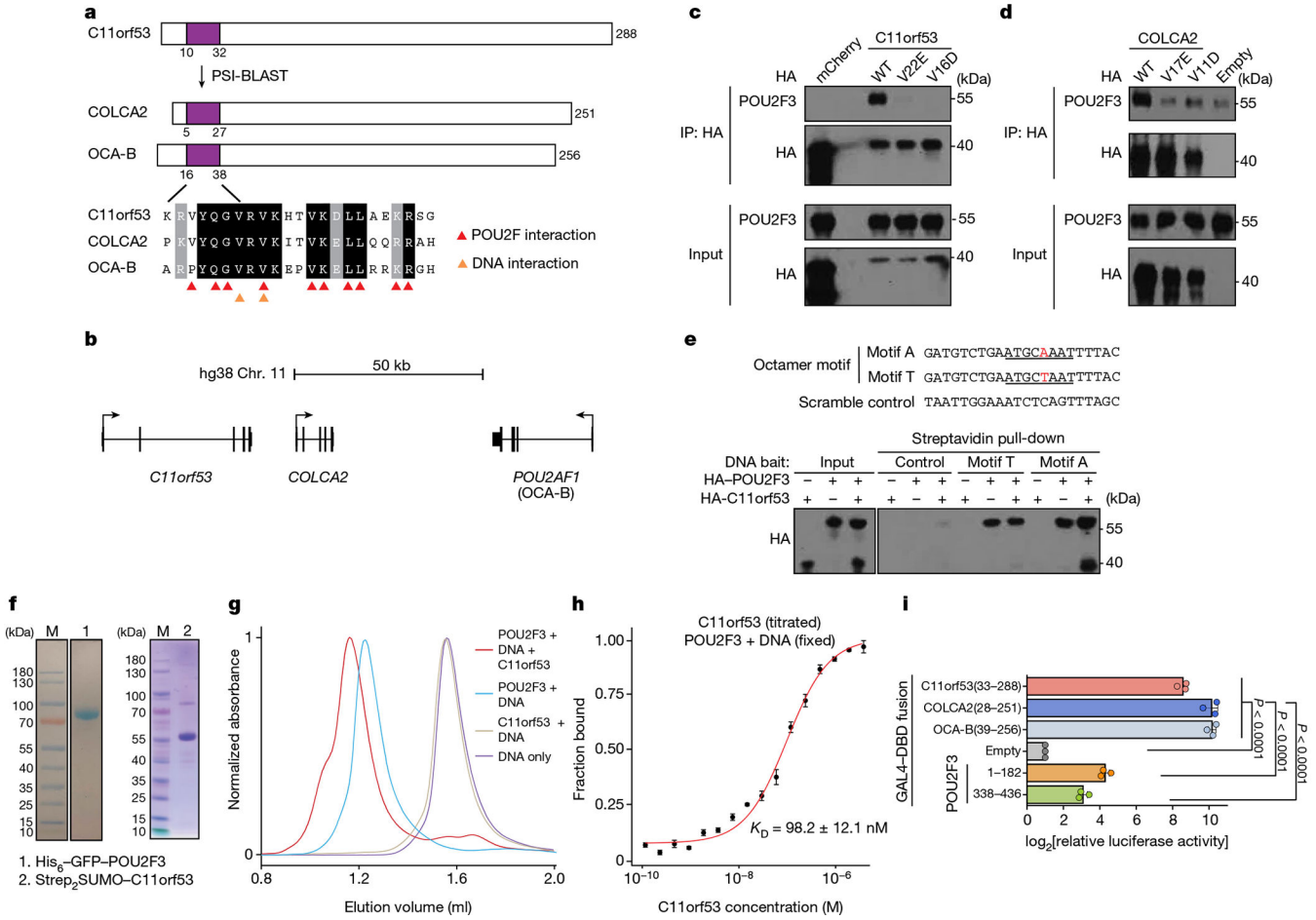


Fig. 2 |. C11orf53/OCA-T1 and COLCA2/OCA-T2 form an OCA-B-like complex with POU2F3 in a DNA-dependent manner.

a, PSI-BLAST search of the conserved peptide of C11orf53 identified OCA-B and COLCA2 as having homology in this region. The arrows indicate residues that are involved in interacting with DNA or POU2F1/OCT1 based on Protein Data Bank (PDB) 1CQT (ref. [25]). **b**, The *C11orf53*, *COLCA2* and *POU2AF1*/OCA-B gene cluster in the human genome (hg38). **c**, Immunoprecipitation (IP)–western blot analysis evaluating the interaction of HA–C11orf53 or HA–mCherry with endogenous POU2F3 in NCI-H211 cells. **d**, IP–western blot analysis evaluating the interaction between HA–COLCA2 and endogenous POU2F3 in NCI-H1048 cells. **e**, DNA pull-down assay evaluating POU2F3 and C11orf53 interactions with biotinylated octamer motif or control DNA. The octamer motif is underlined, bases in red highlight the difference between the two sequences. Immobilized DNA was incubated with cell lysates from transfected HEK293T cells. **f**, Purity assessment of recombinant His_6–GFP–POU2F3 and $Strep_2$SUMO–C11orf53 proteins by SDS–PAGE and Coomassie blue staining. M, marker. **g**, Analytical gel filtration of $Strep_2$SUMO–C11orf53, His_6–GFP–POU2F3 and octamer DNA motif A assemblies. The maximum absorbance at 260 nm for each injection was normalized to 1.0 for ease of comparison between samples. **h**, Microscale thermophoresis analysis of proteins mixed with octamer motif A. His_6–GFP–POU2F3 and DNA were assembled first in a test tube, followed by titration of $Strep_2$SUMO–C11orf53. *n*

= 3. Data are mean ± s.e.m. **i**, Luciferase reporter assay performed in transfected HEK293T cells to evaluate the transactivation activity of indicated segments of C11orf53, COLCA2, OCA-B or POU2F3 that are fused to GAL4 DNA-binding domain (DBD). $n = 3$. The luciferase activity of each construct was first normalized to the empty vector and then transformed ($\log_2$[normalized luciferase activity] + 1). Data are mean ± s.d. Statistical analysis was performed using two-tailed unpaired Student's *t*-tests, comparing the average of luciferase activity of C11orf53(33–288), COLCA2(28–251) and OCA-B(39–256) versus POU2F3(1–182), POU2F3(338–436) and empty vector. For the comparison to empty vector, Welch's correction was performed before the Student's *t*-test. Source data for gels are provided in Supplementary Figs. 1b,d–g and 7, and microscale thermophoresis analysis and GAL4 luciferase assay data are provided in Supplementary Table 1.

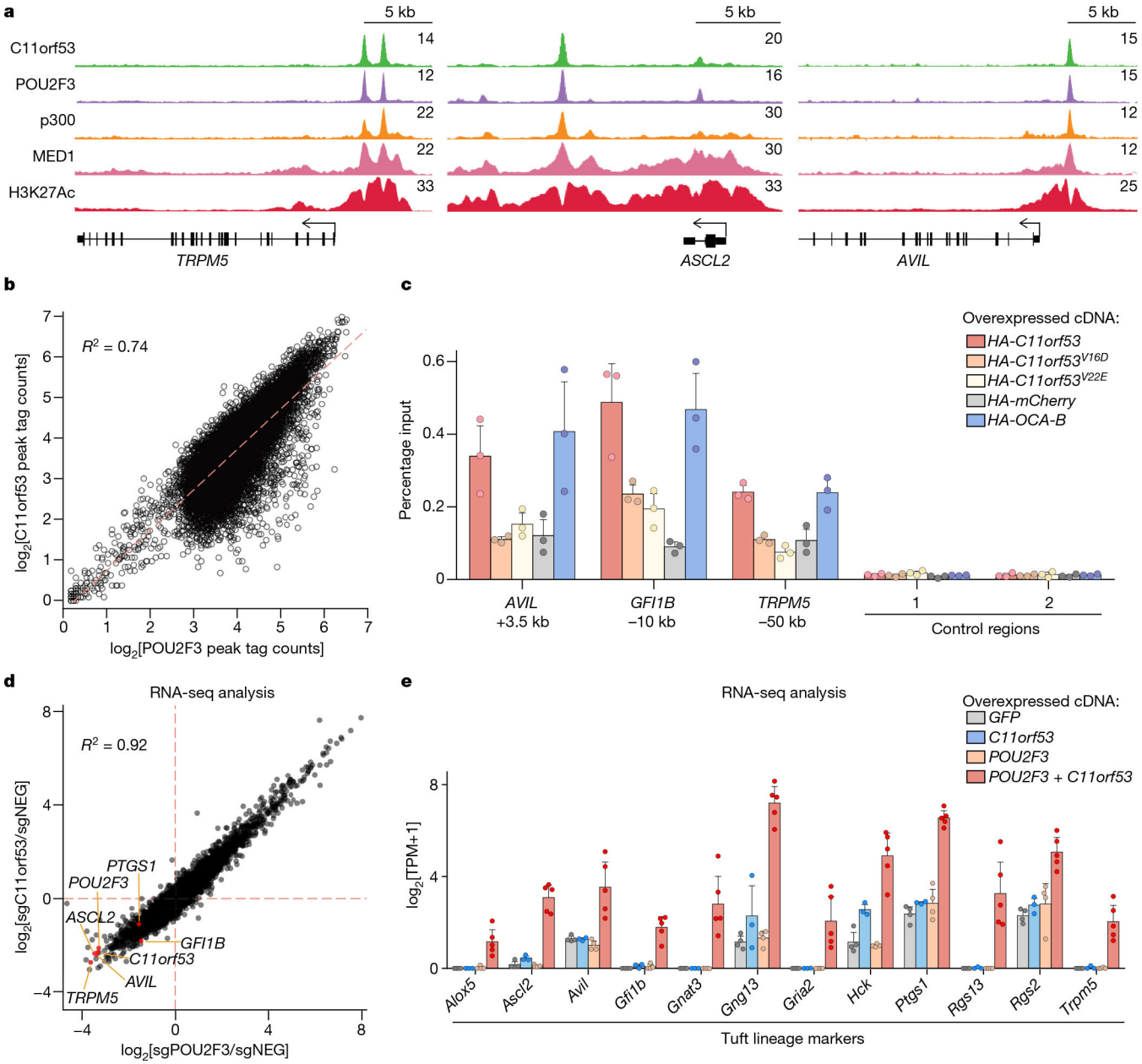


Fig. 3 |. he POU2F3–C11orf53/OCA-T1 complex drives the expression of tuft-cell-specific genes.

a, Chromatin immunoprecipitation sequencing (ChIP–seq) occupancy profiles of the indicated proteins at tuft-cell-specific loci. **b**, The correlation of peak tag counts of POU2F3 and C11orf53, within the genomic intervals of POU2F3 peaks (14,163) consistently identified from three replicates of POU2F3 experiments. $n = 3$. **c**, Anti-HA ChIP–quantitative PCR (qPCR) analysis of the indicated lentivirally expressed HA-tagged proteins. The ChIP–qPCR signal of each sample was normalized to its own input. Protein expression is shown in Extended Data Fig. 6d. Data are mean ± s.d. of three technical replicates. $n = 1$. **d**, RNA-seq analysis comparing mRNA changes after *C11orf53* (3 sgRNAs) or *POU2F3* (2 sgRNAs) knockout compared with control sgRNAs (2 sgRNAs) in NCI-H211 cells. RNA was collected 5 days after sgRNA infection. Each dot represents

the $\log_2$-transformed fold change of a single protein-coding gene. Low-expressed genes were filtered out before performing DESeq2 (read count = 10 as cut-off). Knockout of each protein is shown in Extended Data Fig. 6f. **e**, mRNA levels of the indicated tuft cell marker genes in YT330 mouse SCLC cells lentivirally transduced with the indicated cDNA. Protein expression is shown in Extended Data Fig. 6h. $n = 3$ biological replicates; *GFP* and *POU2F3* had one additional technical replicate and *POU2F3* + *C11orf53* had two additional technical replicates. Data are mean ± s.d. Source data for the gels are provided in Supplementary Fig. 1h,i, gene expression changes ($\log_2$-transformed fold change) after knockout or TPM of the indicated genes after overexpression are provided in Supplementary Table 2, primers and sgRNA sequences are provided in Supplementary Table 6.

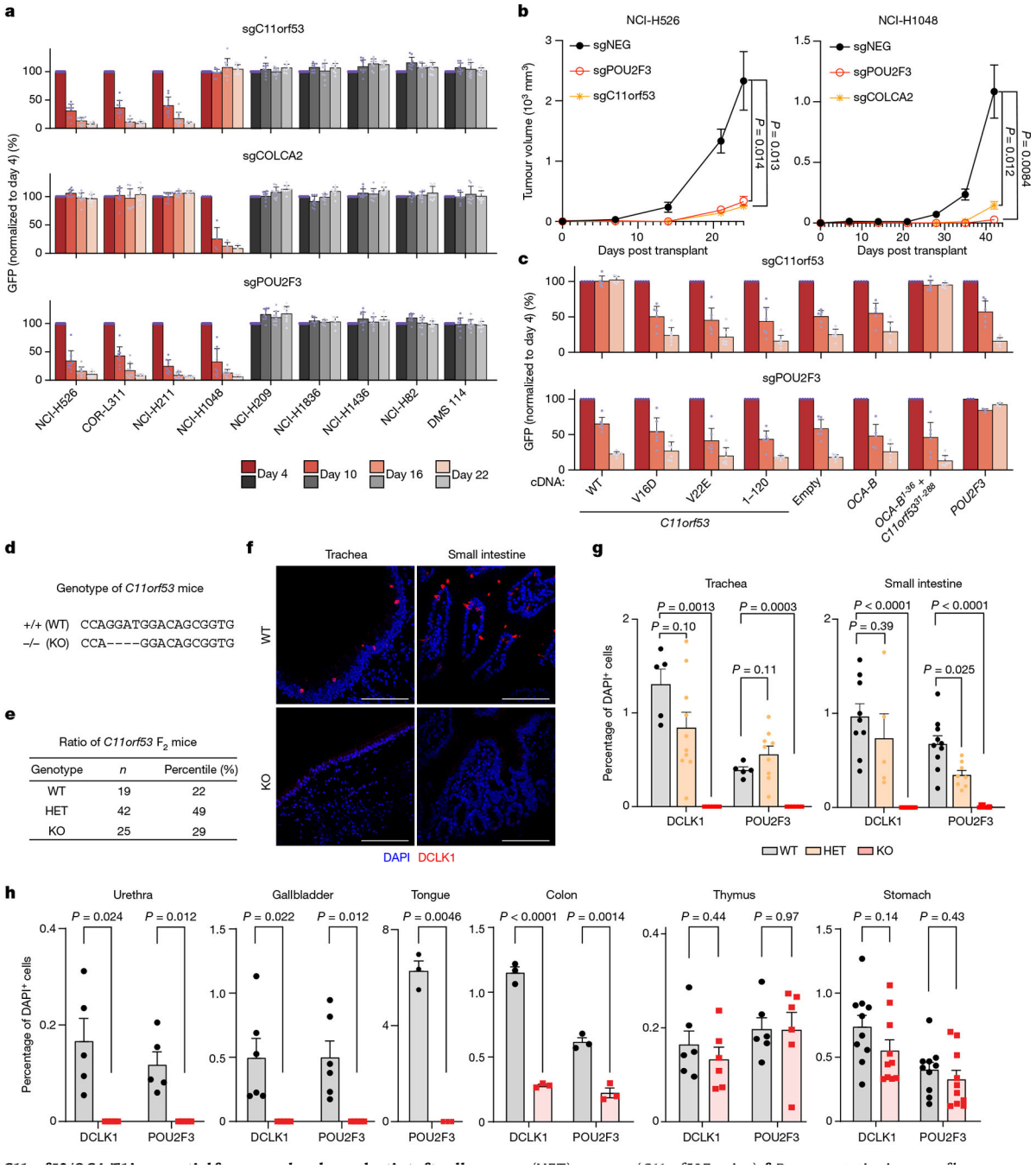


Fig. 4 |. C11orf53/OCA-T1 is essential for normal and neoplastic tuft cell development.

a, Competition-based proliferation assays in *Cas9*-expressing SCLC cells after lentiviral expression of the indicated sgRNAs (linked with GFP). Data are mean ± s.d. normalized percentage of GFP (to day 4 after infection) of two to three sgRNAs. $n = 3$. **b**, The growth kinetics of *Cas9*-expressing NCI-H526 and NCI-H1048 xenografts implanted in immunodeficient mice. Cells were lentivirally transduced with the indicated sgRNAs before injection into mice. $n = 5$, except for *C11orf53*, for which $n = 10$. Data are mean ± s.e.m. sgNEG, sgRNA targeting negative controls. *P* values were calculated using two-tailed

unpaired Student's *t*-tests with Welch's correction. **c**, Competition-based proliferation assays in NCI-H211 cells cotransduced with the indicated cDNAs and sgRNAs lentivirally to assess the functionality of the indicated mutants. cDNAs were engineered to be resistant to Cas9/sgRNA-mediated cutting. Data are mean ± s.d. of GFP percentage (normalized to day 4). Expression of mutants is shown in Extended Data Fig. 6d. **d**, CRISPR–Cas9-induced deletion of *C11orf53* to generate knockout mice. **e**, The ratio of different genotypes of *C11orf53*$^{+/-}$ (HET) progeny (*C11orf53* F_2 mice). **f**, Representative immunofluorescence staining of the tuft cell marker DCLK1 in the trachea and small intestine. Scale bars, 100 μm. **g**,**h**, Quantification of POU2F3-positive and DCLK1-positive cells. Data are mean ± s.e.m. The *P* values for the comparison of tuft cell frequencies of knockout versus WT (DCLK1/POU2F3) or HET versus WT of POU2F3 staining in small intestine, trachea, urethra, gallbladder and tongue were calculated using two-tailed unpaired Student's *t*-tests with Welch's correction; the *P* values for comparing tuft cell frequencies of knockout versus WT in colon, thymus and stomach, or HET versus WT of DCLK1 staining in the small intestine and trachea were calculated using ordinary unpaired two-tailed Student's *t*-tests. The numbers of animals involved in **f–h** are provided in Extended Data Fig. 9. Source data for GFP competition and gene complementary assays are provided in Supplementary Table 3 and mouse data are provided in Supplementary Tables 4 and 5. sgRNA sequences and genotyping primers are provided in Supplementary Table 6.