

Corresponding author(s):	James L. Reading, Sam M. Janes
Last updated by author(s):	17/08/2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	• BD FACSDiva for conventional flow cytometry using FACSymphony™ A5 High-Parameter Cell Analyser (BD Bioscience). • BD FACS Aria Fusion and BD Aria III for cell sorting (BD Bioscience). • Sony Biotechnology ID7000 software for spectral flow cytometry using ID7000 Spectral Cell Analyser (Sony Biotechnology). • Illumina NovaSeq 6000 for scRNA-seq using 10x Genomics kits. • Illumina MiSeq for TCR-seq using Adaptive Biotechnologies and Ampliseq kits. • Nanozoomer (Hamamatsu) for H&E, Leica DMI8 widefield microscope and Zeiss 880 confocal microscope for immunofluorescence imaging. • 10x Genomics Xenium platform using the Mouse 5K Pan-Tissues and Pathways panel for spatial transcriptomics.
Data analysis	Flow cytometry data were analysed using FlowJo v10.8.1 and R-based workflows. Immunofluorescence and histology images were analysed using NDP.view2 v2.9.29 and Fiji/ImageJ2 v2.3.0/1.53q. Statistical analyses and figure generation were performed in R v4.5.1 and GraphPad Prism v11.0.2. Custom R scripts used for data analysis and figure generation will be made publicly available at publication via GitHub (https://github.com/Reading-Lab/preinvasive-lusc-treg) and archived with a DOI through Zenodo (10.5281/zenodo.19599892). Package versions and computational environment details are provided in the accompanying code repository. Computational analyses were performed in R (RStudio IDE v2026.01.1+403). Flow cytometry data were processed and analysed using FlowAI (v1.24), CATALYST (v1.32.1), FlowSOM (v2.2.0) and ConsensusClusterPlus (v1.58.0). Single-cell RNA-seq data were processed using Cell Ranger (v7.1.0). Single-cell RNA-seq data and Xenium spatial transcriptomic were analysed using Seurat (v4.4.1), Azimuth (v0.5.0), SingleCellExperiment (v1.30.1), BPCells (v0.3.0), Harmony (v1.2.3), celda (v1.26.0) and glmGamPoi (v1.22.0). Differential expression, gene-set and signature analyses used edgeR (v4.8.2), limma (v3.66.0), fgsea (v1.34.2) and UCell (v2.12.0). Differential abundance, cell-cell communication and trajectory analyses used miloR (v2.4.1), CellChat (v2.2.0) and Monocle3 (v1.3.4). Statistical and survival analyses used lme4 (v1.1.37), survminer (v0.5.1) and base R functions. Bulk RNA-seq data processing was done using nf-core/maseq (v3.12.0); FastQC (v0.11.9); Trim Galore (v0.6.7); STAR (v2.7.10a); Salmon (v1.10.1); MultiQC (v1.14). Data visualisation used ggplot2 (v4.0.1), ggpubr (v0.6.2), ggrepel (v0.9.6), ggbeeswarm (v0.7.3), ggridges (v0.5.7), ggtern (v4.0.0), scales (v1.4.0), patchwork (v1.3.2), cowplot (v1.2.0), pheatmap (v1.0.13), ComplexHeatmap (v2.24.1) and circlize (v0.4.16).

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing data are available on the European Genome-phenome Archive (EGA) (UCLH Bronchoscopy Surveillance study, tissue and PBMC scRNA+TCR-seq: EGAD50000002939; ASCENT bulk RNA-seq: EGAD50000002852; ASCENT scRNA+TCR-seq: EGAD50000002930; SUMMIT bulk TCR-seq: EGAD50000002950), European Nucleotide Archive ENA (mouse tissue and peripheral blood scRNA+TCR-seq: ERP136782, ERP203132), Human Tumor Atlas Network (HTAN) (PCGA scRNA-seq: dbGaP phs002371). Mouse Xenium spatial transcriptomics data are available from the BioImage Archive under accession S-BIAD3880. The processed data objects and associated metadata are available on Zenodo ([10.5281/zenodo.19599892](https://doi.org/10.5281/zenodo.19599892)). Source data are provided with this paper.

Publicly available datasets reanalysed in this study comprise the bronchial-biopsy transcriptomic dataset from NCBI Gene Expression Omnibus (GEO; GSE33479; [GEO Accession viewer](#)); the preinvasive and early lung adenocarcinoma scRNA-seq dataset from GEO (GSE189357; [GEO Accession viewer](#)); and the LuCA NSCLC single-cell atlas, available from Zenodo under <https://doi.org/10.5281/zenodo.7227571> and from CELLxGENE ([Cellxgene Data Portal](#)). For the publicly available Samsung Medical Center immunotherapy cohort analysed in Fig. 1J, processed transcriptomic data are available from GEO under accession GSE218989 ([GEO Accession viewer](#)), and the corresponding raw sequencing data are available under controlled access from EGA under study accession EGAS50000000324 and dataset accession EGAD50000000469. Clinical response and tumour mutational burden annotations were obtained from the accompanying supplementary data of Shim *et al.* (<https://doi.org/10.1016/j.annonc.2020.04.004>). The Human Lung Cell Atlas reference used for cell-type annotation is available from Zenodo under <https://doi.org/10.5281/zenodo.6337966> and from CELLxGENE ([Cellxgene Data Portal](#)). Human and mouse single-cell sequencing reads were aligned using the 10x Genomics 2020-A reference packages refdata-gex-GRCh38-2020-A (GRCh38; GENCODE v32/Ensembl 98) and refdata-gex-mm10-2020-A (GRCm38/mm10; GENCODE vM23/Ensembl 98), respectively, available from <https://www.10xgenomics.com/support/software/cell-ranger/downloads>. ASCENT bulk RNA-seq reads were aligned to the *Homo sapiens* GRCh38 reference genome using Ensembl release 112 gene annotation.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\)](#), and [sexual orientation and race, ethnicity and racism](#).

Reporting on sex and gender	Sex was considered as a covariate in the analyses where it is mentioned in the manuscript. Gender was not collected or considered in this study. Sex of participants was determined based on their medical record files.
Reporting on race, ethnicity, or other socially relevant groupings	Race, ethnicity, or other socially relevant groupings were not used in any of the analyses included in the manuscript.
Population characteristics	Participants were drawn from related but clinically distinct prospective, observational cohorts comprising lung cancer screening, bronchoscopic surveillance, surgical resection, and control populations.

	The UCLH Bronchoscopy Surveillance cohort comprised adults (≥ 18 years) with histologically confirmed central airway preinvasive lesions, predominantly individuals at high risk of lung cancer with substantial smoking exposure. Lesions ranges from low grade (squamous metaplasia, and mild and moderate dysplasia) to high grade (severe dysplasia, carcinoma in situ, and microinvasive carcinoma in situ). Participants with invasive lung cancer at recruitment were excluded, although those previously treated curatively for lung cancer could participate in the absence of recurrent disease. Relevant clinical characteristics (age, sex, smoking status) of patients from this cohort who contributed samples are listed in Supplementary Table 1. SUMMIT comprised individuals (≥ 18 years) identified through NHS primary-care smoking records who met prespecified age and lung-cancer-risk criteria for low-dose CT screening. ALPINE comprised current or former smokers participating in the North Central London NHS Targeted Lung Health Check programme, with the ALPINE participants selected as controls for this manuscript having no pulmonary nodules on imaging at the time of sampling. ASCENT comprised participants undergoing surgical resection for screen-detected suspected NSCLC. Analysed cancers included lung adenocarcinoma and lung squamous-cell carcinoma, with stages I–III represented and most cases being stage I. Participants selected for neoadjuvant treatment for their current lung malignancy were excluded. Additional blood controls comprised age-matched adults recruited before hip-replacement surgery who had no active infection, history of malignancy, autoimmune condition or abnormal full blood count, together with healthy NHS Blood and Transplant platelet donors. The Molecular Pathogenesis of Lung Disease II cohort comprised adults (≥ 18 years) undergoing investigation or treatment for respiratory disease, including thoracic surgery, bronchoscopy, radiology-guided biopsy, and other respiratory diagnostic procedures. For this study, only participants with a confirmed diagnosis of lung cancer (stage I–III, most cases used in this study being stage I) underwent surgical resection were included, contributing tumour and matched adjacent lung parenchyma samples for analysis. For the above studies, the clinical characteristics that were available and used in the manuscript varied by dataset and analysis. Samples were excluded if they had undergone prior immunotherapy. These included age, smoking status, cumulative tobacco exposure in pack-years, previous cancer diagnosis, bronchial lesion grade, tumour stage. The covariates included in individual statistical models are specified in the Methods and figure legends; technical batch was included as an analytical covariate where appropriate. The available participant-level clinical variables for each analysed dataset are documented in the associated metadata deposited on Zenodo. No germline genotypic information was used for recruitment, participant stratification or covariate adjustment. Participants were drawn from related but clinically distinct prospective, observational cohorts comprising lung cancer screening, bronchoscopic surveillance, surgical resection, and control populations in London, as described above. The UCLH Bronchoscopy Surveillance study recruited consenting adults at UCLH after identification of a preinvasive bronchial lesion. Participation required the ability to undergo autofluorescence bronchoscopy and biopsy. In this study, individuals with respiratory or other medical conditions preventing bronchoscopy, coagulation abnormalities, or inability to provide informed consent were excluded. For SUMMIT, potentially eligible individuals were identified through smoking information recorded in NHS primary care records from participating practices in north central and northeast London and invited to attend a lung health check. Those meeting the study's age, smoking and lung-cancer-risk criteria were invited to consent to low-dose CT screening. ASCENT recruited participants with screen-detected suspected lung cancer from the linked screening pathway who were scheduled to undergo surgical resection. ALPINE recruited current or former smokers attending the North Central London NHS Targeted Lung Health Check programme, with the no-nodule subset selected as screening controls for the present analyses. Additional controls were recruited through a UCLH orthopaedic pre-assessment clinic or obtained from residual NHS Blood and Transplant platelet-donation leukocyte cones. For Molecular Pathogenesis of Lung Disease II, eligible patients were recruited from individuals undergoing a range of respiratory diagnostic and clinical procedures, including bronchoscopy, upper airway endoscopy, thoracic surgery (including VATS and surgical resection), radiology-guided biopsy, respiratory outpatient clinics, or inpatient admissions. Eligible participants are adults (≥ 18 years) undergoing investigation for respiratory symptoms or follow-up of respiratory disease. For this manuscript, only tumour and matched adjacent parenchyma from participants with a confirmed diagnosis of lung cancer (stage I–III, most samples used in the manuscript were stage I) were included. These recruitment strategies introduce several potential sources of selection bias. All participation was consent-based and may favour individuals willing and able to attend screening, undergo longitudinal sampling, or undergo invasive procedures. SUMMIT ascertainment depended on smoking information recorded in primary-care systems and on attendance after invitation, potentially excluding otherwise eligible people with incomplete records or those less likely to engage with screening. SUMMIT and ALPINE were deliberately enriched for older current or former smokers at elevated lung-cancer risk and therefore do not represent younger people, never-smokers or lower-risk populations. The Molecular Pathogenesis of Lung Disease II study is a hospital-based, surgically recruited cohort and therefore enriches for patients with operable disease, good performance status, and earlier-stage tumours. Consequently, findings may not be fully generatable to patients with advanced or unresectable lung cancer or to the wider NSCLC population. The bronchoscopy cohort was additionally selected for known, bronchoscopically detectable central-airway premalignant disease and adequate fitness for repeated bronchoscopy. ASCENT was selected for operable, predominantly early-stage, screen-detected disease in participants fit for surgery and not assigned neoadjuvant treatment; it therefore does not represent advanced, symptomatic, unresectable or medically inoperable NSCLC. Selecting ALPINE controls without any pulmonary nodules may produce a particularly healthy screening comparator group. The orthopaedic and blood-donor controls were also health-selected and may differ from the cancer and premalignancy groups in smoking exposure, comorbidity, medication use and inflammatory status. Recruitment was concentrated in London, and inclusion in individual analyses also depended on sample availability, successful collection and assay quality control. These factors may affect absolute immune-cell frequencies and limit demographic and clinical generalisability. Concordant findings across complementary surveillance, screening, surgical, and control populations reduce dependence on a single sampling context, but do not eliminate the shared geographic, smoking-related and health-selection biases or the partially linked recruitment pathways between SUMMIT, ALPINE, ASCENT and Molecular Pathogenesis of Lung Disease II.
Recruitment	
Ethics oversight	UCLH bronchoscopy study, SUMMIT, and Molecular Pathogenesis of Lung Disease II were approved by the Healthy Research Authority (HRA, England). ASCENT was approved by the HRA and Health and Care Research Wales (HCRW).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Human No formal sample size calculation was performed because this was an observational study using prospectively collected clinical samples with finite availability. Sample sizes were guided by prior cross-sectional and paired studies of patients with NSCLC, in which cohorts of approximately 20–30 patients were sufficient to detect reproducible differences in blood and tumour-infiltrating T cell phenotypes. All available eligible samples were included and are detailed in the corresponding CONSORT-style flow diagrams. Statistical significance was assessed using univariable and multivariable analyses that accounted for relevant clinical and technical covariates and, where appropriate, adjusted for study, repeated measures and multiple testing. The observed effect sizes and statistical significance across independent cohorts further support that the sample sizes were sufficient to detect the biological effects reported. Animal studies For intercession PI3K inhibitor animal studies, a sample size calculation was performed based on previously observed difference in abundance of tumours in NTCU versus control, and were calculated using the <code>pwr.t.test</code> and <code>pwr.anova.test</code> functions from the Bioconductor <code>pwr</code> package using a Cohens <i>f</i> statistic of 0.75 and a Cohens <i>d</i> statistic of 1.5. No formal sample size calculation was performed for the exploratory immune profiling experiments. Sample sizes in longitudinal, cross-sectional and multi-tissue characterisation of the immune landscape were informed by robust immune phenotypes observed in orthogonal human datasets, and the lower expected biological variability of genetically homogeneous mouse cohorts relative to human clinical samples, whilst also minimising animal use in accordance with the principles of animal reduction. Human analyses within this study detected significant longitudinal immune changes using cohorts of five individuals and significant single-cell transcriptomic and TCR repertoire differences using five human samples; accordingly, comparable or larger mouse cohorts were used for profiling experiments. Sample size was also adjusted to adhere to the 3R guidelines of the local Ethics Committee of the Office for Veterinary Affairs. Exact <i>n</i> numbers for each experiment and/or analysis are included in the figure legends and/or methods.
Data exclusions	Human Data was excluded for failing pre-established QC metrics, outlined in consort diagrams in the relevant Extended Data figures/Supplementary tables, and Methods section.

	Animal studies For mouse flow cytometry experimental samples were excluded for analysis based on the following criteria: less than 50 cells in the parent population. Human All human work was validated on either a second external dataset, publicly available scRNA-seq and TCR-seq, and/or on a similar study for flow cytometry. Details of replicates can be found in the main/extended figure legends.
Replication	Animal studies NTCU treatment to study immune cell changes during pre-invasive disease development in the lungs was repeated twice independently (cohort 1 - week 11 and week 24, cohort 2 - week 15 and week 18). Draining and non-draining lymph node analyses were also repeated twice independently, where flow cytometry was performed and replicated. Longitudinal PBMC spectral flow cytometry was also repeated twice independently and replicated. The interception experiment utilising BYL719 and PI-3065 was performed once. Samples for murine scRNA-seq/TCR-seq were collected across several cohorts for different timepoints and aggregated for analyses (cohort 1 – week 15, cohort 2 – week 18, cohort 3 – week 24 for lung data, cohort 1 – week 18, cohort 2 – week 24 for PBMC). Details of replicates can be found in the main/extended figure legends. Human To minimise experimental bias and batch effects, where possible case and control samples were evenly distributed across processing batches. To ensure that batch effect did not confound the data, where appropriate batch was used as a term in covariate analysis.
Randomization	Animal studies For animal studies, following transfer from Charles River and an appropriate acclimatisation period, mice were weighed and cages assigned to control or NTCU groups, ensuring comparable initial average weights among groups. For FTY720/anti-CD25/anti-CTLA-4 experiments, cages were randomly assigned to receive treatment or vehicle control. Similarly, for PI3K delta inhibition studies, BYL719, PI3065, or control diets were randomly assigned within the NTCU-treated arm at week 15 (post-NTCU initiation). Experiments were performed using female FVB/N mice from an inbred strain, minimising biological variability between experimental groups. Human To minimise individual bias, samples were numerically coded for thawing and flow staining (e.g. 1-100). No other blinding was possible/performed.
Blinding	Animal studies Animal studies to assess effects of PI3K inhibitors on NTCU-induced carcinogenesis were conducted with researchers blinded to treatment throughout data collection and analysis of tumour incidence and size. For experiments involving FTY720/anti-CD25/anti-CTLA-4 experiments, investigators weren't blinded because active dosing, health surveillance, and body weight monitoring of mice required knowledge of treatment allocation. To minimise individual bias, intraperitoneal (i.p.) injections and health assessments were distributed randomly among several researchers. Furthermore, to ensure objective reporting the downstream data analysis and outcome assessments were performed with mice numerically coded (e.g. 1-12) instead of explicitly labelled with treatment groups.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☒ No

Field work, collection and transport

Field conditions	
Location	
Access & import/export	
Disturbance	

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- n/a ☐ Involved in the study
- ☐ ☒ Antibodies
- ☒ ☐ Eukaryotic cell lines
- ☒ ☐ Palaeontology and archaeology
- ☐ ☒ Animals and other organisms
- ☐ ☒ Clinical data
- ☒ ☐ Dual use research of concern
- ☒ ☐ Plants

Methods

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

Antibodies used

Anti-human antibodies for human flow cytometry experiments. Details of which are used in which panels/experiments are detailed in the Supplementary Information: CD19, clone LT19, SB UV445 (Bio-Rad, MCA1940SBUV445) CD8, clone RPA-T8, BUV496 (BD, 612942) CD45RA, clone HI100, BUV563 (BD, 612926) HLA-DR, clone WR18, SB UV605 (Bio-Rad, MCA477SBUV605) 4-1BB, clone 4B4-1, BUV661 (BD, 741642) NKG2A, clone 131411, BUV737 (BD, 749682) CD3, clone SK7, BUV805 (BD, 612893) PD-1, clone EH12.2H7, BV421 (BioLegend, 329920) IgM, clone G20-127, V450 (BD, 561286) CD103, clone Ber-ACT8, BV480 (BD, 746472) CD57, clone IA6QA17A04, BV510 (BioLegend, 393313) CD38, clone AT13/5, SB Violet570 (Bio-Rad, MCA1019SBV570) TIGIT, clone TGMAB-2, BV605 (BD, 570447) CD39, clone A1, BV711 (BioLegend, 328228) CD45, clone 2D1, BV750 (BioLegend, 368542) GITR, clone V27-580, BV785 (BD, 747661) NKG2D, clone 1D11, RB545 (BD, 570932) CD25, clone MEM-181, SB Blue675 (Bio-Rad, MCA2127SBB675) ICOS, clone DX29, RB705 (BD, 757469) NKG2C, clone 134591, RB744 (BD, 756942) CD4, clone SK3, PerCP-Fire780 (BioLegend, 344698) IgD, clone W18340F, PerCP-Fire806 (BioLegend, 307818) CD1d tetramer, clone n/a, PE (NIH, n/a) CCR6, clone 433H, RY586 (BD, 753833) CTLA-4, clone REA1003, PE-Vio615 (Miltenyi Biotec, 130-130-454) TCRγδ, clone B1, PE-Fire700 (BioLegend, 331238) KLRG1, clone SA231A2, PE-Fire744 (BioLegend, 367751) CD161, clone REA631, PE-Vio770 (Miltenyi Biotec, 130-114-123) OX40, clone Ber-ACT35 (ACT35), PE-Fire810 (BioLegend, 350043) MR1 tetramer, clone n/a, APC (NIH, n/a) Foxp3, clone 206D, Spark NIR685 (BioLegend, 320130) CD177, clone REA258, APC-Vio770 (Miltenyi Biotec, 130-101-514) CD27, clone O323, APC-Fire810 (BioLegend, 393214) OX40, clone L106, BV605 (BD, 745217) 4-1BB, clone 4B4-1, BV750 (BioLegend, 309843) GITR, clone V27-580, BB515 (BD, 567781) TCF1/7, clone 7F11A10, PE (BioLegend, 655208) Foxp3, clone 259D/C7, PE-CF594 (BD, 562421) Helios, clone D8W4X, PE-Cy7 (Cell Signaling Technology, 45259) BATF, clone REA486, APC (Miltenyi Biotec, 130-107-620) Live/Dead, clone n/a, APC-Cy7 (Thermo Fisher, L34994) CD8, clone SK1, BUV496 (BD OptiBuild, 741199) CD4, clone OKT4, BUV615 (BD OptiBuild, 750975) CD101, clone V7.1, BUV661 (BD OptiBuild, 750530) CD38, clone HB-7, BUV737 (BD OptiBuild, 741902) CD103, clone Ber-ACT-8, BUV805 (BD OptiBuild, 748501) Foxp3, clone 206D, BV421 (BioLegend, 320124) Live/Dead Yellow, clone n/a, BV570 (Thermo Fisher, L34968) CD57, clone QA17A04, BV605 (BioLegend, 393304) CCR7, clone G043H7, BV650 (BioLegend, 353234) CD27, clone M-T271, BV750 (BioLegend, 356430) TIM-3, clone F38-2E2, BV786 (BioLegend, 345032) PD-1, clone EH12.1, BB515 (BD, 564494) Granzyme B, clone GB11, PE-CF594 (BD, 562462) CXCR4, clone 12G5, PE-Cy5 (BioLegend, 306508) TIGIT, clone A15153G, PE-Cy7 (BioLegend, 372714) TOX, clone REA473, APC (Miltenyi Biotec, 130-118-335) HLA-DR, clone G46-6, APC-R700 (BD, 565127) CD3, clone UCHT1, APC-Fire750 (BioLegend, 300470) Live/Dead Blue, clone n/a, LD Blue (Thermo Fisher, L23105) CD95, clone DX2, BUV615 (BD, 752346) CD39, clone Tu66, BUV737 (BD, 612852) Granzyme B, clone GB11, Pacific Blue (BioLegend, 515408) CD14, clone M5E2, V500 (BD, 561391) CD19, clone HIB19, BV510 (BioLegend, 302242) CD27, clone O323, BV605 (BioLegend, 302830) CD11b, clone D12, BV750 (BD OptiBuild, 747210) CXCR4, clone 12G5, BV785 (BioLegend, 306530) Granzyme K, clone GM26E7, FITC (BioLegend, 370508) TCRγδ, clone 11F2, BB700 (BD, 745944) EOMES, clone WD1928, PerCP-eFluor710 (Thermo Fisher, 46-4877-42) CD4, clone OKT4, BB790 (Custom, 624296) CD56, clone MEM-188, PE-Cy5 (BioLegend, 304608) CD16, clone 3G8, PE-Fire640 (BioLegend, 302068) T-bet, clone 4B10, AF660 (Thermo Fisher, 606-5825-82) HLA-DR, clone L243, Alexa Fluor 700 (BioLegend, 307626) CD57, clone REA769, APC-Vio770 (Miltenyi Biotec, 130-116-503) CD38, clone HIT2, APC-Fire810 (BioLegend, 303550) Live/Dead, clone n/a, APC-eFluor780 (Invitrogen eBioscience, 65-0865-18) TCF1/7, clone S33-966, AF488 (BD, 567018) Granzyme B, clone QA16A02, AF700 (BioLegend, 372222) Ki-67, clone B56, BUV395 (BD, 564071) Viability, clone n/a, Zombie NIR (BioLegend, 423106) Mouse antibodies used for flow cytometry experiments: CD4, clone GK1.5, BUV496 (BD, 612952) CD44, clone IM7, BUV563 (BD, 741227) GITR, clone DTA-1, BUV615 (BD, 751532) CD1d, clone 1B1, BUV661 (BD, 741530) CD1d, clone 1B1, BV711 (BD, 740711) TREM2, clone 6E9, PE (BioLegend, 824805) ST2, clone U29-93, BUV661 (BD, 750671) KLRG1, clone 2F1, BUV737 (BD, 741812) CD3, clone 145-2C11, BUV805 (BD, 749276) PD-1, clone RMP1-30, BV421 (BioLegend, 109121) IgD, clone 11-26c.2a, Pacific Blue (BioLegend, 405712) CD23, clone B3B4, BV510 (BioLegend, 101623) CD27, clone L128, BV605 (BD, 562656) TIGIT, clone 1G9, BV650 (BD, 744213) OX40, clone OX-86, BV711 (BD, 740754) IgM, clone II/41, BV750 (BD, 747333) CD62L, clone MEL-14, BV785 (BioLegend, 104440) CD8, clone 53-6.7, Spark Blue 550 (BioLegend, 100780) CD45, clone YW62.3, SB Blue 580 (Bio-Rad, MCA1031SBB580) MHC II, clone M5/114.15.2, NovaFluor B660 (Thermo Fisher, M024T02B08) CD21, clone 7G6, BB700 (BD, 742139) Nkp46, clone 29A1.4, PerCP-eFluor 710 (eBioscience, 46-3351-82) TCRγδ, clone GL3, RB780 (BD, 755465)
--

	CD1d-PBS57 tetramer, clone n/a, PE (NIH, n/a) Foxp3, clone 3G3, PE-CF594 (BD, 567456) CD39, clone Duha59, PE-Fire 640 (BioLegend, 567456) ICOS, clone 15F9, PE-Fire 700 (BioLegend, 107716) 4-1BB, clone REA936, PE-Vio770 (Miltenyi Biotec, 130-115-457) CTLA-4, clone UC10-4B9, PE-Fire 810 (BioLegend, 106335) MR1-5-OP-RU tetramer, clone n/a, APC (NIH, n/a) CCR8, clone SA214G2, Spark NIR685 (BioLegend, 150324) CD25, clone PC61, APC/Fire 750 (BioLegend, 102053) RORyt, clone Q31-378, RB613 (BD, 571115) Nrp1, clone V46-1954, RB670 (BD, 770399) CD69, clone H1.2F3, AF700 (BD, 104539) GATA3, clone L50-823, PE-Cy7 (BD, 560405) CD38, clone 90, APC-Fire 810 (BioLegend, 102745) CD11c, clone N418, SparkPLUS UV395 (BioLegend, 117378) CD19, clone 6D5, SBUV445 (Bio-Rad, MCA1439SBUV445) CD45, clone 30-F11, BUV563 (BD, 612924) CD26, clone H194-112, BUV661 (BD, 741492) PD-L1, clone 10F.9G2, BUV737 (BD, 568826) CD11b, clone M1/70, BUV805 (BD, 568345) CD64, clone X54-5/7.1, BV421 (BioLegend, 139309) PDCA-1, clone 927, Pacific Blue (BioLegend, 127018) CD103, clone M290, BV480 (BD, 566118) Ly6G, clone 1A8, BV570 (BioLegend, 127629) CD40, clone 3/23, BV605 (BD, 745218) CD40, clone 3/23, RB744 (BD, 758113) CCR7, clone 4B12, BV650 (BioLegend, 120137) CD24, clone M1/69, BV750 (BD, 747362) Siglec-F, clone E50-2440, BV786 (BD, 740956) MerTK, clone bb14-16, BB515 (BD, 570399) CD3, clone 17A2, Spark Blue 550 (BioLegend, 100260) CD43, clone S7, BV750 (BD, 747277) CD45.1, clone A20, RB613 (BD, 759389) I-A/I-E (MHC II), clone M5/114.15.2, PerCP (BioLegend, 107624) CD172a, clone P84, RB780 (BD, 755521) CD163, clone S15049F, PE (BioLegend, 155327) CD163, clone S15049F, BV605 (BioLegend, 333616) F4/80, clone T45-2342, BUV615 (BD, 751454) CD86, clone GL1, RY610 (BD, 759497) CXCR4, clone 2B11/CXCR4, RY703 (BD, 771126) XCR1, clone REA707, PE-Vio770 (Miltenyi Biotec, 130-111-186) CD135, clone A2F10, APC (BioLegend, 135310) CD16, clone S17014E, APC (BioLegend, 158006) CD80, clone 16-10A1, Spark NIR685 (BioLegend, 104762) Ly6A, clone E13-167.7, APC-Fire 750 (BioLegend, 122526) Ly6C, clone HK1.4, APC-Fire 810 (BioLegend, 128056) CD23, clone B3B4, BUV661 (BD Biosciences, 741472) CD69, clone H1.2F3, BUV737 (BD Biosciences, 612793) Ki-67, clone B56, BV421 (BD Biosciences, 562899) CD21, clone 7G6, BV480 (BD Biosciences, 746306) CD19, clone 6D5, BV605 (BioLegend, 115539) TCRβ, clone H57-597, BV711 (BioLegend, 109243) CD25, clone PC61, PE (BioLegend, 102007) CD44, clone IM7, PE-Dazzle 594 (BioLegend, 103056) B220, clone RA3-6B2, PE-Cy7 (BioLegend, 103222) Foxp3, clone FJK-16s, APC (Invitrogen, 17-5773-82) Mouse antibodies used for immunofluorescence experiments: KRT5 (BioLegend, 905901) CD4 (Abcam, ab183685) FOXP3 (ThermoFisher, 14-5773-82) CD8a (ThermoFisher, 14-0081-82)
Validation	All new antibodies were titrated to determine the optimal working concentration. Single stain controls were used to calculate compensation matrix in flow cytometry and spectral flow cytometry experiments. Internal control healthy PBMC (human) or splenocytes (mouse) were used for each run to capture technical variabilities.

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	No cell lines were used in this study.
Authentication	
Mycoplasma contamination	
Commonly misidentified lines (See ICLAC register)	

Palaeontology and Archaeology

Specimen provenance	
Specimen deposition	
Dating methods	
☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.	
Ethics oversight	

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	WT Mus musculus on an FVB/N background purchased from Charles River UK. KRT5-CreER;tdTomato (MGI: 4358332, MGI: 3809523) were backcrossed into the FVB/N background and bred and maintained at the Biological Services Unit. All mice were maintained in specific-pathogen free conditions in the Biological Services Unit at University College London. Experiments were commenced when animals were 6-7 weeks. All mice were female.
Wild animals	No wild animals were used in this study.
Reporting on sex	All animal studies were conducted on females only, as females have been reported to have higher susceptibility to NTCU-induced squamous lung carcinogenesis compared to males of the same strain (PMID: 22086679). NTCU-treated and aged matched control animals were housed separately to minimise carcinogen exposure to controls.
Field-collected samples	The study did not include samples collected from the field.
Ethics oversight	Work was approved by the University College London Biological Services Review Committee and conducted following the UK Home Office procedural and ethical guidelines Animal [Scientific Procedures] Act 1986.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	UCLH Bronchoscopy Surveillance study: REC: 01/0148 SUMMIT: REC: 17/LO/2004, NCT03934866 ASCENT: REC: 20/SC/0128, NCT04204499 ALPINE: REC: 23/EM/0114, NCT05902559 An Investigation into the Molecular Pathogenesis of Lung Disease II: REC: 06/Q0505/12
Study protocol	Not currently published. Can be made available upon reasonable request. Summarised information available in relevant clinicaltrials.gov website (NCT03934866, NCT04204499, NCT05902559). UCLH Bronchoscopy Surveillance study has been recruiting at UCLH since 2001. This surveillance study targets individuals at high risk of lung cancer (i.e., heavy smokers) and uses autofluorescence bronchoscopy to screen for preinvasive lesions in the bronchus. Identified lesions are graded based on the worst lesion at the time. Patients with low-grade lesions identified continue to undergo surveillance approximately every 12 months, and those with high-grade lesions identified undergo surveillance approximately every 4 to 6 months. Bronchoscopy biopsies of suspected high-grade carcinoma in situ lesions are collected, along with a separate normal site biopsy. SUMMIT study started in 2019, with the first group target of 25,000 individuals at high risk for lung cancer due to significant smoking history from participating GP practices in north and east London. Patients receive a scan and bloods taken minimum at baseline and after one year. Suspicious cases are also followed up in 3 month intervals with scans, and sometimes bloods.
Data collection	ASCENT recruitment started in 2020, and is a prospective, observational cohort of participants scheduled to undergo surgical resection for low-dose CT screen-detected lung cancer. Participants undergoing surgery in ASCENT had previously taken part in SUMMIT. ALPINE is a prospective, observational cohort of minimally invasive samples from patients undergoing screening within the North Central London Targeted Lung Health Check programme. It is a continuation of SUMMIT which allows for blood / PBMC collection, and also contributes to ASCENT recruitment if / when participants develop NSCLC and undergo surgical resection. An Investigation into the Molecular Pathogenesis of Lung Disease II cohort comprised adults (≥18 years) undergoing investigation or treatment for respiratory disease, including thoracic surgery, bronchoscopy, radiology-guided biopsy, and other respiratory diagnostic procedures. For this study, only participants with a confirmed diagnosis of lung cancer (stage I-III, most cases used in this study being stage I) underwent surgical resection were included, contributing tumour and matched adjacent lung parenchyma samples for analysis.
Outcomes	

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | | |
|-------------------------------------|---|
| No | Yes |
| ☒ | ☐ Public health |
| ☒ | ☐ National security |
| ☒ | ☐ Crops and/or livestock |
| ☒ | ☐ Ecosystems |
| ☒ | ☐ Any other significant area |

Experiments of concern

Does the work involve any of these experiments of concern:

- | | |
|-------------------------------------|--|
| No | Yes |
| ☒ | ☐ Demonstrate how to render a vaccine ineffective |
| ☒ | ☐ Confer resistance to therapeutically useful antibiotics or antiviral agents |
| ☒ | ☐ Enhance the virulence of a pathogen or render a nonpathogen virulent |
| ☒ | ☐ Increase transmissibility of a pathogen |
| ☒ | ☐ Alter the host range of a pathogen |
| ☒ | ☐ Enable evasion of diagnostic/detection modalities |
| ☒ | ☐ Enable the weaponization of a biological agent or toxin |
| ☒ | ☐ Any other potentially harmful combination of experiments and agents |

Plants

Seed stocks	
Novel plant genotypes	
Authentication	

ChIP-seq

Data deposition

- ☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.
- ☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links May remain private before publication.	
Files in database submission	
Genome browser session (e.g. UCSC)	

Methodology

Replicates	
Sequencing depth	
Antibodies	
Peak calling parameters	
Data quality	

Flow Cytometry

Plots

Confirm that:

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

Methodology

Sample preparation

Instrument

Software

Cell population abundance

Gating strategy

Human flow cytometry was performed on cryopreserved PBMC samples isolated from blood by density centrifugation. Full details are outlined in the Methods section of the manuscript but briefly, these samples were thawed in R20 media, washed, and resuspended in FACS buffer (PBS with 2% FBS and 2 mM EDTA) prior to staining.

Mouse flow cytometry was performed on both cryopreserved samples and fresh samples. These were processed according to tissue type (outlined in Methods section of the manuscript).

For both mouse and human samples (depending on panel) samples were stained with chemokine receptor antibody mix containing Fc Receptor Binding Inhibitor Polyclonal antibody and Brilliant Stain Buffer Plus for 15 min at RT, or first incubated with Live/Dead Fixable Blue at RT 20 min prior to chemokine receptor mix for 15 min at 37 °C. In all cases, this was followed by incubation with extracellular markers for 30 min at 4 °C, fixed with FOXP3 Transcription Factor Fixation/Permeabilization Concentrate and Diluent solutions or True-Nuclear Transcription Factor Buffer Set, and left for fixation for 1 hr or overnight at 4 °C, before staining for intracellular markers for 1 hr at RT. Detailed descriptions can be found in the Methods section of the manuscript.

BD FACSymphony A5, Sony ID7000 Spectral Cell Analyzer

FACS Diva and ID7000 for acquisition (and unmixing in the case of spectral flow cytometry), FlowJo v10.8.1. and GraphPad Prism v11.0.2. for analysis.

Human PBMC composition: CD3+ ~70-80% of live cells

Murine lung composition: CD45+ ~50-80% of live cells, murine lymph node composition: CD45+ ~97-100% of live cells, murine cardiac puncture/vein bleed composition: CD45+ ~95-98% of live cells, murine spleen composition ~90 % of live cells.

Gating strategies were assigned based on staining of unstimulated and stimulated controls, as well as single stains.

For human blood, For human PBMCs, Treg cells were identified using the following gating strategy: FSC-A^{low} / SSC-A^{intermediate} (lymphocytes) → FSC-A / FSC-H (singlet gate) → SSC-A / SSC-H (singlet gate) → ZombieNIR⁻ (live cells) → CD3⁺CD19⁻ (T cells) → CD3⁺TCRgd⁻ (TCRab T cells) → CD4⁺CD8⁻ (CD4 T cells) → Foxp3⁺ (Treg) → CCR7⁺CD45RA⁻ (naive), CCR7⁺CD45RA⁻ (Tcm), CCR7⁺CD45RA⁺ (Temra), CCR7⁺CD45RA⁻ (Tem).

For human lung tissue, Treg cells were identified using the following gating strategy: FSC-A^{low} / SSC-A^{intermediate} (lymphocytes) → FSC-A / FSC-H (singlet gate) → SSC-A / SSC-H (singlet gate) → ZombieNIR⁻ CD45⁺ (live lymphocytes) → CD3⁺CD19⁻ (T cells) → CD3⁺TCRgd⁻ (TCRab T cells) → CD4⁺CD8⁻ (CD4 T cells) → Foxp3⁺ (Treg).

For murine tissues, For murine tissues, Treg cells were identified using the following gating strategy: FSC-A^{low} / SSC-A^{intermediate} (lymphocytes) → FSC-A / FSC-H (singlet gate) → SSC-A / SSC-H (singlet gate) → ZombieNIR⁻CD45⁺ (live lymphocytes) → CD3⁺CD19⁻ (T cells) → CD3⁺TCRgd⁻ (TCRab T cells) → CD4⁺CD8⁻ (CD4 T cells) → Foxp3⁺ (Treg).

For murine tissues, cDCs were identified using the following gating strategy: FSC-A^{lo} / SSC-A^{intermediate} (lymphocytes) → FSC-A / FSC-H (singlet gate) → SSC-A / SSC-H (singlet gate) → LiveDead⁻CD45⁺ (live lymphocytes) → CD3⁺CD19⁻ (non-T / B cells) → Ly6G⁻ (non-neutrophils) → PDCA1⁻ (non-pDC) → MHC-II^{hi}CD11c^{lo/int} (mig-DC) or MHC II^{lo/int} CD11c^{hi} (res-DC) → CD26⁺CD64^{lo} (conventional DC) → CD103⁺SIRPa⁻ (cDC1) or CD103⁺SIRPa⁺ (cDC2).

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

Magnetic resonance imaging

Experimental design

Design type

Design specifications

Behavioral performance measures

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)

Correction

Models & analysis

n/a Involved in the study

☐ ☐

Functional and/or effective connectivity

☐ ☐

Graph analysis

☐ ☐

Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis

This checklist template is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>