Corresponding author(s): Ilaria Malanchi
Luigi Ombrato

Last updated by author(s): 30/07/20

# Reporting Summary

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

## 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<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the 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<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*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 | Flow cytometry: samples were run on a BD 671 LSR-Fortessa (BD Biosciences, USA) using the BD FACSDiva software v8.0.1 |
| Data analysis | Analyses were performed using Prism software (version 7.0c, GraphPad Software, USA)<br>Fluorescence imaging: FiJi (version 2.0.0-rc-68/1.52g, ImageJ), ZEN10 (Zeiss), Imaris v 8.3.1 and Adobe Photoshop CC 2018 (version 19.0, Adobe, USA) were used to analyse fluorescence images.<br>Immunohistochemistry: images were acquired using NIS-elements software (version 4.51, Nikon, Japan)<br>Flow cytometry: data analyses were carried out using FlowJo 10.4.2 (FlowJO, LCC 2006-2018, USA). |

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

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

N/A

# 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

# Life sciences study design

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

| | |
|---|---|
| Sample size | No group were compared in these experimental protocol. |
| Data exclusions | no data were exlided |
| Replication | Unless otherwise specified in the figure legends, experiments were reproduced in at least two independent experiments |
| Randomization | The majority of the in vivo data generated in this study involved analysis between different areas of the same tissue in each mouse, therefore both control and experiment cannot be randomized. |
| Blinding | Investigators were not blinded for studies involving the analysis of the Niche versus distant lung cells as the cells were from the same samples and the two subsets could only be discriminated by FACS analysis itself. |

# 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 |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |

## Methods

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

## Antibodies

| | |
|---|---|
| Antibodies used | CD11b-APCCy7_ Biolegend_ 101226_ Rat monoclonal_ (M1/70)_ 1:100 (FC) CD45-BV421_ Biolegend_ 103133_ Rat monoclonal_ (30-F11)_ 1:200 (FC)<br>CD45-APC_ eBioscience_ 17-0451-83_ Rat monoclonal_ (30-F11)_ 1:200 (FC) CD45-APC-eFluor780_ eBioscience_ 47-0451-82_ Rat monoclonal_ (30-F11)_ 1:200 (FC)<br>CD326(EPCAM)-APC_ eBioscience_ 17-5791-81_ Rat monoclonal_ (G8.8)_ 1:200 (FC) CD326(EPCAM)-APC750Fire_ Biolegend_ 118230_ Rat monoclonal_ (G8.8)_ 1:200 (FC)<br>GFP_ Abcam_ ab6673_ Goat polyclonal_ 1:300 (IF)<br>Ly6A/E(SCA-1)-APC_ Biolegend_ 108111_ Rat monoclonal_ (D7)_ 1:200 (FC) Ly6A/E(SCA-1)-APC750Fire_ Biolegend_ 127652_ Rat monoclonal_ (D7)_ 1:200 (FC) Ly6A/E(SCA-1)-BV786_ BD Bioscience_ 563991_ Rat monoclonal_ (D7)_ 1:200 (FC) Ly6G-APC_ BD Bioscience_ 560599_ Rat monoclonal_ (1A8)_ 1:150 (FC)<br>Ly6G-APC750Fire_ Biolegend_ 127652_ Rat monoclonal_ (1A8)_ 1:150 (FC)<br>Ly6G-V450_ BD Bioscience_ 560603_ Rat monoclonal_ (1A8)_ 1:150 (FC)<br>mCHERRY_ Abcam_ ab183628_ Rabbit polyclonal_ 1:750 (IF)<br>TER-119_ Biolegend_ 116233_ Rat monoclonal_ (TER-119)_ 1:200 (FC) |
| Validation | The antibodies used have been validated accordingly to manufacturer's instructions. Mouse lung cell suspensions were used to validate FACS antibodies. Human or mouse lung sections were used to validates the antibodies for IF or IHC straining. |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | 4T1 and HC11 cell lines as well as MS5 and ML-1 cells were provided by the Cell Services Unit at The Francis Crick Institute. HEK 293FT were purchased from ThermoFisher Scientific (R700-07) |
| Authentication | Short Tandem Repeat (STR) was used to identify all cell lines used while SPID was used to confirm the species of origin. |
| Mycoplasma contamination | All cells are routinely tested for Mycoplasma by the Cell Services Unit of The Francis Crick Institute. |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified lines were used. |

## Animals and other organisms

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

| | |
|---|---|
| Laboratory animals | Female Balb/c mice 6-12 weeks old, NSG-NESTIN-GFP mice |
| Wild animals | no wild animals were used |
| Field-collected samples | N/A |
| Ethics oversight | All experiments were approved by Francis Crick ethical review committees and conducted according to UK Home Office Regulations (project license P83B37B3C and PPL 70/8904). |

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

## 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 | Lungs were minced and digested for 30 min in a shaker at 37ºC with a mixture of DNase I (Merck Sigma-Aldrich, Germany) and Liberase TM and TH (Roche Diagnostics, Switzerland) in HBSS solution. Samples were then washed, passed through a 100 um filter and incubated in Red Blood Cell Lysis buffer (Miltenyi Biotec, Germany) for 3-5 min at room temperature. After a wash with MACS buffer (0.5% BSA and 250 mM EDTA in PBS), samples were passed subsequently through a 40 um filter and a 20 um strainer-capped flow cytometry tube for single cell suspension to use for flow cytometric analysis or further purification. |
| Instrument | Flow cytometry analyses were carried out on a BD LSR-Fortessa (BD Biosciences, USA). |
| Software | FlowJo 10.4.2 (FlowJO, LCC 2006-2018, USA) was used for analysis. |
| Cell population abundance | Above 0.4% for the smaller population. |
| Gating strategy | All gating strategy are described in the methods and two typical examples are provided. |

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