## 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](https://www.nature.com/nature/authors.html) and the [Editorial Policy Checklist](https://www.nature.com/nature/authors.html#submission-checklist).

### Statistical parameters

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

<table>
<thead>
<tr>
<th>Item</th>
<th>Confirmed</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>The exact sample size ((n)) for each experimental group/condition, given as a discrete number and unit of measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The statistical test(s) used AND whether they are one- or two-sided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A description of all covariates tested</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For null hypothesis testing, the test statistic (e.g. (F), (t), (r)) with confidence intervals, effect sizes, degrees of freedom and (P) value noted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give (P) values as exact values whenever suitable.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimates of effect sizes (e.g. Cohen's (d), Pearson's (r)), indicating how they were calculated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clearly defined error bars</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State explicitly what error bars represent (e.g. SD, SE, CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Our web collection on [statistics for biologists](https://www.nature.com/nature/research/mar2018/01982-159.html) may be useful.*

### Software and code

**Policy information about availability of computer code**

<table>
<thead>
<tr>
<th>Data collection</th>
<th>No special or proprietary software was used.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data analysis</td>
<td>The following commercial software was used in this study: FlowJo v10, GSnap (version 2014-10-22), TraCeR (version 2015-10-21), gseapy (version 0.9.3), DAVID, cytoscape 3.4.0 and Monocle 2.0. R (version 3.4.3) with following Bioconductor packages: Rtsne v0.13, scran_1.6.6, densityClust v0.3 and ggplot2_2.2.1. R (version 3.3.1) with following Bioconductor packages: VennDiagram_1.6.18, ks_1.10.7, limma_3.28.21, survival_2.41-3 and ComplexHeatmap_1.11.8. Code for clustering is available on GitHub (<a href="https://github.com/Japrin/sscClust">https://github.com/Japrin/sscClust</a>). Other ad hoc scripts for analysing data are available upon request.</td>
</tr>
</tbody>
</table>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](https://www.nature.com/nature/authors.html#submission-checklist) 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

The expression data can be accessed from GEO database (accession number GSE99254), and the raw data can be accessed from EGA database (accession number EGAS00001002430). In addition, we have developed an interactive web server (http://lung.cancer-pku.cn) for analysing, visualising and downloading the single cell data for individual or multiple user-input genes.

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

☒ Life sciences
☐ Behavioural & social sciences

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

Life sciences

Study design

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

<table>
<thead>
<tr>
<th>Sample size</th>
<th>No statistical method was used to predetermine sample size.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data exclusions</td>
<td>For sequencing data, we excluded low-quality cells if abnormalities exist in (1) cell library sizes; (2) the numbers of expressed genes; (3) the proportion of mitochondrial gene counts; (4) CD3/CD4/CD8 gene expression value. After these filtering, 11,769 of total 12,346 cells remained. The details of cut-off line could be checked in Methods.</td>
</tr>
<tr>
<td>Replication</td>
<td>All replications were successful, and the detailed information was provided in corresponding figure legends.</td>
</tr>
<tr>
<td>Randomization</td>
<td>The patients with non-small cell lung cancer were recruited randomly in this study. The CD4/CD8/CD25 T cells were collected by FACS, and the single cells of each subtype were randomly collected.</td>
</tr>
<tr>
<td>Blinding</td>
<td>Not applicable since there was no specific grouping.</td>
</tr>
</tbody>
</table>

Materials & experimental systems

Policy information about availability of materials

n/a Involved in the study
☒ ☐ Unique materials
☒ ☐ Antibodies
☐ ☐ Eukaryotic cell lines
☒ ☐ Research animals
☐ ☒ Human research participants

Unique materials

Obtaining unique materials

Describe any restrictions on the availability of unique materials OR confirm that all unique materials used are readily available from the authors or from standard commercial sources (and specify these sources).

Antibodies

Antibodies used

Antibody (FACS) Clone Colour Catlog # Lot #
CD3 UCHT1 eFluor450 (eBioscience) 48-0038-42 4292976
CD4 OKT4 FITC (eBioscience) 11-0048-42 4307103
CD4 SK3 BV510 (BioLegend) 344634 B237895
CD8a OKT8 APC (eBioscience) 17-0086-42 4309588
CD25 BC96 PE (eBioscience) 12-0259-42 4302545
CX3CR1 2A9-1 FITC (BioLegend) 341605 B202340
CD103 B-Ly7 FITC (eBioscience) 11-1038-42 E12161-1636
PD1 EH12.2H7 PE (BioLegend) 329905 B214405
CTLA4 BNI3 BV605 (BioLegend) 369609 B230402

Antibody (multicolour IHC) Clone Catlog # Lot # Dilution
CD3 SP7 (Abcam) ab16669 GR224983-9 1/400
CD4 EPR6855 (Abcam) ab133616 GR91015-14 1/400
CD8 144B (Abcam) ab14147 GR240737-3 1/500
FOXP3 mAbcam22510 (Abcam) ab22510 GR146107-4 1/500

Validation
All the antibodies used in this study were commercial antibodies, with validation procedures described on the following sites of the manufacturers:

Eukaryotic cell lines
Policy information about cell lines
Cell line source(s) State the source of each cell line used.
Authentication Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.
Mycoplasma contamination Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register) Name any commonly misidentified cell lines used in the study and provide a rationale for their use.

Research animals
Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research
Animals/animal-derived materials For laboratory animals, report species, strain, sex and age OR for animals observed in or captured from the field, report species, sex and age where possible.

Human research participants
Policy information about studies involving human research participants
Population characteristics Fourteen patients who were pathologically diagnosed with NSCLC were enrolled in this study. None of the patients had autoimmune disorder or history of prior cancer. None of the patients was treated with chemotherapy, radiation, or any other anti-tumour medicines prior to tumour resection. Detailed information can be found in the clinical sample collection and preparation section of Methods and Supplementary Table 1.

Method-specific reporting

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
For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission
Provide a list of all files available in the database submission.
## Methodology

### Replicates
Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

### Sequencing depth
Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

### Antibodies
Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

### Peak calling parameters
Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

### Data quality
Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

### Software
Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

## 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
See the clinical sample collection and preparation section of Methods.

#### Instrument
BD FACSariaTM

#### Software
FlowJo v10

#### Cell population abundance
The abundance of the relevant cell populations, determined by testing the sorted cells again by FACS, reached >99%.

#### Gating strategy
See the FACS, reverse transcription, library preparation and sequencing section of Methods.

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
Indicate task or resting state; event-related or block design.

#### Design specifications
Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

#### Behavioral performance measures
State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

### Acquisition

#### Imaging type(s)

#### Field strength
Specify in Tesla

#### Sequence & imaging parameters
Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.

#### Area of acquisition
State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.
**Preprocessing**

<table>
<thead>
<tr>
<th>Task</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion MRI</td>
<td>Used or Not used, specify hardware version and revision number.</td>
</tr>
<tr>
<td>Preprocessing software</td>
<td>Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).</td>
</tr>
<tr>
<td>Normalization</td>
<td>If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.</td>
</tr>
<tr>
<td>Normalization template</td>
<td>Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.</td>
</tr>
<tr>
<td>Noise and artifact removal</td>
<td>Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).</td>
</tr>
<tr>
<td>Volume censoring</td>
<td>Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.</td>
</tr>
</tbody>
</table>

**Statistical modeling & inference**

<table>
<thead>
<tr>
<th>Task</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model type and settings</td>
<td>Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).</td>
</tr>
<tr>
<td>Effect(s) tested</td>
<td>Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.</td>
</tr>
<tr>
<td>Specify type of analysis</td>
<td>Whole brain, ROI-based, Both, or Whole brain and ROI-based.</td>
</tr>
<tr>
<td>Statistic type for inference</td>
<td>Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods. (See Eklund et al. 2016)</td>
</tr>
<tr>
<td>Correction</td>
<td>Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).</td>
</tr>
</tbody>
</table>

**Models & analysis**

<table>
<thead>
<tr>
<th>Task</th>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>n/a</td>
<td>Involved in the study</td>
</tr>
<tr>
<td>Functional and/or effective connectivity</td>
<td>Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).</td>
</tr>
<tr>
<td>Graph analysis</td>
<td>Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).</td>
</tr>
<tr>
<td>Multivariate modeling and predictive analysis</td>
<td>Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.</td>
</tr>
</tbody>
</table>

**Study design**

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

**Study description**

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

**Research sample**

State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.

**Sampling strategy**

Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.

**Data collection**

Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper,
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</td>
</tr>
<tr>
<td>Timing</td>
<td>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</td>
</tr>
<tr>
<td>Data exclusions</td>
<td>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</td>
</tr>
<tr>
<td>Non-participation</td>
<td>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</td>
</tr>
<tr>
<td>Randomization</td>
<td>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</td>
</tr>
</tbody>
</table>