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 our Editorial Policies 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

☐  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 measures 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. 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  SPSS v23 (IBM SPSS Statistics)

Data analysis  SPSS v23 (IBM SPSS Statistics)

For manuscripts utilizing custom algorithms, software that was written in house 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 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

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Field-specific reporting

4/28/2021
Life sciences study design

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

| Sample size | 11,871 samples were examined, comprising primaries (N=5,862), distant (N=5,605) and LNs metastases (N=404) |
|-------------|--------------------------------------------------------------------------------------------------|
| Data exclusions | NA                                                                                               |
| Replication | NA                                                                                               |
| Randomization | NA                                                                                               |
| Blinding | NA                                                                                               |

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 all item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|------------------------------------|---------|
| n/a                               | n/a     |
| ☑ Antibodies                       | ☑ Involved in the study |
| ☑ Eukaryotic cell lines            | ☑ ChiP-seq |
| ☑ Palaeontology and archaeology    | ☑ Flow cytometry |
| ☑ Animals and other organisms      | ☑ MRI-based neuroimaging |
| ☑ Human research participants      |         |
| ☑ Clinical data                    |         |
| ☑ Dual use research of concern     |         |