Statistics

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

- 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
- 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. 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**
Patient data between 2012 and 2016 were retrospectively obtained from the database of Peking University Cancer Hospital

**Data analysis**
Statistical analyses were performed using SPSS 22.0 software (IBM, Armonk, NY, 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 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 data that support the findings of our study are not publicly available, as they contain information that could compromise research participant privacy. However, the data from the corresponding author can be made available upon reasonable request.
**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](http://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 prior sample size calculations were performed. All available subjects who met the inclusion criteria were included. |
| Data exclusions | Patients were excluded based on the following criteria: M1 stage, breast carcinoma in situ only, synchronous bilateral breast cancer, no NAC or trastuzumab treatment, and a history of malignant tumors. |
| Replication | N/A |
| Randomization | There was no randomization to groups. |
| Blinding | No blinding was performed because This is a retrospective cohort study. |

**Behavioural & social sciences 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, 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. |
| Timing | Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort. |
| Data exclusions | 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. |
| Non-participation | 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. |
| Randomization | 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. |

**Ecological, evolutionary & environmental sciences study design**

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

| Study description | Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates. |
| Research sample | Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source. |
### Field work, collection and transport

**Field conditions**
Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).

**Location**
State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).

**Access & import/export**
Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).

**Disturbance**
Describe any disturbance caused by the study and how it was minimized.

### 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

| Method                        | Yes | No |
|-------------------------------|-----|----|
| Antibodies                   |     |    |
| Eukaryotic cell lines         |     |    |
| Palaeontology and archaeology |     |    |
| Animals and other organisms  |     |    |
| Human research participants  |     |    |
| Clinical data                |     |    |
| Dual use research of concern |     |    |

#### Methods

| Method                        | Yes | No |
|-------------------------------|-----|----|
| Involved in the study         |     |    |
| ChIP-seq                      |     |    |
| Flow cytometry                |     |    |
| MRI-based neuroimaging        |     |    |

#### Antibodies

**Antibodies used**
Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.

**Validation**
Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer’s website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

#### Eukaryotic cell lines

**Policy information about cell lines**
State the source of each cell line used.

**Cell line source(s)**

**Authentication**
Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.
Palaeontology and Archaeology

Specimen provenance

Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).

Specimen deposition

Indicate where the specimens have been deposited to permit free access by other researchers.

Dating methods

If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Animals and other organisms

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

Laboratory animals

For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.

Wild animals

Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.

Field-collected samples

For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.

Ethics oversight

Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

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

Not applicable

Study protocol

This is a retrospective study. All details are provided in the Methods section.

Data collection

Clinical data were retrospectively obtained from the database of Peking University Cancer Hospital.

Outcomes

The outcome measures were defined and pre-specified - as described in the manuscript.
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 | [ ] |

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. |
|-------------------|--------------------------------------|
|                   | 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. |
|------------------------------|-------------------------------------------------------------|

| Genome browser session (e.g. UCSC) | no longer applicable |

Methodology

| Replicates | Describe the experimental replicates, specifying number, type and replicate agreement. |
|------------|---------------------------------------------------------------------------------|
| 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
Describe
the
sample
preparation,
detailing
the
biological
source
of
the
cells
and
any
tissue
processing
steps
used.

Instrument
Identify
the
instrument
used
for
data
collection,
specifying
make
and
model
number.

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

Cell population abundance
Describe
the
abundance
of
the
relevant
cell
populations
within
post-sort
fractions,
providing
details
on
the
purity
of
the
samples
and
how
it
was
determined.

Gating strategy
Describe
the
gating
strategy
used
for
all
relevant
experiments,
specifying
the
preliminary
FSC/SSC
gates
of
the
starting
cell
population,
indicating
where
boundaries
between
"positive"
and
"negative"
staining
cell
populations
are
defined.

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)
Specify:
functional,
structural,
diffusion,
perfusion.

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.

Diffusion MRI
- Used
- Not used

Preprocessing

Preprocessing software
Provide
detail
on
software
version
and
revision
number
and
on
specific
parameters
(model/functions,
brain
extraction,
segmentation,
smoothing
kernel
size,
etc.).

Normalization
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.

Normalization template
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.

Noise and artifact removal
Describe
your
procedure(s)
for
artifact
and
structured
noise
removal,
specifying
motion
parameters,
tissue
signals
and
physiological
signals
(heart
rate,
respiration).
### Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

### Statistical modeling & inference

| **Model type and settings** | Specify type (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). |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| **Effect(s) tested**        | Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used. |
| **Specify type of analysis:** | Whole brain | ROI-based | Both |
| **Statistic type for inference** | Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods. |
| **Correction**              | Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo). |

#### Models & analysis

| **n/a Involved in the study** | Functional and/or effective connectivity | Graph analysis | Multivariate modeling and predictive analysis |
|------------------------------|------------------------------------------|----------------|---------------------------------------------|
|                              | Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information). | 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.). | Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics. |