nature research

Corresponding author(s): Sarah-Maria Fendt

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.

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).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of 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 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)
☐	☒	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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	GC-MS data were collected using MSD Chem Station Data Acquisition (vE.02.02.1431) or MassHunter Acquisition (vB.08.02 Build 8.2.8260.0) software. Proteomic data were collected using Q Exactive HF Tune 2.9.
Data analysis	GC-MS data were analyzed using MSD Chem Station Data Analysis (vE.02.02.1431) or Agilent MassHunter (vB.08.02 Build 8.2.8260.0) software followed by an in-house developed MatLab script. Proteomic data were analyzed using MaxQuant (v1.6.1.0). Microsoft Excel 2013 was used for data output. Statistical data analysis was performed using GraphPad Prism 7 software.

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 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 authors declare that all data supporting the findings of this study are available within the article, its extended data files, or from the corresponding author upon reasonable request.

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 ☐ Ecological, evolutionary & environmental 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.

Sample size	In vitro sample sizes were based on previous similar studies that have given statistical results. For in vivo experiments, sample size was determined using power calculations with B=0.8 and P<0.05, based on preliminary data and respects the limited use of animals in line with the 3R system: Replacement, Reduction, Refinement.
Data exclusions	No data were excluded from the study.
Replication	All experiments were carried out at least in triplicates. All attempts at replications were successful.
Randomization	Mice were randomized into control and treatment groups. Human participants were allocated to healthy, HCC or LC patient cohorts, based on health status of the patient.
Blinding	Mice and human participants were given a number (i.e. mouse XX and patient YY) prior to data collection and analysis. Data was collected and analyzed, and subsequently grouped in the corresponding cohorts for statistical analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human HEK293T epithelial cells, RWPE-1 prostate cells, MCF10A breast cells, A549 and H460 lung carcinoma, MDA-MB-468 and T47D breast adenocarcinoma, and DU145 prostate carcinoma cell lines were obtained from ATCC (Manassas, VA, USA). HUH7 liver carcinoma cell line was obtained from the Japanese Collection of Research Bioresources (JCRB) Cell Bank (Osaka, Japan).
Authentication	MCF10A and MDA-MB-468 cells were authenticated via fingerprinting. All other cell lines used in this study were otbained from well-established cell banks (ATCC and JCRB), and therefore, no authentication was performed on these cells.

Mycoplasma contamination	All cell lines used in this study were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

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

Laboratory animals	(1) Subcutaneous HUH7 xenograft model: 8-9 week-old female immunocompromised NMRI nu/nu mice (Taconic M&B AS, Denmark) were injected and sacrificed after 2-3 weeks. (2) DEN-induced HCC model: 2 week-old male C57Bl/6N mice (KU Leuven Animal Facility) were injected and sacrificed after 20-32 weeks. (3) Pten and Stk HCC model: male C57BL/6 mice (Institutional Animal Care at MD Anderson Cancer Center) were sacrificed after 12 months (Pten) or 4-6 months (Stk). (4) myrAKT-N-Ras HCC model: male 6-8 week-old FVB/N mice (Committee for Animal Research, University of California) were injected and sacrificed after 6-9 weeks. (5) Orthotopic HUH7 HCC model: 6 week-old male immunocompromised NMRI nu/nu mice (Taconic M&B AS, Denmark) were injected and sacrificed after 2-3 weeks.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.

Human research participants

Policy information about studies involving human research participants

Population characteristics	(1) Blood samples: healthy participants were 24-59 years old and this patient cohort contained 52% females and 48% males. HCC patients (diagnosed by a clinician) were 34-83 years old and this patient cohort consisted of 37.5% females and 62.5% males. Lung cancer patients (diagnosed by a clinician) were 43-85 years old and the cohort contained 66.5% females and 33.5% males. (2) Tissue samples: liver and HCC samples were obtained via biobanks and did not involve the handling of human patients. Lung cancer and adjacent lung tissue samples involved handling of human patients, which were 55-85 years old. This cohort consisted of 60% females and 40% males.
Recruitment	Healthy participants were recruited via email inquiry and upon informed consent. Cancer patients were recruited through the clinicians based on the presence of HCC or lung cancer, respectively, upon informed consent. Additionally, retrospective analysis of irreversibly anonymized archival tissue samples no longer required for diagnostics was carried out with approval by the local authorities. Healthy volunteers were on average younger than cancer patients this could lead to a bias due to age related effect. Moreover, normal tissue was either adjacent to cancer tissue or from healthy individuals. Thus, the health status of the individual could create a bias.