nature research

Corresponding author(s): DBPR NATMETAB-A18050090B

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	For targeted Mass Spectrometry, peak areas from the total ion current for each metabolite SRM transition were integrated using MultiQuant v2.1 software (AB/SCIEX). The results shown in this manuscript were part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. The genetic mutation status was confirmed by cansar portal (v3.0 beta) (https://cansar.icr.ac.uk/) and cancer Catalogue Of Somatic Mutations In Cancer (COSMIC) (http://cancer.sanger.ac.uk/cosmic/sample/overview?id=722040).
Data analysis	The TIC frequencies were estimated using ELDA software (Version 4.11.0.3973 vom 02.01.2018). For whole transcriptome analysis by microarray, fluorescent images were obtained with a Beadarray reader and processed with the BeadScan software (version: 3.5, Illumina, CA, USA). The whole transcriptome raw data were obtained from the GenomeStudio software (version: 2011.1) with the subtraction of the background. High-throughput small RNA-seq was performed on the HiSeq 2500 platform for 1 x 50-nt single-end sequencing and the sequencing adaptor was trimmed from the raw reads. To identify the known miRNAs, the remaining sequences were aligned to the miRBase (release 21.0) (http://www.mirbase.org/) using Bowtie (http://bowtie-bio.sourceforge.net/manual.shtml) (version: 1.2.1, release: 06/12/2017). Matched sequences with < 1 mismatch were known miRNAs. In addition, the unmatched sequences were used to predict the candidate novel miRNAs using miRDeep2 (Version 2.0.0.8). The hairpin RNA structures containing the unmatched sequences were predicated, complying with the criteria of pre-miRNAs in order to identify the potentially novel miRNAs. The enrichment of Gene Ontology (version: releases/2016-09-30) functional annotations was performed using DAVID Bioinformatics tool (v6.8, Oct. 2016). The heatmap for miRNA expression was generated according to Heatmapper website server (http://www2.heatmapper.ca/expression/). Metabolomics analysis was performed with online MetaboAnalyst 3.0 software (https://www.metaboanalyst.ca/). To predict upstream miRNA candidates that regulate VHL, TargetScan (version: release 7.2, March 2018) was used. The pyrosequencing result was analyzed

using PyroMark Q24 software. Western blot quantification was performed by Image Studio Lite (V5.0, release/March 3, 2015, LI-COR Biosciences). GraphPad Prism v. 6.0.2 were used for the statistical analysis. The 3D structures were measured by ImageJ 1.51s software. The total number of colonies in plate was analyzed by openCFU (http://opencfu.sourceforge.net).

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

Information and data for PDX models from the JAX PDX Resource are publicly available from the PDX Portal hosted by Mouse Tumor Biology Database (MTB; http://www.tumor.informatics.jax.org/mtbwi/pdxSearch.do). Data from this study have been deposited in the Gene Expression Omnibus (GEO) databases under the following accession: GSE103155 (microarray: single cell clones) and GSE103352 (miRNAseq). The results shown in this manuscript were part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. Analyses for an association between miRNA profiles and EGFR mutations as well as an association between VHL and MIR147b on a cohort of human lung adenocarcinoma cell lines were derived from a public RNA-seq dataset in the ArrayExpress database under accession number E-MTAB-2706. The heatmap for miRNA expression was generated according to Heatmapper website server (http://www2.heatmapper.ca/expression/). The genetic mutation status was confirmed by cansar portal (v3.0 beta) (https://cansar.icr.ac.uk/) and cancer Catalogue Of Somatic Mutations In Cancer (COSMIC) (http://cancer.sanger.ac.uk/cosmic/sample/overview?id=722040). The data that support the findings of this study are available from the corresponding author upon 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	Sample size was not predetermined for any experiments in the manuscript. We utilized sample sizes that are consistent with recent studies on a similar topic, and made an effort to avoid needless use of animals. In addition, we used statistical analysis consistent with the sample size of each experiment.
Data exclusions	No data were excluded from the analyses.
Replication	All experiments were performed in three to seven biological replicates, and independently reproduced as indicated in figure legends. And all attempts at replication were successful.
Randomization	Mice were age and sex-matched and randomized where appropriate (e.g. prior to xenograft transplantation for matched conditions).
Blinding	For mouse experiments, the investigators performing tumor volume measurements were blinded. For targeted Mass Spectrometry, whole transcriptome and high-throughput small RNA-seq analysis, the group allocation and data analysis were blinded.

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique materials used are readily available from the authors upon Material Transfer Agreement signature. DMOG (Cat #400091) was from Calbiochem. R59949 (Cat #D5794) and dimethyl malonate (DMM, Cat #136441) were purchased from Sigma-Aldrich. 3D cultures of lung patient-derived xenograft tumors (PDX_LU_10) were established at Beth Israel Deaconess Medical Center of Harvard Medical School.

Antibodies

Antibodies used	For immunofluorescence staining, primary mouse anti-human ZO-1 (1:100, clone #ZO1-1A12, lot #Q1215680, Cat #33-9100) was from Thermo Fisher Scientific. Secondary donkey anti-mouse IgG conjugated with Alexa Fluor 488 (1:500, lot #1741782, Cat #A-21202) was from Life Technologies. For western blot, primary polyclonal rabbit anti-VHL antibody (1:100, lot #RD2197039, Cat #PA5-27322) was from Thermo Fisher Scientific. Mouse anti-β-actin (1:5,000, clone #C4, lot #I1018, Santa Cruz, sc-47778) was used as loading control. IRDye 680RD goat anti-rabbit (1: 20,000, lot #C70406-04, cat #LI-COR926-68171, LI-COR Biosciences) and IRDye 800CW goat-anti-mouse (1: 20,000, lot #C70310-02, cat #LI-COR827-08364, LI-COR Biosciences) were used as secondary antibodies.
Validation	33-9100 targets ZO-1 has been successfully used in ELISA, Immunofluorescence (IF) and Western Blot (WB) applications and shows reactivity with human and canine samples. The validation of ZO-1 antibody is demonstrated on the Thermo Fisher's website (https://www.thermofisher.com/antibody/product/ZO-1-Antibody-clone-ZO1-1A12-Monoclonal/33-9100). PA5-27322 targets VHL in Immunohistochemistry (Paraffin) (IHC (P)), Immunoprecipitation (IP), and WB applications and shows reactivity with human and mouse samples. The PA5-27322 immunogen is recombinant fragment corresponding to a region within amino acids 1 and 213 of Human VHL. The validation of VHL antibody is demonstrated on the Thermo Fisher's website (https://www.thermofisher.com/antibody/product/VHL-Antibody-Polyclonal/PA5-27322). sc-47778 targets β-actin in IHC (P), IP, WB and IF applications and shows reactivity with mouse, rat, human, avian, bovine, canine, porcine, rabbit, Dictyostelium discoideum and Physarum polycephalum samples. The sc-47778 is raised against gizzard Actin of chicken origin. The validation of β-actin antibody is demonstrated on the Santa Cruz's website (https://www.scbt.com/scbt/product/beta-actin-antibody-c4). A-21202 targets mouse IgG specifically in Immunocytochemistry (ICC), IF and IHC and shows minimum cross-reactivity to bovine, chicken, goat, guinea pig, hamster, horse, human, mouse, rat, and sheep serum proteins. The validation of donkey anti-mouse IgG conjugated with Alexa Fluor 488 antibody is demonstrated on the Thermo Fisher's website (https://www.thermofisher.com/antibody/product/Donkey-anti-Mouse-IgG-H-L-Highly-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-21202). RDye 680RD goat anti-rabbit targets rabbit IgG in WB and IHC applications. The validation of RDye 680RD goat anti-rabbit antibody is demonstrated on the LI-COR Biosciences website (https://www.licor.com/bio/products/reagents/secondary_antibodies/irdye_680rd.html). IRDye 800CW goat-anti-mouse targets mouse IgG in WB and IHC applications. The validation of IRDye 800CW goat-anti-mouse antibody is demonstrated on the LI-COR Biosciences website (https://www.licor.com/bio/products/reagents/secondary_antibodies/irdye_800cw.html).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human lung EGFR-wild type cell lines H358, H460, A549, H1299, and H69 were obtained from ATCC as well as EGFR-mutant cell lines H1650, H1975, HCC827, HCC827GR, PC9, PC9ER, and H3255 were provided by Dr. Susumu Kobayashi. Immortalized tracheobronchial epithelial AALE cells were provided by Dr. William C. Hahn. 3D cultures of lung patient-derived xenograft tumor (PDX_LU_10) were established at Beth Israel Deaconess Medical Center of Harvard Medical School.
Authentication	Cell line identities were confirmed by STR fingerprinting.
Mycoplasma contamination	All cell lines were found negative for mycoplasma using the MycoAler Kit (Lonza).
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	All research involving animals complied with protocols approved by the Institutional Animal Care and Use Committees (IACUC) from BIDMC, The Jackson Laboratory and Yale University. For establishing subcutaneous tumors from PDXs, 4-6 weeks old female NSG immunodeficient mice (The Jackson Laboratory, strain # NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, stock # 005557, homozygous for both the Prkdcscid and Il2rgtm1Wjl alleles) were used as part of commercial production of the PDX at The Jackson Laboratory or implanted as part of this study at Yale Cancer Center. No littermate controls were used. For generating xenograft tumors by transplanting H1975 cells with miR-147b knockdown, 4-6 weeks old female nude immunodeficient mice (The Jackson Laboratory, strain # NU/J, stock # 002019, homozygous for Foxn1nu) were used for subcutaneous injections at the BIDMC.

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	Patient characteristics are summarised in Supplementary Table 2. Information and data for PDX models from the JAX PDX Resource are publicly available from the PDX Portal hosted by Mouse Tumor Biology Database (MTB; http://tumor.informatics.jax.org/mtbwi.pdxSearch.do)
Recruitment	Patient tumor material was obtained through a network of collaborating cancer research centers to develop the JAX PDX Resource. The collaborating centers were responsible for any necessary IRB approvals and patient consents to allow their tumor tissue to be used in research. These responsibilities are acknowledged in the JAX PDX Consortium Agreement, which was signed by the participating member prior to submitting their first patient specimen for model development. The JAX PDX Resource received coded patient tumor samples, with all personal identifiers removed. No JAX investigator has access to patient consent forms. Successful PDX models were assigned a unique number with all reference to the donating center and date of specimen collection removed to further protect patient privacy. Besides, patients with advanced NSCLC who developed progression after initial response to EGFR TKI were consented and enrolled to a Yale University IRB approved protocol, in accordance with ethical guidelines, allowing the collection and analysis of clinical data, fresh tissue, and the generation of patient-derived xenografts. All Yale patients received prior radiotherapy. Part of the tissues has been collected in Yale patients whose tumor have developed progression after initial response to EGFR TKI. These clinical samples are primary tissues that are tolerant or resistant to EGFR TKIs. The genetic mutations have been described in Supplementary Table 2 and these are unlikely to have an impact the results.