nature research

Corresponding author(s): Campbell, Hoare

Last updated by author(s): Aug 12, 2021

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 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	Image processing from sequencing data using standard Illumina X10 and NovaSeq pipeline
Data analysis	Alignment and variant calling performed using Sanger Institute's custom pipeline. Single-nucleotide substitutions were called using the CaVEMan (cancer variants through expectation maximization) algorithm (https://github.com/cancerit/CaVEMan). Small insertions and deletions were called using the Pindel algorithm (https://github.com/genome/pindel). Rearrangements were called using the BRASS (breakpoint via assembly) algorithm (https://github.com/cancerit/BRASS). List of programs and softwares: • R: version 3.5.1 • Perl: version 5.3.0 • Python: version 3.8.5 • MATLAB: version R2019b • BWA-MEM: version 0.7.17 (https://sourceforge.net/projects/bio-bwa/) • cgpCaVEMan: version 1.11.2/1.13.14/1.15.1 (https://github.com/cancerit/CaVEMan) • cgpPindel: version 2.2.2/2.2.4/2.2.5/3.2.0/3.3.0 (https://github.com/cancerit/cgpPindel) • Brass: version 5.4.1/6.0.5/6.1.2/6.2.0/6.3.4 (https://github.com/cancerit/BRASS) • ASCAT NGS: version 4.0.1/ 4.1.2/4.2.1 (https://github.com/cancerit/ascatNgs) • JBrowse: version 1.16.1 (https://jbrowse.org/) • cgpVAF: version 2.4.0 (https://github.com/cancerit/vafCorrect) • alleleCount: version 4.1.0 (https://github.com/cancerit/alleleCount) • SigProfiler: version 1.0.0-GRCh37 (https://github.com/AlexandrovLab) • HDP: version 0.1.5 (https://github.com/nicolaroberts/hdp) • dNdScv: version 0.0.1 (https://github.com/im3sanger/dndscv)

• Telomerecat: version 3.4.0 (https://github.com/jhrf/telomerecat)
• STAR: version 2.7.6a (https://github.com/alexdobin/STAR)
• Picard-tools: version 2.20.7 (https://broadinstitute.github.io/picard/)
• Samtools: version 1.12 (http://www.htslib.org/)
• TrimGalore: version 0.6.4 (https://github.com/FelixKrueger/TrimGalore)
• GATK: version 4.1.4.1 (https://gatk.broadinstitute.org/hc/en-us)
• GSEA: version 3.0 (https://www.gsea-msigdb.org/gsea/index.jsp)
• XGBoost: version 0.82.1 (https://xgboost.readthedocs.io/en/latest/)
• NDP.view2 (https://www.hamamatsu.com/eu/en/product/type/U12388-01/index.html)
• label.switching: version 1.8 (https://cran.r-project.org/web/packages/label.switching/index.html)
• philentropy: version 0.3.0 (https://cran.r-project.org/web/packages/philentropy/index.html)
• MCMCglmm: version 2.29 (https://cran.r-project.org/web/packages/MCMCglmm/index.html)
• Magick: version 2.0 (https://cran.r-project.org/web/packages/magick/index.html)
• Pheatmap: version 1.0.12 (https://cran.r-project.org/web/packages/pheatmap/index.html)
• Thermo Fisher software Tracefinder: version 5.0 (https://www.thermofisher.com/uk/en/home/industrial/mass-spectrometry/liquid-chromatography-mass-spectrometry-lc-ms/lc-ms-software/lc-ms-data-acquisition-software/tracefinder-software.html)
• CellProfiler: version 4.0.3 (https://cellprofiler.org/)
• PerkinElmer harmony: version 4.9 (https://www.perkinelmer.com/category/cellular-imaging-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 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

Whole genome sequencing data in the form of BAM files across samples reported in this study have been deposited in the European Genome-Phenome Archive (Accession number EGAD00001006255; https://www.ebi.ac.uk/ega/home). RNA-sequencing data has been deposited in the European Nucleotide Archive (Accession number ERP123192; https://www.ebi.ac.uk/ena/browser/home).

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

Life sciences study design

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

Sample size	Across all patients, we identified 1,322,612 unique somatic substitutions, with 1,946,613 called overall – this means that from the 1586 microdissections, we have the equivalent of ~1078 (68%) unique samples sequenced. From published power calculations for identifying cancer genes, this effective sample size equates to a power of ~90% for detecting a significant excess of mutations in 90% of genes mutated in 2% of clones.
Data exclusions	Samples with low mean coverage (<15x) were excluded due to the inaccuracy of mutation catalogues
Replication	Experiments for metabolomics and RNA-sequencing on cells transfected with wild-type or mutant FOXO1 constructs were performed with 5 replicates. Replicates showed consistent results.
Randomization	Not applicable - this is a descriptive study, not an intervention study.
Blinding	Not applicable - all dependent variables were computationally generated (mutation counts, signatures etc) and statistical analyses were prespecified.

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

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used

anti-β-Actin (clone: AC15) (Sigma, A5441, 1:5000, RRID:AB_476744); anti-Akt (clone: C73H10) (Cell Signaling, 2938, 1:1000, RRID:AB_915788); anti-phospho-Akt(T308) (clone: 244F9) (Cell Signaling, 4056, 1:1000, RRID:AB_331163); anti-GFP (Abcam, ab6556, 1:1000, RRID:AB_305564); anti-FOXO1 (clone: C29H4) (Cell Signaling, 2880, 1:1000, RRID:AB_2106495); anti-phospho-FOXO1 (T24) (Cell Signaling, (9464, 1:1000, RRID:AB_329842).

Validation

anti-β-Actin (clone: AC15): Validated by supplier with the following notes - species reactivity: pig, Hirudo medicinalis, bovine, rat, canine, feline, human, rabbit, carp, mouse, guinea pig, chicken, sheep; application(s): western blot: 1:5,000-1:10,000 using cultured human or chicken fibroblast cell extracts.
anti-Akt (clone: C73H10): Validated by supplier with the following notes - species reactivity: human, mouse, rat, monkey; application(s): suitable for western blot.
anti-phospho-Akt(T308) (clone: 244F9): Validated by supplier with the following notes - species reactivity: human, mouse, rat, monkey; application(s): suitable for western blot.
anti-GFP (Abcam, ab6556): Validated by supplier with following notes - species reactivity: independent; application(s): suitable for Suitable for: IHC-P, Electron Microscopy, ICC, IP, Flow Cyt, IHC-Fr, western blot
anti-FOXO1 (clone: C29H4): Validated by supplier with the following notes - species reactivity: human, mouse, rat, monkey; application(s): suitable for western blot.
anti-phospho-FOXO1 (T24): Validated by supplier with the following notes - species reactivity: human, mouse, rat, monkey; application(s): suitable for western blot.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The 3 HCC cell lines (HepG2, Hep3B and PLC/PRF/5) were all obtained from ATCC.

Authentication

Identification confirmed by SNP genotyping

Mycoplasma contamination

All cell lines were confirmed as Mycoplasma negative

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study.

Human research participants

Policy information about studies involving human research participants

Population characteristics

The dataset comprised 1590 genomes from 34 liver samples, including 5 normal liver controls with no prior neoadjuvant therapy, 10 with alcohol-related liver disease (ARLD) and 19 with NAFLD (Supplementary Table S1). All patients with ARLD or NAFLD had HCC, liver failure or both and tissues were derived from hepatic resection or transplantation. Overall, 9 samples were from patients who had a synchronous HCC and underlying cirrhosis; a further 8 samples had HCC without underlying cirrhosis, including 3 hepatic resection samples from one patient over a 5-year timespan (Extended Figure 1). All samples underwent central histological review by specialist hepatopathologists, and the histological and clinical features of the patients matched those expected for the underlying disease processes (Supplementary Table S1). The average age of research subjects was 61 years, and the male:female split was 29:5.

Recruitment

Recruited through Addenbrooke's Hospital, Cambridge, UK. All patient gave written informed consent, and were typically of advanced stage liver disease. Because explanted liver samples were mostly used, there is a recruitment bias towards high severity of disease.

Ethics oversight

East of England Research Ethics Committee: 15/EE/0351 and 16/NI/0196

Note that full information on the approval of the study protocol must also be provided in the manuscript.