

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](#).

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

FASTQ sequence files for WGS and WES were generated from the Illumina sequencer images using the Illumina RTA 1.18.66.3 (HiSeq 2500) or 2.7.7 (HiSeq 4000) and bcl2fastq 2.17.1.14 software; The Illumina sequencer images for the Chromium library sequencing were processed with the Illumina bcl2fastq 2.17.2.1 software and the 10x Genomics Long Ranger 2.1.5 pipelines. Linked-Read data quality was assessed using the 10x Genome browser Loupe; Ion Torrent WES sequencing used the BaseCaller module of the Torrent SuitTM software package v5.8.0 (Thermo Fisher Scientific Inc). Pacbio Sequel sequencing used the Pacbio SMARTlink software v5.0.0. CytoScan array (Affymetrix) was processed with ChAS software (version 3.3)

Data analysis

BAMSurgeon v1.1
Bowtie2 v2.2.9
BWA MEM v0.7.15
DeepVariants v0.5.1
Freebayes v1.2.0
GATK3 v3.8
HaplotypeCaller v3.8.0
MuSE v1.0rc_c
MuTect2 4.beta.6
NovoAlign v3.07.01
RTG v3.8.4
SomaticSeq v2.8.1 / seqc2_v1.2
SomaticSniper v1.0.5.0
Strelka2 v2.8.4
TMAP v3.0.1
TNScope v201711.02
Trimmomatic v0.30
VarDict v1.5.1

NeuSomatic v0.1.3
 AscatNgs v4.2.1
 Cell Ranger DNA 1.1

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. 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

All raw data (FASTQ files) are available on NCBI's SRA database (SRP162370). The truth set for somatic mutations in HCC1395, VCF files derived from individual WES and WGS runs, bam files of BWA-MEM alignment and source codes are available on NCBI's ftp site (ftp://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/seqc/Somatic_Mutation_WG). Some alignment files (BAM) are also available on Seven Bridges' Cancer Genomics Cloud (CGC) platform under SEQC-II project. Public databases for variant annotation: COSMIC (<https://cancer.sanger.ac.uk/cosmic>), ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>)

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	DNA extracted from two cell lines (HCC1395 and HCC1395BL) were applied to multiple NGS technologies (HiSeq-WES, HiSeq-WGS, NovaSeq-WGS, PacBio-WGS, 10X Genomics-WGS, Ion Torrent-WES, and NextSeq-Ampliseq). Total of 21 WGS on HiSeq (3x3+3) and NovaSeq (9) with coverage ranging 50X to 100X were used to define initial "truth set" of somatic mutations and germline variants. 12 repeats of HiSeq-WES (3x3+3), a single run of Ion Torrent-WES, and three replicates of NextSeq-Ampliseq were used to confirm three subsets of variants in the "truth set". Results from WGS of PacBio were used to resolve contradicting calls of a small number of variants. Single cell sequencing with genomic DNA (gDNA) was performed on HCC1395 (1465 cells) and HCC1395BL (983 cells).
Data exclusions	All data from planned experiments have been included.
Replication	Six sequencing centers performed 12 repeats (3x3+3) of HiSeq-WGS and HiSeq-WES, and 11 repeats of WGS on the 10X Genomics. A single sequencing center performed 9 repeats of NovaSeq-WGS and 3 repeats of NextSeq-Ampliseq. All attempts at replications were successful.
Randomization	Aliquots DNA from the same lot were randomly distributed to each sequencing center.
Blinding	Since materials derived from two well-established cell lines originated from a consented patient, no blinding was necessary.

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

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)	American Type Culture Collection (ATCC), HCC1395 (CRL2324), HCC1395 BL (CRL2325)
Authentication	The cell lines have been authenticated by STR profile, Karyotyping, PCR Mycoplasma, and Sterility testing.
Mycoplasma contamination	No mycoplasma contamination was detected
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.