nature portfolio

Corresponding author(s): Yuexiang Wang, Kui Wu, Hui Cao, Simin Wang

Last updated by author(s): Oct 16, 2024

Reporting Summary

Nature Portfolio 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 Portfolio 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	Software was used for collection and analysis and is listed below
Data analysis	The raw sequencing data was processed with the following steps by SOAPnuke (https://github.com/BGl-flexlab/SOAPnuke) MuSE (vl.O, default parameters), Strelka2 (v2.99, default parameters), MuTect2 (GATK v4.l.O.O, default parameters), SomaticSniper (vl.0.5.0, -q 1-Q 40 -L -G) and Lancet (vl.07, default parameters) were used for single nucleotide variants (SNVs), while Strelka2 (v2.99, default parameters), MuTect2 (GATK v4.l.O.O, default parameters), Lancet (vl.07, default parameters) and SvABA (v0.2.1, default parameters) were used for small indels. All SNVs and indels were subsequently annotated by ANNOVAR (v20170716) SigProfilerMatrixGenerator (vl.1) and BayesNMF algorithm to extract de novo signatures. SigProfiler was used to extract mutational signatures. Copy number variations were estimated by FACETS (v0.5.11) and CNVkit. A critical value (cval) of 150 was used to run FACETS with the recommended parameters. Then, the broad and focal CNVs were identified by GISTIC2.0 Structural variations (SVs) were detected using Manta (vl.5.0) To identify and visualize chromothripsis-like patterns in the cancer genomes, the copy number (CN) and SV data were used as input for ShatterSeek (v0.4) (https://github.com/parklab/ShatterSeek) with default parameters. An open-source tool named SeqKat (v0.0.8) was used to predict kataegis regions. Then KataegisPortal was used to visualize the kataegis events ABSOLUTE (vl.0.6) was used to estimate the purity and ploidy of paired samples. Somatic non-synonymous mutations (SNVs and indels) were extracted to construct phylogenetic tree of each patient based on maximum likelihood by MEGA software SOAPnuke was used to analyse transcriptome sequencing. Fusion genes and their respective fusion points were detected by following algorithms: FusionCatcher (v.l.10), Arriba (vl.1.0), STAR-Fuison (vl.6.0) and SOAPfuse (vl.18). The total immune components for each sample were analyzed using the ESTIMATE (vl.0.13). The telomere content was determined using TelomereHunter.Gene set enrichment analysis (GSEA , v4.0.l) was applied separately to each mRNA subtype. Graph Pad Prism 9 and Adobe Illustrator were used to draw graphs. GraphPad Prism 9 was used to analyze statistical data. Image J was used for quantification of immunoblot intensity

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The sequencing datasets (raw WGS, WES, and WTS data) generated in this study have been deposited in the Genome Sequence Archive in the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: HRA005970) database under accession code HRA005970 [Homepage: https://bigd.big.ac.cn/gsa-human/browse/HRA005970] or [Download page: https://download.cncb.ac.cn/gsa-human/HRA005970]. The expression data of CINSARC cohort and Japanese cohort was downloaded from the Array Express accession: E-MTAB-373 and the NCBI database under accession GSE136755 [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136755]. The somatic mutation datasets (MAF file) of SARC analyzed in this study were downloaded from the GDC Portal [https://gdc.cancer.gov/about-data/publications/sarc_2017]. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The samples used in this study are from both female and male patients
Reporting on race, ethnicity, or other socially relevant groupings	Race was determined by self-identification. Samples were from Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, China. No analysis was performed based on race.
Population characteristics	Population characteristics information has been summarized within the Supplementary Data 1. 113 de-identified tumor specimens and 68 matched normal samples (5 peripheral blood samples and 63 non-cancerous tissues) were collected from 101 GIST patients surgically dissected at Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine.
Recruitment	Participants were not recruited.
Ethics oversight	Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences

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

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	113 de-identified tumor specimens and 68 matched normal samples (5 peripheral blood samples and 63 non-cancerous tissues) were collected from 101 GIST patients surgically dissected at Ren Ji Hospital, Shanghai iao Tong University School of Medicine. In addition, 4 GIST cell lines were also included in the study. All sample size are described in the figure legends.
Data exclusions	No data were excluded.
Replication	All experiments were replicated independently as described in the figure legends and replicated.
Randomization	Mice were randomly allocated among groups
Blinding	Investigators were blinded when performing the mice experiments

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Antibody sources and catalog number were derscribed in the method section. For Western Blotting, PCNA (Santa Cruz, #sc-56, 1:500, RRID:AB_628110), p-KIT Y721 (Cell Signaling Technologies, #3391, 1:1000, RRID:AB_2131153), KIT (Agilent, #R7145, 1:1000, RRID:AB_2131465), p-MAPK Thr202/Tyr204 (Cell Signaling Technology, #9101, 1:1000, RRID:AB_331646), MAPK (Cell Signaling Technology, #9102, 1:1000, RRID:AB_330744), p-AKT Ser473 (Cell Signaling Technology, #9271, 1:1000, RRID:AB_329825), AKT (Cell Signaling Technology, #9272, 1:1000, RRID:AB_329827) , GAPDH (Sigma, #G8795, 1:1000, RRID:AB_1078991), YLPM1(Novus Biologicals, #NBP2-22326, 1:2000). For IHC, CD117 (ready for use; Maixin Bio Co., Ltd., Fuzhou, China), SDHB (ready for use; Maixin Bio Co., Ltd., Fuzhou, China), YLPM1 (1:400; Novus Biologicals, #NBP2-22326) and CD8 (ready for use; Maixin Bio Co., Ltd., Fuzhou, China).
Validation	All antibody were validated according to the respective manufacture's information and citations.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK 293T was obtained from the American Type Culture Collection (ATCC # ACS-4500) and used for functional studies. The following five GIST cell lines were subjected to WES and/or transcriptome sequencing. GIST-Tl (case 90T, KIT exon 11 V560_ Y578del mutation) was generously provided by Dr. Takahiro Taguchi. The remaining 4 cell lines were developed and kindly provided by Dr. Jonathan Fletcher laboratory at Brigham and Women's Hospital as previously reported. GIST-CN2 primary culture was established from a TKI resistant, metastatic GIST patient (case 94T, male, KIT exon 11: L576P plus exon 13: V654A). GIST-CN10 primary culture was established from a metastatic GIST (male) with PDGFRA D842V mutation. GIST-CN16 primary culture was established from a TKI resistant, metastatic GIST patient (Female, KIT exon 9: A502_Y503dup plus exon 17: N822K).
Authentication	All cell lines have been authenticated by Sanger sequencing.
Mycoplasma contamination	All cell lines were routinely tested negative for microbial contamination (including mycoplasma).
Commonly misidentified lines (See ICLAC register)	No cell line adopted in this study is listed in the database of commonly misidentified cell lines maintained by ICLAC

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	BALB/c nude mice aged 6 weeks were used. All the mice were fed standard laboratory diet and maintained in a pathogen-free environment(20-26?, 40-70%) on an 12-h light/12-h dark cycle with food and water supplied and libitum throughout the experiments period.
Wild animals	The study did not involve wild animals.
Reporting on sex	The findings are not limited to a specific sex.
Field-collected samples	The study did not involve filed-collected samples
Ethics oversight	The animal experiments were conducted according to the guidelines for the care and use of laboratory animals and were approved by Institutional Animal Care and Use Committee (IACUC) of the Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences (with.approved ID SIBS-2017-WYX-l).

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

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA