nature portfolio

Corresponding author(s): Tatsuhiro Shibata

Last updated by author(s): Jan 27, 2023

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	No commercial software was used for data collection. All of the software used is reported in the paper.
Data analysis	Open sources used to analyse the data in this study are the following EthSEQ (version 3.8): Ethnicity Annotation from Whole Exome Sequencing Data BWA-MEM (version 0.7.8): Mapping short read sequences against a large reference genome, such as the human genome. NovoAlign (version 3.2.8): Mapping of short reads onto a reference genome. SAMtools (version 1.9): Remove PCR duplications and generate pileup files from mapped BAM files. GISTIC (version 2.0.21): Identifies significant recurrent copy number alterations FACETS (version 0.6.0): Allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing GATK (version 3.5.0): Genome Analysis Toolkit: Variant Discovery in High-Throughput Sequencing Data. MuPeXI (version 1.1): The mutant peptide extractor and informer, a tool for predicting neo-epitopes from tumor sequencing data POLYSOLVER (version 1.0): A tool for HLA typing based on whole exome sequencing (WES) data and infers alleles for the three major MHC class I (HLA-A, -B, -C) genes. xHLA (version 1.2): Fast and accurate HLA typing from short read sequence data HLA-HD (version 1.4.0): HLA-HD (HLA typing from High-quality Dictionary) can accurately determine HLA alleles with 6-digit precision from NGS data (fastq format). RNA-Seq data can also be applied. LOHHLA (version 1.0): A computational tool to evaluate HLA loss using next-generation sequencing data. Fusionfusion (version 0.3.0): Detecting gene fusion using the putative chimeric transcript generated by several well-known transcriptome alignment tools (STAR, MapSplice2 and TopHat2) STAR (version 2.S.2a): Ultrafast universal RNA-seq aligner SAVNet (version 0.4.0): Detecting somatic variants associated with local aberrant splicing alterations from the matched genome and transcriptome sequencing data. deconstructSigs(version 1.8.0): Determine the weights of each mutational signature contributing to an individual tumor sample

MutSigCV (version 1.4): Identifies genes that are significantly mutated in cancer genomes, using a model with mutational covariates
MutSig2CV: version 2 of MutSigCV
dndscv (version 0.0.1.0): driver-gene detection tool evaluating highly recurrent mutations within a gene
OncodriveFM (version 2.2): driver-gene detection tool evaluating high-functional-impact (pathogenic) mutations within a gene.
OncodriveMUT: driver-mutation detection tool evaluating high-functional-impact (pathogenic) mutations within a gene
Kallisto (version 0.43.1): Quantifying abundances of transcripts from bulk and single-cell RNA-Seq data, or more generally of target sequences using high-throughput sequencing reads
CIBERSORT (version 1.06): Provide an estimation of the abundances of member cell types in a mixed cell population, using gene expression data
ANNOVAR (version April 16, 2018): Functional annotation of genetic variants from high-throughput sequencing data.
SIFT: Predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids
PolyPhen2_HDIV: Predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforwardphysical and comparative considerations.
MutationTaster: Integrates information from different biomedical databases and uses established analysis tools
CADD: Scoring the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome
VADAR (version 1.8): A web server for quantitative evaluation of protein structure quality
UCSF Chimera (version 1.14): The interactive visualization and analysis of molecule

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

All data needed to evaluate the conclusions in this paper are presented in this paper, the Supplementary Materials, or are available at the following repositories:

Japanese cohort; WGS/WES/RNAseq data (n=697) from the International Cancer Genome Consortium (ICGC) Japanese gastric cancer cohort is available at the European Genome-phenome Archive (EGA) with the accession number EGAD00001004051 and EGAD00001008610, and the Japanese Genotype-phenotype Archive with the accession codes JGAS000228 (https://humandbs.biosciencedbc.jp/en/hum0227-v1) and JGAS000229 (https://humandbs.biosciencedbc.jp/en/hum0226-v1). These data (JGAS000228 and JGAS000229) are under controlled access because they are personally identifiable data defined by Japan's Personal Information Protection Law. For use, approval is required at the review by the NBDC Human Data Review Board. Data users shall apply for data use in accordance with the data use application procedures (https://humandbs.biosciencedbc.jp/en/data-use).

Singaporean cohort; WES (n=49) is deposited at the EGA with the accession number EGAD00001006957. These data are under controlled access because they are personally identifiable data under Singapore's Ministry of Health HIM (Health Instruction Manual) guidelines, where WES is classified as Restricted (Sensitive, High) information which requires access controls.

Chinese cohort; WGS/WES data (n=169) is available at the EGA with accession numbers EGAS00001000597 and EGAS00001001056.

Vietnamese cohort; WES data (n=48) is available at the EGA with the accession number EGAS00001000736.

South Korean cohort; WES data (n=52) is available at the European Nucleotide Archive database with accession number PRJEB10531.

TCGA cohort (multi-ancestry); WES/RNAseq data (n=442) is available at dbGAP with the accession number phs000178.

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	No sample size calculation was performed because the present study is observational, not interventional
Data exclusions	Samples not passing quality-control have been excluded. Details of the exclusion criteria are reported in the paper.
Replication	We conducted deep resequencing to validate S327R/G/C mutations in TRIM49/49B/49C (please see Extended Data Figure 2g). In all ten cases

Replication	where we detected TRIM49 mutations by whole-exome sequencing, the same mutations were confirmed by this validation analysis.
Randomization	Not applicable for this observational study
Blinding	Not applicable for this observational study

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Six hundred ninety-seven」apanese GC cases and TCGA (442 cases), China (217 cases), Singapore (49 cases), and Korean cohort (52 cases).
Recruitment	Gastric cancer patients in National Cancer Center, Tokyo, Japan, The University of Tokyo, Yokohama City University, National University of Singapore, and an Tock Seng Hospital were recruited with consent.
Ethics oversight	National Cancer Center, Tokyo, Japan, The University of Tokyo, and Yokohama City University

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