nature research

Corresponding author(s): Seishi Ogawa

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	We enrolled 27 ESCC/dysplasia patients and 44 healthy individuals, from whom 233 and 326 oesophageal samples were collected, respectively and subjected to high-throughput sequencing from 21/May/2014 to 1/Oct/2018. In the remaining 136 samples from 68 subjects, cryopreserved oesophageal samples were provided from Kyoto University Hospital Cancer Biobank. As for external data sets, whole-exome sequencing data of paired tumor/germline control samples from ESCC (n=90) and other cancer (n=8,808) patients were downloaded from the TCGA Data Portal (https://tcga-data.nci.nih.gov/docs/publications/tcga/). Five previously published datasets of whole-exome sequencing obtained from in total 504 ESCC patients were also analysed. No software or program was used for downloading the data.
Data analysis	Detailed descriptions of the software and analysis have been provided in the online methods. Sequencing data was processed using our inhouse pipeline Genomon2. External bam files were converted to fastq format using biobambam2 and subjected to mutation calling by GenomonMutationFilter. SNP array data was analysed by CNAG. Copy number analysis using sequencing data was performed by inhouse pipeline CNACS. Significantly mutated genes were identified using MutSigCV and dNdScv. Mutational signature was evaluated using pmsignature MutationalPatterns. Significant focal copy number alterations were detected using GISTIC 2.0. Clonal architecture was inferred by PyClone (0.13.0). Statistical analyses were performed using R (3.4.4). List of programs and softwares: Genomon2: version 2 (https://genomon.readthedocs.io/ja/latest) GenomonMutationFilter: (https://github.com/Genomon-Project/GenomonMutationFilter) Burrows-Wheeler Aligner: version 0.7.10 (https://sourceforge.net/projects/bio-bwa/)

picard-tools: version 1.39 (http://picard.sourceforge.net/)
biobambam2: version 2.0.85 (https://www.sanger.ac.uk/science/tools/biobambam)
EBCall: version 2 (https://github.com/friend1ws/EBCall)
Integrative Genomics Viewer (IGV): version 2.3 (http://software.broadinstitute.org/software/igv/)
MutSigCV: version 1.4. (http://software.broadinstitute.org/cancer/software/genepattern/modules/docs/MutSigCV)
dNdScv: version 0.0.1 (https://github.com/im3sanger/dndscv)
CNAG: version 3.5.1 (http://www.genome.umin.jp/CNAG_DLpage/CNAG_top.html)
CNACS: (http://plaza.umin.ac.jp/kyoto_tumorpatho/CNACS.html)
GISTIC: version 2.0 (https://software.broadinstitute.org/software/igv/GISTIC)
pmsignature: version 0.3.0 (https://github.com/friend1ws/pmsignature)
MutationalPatterns: version 1.4.2 (http://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html)
PyClone: version 0.13.0 (https://github.com/aroth85/pyclone)

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

All the WES, WGS and SNP array data have been deposited in the European Genome-Phenome Archive (http://www.ebi.ac.uk/ega/) under accession numbers EGAS00001003008, EGAS00001003281, and EGAS00001003331, respectively.

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

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	No statistical methods were used to determine sample size since this is an exploratory study. We enrolled as many patients as possible who provided consent for our study during the enrollment period between 1/May/2014 and 1/Sep/2018. We collected 369 and 326 oesophageal samples from 95 ESCC/dysplasia patients and 44 healthy individuals, respectively. We obtained 11 samples from 7 cases by both solitary sampling and high-density sampling. The sample size of our study was sufficient to assess our sequencing data and to confirm novel findings with statistical significance. An additional 9,312 whole-exome sequencing data of 9,312 cases from external datasets validated our results.
Data exclusions	No data were excluded intentionally. Samples with insufficient amount of DNA for sequencing were excluded from this study.
Replication	We did not attempt replication in sequencing experiments. Instead, we validated sequencing results by confirmatory PCR-based deep sequencing of randomly selected variants (whole exome sequencing (n=622 variants) and whole genome sequencing (n=94)).
Randomization	Not applicable since this is a case-series study which was therefore not planned to detect any difference in effects between the cohorts with and without intervention.
Blinding	Two expert pathologists, blinded to clinical information, reviewed all the specimens obtained from cancer and dysplasia patients excluding cryo-preserved ones.

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants were recruited through Kyoto University Hospital and other participating institutes. We enrolled a total of 139 patients who underwent surgery and/or therapeutic or diagnostic endoscopy for upper gastrointestinal symptoms at Kyoto University Hospital and other participating institutes. Among these, 95 had ESCC (n = 93) or high-grade oesophageal dysplasia (n = 2) and the remaining 44 were turned out to have no oesophageal lesions and thought to be healthy individuals. All subjects and samples were anonymized with unique IDs before analysis. The detailed characteristics of cases were provided in Supplementary Table1.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging