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 any 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 software was used to collect data.

Data analysis | All software used are described in the methods section. Publicly available softwares for Next-generation Sequencing analysis:Skewer (v0.2.2), STAR(v2.4.2a), HiseM (1.2.29), GATK4, Mutect2, Platypus(0.8.1), SvABA(1.1.3), Pelly(0.9.1), MutationalPatterns(3.2.0), Patchwork(1.0), ABSOLUTE(1.0), Seqkat(0.0.8), AmpliconArchitect(1.2), NanoPack, Minimap2. The complex SVs were defined by FindRear and the source code has uploaded to github (https://github.com/ZHOUYong0530/FindRear). All statistics analysis and survival analysis were performed in R (version 4.0).

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 raw sequencing data generated in this study have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation (Beijing Institute of Genomics, Chinese Academy of Sciences (NGS-Human: HRA003107 (WGS&RNA-seq), https://ngdc.cnbc.ac.cn/gsa-human/browse/HRA003107), HRA000021 (WGS, https://ngdc.cnbc.ac.cn/gsa-human/browse/...
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.

- [x] 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: We performed whole-genome sequencing on regional tumor samples and adjacent normal tissues from 528 ESCC patients, of which 133 pairs also were sequenced by RNA-seq. In addition, we supplemented the WGS and Nanopore sequencing of two ESCC samples.
- Data exclusions: The ESCC regional tumor samples produced low quality sequencing data were excluded.
- Replication: For all experiments, at least three independent experiments were performed and each experiment was performed in triplicate. All results of duplicates were consistent.
- Randomization: The ESCC patients were collected randomly to form the cohort. All the studies ESCC patients were retrospective and did not need randomly grouping. And cells were allocated into experimental groups randomly.
- Blinding: Investigators were blinded to the group allocation during cell implantation or sample/data collection.

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
- 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: PTHLH antibody (ab197358, Abcam, Cambridge, UK) (1:200 dilution) was used as the primary antibody in immunohistochemistry.
- Validation: Antibodies used IHC were validated using negative and positive controls and underwent an optimization process including titration, variation of antigen retrieval process, and incubation periods, and further testing in a select set of clinical samples before they were applied to a large scale cohorts.

Eukaryotic cell lines

- Policy information about cell lines
- Cell line source(s): ESCC cell lines KYSE180, KYSE150 and KYSE450 cell line were purchased from Cell Bank of Type Culture Collection of Chinese
| Authentication                                                                 | All of the cells were authenticated by short tandem repeat (STR) analysis. |
|------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| Mycoplasma contamination                                                      | All cell lines were routinely tested to ensure they are free of mycoplasma contamination (VenorTM GeM Mycoplasma Detection Kit, Sigma-Aldrich). |
| Commonly misidentified lines (See ISAC register)                              | No commonly misidentified cell lines were used. |