nature portfolio

NCOMMS-21-30906-T
Corresponding author(s): Prof Yaohe Wang
yaohe.wang@qmul.ac.uk
Last updated by author(s): Aug 9, 2021

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

Bcbio-nextgen v1.1.1 https://doi.org/10.5281/zenodo.3564938 https://github.com/bcbio/bcbio-nextgen
BWA v0.7.17 Li and Durbin, 2009 http://bio-bwa.sourceforge.net/
STAR Dobin et al., 2013 https://github.com/alexdobin/STAR
Samtools v1.9 Li and Durbin, 2009 https://www.htslib.org/
GATK 4 v4.0.6 McKenna et al., 2010 https://github.com/broadinstitute/gatk
Mutect2 Cibulskis et al., 2013 https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Strelka2 v2.8.4 Kim et al., 2018 https://github.com/Illumina/strelka
IntOGen v3.0.8 Martínez-Jiménez et al., 2020 https://bitbucket.org/intogen/intogen-pipeline/src/master/
VEP v95.3 McLaren et al., 2016 https://grch37.ensembl.org/info/docs/tools/vep/index.html
Maftools v2.2 Mayakonda et al., 2018 https://www.bioconductor.org/packages/release/bioc/html/maftools.html
Cibersort Chen et al., 2018 https://cibersort.stanford.edu/
Timer Li et al., 2017 http://cistrome.org/TIMER/
xCell Aran et al., 2017 https://xcell.ucsf.edu/
MCP-counter v1.1 Becht et al., 2016 https://github.com/ebecht/MCPcounter
Salmon Patro et al., 2017 https://combine-lab.github.io/salmon/ v0.9.0
survminer v0.4.3 CRAN - Package survminer https://rpkgs.datanovia.com/survminer/index.html
cqn Hansen et al., 2012 https://www.bioconductor.org/packages/release/bioc/html/cqn.html
tximport v1.10 Soneson et al., 2015 https://bioconductor.org/packages/release/bioc/html/tximport.html
limma v3.38 Ritchie et al., 2015 https://www.bioconductor.org/packages/release/bioc/html/limma.html
SOAPnuke v1.5.6 Chen et al., 2018 https://github.com/BGI-flexlab/SOAPnuke
NMFConsensus GenePattern https://cloud.genepattern.org/gp/pages/protocols/ClassDiscovery_nmf.html

ConsensusClusterPlus v1.46 Wilkerson and Hayes, 2010 https://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
Sequenza v2.1.2 Favero et al., 2015 https://cran.r-project.org/web/packages/sequenza/vignettes/sequenza.html
PyClone Roth et al., 2014 https://github.com/Roth-Lab/pyclone

Data analysis

Only open source codes were used for this study, with default setting or parameters were used unless specified otherwise in the manuscript. All analysis was performed in the R statistical environment version 3.6.1. All statistical tests were two-sided and statistical significance was determined if p value or adjusted p value was less than 0.05, unless otherwise stated. All the codes that were used for analyses and the generation of figures are available at https://github.com/Zhong2020/ESCCproject.

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

Sequence data used during the study has been deposited at National Genomics Data Center of China (https://bigd.big.ac.cn/) and the Bioproject Access ID is PRJCA001577. The WES data is available via: https://ngdc.cncb.ac.cn/gsa-human/s/B0ZNK91t. RNA sequencing data is available via: https://ngdc.cncb.ac.cn/gsa-human/s/C1x3s3X0.

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	For discovery cohort, sample size of patients was determined by the availability of tumour specimens. 120 treatment-naive ESCC tumors were prospectively collected under strict protocols (Table S1) for RNA sequencing analysis. Of them, 103 had sufficient DNA quality and were also profiled using whole-exome sequencing. For the validation study of four top genes (WFDC1, SFRP1, LGR6 and VWA2) related to the stemness as prognostic biomarkers, an independent cohort of 63 patients were chosen (Figure 1e).
Data exclusions	Whole-exome sequencing data analysis were performed in 103 of 120 samples with matched RNA-seq.
Replication	To ensure reproducibility of biological functional validation, drug assays (Fig 3f and g) several measures were undertaken. We used a panel of ESCC cell lines known sensitive or resistance cell lines for our drugs (5-FU) to ensure our drug worked as expected. For detection of gene expression either by Western Blot (Extended Data Fig 3 b, c and d) or RT-PCR (Fig 1e, Extended Data Fig 6 a and b), all experiments were at least duplicated. For cell proliferation assay (Extended Data Fig 3c and d, Extended Data Fig 6d), all experiments were performed in triplicates.
Randomization	Randomized by independent animal caretaker blinded to experimental groups (Fig 1f, Extended Data Fig 3 e and f).
Blinding	When the immunohistochemistry staining slides (Extended Data Fig 3a, Fig 5c and d) were reviewed, histopathologists were blinded to molecular subtype group of each ESCC sample.

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
[ ]	[x] Antibodies
[ ]	[x] Eukaryotic cell lines
[x]	[ ] Palaeontology and archaeology
[ ]	[x] Animals and other organisms
[ ]	[x] Human research participants
[ ]	[x] Clinical data
[x]	[ ] Dual use research of concern

Methods

n/a	Involved in the study
[x]	[ ] ChIP-seq
[x]	[ ] Flow cytometry
[x]	[ ] MRI-based neuroimaging

Antibodies

Antibodies used: Rabbit anti-SFRP1 polyclonal antibody(Atlas Antibodies,Cat# HPA064870), Rabbit anti-XCL1 polyclonal antibody(Atlas Antibodies,Cat# HPA057725), Mouse anti-GAPDH antibody(Proteintech, Cat# 60004-1-Ig), Rabbit anti-GPCR/LGR6 antibody(Abcam, Cat# ab126747), Rabbit anti-CD160 antibody (Origene, Cat# TA349762), Mice anti-CD8 antibody (Gene Tech, Cat# GT211207), Mice anti-CD4 antibody (clone sp35)(Maixin Biotechnology, Cat# RMA-0620), Goat anti-Rabbit IgG-HRP (ZSBIO, Cat# ZB-5301), Goat anti-mouse IgG-HRP (ZSBIO, Cat# ZB-5305).

Validation: All antibodies used in this study are validated by the supplier and the statements could be seen on the manufacturer's website. The detailed dilution of all antibodies used in this manuscript can be seen the methods.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s): All ESCC cell lines including KYSE-30, KYSE-70, KYSE-140, KYSE-150, KYSE-180, KYSE-270, KYSE-410, KYSE-450, KYSE-510, KYSE-520 were bought from DSMZ-German Collection of Microorganisms and Cell Cultures GmbH.

Authentication: All cell line authentication are confirmed with short tandem repeat (STR) assay.

Mycoplasma contamination: The mycoplasma contamination were validated negative in all cell lines.

Commonly misidentified lines (See ICLAC register): No cell lines are misidentified.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals: Female BALB/c nude mice were used in the studies of SFRP1 effect on tumour growth (Fig 1f and Extended Data Fig 3 e and f)

Wild animals: Not used

Field-collected samples: 2×10E6 SFRP1 overexpressing/knock-down and the matched control ESCC cells were subcutaneously injected into bilateral flanks of BALB/c nude mice (n=6 per group). Xenograft volumes were measured after 30 days and harvested at 30 days after inoculation (Extended Data Fig 3 e and f).

Ethics oversight: All animal care and experiments were approved by the Ethical Committee of the Zhengzhou University (Wang/2016) and were in accordance with the Provision and General Recommendation of Chinese Experimental Animals Administration Legislation.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics: For the discovery study, 120 patients diagnosed with esophageal squamous cell carcinoma derived from Anyang Cancer Hospital, for the validated cohort of 65 ESCC samples were from The First Affiliated Hospital of Zhengzhou University. None of these patients received any radiotherapy or chemotherapy before surgery and pathology diagnosis were confirmed by three independent pathologists. The clinical pathological information of the patients are shown in Extended Data Table 1.

Recruitment: Patients were recruited after diagnosis of ESCC, tumour samples and adjacent normal tissues at least 5 cm away from paired tumour tissues were collected and placed in liquid nitrogen within 30 minutes after surgery resection.

Ethics oversight: This study was approved by the ethics committees of both Anyang Cancer Hospital (2015 (NO:1) and The First Affiliated Hospital of Zhengzhou University (2019-KY-51).

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	N/A
Study protocol	120 patients diagnosed with esophageal squamous cell carcinoma from 2013 to 2016 were enrolled from Anyang Cancer Hospital under the approve of ethics committee of Both Anyang Cancer Hospital and The First Affiliated Hospital of Zhengzhou University. None of these patients received any radiotherapy or chemotherapy before surgery and pathology diagnosis were confirmed by three independent pathologists.
Data collection	Tumour samples and adjacent normal tissues at least 5 cm away from paired tumour tissues were collected and placed in liquid nitrogen within 30 minutes after surgery operation. Patient clinical data were recorded during patients' hospital visits and telephone consultancy.
Outcomes	Outcome data of more than four years follow up after surgical resections was recorded