nature portfolio

Corresponding author(s): Jae-Ho Shin, Gun Oh Chong

Last updated by author(s): May 30, 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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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	Human reference sequences (GRCh38), National Center for Biotechnology Information (NCBI) RefSeq database, Kyoto Encyclopedia of Genes and Genomes (KEGG) database, Carbohydrate-Active enZYmes (CAZymes) database, Genome Taxonomy Database (GTDB)
Data analysis	SOAPnuke, Bowtie2, Kraken2, Bracken, VALENCIA, MEGAHIT, Prodigal, CD-HIT, CoverM, GhostKOALA, dbcan3, metaWRAP, GTDB-Tk, iTOL, R software

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 shotgun metagenome sequencing raw data (short-read archives, SRA) are available in the GenBank repository with bioproject accession number PRJNA1054643.

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	All participants in the research are women.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	We recruited a total of 68 women and categorized them into three groups: Group 1 consisted of 23 participants, including 11 with negative malignancy findings (normal) and 12 with low-grade squamous intraepithelial lesions (LSIL); Group 2 comprised 23 patients with high-grade squamous intraepithelial lesions (HSIL); and Group 3 included 22 patients with invasive cervical cancer (ICC).
Recruitment	This study included patients with histologically validated cervical lesions who presented to the Department of Obstetrics and Gynecology at Kyungpook National University Chilgok Hospital (Daegu, Republic of Korea).
Ethics oversight	The ethical approval for this study was obtained from the Institutional Review Board of Kyungpook National University Chilgok Hospital (KNUCH 2023-08-054-001).

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	Participants with missing or unknown data on age, BMI or other clinical data were excluded from this analysis. We further excluded samples with low concentration, poor quality, and failed sequencing, resulting in the final sample size of 68 participants.
Data exclusions	Once the sample size was determined, no data were excluded from the analyses.
Replication	No applicable
Randomization	No applicable
Blinding	No applicable