nature portfolio

Corresponding author(s): Sunghoon Kwon

Last updated by author(s): Mar 18, 2022

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	Standard Illumina processing pipelines were used to collect pair-end sequencing data on a Nextseq 550 instrument. In-house source code is available at the https://github.com/BiNEL-SNU/Select-seq
Data analysis	For the basic RNA-seq analysis and quality control: Cutadapt v3.1, STAR v2.7.3a, RSEM v1.2.25, featureCounts v2.0.0, RSeQC v2.6.4, Qualimap v2.2.1, DESeq2 v1.26; For the A-to-I editing analysis: REDItools v2.0; For the immune cell receptor analysis: TraCeR/BraCeR v0.6.0

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 TCGA gene-expression data and patient's metadata are available from NIH genomic data commons (https://portal.gdc.cancer.gov). Any processed RNA-seq data is uploaded to the Sequence Read Archive (BioProject, PRJNA779567)

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	Overall six tumor sections from the five different breast tumor patient's sample were used in the analysis. One replicate sections from the same human sample were used for the FISH validation. We reasoned n=5 will be appropriate to validate the quality and utility of the proposed method on the human sample.
Data exclusions	No data was excluded from the analysis.
Replication	Six tumor sections from five different patient who diagnosed as the triple negative breast cancer were used in the analysis as the replicate for the TNBC tumor sample. One additional replicate sections were used for the validation of FISH experiment
Randomization	Not relevant for the study
Blinding	Not relevant for the 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

Antibodies

Antibodies used	Primary antibodies used for immunofluorescence were anti-CD4 (rabbit, Abcam, ab181724, 1:200), anti-CD8 (rabbit, Abcam, ab251596, 1:200), anti-CD14 (rabbit, Abcam, ab230903, 1:200), anti-CD19 (rabbit, Abcam, ab237772, 1:200), anti-CD44 (rabbit, Abcam, ab194987, 1:200), and anti-ALDH1 (rabbit, Abcam, ab215996, 1:200). And each antibody was directly conjugated with Alexa 488 (Abcam, ab236553), TxRed (Abcam, ab195225), Cy3 (Abcam, ab188287), and Cy5 (Abcam, ab188288).
Validation	Antibodies were purchased based on the information available on the supplier's website and on their suitability for immunofluorescence application. Thus the antibodies were not further validated for the study.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All parental cell lines were from Korean cell line bank (HEK293T, HuT 78, IM-9, NIH3T3).
Authentication	None of the cell lines used were authenticated.
Mycoplasma contamination	All cell lines used were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Six sections from the five different patient samples were used in the analysis. Patient had the following metadata associated: 1. Female, Primary, T2, N0, M0, Stage 2a, Triple negative breast cancer, neoadjuvant_therapy = N 2. Female, Primary, T4, N3, M1, Stage 4, Triple negative breast cancer, neoadjuvant_therapy = Y 3. Female, Recurrence, T4, N2, M0, Stage 2b, Triple negative breast cancer, neoadjuvant_therapy = Y 4. Female, Primary, T4, N3, M0, Stage 3c, Triple negative breast cancer, neoadjuvant_therapy = Y 5. Female, Primary, T2, N2, M0, Stage 2b, Triple negative breast cancer, neoadjuvant_therapy = Y
Recruitment	Tumor sections are selected appropriate for archival as fresh frozen status and there is no special recruitment criteria.
Ethics oversight	This study complied with all relevant ethical regulations regarding experiments involving human tissue samples and samples were collected with informed consent. Ethical permission (SNUH IRB 1910-130-1072 and SNUH IRB 1405-088-580) for the human sample used in this study was granted by the Regional Ethics Committee of Seoul National University Hospital.

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