nature portfolio

Corresponding author(s): Youness Azimzade

Last updated by author(s): Apr 18, 2026

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	Datasets were manually downloaded from their corresponding repositories.
Data analysis	CIBERSORTx (used to estimate cell fractions from bulk gene expression and RNA seq data), SHAP (used for feature importance), Python3, R 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

This study does not generate new experimental data. Processed data generated in this study are publicly available at:https://github.com/YounessAzimzade/Survival-XML-TME-BC/Data and Source data are provided with this paper.scRNA-seq count matrices and annotations from Wu et al. (2021) were obtained from the Broad Institute Single Cell Portal under accession SCP1039:https://singlecell.broadinstitute.org/single_cell/study/SCP1039/a-single-cell-and-spatially-resolved-atlas-of-human-breast-cancers Counts and coarse-grained annotations from Bassez et al. (2021) were downloaded from:https://lambrechtslab.sites.vib.be/en/single-cellIMC

data from Danenberg et al. (2022) are available at Zenodo under accession 5850952:https://doi.org/10.5281/zenodo.5850952IMC data for the TNBC cohort from Wang et al. (2023) are available at Zenodo under accession 7990870:https://doi.org/10.5281/zenodo.7990870Bulk RNA-seq and gene expression datasets were obtained from TCGA-BRCA:https://portal.gdc.cancer.gov/projects/TCGA-BRCAand METABRIC:https://doi.org/10.1038/nature10983Gene expression data were accessed via cBioPortal:https://www.cbioportal.org/Neoadjuvant chemotherapy (NAC) cohorts were obtained from ArrayExpress and Gene Expression Omnibus (GEO) under the following accession numbers:E-MTAB-4439:https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-4439/GSE18728:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18728GSE19697:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19697GSE20194:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20194GSE20271:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20271GSE22093:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22093GSE22358:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22358GSE42822:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42822GSE22513:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22513GSE25066:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25066GSE32603:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32603GSE32646:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32646GSE37946:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37946GSE50948:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50948GSE23988:https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE23988

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	Given the high prevalence of breast cancer in females, sex was not included as a covariate in our study.
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	Here we used 19 public datasets with their charactristics available in corresponding references.
Recruitment	In each type of analysis, we have tried to identify and use all relevant datasets.
Ethics oversight	NA

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	Sample size was determined based on available data for each analysis. Our NAC datasets includes 2007 samples, TCGA with 1000 and METABRIC with 2000 samples where used for survival. Spatial omics data set also included 700 samples and 200 samples. scRNA-seq dataset included 57 samples (184k cells). We believe these sample sizes were comprehensive for given analysis.
Data exclusions	No relevant dataset was excluded during our analysis.
Replication	Data was divided into discovery and validation cohort to improve reproducibility of the results. In survival analysis, TCGA and METABRIC datasets were used for cross validation. Major findings were also validated in other data types.
Randomization	Datasets were randomly divided into discovery and validation cohorts in NAC cohort. TCGA and METABRIC were already distinct datasets.
Blinding	Since the data were divided into two groups afterwards, investigators were blind to group allocation.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.