nature portfolio

Corresponding author(s): Olivier Gevaert

Last updated by author(s): Oct 17, 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.

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	For accessing data from the Genomic Data Commons (GDC) portal (TCGA/CPTAC), we used the GDC Data Transfer Tool Client (GDC-client): https://gdc.cancer.gov/access-data/gdc-data-transfer-tool
Data analysis	Our code is available at https://github.com/gevaertlab/sequoia-pub. All algorithms/tools are explained in the Methods section in the manuscript.

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

Anonymized WSIs, gene expression and clinical data of the The Cancer Genome Atlas' (TCGA) cohorts were retrieved from the publicly available Genomic Data Commons (GDC) portal (https://portal.gdc.cancer.gov).

Gene expression data of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) cohort were downloaded from GDC portal (https://portal.gdc.cancer.gov) and WSIs were obtained from the Cancer Image Archive with the accession URL (https://www.cancerimagingarchive.net/collections).

Gene expression data and WSIs of the Tempus cohort were obtained through a data transfer agreement with Tempus Labs, Inc.

The publicly available spatial transcriptomics data of GBM were acquired from Datadryad using the following accession URL (https://doi.org/10.5061/dryad.h70rxwdmj). Spatial transcriptomics and matched histology images of BRCA were obtained from Jaume et al. (https://doi.org/10.48550/arXiv.2406.16192).

The RNA-seq data and clinical annotations of the SCANB cohort were obtained from the accession URL (https://data.mendeley.com/datasets/yzxtxn4nmd/3), and data of the METABRIC cohort was obtained for cbioportal with accession URL (https://www.cbioportal.org/study/summary?id=brca_metabric).

Source data for all figures/tables in this work are provided as a zipped folder (including main text and supplementary). Each file within the folder is named according to the figure/panel it belongs to.

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	This study does not involve the recruitment of human research subjects. All datasets include both male and female participants, in which the selection criteria have been described in the published studies. No further selection based on sex and gender was made in the current study.
Reporting on race, ethnicity, or other socially relevant groupings	See above.
Population characteristics	See above.
Recruitment	This study does not involve the recruitment of human research subjects.
Ethics oversight	No ethical approval is required for the data used in the current study.

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 sizes for each cohort are listed in Supplementary Tables and the Methods section.
Data exclusions	For all study cohorts, we selected the patients that had both WSIs and gene expression available, no other exclusions were made.
Replication	Results obtained in the TCGA cohort were validated in the CPTAC and Tempus cohorts.
Randomization	This is an observational study where randomization is not applicable.
Blinding	Blinding is not relevant for this 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
☒	☐ 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.