nature portfolio

Corresponding author(s): Christos Sotiriou

Last updated by author(s): Oct 24, 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	In this retrospective study, 94 early-stage TNBC patients treated at the Institut Jules Bordet (Brussels, Belgium) with standard-of-care therapies between 2000 and 2016 were included. Clinical data were collected for all patients undergoing initial surgery, followed by adjuvant chemotherapy and/or radiotherapy. For two patients, surgical samples of locoregional relapse were also collected. In total, we collected 96 TNBC samples from 94 patients. For each sample, frozen surgical tumor tissue sample was collected from the institutional tissue bank and stored at −80 °C. High quality sequencing data were obtained for 94 out of 96 frozen samples. For each sample, we obtained also bulk RNA sequencing data sequenced on the Illumina NovaSeq 6000 platform (total N = 94 samples). Spatial transcriptomics sequencing data were obtained from the second generation of Spatial Transcriptomics 2K arrays sequenced on the NextSeq500 (v2) platform in duplicates (N = 2 x 1 sample) ou triplicates (N = 3 x 93 samples): total N = 281 samples. For the external validation of our results, data from breast cancer cases were collected from three external breast cancer datasets: METABRIC (N = 335 TNBCs), SCAN-B (N = 672 TNBCs) and I-SPY2 (N = 987 BCs including 363 TNBCs). The latter dataset was specifically used to assess response to immunotherapy in early-stage BC patients. For METABRIC, data were downloaded from https://www.cbioportal.org. For SCAN-B, Expression data, obtained by whole transcriptome RNA sequencing, were downloaded from https://data.mendeley.com/datasets/yzxtxn4nmd. For I-SPY2, expression data were downloaded from the GEO database (GSE194040). In addition, fourteen datasets with available gene expression data were downloaded from https://www.orcestra.ca/clinical_icb, for a total of 1073 patients with diverse metastatic cancers treated by immune checkpoint inhibitors.
Data analysis	For the bulk RNA-seq data, reads were trimmed using Trimmomatic v0.38. Genes were quantified using Salmon on the reference human genome GRCh38/hg38 and GENCODE release v38 for the gene positions. Morphological annotation of the hematoxylin and eosin stained ST slide was performed with QuPath software (version 0.2.3). For the ST RNA-seq data, STAR (https://github.com/alexdobin/STAR) was used to align reads to the reference genome (CRCh38) and used HTseq-count (https://github.com/simon-anders/htseq) to generate gene counts based on a Ensembl (v.86) reference file. ST-pipeline (v.1.6.0)

(https://github.com/SpatialTranscriptomicsResearch/st_pipeline) was used to pre-process the spatial data, ie. spot selection and image alignment.
The processed count matrices derived from the raw ST data and the associated brightfield images (H/E-images) are available via the repository https://zenodo.org/doi/10.5281/zenodo.8135721.
Analysis of the spatial data was mainly performed using R package STstuff. Analysis indirectly related to ST analysis such as various plotting and bioinformatics functions was performed with R package nonSTstuff. All codes associated with the manuscript are available at the github repository: https://github.com/BCTL-Bordet/ST ([https://doi.org/10.5281/zenodo.13867936]).
The following additional R packages were used: glmGamPoi R package (for the batch correction), R package xgboost (version 1.6.0.1) (for per-spot regression of morphological categories), GSVA R package version 1.44.5 (for the gene set variation analysis), xCell package (https://link.springer.com/protocol/10.1007/978-1-0716-0327-7_19) (for cell type enrichment analysis), R package NbClust (for the optimal number of clusters <spatial archetypes).
All statistical analyses were performed using the R software (v4.2.1).

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 raw sequencing files for the ST data and bulk RNA-seq of the ST TNBC cohort generated in this study have been deposited at the European Genome-Phenome Archive (EGA) under accession code EGAS50000000475 ([https://ega-archive.org/datasets/EGAD50000000686]). The raw sequencing data are available under restricted access due to data privacy laws. The data can be obtained upon signature of a data access agreement (DAA) between the investigator requesting the access and Institut Jules Bordet (IJB), subject to applicable laws. Access requests can be initiated by email to the corresponding author (christos.sotiriou@hubruxelles.be) with an approximate timeframe to reply of 4 weeks. The conditions related to the access to the data are specified in the DAA. In details, the raw data are accessible for reproducibility purposes and for academic and non-academic investigators aiming to perform original research. The data can be used for a maximum of 3 years after its reception.
The processed count matrices derived from the raw ST data and the associated brightfield images (H/E-images) are available without any restrictions via the repository Zenodo ([https://zenodo.org/doi/10.5281/zenodo.8135721]). QC filtering, IHC images, morphological annotations and all the projections on ST slides (ie. morphological regression, clusters, megaclusters, signatures) are accessible via the repository. The public data sets can all be accessed via https://www.cbioportal.org/study/summary?id=brca_metabric for METABRIC, https://data.mendeley.com/datasets/yzxtxn4nmd for SCAN-B, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE194040 for I-SPY2 and https://www.orcestra.ca/clinical_icb for other ICB-treated metastatic cancers. The same public data sets are also accessible via the repository.
Other data supporting the findings of this study are available within the article. Supplementary information, supplementary data and source data are provided with this paper. MSigDB is available at https://www.gsea-msigdb.org/gsea/msigdb. The sources of other gene signatures are provided in the Supplementary Table 4.

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	Our study was focused on breast cancer patients which were exclusively women in the retrospective cohort. We did not consider sex and gender as stratification factors in our study design due to the higher prevalence of breast cancer among female patients. Clinical characteristics of enrolled patients are provided in the Supplementary Data 1 and summarized in the Supplementary Table 1.
Reporting on race, ethnicity, or other socially relevant groupings	Data on race, ethnicity or other socially relevant groupings were not routinely collected as part of the study and were not controlled for.
Population characteristics	Our study analyzed data from 94 samples of 92 early-stage TNBC patients treated with standard-of-care therapies between 2000 and 2016. All patients underwent initial surgery, followed by adjuvant chemotherapy and/or radiotherapy. Clinical characteristics of enrolled patients is provided in the Supplementary Data 1 and summarized in the Supplementary Table 1.
Recruitment	For each patient, a frozen surgical tumor tissue sample stored at −80 °C was collected from the institutional tissue bank. The patients provided written informed consent for the storage of their tumor tissue, which was not dedicated for routine use, at the tissue bank.
Ethics oversight	The samples were collected in concordance with the Declaration of Helsinki. The study was approved by the Local Ethical Review Board of the Institut Jules Bordet (Brussels, Belgium) before initiation of the work (research ethics reference: CE2862).

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

Field-specific reporting

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was not predetermined. TNBC samples were chosen based on the availability of banked materials and their negative status for estrogen (ER) and progesterone (PR) receptors and the absence of HER2 amplification. No other selection criteria were used.
Data exclusions	Two collected samples were excluded from analysis due to ST sequencing failure.
Replication	We verified reproducibility of gene expression profiles across duplicates or tirplicates of ST slides (or subarrays) for each sample (ie. spatial distribution of ST TLS signature or individual clusters). Two or three consecutive ST tissue sections, named subarrays were available for each TNBC sample. These tissue sections were distant from each other only by 16 microns, and we found that there were relatively little biological differences between them. For the analysis, sections were considered as technical replicates. In order to compare them, we first needed to make them as stackable as possible using rotation, translation, and possibly reversion (if the tissue section was flipped upside down before being placed on the glass slide). These transformations allowed the analysis for each TNBC sample across the 3D axis and contributed to the identification of the 'spatial genes' across the 2 or 3 consecutive ST tissue sections (described in Methods: "Intrapatient clustering"). The transformations were performed compared to an unmodified reference slide, that was always the one that had been annotated by the pathologist. To validate our ST results, we performed also bulk RNA-seq on the ST TNBC samples and correlated the results between bulk and ST pseudobulk RNA data (ie. distribution of TNBC molecular subtypes, spatial archetypes or megaclusters). Then, we externally validated our findings (ie. predictive and pronostic value of our ST TLS signature, molecular features of spatial archetypes across different external breast cancer datasets: METABRIC (N = 335 TNBCs), SCAN-B (N = 672 TNBCs) and I-SPY2 (N = 987 BCs including 363 TNBCs) and fourteen datasets with diverse metastatic non-BC cancers (N = 1073 patients).
Randomization	Given the exploratory nature of our research, randomization of patients was not performed.
Blinding	Due to the absence of experimental groups, no blinding was performed.

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

Antibodies

Antibodies used	-anti-CD20 antibody (ready-to-use mouse monoclonal, IR604, Dako, USA) -anti-CD3 antibody (ready-to-use rabbit monoclonal, IR503, Dako, USA).
Validation	The double immunohistochemistry CD20 and CD3 was performed according to the previously published protocol by Buisseret, L. et al (doi: 10.1080/2162402X.2016.1257452).

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.