nature research

Corresponding author(s): A/Prof. Alexander Swarbrick

Last updated by author(s): Sep 11, 2020

Reporting Summary

Nature Research 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 Research 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	No specific software was used for the acquisition of the data.
Data analysis	Data analysis was performed using R (v3.5.0 and v3.6.1) with the following packages: Seurat (v3.0.0- for scRNA-Seq and v3.2.0 for ST analysis), DropletUtils (v1.2.2; for EmptyDrops), Garnett (v0.1.4), cluster (v2.0.7-1), CITE-seq-count (v.1.4.3), cola (v1.2.0), MAST (v1.12.0 - as part of DWLS method, v1.8.2 for differential gene expression), clusterProfiler (3.14.0), pheatmap (v1.0.12), Monocle (v2.10.1), STUtility (v0.1.0), AUCell (v1.4.1), survival (v2.44-1.1), survminer (v0.4.7), DWLS (code obtained from https://github.com/dtsoucas/DWLS on 30/11/2020). Additional software included Cellranger (v2.0, 10X Genomics), Space Ranger (v1.0.0, 10X Genomics), Loupe (v4.0.0, 10X Genomics) CIBERSORTx (v1.0), snakemake (v5.5.4), stereoscope (v0.2.0) and QuPath (v0.2.0). Description of the data analysis described in methods can be found at https://github.com/Swarbricklab-code/BrCa_cell_atlas.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

All processed scRNA-seq data is available for in-browser exploration and download through the Broad Single-Cell portal at https://singlecell.broadinstitute.org/single_cell/study/SCP1039. Raw scRNA-Seq data from this study has been deposited in the European Genome-Phenome Archive (EGA), which is hosted by the EBI and the CRG, under the accession code EGAS00001005173. All ST data from this study is available from the Zenodo data repository (DOI: 10.5281/zenodo.4739739).

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	Given the discovery and exploratory nature of our study, no statistical methods were applied to determine sample size. The number of tumor tissues analyzed in this study (26x scRNA-Seq and 6x spatial transcriptomics) reflect the availability and accessibility of the patient cohort to represent the major subtypes of breast cancer, as well as the funding limitations of the project grant.
Data exclusions	For filtering our scRNA-Seq datasets, we excluded cell barcodes determined as 'emptyDroplets' using the DropletUtils (v1.2.2) package. In addition, we excluded poor quality cells with genes and unique molecular identifier (UMI) counts less than 200 and 250, respectively, and a mitochondrial percentage less than 20%. For spatial data, all spatial barcodes that did not fall under tissue, as determined using the Space Ranger software (10X Genomics) were excluded. These steps are described in the manuscript
Replication	The single-cell RNA-Sequencing and spatial transcriptomics experiments described in this study consisted of a an independent single replicate per patient tumor. This was primarily due to the limited tissue samples collected from clinical specimens, as well as funding limitations, and is typical for this field
Randomization	Considering the exploratory nature of our clinical multi-omics study, randomization was not generally relevant for our study.
Blinding	The investigators had no prior knowledge of patient identities and clinical information prior to the collection of scRNA-Seq data. For scRNA-Seq data analysis, investigators were not blinded to this clinical information, as this was required to guide and design the analysis. For the spatial transcriptomics analysis, investigators were not blinded to clinical information prior to collection of data, as this was required for the selection of samples.

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	Details of all commercial antibodies used in this study can be found in Supplementary Table 11. All antibodies were used at a dilution of 1:100, as recommended by the manufacturer.

Validation	All commercial antibodies (Supplementary Table 11) used for CITE-Seq in this study were acquired from Biolegend and validated by the vendor by flow cytometry.

Human research participants

Policy information about studies involving human research participants

Population characteristics	All patients involved in this study were females aged between 35-88 with breast cancer across the three major clinical subtypes (ER+, HER2+ and TNBC). Most of the patients in our study had not undergone treatment. Detailed clinical information can be found in Supplementary Table 1.
Recruitment	Patients who fit the clinical criteria and consented to the study were selected for inclusion for multi-omics analysis. Our study had no self-selection bias or other biases in the recruitment of patients.
Ethics oversight	Primary untreated breast cancers (Supplementary Table 1) were collected with written consent from all patients under the protocols x13-0133, x19-0496, x16-018 and x17-155 with approval from all relevant human research ethics committees (Sydney Local Health District Ethics Committee, Royal Prince Alfred Hospital zone, and the St Vincent's hospital Ethics Committee). Consent included the use of all de-identified patient data for publication. Participants were not compensated.

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