nature portfolio

Corresponding author(s): Gabe S. Sonke

Last updated by author(s): Marte C. Liefaard

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.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

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	Data was collected from the TRAIN-2 clinical trial database and by requesting tumor biopsy material and related data through PALGA and the NKR. RNA-sequencing data was provided by Agendia.
Data analysis	Gene expression data was obtained by RNA-sequencing by Agendia. All analyses were performed with R version 4.2.1

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 clinical data and TIL dataset generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.
The raw RNAseq data for this study were generated by Agendia. This data is available upon reasonable request from the corresponding author with the permission of Agendia NV.

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	All patients included in the TRAIN-2 trial and thus in this study, were female.
Reporting on race, ethnicity, or other socially relevant groupings	In our study, race and ethnicity was not available and is not used in analysis.
Population characteristics	The population consists of adult patients with stage II-III HER2 positive breast cancer who were treated in context of the TRAIN-2 trial.
Recruitment	Patients were recruited in multiple centers across the Netherlands for the TRAIN-2 trial.
Ethics oversight	The TRAIN-2 trial was approved by the ethics committee of the Netherlands Cancer Institute. In addition, this specific project was approved by the Institutional Review Board of the Netherlands Cancer Institute (CFMPB672).

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.

[X] 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	Tumor biopsies were requested for all patients of the TRAIN-2 trial, and for 389 TIL scores were available.
Data exclusions	Patients were excluded if they had no tumor biopsy available or if the TILs could not be scored. In some analyses, patients were excluded because of missing covariates or outcome variables.
Replication	In this study, validation of findings from a TRYPHAENA sub study was attempted.
Randomization	In the TRAIN-2 study, patients were randomized between two treatment arms (neoadjuvant treatment with trastuzumab and pertuzumab combined with either anthracycline-free vs. anthracycline containing neoadjuvant chemotherapy)
Blinding	Pathologists that scored the TILs were blinded for outcome. No other blinding is applicable.

Behavioural & social sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing	
Data exclusions	
Non-participation	
Randomization	

Ecological, evolutionary & environmental sciences study design

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

Study description	
Research sample	
Sampling strategy	
Data collection	
Timing and spatial scale	
Data exclusions	
Reproducibility	
Randomization	
Blinding	

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	
Location	
Access & import/export	
Disturbance	

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

Type text here

Methods

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

Antibodies

Antibodies used	
Validation	

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	
Authentication	
Mycoplasma contamination	
Commonly misidentified lines (See ICLAC register)	

Palaeontology and Archaeology

Specimen provenance	
Specimen deposition	
Dating methods	

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	
Wild animals	
Reporting on sex	
Field-collected samples	
Ethics oversight	

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT01996267
Study protocol	This is a retrospective translational in patient materials from the TRAIN-2 clinical trial. The trial itself has been previously published.
Data collection	Clinical data was collected as part of the trial. Tumor biopsies were requested for this specific research question.
Outcomes	Outcomes of the TRAIN-2 trial were previously published. The current study shows that TILs may be of prognostic value in this patient group.

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
X	☐	Public health
X	☐	National security
X	☐	Crops and/or livestock
X	☐	Ecosystems
X	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
X	☐	Demonstrate how to render a vaccine ineffective
X	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
X	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
X	☐	Increase transmissibility of a pathogen
X	☐	Alter the host range of a pathogen
X	☐	Enable evasion of diagnostic/detection modalities
X	☐	Enable the weaponization of a biological agent or toxin
X	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Novel plant genotypes

Authentication

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Files in database submission

Genome browser session
(e.g. UCSC)

Methodology

Replicates

Sequencing depth

Antibodies

Peak calling parameters

Data quality

Software

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	
Instrument	
Software	
Cell population abundance	
Gating strategy	

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	
Design specifications	
Behavioral performance measures	

Imaging type(s)	
Field strength	
Sequence & imaging parameters	
Area of acquisition	

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	
Normalization	
Normalization template	
Noise and artifact removal	
Volume censoring	

Statistical modeling & inference

Model type and settings	
Effect(s) tested	

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

(See Eklund et al. 2016)

Correction

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis