nature portfolio

Corresponding author(s): Viktor A. Adalsteinsson, Todd R. Golub and G. Mike Makrigiorgos

Last updated by author(s): Jan 28, 2022

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	biopython 1.73, bwa 0.7.17, click 7.0, fgbio 0.6.1, gatk 3.8, numpy 1.16.2, pandas 0.25.3, picard 2.18.26, pysam 0.15.2, python 3.6.8, samtools 1.9, snakemake S.5.2. Custom code for designing MAESTRO probes from somatic variant calls can be found at https://github.com/broadinstitute/MAESTRO-probe_designer.
Data analysis	jupyter 1.0.0, jupyterlab 0.35.4, matplotlib 3.0.2, matplotlib-venn 0.11.5, numpy 1.16.2, pandas 0.25.3, python 3.6.8, scipy 1.2.1, seaborn 0.9.0.

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

The data supporting the results in this study are available within the paper and its Supplementary Information. Sequencing data have been deposited into the controlled-access database Data Use Oversight System (DUOS; http://duos.broadinstitute.org) under the accession number DUOS-000135.

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	We chose 18x replicates of 1/100k, and 18x replicate negative controls. These were the maximum amount of samples that we could fit on the remaining open lanes of the HiSeqX flow cells that we were preparing.
Data exclusions	Samples used in Fig. 2 used a panel containing 440 sites, whereas samples in Fig. 3 and Supplementary Fig. 5 used a panel containing 466 sites for MAESTRO and 489 sites for Conventional. To simplify comparisons, we only included sites in common amongst the three panels, which resulted in 438 evaluable sites. This is described in Methods.
Replication	We created biological and technical replicates where possible, to assess the reproducibility of results. This information is included in the relevant figure captions for each experiment.
Randomization	All patients completed the following course of neoadjuvant Phase II therapy: Bevacizumab x 1 dose; Doxorubicin/Cyclophosphamide x 4 cycles plus Bevacizumab; Paclitaxel x 4 cycles plus Bevacizumab. We selected for analysis all enrolled patients with ER-negative, PR-negative, HER2-negative (triple-negative) breast cancer (TNBC).
Blinding	We were blinded to the outcomes when we performed MRD testing using the WES fingerprints. With WGS fingerprints, this was not possible, as we only performed MRD testing on patients who experienced recurrence.

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	NA19238, NA12878 (purified DNA from both purchased from Coriell).
Authentication	Coriell uses multiplexed PCR for 6 autosomal microsatellite markers.
Mycoplasma contamination	Cultures from Coriell are free of mycoplasma..
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Adult patients with HER2-negative breast cancer and a tumor size larger than 1.5 cm were prospectively enrolled to the Dana-Farber Cancer Institute IRB-approved treatment protocol 07130 (NCT00546156). All patients received preoperative chemotherapy with doxorubicin, cyclophosphamide, paclitaxel and bevacizumab. Healthy donor plasma and whole blood

	were obtained from Research Blood Components. DNA extracted from these samples were used as negative controls or within patient or healthy-donor cfDNA admixtures. Donors were unrelated and samples were de-identified before being sent to us.
Recruitment	Patients were recruited from the Dana-Farber Cancer Institute in Boston, MA. Healthy donor biological material was obtained from Research Blood Components. We selected for analysis all enrolled patients with ER-negative, PR-negative, HER2-negative (triple-negative) breast cancer (TNBC). For patients with sufficient tumor tissue, exome-sequencing identified mutations that we captured using the conventional assay. From within this cohort we identified four TNBC patients who had tested MRD-negative using the exome-wide panel but who experienced metastatic recurrence. For these patients, we applied MAESTRO to analyse genome-wide tumor mutations.
Ethics oversight	The IRB of the Dana-Farber Cancer Institute approved the study protocol.

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	https://clinicaltrials.gov/ct2/show/NCT00546156
Study protocol	Ths study protocol is available on request.
Data collection	Data collection occurred from 2007 through 2017 at the Dana-Farber Cancer Institute and Massachusetts General Hospital in Boston, MA.
Outcomes	The primary clinical outcome was a pathological complete response.