nature portfolio

Corresponding author(s): Fabrice André

Last updated by author(s): September 22nd 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	Data collection was perfomed using FastQC v0.11.8, Fastp v0.20, BWA-MEM v0.7.17, Picard v2.20.3, SAMtools v1.9, GATK v4.1.8.1, mosdepth v0.2.5, SAMtools v1.9, Mutect2, FACETS R package v0.5.143, GenomicDataCommons (TCGA data)
Data analysis	For transcriptomic analyses GeoMX DSP Control Center v2.4.2.2, DESeq 2 v1.36.0, R v4.2.1, clusterProfiler v4.4.4 were used. Data regarding artificial intelligence analysis is available at the following code: https://github.com/loic-lb/Daisy-ML For the experiments in cell lines, statistical analysis was performed using GraphPad Prism 9 (GraphPad Software). Statistical analyses were carried out using Stata software v16 (StataCorp, Texas, US).

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

Artificial intelligence data and modalities for access are available at https://github.com/loic-lb/Daisy-ML
Genomic data and transcriptomic data with modalities for access are available at GDC portal , https://gdc.cancer.gov/about-data/publications/pancanatlas, TCGA_mastercalls.abs_segtabs.fixed.txt and TCGA_mastercalls.abs_tables_JSedit.fixed.txt.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Findings from DAISY trial apply to both sexes. Sex was determined based on self-reporting. Gender was not considered in study design.
Population characteristics	Patients presented metastatic breast cancer with any hormone receptor and HER2 status. For cancers, with expression of hormone receptors, the patients were included if they presented resistance to endocrine therapy and CDK4/6 inhibitors. For patients with HER2-overexpressing breast cancer they must have progressed on Trastuzumab and TDM1. Patients must have received at least one line of chemotherapy in the metastatic setting. Patients must have received at least one line of chemotherapy in the metastatic setting. In the DAISY trial, 186 patients were enrolled and 179 were included in the safety popuation. The median age of the population is 55 years. 99.4% of the patientes are women and 0.6% men. At baseline, 43% of patients presented a PS0 and 71.5% presented hormone receptor positive breast cancer. 64.8% of patients presented 3 or more metastatic sites at baseline and 57.6% of patients presented liver metastasis. 53% of patients received 5 or more lines of treatment in the metastatic setting. The median interval from initial diagnosis to metastatic disease was 25.8 months and the median interval from metastatic diagnosis to inclusion was 43.7 months.
Recruitment	Patients were recruited by oncologists from 15 investigator centers. We did not identify self-selection bias or other bias that could impact the results.
Ethics oversight	All patients who entered in DAISY trial signed an informed consent. DAISY trial was approved by the French ethic committee, CPP-Ile de France on September 05th 2019 and the French health authorities, ANSM, on July 08th 2019.

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	The primary endpoint was the confirmed objective response rate evaluated using RECIST 1.1 criteria. The required number of assessable patients for cohort 1 (n=67) and 2 (n=40) was determined using the A'Hern design with the following hypothesis: Cohort 1 (p0 = 30%; p1 = 45%, alpha = 5%, 1-beta = 80%) and Cohort 2 (p0 = 20%; p1 = 40%, alpha = 5%, 1-beta = 85%). The regimen would be declared promising in cohort 1 if 27 patients present a confirmed objective response among 67 and in cohort 2 if 13 confirmed objective response were observed among 40. Cohort 3 was designed using an optimal two stage design (alpha = 5%, 1-beta = 85%) with non-progression at 3 months as short-term endpoint (p20=30% and p21=50%) and confirmed objective response as primary endpoint (p10=20%, p11=40%). A stop for non-promising activity was planned to be declared if four patients or less among the first 16 present non-progressive disease at 3 months. At final analysis of cohort 3, the regimen would be defined as promising if 13 patients or more present a confirmed objective response. For each cohort, it was assume a rate of 10% non-evaluable patients and sample size was increased: Cohort 1: n=74, Cohort 2: n=44, Cohort 3: n=44. Full details are provided in statistical analysis plan.
Data exclusions	Of the 186 patients included in DAISY trial, seven patients who did not receive at least one dose of T-DXd were excluded from the safety population and two additional patients who did not have a valid first post-baseline assessment of disease status or who did not have progressive disease were excluded from the Full Analysis Set population as planned in the statistical analysis plan. Annex 3 includes all analyses that have been done by the statistician. For artificial intelligence analyses, six patients with no digitized HER2 slide at baseline or not exploitable and one patient presenting HER2 slide

at baseline scanned with another scanner were excluded from the analyses.
For T-Dxd distribution on-treatment, three pairs of biopsies were not analyzable because of absence of cancer cells.
For the RT-PCR tumor samples with <30% of tumor cells were excluded.
For the genomic analyses, samples that a) failed during the processing; b) tumor cellularity <30% were excluded

Replication: Since the paper reports a prospective clinical trial, there is no attemp to replicate the finding in the same paper. Nevertheless, the primary objective of the study was predefined and thus limits the risk of non replication.

Randomization: The trial was a phase II ,single arm, not randomized trial. Patients were allocated into the different cohorts according to HER2 level expression.

Blinding: Investigators were not blinded regarding cohort assignment.

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

Methods

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

Antibodies

Antibodies used: 4B5 pre-diluted kit (VENTANA pathway HER2, clone: 4B5, Roche Diagnostics) was used for HER2 staining on biopsies at progression and for T-DXd distribution in on-treatment samples. DXd-IgG was detected using primary antibody against DXd (antiXAFG5737-1A3-ocChimera, Daiichi Sankyo). The antibodies to perform the multiplex IF were directed against CD3 (clone BC33), CD4 (clone SP35), CD8 (clone C8/144B), CD68 (clone KP-1), FoxP3 (clone 236A/E7), PD-1 (clone CAL20), PD-L1 (clone 73-10), PanCK/SOX10 (clone AE1/AE3/BC34).

Validation: Antibodies used in the biomarker analyses were commercial and validated by the vendor

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s): MCF-7 and SK-BR-3 cells were purchased from DSMZ (Germany). These cell lines are derived from two female patients respectively with breast cancer.

Authentication: None of these cell lines were authenticated.

Mycoplasma contamination: The cell lines were not tested for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register): Not applicable

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: NCT04132960

Study protocol: Annexes 1 and 2

Data collection: The trial included patients from November 2019 to March 2021. Data were captured by clinical research assistants in the centers that participated to the trial. Trial monitoring was done on a regular basis by the sponsor UNICANCER.

Outcomes: The confirmed objective response rate evaluated in each cohort, is defined as the presence of a confirmed partial or complete response assessed by investigators. The investigator evaluated the objective response using RECIST v1.1. For the cohort 3 (IHC0+), a

short term primary endpoint is used for the interim analysis. The short term primary endpoint is the rate of patient without progression at 3 months. The investigator evaluated the progression using RECIST v1.1.
Progression-free survival is defined as the time from inclusion until progression or death from any cause. At the time of analysis, patients alive and without progression were censored at the date of the last tumor assessment.
Duration of response is applicable to subject with objective response, either complete response (CR) or partial response (PR), and is defined as the time from the first documented CR or PR until the date of disease progression, or until the date of death.
More details are available in the statistical analysis plan.