nature research

Corresponding author(s): David Hyman

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

The sample size required depends on the objective response rates at 8 weeks (ORR8). For each cohort, using Simon's optimal 2-stage design (with significance level 10% and power of 80%), a true ORR8 of 10% or less will be considered unacceptable (null hypothesis) whereas a true ORR8 of minimally 30% (alternative hypothesis) will merit further study. In the first stage, enrollment will continue until 7 patients per the Simon 2-stage optimal design. If no responses are observed, enrollment in the second stage for the cohort may be discontinued. Otherwise, the second stage will open and 11 additional response evaluable patients will be assessed for a total of 18 patients in the cohort. The null hypothesis will be rejected (for each cohort separately) if at least 4 responses are observed in each cohort. Once the Simon 2-stage criteria are met for the ERBB2 mutant breast cohort, enrollment into this cohort may continue until approximately 50 patients.

2. Data exclusions

Describe any data exclusions.

The clinical data presented represent an intention to treat population with all patients who received at least one dose of neratinib included in analysis. The genomic data presented include all samples that underwent sequencing and passed routine QA/QC procedures as described and referenced.

3. Replication

Describe whether the experimental findings were reliably reproduced.

As this is a clinical trial, no replication was possible or performed.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

There was no randomization.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

This was an open label study with no blinding.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	SAS EG5.1 by SAS Institute Inc., ABSOLUTE v. 1.0.6, FACETS v. 0.3.9, R v. 3.3.1, MSIsensor v. 0.2

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used
b. Describe the method of cell line authentication used.	No eukaryotic cell lines were used
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell lines were used
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No eukaryotic cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	A detailed summary of the demographics of the human research participants are included in Table 1 and Extended Data Table 1. In addition, patient level demographic data are provided in the cBioPortal project associated with this manuscript.