## 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 possible.*
- [ ] 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 software was used |
|-----------------|---------------------|
| Data analysis   | SAS Version 9.4     |

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 collected for the study will not be made publicly available. We encourage investigators interested in data sharing and collaboration to contact the corresponding authors (MO, AT)
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 | As the study was not prospective, the sample size was not predetermined. Greater than 1000 samples were analyzed which was felt, in discussion with the statistical team, felt to be sufficient. |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data was excluded from the analysis |
| Replication | Replication of core biopsy assessment was not possible as cores were sent for molecular testing after review by the pathologist |
| Randomization | Randomization was not relevant in this study as it was not a prospective study |
| Blinding | Blinding was not relevant in this study as it was not a prospective study |

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 | Methods |
|---------------------------------|---------|
| 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 |
| | n/a |
| | Involved in the study |
| | ChiP-seq |
| | Flow cytometry |
| | MRI based neuroimaging |

Clinical data

Policy information about clinical studies: nature.com/documents/nr-reporting-summary-flat.pdf

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 | This study was not a clinical trial |
|----------------------------|-----------------------------------|
| Study protocol | This was not a clinical trial; thus there was no protocol generated. |
| Data collection | Data was collected on site at MD Anderson Cancer Center in Houston, Texas between June 2019 and June 2020. Biopsies assessed for this analysis were collected between 2016 and 2019. |
| Outcomes | Outcome measures included malignant percentage (where cores <10% were described as inadequate) and malignant area. These endpoints were identified based on molecular testing requirements and discussion with pathology. |