

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 Any Processed data used in our analysis is available at <https://doi.org/10.6084/m9.figshare.25886674>

Data analysis Any custom code or software used in our analysis is available at https://github.com/isglobal-brge/Supplementary-Material/tree/master/Caceres_2024

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](#)

All public data, code, and materials used in the analysis are cited and referenced in the manuscript. The entire computer code with required data to replicate all reported findings can be found at https://github.com/isglobal-brge/Supplementary-Material/tree/master/Caceres_2024. The data from the TruDiagnostic Biobank is available upon reasonable request due to ensure the privacy of the participants. Please e-mail varun@trudiagnostic.com for data requests. The computer software, documentation and licensing to calculate X-Ra is available at <https://xra.isglobal.org>.

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 analyzed tumor data from 3,741 females in The Cancer Altas Genome (TCGA). We analyzed blood data from 1,414 females in the TruDiagnostic Biobank.
Data exclusions	As pre-established, samples with discordant sex, duplicated, high heterozygosity, and relatives were filtered out. Only Female data were analyzed.
Replication	For breast cancer replications, we used Gene Expression Omnibus data with accession codes GSE78754 (n=70), GSE141441 (n=95), GSE184159 (n=63), GSE225845 (n=224). For replications in blood we used GSE56046 (n=605), GSE210255 (n=986) and GSE142536 (n=51)
Randomization	The TruDiagnostic DNA biobank is a USA population-based cohort aged between 13 and 97 years old. As explained in the manuscript, we adjusted for all the necessary covariates.
Blinding	Non applicable.

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	n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies	<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines	<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology	<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms		
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data		
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern		

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Female cancer patients with tumor and undiseased biopsies (TCGA), and women from the general population tested for biological aging in blood samples (TruDiagnostic).
Recruitment	This study is based in public data and the TruDiagnostic Biobank data, which has been described elsewhere.
Ethics oversight	All studies received approval from the ethics committees of the centers involved and written informed consent was obtained from all participants.

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