

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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 were collected using Illumina sequencing machines
Data analysis	Burrows-Wheelers Aligner version 0.7.12, Samtools (v1.17), TelSeq (v0.0.1), TelomereHunter (v1.1.0), Picard (v4.0.1.2), Bedtools (v2.27.1), R (v4.0.0), Code is accessible at https://github.com/pughlab/TGL49_CHARM_LFS_Fragmentomics

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 sequencing data generated in this study have been deposited in the European Genome-Phenome Archive (EGA) database under accession code EGAS00001006539. The datasets are available under restricted access in compliance with patient consent for data sharing, access can be obtained by approval from the University Health Network data access committee (Contact person: Natalie Stickle, Email: natalie.stickle@uhn.ca). The remaining data, including de-identified

clinical data, are available within the Article, Supplementary Information, the Zenodo database under the accession code <http://doi.org/10.5281/zenodo.7448380>, or Source Data file. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	Sex for participants is reported in the Supplementary tables. Further analysis looking into sex and gender was not within the scope of this study.
Reporting on race, ethnicity, or other socially relevant groupings	Race, ethnicity, and other relevant social groupings were not reported or looked into in this study.
Population characteristics	Sex, genotype for TP53, and cancer status/treatment were recorded and reported for this study.
Recruitment	Patients enrolled in the LFS screening programs at the Princess Margaret Cancer Center and SickKids Hospital in Toronto were approached for enrollment in this study. No other selection was performed.
Ethics oversight	This study was approved by the UHN institutional review board (REB# 19-6239). All patients underwent routine clinical care by board certified clinicians as per the standard-of-care. All samples were collected with informed consent for research directly by the patient for patients >19 years of age or by a legal guardian for patients <19 years of age.

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](https://www.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	Sample size was determined based on the patients that could be recruited. Through this process we exhausted the eligible patient lists at our two institutions.
Data exclusions	Data were only excluded from analysis if they failed sequencing
Replication	We have included all code and data required to replicate all of our analyses
Randomization	Samples were binned based on cancer status and thus non random. Selection for machine learning training was randomized using R
Blinding	Analysis was not blinded as we were comparing between the groups. However, machine learning was blinded through random selection and 10-fold cross validation

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
☒	☐ Plants

Methods

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

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	Not a registered clinical trial
Study protocol	Not a clinical trial, prospective study utilizing clinical data
Data collection	Recruitment for patients was ongoing and sample collection was performed during their routine clinical followups which ranged from quarterly to yearly
Outcomes	Outcomes were defined based on their clinical cancer status as determined by routine surveillance methods

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>