nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

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
Sequencing data was generated from archived newborn dried blood spots using targeted next-generation sequencing. Raw sequencing reads were produced on an Illumina NovaSeq 6000 platform, generating FASTQ files that were processed and converted to variant call format (VCF) using the Sention bioinformatics pipeline within Fabric Enterprise (see below). Clinical and demographic data were obtained through linkage with the Michigan Cancer Surveillance Program and provided to investigators in de-identified form.
Variant annotation, interpretation, and case-level review were performed using Fabric Enterprise, release 6.50 (December 2024). Fabric platform components included Enterprise Versions API v3.2.0 and v4.0.0, Secondary v2.3, ACE Program v1.3.6, ACE Data v1.4.6, and Analysis Pipeline v7.7.0.
Software and database versions
Variant annotation and interpretation were performed using Fabric Enterprise (GRCh38), release 6.50 (December 2024), with the following software components and reference databases:
Fabric software components
• Fabric Enterprise Versions API: v3.2.0 and v4.0.0
• Fabric Enterprise Secondary: v2.3
• ACE (Artificial Intelligence–driven Classification Engine): Program v1.3.6; Data v1.4.6
• Analysis Pipeline: v7.7.0
Reference databases
• ClinVar: 2024-09-10
• OMIM: 2024-12-15
• Genomenon: 2024-07-03
• Ensembl: v113

- dbSNP: v150
- gnomAD: r3.2.1
- gnomAD-Mt: r3.1
- dbNSFP: v4.0a
- UMLS: 2024-AB
- GEM package: 2024-10-31_GEM_RC-D
- Human Phenotype Ontology (HPO): 2024-08-13
- Gene Ontology (GO): 2024-11-03

Data analysis

Data Analysis and Code
Single-nucleotide variants and small insertions and deletions (≤20 bp) were called using the Sention pipeline. Copy number variant detection was performed using CNVkit with custom parameters optimized for exon-level heterozygous deletion events affecting pediatric cancer predisposition genes. Variant annotation, prioritization, and classification were conducted using the Fabric Genomics Artificial Intelligence–driven Classification Engine (ACE) that utilize ACMG 2015 Classification. Sequencing enrichment and coverage metrics were assessed using Picard tools. Statistical analyses were performed using SAS version 9.4, with additional visualization and exploratory analyses conducted using cBioPortal.
No custom-written code was developed for this study. Data processing and analysis relied on validated commercial software platforms and established open-source tools. Open-source components (e.g., CNVkit, Picard) are publicly available, whereas proprietary software (Fabric Enterprise, ACE, and Sention) is available from the respective vendors.

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

Provide your data availability statement here.

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	Reporting on sex and gender Sex was recorded as assigned at birth in the Michigan Cancer Surveillance Program. Gender identity was not collected. Analyses were not stratified by sex or gender, as this was not a primary objective of the study.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity data were obtained from the Michigan Cancer Surveillance Program. These variables were collected through medical records and administrative sources. Race and ethnicity were not used as primary variables in the analyses but are reported descriptively where available.
Population characteristics	The study population consisted of children born in Michigan between 1987 and 2020 who were diagnosed with a malignant solid or central nervous system tumor before age 8 years and had an available archived newborn dried blood spot. The cohort included both males and females and represented the racial and ethnic diversity of the statewide population.
Recruitment	Participants were identified retrospectively through linkage of the Michigan Cancer Surveillance Program with the Michigan BioTrust for Health repository. No active recruitment was performed. Inclusion was based on predefined eligibility criteria, including cancer diagnosis, birth in Michigan, and availability of a newborn dried blood spot.
Ethics oversight	The study was approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board, the Michigan Department of Health and Human Services Institutional Review Board, and the Michigan BioTrust for Health Scientific Review Committee. Use of archived newborn dried blood spots was conducted under a waiver of consent for samples collected prior to 2010 and under parental consent for samples collected thereafter, in accordance with Michigan BioTrust policy.

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	Sample size was determined based upon a decision to assemble a population-based cohort. That is, all babies with DBS born in Michigan between 1987 and 2020 who developed a solid or brain malignancy by age 8 years. Only exclusions were: germ cell tumors and skin cancers and, for babies born after 2010, only those whose parents gave permission for the DBS to be used in research (about 30% declined).
Data exclusions	See above
Replication	No attempts at replication. Review of data in well newborns confirms genetic risk prevalence (but scant data).
Randomization	N/A
Blinding	N/A

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

Plants

Seed stocks	none
Novel plant genotypes	none
Authentication	none