nature portfolio

Corresponding author(s): Kang Zhang, Wenhao Zhou, Hao Li, and Fei Liu

Last updated by author(s): Mar 2, 2026

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	No special software or code was used to collect the data
Data analysis	Statistical analysis was performed using Stata version 20 (StataCorp; College Station, TX, USA). The annotation tool CLEVER used in this study is deposited on Zenodo (DOI: 10.5281/zenodo.15534528). The home-made pipeline for SNV/CNV variant calling and annotation from fastq files is deposited on Zenodo (DOI: 10.5281/zenodo.16892762). The analysis pipeline and code to reproduce main results is deposited on Zenodo (DOI: 10.5281/zenodo.17032133). Besides, the code is available from the developer upon reasonable request made to the corresponding author.

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

Data may be available for research purpose from the corresponding authors on reasonable request.

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	We used "sex" in our manuscript because we considered the biological attributes.
Reporting on race, ethnicity, or other socially relevant groupings	Information on race, ethnicity or other socially relevant groupings was not available for all included datasets.
Population characteristics	We established an integrated cohort by combining retrospective clinical features and existing NGS databases with prospective follow-up data, comprising a large cohort of pediatric patients who underwent genetic testing for suspected hereditary diseases at the Clinical Genetics Laboratory of the Children's Hospital of Fudan University or Guangzhou Women and Children's Medical Center between January 1, 2016, and January 31, 2025. Clinical data and follow-up information were collected through electronic medical records and telephone interviews. Totally, 75602 cases were enrolled. And clinical data and genetic informations were used for analyzed. For the publication of findings, all data presented in this manuscript have been rigorously aggregated and anonymized to prevent patient re-identification. All of the participants agreed to publish their results anonymously. Sex analysis was not carried out in this study. The primary objective of this study was to define the spectrum and penetrance of germline variants in a large, unselected pediatric clinical cohort and to estimate the associated population-level cancer risk. The study was not specifically powered or designed to test hypotheses regarding sex-based biological differences in cancer predisposition. Therefore, all primary analyses (e.g., variant prevalence, incidence rates, survival analyses) were performed on the aggregate cohort.
Recruitment	All participants provided written informed consents by guardians before enrollment. Participants were recruited as a consecutive, unselected clinical referral cohort comprising all pediatric patients (n=75,602) referred for genetic testing of suspected hereditary disorders at the participating centers between 2016 and 2025. This design inherently introduces a referral or ascertainment bias, as the cohort represents children with clinical indications severe enough to warrant genetic testing, not the general pediatric population. This bias likely leads to an overestimation of the prevalence of pathogenic germline variants and the associated cancer risks compared to the general population, but accurately reflects the risk within this high-risk clinical subgroup.
Ethics oversight	The study was approved by the Ethics Committee of Guangzhou Women and Children's Medical Center (approval numbers: 2025111A01) and the Ethics Committee of the Children's Hospital of Fudan University (approval number: 2024-286)

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	Totally, 75,602 cases were enrolled in this study. A formal a priori sample size calculation was not performed. This was a retrospective analysis of a consecutive, unselected clinical cohort comprising all pediatric patients referred for genetic testing at the participating centers between 2016 and 2023, constituting the entire available population during that period. The exceptionally large sample size provides substantial statistical power to detect the effects of interest (e.g., differences in tumor incidence rates) and ensures precise estimates for variant prevalence and associated risks.
Data exclusions	The exclusion criteria were patients who lack of available clinical data.

Replication	Not relevant. The present study is an observational analysis of a unique, real-world clinical cohort. Its core data, the patient population, their inherent germline variants, and the occurrence of tumors, are intrinsic clinical facts, not the outcome of a controlled experiment, and therefore cannot be replicated or verified through repeated experimentation.
Randomization	Not relevant. This study was an observational cohort study, not an intervention-based experiment. Therefore, participants were not "allocated" into groups. They were categorized based on their inherent clinical status at the time of genetic testing (forming the "pre-existing tumor group" vs. the "follow-up group") or based on the objective results of their genetic testing (e.g., carriers of P/LP variants). As this grouping reflects real-world clinical presentation and laboratory findings, random allocation was neither performed nor applicable.
Blinding	All data were de-identified to remove any patient-related information. A formal blinding procedure was not applicable or relevant in this study, which was an observational, retrospective-prospective cohort study that analyzed pre-existing, objective data: germline variant classifications were determined by standardized pipelines and ACMG guidelines, and tumor diagnoses were based on clinical records, not on investigator assessment. Since the core data points (genetic status and cancer occurrence) were not subject to investigator interpretation or measurement bias, blinding to group allocation was not a methodological concern.

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	Our study retrospective analyzed cases with genetic test in Children's hospitals. Given its non-interventional and absence of controlled cases, we did not register this study in ClinicalTrials.gov or any equivalent agency.
Study protocol	Study protocol is available in our previous studies.
Data collection	We established an integrated cohort by combining retrospective clinical features and existing NGS databases with prospective follow-up data, comprising a large cohort of pediatric patients who underwent genetic testing for suspected hereditary diseases at the Clinical Genetics Laboratory of the Children's Hospital of Fudan University or Guangzhou Women and Children's Medical Center between January 1, 2016, and January 31, 2025. Patients younger than 3 years were included. The exclusion criteria were individual patients who lacked available clinical data. The aim is to re-analyze cancer risk in pediatric patients focusing on a pediatric cancer-predisposition gene panel. Patients were prospectively followed from the time of genetic testing until either tumor onset or May 31, 2025, whichever occurred first. Retrospective clinical data from birth to the time of NGS testing were retrieved from the electronic medical record system.
Outcomes	The main focus of our study was to assess the rate of tumor occurence in patients with germline variants. Additionally, as part of our secondary outcomes, we sought to determine the genetic spectrums in tumor patients.

Plants

Seed stocks	*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.*
Novel plant genotypes	*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.*
Authentication	*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, mosiacism, off-target gene editing) were examined.*