nature portfolio

Corresponding author(s): Christina Curtis

Last updated by author(s): Oct 28, 2025

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 software was used to collect the sequencing data.
Data analysis	As described in the Methods, published analysis tools and standard Python and R statistics and visualization packages were used to analyze the data and produce the figures. These include: BWA MEM(v1.10) GATK(v3.4-46) Picard-MarkDuplicates(v2.18.7) nf-core/sarek(v2.7.1) Mutect2 (GATK)(v4.1.7.0) Strelka2(v2.9.10) Manta(v1.6.0) vcf2maf(v1.6.19) VariantFilter- https://github.com/rschenck/VariantFilter/tree/strelka2 VEP(v103.1) FACETS-SUITE(v2.0.8) FACETS(v0.6.1) Sequenza(v3.0.0) mosdepth(0.3.2)

maftools(v2.17.0)
MutationalPatterns(v3.9.1)
TelSeq(v0.0.2)
Biopython(v1.85)

Source code is available via GitHub: https://github.com/cancersysbio/HTAN-FAP and an archived copy of the code is available on Zenodo (https://doi.org/10.5281/zenodo.17372231)

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

Bulk DNA sequencing data and metadata are available at the HTAN portal: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002371.v3.p1. Additional intermediate data files used for downstream analyses are available in Zenodo (doi: 10.5281/zenodo.13228022). Single-crypt WGS data is available at https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA023981.

Whole-exome sequencing data for the sporadic CRC cohort are available through the European Genome-phenome Archive (EGA) under accession number EGAS00001003066 (https://ega-archive.org/studies/EGAS00001003066). The whole-exome sequencing data for the FAP multi-region cohort are available through the Genome Sequence Archive (GSA) under accession number HRA000127 (https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000127). To facilitate comparisons, both of these published cohorts were uniformly reprocessed using the same software versions within the Isabl platform, as was used for the HTAN Pre-cancer Atlas bulk WGS FAP cohort.

WGS/WES was mapped to human reference genome assembly GRCh38 using the following annotations:
gnomAD(r2.1.1) - https://gnomad.broadinstitute.org/
dbSNP(Release 138) - https://console.cloud.google.com/storage/browser/genomics-public-data/resources/broad/hg38/v0/
COSMIC(v3.2) - https://cosmic-blog.sanger.ac.uk/
Mills_and_1000G_gold_standard.indels: https://console.cloud.google.com/storage/browser/genomics-public-data/resources/broad/hg38/v0/

Exome capture kits:
- HTAN Pre-cancer atlas: SureSelect Human All Exon v6(V6)
- External FAP (multi-region) cohorts(HRA000127)
- Sporadic CRC cohort: Truseq DNA exome v1-2

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	Familial adenomatous polyposis (FAP) occurs with equal incidence based on sex and is not considered as a covariate for any analysis within the submitted manuscript. Any further sex and gender based analysis are not considered within the scope of this study.
Reporting on race, ethnicity, or other socially relevant groupings	No groupings based on race, ethnicity, or other social groupings were conducted as part of this study.
Population characteristics	Only patients with a confirmed diagnosis of familial adenomatous polyposis (FAP) were included within this study. No considerations around race, ethnicity, or other population based covariates were warranted.
Recruitment	Patients with familial adenomatous polyposis (FAP) receiving care at Stanford Health Care were considered for inclusion and consented to participate in an IRB-approved protocol prior to receiving a partial or total colectomy. This protocol support the generation of PreCancer Atlas of Familial Adenomatous Polyposis (FAP), including the bulk WGS reported in this study. Additional samples from FAP patients undergoing colectomy were collected under an IRB-approved protocol and the Sixth Affiliated Hospital of Sun Yat-sen University.
Ethics oversight	Ethical approval for the PreCancer Atlas of Familial Adenomatous Polyposis (FAP) protocol was granted through Stanford University's Research Compliance Office International Review Board (IRB) under protocol IRB-47044. Additional data The single-crypt WGS sample collection from FAP patients was overseen by the IRB of the Sixth Affiliated Hospital of Sun Yat-sen University (2019ZSLYEC-06).

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	Eligible patients with FAP who consented for inclusion in the IRB-approved protocols were included in the study. This includes six patients, n=123 samples with WGS and/or WES from the HTAN PreCancer Atlas of FAP and two patients, n=129 samples with single gland WGS from Sun Yat-sen University.
Data exclusions	No WGS/WES samples from the bulk FAP HTAN PreCancer Atlas were excluded. Due to the challenges in generating high-quality single gland WGS, we implemented rigorous sample-level quality filters (described in the Supplementary Methods and Supplementary Note 3) which excluded 85/129 sgWGS samples.
Replication	Replicates were not appropriate for this observational study based on the genomic profiles of individual patients. However, the genomic landscape of the HTAN PreCancer Atlas FAP cohort was compared with bulk genomic data from a published FAP cohort and a sporadic colorectal cancer cohort.
Randomization	Randomization was not warranted for this observational study; interventions were not assessed. We report participants age and gender but do not control for these covariates given the small sample sizes available for studying this rare condition (FAP).
Blinding	Blinding was not relevant to this observational 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

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	*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.*