nature portfolio

Corresponding author(s): NCOMMS-23-39771

Last updated by author(s): Sep 4, 2023

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
Data analysis	Metagenome-Atlas 2.4.3, VirSorter 2.2.2, CheckV 0.8.1, Galaha 0.3.1, BBMap 38.96, SAM tools 1.15.1, vConTACT 0.11.0, Prodigal 2.6.3, DRAM1, graphanalyzer 1.4.6, Snakemake, Cytoscape 3.9.0, R-vegan 2.6.2, R-MaAslin2 1.12, R-micEco 0.9.15, custom: VirMake (https://github.com/uio-bmi/VirMake), custom scripts (available at https://github.com/Rounge-lab/CRCbiome_virome_2023)

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

FASTA files of CRCbiome vOTUs detected in 5 or more individuals, are available at ENA with accession number (data submission in process). Statistics on vOTUs are available as supplementary Data 1. Due to the sensitive nature of the remaining data derived from human subjects, processing of data and/or biological material

from this project must comply with the General Data Protection Regulation (GDPR). Data processing must have approval from the Regional Committee for Medical Research in Norway (REC). Furthermore, the processing needs legal basis according to GDPR Article 6 and 9 and the need for a Data Protection Impact Assessment (DPIA) according to GDPR article 35 must be considered. Requests for data access can be directed to Trine B Rounge, trinro@uio.no.

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	The study comprised 582 males (56.3 %) and 452 females (43.7 %). The data on sex was retrieved from the national population registry. All participants signed an informed consent which included permission to retrieve information from national registries. No data on gender was collected or analyzed. Sex data were used as an outcome variable in the differential vOTU abundance and alpha- and beta-diversity analyses.
Reporting on race, ethnicity, or other socially relevant groupings	Self-reported information on nationality (native Norwegian/non-native heritage), marital status (married or cohabiting/not married and not cohabiting), level of education (primary school/high school/university or college) and working status (employed/retired or unemployed) was collected. These data were used as outcome variables in the differential abundance and alpha- and beta-diversity analyses.
Population characteristics	The CRCbiome enrolled individuals aged 55-76 who had a positive FIT test, were scheduled for the subsequent colonoscopy and had no previous medical record of colorectal cancer. After colonoscopy, participants were grouped into four main categories: no confirmed neoplastic findings (n = 169); non-advanced lesions (n = 423); advanced lesions (n = 376); and CRC (n = 66).
Recruitment	The recruitment protocol, as well as inclusion and exclusion criteria are provided in detail in the CRCbiome study design paper (https://bmccancer.biomedcentral.com/articles/10.1186/s12885-021-08640-8) To comply with the overall aim of the CRCbiome study, the proportion of individuals with premalignant or malignant colorectal cancer lesions was higher than in the general population. Sensitivity analyses excluding participants with a malignancy did not impact the outcomes reported in this paper.
Ethics oversight	The Norwegian Regional Committees for Medical and Health Research Ethics (Approval no.: 63148)

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	Details on the sample size selection are provided in the CRCbiome study design paper (https://bmccancer.biomedcentral.com/articles/10.1186/s12885-021-08640-8). No other sample size considerations were made for this paper. The final dataset included 1034 individuals.
Data exclusions	Samples with < 0.7 ng/ul DNA concentration (n = 14), or a depth of less than 1 Gb sequencing data after quality filtering (n = 9) were excluded from the analysis.
Replication	The ZymoBIOMICS Microbial Community Standard was used as a positive control (n = 2). We reproducibly generated microbial profiles for these samples but they did not contain viruses and were thus not reported in this paper. Negative controls for DNA extraction (n = 6) and for library prep (n = 2) yielded a total of 32 and 3 QC sequencing reads, respectively.
Randomization	No randomization was performed for this descriptive study.
Blinding	The lab personnel was blinded to the colonoscopy result during the DNA extraction, sequencing library prep and bioinformatics analysis.

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	The BCSN study (parent study to the CRCbiome): NCT01538550
Study protocol	The CRCbiome (nested within BCSN study): https://bmccancer.biomedcentral.com/articles/10.1186/s12885-021-08640-8 The BCSN study: https://classic.clinicaltrials.gov/ct2/show/NCT01538550
Data collection	Locales: two hospital catchment areas in South-East Norway; recruitment: 2017 - March 2021; data collection: 2017-2021
Outcomes	Primary outcome of the CRCbiome study relates to the colonoscopy result. For this paper, no outcome was defined.