nature portfolio

Corresponding author(s): COMMSMED-24-0216-A
Insa Korten

Last updated by author(s): Jun 24, 2024

# 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<br>*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<br>*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 | Clinical data was stored in a secured Microsoft Access database. Microbiome raw reads were stored on High Performance Computing (HPC) of the University of Bern (UBELIX). |
| Data analysis | All stastical analyses were performed with R version 4.1.2.<br>Bioinformatic processing of 16S-rRNA data was performed using DADA2 (version 1.16.0) and taxonomy was assigned using a naïve Bayesian classifier using the Silva reference database (version 138.2) on UBELIX HPC. Decontamination was performed using decontam (version 1.12.0). Subsequent data-analysis was performed using the following R-packages: MaAsLin2 (version 1.6.0), vegan (version 2.5.7), phyloseq (version 1.46.0), mgcv (version 1.9-0), gtsummary (version 1.7.2), lme4 (version 1.1). Plots were generated using ggplot2 (version 3.4.1), ggrepel (version 0.9.5).<br>Figure 1 was designed with biorender.com |

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

Sequencing data generated during this study have been stored accessible for the public in the NCBI bioproject repository (https://www.ncbi.nlm.nih.gov/bioproject/ with the accession code "PRJNA1019921"). Study participant and further metadata is available upon 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 | In study design, participants of sex male and female were included. Sex was included in our statistical analyses as a fixed effect. We included infants, thus gender was not investigated. |
| Reporting on race, ethnicity, or other socially relevant groupings | Race: we did only include white people. This is unfortunate, but CF is a rare disease and even rarer in races other than white. This is stated in the Ethics statements and displayed in the previously published cohort profiles: BILD (PMID: 38061036) and SCILD (PMID: 29698544) study. No other socially relevant grouping. |
| Population characteristics | All population characteristics are available in table 1 of the manuscript. |
| Recruitment | SCILD-cohort (infants with CF): https://www.scild.ch/ Recruitment after diagnosis. Treating paediatricians from at CF centers in Switzerland ask all parents to participate in the study immediately after diagnosis (no selection bias expected).<br>BILD-cohort (healthy infants): https://www.bild-cohort.ch/ Pregnant women are recruited via their treating physicians / flyer / homepage. |
| Ethics oversight | Ethics committee of the canton Bern, Switzerland https://www.gsi.be.ch/de/start/ueber-uns/kommissionen-gsi/ethikkommission.html<br>SCILD 2017-02139<br>BILD 2019-01072 |

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 | The primary aim of this population-based study was to investigate the development and dynamics of the microbiota in infants with cystic fibrosis (CF) with special interest in the impact of the microbiota on susceptibility of lower respiratory tract infections (LRTI). The study had originally been powered based on the abundance and distribution of previously published microbiota data from infants (also our own), ensuring a power of 0.8 to detect at least significant differences in alpha- and beta-diversity between groups, as well as differences in abundance of the most abundant bacterial families (grouped amplicon sequence variants [ASVs]). We enrolled 50 infants with CF (SCILD cohort) and 30 carefully matched healthy controls (BILD cohort). Finally, we had complete datasets up to one year of age for 80 children, of whom all children had at least 15 high-quality samples available after microbiota laboratory workup (1,511samples).<br>Exemplary sample size calculations were done based on differences seen in the literature (e.g. Biesbroek et al. AJRCCM, 2014) Regarding relative effect sizes in this study (e.g. for increased abundance of lactic acid bacterium Dolosigranulum 2.61 or Corynebacterium 1.98), a statistical power of 0.9 and a significance level of 0.05 the sample size should be at least 7 infants per group. In our study, we compared 61 breastfed infants with 19 infants fed with formula milk.<br>The number of samples available for microbiota analyses reported in the current paper is larger compared to the initial paper (Mika et al., Lancet Respiratory Medicine, 2016). In this study, we could already show significant differences between the microbiota profiles of healthy infants and infants with CF (permutational multivariate ANOVA p=0.001) with even lower sample size (1300 nasal swabs vs 1500 nasal swabs).<br>To disentangle longitudinal relationships between LRTIs, disease and eg. antibiotic treatment on microbiota profiles, we increased the sample size of CF patients from 30 to 50 (antibiotic treatments in first year of life from n=50 to n=70) compared to our previous study. |
| Data exclusions | We excluded only nasal swabs that did not pass the quality control and filtering steps explained in the methods section of the paper. |

| | |
|---|---|
| Replication | We could replicate our own previous findings and published data regarding difference between the nasal microbiota of infants with CF and healthy controls (see manuscript section I). There was no data regarding differences between infants with CF with higher number of LRTIs compared to lower number of LRTIs. |
| Randomization | Randomization was not applicable for this observational study. |
| Blinding | Blinding of investigators was not performed, because this was not relevant to our 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 |

## 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 | This is no clinical trial, but the study cohorts are registered with the ethics project numbers.<br>SCILD 2017-02139<br>BILD 2019-01072 |
| Study protocol | A full study protocol can be assessed upon request. |
| Data collection | All data was collected in Switzerland between 2011-2019. All healthy infants were recruited in Bern. Infants with CF were recruited in<br>Bern (16 infants)<br>Zürich (11 infants)<br>Lausanne (7 infants)<br>Aarau (6 infants)<br>Basel (4 infants)<br>Genf (2 infants)<br>Lugano (2 infants)<br>St. Gallen (1 infants)<br>Luzern (1 infants) |
| Outcomes | This is a descriptive study of microbiota dynamics over time. Early-life microbiota composition and development were studied using PERMANOVA tests to assess the global impact of host/environmental factors on microbiota profiles and differences between groups. In addition smoothing spline generalized additive mixed models were used to assess the associations between LRTI susceptibility and alpha or beta diversity. Differential abundance of specific amplicon sequence variants (ASVs) was tested with longitudinal mixed models. |

## Plants

| | |
|---|---|
| Seed stocks | na |
| Novel plant genotypes | na |
| Authentication | na |