

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                                                                                                                                                                                                                                                                                      |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( <i>n</i> ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted<br><i>Give P values as exact values whenever suitable.</i>                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                                      |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i> ), 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 | Taxonomic annotation was performed using USEARCH by mapping the reads to the Ribosomal Database Project database version 18. Taxonomic annotation were verified using Blast RDP classifier and in cases of disagreement, the Blast annotation was used.                                                                                                                                                                                                 |
| Data analysis   | The statistical analysis was performed using R version 4.2.2. PCA was performed using vegan package v. 2.6.4. K-means clustering performed with the clustering package v.2.1.4. Partitioning tree for cluster membership was created using the rpart library v. 4.1.19. Developmental trajectories were created using gbmt library v0.1.3. The MARE library (V.1.0) which used the glm.nb function (MASS v.7.3.58.4) and gls function (nlme v.3.1.162). |

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](#)

The HEMLi microbiome 16s rRNA gene sequences in this study have been deposited in the European Nucleotide Archive (ENA) under accession code PRJEB5543

[<https://www.ebi.ac.uk/ena/browser/view/PRJEB55243>], along with limited metadata. Additional raw metadata are protected and are not available due to data privacy laws.

## 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                                        | Sex was self-reported by the parents. Infant sex was tested for associations with the microbiota wellbeing index, cluster membership and developmental trajectories, however no significant associations were reported. Additionally, infant sex was used as a confounder in analysing infant gut microbiota, cluster membership and developmental trajectory membership associations with later life health outcomes. The data consisted of 50.7% male and 49.3% female, 967 infants, 768 mothers, and 515 fathers.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Reporting on race, ethnicity, or other socially relevant groupings | Race, ethnicity or other social groupings were not considered in this study. At least one parent in each family had to be Finnish speaking as extensive questionnaires were an integral part of the study, and were recruited from the general population, mainly in the capital region of Finland. All information was self-reported.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Population characteristics                                         | Eight stool samples were collected at weeks 3, 6, 9, 12 and months 6, 9, 12, 18 and 24. Parents were asked to provide a spot stool sample around the time of childbirth, preferably within two weeks of the due date. Healthy term babies born on gestational weeks 37-42 without known congenital defects were included in the study.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Recruitment                                                        | Pregnant women with singleton gestation were recruited from the general population mainly in the capital region of Finland. Three methods were used to recruit participants: (1) displaying recruitment posters and flyers on antenatal clinics within the public health care system; (2) directly approaching potential participants at the antenatal clinics and perinatal ultrasound screenings, and (3) advertising the study on social media. Healthy term babies born on weeks 37-42 without known congenital defects were included in the study. At least one parent in each family had to be Finnish speaking as extensive questionnaires were integral to the study. A total of 1587 families expressed interest in participating in the HELMi study during the recruitment period (February 2016-March 2018). Out of the 1149 consented families, 86 families either withdrew study before the child was born or were excluded due to preterm, birth or birth defect. Altogether 1063 infants fulfilling the inclusion criteria were born, of which 1055 completed the data for childbirth details and infant gender. Study retention was 96.2% at the age of 3 months. In March 2019, when the youngest study children were 12 months, the retention rate was 91.7% (n=967). |
| Ethics oversight                                                   | The study was approved by the ethical committee of The Hospital District of Helsinki and Uusimaa, Finland (263/13/03/03 2015) and performed in accordance with the principles of the Helsinki Declaration.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

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](https://www.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     | 967 infants, 768 mothers, and 515 fathers, comprising 7211 fecal samples                                                                                                                                                                                                                                            |
| Data exclusions | The sequencing cut-off was chosen to be 3000 reads after QC to samples at 3 months or before and to 5000 paired reads for the remaining samples based on species richness to sequencing depth evaluations.                                                                                                          |
| Replication     | For replication all required R code used in the study is available at <a href="https://github.com/bhick001/Wellbeing-Index">https://github.com/bhick001/Wellbeing-Index</a> . HELMi microbiome 16 rRNA gene sequences have been deposited in the European Nucleotide Archive (ENA) with accession number PRJEB55243 |
| Randomization   | This was an observational study of a longitudinal cohort of infants. Randomization was therefore not relevant to this study                                                                                                                                                                                         |
| Blinding        | This was an observational study of a longitudinal cohort of infants. Randomization was therefore not relevant to this 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 &amp; experimental systems

|                                     |                                                        |
|-------------------------------------|--------------------------------------------------------|
| n/a                                 | Involved in the study                                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants                        |

## Methods

|                                     |                                                 |
|-------------------------------------|-------------------------------------------------|
| n/a                                 | Involved in the study                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> 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, mosaicism, off-target gene editing) were examined.