

## Reporting Summary

Nature Research 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 Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
- ☐ ☒ An indication of 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 statistics 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
- ☐ ☒ Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on [statistics for biologists](#) may be useful.

### Software and code

Policy information about [availability of computer code](#)

Data collection

FACSDiva software v. 6.1.3 (BD Biosciences, San Diego, CA, USA) was used for collection of flow cytometer data.

Data analysis

FCAP Array software (version 2.0, Soft Flow, Burnsville, MN, USA, <http://www.softflow.com>), FlowJo X 10.0.7r2 software (Treestar, Ashland, OR, USA), FLASH (Magoc, 2011), QIIME (Caporaso, 2010), OTU-picking algorithm (Rideout, 2014), Deblur (Amir, 2017), GUniFrac R-package (Chen et al., 2012), vegan R-package (<https://cran.r-project.org>, <https://github.com/vegandevs/vegan>), ANCOM v1.1-3 (Mandal et al., 2015), Oligotyping pipeline script v. 1.7 (Eren et al., 2014), mothur v. 1.35 (Schloss et al., 2009), SourceTracker v.1.0 (<https://github.com/danknights/sourcetracker>); PICRUSt (Langille et al., 2013; <http://picrust.github.io/picrust/>), STAMP software version 2.1.3 (Parks et al., 2014), SAS Enterprise Guide 5.1 (SAS Institute Inc., Cary, NC, USA)

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [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 list of figures that have associated raw data
- A description of any restrictions on data availability

The bacterial and fungal sequences were submitted in NCBI Sequence Read Archive (SRA) under the Bioproject ID PRJNAxxxxxx with the following BioSample numbers: SAMNxxxxxxx-SAMNxxxxxxx for bacterial and SAMNxxxxxxx-SAMNxxxxxxx for fungal data.

## Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample size     | In primary analyses all available specimens, and in the exploratory analyses all available relevant data, from the time points of interests were studied. While the LUKAS birth cohorts were not designed for these specific statistical analyses, our previous studies have demonstrated that the cohorts have an adequate power to study the effects of differences in indoor microbial exposures on the development of asthma (Karvonen 2009, 2014). |
| Data exclusions | Samples with no or too small amounts of dust as well as samples with low sequence yield in PCR amplification (threshold determined based on negative controls) or missing data on covariates or end outcome were excluded from sequencing and/or data analyses.                                                                                                                                                                                         |
| Replication     | The primary finding of the study on the asthma-protective effect of LUKAS1 farmhome-like indoor microbiota composition, as defined by Farm home Resembling Microbiota Index (FaRMI) and calculated as described in the paper, were reproduced in larger, cross-sectional GABRIELA cohort of which comparable microbiological data on indoor microbiota, asthma and covariates was available.                                                            |
| Randomization   | The participants were not allocated into experimental groups.                                                                                                                                                                                                                                                                                                                                                                                           |
| Blinding        | Blinding is not relevant to the study because there were no experimental group allocations.                                                                                                                                                                                                                                                                                                                                                             |

## Reporting for specific materials, systems and methods

### Materials & experimental systems

|                                     |                                                                 |
|-------------------------------------|-----------------------------------------------------------------|
| n/a                                 | Involved in the study                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Unique biological materials            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology                          |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms            |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Human research participants |

### Methods

|                                     |                                                    |
|-------------------------------------|----------------------------------------------------|
| n/a                                 | Involved in the study                              |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq                  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging    |

## Antibodies

Antibodies used

Antibody used in flow cytometry; Conjugate; Purpose; Clone; Isotype; Manufacturer;  
 CD14; PE-Cy5.5; marker for monocytes; TüK4; Mouse IgG2a; Invitrogen (Camarillo, CA, USA);  
 CD19; PE-Cy5.5; marker for B cells; SJ25-C1; Mouse IgG1; Invitrogen;  
 CD11c; PE-Cy7; specific to mDC1 and mDC2; 3.9; Mouse IgG1 κ; eBioscience (San Diego, CA, USA);  
 CD1c (BDCA-1); APC; specific to mDC1; AD5-8E7; Mouse IgG2a; Miltenyi Biotec (Bergisch Gladbach, Germany);  
 CD303 (BDCA-2); PE; specific to pDC; AC144; Mouse IgG1; Miltenyi Biotec ;  
 CD141 (BDCA-3); APC; specific to pDC (low), mDC1(low) and mDC2 (high); AD5-14H12; Mouse IgG1; Miltenyi Biotec ;

ILT4 (CD85d); FITC; tolerogenic properties; 27D6; Rat IgM; eBioscience;  
 ILT3-biotin; ; tolerogenic properties; ZM4.1; Mouse IgG1,  $\kappa$ ; Biolegend (San Diego, CA, USA);  
 IgG1; FITC; isotype control; X40; Mouse IgG1,  $\kappa$ ; BD Biosciences;  
 Rat IgM; FITC; isotype control; eBRM; Rat IgM  $\kappa$ ; eBioscience;  
 IgG1-biotin; ; isotype control; MOPC-21; Mouse IgG1,  $\kappa$ ; Biolegend;  
 IgG1-biotin; ; isotype control; P3.6.2.8.1; Mouse IgG1,  $\kappa$ ; eBioscience;  
 Streptavidin; APC-eFluor780; secondary antibody for biotin-conjugated antibodies; eBioscience;

Antibodies for interleukins (IL) 1b, 4, 5, 6, 10, 12p70, 13, 17A and interferon- $\gamma$  were those provided by BD Biosciences (Heidelberg, Germany) in the respective Opteia ELISA and BD human CBA flex kits.

## Validation

All antibodies used have been validated by the manufacturer for the purpose used in this study.

## Human research participants

### Policy information about studies involving human research participants

#### Population characteristics

LUKAS1 and 2 are birth cohorts described in Karvonen et al., 2009. LUKAS1 represents children living in rural Eastern or Middle Finland with one half living on a farm with livestock. LUKAS2 represents unselected population of mostly suburban children from Eastern Finland. The children were followed up from the third trimester of pregnancy to 6 years of age. Of the subjects with the indoor microbiota assessment at 2 months (N=395) 49.1% were girls and 50.9% boys.

Phase II GABRIELA is a cross-sectional study with population characteristics described in detail in Ege et al., 2011. The children were 6 to 10 years old and lived in the rural regions of Southern Germany around Munich or Ulm. Stratified, random subsampling has been performed for microbiological analyses based on three different levels of farm exposure from no exposure to intermediate exposure to living on a farm. Of the subjects with the indoor microbiota assessment (N=1031) 48.6% were girls and 51.4% boys.

#### Recruitment

The recruitment of LUKAS subjects was carried out as described before in Karvonen et al., 2009. In the LUKAS1 study, 214 children were recruited between September 2002 and May 2004. All pregnant women who lived on farms with livestock and an equal number of non-farming women who lived in rural areas were invited to the study at 20–34 weeks of gestation in the areas of four central hospitals in Eastern and Middle Finland (Kuopio, Jyväskylä, Joensuu and Iisalmi). Criteria for inclusion in the study were maternal age of  $\geq 18$  years, singleton pregnancy, mother tongue Finnish, and no plans to move from the study area. Exclusion criteria were the delivery before 36 weeks of gestation, congenital abnormalities in the infants, and a failure to obtain cord blood samples.

For LUKAS2 all the women who were estimated to give birth in the Kuopio University Hospital between May 2004 and May 2005 and who did not live in apartments were invited to participate in the study. For logistic reasons, the women delivering at the end of a week (from Thursday afternoon to Sunday morning) were excluded. Other inclusion and exclusion criteria were the same as in LUKAS1.

The recruitment for GABRIELA was performed as described in Ege et al., 2011 in summer 2006 to spring 2008. In Phase I the prevalence of asthma and allergic diseases, some farm related exposures and indicators of potential participation bias were assessed on a population level by a short questionnaire distributed to all parents of 6–12 year old school children through their elementary schools. In phase II stratified random samples of all children whose parents had given written informed consent to blood sampling, genetic analyses and dust sampling were studied ("GABRIELA study population").

## Flow Cytometry

### Plots

Confirm that:

- ☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- ☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- ☒ All plots are contour plots with outliers or pseudocolor plots.
- ☒ A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

#### Sample preparation

The samples were prepared as described previously in Martikainen et al., 2015. Peripheral blood mononuclear cells (PBMCs) were isolated from EDTA blood (10 ml, Vacutainer, BD, Plymouth, UK) using Ficoll-Paque PLUS (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) density gradient centrifugation. Cells were cryopreserved for later use by resuspending with 10% FBS Gold (PAA Laboratories GmbH, Pasching, Austria) in RPMI 1640 (Invitrogen, Grand Island, USA) and mixed 1:1 with 15% DMSO (Sigma Aldrich Company, St. Louis, USA) in FBS Gold. Cryotubes were stored in  $-80^{\circ}\text{C}$  overnight for gradual temperature reduction and transferred to a liquid nitrogen cryopreservation tank until thawing and staining. Cells were thawed with washing medium (1% L-glutamine + 1% Antibiotic/Antimycotic in RPMI 1640 (Invitrogen, Grand Island, USA)) and resuspended with 10% human AB serum (Innovative Research, Novi, USA) in washing medium. Cell number and viability of PBMCs were determined by trypan blue exclusion. The mean viability of thawed cells was 93.9%.

The following steps were performed on ice and protected from light. After resuspending, cells were washed with FACS buffer

(5% FBS Good, 0.02% NaN<sub>3</sub> in PBS) and centrifuged at 1300 rpm (300x g) for 5 min in +4°C. The supernatant was discarded, and the cells were suspended in human AB FACS buffer (10% human AB, 0.02% NaN<sub>3</sub> in PBS). Cell suspensions were incubated with fluorochrome-labeled antibody cocktails for 20 min at +4°C. The cells were then washed with FACS buffer and centrifuged at 1300 rpm (300x g) for 5 min in +4°C twice. Cell pellets from tubes without biotin-conjugated antibodies were resuspended in FACS buffer and moved to +4°C to wait for flow cytometry analysis. Cell pellets from tubes with biotin-conjugated antibodies were resuspended in Streptavidin-APC-eFluor780-FACS reagent buffer and incubated for 20 min at +4°C. After incubation cells were washed twice with FACS buffer and centrifuged at 1300 rpm (300x g) for 5 min in +4°C. After washing, cell pellets were resuspended in FACS buffer to wait for flow cytometry analysis. 7-AAD viability staining solution was added to the samples 5 minutes before the analysis. Samples were analyzed by flow cytometry immediately after staining.

Instrument

FACSCantoII cytometer (BD Biosciences, San Diego, CA, USA)

Software

FACSDiva software v. 6.1.3 (BD Biosciences, San Diego, CA, USA) was used for data collection and FlowJo X 10.0.7r2 software (Treestar, Ashland, OR, USA) for data analysis.

Cell population abundance

Cell sorting was not included in the protocol.

Gating strategy

Erythrocytes and debris were gated out according to the size (forward scatter) and cytoplasmic granularity (sideward scatter) (supplementary figure I). 7-AAD positive dead cells, CD19+ B cells, and CD14+ monocytes were excluded by high PE-Cy 5.5 fluorescence. The main peripheral blood DC subsets were identified as BDCA2+ pDCs, BDCA1+CD11c+ mDC1s, and BDCA3+high mDC2s. Gating adjustments were based on fluorescence minus one (FMO) tubes. After identification of DC subtypes, surface expression of functional markers ILT3 and ILT4 were analyzed in mDC1 and pDC subsets. Functional markers were not analyzed in mDC2s due to small population size. Doublets were excluded using forward scatter area (FSC-A) and forward scatter height (FSC-H) to confirm that doublets did not interfere with analysis.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.