Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size
   Describe how sample size was determined.
   No sample size calculation was performed. We used the samples available from the RISK study, an inceptional prospective cohort without a predicted target regarding sample size (all pediatric individuals with symptoms of Crohn’s disease were sampled at 28 clinics over a period of three years).

2. Data exclusions
   Describe any data exclusions.
   No data or sample was excluded. We used the 245 samples from the RISK study that had i) RNA-Seq available, ii) SNP data available and iii) had been profiled over 3 years (as described in Kugathasan et al., Lancet, 2017).

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   We performed a single replication attempt, as described in the main text ("Applying the approach to an independent sample of peripheral blood gene expression, the TRS also distinguished 61 pediatric Crohn’s disease cases and 12 controls (ΔSD=1.2; P=4×10^-5)."

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Samples were allocated into three different groups: i) non-IBD controls; ii) CD patients that remain in B1 status over a three year period and iii) CD patients that developed either B2 and/or B3 complications in the window from 90-days after diagnosis to 3-year after diagnosis. No covariates were controlled for. The details are available in the 1st/2nd paragraph of the Online Methods.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Blinding was not considered, given that group allocation had been previously done previous to this study by the investigators from the RISK study (Kugathasan et al. Lancet, 2017).

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

- n/a
- **Confirmed**

- The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

*See the web collection on statistics for biologists for further resources and guidance.*

7. Software
Policy information about availability of computer code

- **Software**

- Describe the software used to analyze the data in this study.

  For all our analyses we used the R software environment for statistical computing.

  For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

8. Materials and reagents
Policy information about availability of materials

- **Materials and reagents**

- Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

  All used samples are available for public use (SNP data is available upon contact with corresponding author).

9. Antibodies
Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- No antibodies were used

10. Eukaryotic cell lines
a. State the source of each eukaryotic cell line used.

  No eukaryotic cell lines were used

b. Describe the method of cell line authentication used.

  No eukaryotic cell lines were used
c. Report whether the cell lines were tested for mycoplasma contamination.

  No eukaryotic cell lines were used
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

  No eukaryotic cell lines were used

11. Description of research animals
Provide details on animals and/or animal-derived materials used in the study.

- No animals were used
12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All used samples belong to the cases of pediatric Crohn’s disease previously reported in Kugathasan et al. (Lancet, 2017).