Author’s response to reviews

Title: The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients; Data release 1

Authors:

Floris Imhann (florisimhann@gmail.com)
Joeri van der Velde (joeriv@gmail.com)
Ruggero Barbieri (rbarbieri1986@gmail.com)
Rudi Alberts (r.alberts@umcg.nl)
Michiel Voskuil (m.d.voskuil@umcg.nl)
Arnau Vich Vila (arnauvich@gmail.com)
Valerie Collij (v.collij@umcg.nl)
Lieke Spekhorst (l.m.spekhorst@umcg.nl)
Kimberly van der Sloot (k.w.j.van.der.sloot@umcg.nl)
Vera Peters (v.peters@umcg.nl)
Hendrik van Dullemen (h.m.dullemen@umcg.nl)
Marijn Visschedijk (m.c.visschedijk@umcg.nl)
Eleonora Festen (e.a.m.festen@umcg.nl)
Morris Swertz (m.a.swertz@gmail.com)
Gerard Dijkstra (gerard.dijkstra@umcg.nl)
Rinse Weersma (r.k.weersma@umcg.nl)

Version: 1 Date: 08 Nov 2018

Author’s response to reviews:

Point-to-point response to the reviewers

BMGE-D-18-00322
The 1000IBD project: multi-omics data of 1000 inflammatory bowel disease patients; Data release 1

Reviewer reports: Shin Takasawa (Reviewer 1)

I found that this is an excellent "protocol paper" and I have no serious criticism regarding its purpose and methodology. I hope new findings that surpass previous results concerning IBD (Ex. Roche-Lima, A. et al. Front. Genet. 9, 116, 2018; Friedman, M.F. et al. Clin. Epidemiol. 10, 671-681, 2018) are soon followed using 1000IBD project data.

We thank the reviewer for his careful assessment of our manuscript and for the kind remarks!

Nicholas Alexander Kennedy (Reviewer 2)

This paper provides a basic description of the phenotype and genotype data generated as part of the 1000IBD project. It does not, in itself, contain any new findings, but the cohort described will no doubt be used by the submitting authors and others to generate a wide range of interesting analyses.

We thank the reviewer for his careful assessment of our manuscript.

I have a few specific questions:

1. It's not clear to me why only a subset of the data generated is being released at this stage (e.g. 314 of the 1215 patients with ImmunoChip data). This could be better explained in the manuscript.

We agree and have added a clearer explanation. (page 15, lines 225-229)

The molecular data of the 1000IBD cohort is comprised of both newly generated and previously generated molecular data from samples from the same IBD patients that has now been added to the 1000IBD project.

For example, the ImmunoChip is no longer available. Hence, not all 1000IBD participants are genotyped using the Illumina ImmunoChip. All available ImmunoChip data is made available.
Genotyping efforts continue using the newer Illumina Global Screening Array. In addition, while going through our data archive, we were able to identify more ImmunoChip data and expanded the total number of 1000IBD participants for which ImmunoChip genotypes are available to 427 IBD.

2. The website for EGA is https://ega-archive.org (i.e. without the www)

We have corrected this error (page 4, line 37).

3. The website https://1000ibd.org is not live. The Molgenis github page (https://github.com/molgenis/molgenis-projects) suggests the address should be https://1000ibd.com, but again this is not live.

We apologize for the inconvenience. Probably due server maintenance or software updates, the server was offline.

Both the URLs https://1000ibd.org, https://1000ibd.nl and https://1000ibd.com are registered and refer to the same 1000IBD website, which is now live. Throughout the manuscript the we now refer to https://1000ibd.org. We have also made sure that the URL works both with and without ‘www.’

(page 24, lines 422).

4. I would have liked to see some top-level summary statistics of the cohort presented (e.g. diagnoses, age at diagnosis, sex, Montreal classification, disease duration at recruitment, medication exposure, disease activity, etc.)

We agree and have expanded table 1 including the below-mentioned items (pages 12-13).

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of participants</td>
<td>1215</td>
</tr>
<tr>
<td>Age (Median ± IQR)</td>
<td>41 ± 25</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male (%)</td>
<td>510 (41.97)</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Female (%)</td>
<td>705 (58.03)</td>
</tr>
<tr>
<td>Crohn’s Disease (%)</td>
<td>615 (50.62)</td>
</tr>
<tr>
<td>Ulcerative Colitis (%)</td>
<td>495 (40.74)</td>
</tr>
<tr>
<td>IBDU (%)</td>
<td>61 (5.02)</td>
</tr>
<tr>
<td>Other (microscopic colitis, IBDI, reconsidering IBD diagnosis) (%)</td>
<td>44 (3.62)</td>
</tr>
</tbody>
</table>

**Montreal Classification**

<table>
<thead>
<tr>
<th>A: Age of Onset</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 (%)</td>
<td>159 (13.09)</td>
</tr>
<tr>
<td>A2 (%)</td>
<td>710 (58.44)</td>
</tr>
<tr>
<td>A3 (%)</td>
<td>253 (20.82)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>L: Disease Location (CD only)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 (%)</td>
<td>224 (36.42)</td>
</tr>
<tr>
<td>L2 (%)</td>
<td>120 (19.51)</td>
</tr>
<tr>
<td>L3 (%)</td>
<td>255 (41.46)</td>
</tr>
<tr>
<td>L4 (%)</td>
<td>65 (10.57)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B: Disease Behaviour (CD only)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>B1 (%)</td>
<td>301 (48.94)</td>
</tr>
<tr>
<td>B2 (%)</td>
<td>208 (33.82)</td>
</tr>
<tr>
<td>B3 (%)</td>
<td>102 (16.58)</td>
</tr>
</tbody>
</table>

| Perianal                             | 189 (30.73) |

<table>
<thead>
<tr>
<th>E: Disease Extent (UC only)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E1 (%)</td>
<td>57 (10.25)</td>
</tr>
<tr>
<td>E2 (%)</td>
<td>162 (29.13)</td>
</tr>
<tr>
<td>E3 (%)</td>
<td>299 (53.78)</td>
</tr>
</tbody>
</table>
S: Disease Severity (UC only)
S1 (%) 29 (5.22)
S2 (%) 139 (25.00)
S3 (%) 191 (34.35)
S4 (%) 119 (21.40)
Age at Diagnosis in years (Median ± IQR) 27 ± 19
Disease Duration at Recruitment in years (Median ± IQR) 8 ± 12
Medication Exposure
Steroids % 90.07
Steroids CD % 91.99
Steroids UC % 88.28
Steroids IBDU % 88.33
Immunosuppressors % 68.32
Immunosuppressors CD % 79.08
Immunosuppressors UC % 56.97
Immunosuppressors IBDU % 56.67
Biologica] s % 37.30
Biologica] s CD % 55.07
Biologica] s UC % 17.37
Biologica] s IBDU % 25.00
Mesalazines % 44.34
Mesalazines CD % 18.06
Mesalazines UC % 70.99
Mesalazines IBDU % 83.33
Average Disease Activity*
HBI (Average ± Standard Deviation) 2.99 ± 3.18
SSCAI (Average ± Standard Deviation) 1.61 ± 1.97

* For each patient, the median disease activity was determined. For the entire group the average of the individual medians is presented here

IQR: interquartile range, CD: Crohn’s disease, UC: ulcerative colitis, IBDU: inflammatory bowel disease undetermined, IBDI: inflammatory bowel disease intermediate, HBI: Harvey-Bradshaw Index, SSCAI: Simple Clinical Colitis Activity Index

5. As new, unpublished instruments, details on the FFQ and environmental questionnaire is rather limited.

We agree and have added the following additional information.

Regarding the FFQ:

The Groningen IBD-specific Food Frequency Questionnaire (GrIB FFQ) was designed to assess the current dietary habits and nutritional intake of IBD patients. It consists of 119 questions on food items that are grouped into categories: breakfast, lunch, dinner, snacks and drinks. Since IBD patients often follow unguided dietary habits, i.e., those made without consulting a physician or dietician first, population-specific and more extensive items (e.g., dairy substitutes, meat replacers and supplements) are included in the questionnaire. When using this nutritional tool, patients report the intake of foods consumed during the previous month. The food data obtained via the GrIB FFQ will be converted into energy and nutrient intake (in grams/day) using the NEVO food composition database of 2016 (NEVO 2016, RIVM, Bilthoven, the Netherlands). The nutritional intake part of the GrIB FFQ was developed in collaboration with, and validated by, the division of Human Nutrition of Wageningen University using standardized procedures. (1,2)

The GrIB FFQ provides a broader overview than traditional food questionnaires. It also assesses factors that influence nutrition expenditure but are often disregarded. To complement the questions on nutritional intake, items on patient’s conceptions about the role of nutrition in IBD have been added. Since these additional items could not be included in the standard validation procedure of the Wageningen University, the entire GrIB FFQ will be validated with data collected in an upcoming randomized controlled trial.


   (page 14, lines 196-214)

Regarding the environmental questionnaire:

The environmental questionnaire has recently been validated, and the paper on the validation has just been published:


   (page 15, lines 222-223)

6. When are these patients recruited? At diagnosis? Opportunistically? How are they selected? What are the inclusion/exclusion criteria?

We have added additional information regarding patient recruitment, selection, inclusion and exclusion criteria to the manuscript (page 9, lines 131-147).

All IBD patients treated in the specialized IBD Center of the Department of Gastroenterology and Hepatology of the University Medical Center Groningen (UMCG) are asked to participate in the 1000IBD project. Patients recruited for the 1000IBD project include new patients with new-onset IBD, new patients with existing IBD newly referred to our hospital, and patients with IBD already being treated in our hospital. Since there are more than 2000 IBD patients already being treated in our hospital, this last group is gradually being asked to participate to allow our research nurses to obtain the data and samples in an organized way. Since our university hospital is a tertiary referral centre, the number of new-onset IBD patients is limited.
The only inclusion criteria for the 1000IBD project are that patients need to: (1) be at least 18 years old, (2) have IBD based on accepted radiological, laboratory and endoscopic findings, (3) have provided informed consent and (4) able to speak, read and write Dutch in order to be able to fill in the questionnaires. Paediatric patients are not included in 1000IBD because our institutional review board (IRB) approval does not cover them. There are no additional inclusion or exclusion criteria. Inclusion is still on-going, and the project had enrolled 1,215 IBD patients as of September 1, 2017.

7. Are patients recallable for resampling? If so, is this available to external researchers, or just to the UMCG group?

Yes, patients are recallable for sampling, and all patients in the 1000IBD cohort are followed up. The possibility to recall patients for additional material is only available to UMCG researchers. However, collaborations in which extra material is collected are possible.

(page 25, lines 432-435)

James John Ashton (Reviewer 3)

Background- The authors present a succinct overview of the importance of developing multi-omic data in IBD in order to move to a systems biology and personalised medicine approach to disease. It would be interesting to discuss the recent publications from the RISK cohort where use of these data has enabled modelling to take a first step towards precision medicine in IBD. I feel this would place the 1000IBD project in context and emphasise it's importance.

We thank the reviewer for his careful assessment of our manuscript.

We agree that adding a short discussion of how research from previous efforts from the RISK cohort, the PRISM cohort and our own cohort has enabled precision medicine in IBD is a good addition to the paper.

We have now added the following to the introduction:

Previous efforts from the RISK cohort, the PRISM cohort, and samples from patients treated in our university hospital, as well as from consortia in which these cohorts participate, have already enabled the first steps towards precision medicine in IBD. For example, the discovery of genetic
variants enables the prediction of the risk of pancreatitis as a severe side-effect of azathioprine, a commonly used immunosuppressant in IBD. (1) Microbial DNA profiles and RNA-sequencing profiles from the intestinal biopsies of the RISK cohort have uncovered RNA-microbe interactions and shown that biopsies taken from the distal colon can predict the IBD disease location higher up in the intestine. (2,3) In addition, stool samples from the PRISM cohort have also been used to discover microbial profiles that can predict the efficacy of vedolizumab, a biological drug regulating T-cell homing to the gut (4), while a genetic variant in the WWOX gene discovered using genotypes of IBD patients treated in our hospital can be used to assess the risk of stricturing and penetrating Crohn’s disease behavior. (5)


(pages 7-8, lines 98-110)

What will be the arrangements for publication of any results developed from these data by institutions external to Groningen? Will co-authorship be required or just citation?
The arrangements for publication regarding co-authorship or just citation will depend on the extent of cooperation required by the Groningen team. If IBD participants need to be recalled for additional sampling or if large data processing efforts by the Groningen team are required, co-authorship will be required. However, if a limited amount of data is requested from the EGA, citation of the current manuscript would suffice.

Methods- This is clearly an extremely ambitious and important project, with very important data likely to be generated by this work (and already developed). Whilst I do not include specific comments about the outputs yet several questions do arise which would be important to answer prior to using these data.

We thank the reviewer for his kind remarks concerning the 1000IBD project. We have tried to clarify the methods in the revised manuscript.

Additionally, the use of patient-based reporting systems has previous had flaws (reporting bias, subjective etc.) but also represents a very interesting way of collecting data.

The patient-based reporting systems have their flaws. However, we think they are an addition to the nurse- and doctor-based reporting systems in place.

What were the ages of the patients included?
We have expanded the descriptive statistics table that includes data on age and age at diagnosis (see question from reviewer 2 and pages 12-13.

Were paediatric samples included?
Paediatric patients are not included because our IRB approval does not cover paediatric IBD patients.

(page 9, lines 144-145)
Were the two new questionnaires developed for this study validated, reviewed externally or piloted prior to use? The importance of the quality of the clinical phenotypic data here cannot be understated.

We agree that quality of the clinical phenotypic data is pivotal. Therefore, we are currently validating the FFQ and have validated the environmental questionnaire. We have provided an elaborate answer to the same question from reviewer 2. A better description has been added to the manuscript.

Will all raw fastq files from the WES, microbiome and transcriptomic sequencing be available to run through custom pipelines?

Yes, all raw sequencing data will be made available. A brief remark is added on page 25, lines 428-430.

RNAseq and microbiome (MGS) data is notorious for being subject to batch effects, could the authors comment on how the 300 samples were grouped and how batch effect was minimized?

We agree that batch effects are an important aspect of transcriptome and microbiome studies. Therefore, we have taken the following precautions:

Regarding the RNA-sequencing:

The RNA samples were pseudo-randomized on plates to assure that no single factor was dominant on one plate (IBD diagnosis, disease location or disease activity). RNA-sequencing was conducted in two batches comprising one pilot batch of one plate of 20 samples, and one batch of one plate of 80 samples and two plates containing 100 samples each. When a principal coordinate analysis was executed as part of the QC, no relevant batch effect was detected.

pages 18-19, lines 306-312

Regarding the sequencing of microbial DNA:
The microbial DNA samples were randomized on 96-well plates so that age and IBD diagnosis were mixed. The plates containing microbial DNA were sent to the Broad Institute, Boston, USA for 16S sequencing in one batch. These microbial DNA samples were stored at the Broad Institute at -80°C, and whole genome metagenomic sequencing was performed at a later stage. A second batch of microbial DNA was sent to the Broad Institute for whole genome shotgun metagenomic sequencing using the same procedure. When a principal coordinate analysis (PCA) was executed as part of the QC, no relevant batch effects were detected.

Previous studies have used RNAlater to preserve bacterial DNA and human RNA to prevent degradation, did the authors consider this?

We did not use RNAlater for the current faecal samples or the intestinal biopsies. However, the use of RNAlater is implemented in our new faecal sample collection protocols. The biopsies were immediately snap-frozen using liquid nitrogen so they can be used for DNA-isolation, RNA-isolation, but also for Western Blot and immune-histochemistry. We were able to successfully isolate both DNA and RNA from the snap-frozen samples. Results using the bacterial DNA from biopsies have already been published. (1)


A brief remark is mentioned on page 18, lines 295-299.

How quickly were biopsies snap-frozen after sampling?

The biopsies were immediately snap-frozen by the endoscopy nurse or research technician present during the endoscopy procedure using liquid nitrogen which was readily available in the endoscopy room.

This is discussed briefly on page 18, lines 295-299 of the manuscript.

Similarly, how many of the stool samples were frozen within 15 minutes? The exact conditions for samples is important in determining the quality of the output data.

We agree that the exact sampling conditions are very important and our protocols ensure good quality samples.
We surveyed 248 IBD patients who took part in the stool sampling. Of these 248 patients, only three patients required more than 15 minutes to store their faecal sample in their freezer, while 13 patients did not fill in this particular question. A large majority reported being able to freeze their faecal sample within 10 minutes after production.

A separate paper on the patient experiences and attitudes towards faecal sampling has been submitted and will be published soon.

A brief remark is mentioned on page 19, lines 318-321.

At what point in the disease course were patients recruited? Are any treatment naive? Are any followed longitudinally with multiple samples available?

Patients are recruited during all stages of their disease course, including at the beginning when they are treatment-naïve. However, since our hospital is a tertiary referral centre, the number of new-onset IBD patients is limited. A more extensive description of the selection and recruitment of IBD patients is provided in the new manuscript on page 9, lines 131-147.

All 1000IBD participants are followed up longitudinally. Every time they visit the outpatient department, the gastroenterologist updates the electronic health record, which automatically updates the research database with the same information (page 11, lines 166-172).

In addition to the current 1000IBD biomaterial collections, subsets of the 1000IBD cohort will be sampled longitudinally. In the IBD Tracker project, a selection 1000IBD participants will undergo weekly stool sampling. A separate IRB approval has been obtained for this project. (page 20, lines 347-351).

Overall - This is a good report and the publication would benefit researchers to understand the access to these data, clearly further work is ongoing.

We thank the reviewer for his careful assessment of our manuscript and his kind remarks.
There are several typos that should be corrected including:

Page 10, line 43- GENERATION OF DIETART AND ENVIRONMENTAL DATA

We have reread the article and corrected the typos.

Editorial Policies

Please read the following information and revise your manuscript as necessary. If your manuscript does not adhere to our editorial requirements this will cause a delay whilst the issue is addressed. Failure to adhere to our policies may result in rejection of your manuscript.

In accordance with BioMed Central editorial policies and formatting guidelines, all submissions to BMC Gastroenterology must have a Declarations section which includes the mandatory subsections listed below. Please refer to the journal's Submission Guidelines web page for information regarding the criteria for each subsection (https://bmcgastroenterol.biomedcentral.com/).

Where a mandatory section is not relevant to your study design or article type, for example, if your manuscript does not contain any individual persons data, please write "Not applicable" in these sections.

For the 'Availability of data and materials' section, please provide information about where the data supporting your findings can be found. We encourage authors to deposit their datasets in publicly available repositories (where available and appropriate), or to be presented within the manuscript and/or additional supporting files. Please note that identifying/confidential patient data should not be shared. Authors who do not wish to share their data must state that data will not be shared, and provide reasons for this in the manuscript text. For further guidance on how to format this section, please refer to BioMed Central's editorial policies page - http://www.biomedcentral.com/submissions/editorial-policies#availability+of+data+and+materials.
Declarations

- Ethics approval and consent to participate
- Consent to publish
- Availability of data and materials
- Competing interests
- Funding
- Authors' Contributions
- Acknowledgements
- Authors' Information

Our manuscript fulfils all the above-mentioned criteria.