Inventory of Supporting Information

Manuscript #: NG-A57877

Corresponding author name(s): Hailiang Huang, Carl A. Anderson, Mark J. Daly

nature portfolio

Please complete each of the Inventory Tables below to outline your Extended Data and Supplementary Information items.

There are four sections:

- *Extended Data*
- *Supplementary Information: Flat Files*
- *Supplementary Information: Additional Files*
- *Source Data*

Each section includes specific instructions. Please complete these tables as fully as possible. We ask that you avoid using spaces in your file names, and instead use underscores, i.e.: Smith_ED_Fig1.jpg not Smith ED Fig1.jpg

Please note that titles and descriptive captions will only be lightly edited, so please ensure that you are satisfied with these prior to submission.

If you have any questions about any of the information contained in this inventory, please contact the journal.

1. Extended Data

Complete the Inventory below for all Extended Data figures.

- Keep Figure Titles to one sentence only

- Upload your files as 'Figure Files' in our Manuscript Tracking system
- File names should include the Figure Number. i.e.: *Smith_ED_Fig1.jpg*
- Please be sure to include the file extension in the Filename. Note that Extended Data files must be submitted as .jpg, .tif or .eps files *only*, and should be approximately 10MB
- All Extended Data figure legends must be provided in the Inventory below and should not exceed 300 words each *(if possible)*
- Please include Extended Data *ONLY* in this table

Figure #	**Figure title** One sentence only	**Filename** This should be the name the file is saved as when it is uploaded to our system. Please include the file extension. i.e.: *Smith_ED_Fig1.jpg*	**Figure Legend** If you are citing a reference for the first time in these legends, please include all new references in the main text Methods References section, and carry on the numbering from the main References section of the paper. If your paper does not have a Methods section, include all new references at the end of the main Reference list.
Extended Data Fig. 1	Overview of the study design	EDF1.jpg	We utilized a logistic mixed-model for the association analysis, followed by the fixed-effect meta-analysis to combine multiple cohorts. Multiple cohorts serve the purpose of replication. Two large cohorts at Broad Institute of different exome capture platforms were used to discover candidate variants (Nextera WES and Twist WES). Two independent cohorts at Sanger (Sanger WGS and Sanger WES) and one Kiel/Regeneron cohort (Regeneron WES) were used to replicate the findings.
Extended Data Fig. 2	Quality control procedures applied in the Broad sequencing pipeline	EDF2.jpg	We show as an example the quality control steps performed on variants and subjects from the Broad sequencing platform. Quality controls performed on data from other platforms follow a similar plan and are described in Methods. Quality control steps using external information from gnomAD were colored green. Thresholds and details can be found in Methods.
Extended Data Fig. 3	QQ plots of the heterogeneity-of-effect test between the	EDF3.jpg	Only QC passed variants with minor allele frequency in NFE between 0.0001 and 0.10 were included. **a,** all variants. **b,** non-synonymous variants. **c,** synonymous variants. In **a** and **b**,

	Nextera and Twist discovery cohorts		the y axis is capped at -$\log_{10} p$ = 30 while the top four variants (three in *NOD2* and one in *IL23R*) have -$\log_{10} p$ > 100. In **c,** to remove the synonymous variants that tag causal non-synonymous variants and artifacts through LD, we removed loci hosting large-effect coding variants (*IL23R*, *NOD2*, *LRRK2*, *TYK2*, *ATG16L1*, *SLC39A8*, *PTGER4*, *IRGM*, *CARD9*), implicated by variants removed in the heterogeneous test (*AHNAK2, LILRA*), and with long range LD (MHC).
Extended Data Fig. 4	Power to detect single variant associations.	EDF4.jpg	We performed a series of power calculations using the methodology described by Johnson and Abecasis (2017). Our initial 'exome-wide scan' (two cohorts) had fewer samples and a more lenient significance threshold than subsequent meta-analysis (five cohorts). However, both analyses had similar power to detect true associations at their respective significance levels. Our single-variant association analyses did not have the power to uncover association to variants with a MAF = 0.0001 and below (unless the variant has a very strong effect, e.g. 0.76 power at OR = 8). Similarly, the exome-wide scan had limited power to detect association to variants with a MAF = 0.001 and OR < 2, but was well-powered above these thresholds. **a,** Power of the exome-wide scan analysis **b,** Power of the meta-analysis. **c,** Power to detect single-variant associations at different minor allele frequencies at α = 0.0002 ('scan'; dashed lines) and 3 x 10-7 ('meta'; solid lines) and assuming Crohn's disease population prevalence of 276 in 100,000, and an additive effect model.
Extended Data Fig. 5	Relation to known IBD associations	EDF5.jpg	Numbers in brackets are the number of variants assigned to the categories out of the 45 exome-wide significant variants.
Extended Data Fig. 6	WES variants from this study implicating known IBD loci	EDF6.jpg	**a-c:** a novel CD variant implicating *TAGAP*. **d-g:** CD variants tagging fine-mapped IBD associations in *LRRK2*. **a** and **d,** P-value for variants from the fine-mapping study[5]. **b** and **e**, PIP from fine-mapping. **c**, **f** and **g,** P-value for variants from this study. Open circle indicating LD information is missing. LD calculated between the plotted variant and the best variant in

			b for panel **c**, and variants with best PIP in credible sets 1 and 2 (panel **e**) respectively for panels **f** and **g.**
Extended Data Fig. 7	Nextera and Twist callset population assignment.	EDF7.jpg	Principal components for **a, c,** before removing non-European samples for Twist and Nextera respectively. **b, d,** after removing non-European samples for Twist and Nextera respectively. Principal components generated from the 1000 Genome Project Phase III data and different colors stand for different continental / superpopulations. Study subjects (black dots) were projected onto principal components.

Delete rows as needed to accommodate the number of figures (10 is the maximum allowed).

2. Supplementary Information:

A. Flat Files

Complete the Inventory below for all additional textual information and any additional Supplementary Figures, which should be supplied in one combined PDF file.

- **Row 1:** A combined, flat PDF containing any Supplementary Text, Discussion, Notes, Additional Supplementary Figures, Supplementary Protocols, simple tables, and all associated legends. Only one such file is permitted.

- **Row 2:** Nature Research's Reporting Summary; if previously requested by the editor, please provide an updated Summary, fully completed, without any mark-ups or comments. **(Reporting Summaries are not required for all manuscripts.)**

Item	**Present?**	**Filename** **This should be the name the file is saved as when it**	**A brief, numerical description of file contents.** **i.e.:** ***Supplementary Figures 1-4, Supplementary Discussion, and Supplementary Tables 1-4.***

		is uploaded to our system, and should include the file extension. The extension must be .pdf	
Supplementary Information	Yes	Supplementary Information.pdf	Details of individual participating IBD cohorts Supplemental acknowledgments of participating consortia and programs
Reporting Summary	Yes	nr-reporting-summary.pdf	
Peer Review Information	No	*OFFICE USE ONLY*	

B. Additional Supplementary Files

Complete the Inventory below for all additional Supplementary Files that cannot be submitted as part of the Combined PDF.

- Do not list Supplementary Figures in this table (see section 2A)
- Where possible, include the title and description within the file itself
- Spreadsheet-based tables & data should be combined into a workbook with multiple tabs, not submitted as individual files.
- Compressed files are acceptable where necessary. ZIP files are preferred.
- Please note that the *ONLY* allowable types of additional Supplementary Files are:
    - Supplementary Tables ○ Supplementary Audio ○ Supplementary Videos ○ Supplementary Software
    - Supplementary Data, for example: raw NMR Data, Cryo-EM Data, Computational Data, Crystallographic Data, etc.

Type	**Number** **If there are multiple files of the same type this should be the numerical**	**Filename** **This should be the name the file is saved as when it is uploaded to our**	**Legend or Descriptive Caption** **Describe the contents of the file**

	indicator. i.e. "1" for Video 1, "2" for Video 2, etc.	system, and should include the file extension. i.e.: *Smith_ Supplementary_Video_1.mov*	
Supplementary Tables	1	Supplementary Tables.xlsx	Supplementary Tables 1-8
Supplementary Data 1	1	Supplementary Data 1.pdf	Principal components for subjects in the Nextera and Twist cohorts. Cases and controls are plotted as on the first two principal components for exome-wide significant CD variants. Carriers of the minor alleles are highlighted for cases and controls respectively.

Add rows as needed to accommodate the number of files.

3. Source Data

Complete the Inventory below for all Source Data files.

- Acceptable types of Source Data for Main Figures and Extended Data Figures are:
    - Statistical Source Data
        - Plain Text (ASCII, TXT) or Excel formats only
        - One file for each relevant Figure, containing all source data
    - Full-length, unprocessed Gels or Blots
        - JPG, TIF, or PDF formats only
        - One file for each relevant Figure, containing all supporting blots and/or gels
- 'Source Data' is only allowed for Main Figures and Extended Data Figures.
    - Include Unprocessed Gels or Blots for Supplementary Figures as additional Supplementary Figures.
    - Include Statistical Source Data for Supplementary Figures as 'Supplementary Data' files and list them in section 2B.
    - Please see this example of Source Data in a publication.

Parent Figure or Table	**Filename** This should be the name the file is saved as when it is uploaded to our system, and should include the file extension. i.e.: *Smith_SourceData_Fig1.xls,* or *Smith_Unmodified_Gels_Fig1.pdf*	**Data description** i.e.: Unprocessed Western Blots and/or gels, Statistical Source Data, etc.
Source Data Fig. 1		
Source Data Fig. 2		
Source Data Fig. 3		
Source Data Fig. 4		
Source Data Fig. 5		
Source Data Fig. 6		
Source Data Fig. 7		
Source Data Fig. 8		
Source Data Extended Data Fig. 1		
Source Data Extended Data Fig. 2		

Source Data Extended Data Fig. 3		
Source Data Extended Data Fig. 4		
Source Data Extended Data Fig. 5		
Source Data Extended Data Fig. 6		
Source Data Extended Data Fig. 7		
Source Data Extended Data Fig. 8		
Source Data Extended Data Fig. 9		
Source Data Extended Data Fig. 10		