Supporting Information: Cytometric fingerprints of gut microbiota predict Crohn's disease state

Peter Rubbens[a,b,1], **Ruben Props**[c], **Frederiek-Maarten Kerckhof**[c], **Nico Boon**[c], **and Willem Waegeman**[a,1]

[a]KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Coupure Links 653, B-9000, Gent, Belgium; [b]Current address: Flanders Marine Institute (VLIZ), Wandelaarkaai 7, B-8400, Ostend, Belgium; [c]Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, B-9000, Gent, Belgium

Variations in the gut microbiome have been associated with changes in health state such as Crohn's disease. Most surveys characterize the microbiome through analysis of the 16S rRNA gene. An alternative technology that can be used is flow cytometry. In this report we analyzed a disease cohort that has been characterized by both technologies. Changes in microbial community structure are reflected in both types of data. We demonstrate that cytometric fingerprints can be used as a diagnostic tool in order to classify samples according to Crohn's disease state. These results highlight the potential of flow cytometry to perform rapid diagnostics of microbiome-associated diseases.

16S sequencing | Crohn's disease | Diversity | Flow cytometry | Gut microbiota |

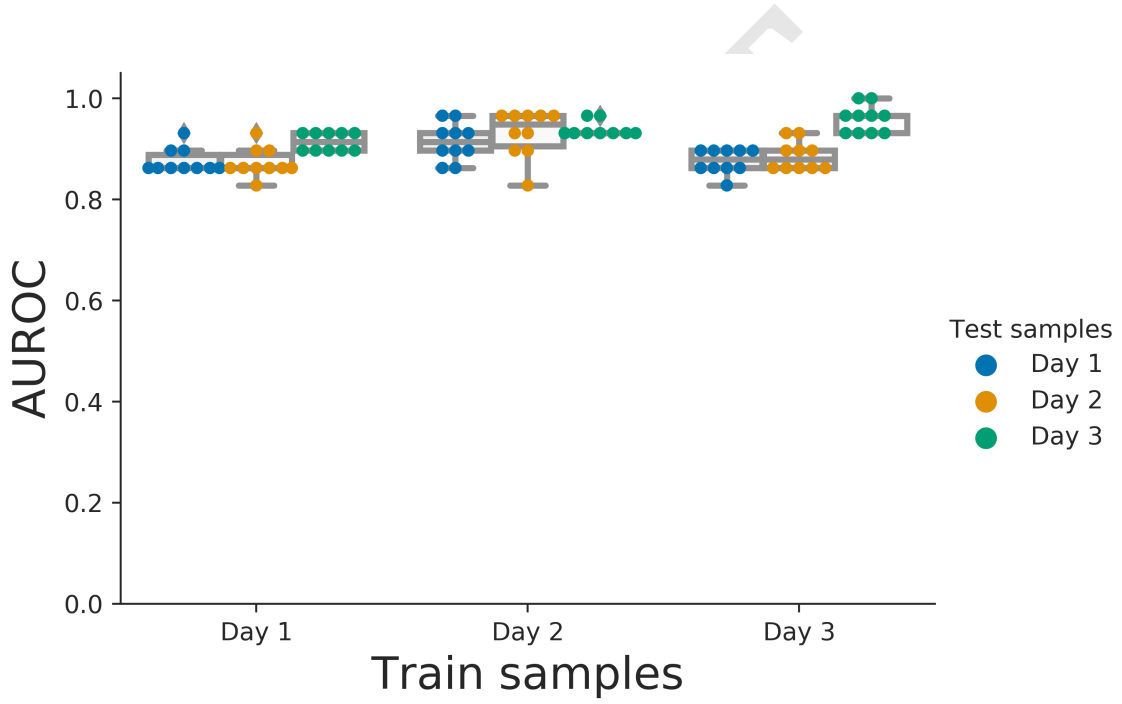

Fig. 1. Summary of flow cytometry predictions when samples are split according to the day at which they were measured (day one, two or three). Each batch of samples was part of the training set, to predict the disease state of samples in three different test batches, again split up according to the day at which they were measured. The performance was evaluated using the area under the ROC (AUROC) curves, based on pairwise averaging of the test set. Each dot represents the AUROC for an individual run (ten in total), along with a visualization of the median. A boxplot displays the first and third quartile and the median line. Whiskers extend from the quartiles to 1.5 times the interquartile range.

The authors have no conflicts of interest to declare.

[1]To whom correspondence should be addressed. E-mail: Peter.Rubbens@vliz.be.

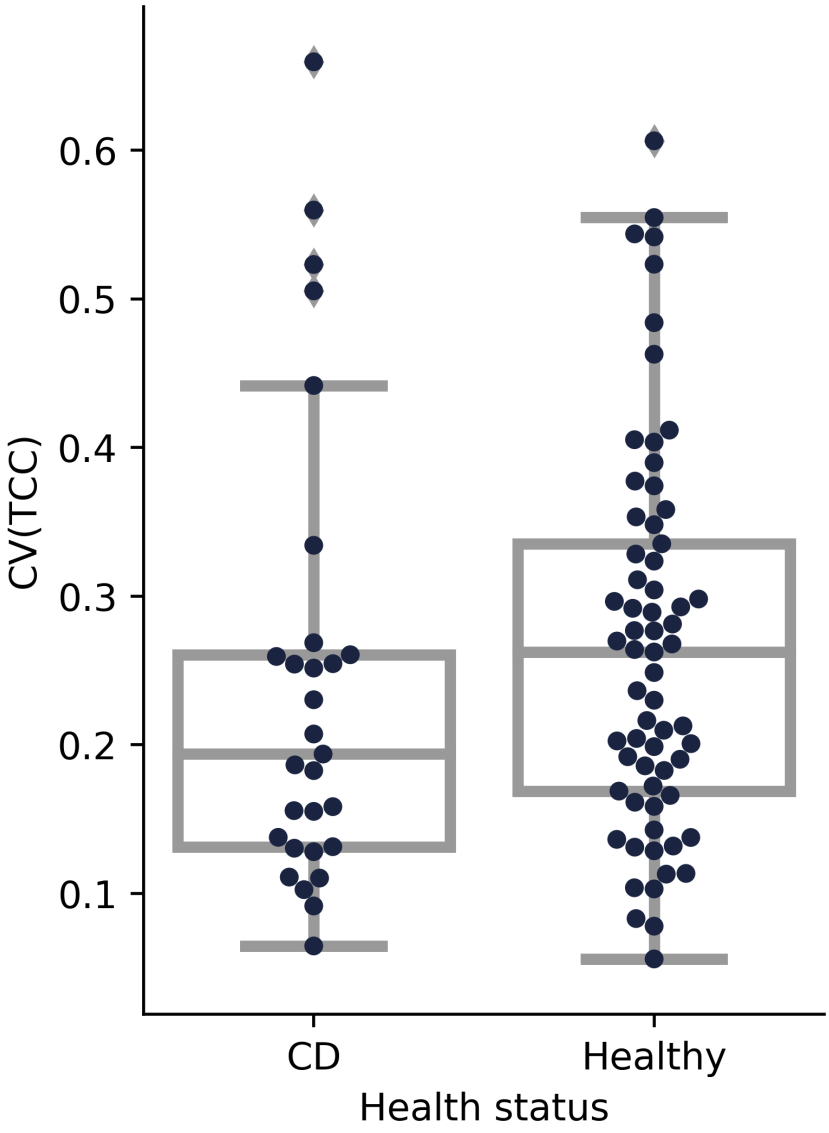

Fig. 2. Distribution of the coefficient of variation (CV) values of the total cell counts (TCC) per sample, according to patient status (CD or HC). Each boxplot displays the first and third quartile and the median line. Whiskers extend from the quartiles to 1.5 times the interquartile range. Points that lie outside this range are visualized as outliers.

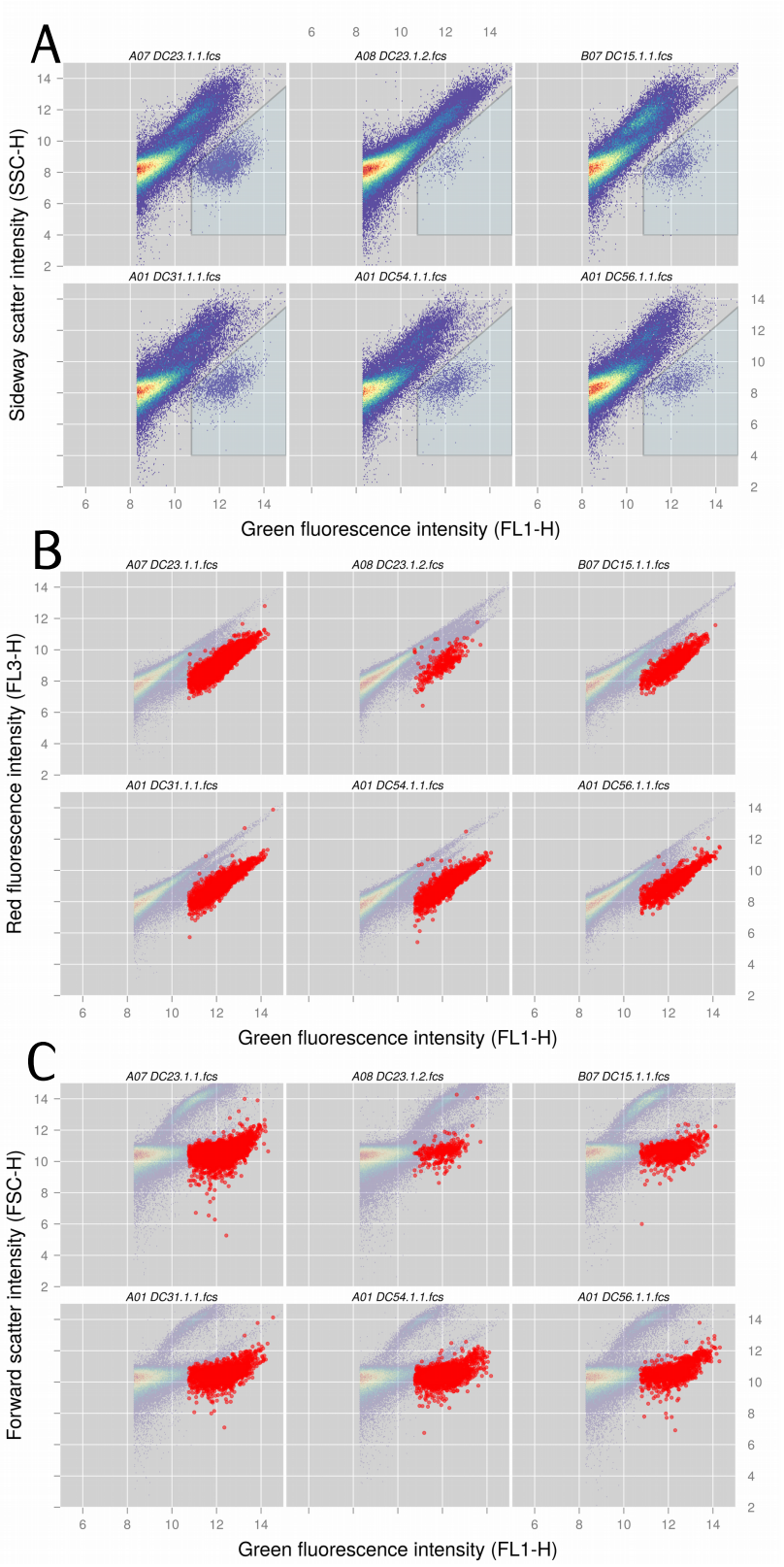

Fig. 3. Visualization of the gating strategy to exclude bacterial cells from noise using multiple scatter plots. The visualizations indicate that the cells selected within the gate are separated from noise in the multiple channels (visualized in red). **A**: Gating template drawn in the $f(x) = \text{asinh}(x)$ transformed FL1-H – SSC-H space. **B**: Visualization of the data in the $f(x) = \text{asinh}(x)$ transformed FL1-H – FL3-H space. Cells selected within the gate in **A** are visualized in red. **C**: Visualization of the data in the $f(x) = \text{asinh}(x)$ transformed FL1-H – FSC-H space. Cells selected within the gate in **A** are visualized in red.

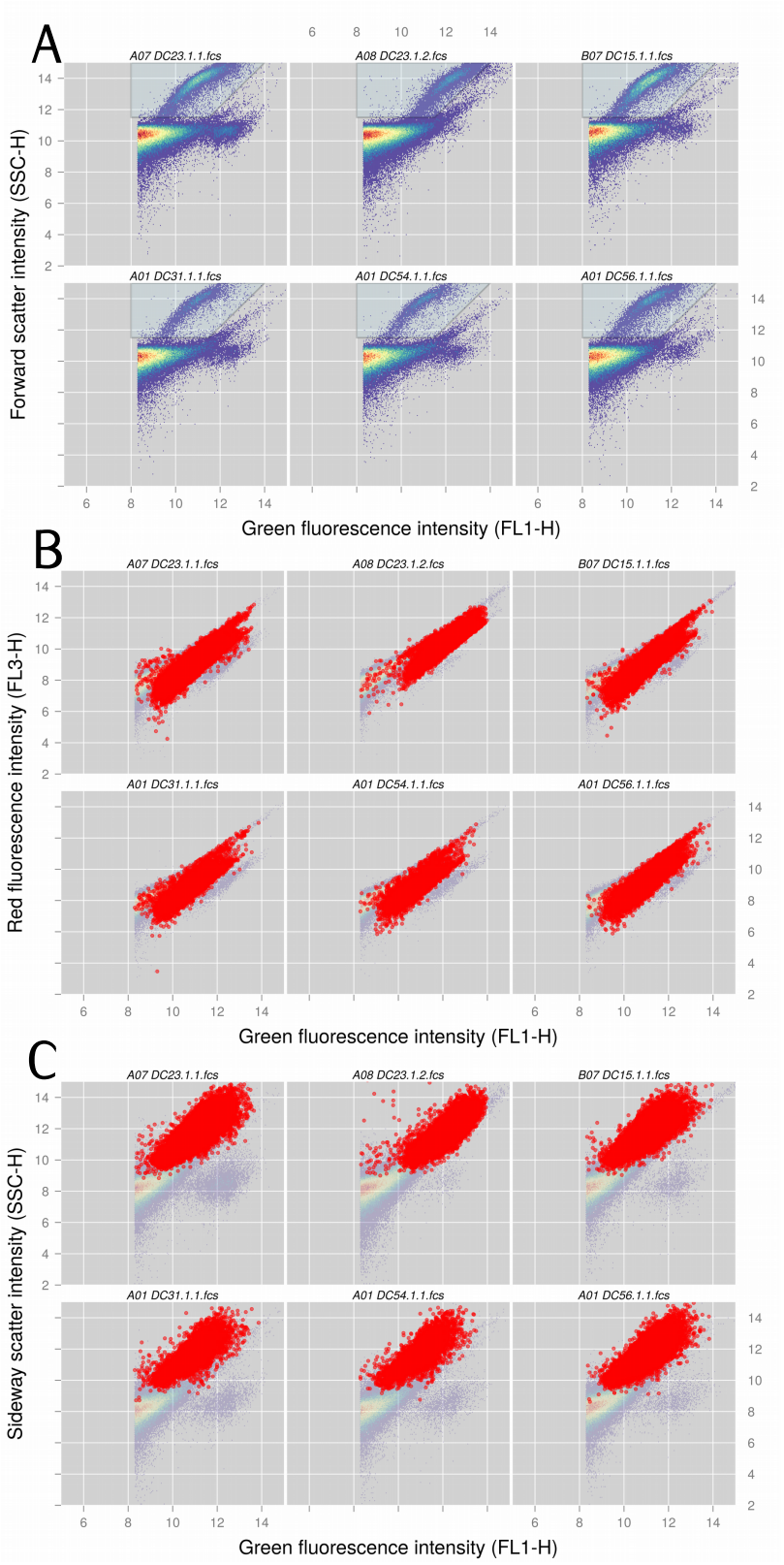

Fig. 4. Visualization of additional background that is excluded from the data using multiple scatter plots. The visualizations indicate that the cells selected within the gate are part of the background in the FL1-H – FL3-H scatter plot (visualized in red), and therefore need to be removed from the data. **A**: Gating template drawn in the $f(x) = \text{asinh}(x)$ transformed FL1-H – SSC-H space. **B**: Visualization of the data in the $f(x) = \text{asinh}(x)$ transformed FL1-H – FL3-H space. Cells selected within the gate in **A** are visualized in red. **C**: Visualization of the data in the $f(x) = \text{asinh}(x)$ transformed FL1-H – FSC-H space. Cells selected within the gate in **A** are visualized in red.

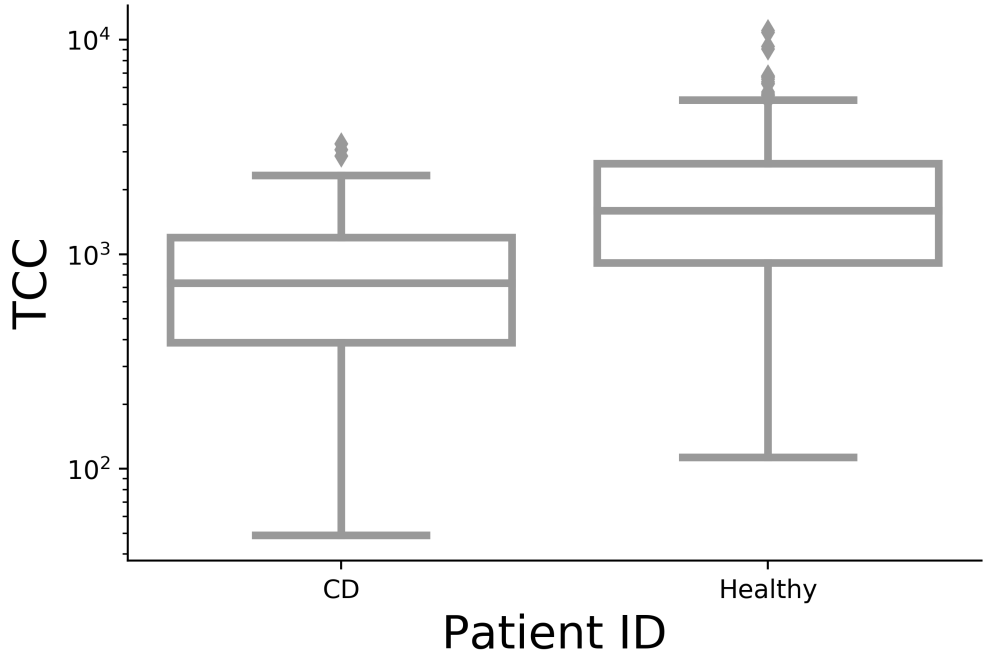

Fig. 5. Distribution of total cell counts (TCC) per replicate sample, according to patient status (CD or HC). Each boxplot displays the first and third quartile and the median line. Whiskers extend from the quartiles to 1.5 times the interquartile range. Points that lie outside this range are visualized as outliers.