nature portfolio

https://doi.org/10.1038/s41588-021-00983-z

## Supplementary information

# Challenges and future directions for studying effects of host genetics on the gut microbiome

In the format provided by the authors and unedited

**Challenges and future directions for studying effects of host genetics on the gut microbiome**

Serena Sanna[1,2], Alexander Kurilshikov[2], Adriaan van der Graaf[2,3], Jingyuan Fu[2,4], Alexandra Zhernakova[2]

[1] Institute for Genetic and Biomedical Research (IRGB), National Research Council (CNR), Cagliari, Italy

[2] University of Groningen and University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands

[3] Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

[4] Department of Pediatrics, University of Groningen and University Medical Center Groningen, Groningen, the Netherlands

Correspondence to:
Prof. A. Zhernakova: sasha.zhernakova@gmail.com
Dr. S. Sanna: serena.sanna@irgb.cnr.it

**Supplementary Note:**

To identify shared loci, we proceeded as follows. For each of the 12 studies, we extracted all the SNPs reported to be associated with abundances of taxonomies, microbial pathways or beta-diversity at genome-wide significance (**Supplementary Table 1**). SNP names were matched to their genomic position in build37. For the Finnrisk study, the only study analyzing rare and less frequent variants, we selected only those SNPs with MAF > 0.05 based on frequency information reported in this study. For each study, we defined the associated regions by looking at consecutive SNPs. If they were closer than 100 Kb, they were used to define the upper and lower limits of the locus. Otherwise, the two SNPs were considered independent. This process was repeated recursively for all SNPs in a study. We then considered two loci from two studies to be the same if they were closer than 100Kb, and in that case merged the two loci into one region. Given the extensive linkage disequilibrium around *LCT*, we merged regions around this gene if they were closer than 200Kb. The number and names of studies reporting each locus, as well as their genomic position, can be found in **Supplementary Table 2**.