# Population-level analysis of gut microbiome variation

Gut microbiome:

An essential component of human health

A complex trait

An adaptive and dynamic ecosystems

Many current studies focus on the microbiome has been on understanding **disease**:

.
Case-control design

Dysbiosis

Diagnostic and therapeutic application

| Dysbiosis-Associated Diseases or Conditions |
|---|
| Obesity |
| Metabolic syndrome |
| Nonalcoholic steatohepatitis |
| Inflammatory bowel diseases (Crohn's disease, ulcerative colitis, pouchitis) |
| Irritable bowel syndrome, functional bowel disorders |
| Atherosclerosis |
| Type 1 diabetes |
| Autism |
| Allergy |
| Asthma |
| Celiac disease |

Major challenges:

1, huge inter-individual variation requires large-scale sample sizes

2, configuration of a universal pattern of healthy microbiome

3, interaction between microbiota, host, and environmental factors

population-scale metagenome project

Configuration of healthy microbiome
-- features that broadly distinguish healthy from unhealthy microbiomes

Prevalent organisms: a core set of microbial taxa

Molecular pathways: a healthy “functional core”

Ecological properties: diversity
stability

Impacts of host and environmental factors:

Geography

Early Life exposures

Diet

Medication

Genetics

…

Increasing throughput and cost-effectiveness, large-scale projects have since been launched:

| | |
|---|---|
| Metagenomes of the Human Intestinal Tract (Meta-HIT) (2010, Europe) | 124 |
| Human Microbiome Project (HMP) (2012,US) | 242 |
| LifeLines-DEEP Study (LLDeep) (2016, Dutch) | 1135 |
| Flemish Gut Flora Project (FGFP) (2016, Belgian) | 1106 |
| … | … |

Total: nearly 4,000 well-profiled individuals

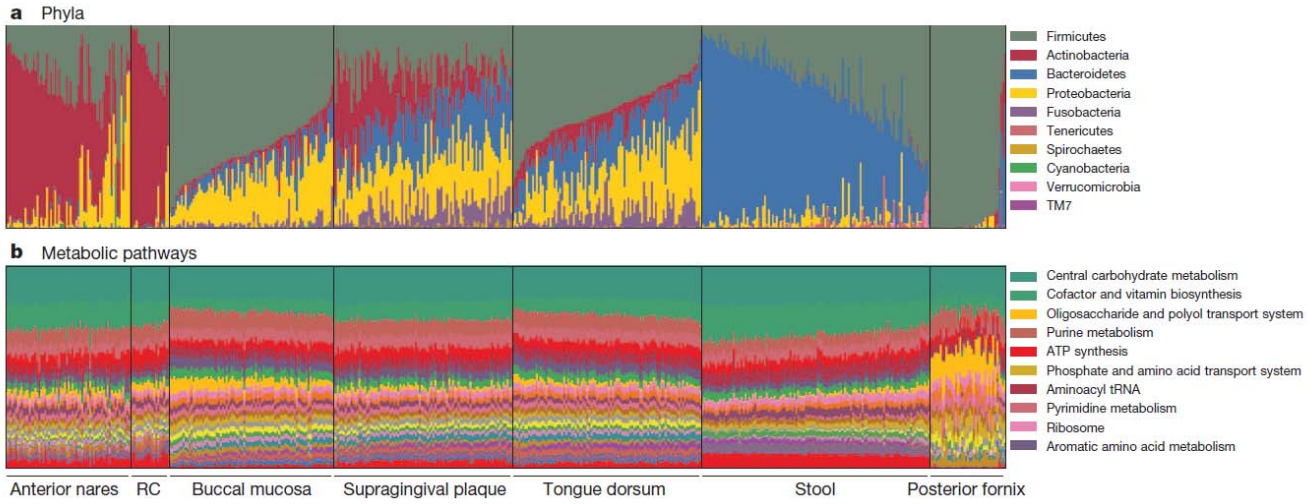
HMP
(Introduction, project objectives)

Identification of core microbiome

Identify a “core” set of microbial taxa universally present in healthy individuals:

Hypothesis: Absences of such microbes would indicate dysbiosis.

An alternative hypothesis: a “healthy functional core”



Impacts of host and environmental factors:

HMP projects show both clades and metabolism in the microbiota are associated with hosts' age, gender, body max index (BMI) and other clinical metadata.

However, the correlation strength is mild, even with the most significant associations possessing low effect size and considerable unexplained variance

Other potentially important factors short/long-term diet, daily cycles, founder effects such as mode of delivery and host genetics should be considered in future analyses

The LifeLines-DEEP cohort, a Dutch population-based study.

REPORTS

MICROBIOME

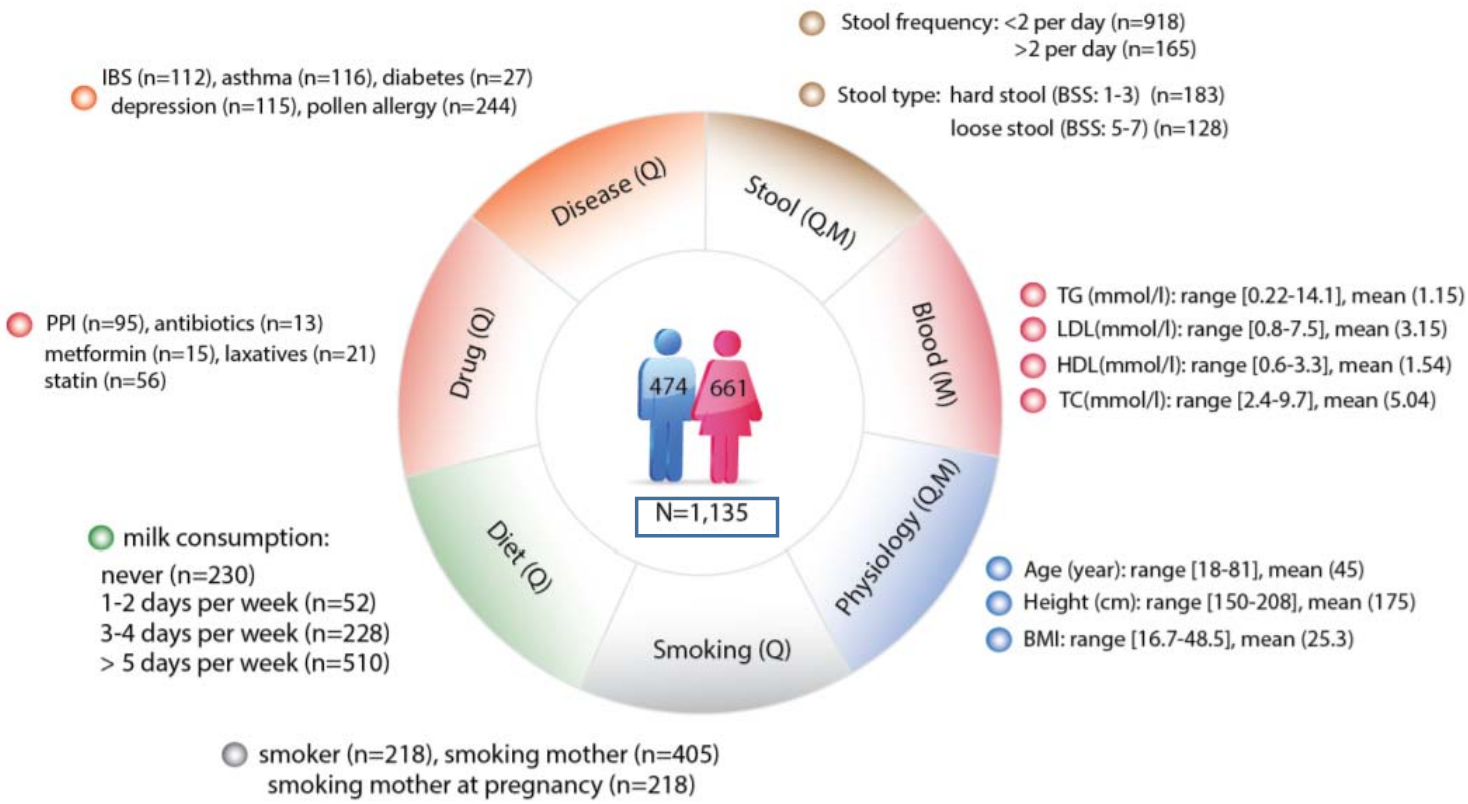
# Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity

Alexandra Zhernakova,[1,2*] Alexander Kurilshikov,[3,4†] Marc Jan Bonder,[1†] Ettje F. Tigchelaar,[1,2†] Melanie Schirmer,[5,6] Tommi Vatanen,[5,7] Zlatan Mujagic,[2,8] Arnau Vich Vila,[9] Gwen Falony,[10,11] Sara Vieira-Silva,[10,11] Jun Wang,[10,11] Floris Imhann,[9] Eelke Brandsma,[12] Soesma A. Jankipersadsing,[1] Marie Joossens,[10,11,13] Maria Carmen Cenit,[1,14,15] Patrick Deelen,[1,16] Morris A. Swertz,[1,16] LifeLines cohort study, Rinse K. Weersma,[9] Edith J. M. Feskens,[2,17] Mihai G. Netea,[18] Dirk Gevers,[5‡] Daisy Jonkers,[8] Lude Franke,[1] Yurii S. Aulchenko,[4,19,20,21] Curtis Huttenhower,[5,6] Jeroen Raes,[10,11,13] Marten H. Hofker,[12] Ramnik J. Xavier,[5,22,23,24] Cisca Wijmenga,[1*§] Jingyuan Fu[1,12*§]

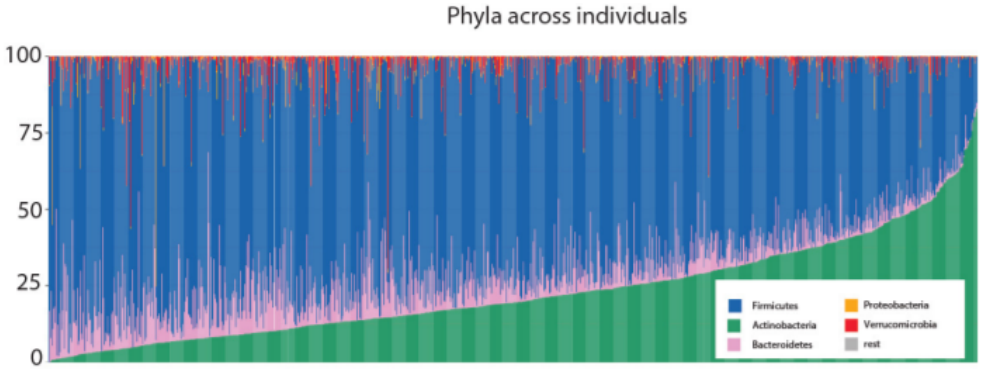
Deep sequencing of the gut microbiomes of 1135 participants from a Dutch population-based cohort shows relations between the microbiome and 126 exogenous and intrinsic host factors, including 31 intrinsic factors, 12 diseases, 19 drug groups, 4 smoking categories, and 60 dietary factors. These factors collectively explain 18.7% of the variation seen in the interindividual distance of microbial composition. We could associate 110 factors to 125 species and observed that fecal chromogranin A (CgA), a protein secreted by enteroendocrine cells, was exclusively associated with 61 microbial species whose abundance collectively accounted for 53% of microbial composition. Low CgA concentrations were seen in individuals

these participants: 41 intrinsic factors of various physiological and biomedical measures, 39 self-reported diseases, 44 categories of drugs, 5 categories of smoking status, and 78 dietary factors (fig. S1 and table S1). These factors cover dietary habits, lifestyle, medication use, and health parameters. Most of the factors showed a low or modest intercorrelation (table S2, A to C, and fig. S2, A to D); many are highly variable, including, as expected in the Dutch population, the high consumption of milk products and low use of antibiotics. Antibiotic use in the Netherlands is the lowest in Europe, at a level half that of the UK and one-third that of Belgium. To cover health-domain factors relevant to the host immune system and gut health, we collected cell counts for eight different blood cell types, measured blood cytokine concentrations, assessed stool frequency and stool type by Bristol stool score, and measured fecal levels of several secreted proteins, including calprotectin as a marker for the immune system activation, human β-defensin-2 (HBD-2) as a marker for defense against invading microbes, and chromogranin A (CgA) as a marker for neuroendocrine system activation.

After quality control and removal of sequence reads mapping to the human genome, the microbiome sequence reads were mapped to ~1 million microbial-taxonomy–specific marker genes with MetaPhlAn 2.0 (*5*) to predict the abundance of microorganisms (fig. S3A). For each participant, we predicted the abundances for 1649 microbial taxonomic clades ranging from four different do-

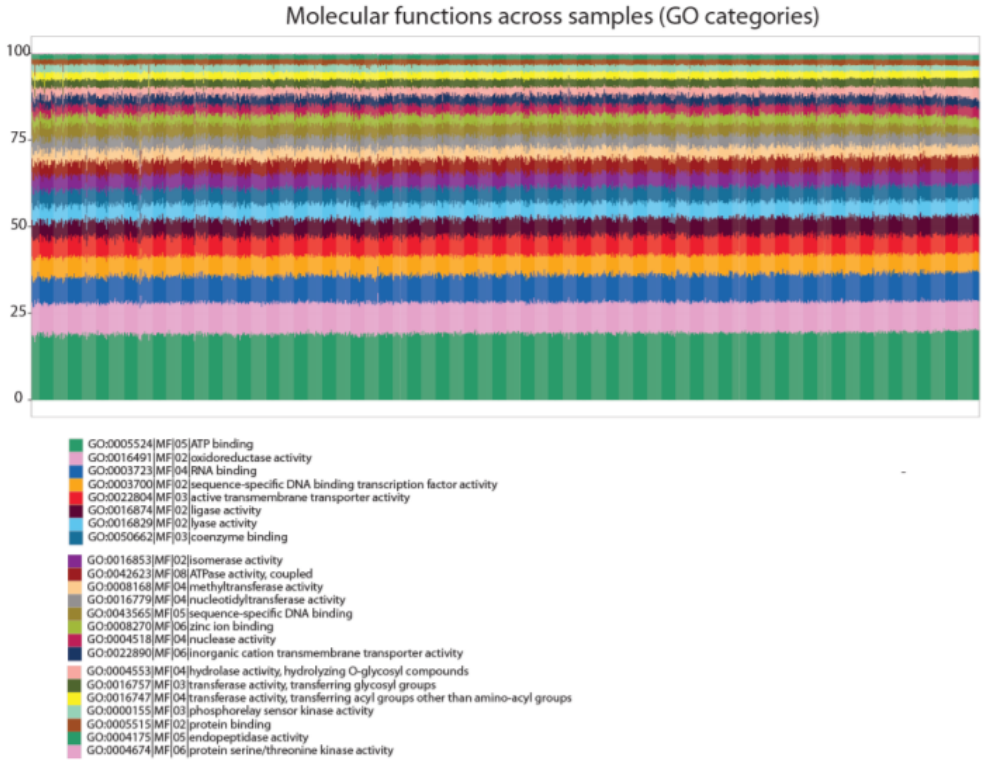

Stool frequency: <2 per day (n=918)
>2 per day (n=165)
Stool type: hard stool (BSS: 1-3) (n=183)
loose stool (BSS: 5-7) (n=128)
IBS (n=112), asthma (n=116), diabetes (n=27)
depression (n=115), pollen allergy (n=244)
Disease (Q)
Stool (Q,M)
Drug (Q)
Blood (M)
474
661
N=1,135
Diet (Q)
Physiology (Q,M)
Smoking (Q)
PPI (n=95), antibiotics (n=13)
metformin (n=15), laxatives (n=21)
statin (n=56)
TG (mmol/l): range [0.22-14.1], mean (1.15)
LDL(mmol/l): range [0.8-7.5], mean (3.15)
HDL(mmol/l): range [0.6-3.3], mean (1.54)
TC(mmol/l): range [2.4-9.7], mean (5.04)
milk consumption:
never (n=230)
1-2 days per week (n=52)
3-4 days per week (n=228)
> 5 days per week (n=510)
Age (year): range [18-81], mean (45)
Height (cm): range [150-208], mean (175)
BMI: range [16.7-48.5], mean (25.3)
smoker (n=218), smoking mother (n=405)
smoking mother at pregnancy (n=218)

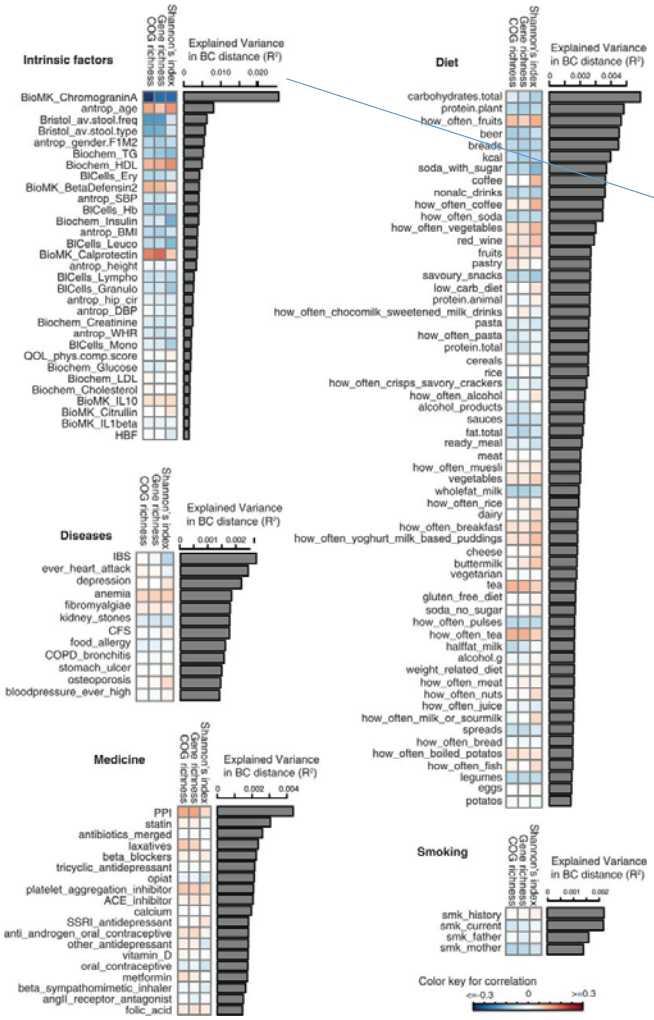
The microbial composition of each individual is shown at phylum level. The individuals are sorted by the abundance of *Actinobacteria*.

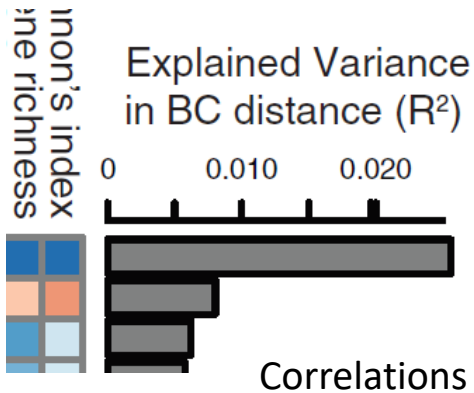



High interindividual variation of community composition

High function stability across individuals based on GO categories

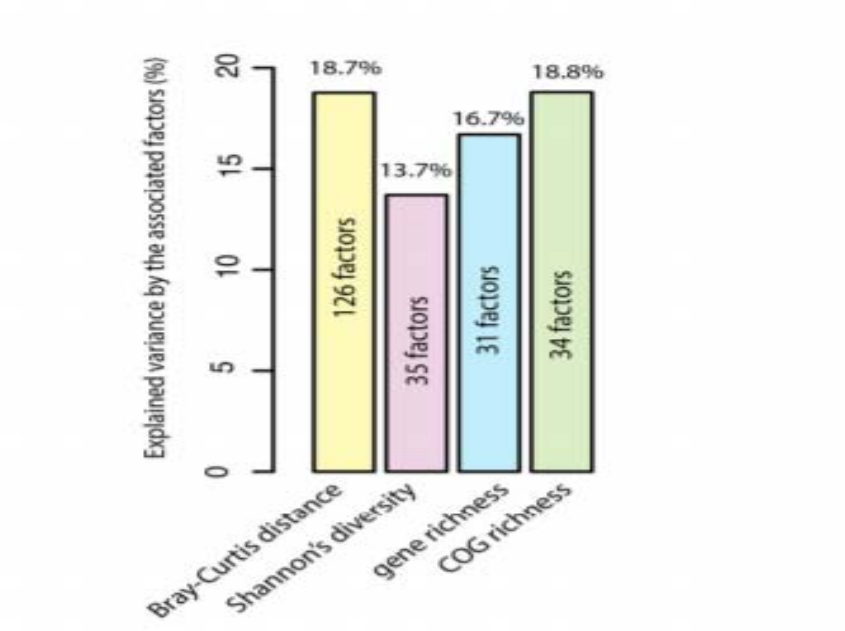





Correlations for 207 phenotype factors with the interindividual variation in microbial composition, diversity, richness of genes and COGs

No more than 18.8% variation could be explained by the associations above.



The cross-sectional studies

Associative studies:

the associations cannot be equated with proof of causation.

Still important to the exploration of the complex relationship between the gut microbiome and human health

Throw the light on further study

To increase the power of such studies, a longitudinal design is expected.

long-term effect of diet, medication factors, and the respective microbiota characteristics or fluctuations.

While more expensive and logistically challenging, be necessary to facilitate predictive models and establish the causality of dysbioses.