Population-level analysis of gut microbiome variation

Department of Microbiology, CUHK

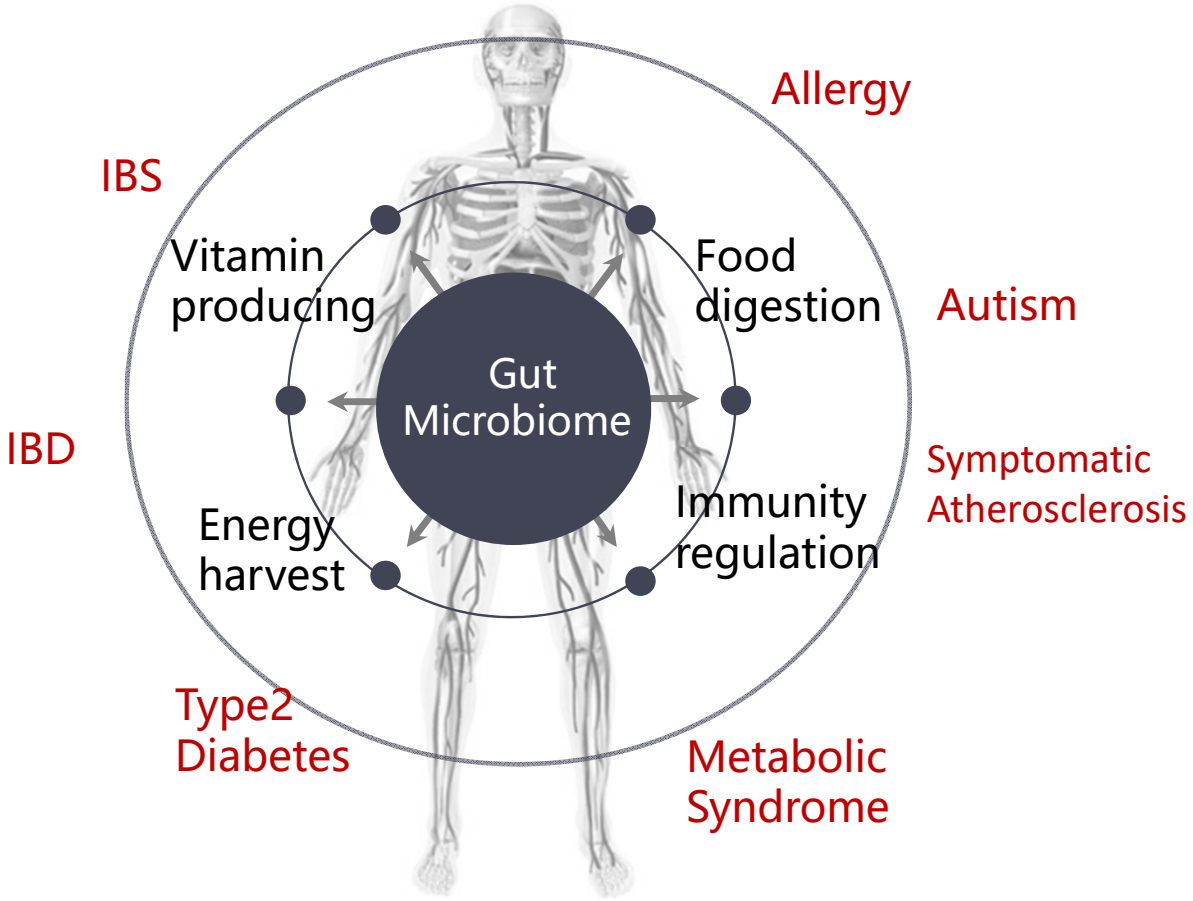
Ph.D. Student: WEI, Yuchen

Supervisor: Prof. Guoping ZHAO

1st, Dec, 2016

Background

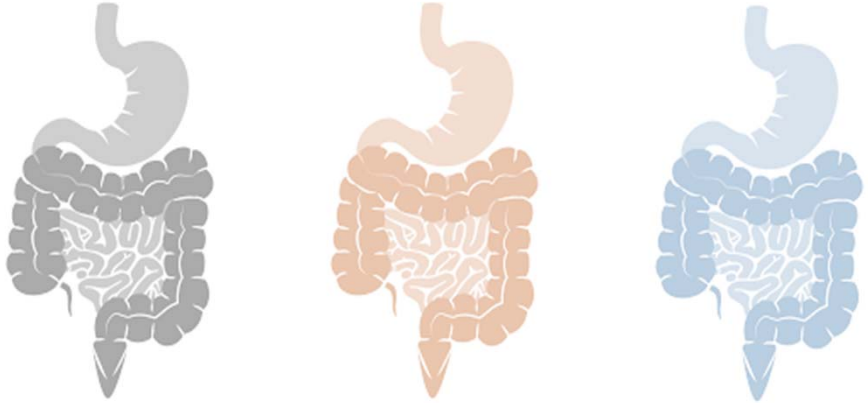
Gut microbiome: an essential component of human health



Enterotype (Arumugam M, 2011)

"An enterotype is a classification of living organisms based on its bacteriological ecosystem in the gut microbiome"

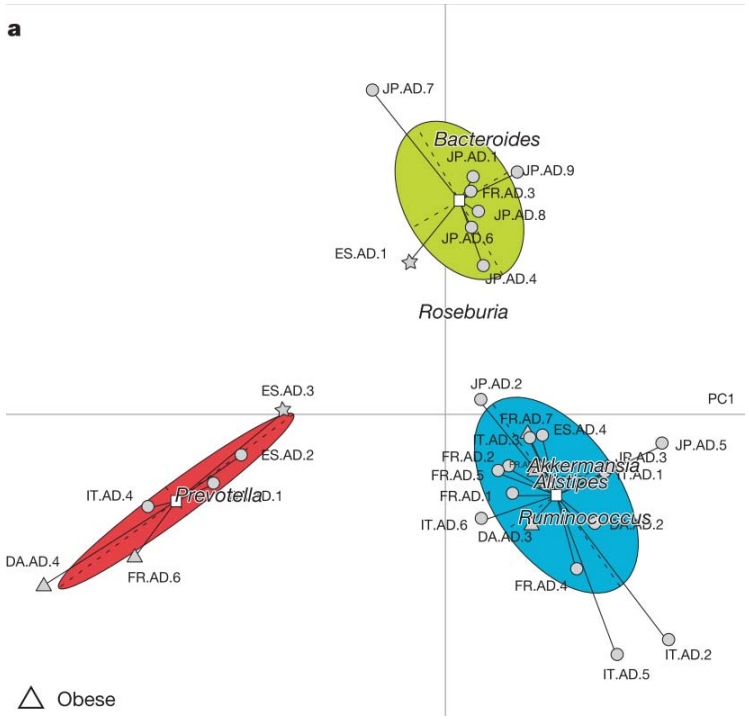
-- *Wikipedia*



Bacteroides	*Prevotella*	*Ruminococcus*

Three enterotypes have been proposed by the original study

Enterotype (Arumugam M, 2011)



"Several measured host properties, namely nationality, gender, age or body mass index (BMI), do not seem to significantly correlates with the enterotypes"

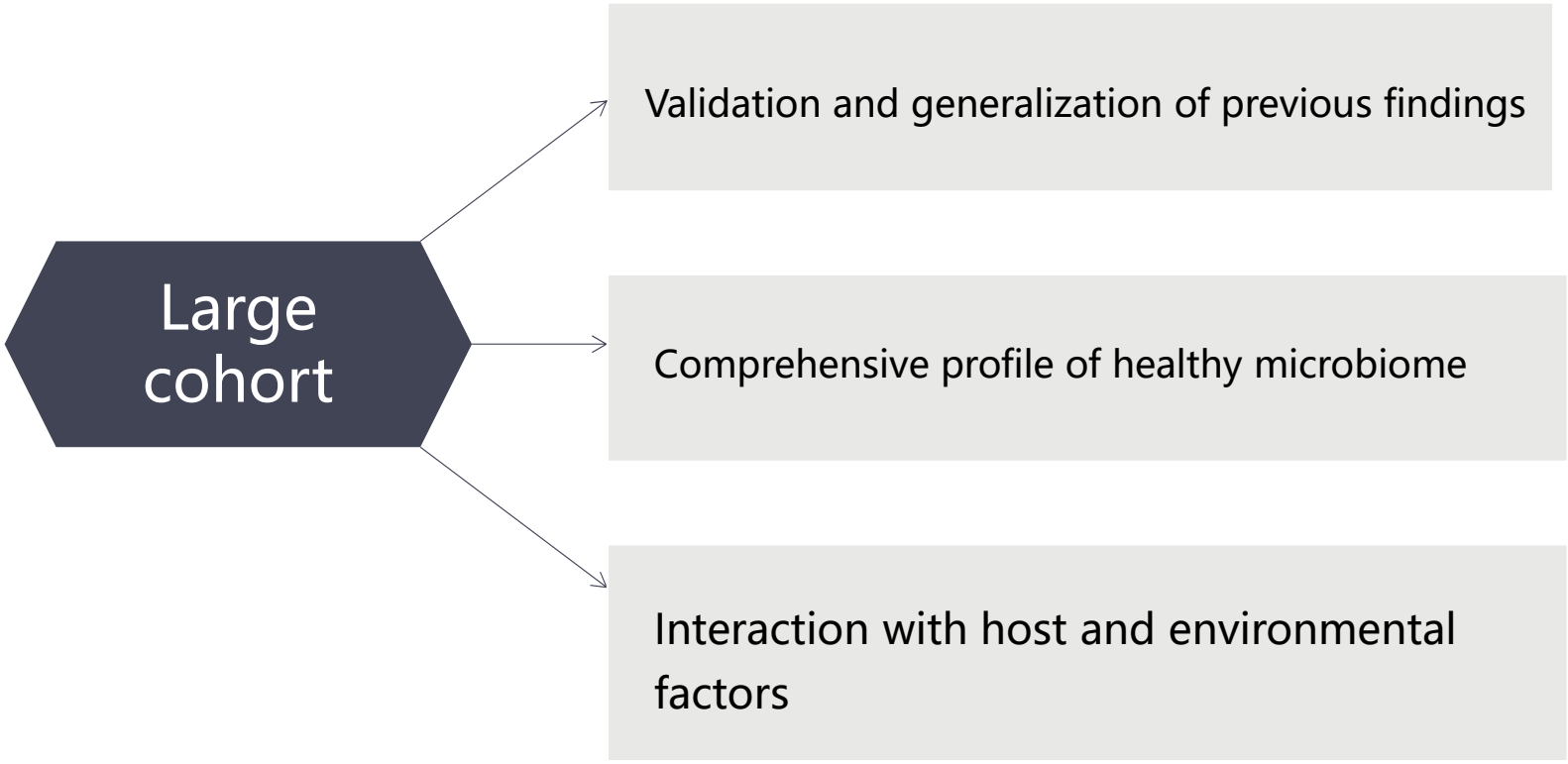
Arumugam M, Raes J, Pelletier E, Le P, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature [Internet]. 2011;473(7346):174–80.

The consistent enterotype have been observed by the following studies.

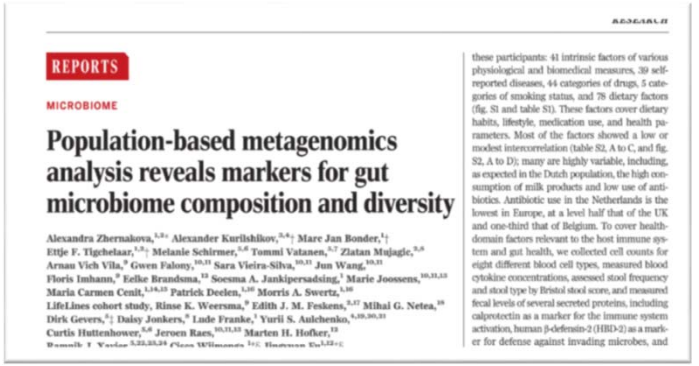
The long-term dietary(1), early year experiences(2), residence types(2), have been reported to associated with enterotypes

The impacts of BMI(1)(3), family nesting(1)(2), etc. are still under debates

1, Wu GD, Chen J, Hoffmann C, et al. Linking Long-Term Dietary Patterns with Gut Microbial Enterotypes. Science (New York, N.y). 2011;334(6052):105-108.
2,Moeller AH, et al. Chimpanzees and humans harbour compositionally similar gut enterotypes. Nat Commun 2012 Nov 13
3, Lim MY, Rho M, Song Y-M, Lee K, Sung J, Ko G. Stability of gut enterotypes in Korean monozygotic twins and their association with biomarkers and


Large cohort
Validation and generalization of previous findings
Comprehensive profile of healthy microbiome
Interaction with host and environmental factors

LifeLines-Deep Cohort, Netherland (N=1135)



REPORTS

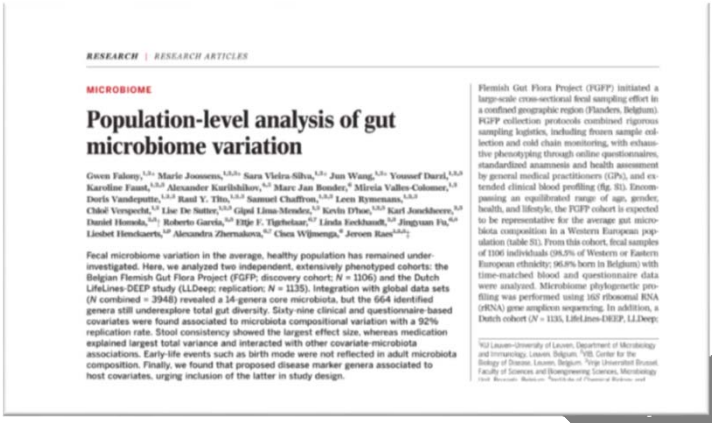
MICROBIOME

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

Alexandra Zhernakova, Alexander Kurilshikov, Marc Jan Bonder, Ettje F. Tigchelaar, Melanie Schirmer, Tommi Vatanen, Zlatan Mujagic, Arnau Vich Vila, Gwen Falony, Sara Vieira-Silva, Jun Wang, Floris Imhann, Eelke Brandsma, Soesma A. Jankipersadsing, Marie Joossens, Maria Carmen Cenit, Patrick Deelen, Morris A. Swertz, LifeLines cohort study, Rinse K. Weersma, Edith J. M. Feskens, Mihai G. Netea, Dirk Gevers, Daisy Jonkers, Lude Franke, Yurii S. Aulchenko, Curtis Huttenhower, Jeroen Raes, Marten H. Hofker

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

Flemish Gut Flora Project, Belgium (N=1106)



RESEARCH | RESEARCH ARTICLES

MICROBIOME

Population-level analysis of gut microbiome variation

Gwen Falony, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, Marc Jan Bonder, Mireia Valles-Colomer, Doris Vandeputte, Raul Y. Tito, Samuel Chaffron, Leen Rymenans, Chloë Verspecht, Lise De Sutter, Gipsi Lima-Mendez, Kevin D'hoe, Karl Jonckheere, Daniel Homola, Roberto Garcia, Ettje F. Tigchelaar, Linda Eeckhaudt, Jingyuan Fu, Liesbet Henckaerts, Alexandra Zhernakova, Cisca Wijmenga, Jeroen Raes

Flemish Gut Flora Project (FGFP) initiated a large-scale cross-sectional fecal sampling effort in a confined geographic region (Flanders, Belgium). FGFP collection protocols combined rigorous sampling logistics, including frozen sample collection and cold chain monitoring, with exhaustive phenotyping through online questionnaires, standardized anamnesis and health assessment by general medical practitioners (GPs), and extended clinical blood profiling (fig. S1). Encompassing an equilibrated range of age, gender, health, and lifestyle, the FGFP cohort is expected to be representative for the average gut microbiota composition in a Western European population (table S1). From this cohort, fecal samples of 1106 individuals (98.5% of Western or Eastern European ethnicity; 96.6% born in Belgium) with time-matched blood and questionnaire data were analyzed. Microbiome phylogenetic profiling was performed using 16S ribosomal RNA (rRNA) gene amplicon sequencing. In addition, a Dutch cohort (N = 1135, LifeLines-DEEP, LLDeep;

Fecal microbiome variation in the average, healthy population has remained underinvestigated. Here, we analyzed two independent, extensively phenotyped cohorts: the Belgian Flemish Gut Flora Project (FGFP; discovery cohort; N = 1106) and the Dutch LifeLines-DEEP study (LLDeep; replication; N = 1135). Integration with global data sets (N combined = 3948) revealed a 14-genera core microbiota, but the 664 identified genera still underexplore total gut diversity. Sixty-nine clinical and questionnaire-based covariates were found associated to microbiota compositional variation with a 92% replication rate. Stool consistency showed the largest effect size, whereas medication explained largest total variance and interacted with other covariate-microbiota associations. Early-life events such as birth mode were not reflected in adult microbiota composition. Finally, we found that proposed disease marker genera associated to host covariates, urging inclusion of the latter in study design.

MICROBIOME

Population-level analysis of gut microbiome variation

Gwen Falony,[1,2]* Marie Joossens,[1,2,3]* Sara Vieira-Silva,[1,2]* Jun Wang,[1,2]* Youssef Darzi,[1,2,3] Karoline Faust,[1,2,3] Alexander Kurilshikov,[4,5] Marc Jan Bonder,[6] Mireia Valles-Colomer,[1,2] Doris Vandeputte,[1,2,3] Raul Y. Tito,[1,2,3] Samuel Chaffron,[1,2,3] Leen Rymenans,[1,2,3] Chloë Verspecht,[1,2] Lise De Sutter,[1,2,3] Gipsi Lima-Mendez,[1,2] Kevin D'hoe,[1,2,3] Karl Jonckheere,[2,3] Daniel Homola,[2,3]† Roberto Garcia,[2,3] Ettje F. Tigchelaar,[6,7] Linda Eeckhaudt,[2,3] Jingyuan Fu,[6,8] Liesbet Henckaerts,[1,9] Alexandra Zhernakova,[6,7] Cisca Wijmenga,[6] Jeroen Raes[1,2,3]‡

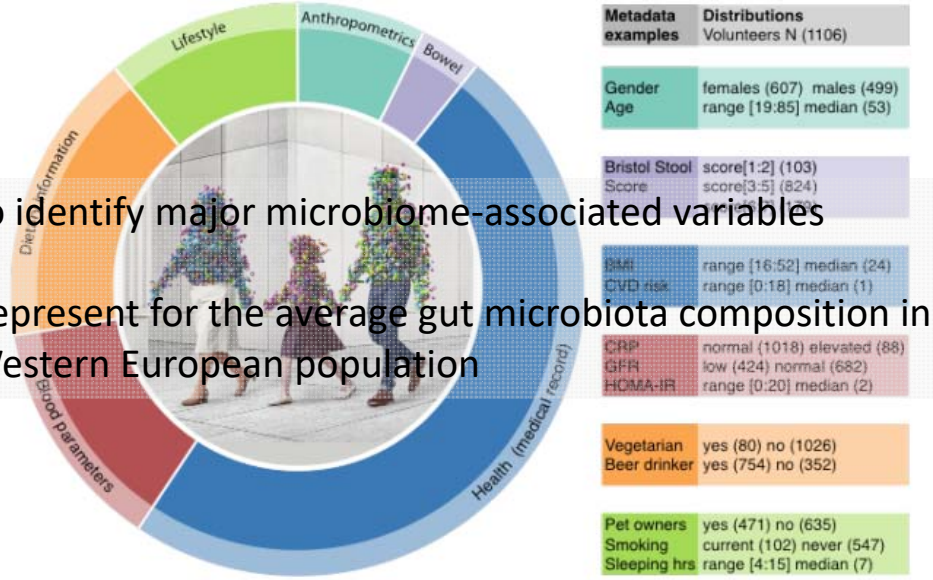
Fecal microbiome variation in the average, healthy population has remained under-investigated. Here, we analyzed two independent, extensively phenotyped cohorts: the Belgian Flemish Gut Flora Project (FGFP; discovery cohort; *N* = 1106) and the Dutch LifeLines-DEEP study (LLDeep; replication; *N* = 1135). Integration with global data sets (*N* combined = 3948) revealed a 14-genera core microbiota, but the 664 identified genera still underexplore total gut diversity. Sixty-nine clinical and questionnaire-based covariates were found associated to microbiota compositional variation with a 92% replication rate. Stool consistency showed the largest effect size, whereas medication explained largest total variance and interacted with other covariate-microbiota associations. Early-life events such as birth mode were not reflected in adult microbiota composition. Finally, we found that proposed disease marker genera associated to host covariates, urging inclusion of the latter in study design.

Flemish Gut Flora Project (FGFP) initiated a large-scale cross-sectional fecal sampling effort in a confined geographic region (Flanders, Belgium). FGFP collection protocols combined rigorous sampling logistics, including frozen sample collection and cold chain monitoring, with exhaustive phenotyping through online questionnaires, standardized anamnesis and health assessment by general medical practitioners (GPs), and extended clinical blood profiling (fig. S1). Encompassing an equilibrated range of age, gender, health, and lifestyle, the FGFP cohort is expected to be representative for the average gut microbiota composition in a Western European population (table S1). From this cohort, fecal samples of 1106 individuals (98.5% of Western or Eastern European ethnicity; 96.8% born in Belgium) with time-matched blood and questionnaire data were analyzed. Microbiome phylogenetic profiling was performed using 16*S* ribosomal RNA (rRNA) gene amplicon sequencing. In addition, a Dutch cohort (*N* = 1135, LifeLines-DEEP, LLDeep;

[1]KU Leuven–University of Leuven, Department of Microbiology and Immunology, Leuven, Belgium. [2]VIB, Center for the Biology of Disease, Leuven, Belgium. [3]Vrije Universiteit Brussel, Faculty of Sciences and Bioengineering Sciences, Microbiology Unit, Brussels, Belgium. [4]Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia.

The FPGP study

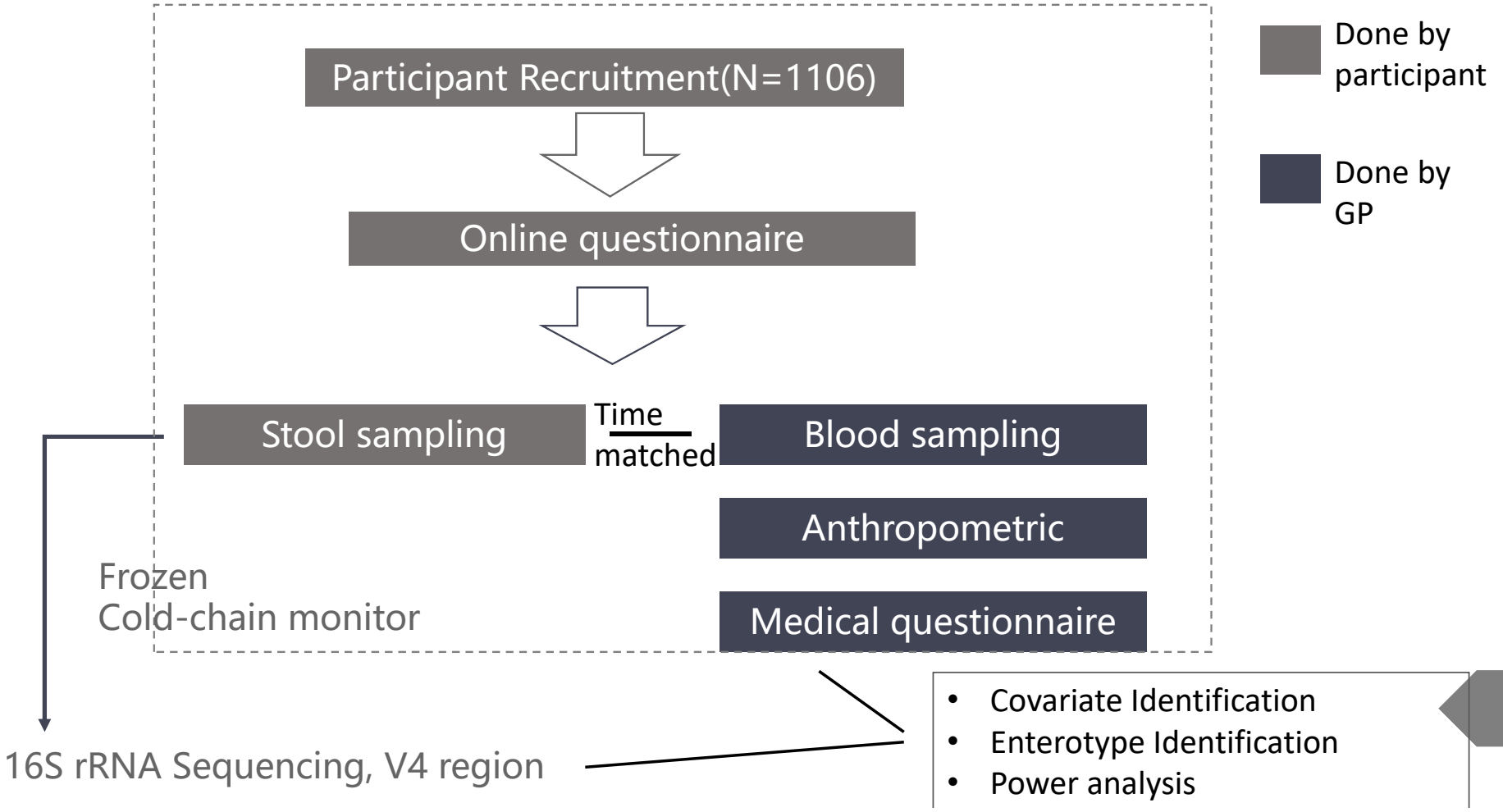
- Large-scale cross-sectional study
- Geographical confined region:
  Flander,Belgium
  (±6.5 million inhabitants, ±13,500 km2)
- Mono-ethnic region:
  96.8% Eastern or Western ethnicity
- Rigorous sampling protocol
- Exhaustive phenotyping



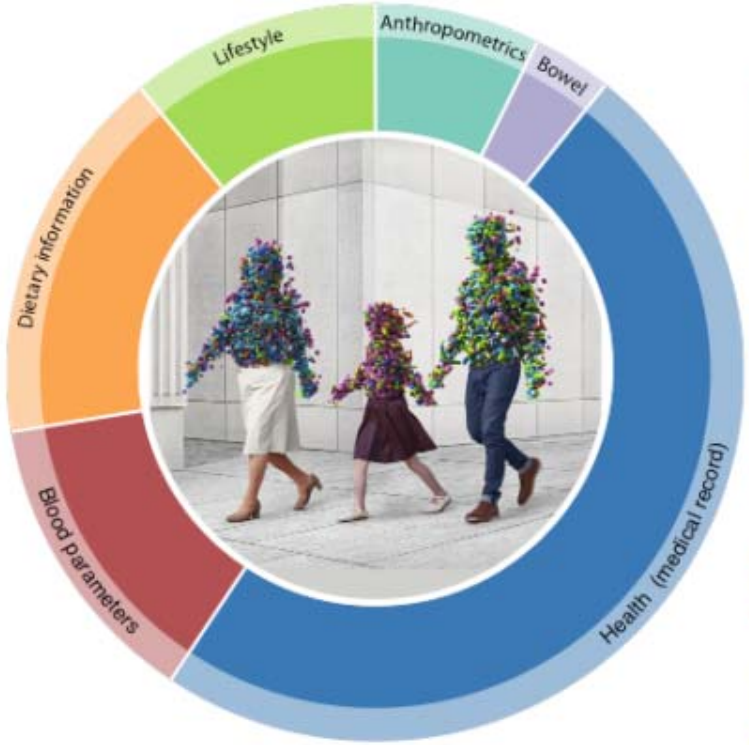
To identify major microbiome-associated variables

Represent for the average gut microbiota composition in a Western European population

The FPGP study



The FPGP study



Metadata examples	Distributions Volunteers N (1106)
Gender Age	females (607) males (499) range [19:85] median (53)
Bristol Stool Score	score[1:2] (103) score[3:5] (824) score[6:7] (179)
BMI CVD risk	range [16:52] median (24) range [0:18] median (1)
CRP GFR HOMA-IR	normal (1018) elevated (88) low (424) normal (682) range [0:20] median (2)
Vegetarian Beer drinker	yes (80) no (1026) yes (754) no (352)
Pet owners Smoking Sleeping hrs	yes (471) no (635) current (102) never (547) range [4:15] median (7)

Falony G, Joossens M, Vieira-silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation.

The FPGP study

Identifying microbiome covariate

503 metadata variable

MANOVA test, FDR < 10%
Collinear variables removed

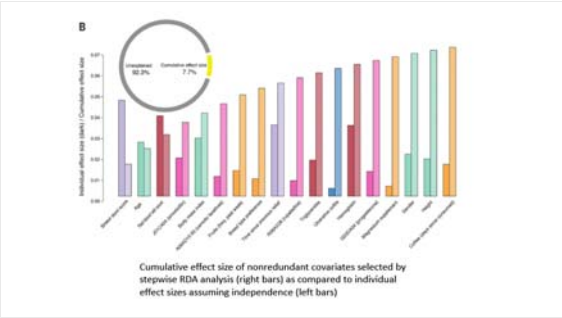
69 factors

RDA

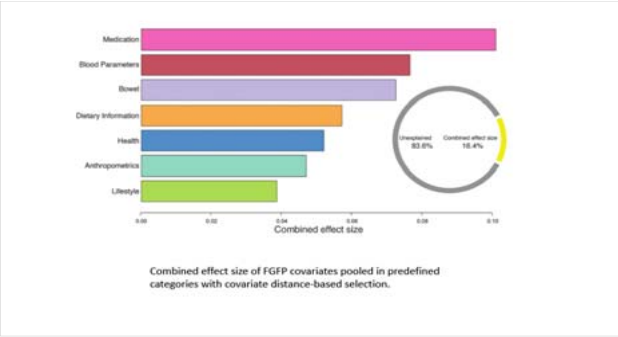
Pool into predefined categories

18 factors

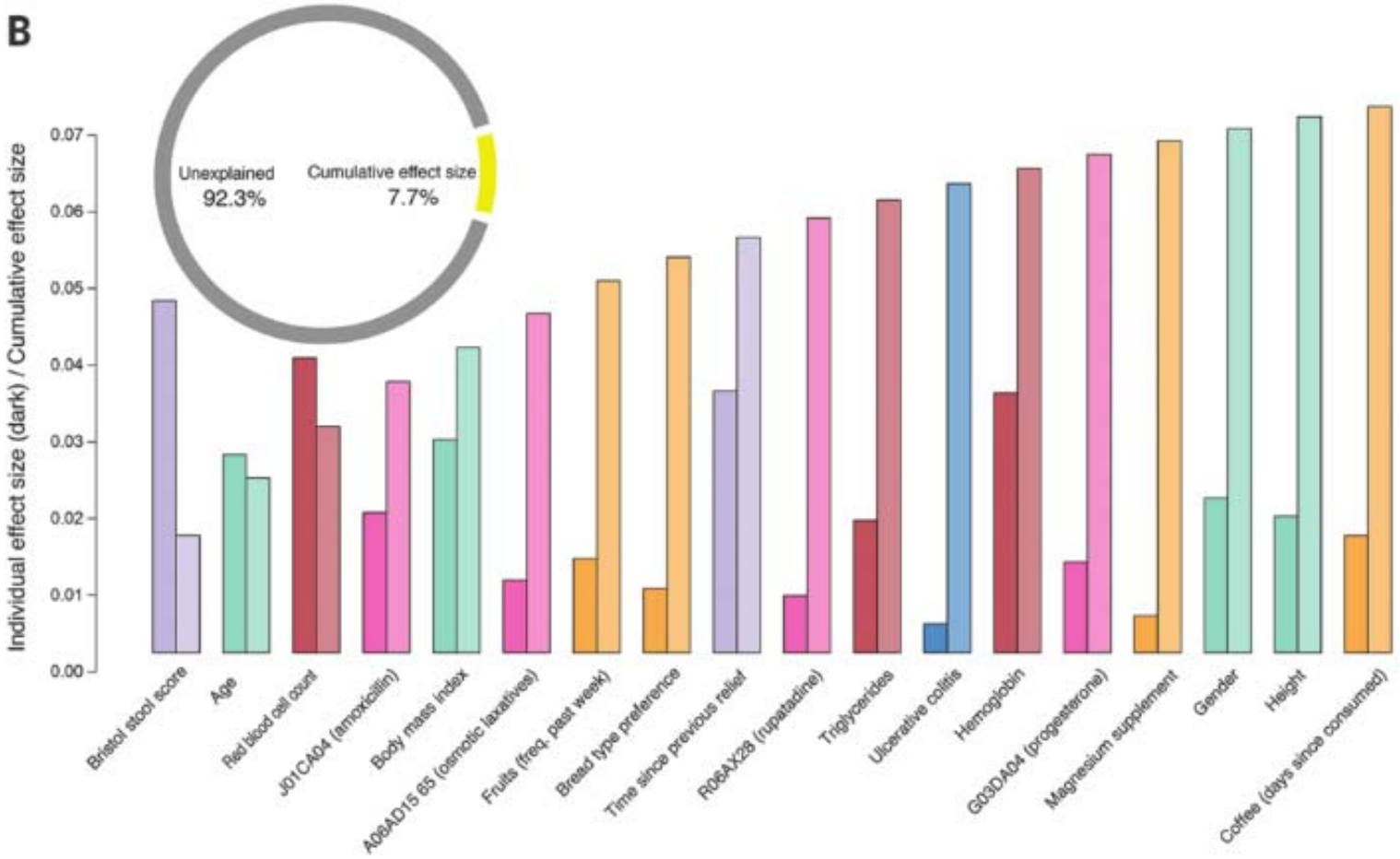
Covariates' combined effect size per phenotypical category revealed



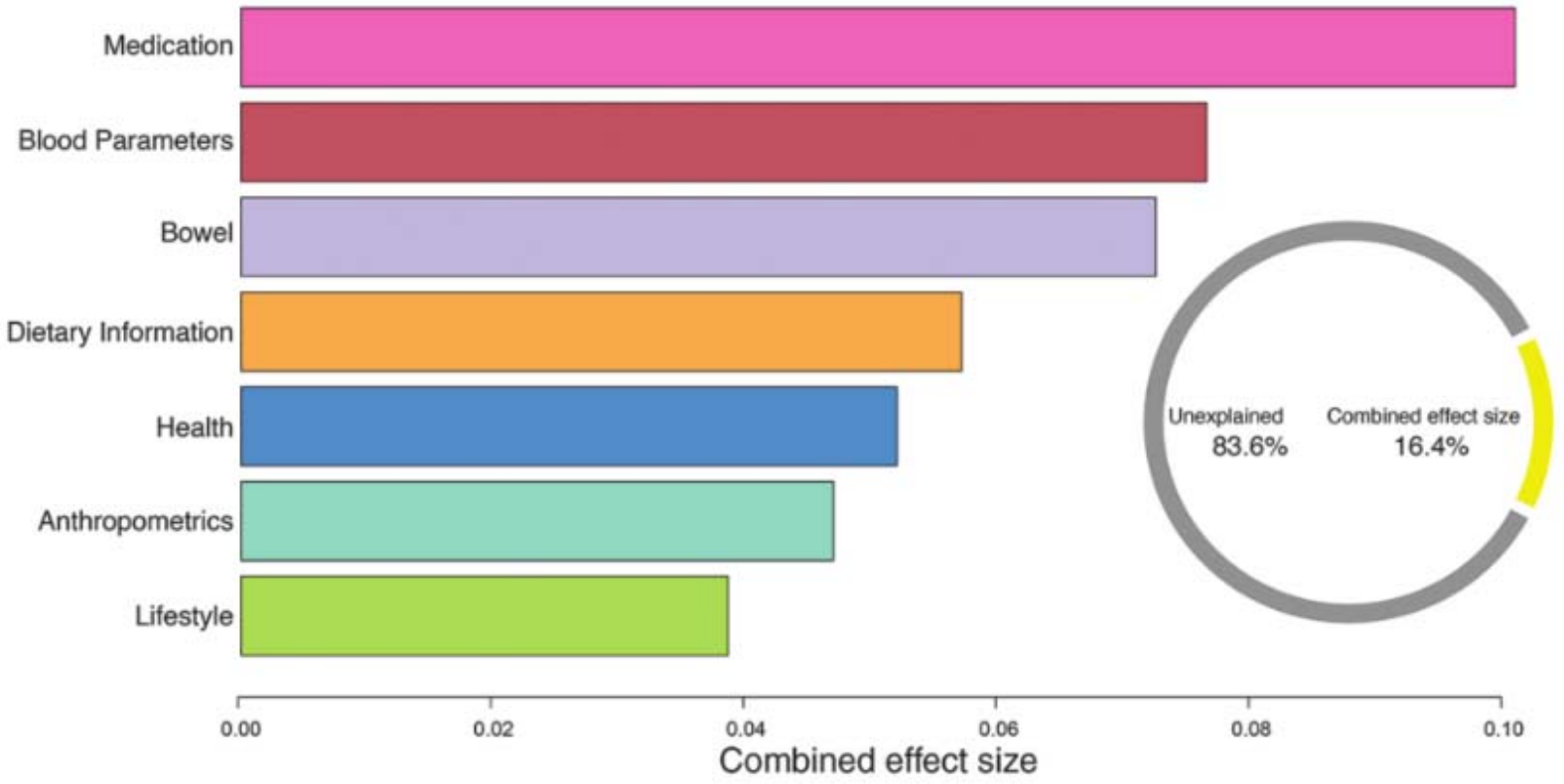
Cumulative effect size of nonredundant covariates selected by stepwise RDA analysis (right bars) as compared to individual effect sizes assuming independence (left bars)



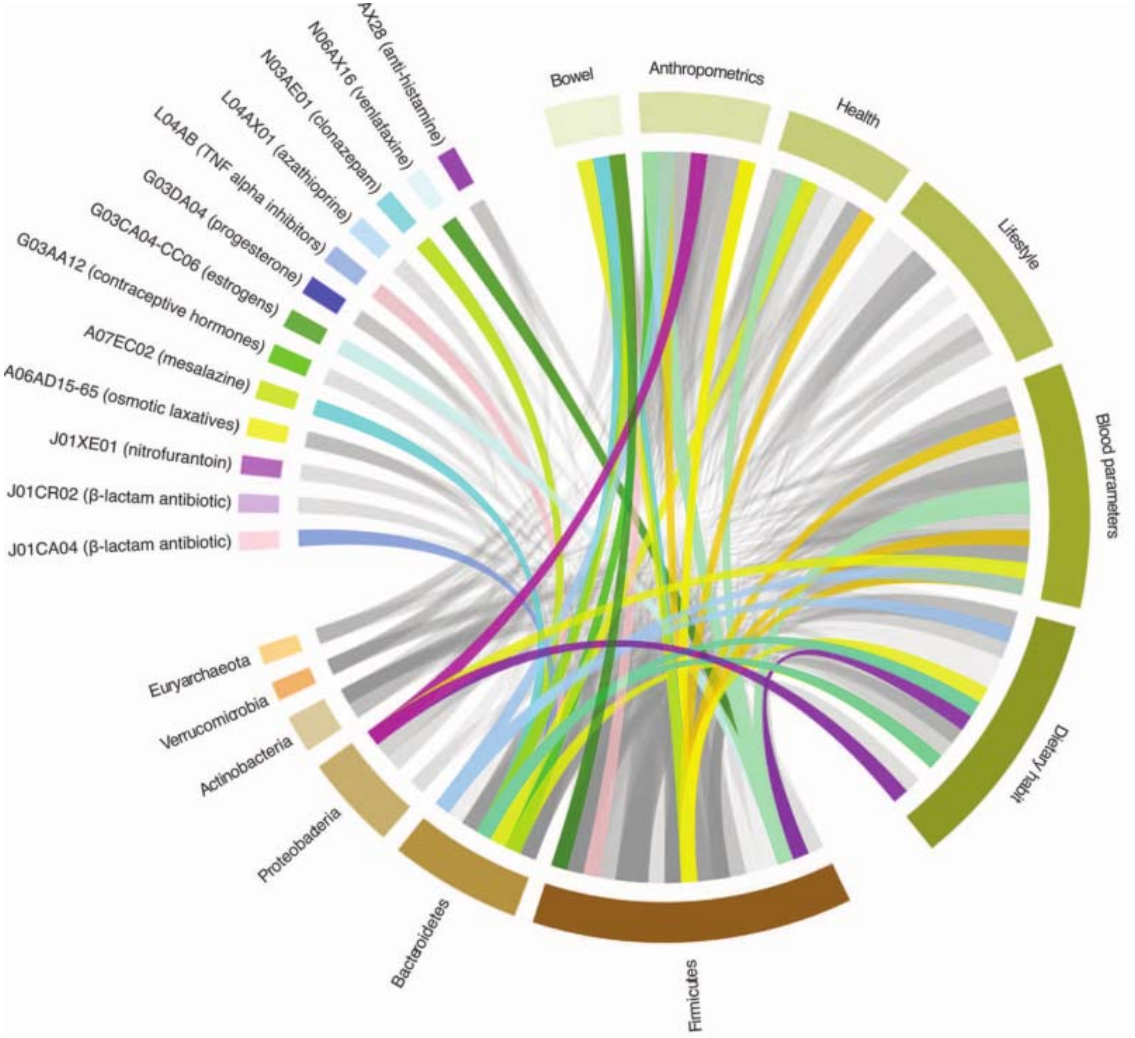
Combined effect size of FGFP covariates pooled in predefined categories with covariate distance-based selection.



Cumulative effect size of nonredundant covariates selected by stepwise RDA analysis (right bars) as compared to individual effect sizes assuming independence (left bars)



Combined effect size of FGFP covariates pooled in predefined categories with covariate distance-based selection.

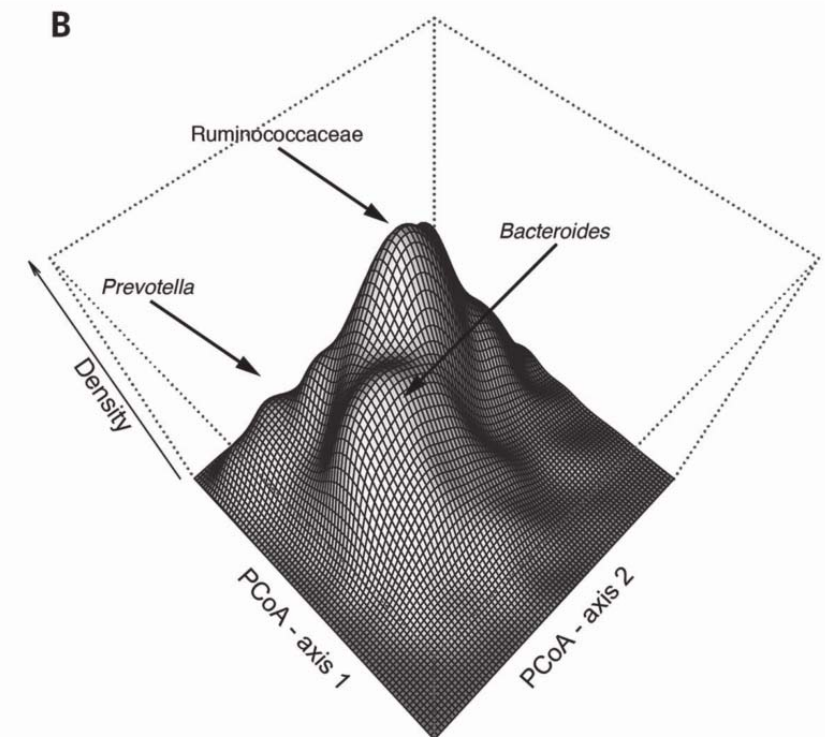


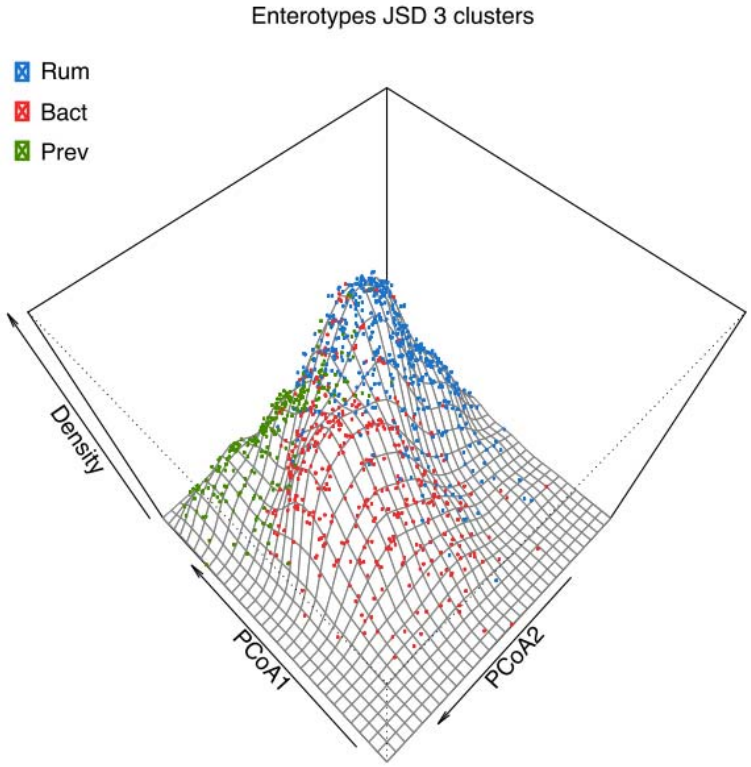
Of the covariate interactions detected, 63% were driven by medication.

Drug-microbiome associations as potentially confounding factors in clinical studies

The FPGP study

Identifying the enterotypes





The FPGP study

An alternative method for identifying the sample subsets:
Bi-clustering approach: group the taxa and sample simultaneously.

Two stable bi-clusters were detected, spanning 410 and 374 samples, respectively, with an intersection of 92

Cluster 1

- Clostridia
- Women
- Lower Weight
- Elevated microbiota richness

Ruminococcus enterotype

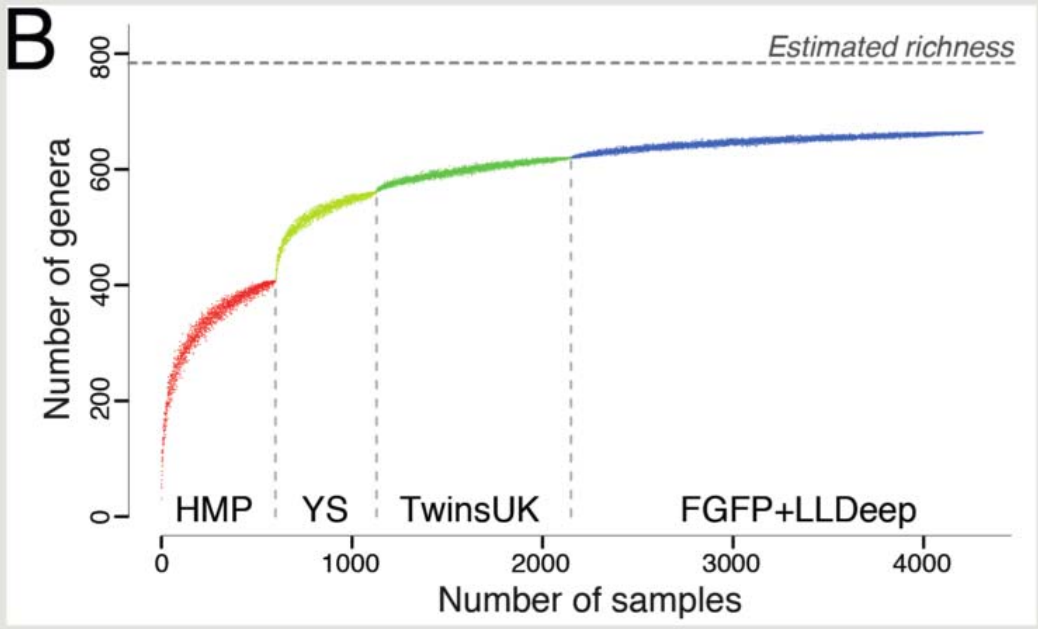
Cluster 2

- Bacteroides
- Reduced microbiome diversity
- Preference for white, low-fiber bread
- High prevalence of recent amoxicillin treatment

Bacteroides enterotype

The FPGP study

Collector's curve and power analysis



Combining the FGFP and LLDeep data with other U.K. and U.S. studies, yielding nearly 4000 well-profiled individuals

Total western richness is still under-sampled

The FPGP study

Statistical power analysis:

To detect an unknown shift (an unstudied disease)

A 9% difference between taxon proportions with 400 samples per group at a power above 95% and a 5% difference with 500 samples per group at a power of 80%

a known association in a background of other factors (Take BMI as an example)

It is estimated that 865 lean (BMI <25) and 865 obese (BMI ≥30) volunteers would be necessary to study microbiota compositional shifts with P < 5%significance level and a power of 80%

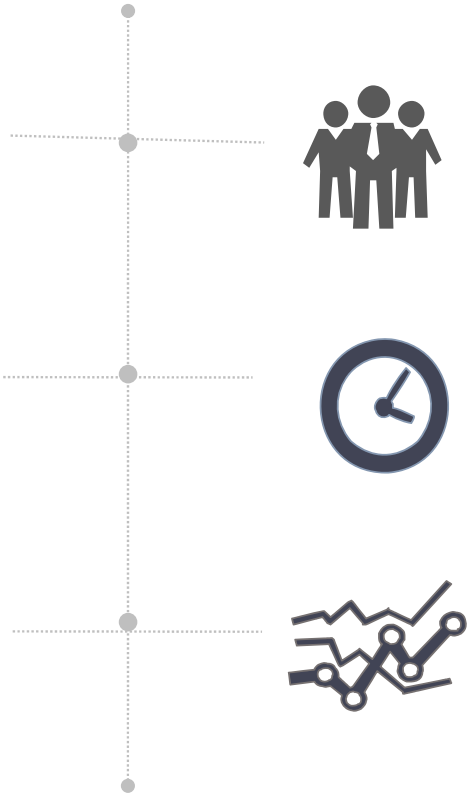
The FPGP study

Total gut diversity is not yet covered

The variation of microbiome that can be explained by phenotype factors is still modest

Large-scale study design is indispensable

Look into future:

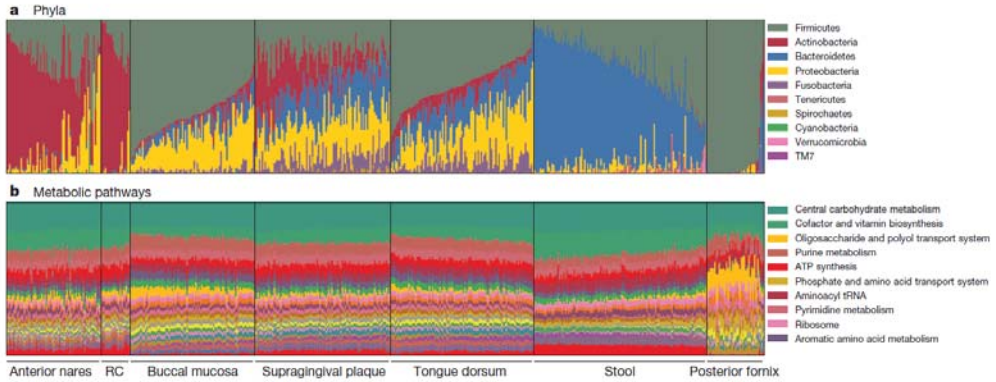


Base on the existed cohort

Longitudinal study design

Computational method to narrow down the subsets of general population

Thank you



MetaHIT

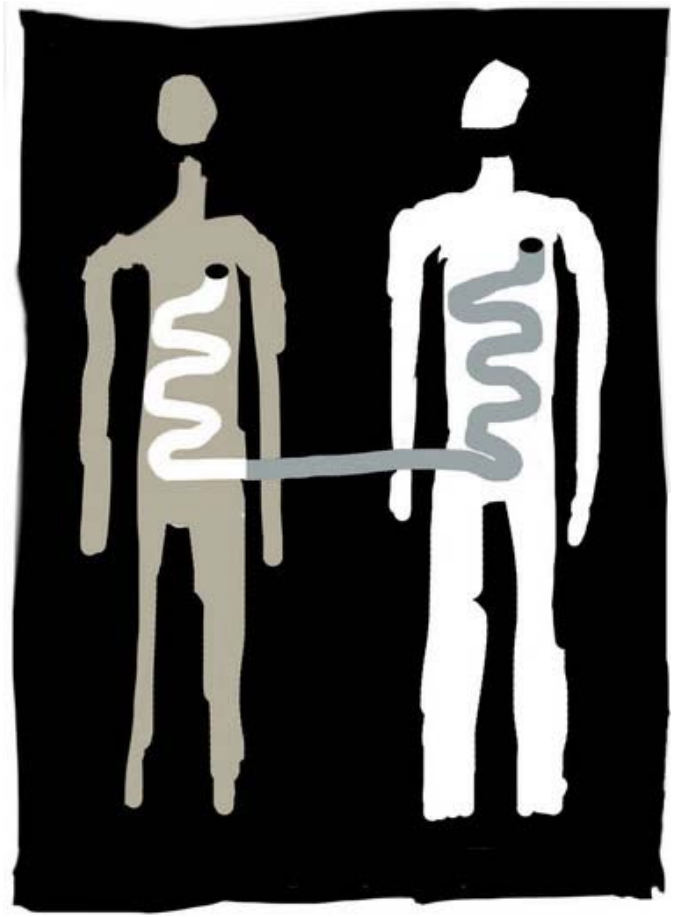
您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。您的内容打在这里，或者通过复制您的文本后，在此框中选择粘贴，并选择只保留文字。

Reference population
Functional core microbiome

HMP

North American

Reference population
Organismal and functional core microbiome
Several phenotypes information



- Chronic gastrointestinal diseases
  - IBD
  - IBS
- Systemic diseases
  - Metabolic syndrome
  - Type2 diabetes
  - Auto-immune diseases
- Psychiatric diseases