### REPORTS

#### MICROBIOME

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

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

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 with a more diverse microbiome. These results are an important step toward a better understanding of environment-diet-microbe-host interactions.

he human gut microbiome plays a major role in the production of vitamins, enzymes, and other compounds that digest and metabolize food and regulate our immune system (1). It can be considered as an extra organ, with remarkable dynamics and a major impact on our physiology. The composition of the gut microbiome can be considered as a complex trait, with the quantitative variation in the microbiome affected by a large number of host and environmental factors, each of which may have only a small additive effect, making it difficult to identify the association for each separate item. In this study, we present a systematic metagenomic association analysis of 207 intrinsic and exogenous factors from the LifeLines-DEEP cohort, a Dutch population-based study (2, 3). Our study reveals covariates in the microbiome and, more importantly, provides a list of factors that correlate with shifts in the microbiome composition and functionality.

This study includes stool samples from 1179 LifeLines-DEEP participants from the general population of the northern part of the Netherlands (2). The cohort comprised predominantly Dutch participants; 93.7% had both parents born in the Netherlands. The gut microbiome was analyzed with paired-end metagenomic shotgun sequencing (MGS) on a HiSeq 2000, generating an average of 3.0 Gb of data (about 32.3 million reads) per sample (4). After excluding 44 samples with low read counts, 1135 participants (474 males and 661 females) remained for further analysis. We tested 207 factors with respect to the microbiomes of

these participants: 41 intrinsic factors of various physiological and biomedical measures, 39 selfreported 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 healthdomain 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  $\beta$ -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 domains to 632 species (Fig. 1A). Most of the reads (97.6%) came from *Bacteria*; 2.2% were from Archaea, 0.2% from Viruses, and <0.01% from Eukaryotes. Comparison to previous taxonomic profiles of the same subjects by 16S ribosomal RNA (rRNA) gene sequencing (Fig. 1B) showed that MGS predicted more microbial species but fewer families and genera. At the phylum level, the abundances of dominant bacterial phyla Firmicutes (63.7%) and Bacteroidetes (8.1%) were similar to estimates based on 16S rRNA gene sequencing, but the abundance of Actinobacteria was higher in MGS (22.3%) than 16S (12.3%) (fig. S4). The microbiome quality control project has recently suggested that microbial composition estimates may not be comparable between studies if sample preparation and data analysis are not done in the same way (6). For instance, compared to the composition reported in other studies of a similar size that used different methods (7, 8), our study

<sup>1</sup>University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, Netherlands. <sup>2</sup>Top Institute Food and Nutrition, Wageningen, Netherlands. <sup>3</sup>Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia. <sup>4</sup>Novosibirsk State University, Novosibirsk, Russia. <sup>5</sup>The Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>6</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA. <sup>7</sup>Department of Computer Science, Aalto University School of Science, Espoo, Finland. <sup>8</sup>Division of Gastroenterology-Hepatology, Department of Internal Medicine, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Center, Maastricht, Netherlands. <sup>9</sup>University of Groningen, University Medical Center Groningen, Department of Gastroenterology and Hepatology, Groningen, Netherlands. <sup>10</sup>KU Leuven–University of Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Center Groningen, Netherlands. <sup>11</sup>Vrije Universiteit Brussel, Faculty of Sciences and Bioengineering Sciences, Microbiology Unit, Brussels, Belgium. <sup>11</sup>Vrije University of Groningen, Netherlands. <sup>13</sup>Drije University of Groningen, Netherlands. <sup>11</sup>Division of Human Nutrition, Wageningen University Hospital, Valencia, Spain. <sup>16</sup>University of Groningen, Netherlands. <sup>11</sup>Division of Human Nutrition, Wageningen University, Wageningen, Netherlands. <sup>12</sup>Department of Internal Medicine and Radboud Center for Infectious Disease, Radboud University Medical Center, Nijmegen, Netherlands. <sup>12</sup>Division of Human Nutrition, Wageningen University, Wageningen, Netherlands. <sup>12</sup>Department of Computers, University of Gooling, Resa, Bacha University Medical Center, Nijmegen, Netherlands. <sup>13</sup>Division of Human Nutrition, Wageningen University, Wageningen, Netherlands. <sup>13</sup>Department of Internal Medicine and Radboud Center for Infectious Disease, Radboud University Medical Center, Nijmegen, Netherlands. <sup>13</sup>Division of Human Nutrition,

\*Corresponding author. Email: a.zhernakova@umcg.nl (A.Z.); c.wijmenga@umcg.nl (C.W.); j.fu@umcg.nl (J.F.) †These authors contributed equally to this work. ‡Present address: Janssen Human Microbiome Institute, Janssen Research and Development, Cambridge, MA, USA. §These authors contributed equally to this work.

detected a higher abundance of *Actinobacteria* but a lower abundance of *Bacteroidetes*. Notably, all samples in our study were isolated and processed with the same pipeline, ensuring low technical variation and high analysis power to access the association of multiple factors with the microbiome.

The high interindividual variation reflects the community composition (fig. S5) and is clearly driven by the abundance of the dominant phyla (Fig. 2A). Our further analysis of microbial composition was confined to the 632 unique species (table S3). For the functional profiling, the abundances of 568,874 UniRef gene families were grouped into clusters of orthologous groups (COG) on the basis of the EggNOG database and MetaCyc pathways (fig. S3A). Although the distribution of diversity, genes, and COG richness showed high intervariability (Fig. 2, B to D), functional



**Fig. 1. The taxonomic tree of microbial taxonomies predicted by MGS and 16S rRNA gene sequen-cing. (A)** Taxonomic tree based on MGS shotgun sequencing data. (B) Taxonomic tree based on 16S rRNA gene sequencing data. Each dot represents a taxonomic entity. From the inner to outer circles, the taxonomic levels range from domains to species. Different colors of dots indicate different taxonomy levels according to the color key shown. Numbers in parentheses indicate the total number of unique taxonomies detected at each level.



Fig. 2. The interindividual variation of microbial composition and function profile. (A) Principal coordinates analysis (PCoA) plots of Bray-Curtis distance of microbial composition. The composition was driven by the most dominant phyla: *Firmicutes, Actinobacteria,* and *Bacteroidetes.* Each dot represents one individual. Color indicates the relative abundance of each phylum. (B) Distribution of Shannon's diversity index. (C) Distribution of the gene richness. (D) Distribution of the clusters of orthologous groups (COG) richness.

profiles based on 23 nonredundant, Gene Ontology molecular function categories remained stable (fig. S6) within our cohort, similar to previous reports (9).

We correlated 207 factors to the interindividual variation in microbial composition, diversity, richness of genes, and COGs (fig. S3B). At a false discovery rate (FDR) of <0.1, 126 factors were associated with interindividual distance of microbial composition (Bray-Curtis distance) (Fig. 3 and table S4), of which 90% could be replicated in 16S rRNA data from the same subjects (table S5 and fig. S7), together explaining 18.7% of the variation in composition distance (fig. S8A). A total of 35 factors were associated with Shannon's diversity index of microbial composition (together explaining 13.7% variation; table S6 and fig. S8A), of which 80% were replicated in 16S rRNA data from the same subjects (table S7); 31 factors were associated with gene richness (together explaining 16.7% of variation; table S8) and 34 factors with COG richness (explaining 18.8% of variation) (table S9 and fig. S8A; for replication rates, see table S10). We saw a large overlap between different diversity and richness analyses, and most of them were also associated with composition distance (fig. S8B).

We performed multivariate association analyses between each factor with 170 abundant species (>0.01% of total microbial composition and present in at least 10 individuals) and 215 MetaCyc pathways (fig. S3C). When corrected for age, gender, and sequence depth, we found 485 associations at FDR<0.1 between 110 factors and 125 species (table S11) and 524 associations between 71 factors and 176 MetaCyc pathways (table S12). By correcting the correlation structures among all 207 factors, the number of associations was reduced to 128 independent associations with species (table S13) and 215 associations with pathways (table S14).

Our data confirmed some previous findings and also yielded novel associations. In our study, age and gender were correlated not only with microbial composition distance and diversity but also with functional richness. Women showed higher COG richness than men (adjusted P = 0.03), and COG richness increased with age (adjusted P =0.002) (fig. S9). Multiple intrinsic parameters, such as blood cell counts and lipid concentrations, were associated to composition and function levels as well. For example, a higher amount of hemoglobin was consistently associated with lower diversity and functional richness (Fig. 3 and tables S6 to S9). The strongest associations that we found were for the fecal levels of several secreted proteins, including human β-defensin-2 (HBD-2), calprotectin (10, 11), and chromogranin A (CgA), with microbial composition, diversity, and functional richness (Figs. 3 and 4A), as well as with specific species (table S11) and pathways (table S12). Among these associations, CgA showed the strongest association with composition distance (adonis  $R^2$  = 0.03, adjusted P = 0.0006), microbial diversity (Spearman r = -0.22, adjusted  $P = 1.49 \times 10^{-12}$ ), gene richness (Spearman r = -0.23, adjusted  $P = 9.4 \times 10^{-13}$ ), and COG richness (Spearman

r = -0.285, adjusted  $P = 2.53 \times 10^{-20}$ ) (tables S4 to S9). The association of CgA with composition distance was then validated in an independent cohort of 19 individuals for whom 16S rRNA gene sequencing data were available (P = 0.0065) (fig. S10). A lower amount of CgA was associated with higher diversity, with functional richness, with high concentrations of high-density lipoprotein (HDL), and with intake of fruits and vegetables. In contrast, elevated fecal CgA was associated with high fecal levels of calprotectin, high blood concentrations of triglycerides, high stool frequency, soft stool type, and self-reported irritable bowel syndrome (IBS) (Fig. 4B). After correcting for the confounding effect of all other factors, our analysis revealed 61 species exclusively associated with CgA (Fig. 4, C and D, and table S13) whose abundances collectively accounted for 53% of the total abundance of the microbiome on average, and with 40 MetaCyc pathways (table S14) that accounted for 34.6% of the pathway profiles. The strongest association to CgA was observed for the Archaea species Methanobrevibacter smithii (fig. S11A), which plays an important role in the digestion of polysaccharides by consuming the end products of bacterial fermentation and methanogenesis (12) (fig. S11B). A negative association with CgA abundance was observed for 24 out of 36 species from phylum Bacteroidetes (Fig. 4, C and D).

CgA is a member of the granine peptides, which are secreted in nervous, endocrine, and immune cells under stress (13) and during active periods of gut-related diseases such as IBS and inflammatory bowel disease, although some findings are contradictory (14-16). Many different functions have been proposed for CgA and other granine peptides, including roles in neurological pathways, pain regulation, and antimicrobial activity against bacteria, fungi, and yeasts (17, 18). However, their mechanism of action and physiological importance need further detailed investigation. To test whether genetic variants that influence CHGA gene expression (encoding CgA) can affect fecal CgA level and the gut microbiome, we tested the effect of six single-nucleotide polymorphisms known to regulate gene expression of CHGA on fecal CgA and abundances of species (table S15). No significant association was observed, suggesting that genetic variation in CHGA expression does not explain the variation observed in the fecal CgA levels and microbiome composition (tables S16 and S17). Our observation that CgA strongly correlates with microbiome composition, especially with a large number of species from Bacteroidetes phylum, and with diversity will hopefully encourage studies to unravel the role of CgA in gut health.

We also observed associations (FDR<0.1) between 63 dietary factors and interindividual distances in microbiota composition, including energy (kilocalories); intake of carbohydrates, proteins, and fats; and intake of specific food items such as bread and soft drinks (Fig. 3 and table S4). Drinking buttermilk (sour milk with a low fat content) was associated with high diversity, whereas drinking high-fat (whole) milk (3.5% fat content) was associated with lower diversity (table S6). Two of the species most strongly associated with drinking buttermilk are *Leuconostoc mesenteroides* ( $q = 9.1 \times 10^{-46}$ ) and *Lactococcus lactis* ( $q = 2.5 \times 10^{-8}$ ), both used as a starter culture for industrial fermentation (table S11). The abundance of dairyfermentation–related bacteria increased with increasing dairy consumption, indicating potential for the use of probiotic drinks to augment and alter the gut microbiome composition. Consumption of alcohol-containing products, coffee, tea, and sugar-sweetened drinks was also correlated with microbial composition. Consumption of sugar-sweetened soda had a negative effect on microbial diversity (adjusted  $P = 5 \times 10^{-4}$ ), whereas consumption of coffee, tea, and red wine, which all have a high polyphenol content, was associated with increased diversity (19–21). Red wine consumption correlated with *Faecalibacterium prausnitzii* abundance, which has antiinflammatory properties, correlates negatively with inflammatory bowel disease (22), and shows higher abundance in high-richness microbiota (23). Apart from the negative associations between sugar-sweetened soda and bacterial diversity, other features of a Western-style diet, such as higher intake of total energy, snacking, and highfat (whole) milk, were also associated with lower microbiota diversity (Fig. 3). A higher amount of



**Fig. 3. Factors associated with interindividual variation of gut microbiome.** A total of 126 factors (FDR<0.1) were associated with interindividual variation of the gut microbiome. The bar plot indicates the explained variation of each factor in the interindividual variation of microbial composition [Bray-Curtis (BC) distance]. The heatmap next to the bar plot shows the correlation coefficients of each factor with Shannon's index of diversity, gene richness, and COG richness, respectively. Color key for correlation is shown.

carbohydrates in the diet was associated with lower microbiome diversity. Total carbohydrate intake was positively associated with Bifidobacteria but negatively associated with *Lactobacillus*, *Streptococcus*, and *Roseburia* species. A lowcarbohydrate diet consistently showed opposite directions of association for these species. We did not observe an association of carbohydrate intake to *prevotella* species, as has been described previously (24).

As expected, the use of antibiotics was significantly associated with microbiome composition, in particular with strong and significant decreases in two species from the genus *Bifidobacterium* (*Actinobacteria* phylum) (table S11), in line with previous studies (25). Several other drug categories, such as proton-pump inhibitors (PPIs) (95 users), metformin (15 users), statins (56 users), and laxatives (21 users), also had a strong effect on the gut microbiome. PPI users were found to have profound changes in 33 bacterial pathways (table S12). The most significant positive correlation of PPIs was observed with the pathway of 2,3but anediol biosynthesis ( $q = 5.3 \times 10^{-14}$ ). We also observed overlap between species and pathways associated to PPI and with calprotectin levels, particularly for bacteria typical of the oral microbiome (table S2, A to C; table S11; and fig. S12). This is in line with the correlations of PPIs with calprotectin levels reported in the literature (26). Even after excluding the 95 PPI users from our analysis, the positive correlation of calprotectin to most oral





bacteria remained significant, indicating that this association is not due to the confounding effect of PPIs (fig. S12). Furthermore, the amounts of calprotectin were positively correlated with age and metabolic phenotypes [body mass index (BMI), diabetes, use of statins and metformin, glycated hemoglobin (HbAIc), and systolic blood pressure], but negatively correlated with the consumption of vegetables, plant proteins, chocolate, and breads. Multivariate analysis correcting for all factors revealed 14 species (table S13) and 114 bacterial metabolic pathways (table S14) exclusively associated with calprotectin, suggesting that calprotectin is robustly associated with the gut microbiome.

Metformin is commonly used to control blood sugar concentrations for treating type 2 diabetes, but can cause gastrointestinal intolerance (27). In 15 metformin users, we observed an increased abundance of Escherichia coli and a positive correlation with specific pathways, including the degradation and utilization of D-glucarate and Dgalactarate and pyruvate fermentation pathways. Previous studies in Caenorhabditis elegans indicated the specific drug-bacteria interaction of metformin and E. coli (28). Our results are in line with recent observations in humans (29) that suggest that metformin can affect the microbiome through short-chain fatty acid (SCFA) production. To confirm this observation, we profiled acetate, propionate, and butyrate in 24 type 2 diabetes patients in our cohort-9 nonmetformin users and 15 users (4)--and found that SCFA concentrations were consistently higher in metformin users, especially for propionate (Wilcoxon test, P = 0.035) (fig. S13).

We assessed the effect of current smoking status, smoking history, parental smoking, and maternal smoking during pregnancy on the gut microbiome. These parameters were associated with Bray-Curtis distance, albeit with very modest effect. We did not detect significant associations for individual species or at pathways. In this study, we included 39 self-reported diseases, for which participants had reported at least five cases. IBS was reported by 9.9% of participants (n = 112, table S1) and was associated with changes in the gut microbiome and a lower microbial diversity (adjusted P = 0.05) (table S6). Species from the Eggerthella and Coprobacillus genera were positively associated with medication and food allergies, respectively. Individuals who had suffered a heart attack (n = 10) in the past had a significantly lower abundance of Eubacterium eligens bacterium, even after correcting for all other factors ( $q = 4.6 \times 10^{-4}$ ).

Linking the deep-sequenced MGS data to various intrinsic and exogenous factors from the same individual not only allowed us to detect associations at species level, but also provided new insights into the interaction between the host, microbiota, and environmental factors, including diet. For instance, we have replicated and expanded our association of BMI and blood lipid concentrations with the gut microbiota based on 16S rRNA gene sequencing data (*30*) by showing associations with four specific species of the family Rikenellaceae. We previously associated this family with BMI and triglycerides in 16S rRNA data. In this study, we observed that a higher BMI was associated with

a lower abundance of two species from the family Rikenellaceae, Alistipes finegoldii, and Alistipes senegalensis, whereas blood lipids were associated with two other species, Alistipes shahii and Alistipes putredinis (table S11). Notably, these species were also associated to certain dietary factors and drugs. For instance, a high level of A. shahii, which was associated to low triglyceride (TG) levels, was linked to higher fruit intake (q = 0.00027). Individuals with a higher abundance of A. shahii had a higher number of different species in the gut (species richness) (Spearman r = 0.2, adjusted  $P = 3.96 \times 10^{-11}$ ), suggesting a beneficial effect on the microbial ecosystem (table S18). Correlations with the number of different species were also found for other bacteria, including Roseburia hominis, Coprococcus catus, and Barnesiella intestinihominis and unclassified species from genus Anaerotruncus that also showed correlation both with fruit, vegetable, and nut consumption and with intrinsic phenotypes like HDL, triglycerides, and quality of life. On the basis of these data, it would be interesting to explore the potential to modulate disease-associated species through medication or diet, although we still need to address the causality and underlying mechanism.

Our study revealed significant associations between the gut microbiome and various intrinsic, environmental, dietary and medication parameters, and disease phenotypes, with a high replication rate between MGS and 16S rRNA gene sequencing data from the same individuals. Moreover, our study provides many new intrinsic and exogenous factors that correlate with shifts in the microbiome composition and functionality that potentially can be manipulated to improve microbiome-related health, and we hope our results will inspire further experiments to explore the biological relevance of associated factors. Although most of the factors that we assessed exerted a very modest effect, fecal levels of CgA showed a high potential as a biomarker for gut health.

#### **REFERENCES AND NOTES**

- 1. J. C. Clemente, L. K. Ursell, L. W. Parfrey, R. Knight, Cell 148, 1258-1270 (2012).
- E. F. Tigchelaar et al., BMJ Open 5, e006772 (2015). 2.
- S. Scholtens et al., Int. J. Epidemiol. 44, 1172-1180 (2015). 3 Information on materials and methods is available at Science 4. Online
- 5 N. Segata et al., Nat. Methods 9, 811-814 (2012).
- R. Sinha, C. C. Abnet, O. White, R. Knight, C. Huttenhower, 6. Genome Biol. 16, 276 (2015).
- J. K. Goodrich et al., Cell 159, 789-799 (2014). 7
- 8. G. Falony et al., Science 352, 560-564 (2016).
- 9 C. Huttenhower et al., Nature 486, 207-214 (2012).
- 10. F. Hildebrand et al., Genome Biol. 14, R4 (2013).
- 11. C. Hedin et al., Gut 10.1136/gutjnl-2014-308896 (2015).
- 12. T. L. Miller, M. J. Wolin, E. Conway de Macario, A. J. Macario, Appl. Environ. Microbiol. 43, 227-232 (1982).
- 13. T. Lee et al., Acta Paediatr. 95, 935-939 (2006)
- 14. L. Öhman, M. Stridsberg, S. Isaksson, P. Jerlstad, M. Simrén, Am. J. Gastroenterol. 107, 440-447 (2012).
- 15. V. Sciola et al., Inflamm. Bowel Dis. 15, 867-871 (2009).
- 16. M. Wagner et al., Inflammation 36, 855-861 (2013).
- 17. R. Aslam et al., Curr. Med. Chem. 19, 4115-4123 (2012).
- 18. A. Bartolomucci et al., Endocr. Rev. 32, 755-797 (2011).
- 19. M. I. Oueipo-Ortuño et al., Am. J. Clin. Nutr. 95, 1323-1334 (2012) 20. A. Duda-Chodak, T. Tarko, P. Satora, P. Sroka, Eur. J. Nutr. 54, 325-341 (2015).
- 21. C. E. Mills et al., Br. J. Nutr. 113, 1220-1227 (2015)
- 22. H. Sokol et al., Proc. Natl. Acad. Sci. U.S.A. 105, 16731-16736 (2008).

- 23. E. Le Chatelier et al., Nature 500, 541-546 (2013).
- 24. G. D. Wu et al., Science 334, 105-108 (2011).
- 25. K. Korpela et al., Nat. Commun. 7, 10410 (2016).
- 26. A. Poullis, R. Foster, M. A. Mendall, D. Shreeve, K. Wiener, Eur. J. Gastroenterol. Hepatol. 15, 573-574 (2003)
- 27. J. H. Burton et al., J. Diabetes Sci. Technol. 9, 808-814 (2015).
- 28 E Cabreiro et al. Cell 153 228-239 (2013)
- 29. K. Forslund et al., Nature 528, 262-266 (2015).
- 30. J. Fu et al., Circ. Res. 117, 817-824 (2015).

#### ACKNOWLEDGMENTS

We thank the LifeLines-DEEP participants and the Groningen LifeLines staff for their collaboration. We thank J. Dekens. M. Platteel. and A. Maatman for management and technical support. We thank J. Senior and K. McIntvre for editing the manuscript. This project was funded by grants from the Top Institute Food and Nutrition, Wageningen, to C.W. (TiFN GH001); the Netherlands Organization for Scientific Research to J.F. (NWO-VIDI 864.13.013), L.F. (ZonMW-VIDI 917.14.374), and R.K.W. (ZonMW-VIDL016136308); and CardioVasculair Onderzoek Nederland to M.H.H. and A.Z. (CVON 2012-03). A.Z. holds a Rosalind Franklin Fellowship (University of Groningen), and M.C.C. holds a postdoctoral fellowship from the Fundación Alfonso Martín Escudero. This research received funding from the European Research Council (ERC) under the European Union's Seventh Framework Program: C.W. is supported by FP7/2007-2013)/ERC advanced Grant Agreement no. 2012-322698 M.G.N. is supported by an ERC Consolidator Grant (no. 310372). L.F. is supported by FP7/2007-2013, grant agreement 259867, and by an ERC Starting Grant, grant agreement 637640 (ImmRisk). J.R. and G.F. are supported by FP7 METACARDIS HEALTH-F4-2012-305312, VIB, FWO, IWT (Agency for Innovation by Science and Technology), the Rega institute for Medical

Research, and KU Leuven. S.V.-S. and M.J. are supported by postdoctoral fellowships from FWO. T.V., M.S., and R.J.X. are supported by NIH, JDRF, and CCFA. A.Z., C.W., and J.F. designed the study. A.Z., E.F.T., L.F., and C.W. initiated the cohort and collected cohort data. A.Z., E.F.T., Z.M., S.A.J., M.C.C., and D.G. generated data. A.Z., A.K., M.J.B., E.F.T., M.S., T.V., A.V.V., G.F., S.V.-S., J.W., F.L. P.D., M.A.S., C.H., R.J.X., and J.F. analyzed data. G.F. S.V.-S., J.W., E.B., M.J., R.K.W., E.J.M.F., M.G.N., D.G., D.J. L.F., Y.S.A., C.H., J.R., R.J.X., and M.H.H. participated in integral discussions. A.Z., A.K., M.J.B., R.J.X., C.W., and J.F. wrote the manuscript. The authors have no conflicts of interest to report. The raw sequence data for both MGS and 16S rRNA gene sequencing data sets, and age and gender information per sample are available from the European genome-phenome archive (https://www.ebi.ac.uk/ega/) at accession number EGAS00001001704. Other phenotypic data can be requested from the LifeLines cohort study (https://lifelines.nl/lifelines-research/access-to-lifelines) following the standard protocol for data access. The study was approved by the institutional review board of UMCG, ref.M12. 113965. D.J. has additional funding from EU FP7/ no. 305564 and EU FP7/ no. 305479. C.H. is on the Scientific Advisory Board for Seres Therapeutics. Y.S.A. is a director and co-owner of PolyOmica, which provides services in statistical (gen)omics.

#### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6285/565/suppl/DC1 Materials and Methods Figs. S1 to S13 Tables S1 to S19 References (31-54)

1 September 2015: accepted 11 March 2016 10.1126/science.aad3369

#### **ORGANIC CHEMISTRY**

# Kinetically controlled *E*-selective catalytic olefin metathesis

Thach T. Nguyen,<sup>1</sup> Ming Joo Koh,<sup>1</sup> Xiao Shen,<sup>1</sup> Filippo Romiti,<sup>1</sup> Richard R. Schrock,<sup>2</sup> Amir H. Hoveyda<sup>1\*</sup>

A major shortcoming in olefin metathesis, a chemical process that is central to research in several branches of chemistry, is the lack of efficient methods that kinetically favor E isomers in the product distribution. Here we show that kinetically E-selective cross-metathesis reactions may be designed to generate thermodynamically disfavored alkenyl chlorides and fluorides in high yield and with exceptional stereoselectivity. With 1.0 to 5.0 mole % of a molybdenum-based catalyst, which may be delivered in the form of air- and moisture-stable paraffin pellets, reactions typically proceed to completion within 4 hours at ambient temperature. Many isomerically pure E-alkenyl chlorides, applicable to catalytic cross-coupling transformations and found in biologically active entities, thus become easily and directly accessible. Similarly, E-alkenyl fluorides can be synthesized from simpler compounds or more complex molecules.

lefin metathesis is an enormously enabling chemical process for which well-defined catalysts were discovered nearly three decades ago (1, 2). Kinetically controlled Z-selective reactions were introduced in 2009 (3), but there are no corresponding transformations that are broadly applicable and through which E isomers can be synthesized in high yield.

\*Corresponding author. Email: amir.hoveyda@bc.edu

Although E-selective cross-metathesis (CM) reactions involving Ru catechothiolate complexes (4) were reported very recently in 2016, only the thermodynamically preferred E isomers of simple (unfunctionalized) 1,2-disubstituted aliphatic alkenes could be obtained in 3 to 31% yield (5). E alkenes are often lower in energy and thus generated preferentially; nonetheless, olefin metathesis strategies that furnish them are needed for several reasons: The energy gap between the geometric forms is often too small to ensure high selectivity; E olefin isomers are not always thermodynamically preferred; and, in many cases,

<sup>&</sup>lt;sup>1</sup>Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, MA 02467, USA. <sup>2</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.



### 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, Ramnik J. Xavier, Cisca Wijmenga and Jingyuan Fu (April 28, 2016)

*Science* **352** (6285), 565-569. [doi: 10.1126/science.aad3369]

Editor's Summary

#### "Normal" for the gut microbiota

For the benefit of future clinical studies, it is critical to establish what constitutes a "normal" gut microbiome, if it exists at all. Through fecal samples and questionnaires, Falony *et al.* and Zhernakhova *et al.* targeted general populations in Belgium and the Netherlands, respectively. Gut microbiota composition correlated with a range of factors including diet, use of medication, red blood cell counts, fecal chromogranin A, and stool consistency. The data give some hints for possible biomarkers of normal gut communities.

Science, this issue pp. 560 and 565

This copy is for your personal, non-commercial use only.

Article Tools	Visit the online version of this article to access the personalization and article tools: http://science.sciencemag.org/content/352/6285/565
Permissions	Obtain information about reproducing this article: http://www.sciencemag.org/about/permissions.dtl

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.