nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-017-0096-0

Stability of the human faecal microbiome in a cohort of adult men

Raaj S. Mehta[1,2], Galeb S. Abu-Ali[3,4], David A. Drew[1,2], Jason Lloyd-Price[3,4], Ayshwarya Subramanian[3,4], Paul Lochhead[1,2], Amit D. Joshi[1,2], Kerry L. Ivey[5,6], Hamed Khalili[1,2], Gordon T. Brown[1,2], Casey DuLong[3], Mingyang Song[1,2], Long H. Nguyen[1,2], Himel Mallick[3,4], Eric B. Rimm[5,7], Jacques Izard[8], Curtis Huttenhower[3,4*] and Andrew T. Chan[1,2,7*]

Characterizing the stability of the gut microbiome is important to exploit it as a therapeutic target and diagnostic biomarker. We metagenomically and metatranscriptomically sequenced the faecal microbiomes of 308 participants in the Health Professionals Follow-Up Study. Participants provided four stool samples—one pair collected 24–72 h apart and a second pair ~6 months later. Within-person taxonomic and functional variation was consistently lower than between-person variation over time. In contrast, metatranscriptomic profiles were comparably variable within and between subjects due to higher within-subject longitudinal variation. Metagenomic instability accounted for ~74% of corresponding metatranscriptomic instability. The rest was probably attributable to sources such as regulation. Among the pathways that were differentially regulated, most were consistently over- or under-transcribed at each time point. Together, these results suggest that a single measurement of the faecal microbiome can provide long-term information regarding organismal composition and functional potential, but repeated or short-term measures may be necessary for dynamic features identified by metatranscriptomics.

Understanding the temporal dynamics of the healthy adult gut microbiome is integral in leveraging these microbial communities to promote human health. Large-scale changes in microbial composition have been associated with host health overall[1–3], but inferring causality and developing personalized therapies will require large-scale prospective cohort studies. Furthermore, to exploit the faecal microbiome as a predictive biomarker or eventually as a diagnostic tool in clinical settings, it is critical to be able to discriminate between normal versus pathological variation over time[4].

Previous efforts have provided excellent characterization of the ecological stability of the adult faecal microbiome[4–13]. All measures of stability in microbial communities must be in the context of relative differences since, despite daily variability in species' relative abundances, microbial communities in the gut microbiome have been observed to be generally consistent over time, even on the scale of years or decades[4,5,14,15]. This relative stability appears to be due to individually persistent strains within individual hosts[5,16]. Moreover, specific inter-individual differences in community structures appear to be preserved over the long-term[17], allowing an individual's faecal microbiome to be uniquely distinguished from that of others to serve as a faecal microbial fingerprint[13]. Nonetheless, despite the relative stability of the community profile over the long-term, recent studies have shown that host lifestyle or exposures such as a sudden change in diet, the initiation of antibiotics or the acquisition of pathogenic species can lead to profound disruptions in the microbiome[9,18,19]. When such pressures are lifted, the host's faecal microbiome generally recovers to a composition comparable to its original state[9].

An understanding of microbiome stability as it affects taxonomic and functional features is vital for applying it diagnostically or prognostically in long-term public health studies. Different molecular features, including strain membership, species abundances, functional profiles or metatranscription, may all prove to be informative regarding host health conditions and they are likely to differ dramatically in their relative stability within and between subjects over time[20]. The Human Microbiome Project (HMP), for example, found that in the absence of perturbations from disease or overt xenobiotics, metagenomic functional profiles were more comparable between individuals, while strains were stable within subjects[17]. Even fewer studies have focused on the metatranscriptome, an indicator of different aspects of microbial functional activity[21,22]. Finally, the dynamics of response in any of these features to known dietary and xenobiotic perturbations are themselves not yet fully known[23,24]. Thus, the magnitude of changes in microbial composition, functional potential and gene expression that occur over various time intervals and their utility for molecular epidemiology are unclear.

To address these knowledge gaps, we deeply characterized the faecal microbiome among 308 individuals enroled in the Men's Lifestyle Validation Study (MLVS), nested within the ongoing population-based Health Professionals Follow-up Study (HPFS)—a prospective cohort of 51,529 men followed since 1986. This cohort provided an unparalleled opportunity to apply insights from community ecology and epidemiology to characterize the stability of the faecal microbiome structure and function over time. Specifically, we hypothesized that inter-individual differences in faecal microbiome

[1]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [2]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [3]Biostatistics Department, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [4]The Broad Institute, Cambridge, MA, USA. [5]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [6]South Australian Health and Medical Research Institute, Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia. [7]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA. [8]University of Nebraska, Lincoln, NE, USA. Raaj S. Mehta, Galeb S. Abu-Ali, Curtis Huttenhower and Andrew T. Chan contributed equally to this work. *e-mail: chuttenh@hsph.harvard.edu; achan@mgh.harvard.edu

communities and metagenomes would persist over short- (24–72 h) and intermediate-term (6 months) intervals, but that metatranscriptomes would demonstrate greater variability over time. The work described here characterizing faecal microbial stability is accompanied by further research on molecular function in the faecal metagenome and metatranscriptome population[25]. These studies will provide a foundation for leveraging the faecal microbiome as a biomarker for population-based cohort studies and provide insight into how metagenomics and metatranscriptomics relate to each other over time.

Results

The faecal metagenome is more stable than the metatranscriptome. We first quantified the longitudinal stability of metagenomic taxonomic profiles, functional profiles and metatranscriptomes in MLVS subjects. Our approach is detailed in Fig. 1. Participants provided up to four stool samples—a set of stool samples collected 24–72 h apart followed by collection of a second set approximately 6 months later. DNA was extracted from 929 samples collected from all 308 men. RNA was extracted and reverse-transcribed to complementary DNA from 378 samples from participants who provided a stool at both sampling time points and who did not report the use of antibiotics within the past year. All samples were then sequenced using the Illumina HiSeq platform (see Methods). Raw sequence data were filtered to remove low-quality and human host reads and features with low overall relative abundance. Metagenomic and metatranscriptomic reads were profiled for functional and taxonomic composition using HUMAnN2 (ref. [26]) and MetaPhlAn2 (ref. [27]), respectively.

We compared the within-person stability of the faecal metatranscriptome (functional elements from RNA) with the stability of taxonomic composition (species from DNA) and the metagenome (functional elements from DNA) over short (24–72 h) and intermediate (6 month) time intervals. This was assessed by calculating two β-diversity metrics—the Jaccard index (for membership) and Bray–Curtis (BC) dissimilarity (for abundance)—between an individual's samples collected over the study period. In accordance with our earlier pilot study[21], the metagenomic functional potential (DNA) was more stable than the metatranscriptome (RNA), which was in turn more stable than taxonomic profile abundances (species) (Fig. 2a).

In addition, we tested whether the stability of microbiome features was consistent over short- and intermediate-term intervals. Species, DNA and RNA features that remained stable within a subject over a 24–72 h period tended to be similarly stable over 6 months (Spearman's $r = 0.99$, two-tailed $P < 2.2 \times 10^{-16}$), suggesting the presence of a subset of features that can be consistently measured over time.

Between-person exceeds within-person metagenomic variation. As in previous studies of the faecal microbiome[17], individuals' microbial community structure and metagenomic functional potential were more self-similar (red boxes) over time than between subjects (blue boxes) (Fig. 2b). This remained consistent over short- (24–72 h) and intermediate-term (6 month) periods (within-person variation versus between-person variation for all four Mann–Whitney *U* tests: $P < 2.2 \times 10^{-16}$). In contrast, we found that within- and between-subject variation over short- and intermediate-term periods was comparable for metatranscriptomes (short term Mann–Whitney *U* test: $P = 0.25$; intermediate term Mann–Whitney *U* test: $P = 0.50$).

To additionally quantify the proportion of variation that can be attributed to between-subject variation for the purposes of prospective population-based cohort studies, we estimated intra-class

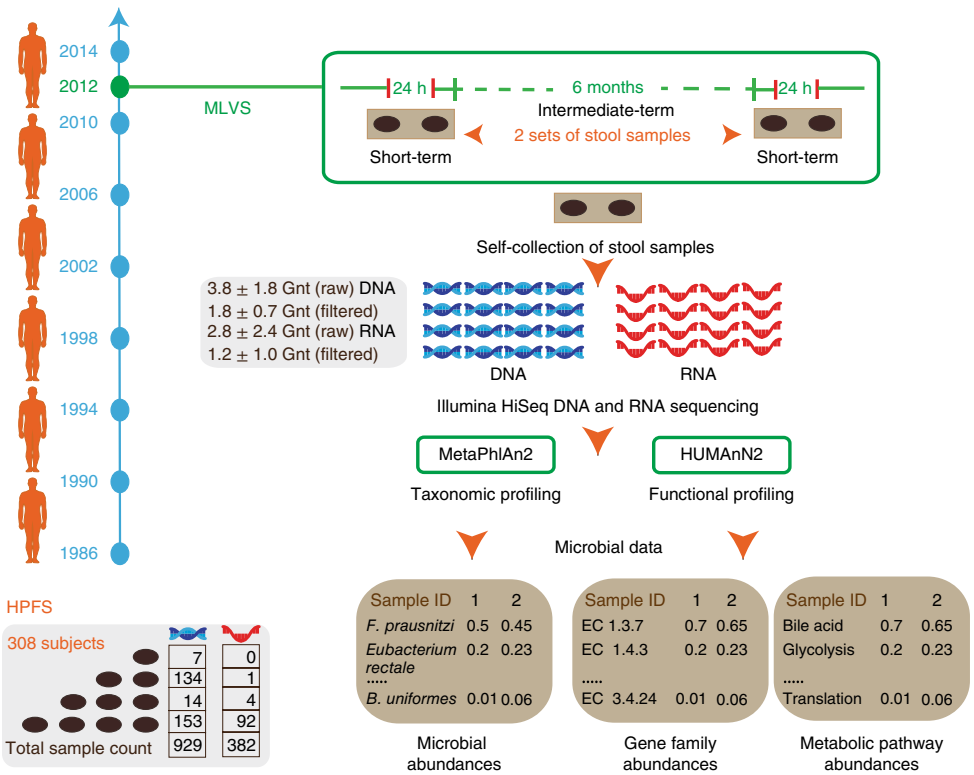

Fig. 1 | Experimental design. 308 participants from the MLVS, nested within the HPFS cohort, were recruited to assess the stability of microbiome communities, metagenomes and metatranscriptomes. Participants provided up to four stool samples using a previously validated self-sampling method[21]. One set of stool samples was collected 24–72 h apart followed by a second set approximately 6 months later. Metagenomic and metatranscriptomic reads were generated to provide taxonomic, functional metagenomic and transcriptional features for stability assessment.

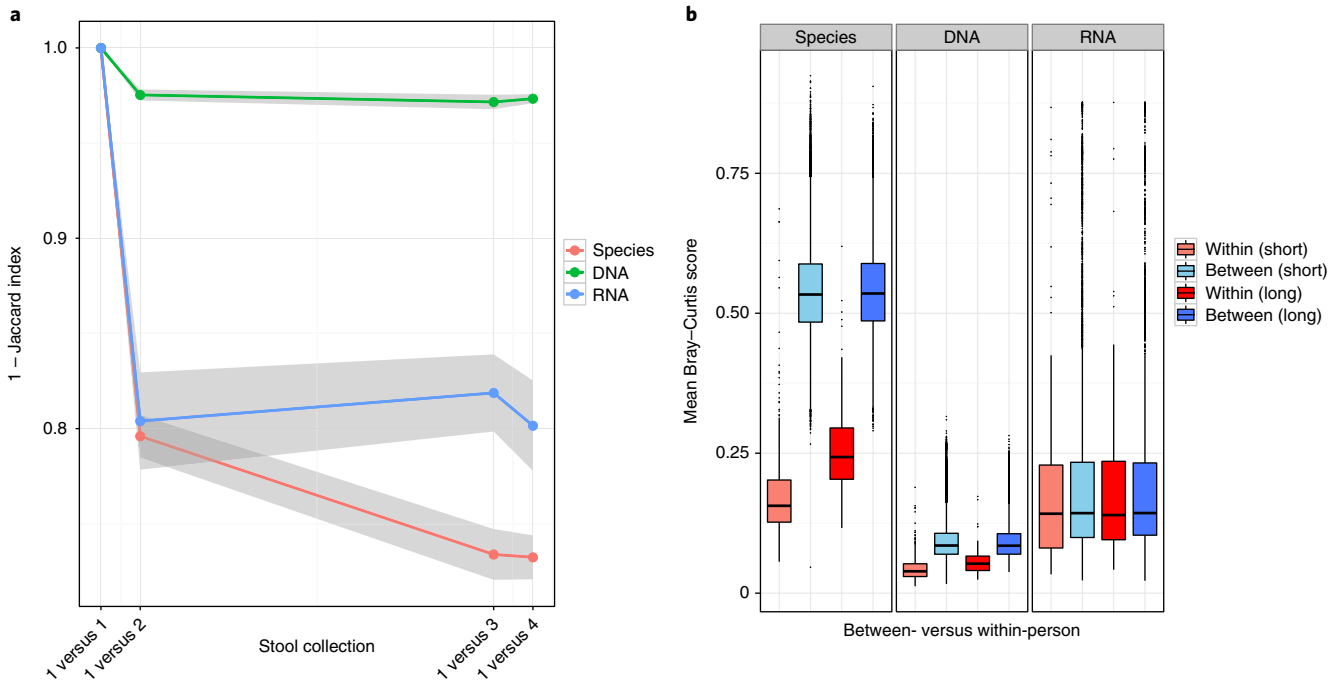

Fig. 2 | Inter-individual differences in organismal composition and functional potential appear to be preserved, unlike the more variable metatranscriptomes. a, 1 – Jaccard Index (fraction of shared features) between all possible pairwise combinations of the first faecal sample with the other three samples collected from each individual ($n = 308$). 95% confidence intervals are shown in grey. **b**, Bray–Curtis β-diversity scores within and between subjects for short- (24–72 h; $n = 308$ individuals) and intermediate-term intervals (6 months; $n = 160$ individuals). Here, species represents taxonomic profile abundances, DNA represents metagenomic functional profiles, and RNA represents metatranscriptomes. Boxplot whiskers include observations within 1.5 interquartile range of the upper and lower quartiles.

correlation coefficients (ICCs). The majority of species and genes (86.8 and 92.8%, respectively) retained ICCs greater than 0.40 over the longest sampling interval in our study. In contrast, only a small fraction of transcripts (0.79%) had an ICC greater than 0.40 (Supplementary Table 1).

Taken together, these data suggest that specific inter-individual differences in faecal microbiome organismal composition and functional potential appear to be preserved over intervals up to at least six months and thus may be more reliably associated with long-term health outcomes using a limited number of measurements. In contrast, the more dynamic features of the metatranscriptome may be more informative regarding short-term events.

Feature abundance and prevalence are associated with stability. Feature consistency over short- and intermediate-term periods led us to consider if there may be properties inherent to features that contribute to stability across all hosts. Across a range of body sites, overall taxonomic stability of the human microbiome is correlated with feature abundance and prevalence[13]. We extended these results by exploring the correlations between these two ecological properties and the stability of individual organisms in the faecal microbiome, as well as the stability of the genomic and transcriptomic levels of individual enzymes. As in previous analyses where dissimilarity metrics have been used for individual features[5], using the complement of within-person BC dissimilarity score (1 – BC) over the longest interval in our study (time point 1 versus time point 4; ~6 months), we similarly found that baseline feature prevalence and abundance were also strongly and positively correlated with not only organismal composition, but also genomic and transcriptional stability (Fig. 3). Many species that were highly abundant and prevalent in this population were highly stable, including the well-characterized *Bacteroides uniformis* (dark green circle) and *Faecalibacterium prausnitzii* (orange circle)[28]. In contrast, species that were of low prevalence and abundance were less stable, including those that were probably carried from the oral cavity including *Lactococcus lactis* (black circle) and *Streptococcus thermophilus* (dark purple circle)[21,29]. Similarly, enzymes such as RNA polymerase (red circle) (EC (Enzyme Commission) 2.7.7.6), which are essential for prokaryotic function, were universally prevalent, highly abundant and stably expressed at the genomic and transcriptional level.

Notably, however, there were select features for which the stability levels did not correspond with their prevalence or abundance. For example, *Sutterella wadsworthensis* (blue circle), was in the bottom quintile for prevalence and the bottom half for abundance, but it was in the top quintile for stability. This may be due to the role of this species' lone production of the adenosine 3′-phosphate-5′-phosphosulfate-independent sulfuryl transferase (EC 2.8.2.22)—an important microbial enzyme involved in the detoxification of phenolic compounds[30,31]. Similarly, *Bacteroides plebius* (light purple circle), found to occupy a unique role in starch utilization, was found in the bottom quintile for prevalence but in the top quintile for stability[32]. Additionally, as observed in the HMP, there were genes that were universally prevalent at the DNA level but of relatively low genomic abundance[17]. Many of these enzymes had some of the most variable gene expression. For example, the pair of enzymes thiaminase I and II (EC 2.5.1.2 and EC 3.5.99.2), which are involved in the microbial degradation of thiamin, possessed DNA sequences detected in 100% of participants, but were within the bottom 5% of stable transcripts[33]. Similarly, the enzyme ErmC (EC 2.1.1.184), which encodes one of the macrolide resistance genes was universally prevalent at the genomic level but was unstable at the RNA level and of intermediate stability at the genomic level[34].

Host factors that may influence faecal microbiome stability. In addition to examining determinants of stability for specific features,

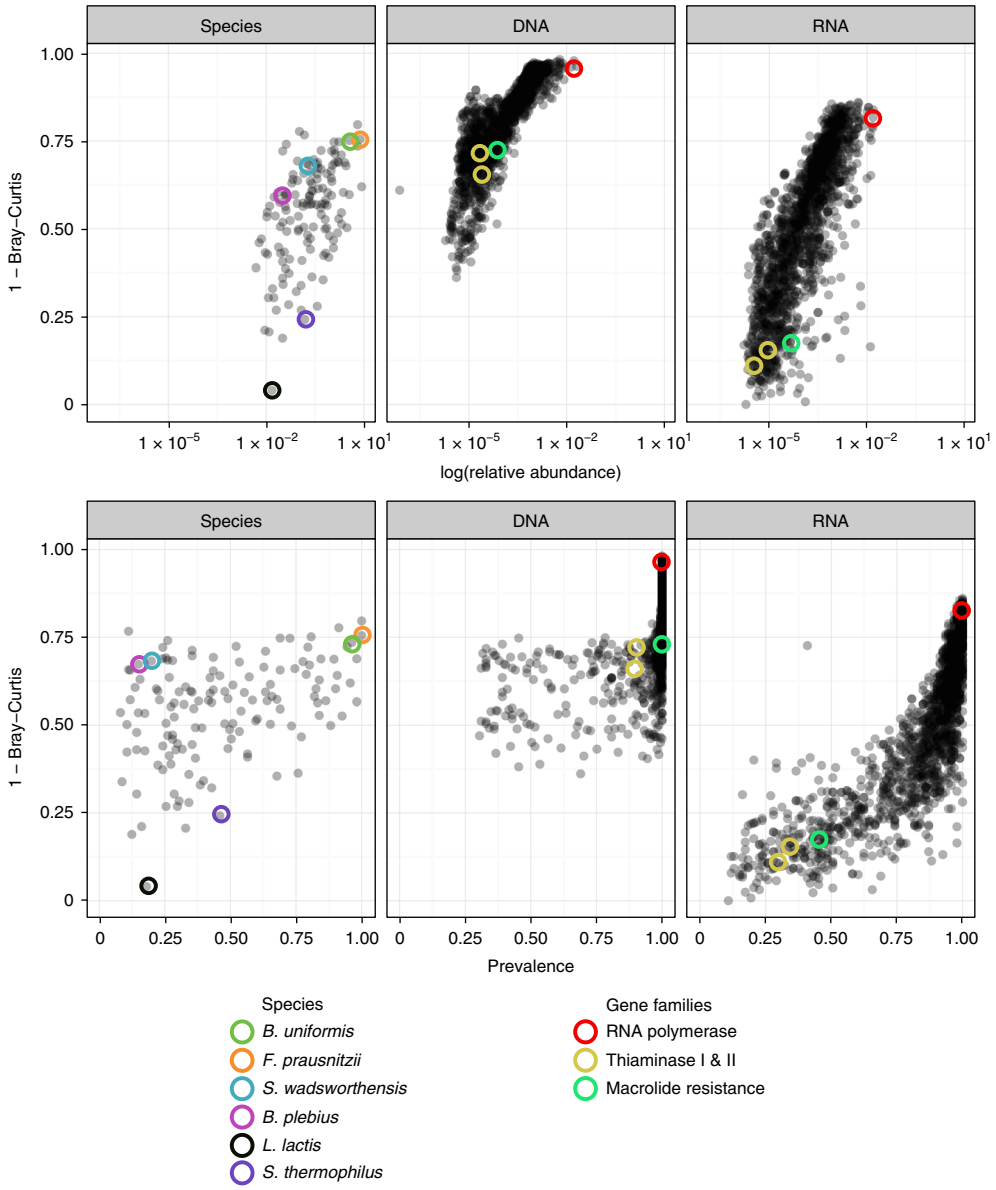

Fig. 3 | Stability of individual species, genes and transcripts over 6 months is correlated with average baseline relative abundance and prevalence. Each point represents an individual feature—species ($n=139$), genes ($n=1,952$) or transcripts ($n=1,803$). Coloured circles highlight species and gene families discussed in the main text, according to the figure legend.

we assessed whether there were lifestyle factors or behaviours at the time of stool collection that could explain global instability of the faecal microbiome for each individual. There have been inconsistent data on the effects of antibiotics and bowel preparation on the faecal microbiome composition and limited data on temporal change in the metagenome[19,24,35–38]. In addition, ecological theory suggests that species-rich communities may have a greater compensatory capacity for disturbances in the ecosystem[39,40].

In an exploratory analysis, we examined the association of putative host or environmental factors that may influence the stability of microbial community structure or pathways over a longer period. To do this, we calculated BC dissimilarity values for each individual over the longest sampling interval (around six months) using species, genes and transcripts as our response variables. As shown in Table 1, there were few factors that consistently explained instability. However, common predictive factors across feature types included mean Shannon index values (α-diversity) and exposure to a bowel laxative preparation within the two months before collection. Antibiotic use during the six-month period between sample collection seemed to affect genomic instability to a slightly greater extent than species instability. However, the overall low correspondence between antibiotic usage and dissimilarity suggests that faecal ecological variability may either be attributed to additional exogenous factors not measured here or to endogenous and perhaps stochastic host–microbial interactions.

Relating metagenomic and metatranscriptomic stability over time. Next, we considered the degree to which transcriptomic stability is related to metagenomic stability, since the central dogma suggests that DNA copy number changes are likely to influence transcript levels. Recent evidence from a pilot study involving a subset of these participants ($n=8$) found that many gene and transcript relative abundances were well correlated and that a substantial fraction (41%) of microbial transcripts were basally regulated[21]. Again using the complement of the within-person BC dissimilarity score over six months as our metric of stability, we found that genomic

Table 1 | Association of putative factors that may influence the stability of human faecal microbiome communities and pathways over a six-month period

	Species ($n=160$)		DNA ($n=160$)		RNA ($n=93$)	
Median dissimilarity over 6 months[a]	**Parameter estimate**	***P* value**[b]	**Parameter estimate**	***P* value**[b]	**Parameter estimate**	***P* value**[b]
Mean Shannon index (α-diversity)	**−0.096**	**0.006**	**−0.037**	**<0.0001**	−0.076	0.13
Acid-lowering medication use[c]	0.020	0.33	0.007	0.07	0.008	0.78
Bowel preparation[d]	**0.087**	**0.01**	**0.015**	**0.04**	**0.13**	**0.03**
Body mass index (kg m^{-2})	0.002	0.51	0.0007	0.15	0.0035	0.33
Age	−0.0004	0.84	0.0002	0.69	0.005	0.08
Any antibiotic use[e]	0.020	0.33	0.0004	0.92	NA	NA
New antibiotic use[e]	0.033	0.34	**0.022**	**0.002**	NA	NA
Bristol category change[f]	0.0004	0.56	−0.00001	0.52	0.0002	0.12

In an exploratory analysis, linear models were used to determine which of the measured factors predicted individuals' dissimilarity in faecal microbiome communities and faecal microbiome pathways between the first and last sample. All models were adjusted by the mean total number of reads that passed filtration per identification. *n* refers to the number of participants in each analysis. Statistically significant values appear in bold.[a]As estimated by the BC dissimilarity score between the first and fourth stool sample.[b]All models were adjusted for median read depth (pass-filter reads).[c]Use of proton-pump inhibitors or H_2 receptor antagonists more than once per week in the past two months.[d]Any reported bowel preparation (two months before collection).[e]The reference level was no antibiotic use. 'Any antibiotic use' was defined as any use in the past 12 months before the second stool collection. 'New antibiotic use' was defined as new use of antibiotics before the second collection. We did not assess transcriptomics in individuals who reported antibiotic use.[f]Bristol scores were assigned to three categories: 1–2 = hard; 3–4 = normal; 5–7 = soft or liquid.

instability appears to explain a large proportion of transcriptomic instability (Spearman's $r = 0.86$; $r^2 = 0.74 = 74\%$; $P < 2.2 \times 10^{-16}$). To further investigate whether this trend is driven by a small number of highly expressed genes, we divided our dataset into genes that were prevalently expressed (defined as transcripts detected in > 90% of samples; $n = 1{,}049$) and genes that were variably detected ($n = 754$). When comparing r^2 values for these two groups, we did not observe an appreciable difference (0.56 versus 0.48). However, we found that among only the transcripts in the bottom 10% of prevalence ($n = 176$), a similar strength in association was not seen ($r^2 = 0.01$), suggesting that measured expression of these genes is influenced by other factors.

In addition, we investigated whether dominant expression of stable versus unstable genes varied according to species. For each transcript, a dominant contributing species was identified based on the maximal average contributing organism for each transcript from HUMAnN2 (see Methods). We then ranked species according to the total number of genes that they dominantly expressed (Fig. 4). As shown, *Bacteroides* species are dominant contributors to the greatest number of genes expressed, as well as dominant contributors to the expression of the most stable transcripts. In contrast, *Streptococcus* species are dominant contributors to the expression of some of the most unstable transcripts. Interestingly, these are not typically viable gut residents[21], suggesting that some species play more variable roles because they are rare, transitory, more prone to influence by environmental factors or possibly under-sampled.

Stability of gene expression over time. The previous section highlighted that much of the transcriptional variability over time for each feature was associated with changes in the metagenome. Thus, next we queried what proportion of the remainder of transcriptional variation could be explained by differences in the transcriptional expression of features over time. In an approach similar to that performed in the pilot study[21], we identified significantly differentially transcribed pathways, defined as having a mean log RNA/DNA abundance ratio > 2 for all four time points and then explored where they overlapped. Of the 218 pathways that were significantly differentially expressed, 79% were consistently over- or under-expressed across three or four time points (Fig. 5a).

To better understand which pathways were most stably expressed, for each time point, we computed the mean RNA/DNA ratio (see 'RNA/DNA normalization' in the Methods of the accompanying manuscript by[48] for each pathway across the 85 participants who gave 4 samples for transcriptomic analysis. We then calculated the coefficient of variation for these means and ranked the pathways by their coefficient of variation (Fig. 5b). The top eight resulting most stable pathways in terms of transcriptional regulation appear to be involved in cellular housekeeping, such as carbon metabolism (stably over-expressed) and amino acid synthesis (stably under-expressed).

Discussion

Here, we quantify faecal microbiome stability for prospective population-based cohort studies, with a particular focus on the relationship between metatranscriptomic features and the metagenome (both organismal composition and functional potential). See also ref. [48] for a discussion of the molecular and functional activity. As expected from single organisms and human studies, metatranscriptomic variation exceeded that of the metagenome. In addition, unlike the metagenome, metatranscriptomic variation within individuals over time was higher, to the point that it was comparable to variation between individuals. Taken in combination with our accompanying report describing a relatively stable 'core' metatranscriptome[48], these findings, combined with previous reports[21,23,24], suggest a model in which gut microbial community function mirrors that of differentiated human cell types: a subset of transcripts are prevalently active in the gut from varied microbes and conditions, while the remainder respond dynamically to intrinsic or external factors, such as diet.

In addition, we found the striking result that approximately 74% of the instability in the faecal metatranscriptome can be explained by instability in the faecal metagenome. These findings are similar to those from our pilot study, which showed that at a single time point, more than half of the variation in microbial community gene expression can be explained by metagenomic composition. This is best illustrated by our finding that some of the most highly stable (as well as highly abundant and prevalent) species, such as *B. uniformis*, were responsible for the dominant expression of a large number of stable genes. In contrast, *Streptococcus salivarius*—an unstable species that is probably transiting from the oral microbiome[21]—dominantly expresses many of the most variable genes. Similar findings were found in relation to the regulation of gene expression. Of the genes we identified as significantly differentially transcribed, nearly 80% were consistently over- or under-expressed. Many of these stably expressed pathways appeared to be essential for cellular function or involved in glycolysis, the synthesis of amino acids or the production of short-

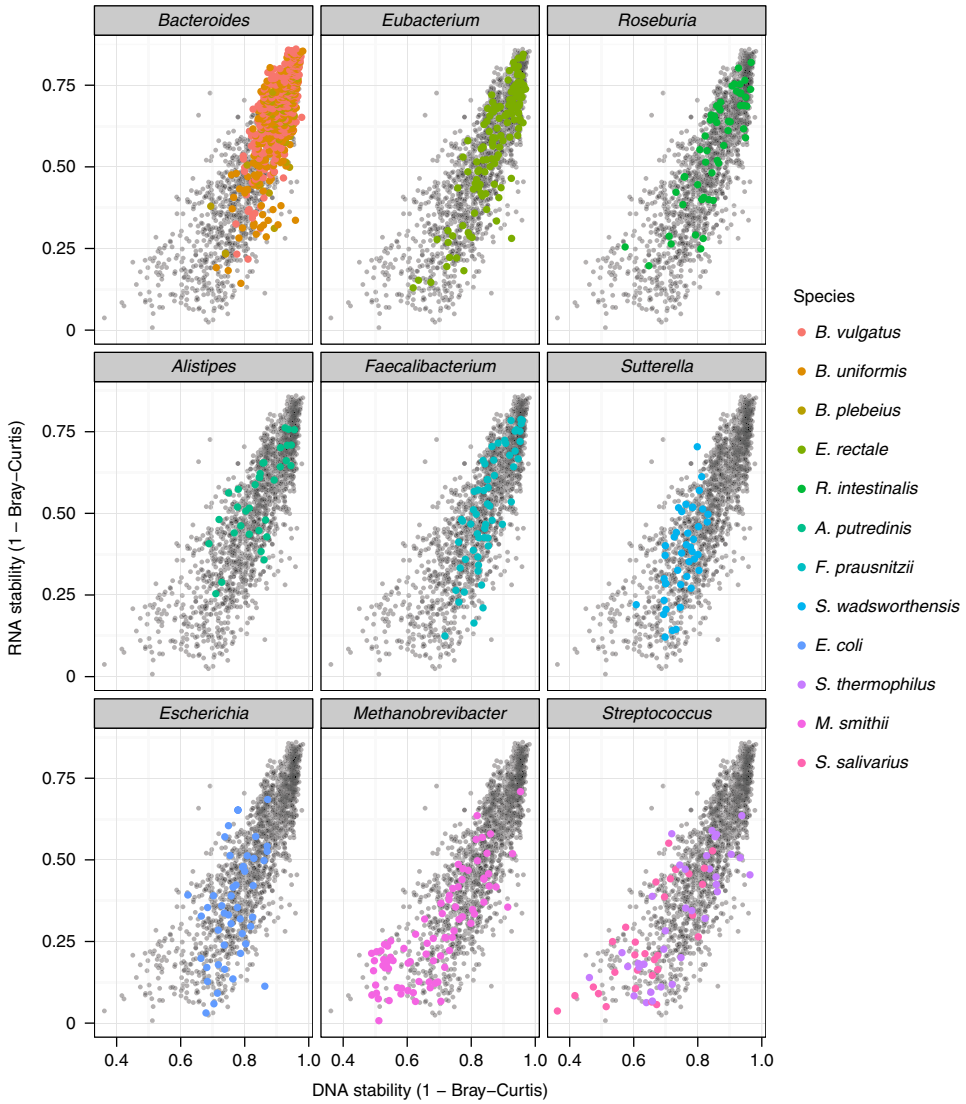

Fig. 4 | Relating metagenomic and metatranscriptomic stability over time. Genomic instability appears to explain a large proportion of transcriptomic instability over intermediate-term (6 month) periods ($r^2 = 0.74$); the metatranscriptome should typically be no more stable than the metagenome. In addition, we investigated whether dominant expression of stable genes versus expression of unstable genes varied according to species. For each transcript ($n = 1{,}803$), a dominant contributing species was identified based on the maximal average contribution to each transcript. We then ranked the species according to the total number of genes that they dominantly expressed and (for presentation purposes) selected those contributing to 30 or more and overlaid them on the scatterplot in colour (grouped by genus). Each point represents an individual feature.

chain fatty acids. Taken together, it appears that much of the metatranscriptomic variation in stools over time might reflect changes in gene family copy number, driven by factors such as the proliferation of microbes, introduction of new species or host-induced shifts in community structures. In contrast, in the face of relative metagenomic stability, substantial deviations from a common genomic profile may be associated with shifts in disease state[41,42].

Of course, we observed exceptions to these patterns, which may additionally explain some of the instability in the function of the faecal microbiome. Some species, such as *S. wadsworthensis* and *Methanobrevibacter smithii*, despite their stable presence, appear to be dominant contributors to the expression of unstable genes. Especially in the case of functionally unique taxa such as the archaea, this suggests specialized roles associated with targeted, highly variable activity over time. In addition, we found gene families with levels of RNA stability that were discordant with DNA stability, as well as several pathways that were differentially transcribed (up or down) only at a single time point. Given the large relative time scales of even our shortest gap (1–3 days) compared with transcriptional regulation (minutes), these changes are difficult to interpret without further data, but they may indicate pathways particularly susceptible to modulation by xenobiotics, diet or other targeted perturbations[23,24]. These outlying genes with high metagenomic stability but variable metatranscriptomic levels, or the stable microbes leading to the expression of unstable genes, are likely to play a critical role in our understanding of the impact of host behaviour on the faecal microbiome[48].

Finally, for the purposes of prospective, longitudinally followed cohort studies, our study is consistent with previous work and indicates that many long-term features of the faecal microbiome are well-represented in a single self-collected stool sample[4,5,17,21]. These include strain composition, general taxonomic abundances and core metagenomic functional profiles. Furthermore, our results suggest that, as expected, metatranscriptomes vary over time and are appropriate markers for short-term exposures, and not necessarily long-term exposure–disease relationships. They do, however,

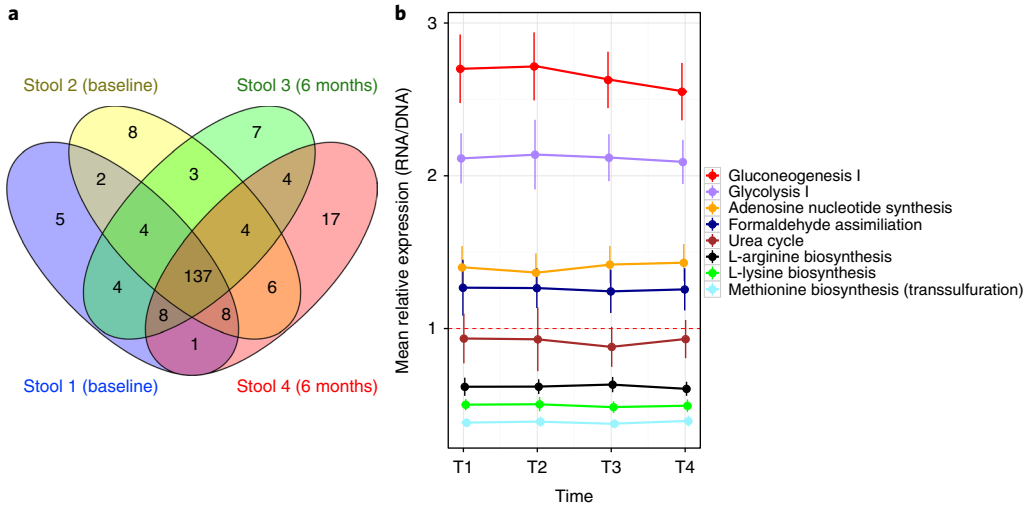

Fig. 5 | Exploring the stability of gene expression. a, For a large proportion of transcriptional pathways over- or under-expressed over time in the stool, this differential expression may be consistent. The Venn diagram shows consistency of differentially transcribed pathways (mean RNA/DNA ratios >2 across at least one time point; $n = 218$) across 4 stool samples. **b**, A subset ($n = 8$) of stably transcribed pathways is highlighted, according to those with the lowest coefficient of variance of mean RNA/DNA expression ratios across the four time points. Many of the stable pathways appear to be involved in cellular housekeeping, such as carbon metabolism (over-expressed; above dashed red line; >1) and amino acid synthesis (under-expressed; below dashed red line; <1). Error bars indicate 95% confidence intervals.

remain somewhat personalized, and subsets of microbial transcripts are individually more stable. It remains to be determined which of these microbiome features, if any, represent the best biomarkers for diagnosis or prognosis of health conditions in diverse human populations.

Moving forward, the large-scale, population-based collection of stools is critical to exploring the factors that promote a stable core functional microbiome, yet yield unique species and metagenomic profiles. Prospective cohort studies characterizing the microbiome in relation to lifestyle data, such as dietary and medication information, have already begun to explore inter-individual variation and will be essential in mechanistically understanding the interactions between the faecal microbiome and the host in the context of health and disease[17,43–45]. Continuing to improve our understanding of the stability of the faecal microbiome function is critical not only in determining the ecological dynamics of the human microbiome[25], but also for the promotion of human health.

Methods

Study population. The HPFS is an ongoing prospective cohort study that began in 1986 among 51,529 US male podiatrists, dentists, osteopathic physicians, veterinarians, pharmacists and optometrists aged 40 to 75 years at enrolment. In this study, participants returned questionnaires every 2–4 years with greater than 90% follow-up to provide information about lifestyle and dietary factors, medication use, and diagnoses of colorectal cancer and other diseases. The MLVS was established among 700 men aged 52–81 years (median 69 years) nested within HPFS who had completed the 2010 food frequency questionnaire and previously provided a blood sample. Men with coronary artery disease, stroke or transient ischaemic attack, cancer (except squamous or basal cell skin cancer) or major neurological disease (amyotrophic lateral sclerosis, Alzheimer's, Parkinson's, epilepsy or multiple sclerosis) were excluded. The 308 individuals sampled in this study were recruited into the MLVS starting in July 2012 and ending in July 2013. Participant recruitment and protocols were approved by the Harvard T. H. Chan School of Public Health Institutional Review Board #HSPH 22067-102. All participants provided informed consent for the study.

Sample size rationale. A tiered study design with metagenomic sequencing and a smaller batch of metatranscriptomic sequencing was used to determine both the composition and metabolic potential of the faecal microbiota. From the HMP findings of ~300 taxa per sample, our sample was estimated to yield relative abundances with standard deviation quartiles of (0.016, 0.027, 0.058) among rare taxa (present in < 50% of samples) and (0.081, 0.10, 0.15) among common taxa (arcsine square root-transformed relative abundance units).

Sample collection. Participants provided up to four stool samples—a set of samples from two consecutive bowel movements 24–72 h apart followed by collection of a second set of two such samples approximately 6 months later. Our collection protocol had previously been validated (see Supplementary Discussion)[21]. Briefly, participants deposited each bowel movement into a plastic commode collection bowl and then, using a specially designed spoon attached to a collection tube cap, faeces were scooped into a tube containing RNAlater. At each stool collection, participants documented the date and time of defecation as well as the Bristol score[46]. In addition, concurrent with each set of samples, participants completed a brief questionnaire collecting information regarding their recent use of antibiotics, gastric-acid-reducing medications, laxatives and probiotics, as well as other relevant exposures. At each time point, participants stored the specimen in the RNAlater fixative at ambient temperature until the specimen collected from the second bowel movement was produced. Each set of stool specimens was then placed in a special mailing kit and returned at ambient temperature by overnight mail. Upon receipt, the collection tubes were immediately placed into −80 °C freezers.

Sample handling and nucleic acid extraction. DNA and RNA extraction, processing and sequencing were as published[21]. Briefly, 100 mg stool aliquots were centrifuged at maximum speed to remove the excess of RNAlater. To the pellet, 110 µl of Tris-EDTA buffer with Proteinase K (Qiagen) and lysozyme (Sigma–Aldrich) (15 mg ml^{-1}) were added with incubation on a laboratory shaker for 10 min. The mechanical lysis was performed by the addition of a 1.2 ml RLT buffer with 2-mercaptoethanol (Qiagen) and 1 ml of 0.1 mm glass beads (BioSpec Products), followed by bead beating for 3 min. The debris were removed by centrifugation and the supernatant was used in Qiagen AllPrep spin columns according to the manufacturer's protocol (Qiagen). The only exception was in the preparation of the RNA: 60 µl DNase solution was added to buffer RW1 and incubated at room temperature for 15 min, followed by another wash with RW1 buffer. A NanoDrop 1000 (Thermo Fisher Scientific) was used to determine the DNA and RNA concentrations, quality and purity. SUPERase In was added to the aliquot before freezing. RNA was extracted and reverse-transcribed to complementary DNA only from stool samples spanning 6 months from participants who did not report the use of antibiotics within the past year (Fig. 1). RNA sequencing (RNA-Seq) libraries were depleted for ribosomal RNA using Ribo-Zero (Epicentre) as previously described[21].

DNA was extracted from all 929 samples. RNA was extracted from 378 samples derived from a subset of 96 individuals who provided all four stool samples spanning the 6-month study interval and who did not report the use of antibiotics within the past year.

Library prep, sequencing and taxonomic/functional profiling. We used the Nextera XT DNA Library Preparation Kit for whole-genome sequencing. For RNA-Seq, we used RNAtag-Seq—a method to create an RNA-Seq library containing large numbers of RNA samples that are bar-coded and pooled before library construction[47]. This approach does not require poly(A) capture/enrichment or random priming.

Using Illumina HiSeq paired-end (2×101 nucleotides) shotgun sequencing, 929 metagenomes and 378 metatranscriptomes were obtained. Six RNA samples were not sequenced due to platform constraints. After sequencing and before biocomputing quality control, the sequencing depth (mean ±s.d.) for DNA was 3.8 ± 1.6 giganucleotides (Gnt). After quality control, it was 1.8 ± 0.7 Gnt. For RNA, the mean sequencing depth was 2.8 ± 2.4 Gnt before quality control and 1.2 ± 1.0 Gnt after. The quality control step included the removal of human sequences, quality trimming and depletion of duplicate reads using HMP protocols (HMP) with KneadData (http://huttenhower.sph.harvard.edu/kneaddata).

We then performed taxonomic profiling with MetaPhlAn2 (ref. [27]) and functional profiling of genes and transcripts using HUMAnN2 (ref. [26]). Outlier identification was performed and eight samples were removed because their ordination scores in either dimension (NMDS1 or NMDS2) were outside the range of the median ± 3 interquartile ranges. Notably, four of these samples were provided by a participant who reported a history of colectomy (Supplementary Fig. 1). Six RNA samples were removed because they did not have matching DNA-sequenced samples. We subsequently filtered all taxonomic features with a relative abundance less than 10^{-4} (0.01%) in greater than 10% of all samples. Similarly, for DNA and RNA, we filtered all gene families with a relative abundance less than 10^{-5} (0.001%) in greater than 10% of all samples.

Statistical analysis. Variability in the relative abundance values for community composition, metagenomes and metatranscriptomes within and between samples was determined by calculating the BC dissimilarity metric for each individual over time[17]. Short-term intervals, as previously described, were defined as consecutive bowel movements (24–72 h apart), whereas the intermediate interval was six months—the longest sampling interval in the study. Similarly, the Jaccard index was calculated within samples across time, where comparisons were drawn between the first sample and the remaining three[5]. With these metrics, higher values signify more variable features, whereas lower values indicate more stable features.

To estimate the reliability or reproducibility in measuring relative abundance values of features over time, we calculated ICCs. Relative abundance values were first arcsine square root-transformed to variance-stabilize data and better approximate normality and then, using linear mixed-effects models with restricted maximum likelihood estimation, we divided the between-person variance by the sum of the within- and between-person variances[48]. In a previous study, in simulations on proportion data, this approach performed similarly to generalized linear mixed model-based estimates of ICCs using multiplicative models fitted by penalized quasi-likelihood estimation and from additive models fitted by Markov chain Monte Carlo sampling[48].

We explored whether several putative factors (bowel preparation use, antibiotic use, acid-lowering medication, and so on) were associated with within-person community, metagenome and metatranscriptome stability across the longest interval in our study using linear models. The Shannon index was calculated to be the mean of the within-sample diversity at time point 1 and time point 4. Acid-lowering medication use was determined by the use of proton-pump inhibitors or H_2 receptor antagonists more than once a week for two months. Bowel preparation was determined as any use of bowel preparation in the two months before stool collection. Antibiotic use was defined as the oral or intravenous administration of antibiotics in the 12 months before stool collection. Those who never used antibiotics were set as the reference level. 'Any antibiotics' users were defined as participants using antibiotics at any point during the 12 months before the second stool collection and 'new antibiotics' users were defined as non-users of antibiotics before collection time point 1 who had initiated the use of antibiotics after the first collection and before the second. The Bristol category change was determined by categorization of participant-reported Bristol stool scores into hard (1–2), normal (3–4) and soft or liquid (5–7) categories and then calculating the difference between the categories over time[46].

We examined associations between the baseline mean relative abundance and prevalence of each feature versus feature stability, and how metagenomic variability correlated with metatranscriptomic variability for each feature over time. As part of this, we investigated whether dominant expression of stable genes versus the expression of unstable genes varied according to species. For each transcript, a dominant contributing species was identified based on the maximal average contribution to each transcript. We then ranked the species according to the total number of genes that they dominantly expressed and (for presentation purposes) selected those contributing to 30 or more and overlaid them on a scatterplot in colour (grouped by genus). Examples of this process are presented in Figs. 3 and 4.

To test the stability of gene expression over time, we identified 740 pathways found in the 340 samples for which we had DNA and RNA data across all four time points. In total, 218 differentially transcribed pathways were defined as having a mean $\log_{10}$(RNA/DNA) significantly different from zero at each later time point using a linear model. We subjected these nominal *P* values to false discovery rate correction following the Benjamini–Hochberg method with $\alpha = 0.05$. We then explored the consistency of the differential regulation of each pathway by exploring the overlap at each time point.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. Sequence data have been deposited in the Sequence Read Archive under BioProject ID: PRJNA354235. Data from the Health Professionals Follow-up Study, including metadata not included in the current manuscript but collected as a part of the MLVS, can be obtained through written application. As per standard controlled access procedure, applications to use HPFS resources will be reviewed by our External Collaborators Committee for scientific aims, evaluation of the fit of the data for the proposed methodology and verification that the proposed use meets the guidelines of the Ethics and Governance Framework and the consent that was provided by the participants. Investigators wishing to use HFPS or MLVS cohort data are asked to submit a brief (two pages) description of the proposed project ('letter of intent') to E.B.R. (HPFS Director; erimm@hsph.harvard.edu).

Received: 27 March 2017; Accepted: 11 December 2017;
Published online: 15 January 2018

References

1. Hsiao, E. Y. et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* **155**, 1451–1463 (2013).
2. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).
3. Blumberg, R. & Powrie, F. Microbiota, disease, and back to health: a metastable journey. *Sci. Transl. Med.* **4**, 137rv7 (2012).
4. Caporaso, J. G. et al. Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
5. Faith, J. J. et al. The long-term stability of the human gut microbiota. *Science* **41**, 1237439 (2013).
6. Flores, G. E. et al. Temporal variability is a personalized feature of the human microbiome. *Genome Biol.* **15**, 531 (2014).
7. Ding, T. & Schloss, P. D. Dynamics and associations of microbial community types across the human body. *Nature* **509**, 357–360 (2014).
8. Jeffery, I. B., Lynch, D. B. & O'Toole, P. W. Composition and temporal stability of the gut microbiota in older persons. *ISME J.* **10**, 170–182 (2016).
9. David, L. A. et al. Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* **15**, R89 (2014).
10. Rajilić-Stojanović, M., Heilig, H. G. H. J., Tims, S., Zoetendal, E. G. & De Vos, W. M. Long-term monitoring of the human intestinal microbiota composition. *Environ. Microbiol.* **15**, 1146–1159 (2013).
11. Zoetendal, E. G., Akkermans, A. D. L. & De Vos, W. M. Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl. Environ. Microbiol.* **64**, 3854–3859 (1998).
12. Costello, E. K. et al. Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
13. Franzosa, E. A. et al. Identifying personal microbiomes using metagenomic codes. *Proc. Natl Acad. Sci. USA* **112**, E2930–E2938 (2015).
14. Dubos, R. J. & Schaedler, R. W. Reversible changes in the susceptibility of mice to bacterial infections. *J. Exp. Med.* **104**, 53–65 (1956).
15. Schaedler, R. W. & Dubos, R. J. Reversible changes in the susceptibility of mice to bacterial infections. *J. Exp. Med.* **104**, 67–84 (1956).
16. Schloissnig, S. et al. Genomic variation landscape of the human gut microbiome. *Nature* **93**, 45–50 (2013).
17. Consortium, T. H. M. P. Structure, function and diversity of the healthy human microbiome. *Nature* **86**, 207–214 (2012).
18. David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
19. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108**, 4554–4561 (2011).
20. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: networks, competition, and stability. *Science* **350**, 663–666 (2015).
21. Franzosa, E. A. et al. Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl Acad. Sci. USA* **111**, E2329–E2338 (2014).
22. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220–230 (2012).
23. McNulty, N. P. et al. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* **3**, 106ra106 (2011).
24. Maurice, C. F., Haiser, H. J. & Turnbaugh, P. J. Xenobiotics shape the physiology and gene expression of the active human gut microbiome. *Cell* **152**, 39–50 (2013).
25. Abu-Ali, G. S. et al. Metatranscriptome of human faecal microbial communities in a cohort of adult men *Nat. Microbiol.* https://doi.org/10.1038/s41561-017-0084-4 (2018).
26. Abubucker, S. et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).

27. Truong, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
28. Lopez-Siles, M., Duncan, S. H., Garcia-Gil, L. J. & Martinez-Medina, M. *Faecalibacterium prausnitzii*: from microbiology to diagnostics and prognostics. *ISME J.* **11**, 841–852 (2017).
29. Mahony, J., McDonnell, B., Casey, E. & van Sinderen, D. Phage–host interactions of cheese-making lactic acid bacteria. *Annu. Rev. Food Sci. Technol.* **7**, 267–285 (2016).
30. Kim, D. H., Konishi, L. & Kobashi, K. Purification, characterization and reaction mechanism of novel arylsulfotransferase obtained from an anaerobic bacterium of human intestine. *Biochim. Biophys. Acta* **872**, 33–41 (1986).
31. Malojčić, G. et al. A structural and biochemical basis for PAPS-independent sulfuryl transfer by aryl sulfotransferase from uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 19217–19222 (2008).
32. Hehemann, J.-H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
33. Jenkins, A. H., Schyns, G., Potot, S., Sun, G. & Begley, T. P. A new thiamin salvage pathway. *Nat. Chem. Biol.* **3**, 492–497 (2007).
34. Bussiere, D. E. et al. Crystal structure of ErmC', an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* **37**, 7103–7112 (1998).
35. Jalanka, J. et al. Effects of bowel cleansing on the intestinal microbiota. *Gut* **64**, 1562–1568 (2015).
36. O'Brien, C. L., Allison, G. E., Grimpen, F. & Pavli, P. Impact of colonoscopy bowel preparation on intestinal microbiota. *PLoS ONE* **8**, e62815 (2013).
37. Modi, S. R., Collins, J. J. & Relman, D. A. Antibiotics and the gut microbiota. *J. Clin. Invest.* **124**, 4212–4218 (2014).
38. Jakobsson, H. E. et al. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS ONE* **5**, e9836 (2010).
39. McCann, K. S. The diversity–stability debate. *Nature* **405**, 228–233 (2000).
40. Schindler, D. E. et al. Population diversity and the portfolio effect in an exploited species. *Nature* **465**, 609–612 (2010).
41. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
42. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Med.* **8**, 51 (2016).
43. Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
44. Falony, G. et al. Population-level analysis of gut microbiome variation. *Science* **352**, 560–564 (2016).
45. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
46. Lewis, S. J. & Heaton, K. W. Stool form scale as a useful guide to intestinal transit time. *Scand. J. Gastroenterol.* **32**, 920–924 (1997).
47. Shishkin, A. A. et al. Simultaneous generation of many RNA-Seq libraries in a single reaction. *Nat. Methods* **12**, 323–325 (2015).
48. Nakagawa, S. & Schielzeth, H. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biol. Rev.* **85**, 935–956 (2010).

Acknowledgements

We thank the participants who graciously participated in this research, K. Stewart and G. Gupta at the Massachusetts General Hospital (MGH) who assisted with recruitment for the study, and S. Sawyer (Brigham and Women's Hospital), M. Atar (MGH), C. Dulong (MGH and the Harvard T. H. Chan School of Public Health) and T. Poon (Broad Institut) for their assistance with project logistics, sample handling, nucleic acid extractions and sequencing. This work was supported by National Institutes of Health grants U54DE023798, UM1 CA167552, U01CA152904, R01 HL35464, R01CA202704 and K24DK098311, as well as by the Starr Cancer Consortium. A.T.C. was in part supported by the Stuart and Suzanne Steele MGH Research Scholars Program. J.I. was in part supported by the Nebraska Tobacco Settlement Biomedical Research Development Fund. R.S.M. was supported by a Howard Hughes Medical Institute Medical Research Fellowship and an AGA–Eli and Edythe Broad Student Research Fellowship.

Author contributions

J.I., A.T.C. and C.H. designed and managed the study. R.S.M., D.A.D., K.L.I., G.T.B., C.D., E.B.R. and J.I. collected the samples and generated the data. R.S.M., G.S.A.-A., D.A.D., J.L.-P., A.S., P.L., A.D.J., H.K., G.T.B., M.S., L.H.N. and H.M. analysed the data. R.S.M., G.S.A.-A., D.A.D., K.L.I., J.I., C.H. and A.T.C. prepared and wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-017-0096-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to C.H. or A.T.C.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

nature research

Corresponding author(s): Andrew T. Chan; Curtis Huttenhower

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

A tiered study design with metagenomic sequencing and a smaller batch of metatranscriptomic was used to determine both the composition and metabolic potential of the fecal microbiota. From the HMP1 findings of ~300 taxa/sample, our sample was estimated to yield relative abundances with standard deviation quartiles of [0.016, 0.027, 0.058] among rare taxa (present in <50% of samples) and [0.081, 0.10, 0.15] among common taxa (arc sin-sqrt relative abundance units).

2. Data exclusions

Describe any data exclusions.

Outlier identification was performed and eight samples were removed because their ordination scores in either dimension (NMDS1 or NMDS2) were outside the range of the median ± 3 IQRs. Notably, four of these samples were provided by a participant who reported a history of colectomy (Supplementary Fig. 1).

3. Replication

Describe whether the experimental findings were reliably reproduced.

There were no formal technical or biological replicates used, although participants provided up to 4 stool samples over the course of the study

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Not applicable. There were no experimental groups created.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Not applicable.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	MetaPhlAn2 (taxonomic profiling); HUMAnN2 (functional profiling); KneadData (quality control of sequence data); R (statistical software).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No cell lines used.
b. Describe the method of cell line authentication used.	No cell lines used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No cell lines used.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	N/a

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	Not applicable

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Our study was comprised of 308 healthy, mostly Caucasian health professionals (aged 65-81, median 69).