ARTICLE

doi:10.1038/nature11319

Gut microbiota composition correlates with diet and health in the elderly

Marcus J. Claesson[1,2]*, Ian B. Jeffery[1,2]*, Susana Conde[3], Susan E. Power[1], Eibhlís M. O'Connor[1,2], Siobhán Cusack[1], Hugh M. B. Harris[1], Mairead Coakley[4], Bhuvaneswari Lakshminarayanan[4], Orla O'Sullivan[4], Gerald F. Fitzgerald[1,2], Jennifer Deane[1], Michael O'Connor[5,6], Norma Harnedy[5,6], Kieran O'Connor[6,7,8], Denis O'Mahony[5,6,8], Douwe van Sinderen[1,2], Martina Wallace[9], Lorraine Brennan[9], Catherine Stanton[2,4], Julian R. Marchesi[10], Anthony P. Fitzgerald[3,11], Fergus Shanahan[2,12], Colin Hill[1,2], R. Paul Ross[2,4] & Paul W. O'Toole[1,2]

Alterations in intestinal microbiota composition are associated with several chronic conditions, including obesity and inflammatory diseases. The microbiota of older people displays greater inter-individual variation than that of younger adults. Here we show that the faecal microbiota composition from 178 elderly subjects formed groups, correlating with residence location in the community, day-hospital, rehabilitation or in long-term residential care. However, clustering of subjects by diet separated them by the same residence location and microbiota groupings. The separation of microbiota composition significantly correlated with measures of frailty, co-morbidity, nutritional status, markers of inflammation and with metabolites in faecal water. The individual microbiota of people in long-stay care was significantly less diverse than that of community dwellers. Loss of community-associated microbiota correlated with increased frailty. Collectively, the data support a relationship between diet, microbiota and health status, and indicate a role for diet-driven microbiota alterations in varying rates of health decline upon ageing.

The gut microbiota is required for development and for homeostasis in adult life. Compositional changes have been linked with inflammatory and metabolic disorders[1], including inflammatory bowel disease[2,3], irritable bowel syndrome[4,5] and obesity[6] in adults. The composition of the human intestinal microbiota is individual-specific at the level of operational taxonomic units (OTUs) and stable over time in healthy adults[7]. The composition of the intestinal microbiota in older people (>65 years) is extremely variable between individuals[8], and differs from the core microbiota and diversity levels of younger adults[8,9]. A feature of the ageing process is immunosenescence, evidenced by persistent NF-κB-mediated inflammation and loss of naive $CD4^+$ T cells[10]. The microbiota is pivotal for homeostasis in the intestine[11], and chronic activation of the innate and adaptive immune system is linked to immunosenescence[12]. Correlations have previously been made between specific components of the microbiota and pro-inflammatory cytokine levels, but these did not separate young adults from older people[9]. Alterations in the microbiota composition have also been associated with frailty[13], albeit in a small cohort from a single residence location.

Deterioration in dentition, salivary function, digestion and intestinal transit time[14] may affect the intestinal microbiota upon ageing. A controllable environmental factor is diet, which has been shown to influence microbiota composition in animal models, in small-scale human studies[15–20] and over the longer term[21]. However, links between diet, microbiota composition and health in large human cohorts are unclear. To test the hypothesis that variation in the intestinal microbiota of older subjects has an impact on immunosenescence and frailty across the community, we determined the faecal microbiota composition in 178 older people. We also collected dietary intake information, and measured a range of physiological, psychological and immunological parameters. Dietary groupings were associated with separations in the microbiota and health data sets; the healthiest people live in a community setting, eat differently and have a distinct microbiota from those in long-term residential care. Measures of increased inflammation and increased frailty support a diet–microbiota link to these indicators of accelerated ageing, and suggest how dietary adjustments could promote healthier ageing by modulating the gut microbiota.

Microbiota and residence location

We previously identified considerable inter-individual variability in the faecal microbiota composition of 161 older people (≥65 years), including 43 receiving antibiotics[8]. To investigate links between diet, environment, health and microbiota, we analysed 178 subjects, non-antibiotic-treated, for whom we also had dietary information, and stratified by community residence setting: (1) community-dwelling, $n = 83$; (2) attending an out-patient day hospital, $n = 20$; (3) in short-term (<6 weeks) rehabilitation hospital care, $n = 15$; (4) in long-term residential care (long-stay), $n = 60$. The mean subject age was 78 (± 8 s.d.) years, with a range of 64 to 102 years, and all were of Caucasian (Irish) ethnicity. We included 13 young adults with a mean age of 36 (± 6 s.d.) years. We generated 5.4 million sequence reads from 16S rRNA gene V4 amplicons, with an average of 28,099 (± 10,891 s.d.) reads per subject.

UniFrac β-diversity analysis indicates the extent of similarity between microbial communities[22]. UniFrac PCoA (principal co-ordinate) analysis of 47,563 OTUs (grouped at 97% sequence identity) indicated a clear separation between community-dwelling and long-stay subjects using both weighted and un-weighted analysis (Fig. 1a, b). Microbiota from the 13 younger controls clustered with

[1]Department of Microbiology, University College Cork, Ireland. [2]Alimentary Pharmabiotic Centre, University College Cork, Ireland. [3]Department of Statistics, University College Cork, Ireland. [4]Teagasc, Moorepark Food Research Centre, Moorepark, Fermoy, Co, Cork, Ireland. [5]Cork University Hospital, Wilton, Cork, Ireland. [6]St. Finbarr's Hospital, Douglas Road, Cork, Ireland. [7]Mercy University Hospital, Grenville Place, Cork, Ireland. [8]South Infirmary, Victoria University Hospital, Cork, Ireland. [9]Institute of Food and Health, University College Dublin, Ireland. [10]School of Biosciences, Cardiff University, Museum Avenue, Cardiff CF10 3AT, UK. [11]Department of Epidemiology and Public Health, University College Cork, Ireland. [12]Department of Medicine, University College Cork, Ireland.
*These authors contributed equally to this work.

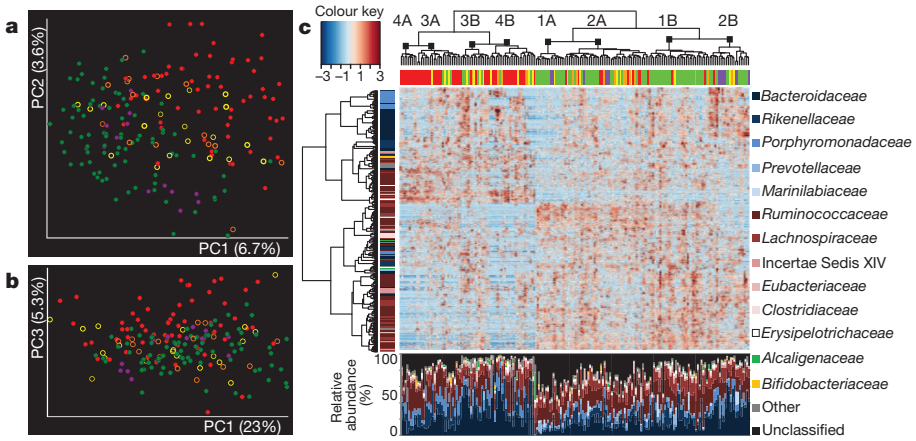

Figure 1 | Microbiota analysis separates elderly subjects based upon where they live in the community. **a**, Unweighted and **b**, weighted UniFrac PCoA of faecal microbiota from 191 subjects. Subject colour coding: green, community; yellow, day hospital; orange, rehabilitation; red, long-stay; and purple, young healthy control subjects. **c**, Hierarchical Ward-linkage clustering based on the Spearman correlation coefficients of the proportion of OTUs, filtered for OTU subject prevalence of at least 20%. Subjects colour coding as in **a**. Labelled clusters in top of panel **c** (basis for the eight groups in Fig. 4) are highlighted by black squares. OTUs are clustered by the vertical tree, colour-coded by family assignments. *Bacteroidetes* phylum, blue gradient; *Firmicutes*, red; *Proteobacteria*, green; and *Actinobacteria*, yellow. Only 774 OTUs confidently classified to family level are visualized. The bottom panel shows relative abundance of family-classified microbiota.

community-dwelling subjects. Eighteen other non-UniFrac β-diversity metrics supported microbiota separation by residence location (Supplementary Fig. 1).

When we examined OTU abundance, we identified a cluster comprised of the majority of the long-stay subjects, separated from the majority of the community-dwelling and young healthy subjects (Fig. 1c). Family-level microbiota assignments showed that long-stay microbiota had a higher proportion of phylum *Bacteroidetes*, compared to a higher proportion of phylum *Firmicutes* and unclassified reads in community-dwelling subjects (Fig. 1c). At genus level, *Coprococcus* and *Roseburia* (of the *Lachnospiraceae* family) were more abundant in the faecal microbiota of community-dwelling subjects (Supplementary Table 1 shows complete list of genera differentially abundant by community location). Genera associated with long-stay subjects included *Parabacteroides*, *Eubacterium*, *Anaerotruncus*, *Lactonifactor* and *Coprobacillus* (Supplementary Table 2). The genera associated with community belonged to fewer families, *Lachnospiraceae* were the most dominant. Thus, the microbiota composition of an individual segregated depending on where they lived within a single ethnogeographic region, in a homogeneous cohort where confounding effects of climate, culture, nationality and extreme environment were not a factor.

Concordance of diet and microbiota

Dietary data (for 168 of the 178 subjects, plus five percutaneous endoscopic gastrostomy (PEG)-fed subjects) was collected through a semiquantitative, 147-item, food frequency questionnaire (FFQ), weighted by 10 consumption frequencies. The data were visualized with correspondence analysis (CoA; Fig. 2a). The first CoA axis

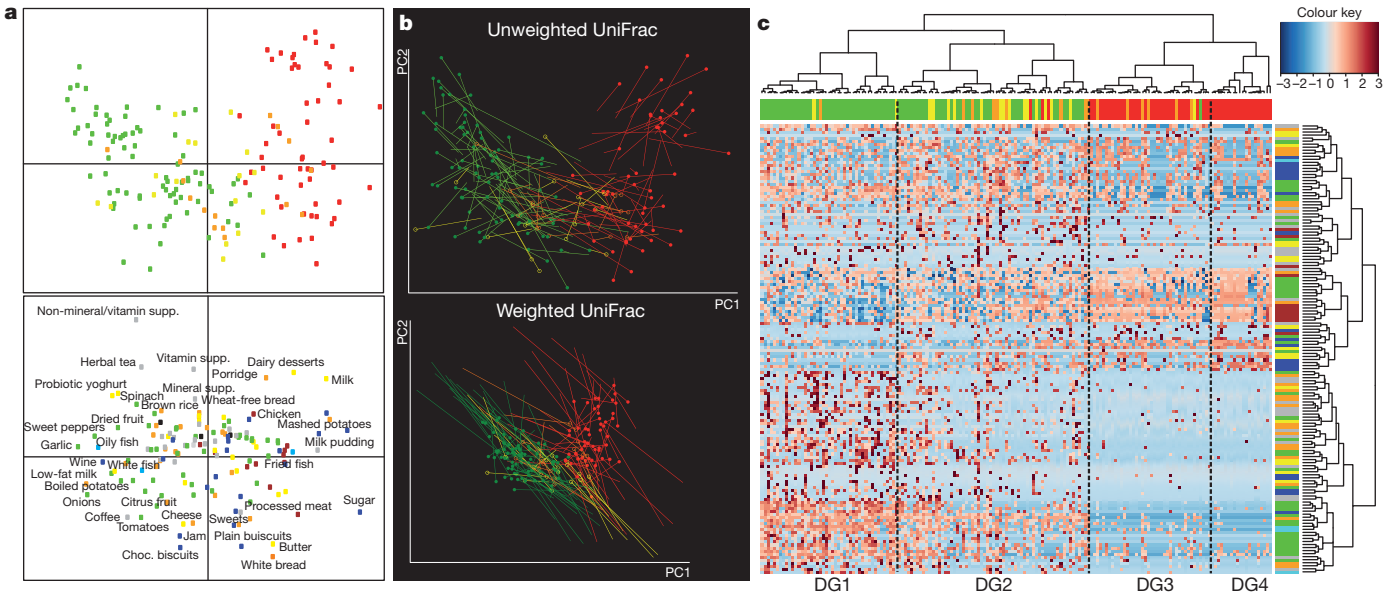

Figure 2 | Dietary patterns in community location correlate with separations based on microbiota composition. **a**, Food correspondence analysis. Top panel, FFQ PCA; bottom panel, driving food types. **b**, Procrustes analysis combining unweighted and weighted UniFrac PCoA of microbiota (non-circle end of lines) with food type PCA (circle-end of lines). **c**, Four dietary groups (DG1, DG2, DG3 and DG4) revealed through complete linkage clustering using Euclidean distances applied to first eigenvector in correspondence analysis. Colour codes in **a**, and horizontal clustering in **b** and **c**, are community location, as per Fig. 1. Food labelling in lower panel in **a**, and vertical clustering in **c**: green, fruit and vegetables; orange, grains such as potatoes, cereals and bread; brown, meat; cyan, fish; yellow, dairy products; blue, sweets, cakes and alcohol; grey vitamins, minerals and tea. Only peripheral and most driving foods are labelled; for a complete list see Supplementary Table 2.

described over 11% of the data set variance and most differences in food consumption between community-dwelling and long-stay subjects. The most discriminating food types were vegetables, fruit and meat, whose consumption changed in a gradual manner along the first eigenvector. Procrustes analysis of the FFQ and the microbiota β-diversity was used to co-visualize the data (Fig. 2b). Separations based on either diet or microbiota co-segregated along the first axis of both data sets (unweighted and weighted UniFrac, Fig. 2b; Monte-Carlo P value < 0.0001). Application of complete linkage clustering and Euclidean distances to the first eigenvector (Fig. 2c) revealed four dietary groups (DGs). DG1 ('low fat/high fibre') and DG2 ('moderate fat/high fibre') included 98% of the community and day hospital subjects, and DG3 ('moderate fat/low fibre') and DG4 ('high fat/low fibre') included 83% of the long-stay subjects. For a complete description of dietary groups, see Supplementary Notes and Supplementary Table 3.

The healthy food diversity index (HFD[23]) positively correlated with three microbiota diversity indices (Supplementary Fig. 2a), and all four indices showed significant differences between community and long-stay subjects (Supplementary Fig. 2b), indicating that a healthy, diverse diet promotes a more diverse gut microbiota. Analysing by dietary groups rather than residence location confirmed that both microbiota and diet were most diverse in DG1, and least diverse in DG3 and DG4 (Supplementary Fig. 3). Procrustes analysis similarly showed that the dietary groups were associated with separations in microbiota composition (Supplementary Fig. 3). Furthermore, the microbiota was associated with the duration in long-stay, with residents of more than a year having a microbiota that was furthest separated from community-dwelling subjects (Supplementary Fig. 4). For the majority of these longer-term residents, the diet was different from that in more recently admitted subjects (Supplementary Fig. 4). Examination of duration of care (Supplementary Fig. 4c) showed that diet changed more quickly than the microbiota did; both diet and microbiota moved in the direction away from the community types. After 1 month in long stay, all subjects had a long-stay diet, but it took a year for the microbiota to be clearly the long-stay type. Collectively the data indicate that the composition of the microbiota is determined by the composition and diversity of the diet.

Community setting and faecal metabolome

Faecal metabolites correlate with microbiota composition and inflammatory scores in Crohn's disease[24]. We therefore performed metabolomic analysis (NMR spectroscopy) of faecal water from 29 subjects, representative (by UniFrac) of three community settings. (Day-hospital subjects grouped closely to community dwellers by microbiota and dietary analysis, and were not included.) A representative NMR profile is presented in Supplementary Fig. 5. Initial PCA (principal component analysis) analysis showed a trend for separation according to community setting (data not shown). Pair-wise statistical models were therefore constructed according to the cluster groups. Valid and robust models were obtained for comparison of NMR spectra from community and long-stay subjects, and community and rehabilitation subjects (Fig. 3). The major metabolites separating community from long-stay subjects were glucose, glycine and lipids (higher levels in long-stay than community subjects), and glutarate and butyrate (higher levels in community subjects). Co-inertia analysis of the genus-level microbiota and metabolome data revealed a significant relationship (P value < 0.01) between the two data sets (Supplementary Fig. 6 and Supplementary Notes). Notwithstanding three long-stay subjects, a diagonal separated community from long-stay in both microbiota and metabolome data sets. Other metabolites of interest were acetate, propionate and valerate, which were more abundant in community dwellers (Supplementary Fig. 6).

To investigate microbial short-chain fatty acid (SCFA) production further, the frequency of microbial genes for SCFA production was investigated by shotgun metagenomic sequencing. We sequenced 125.9 gigabases (Gb) of bacterial DNA from 27 of the 29 subjects, and assembled contigs with a total length of 2.20 Gb, containing 2.51 million predicted genes (Supplementary Table 4). Consistent with reduced microbiota diversity (Supplementary Fig. 3), there were significantly fewer total genes predicted, and higher N50 values (N50 is the length of the smallest contig that contains the fewest (largest) contigs whose combined length represents at least 50% of the assembly), in the assembled metagenomic data of long-stay subjects compared to rehabilitation or community subjects (Supplementary Fig. 7). The metagenomes were then searched for key microbial genes in butyrate, acetate and propionate production, revealing significantly higher gene counts and coverage for butyrate- and acetate-producing enzymes (BCoAt and ACS, respectively) in community and rehabilitation compared to long-stay subjects (Supplementary Fig. 8 and Supplementary Table 5). There was also significantly higher coverage of the propionate-related genes (PCoAt) in community compared to long-stay subjects, but the higher gene count was not significant (Supplementary Table 5). These observations are consistent with the association of butyrate, acetate and propionate and the direction of the main split between long-stay and community subjects in the metabolome; candidate

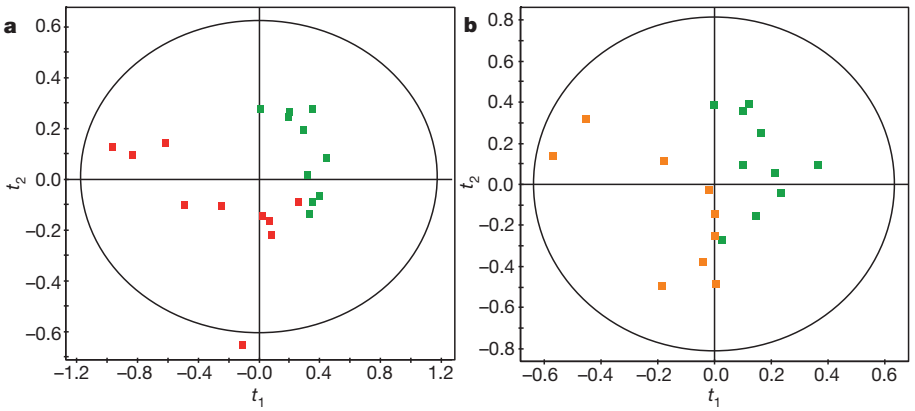

Figure 3 | PLS-DA plots of ^{1}H NMR spectra of faecal water from community, long-stay and rehabilitation subjects. **a**, Community subjects (green) versus long-stay subjects (red); $R^2 = 0.517$, $Q^2 = 0.409$, two-component model. **b**, Community subjects (green) versus rehabilitation subjects (orange); $R^2 = 0.427$, $Q^2 = 0.163$, two-component model. The ellipses represent the Hotellings T2 with 95% confidence. To confirm the validation of the model, permutation tests ($n = 1{,}000$) were performed. For model **a**, the 95% confidence interval for the misclassification error rate (MER) was (0.43, 0.57). Using the PLS-DA model on the data resulted in an MER of 0.2 which is outside the 95% confidence interval obtained for random permutation tests, thus validating the model. For model **b**, using permutation testing the 95% confidence interval for the MER was (0.45, 0.55). Using the PLS-DA model on the data resulted in an MER of 0.16 which is outside the 95% confidence interval obtained for random permutation tests.

genera associated include *Ruminococcus* and *Butyricicoccus* for butyrate production (Supplementary Fig. 6), but require validation in larger cohorts. Microbiota function deduced from the metagenome thus corresponded to the measured metabolome for at least one key metabolite that can affect health[25].

Microbiota–health correlations

Markers of inflammation (serum TNF-α, IL-6 and IL-8 and C-reactive protein (CRP)) had significantly higher levels in long-stay and rehabilitation subjects than in community dwellers (Supplementary Fig. 9). Long-stay subjects also scored poorly for diverse health parameters (Supplementary Tables 6 and 7), including the Charlson comorbidity index (CCI, a robust predictor of survival encompassing 19 medical conditions[26]), the geriatric depression test (GDT), the Barthel index[27], functional independence measure (FIM[28]), mini-mental state exam (MMSE[29]) and mini nutritional assessment (MNA[30]).

Correlations between health parameters and microbiota composition were examined using quantile (median) regression tests, adjusted for gender, age and community setting with an additive model (Supplementary Methods). Median regression gives less weight to extreme values than the linear regression based on ordinary least squares and consequently, is less influenced by outliers. The model was adjusted for medications that might influence the tested parameters (Supplementary Table 8). The effect of medication was generally small (Supplementary Table 8). Because ethnicity was exclusively Irish Caucasian it did not require model adjustment. The microbiota composition did not differ for males and females after adjusting for age and location.

Significant associations between several health/frailty measurements and the major separations from microbiota UniFrac analysis (Fig. 1) are shown in Table 1. For example, a positive change in microbiota along the full range of the PC1 axis in the un-weighted UniFrac PCoA for long-stay-only subjects was associated with inflammation (CRP increase of 13.9 mg l^{-1}), and other inflammatory markers significantly correlated with microbiota (IL-6 and IL-8, whole cohort). As expected, there was minimal variability amongst community-dwelling subjects, but within the long-stay subjects the most significant associations were related to functional independence (FIM), Barthel index and nutrition (MNA), followed by blood pressure and calf circumference. The latter may be attributable to the influence of diet and/or the microbiota on muscle mass, sarcopaenia[31] and thereby on frailty. This was supported by investigation of linkage between frailty and faecal metabolites (probabilistic principal components and covariates analysis; PPCCA[32]). Thus, the FIM and Barthel indices were significant covariates with the faecal water metabolome (Supplementary Fig. 10) and levels of acetate, butyrate and propionate increased with higher values of both indices (that is, less frail subjects). Among community-dwelling subjects, there was also a strong association between microbial composition and nutrition (MNA) and a weaker link with blood pressure, for which a relationship with the microbiota has previously been established[33]. There was no correlation between the *Bacteroidetes*:*Firmicutes* ratio and body mass index (BMI), although there was a correlation with overall microbiota in long-stay subjects. Measures for the geriatric depression test (GDT) showed significant microbiota association with PCoA axis 2 (Table 1). We detected no significant confounding of microbiota–health correlations due to medications, antibiotic treatment (before the 1-month exclusion window), and diet–health correlations separate from dietary impact on microbiota (Supplementary Notes).

Taken together, the major trends in the microbiota that separated healthy community subjects from less healthy long-stay subjects were

Table 1 | Regression tests of associations between clinical measurements and microbiota composition.

a Unweighted UniFrac PCoA for all four residence locations

Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*
GDT	–0.42	–0.11	0.6	**–2.7**	**–0.54**	**0.037**	0.18	0.04	0.84
Diastolic blood pressure	0.97	0.25	0.81	**–10.1**	**–2.02**	**0.033**	**–14.2**	**–3.1**	**0.001**
Weight	**–14.6**	**–3.8**	**0.033**	–7.16	–1.43	0.27	–1.57	–7.2	0.18
CC	**–3.9**	**–1.01**	**0.022**	–2.9	–0.58	0.19	**–3.2**	**–0.7**	**0.047**
IL-6	**6.71**	**1.7**	**0.006**	**6.1**	**1.22**	**0.007**	2.08	0.45	0.2
IL-8	4.23	1.1	0.43	**13.6**	**2.7**	**0.03**	4.06	0.89	0.47376716
TNF-α	1.1	0.28	0.31	0.62	0.13	0.72	**3.9**	**0.9**	**0.0005**

b Unweighted UniFrac PCoA for community-only subjects

Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*
MNA	–1.1	–0.26	0.29	**1.9**	**0.5**	**0.006**	0.7	0.14	0.59
Diastolic blood pressure	–8.4	–1.98	0.08	**14.3**	**3.4**	**0.035**	–15.72	–3.26	0.13
GDT	–0.13	–0.03	0.8	**–1.5**	**–0.35**	**0.02**	–0.8	–0.16	0.4

c Unweighted UniFrac PCoA for long-stay-only subjects

Parameter	PC1			PC2			PC3		
	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*	RC range	RC s.d.	*P*
Barthel	**–6**	**–1.5**	**0.004**	**–4.8**	**–1.3**	**0.036**	–0.6	–0.15	0.71
FIM	**–30.8**	**–7.8**	**0.046**	**–33.3**	**–4.7**	**0.024**	–2.42	–0.6	0.86
MMSE	–12.15	–3.08	0.14	**–18.4**	**–4.8**	**0.009**	3.22	0.8	0.63
MNA	–3.87	–0.98	0.23	**–11.2**	**–3**	**0.004**	–0.02	–0.005	0.99
BMI	–1.2	–0.31	0.69	**–5**	**–1.3**	**0.047**	–0.24	–0.06	0.92
CC	0.2	0.05	0.93	**–6.8**	**–1.77**	**0.0016**	0.45	0.11	0.82
Diastolic blood pressure	**19.3**	**4.9**	**0.015**	**–12.4**	**–3.24**	**0.034**	**–15.4**	**–3.81**	**0.007**
Systolic blood pressure	**36.5**	**9.3**	**0.007**	–1.57	–0.41	0.83	–2.05	–0.51	0.87
Weight	–3.2	–0.81	0.69	**–12.7**	**–3.3**	**0.024**	–2.48	–0.61	0.72
IL-8	–2.56	–0.65	0.78	**22.31**	**5.84**	**0.006**	1.14	0.28	0.93
CRP	**13.9**	**3.53**	**0.02**	–3.01	–0.8	0.27	–2.54	–0.63	0.61

Quantile (median) regression tests of associations between clinical measurements and microbiota composition as measured by unweighted UniFrac PCoA across all four residence locations (that is, all subjects (**a**), community-only subjects (**b**) and long-stay-only subjects (**c**)). Column headings are: RC range, regression coefficients scaled to the full variation along each PCoA axis, thus indicating relative magnitude and direction of the health association; RC s.d., regression coefficients scaled to one standard deviation; *P*, quantile regression *P* values generated by boot-strap analysis. Significant associations are in bold. An additive model was used to adjust for the effects of age, sex, residence location, relevant medication and the two other principal coordinates. CC, calf circumference; IL, interleukin; MMSE, mini-mental state examination.

associated with markers for increased frailty and poorer health, having adjusted for gender, age and location. Because location largely determines diet (Fig. 2), adjusting for location reduces the effect of diet, and as there was also clear evidence for microbiota–health associations within the long-stay setting, we infer that the causal relationship is in a diet–microbiota–health direction.

Microbiota structure and healthy ageing

Gut microbiota can be assigned to one of three enterotypes[34], driven by *Bacteroides*, *Prevotella* and *Ruminococcus* species. A recent study detected only the *Bacteroides* and *Prevotella* enterotypes, which were associated with diets rich in protein and carbohydrate, respectively[21]. Using those methods, we predicted an optimal number of two clusters using five out of six methodologies, albeit with weaker support than previous studies (Supplementary Fig. 11). In line with a previous study[21], the two clusters associated with *Bacteroides* and *Prevotella*, but not with *Ruminococcus*. Although enterotype assignments from the three approaches were very different (Supplementary Fig. 11), community subjects were more frequently of the *Prevotella* enterotype.

To identify patterns in the microbiota, we established co-abundance associations of genera (Supplementary Fig. 12a), and then clustered correlated genera into six co-abundance groups (CAGs) (Supplementary Fig. 12b). These are not alternatives to enterotypes, which are subject-driven and poorly supported in this elderly cohort, but they describe the microbiota structures found across the subject groups in statistically significant co-abundance groups (Supplementary Notes). The dominant genera in these CAGs were *Bacteroides*, *Prevotella*, *Ruminococcus*, *Oscillibacter*, *Alistipes* and the central *Odoribacter* CAG. These CAG relationships are termed Wiggum plots, in which genus abundance can be represented as discs proportional to abundance (Supplementary Fig. 12), to normalized over-abundance (Fig. 4), or to differential over-abundance (Supplementary Fig. 13). In the Wiggum plot corresponding to the whole cohort (Supplementary Fig. 12), the path away from the *Ruminococcus* CAG towards the *Oscillibacter* CAG shows a reduced number of genera that make butyrate, and an increased number able to metabolize fermentation products.

To simplify the microbiota data for health correlation, we used the eight subject divisions identified by OTU clustering (Fig. 1c). These eight divisions were superimposed on a UniFrac PCoA analysis of the data in Fig. 1a, defining 8 subject groups (Fig. 4, Groups 1A through 4B). These are separation points within a microbiota composition spectrum that represent groups of individuals who have significantly different microbiota as defined by the permutation multivariate analysis of variance (MANOVA) test on unweighted UniFrac data. We then constructed individual Wiggum plots for the microbiota in these 8 groups (Fig. 4). The transition from healthy community-dwelling subjects, to frail long-term care residents, is accompanied by distinctive CAG dominance, most significantly in abundances of *Prevotella* and *Ruminococcus* CAGs (community associated CAGs) and *Alistipes* and *Oscillibacter* CAGs (long-stay-associated CAGs).

Our analysis of Fig. 4 suggested two paths from community-associated health to long-stay-associated frailty (plot 1A–4A, and 1B–4B), which were examined with reference to health correlations

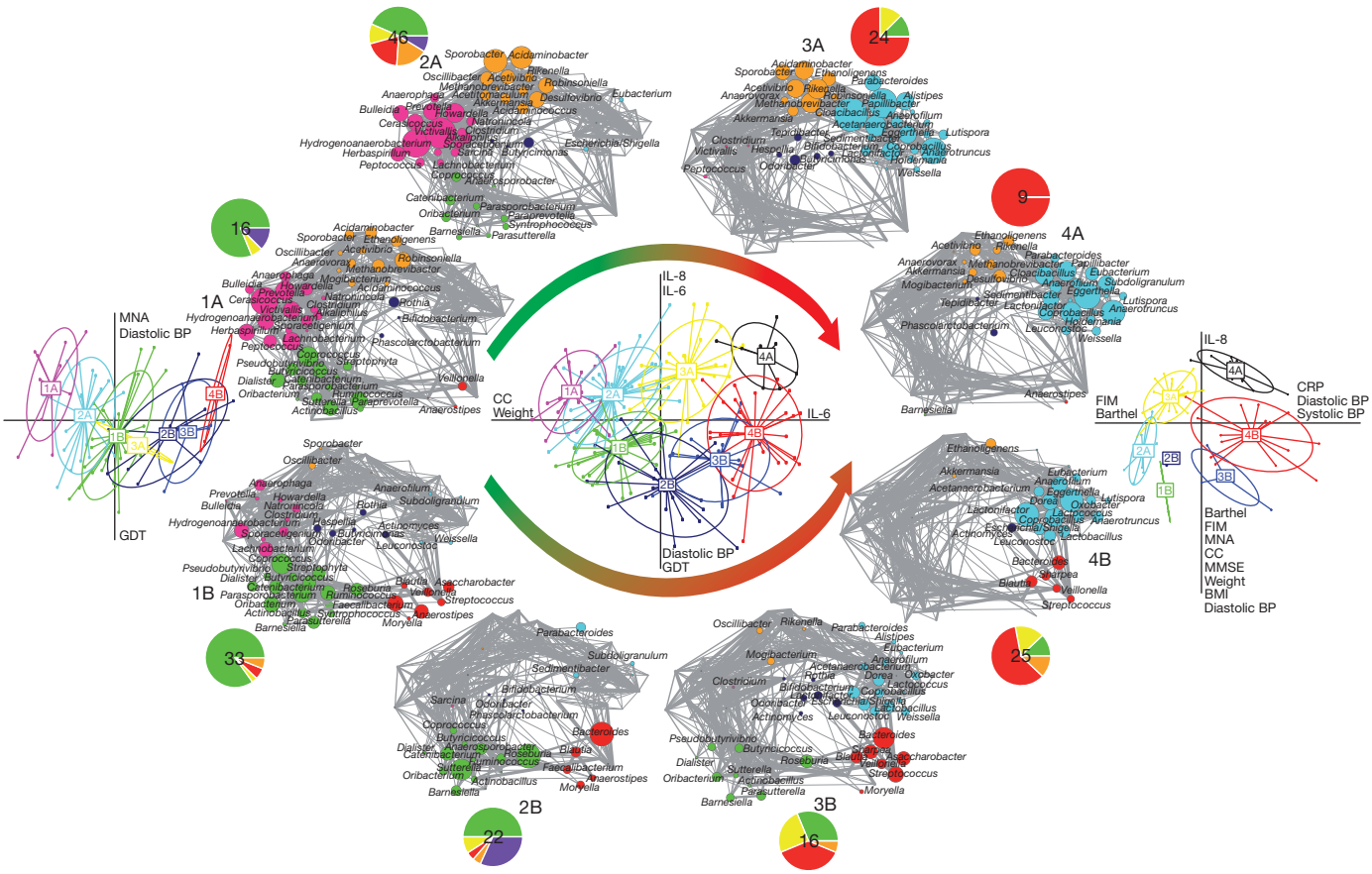

Figure 4 | Transition in microbiota composition across residence location is mirrored by changes in health indices. The PCoA plots show 8 groups of subjects defined by unweighted UniFrac microbiota analysis of community subjects (left), the whole cohort (centre), and long-stay subjects (right). The main circle shows the Wiggum plots corresponding to the 8 groups from whole-cohort analysis, in which disc sizes indicate genus over-abundance relative to background. The pie charts show residence location proportions (colour coded as in Fig. 1c) and number of subjects per subject group. Curved arrows indicate transition from health (green) to frailty (red). FIM, functional independence measure; MNA, mini nutritional assessment; GDT, geriatric depression test; CC, calf circumference; CRP, C-reactive protein; IL, interleukin; BP, blood pressure; MMSE, mini-mental state examination.

in Table 1, plus separate PCoAs for the community-only, and long-stay-only subjects. The community and whole-cohort analyses identified an association of depression with axis 2—subjects in the lower path had higher GDT scores. IL-6 and IL-8 levels were higher in the upper path by whole-cohort analysis (Fig. 4 and Supplementary Fig. 14), whereas CRP levels were higher in the lower path in long-stay-only analysis. Furthermore, subjects in the lower path had higher systolic and diastolic blood pressure, except in the community-only analysis. This apparent inconsistency is explained by a highly significant change in diastolic blood pressure along the primary PCoA axis in the long-stay subjects, emphasizing the value of a stratified cohort. The subjects in the upper path were older but had higher Barthel and FIM scores than subjects of a similar age in the lower path (Supplementary Fig. 14), consistent with healthier ageing. Movement along PCoA axis 1 of the whole cohort (that is, from community to long-stay, left to right, Fig. 4) is associated with a reduction in abundance of *Ruminococcus* and *Prevotella*, and increased abundance of the *Oscillibacter* CAG, accompanied by calf circumference decrease and weight decrease (Table 1), and increase in IL-6 levels. Moving along axis 1 of the long-stay PCA (that is, between the two right-ward arms), there is a reduction in the *Oscillibacter* CAG, increase in abundance of the *Bacteroides* CAG, reduced FIM and Barthel indices, and increased levels of CRP (Fig. 4). Consideration of the microbiota–health correlations in the long-stay cohort (Fig. 4), upwards along axis 2, highlights the association with increased frailty, reduced muscle mass, and poorer mental activity moving away from community-type microbiota.

Health–microbiota associations were statistically significant, even when regression models were adjusted for location. Although other factors undoubtedly contribute to health decline, and are difficult to completely adjust for in retrospective studies, the most plausible interpretation of our data is that diet shapes the microbiota, which then affects health in older people. Diet-determined differences in microbiota composition may have subtle impacts in young adults in developed countries. These would be difficult to correlate with health parameters, but become far more evident in the elderly who are immunophysiologically compromised. This is supported by the stronger microbiota–health associations evident in the long-stay cohort, and there is now a reasonable case for microbiota-related acceleration of ageing-related health deterioration. An ageing population is now a general feature of western countries[35,36] and an emerging phenomenon even among developing countries. The association of the intestinal microbiota of older people with inflammation[12] and the clear association between diet and microbiota outlined in this and previous studies[20,21,37,38] argue in favour of an approach of modulating the microbiota with dietary interventions designed to promote healthier ageing. Dietary supplements with defined food ingredients that promote particular components of the microbiota may prove useful for maintaining health in older people. On a community basis, microbiota profiling, potentially coupled with metabolomics, offers the potential for biomarker-based identification of individuals at risk for, or undergoing, less-healthy ageing.

METHODS SUMMARY

Amplicons of the 16S rRNA gene V4 region were sequenced on a 454 Genome Sequencer FLX Titanium platform. Sequencing reads were quality filtered, OTU clustered, ChimeraSlayer filtered and further analysed using the QIIME pipeline[39] and RDP-classifier[40]. Statistical analysis was performed using Stata and R software packages. Nuclear magnetic resonance (NMR) spectroscopy was performed on a 600 MHz Varian NMR Spectrometer as previously described[41].

Habitual dietary intake was assessed using a validated, semiquantitative, FFQ, administered by personnel who received standardized training in dietary assessment. FFQ coding, data cleaning and data checks were conducted by a single, trained individual to ensure consistency of data.

Full Methods and any associated references are available in the online version of the paper.

Received 10 January; accepted 14 June 2012.
Published online 13 July 2012.

1. O'Toole, P. W. & Claesson, M. J. Gut microbiota: changes throughout the lifespan from infancy to elderly. *Int. Dairy J.* **20,** 281–291 (2010).
2. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104,** 13780–13785 (2007).
3. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
4. Kassinen, A. *et al.* The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* **133,** 24–33 (2007).
5. Jeffery, I. B. *et al.* An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61,** 997–1006 (2012).
6. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444,** 1022–1023 (2006).
7. Rajilić-Stojanović, M. *et al.* Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. *Environ. Microbiol.* **11,** 1736–1751 (2009).
8. Claesson, M. J. *et al.* Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl Acad. Sci. USA* **108** (Suppl 1), 4586–4591 (2011).
9. Biagi, E. *et al.* Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5,** e10667 (2010).
10. Franceschi, C. *et al.* Inflamm-aging: an evolutionary perspective on immunosenescence. *Ann. NY Acad. Sci.* **908,** 244–254 (2000).
11. Garrett, W. S., Gordon, J. I. & Glimcher, L. H. Homeostasis and inflammation in the intestine. *Cell* **140,** 859–870 (2010).
12. Guigoz, Y., Dore, J. & Schiffrin, E. J. The inflammatory status of old age can be nurtured from the intestinal environment. *Curr. Opin. Clin. Nutr. Metab. Care* **11,** 13–20 (2008).
13. van Tongeren, S. P., Slaets, J. P., Harmsen, H. J. & Welling, G. W. Fecal microbiota composition and frailty. *Appl. Environ. Microbiol.* **71,** 6438–6442 (2005).
14. Lovat, L. B. Age related changes in gut physiology and nutritional status. *Gut* **38,** 306–309 (1996).
15. Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137,** 1716–1724 (2009).
16. Mai, V., McCrary, Q. M., Sinha, R. & Glei, M. Associations between dietary habits and body mass index with gut microbiota composition and fecal water genotoxicity: an observational study in African American and Caucasian American volunteers. *Nutr. J.* **8,** 49 (2009).
17. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332,** 970–974 (2011).
18. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107,** 14691–14696 (2010).
19. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. **1,** 6ra14 (2009).
20. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* **333,** 101–104 (2011).
21. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334,** 105–108 (2011).
22. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71,** 8228–8235 (2005).
23. Drescher, L. S., Thiele, S. & Mensink, G. B. A new index to measure healthy food diversity better reflects a healthy diet than traditional measures. *J. Nutr.* **137,** 647–651 (2007).
24. Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **4,** e6386 (2009).
25. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS Microbiol. Lett.* **217,** 133–139 (2002).
26. de Groot, V., Beckerman, H., Lankhorst, G. J. & Bouter, L. M. How to measure comorbidity. a critical review of available methods. *J. Clin. Epidemiol.* **56,** 221–229 (2003).
27. Mahoney, F. I. & Barthel, D. W. Functional evaluation: the Barthel index. *Md. State Med. J.* **14,** 61–65 (1965).
28. Kidd, D. *et al.* The functional independence measure: a comparative validity and reliability study. *Disabil. Rehabil.* **17,** 10–14 (1995).
29. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12,** 189–198 (1975).
30. Bauer, J. M., Kaiser, M. J., Anthony, P., Guigoz, Y. & Sieber, C. C. The mini nutritional assessment–its history, today's practice, and future perspectives. *Nutr. Clin. Pract.* **23,** 388–396 (2008).
31. Cruz-Jentoft, A. J. *et al.* Sarcopenia: European consensus on definition and diagnosis: report of the European Working Group on Sarcopenia in Older People. *Age Ageing* **39,** 412–423 (2010).
32. Nyamundanda, G., Brennan, L. & Gormley, I. C. Probabilistic principal component analysis for metabolomic data. *BMC Bioinformatics* **11,** 571 (2010).

33. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472,** 57–63 (2011).
34. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 (2011).
35. European Commission. *Population structure and ageing* http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_structure_and_ageing (2011).
36. Kinsella, K. & He, W. *An Aging World: 2008* (US Government Printing Office, 2009).
37. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474,** 327–336 (2011).
38. Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5,** 220–230 (2011).
39. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335–336 (2010).
40. Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* **73,** 5261–5267 (2007).
41. O'Sullivan, A., Gibney, M. J. & Brennan, L. Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am. J. Clin. Nutr.* **93,** 314–321 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Government of Ireland National Development Plan by way of a Department of Agriculture Food and Marine, and Health Research Board FHRI award to the ELDERMET project, as well as by a Science Foundation Ireland award to the Alimentary Pharmabiotic Centre. M.J.C. is funded by a fellowship from the Health Research Board of Ireland. We thank K. O'Donovan and P. Egan for clinical assistance, staff in Cork City and County hospitals for facilitating subject recruitment, S. Wong and B. Clayton for supercomputer access.

Author Contributions All authors are members of the ELDERMET consortium (http://eldermet.ucc.ie). P.W.O.T., E.M.O.C., S.Cu.[1] and R.P.R. managed the project; D.v.S., G.F.F., C.S., J.R.M., F.S., C.H., R.P.R. and PWOT designed the analyses; M.J.C., I.B.J., S.Co.[3], E.M.O.C., H.M.B.H., M.C., B.L., O.O.S., A.P.F., S.E.P., M.W. and L.B. performed the analyses; J.D. performed DNA extraction and PCR; M.W. and L.B. performed NMR metabolomics; M.O.C., N.H., K.O.C. and D.O.M. performed clinical analyses; M.J.C., I.B.J., S.Co.[3], E.M.O.C., L.B., J.R.M., A.P.F., R.P.R., C.H., F.S. and P.W.O.T. wrote the manuscript.

Author Information Amplicon sequence data, and shotgun sequence data, contigs, genes and annotations, have been deposited in MG-RAST under the Project ID 154 (http://metagenomics.anl.gov/linkin.cgi?project154). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.W.O.T. (pwotoole@ucc.ie).

METHODS

Subject recruitment and sample collection. This study was approved by the Cork Clinical Research Ethics Committee. Subjects older than 64 years were recruited and clinically investigated in two local hospitals, which serve a population base of ~481,000 in the Cork city and county region. They were defined as (1) community-dwelling (community); (2) attending an out-patient day hospital (out-patient); (3) in short-term rehabilitation hospital care (rehabilitation; under 6 weeks stay) or (4) in long-term institutionalized care (long stay; more than 6 weeks). The mean age of the subjects was 78 (± 8) years, with a range of 64 to 102 years. The subjects were all of Irish ethnicity. None of the faecal samples from elderly subjects from our previous study[8] were analysed in the current analysis, because we did not have food frequency data for all that cohort. Exclusion criteria were a history of alcohol abuse, participation in an investigational drug evaluation or antibiotic treatment within the previous 30 days, or advanced organic disease. Informed consent was obtained from all subjects or, in cases of cognitive impairment, by next-of-kin in accordance with the local research ethics committee guidelines. Data collected included anthropometric measurements, clinical history and status and medication history. Antibiotic use before the one-month exclusion period was also recorded for each subject. Thirteen younger adult subjects of age ranging 28–46 years, which had not been treated with antibiotics within 30 days, were also recruited by informed consent.

Clinical and nutritional data collection. Habitual dietary intake was assessed using a validated, semiquantitative, food frequency questionnaire (FFQ) based upon the SLAN study[42]. Food properties were determined using the UK Food Standards Agency Nutrient databank[43]. The mini nutritional assessment (MNA) was used as a screening and assessment tool to identify subjects at risk of malnutrition.

Non-fasted blood samples were collected and analysed at Cork University Hospital clinical laboratories. Cytokines were measured using validated, commercial multi-spot microplates (Meso Scale Diagnostics). Anthropometric measures included height, weight, calf and mid-arm circumference. Charlson comorbidity index, mini mental state exam, geriatric depression test, Barthel score and functional independence measures were carried out on all participants. For long-term care, day-hospital and rehabilitation subjects, a research nurse reviewed the medical records for information on disease and current medication usage.

Molecular methods and bioinformatics. DNA was extracted from faecal samples, and the V4 region of the 16S rRNA gene was amplified, sequenced and analysed, as described previously[44]. Briefly, V4 amplicons were sequenced on a 454 Genome Sequencer FLX Titanium platform (Roche Diagnostics and Beckman Coulter Genomics). Raw sequencing reads were quality trimmed using the QIIME pipeline[39] according to the following criteria: (1) exact matches to primer sequences and barcode tags, (2) no ambiguous bases (Ns); (3) read-lengths not shorter than 150 base pairs (bp) or longer than 350 bp; (4) the average quality score in a sliding window of 50 bp not to fall below 25. For large-scale assignments into the new Bergey's bacterial taxonomy[45] we used the RDP-classifier version 2.2 with 50% as confidence value threshold. This was based on what was found suitable for V4 amplicons from the human gut environment[44]. RDP classifications were imported into a MySQL database for efficient storage and advanced querying.

The amplicon reads were clustered into OTUs at 97% identity level, and filtered for chimaeric sequences using ChimeraSlayer (http://microbiomeutil.sourceforge.net/#A_CS). Representative sequences (the most abundant) for each OTU were aligned using PyNAST[46] before tree building using FastTree[47]. These phylogenies were combined with absence/presence or abundance information for each OTU to calculate unweighted or weighted UniFrac distances, respectively[48]. Principal co-ordinate analysis and Procrustes superimposition were then performed from the UniFrac distances and Food Frequency data. The amplicon sequences were deposited in MG-RAST under the Project ID 154.

Metagenomes were sequenced from libraries with 91 bp paired-end Illumina reads and 350 bp insert size and assembled using MetaVelvet[49]. Samples EM039 and EM173 were sequenced from libraries of 101 bp paired-end Illumina reads with a 500 bp insert size, and subsequently assembled using MIRA[50] in hybrid with 551,726 and 665,164 454 Titanium reads, respectively. Protein sequences from enzymes were screened against the assembled metagenomes using TBLASTN with an amino acid identity cut-off of 30% and an alignment length cut-off of 200 bp. We screened the metagenome data for enzymes associated with production of butyrate (butyryl-CoA transferase/acetyl-CoA hydrolase), acetate (acetate-formyltetrahydrofolate synthetase/formate-tetrahydrofolate ligase), and propionate (propionyl-CoA:succinate-CoA transferase/propionate CoA-transferase). Genes were predicted using MetaGene[51].

NMR analysis of the faecal water metabolome. Faecal water samples were prepared by the addition of 60 µl D_2O and 10 µl tri-methylsilyl-2,2,3,3-tetradeuteriopropionate to 540 µl faecal water. Spectra of samples were acquired by using a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence with 32k data points and 256 scans. Spectra were referenced to TSP at 0.0 p.p.m., phase and baseline corrected with a line broadening of 0.3 Hz using the processor on Chenomx NMR suite 7 (Chenomx). The spectra were integrated at full resolution for data analysis (PCA, PLS-DA, CIA) with the water region (4–6 p.p.m.) excluded and the data was normalized to the sum of the spectral integral. For PPCCA data analysis, the spectra were integrated into spectral regions (0.01 p.p.m.). Two-dimensional ^1H–^1H correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) were acquired on a 600 MHz NMR spectrometer. TOCSY spectra were acquired with a spin lock of 65 ms. All two-dimensional data were recorded with standard Varian pulse sequences collecting 1,024 × 128 data points with a sweep width of 9.6 kHz and 32 scans per increment.

Statistical methods and metabolome data analysis. Statistical analysis was carried out using R (version 2.13.2) or Stata (version 11) software packages. Kruskal–Wallis and Mann–Whitney tests were used to find significant differences in microbial taxa, clinical and biochemical measures, alpha diversity, and Healthy Food Diversity (HFD). Data were visualized by boxplots. Unless stated otherwise, box plots represented the median and interquartile ranges, with the error bars showing the last datum within 1.5 of the interquartile range of the upper and lower quartiles. We used least square linear regression for comparing alpha diversity and HFD. Median regression[52] was used to compare clinical measures and microbiota, while adjusting for age, gender, medications, and when appropriate residence location. For median regression, the median was modelled as a linear function of independent variables. Model parameters are estimated such that they minimised the sum of the absolute differences between observed and predicted values. *P* values were generated using the wild bootstrap method[53] to estimate variance.

A linear quantile (median) regression for two variables—a response variable (y) and a predictor variable (x)—is the following: median (y) = $\beta_0 + \beta_1 x$ where β_0 is the intercept (value when $y = 0$) and β_1 is the slope (change in median of y for a unit change in x). Together, these parameters describe the association between y and x, where x is a predictor of y. In the case of multiple predictor variables, each one is added to the regression equation and so the equation becomes median (y) = $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ and now the slope β_1 is interpreted as the median change in x_1 after adjusting for x_2. This can be likened to a laboratory experiment where the specific effect of one variable on another is isolated by holding all other relevant variables constant.

Following statistical analysis of the taxonomic classifications, we estimated FDR values using the Benjamini–Hochberg method[54] to control for multiple testing. The exception to this were analyses at the genus level where we estimated the proportion of true null hypotheses with the *Q*-value function unless the estimated π_0 was less than or equal to zero[55].

Statistical analysis of the NMR data was performed using diverse software packages: PCA and PLS-DA analysis was performed in SIMCA-P+ (Umetrics); permutation testing was performed in R and PPCCA was performed in R using the MetabolAnalyze package. The NMR data was Pareto-scaled before data analysis. Assignment of the spectral peaks (Supplementary Table 9) was performed using in-house libraries, statistical correlation analysis and two-dimensional NMR spectra (TOCSY and COSY).

42. Harrington, J. *et al.* Sociodemographic, health and lifestyle predictors of poor diets. *Public Health Nutr.* **14,** 2166–2175 (2011).
43. McCance, R. A. & Widdowson, E. M. *The composition of foods* 6th edn (Royal Soc. Chemistry, 2002).
44. Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4,** e6669 (2009).
45. Lilburn, T. G. & Garrity, G. M. Exploring prokaryotic taxonomy. *Int. J. Syst. Evol. Microbiol.* **54,** 7–13 (2004).
46. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26,** 266–267 (2010).
47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5,** e9490 (2010).
48. Hamady, M., Lozupone, C. & Knight, R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J.* **4,** 17–27 (2010).
49. Namiki, T., Hachiya, T., Tanaka, H. & Sakakibara, Y. in *ACM Conference on Bioinformatics Computational Biology and Biomedicine* (Association for Computing Machinery, 2011).
50. Chevreux, B. *et al.* Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14,** 1147–1159 (2004).
51. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34,** 5623–5630 (2006).
52. Koenker, R. & Basset, G. Regression quantiles. *Econometrica* **46,** 33–50 (1978).
53. Feng, X. D., He, X. M. & Hu, J. H. Wild bootstrap for quantile regression. *Biometrika* **98,** 995–999 (2011).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57,** 289–300 (1995).
55. Dabney, A., Storey, J. D. & Warnes, G. R. qvalue: *Q*-value estimation for false discovery rate control; R package version 1.24.20 (2010).