Supplementary Table 1: Numbers and demographics of subjects in control and intervention cohorts

	Controls						MedDiet Intervention					
	All Countries	Italy	UK	Netherlands	Poland	France	All Countries	Italy	UK	Netherlands	Poland	France
Individuals with sequenced microbiome	289	91	16	37	105	40	324	112	32	38	112	39
Median Age (Min-Max)	71 (65-79)	72 (65-79)	70.5 (65-79)	71.5 (65-79)	72 (65-79)	68 (65-77)	71 (65-79)	72 (65-79)	70.5 (65-79)	71.5 (65-79)	72 (65-79)	68 (65-72)
Gender (Male:Female)	145:144	46:45	7:9	21:16	46:59	25:15	141:182	56:56	12:20	13:25	45:67	15:24
Median BMI (Min-Max)	26.8 (18.8-44.6)	26.4 (18.8-44.6)	27.20 (20-31.5)	26.7 (19.7-35.1)	26.4 (21-37.6)	25.05 (18.9 – 37.5)	26.9 (18.5-46)	27.15 (18.7-37.9)	25.95 (18.5-33.7)	25 (20.3-35.8)	28.1 (19.5-46)	24.3 (19.7-31.3)

*: Baseline

SUPPLEMENTARY TEXT 1

MEASUREMENT OF INFLAMMATORY AND ADIPOSITY RELATED HORMONES

Briefly, fresh blood from each participant was collected after fasting in each recruiting centre. Blood was immediately centrifuged at 2000 x g for 10 min at 4°C and separated into plasma and serum according to a standardized operating procedure. All the specimens were stored at -80 until the time of analysis and sent to the project partners responsible for the analyses. A magnetic bead-based multiplex immunoassay (Bio-Plex) (BIO-RAD laboratories, Milan, Italy) was used to measure the inflammatory and adiposity related markers according to the manufacturer's instructions. In particular, Interleukin (IL) 1beta, 1Ra, 2, 4, 5, 6, 7, 8 10, 12p70, 13, 17, 17A, 18, Tumor Necrosis Factor alpha (TNFα), Interferon gamma (INFγ), Granulocyte Macrophage Colony-Stimulating Factor (GM-CSF), Granulocyte Colony-Stimulating Factor (G-CSF), Macrophage inflammatory protein-1beta (MIP1β) and Monocyte Chemotactic Protein-1 (MCP-1), were measured in multiplex with Bio-Plex Pro Cytokine, Chemokine, and Growth Factor Assays (intra-assay coefficient of variation (CV) was lower than 4.55% for all the molecules); Transforming Growth Factor beta1 (TGF-β1 intra-assay CV, 3.83%) with Bioplex Pro TGF- beta assay; Ghrelin (inter-assay CV, 2%) and Resistin (inter-assay CV, 4%) in multiplex with Bio-Plex Pro human diabetes assay. Plates were read and analyzed by Bio-Plex Manager Software. The level of Interleukin 6 receptor alpha (IL6rα, inter-assay CV, 3.1%)), Glycoprotein 130 (gp130, inter-assay CV, 5.9%), Pentraxin-3 (inter-assay CV, 6.8%) and soluble TNFalpha receptors R1 (TNF-R1, inter-assay CV, 6.1%) and R2 (TNF-R2, inter-assay CV, 7.7%) were assessed in multiplex in a subgroup of 360 samples with Bioplex Pro human inflammation assay (gp-130, inter-assay %CV 5.9). The quantitative determination of hsCRP, leptin, adiponectin has been performed by ProcartaPlexTM Immunoassay (eBioscience, Hatfield, UK) according to the manufacturer's instructions. Analysis was performed using Luminex 200 instrumentation (Luminex Corportation, The Netherlands). Assay sensitivities were 19.31 pg/mL for Leptin, 4.39 pg/mL for hsCRP, and 47.46 pg/mL for adiponectin.
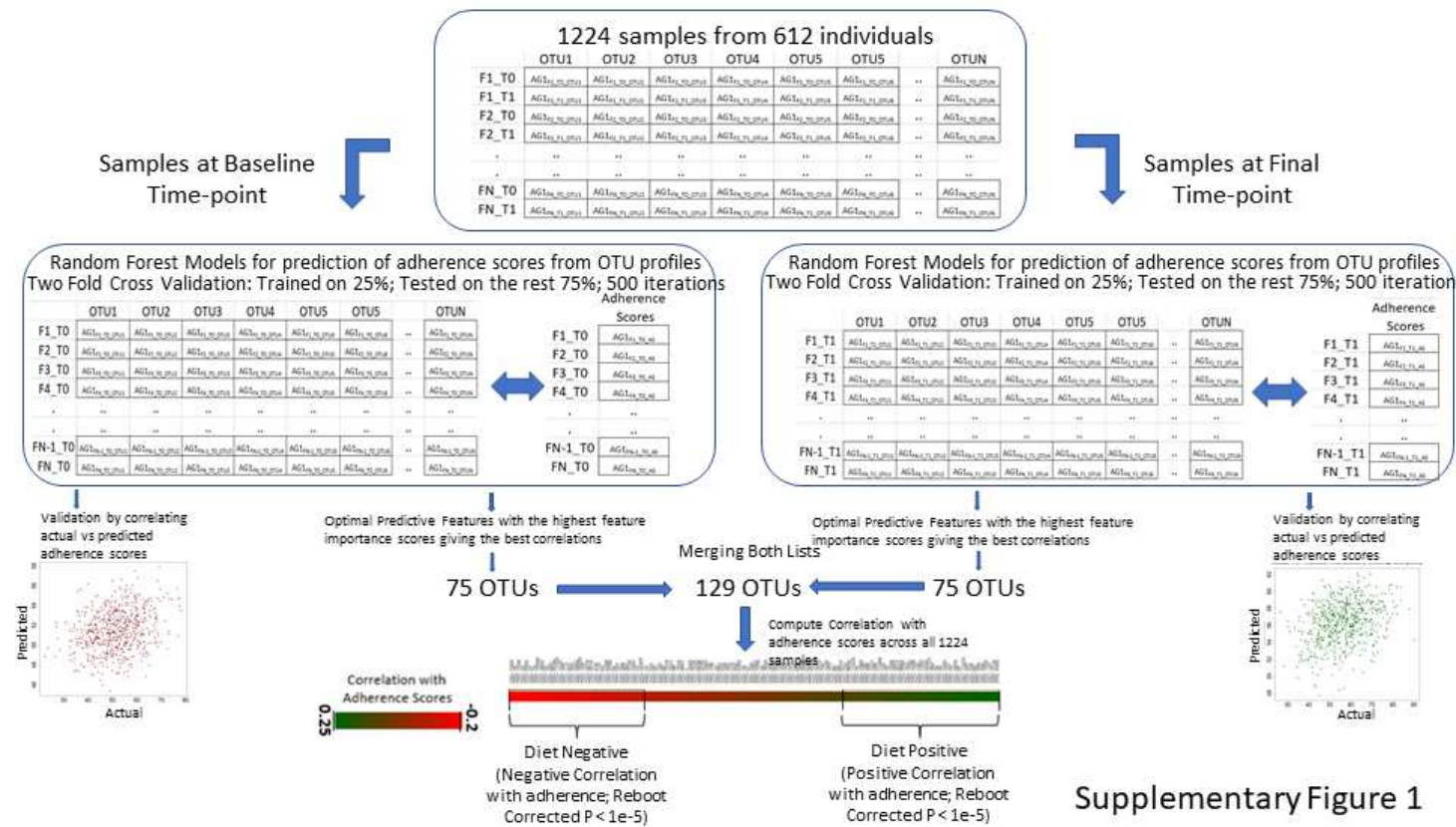
SUPPLEMENTARY TEXT 2

METHODOLOGY OF DNA EXTRACTION AND GENERATION OF 16S rRNA AMPLICON READS

A 250 mg stool sample was incubated with 1 ml lysis buffer (500 mM NaCl, 50 mM tris-HCl, pH 8.0, 50 mM EDTA and 4% sodium dodecyl sulphate (SDS)) in a 2-ml screw cap tube with 0.5 g sterile 0.1 mm zirconia beads and four sterile 3.5 mm glass beads (BioSpec Products, Bartlesville, OK). This was homogenised three times for 60 s at maximum speed (Mini-Beadbeater™, BioSpec Products), with cooling on ice for 60 s between homogenisation cycles. Samples were incubated at 95 °C for 15 min to further lyse the cells. Samples were centrifuged (16,000g) at 4 °C for 5 min and the supernatant was collected. For increased yield, an additional 300 µl of RBB lysis buffer was added to the pellet and the RBB steps were repeated as before. The supernatants were pooled and incubated with 350 µl of 7.5 M ammonium acetate (Sigma Aldrich, …) for 10 min. The protein-free DNA was precipitated with isopropanol at 4 °C and centrifuged at 16,000g. The pellet was washed with 70% (v/v) ethanol, allowed to dry, re-suspended in TE buffer, and treated with 10 mg/ml RNAse A (Thermo Scientific, Ireland). Proteinase K treatment and remaining DNA isolation was performed on-column using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany) according to manufacturers' instructions leading to 200 µl of DNA eluted in AE buffer. DNA was visualised on a 0.8% agarose gel for quality assessment and quantified using a NanoDrop 2000 system (Thermo Scientific). DNA was stored at −20 °C until use.
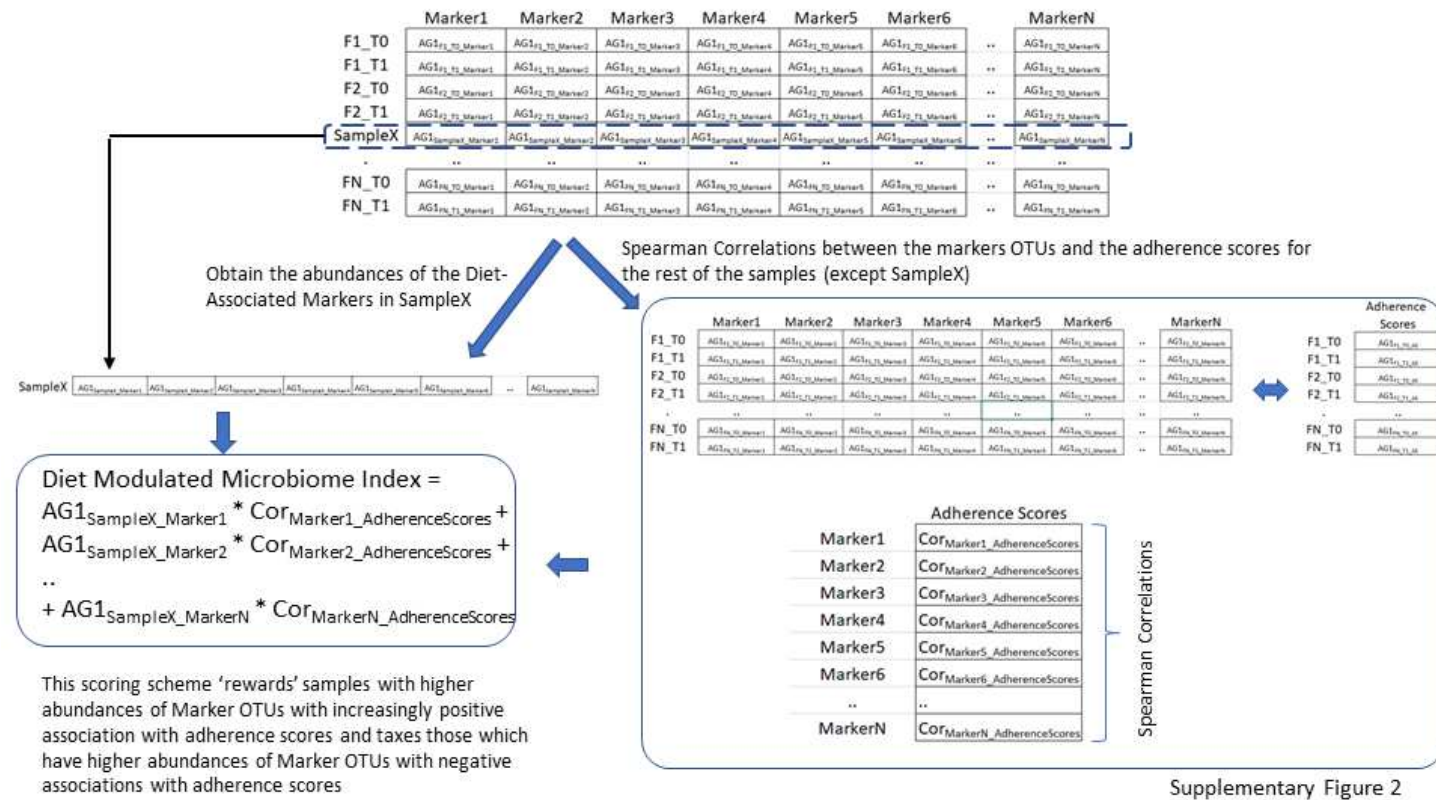
16S rRNA gene libraries for the Illumina MiSeq System were prepared manually following the manufacturer's protocol (15031942; Illumina, San Diego, CA, USA), with some modifications. V3 and V4 region of 16S rRNA genes were amplified using 15 ng of DNA template, Phusion HF Master Mix (Thermo Scientific) and 0.2 µM primers (98 °C 30 s; 25 cycles of 98 °C 10 s, 55 °C 15 s, 72 °C 20 s; 72 °C 5 min)(60). Amplicons were cleaned up using SPRIselect magnetic beads (Beckman Coulter, Indianapolis, IN) and checked for quality on a 1.2% agarose gel. Cleaned amplicons (5 µl) were used as template for Index PCR using Phusion HF Master Mix and Nextera XT Index Kit v2 Set A and D (Illumina) (98 °C 30 s; 8 cycles of 98 °C 30 s, 55 °C 30 s, 72 °C 30 s; 72 °C 5 min). Indexed amplicons were cleaned up using SPRIselect magnetic beads, run on a 1.2% agarose gel and quantified by Qubit dsDNA HS Assay (Thermo Scientific). The samples were pooled in equimolar amounts (40 ng DNA per sample) with up to 288 samples per library. Final library sizes were validated using Bioanalyzer DNA 1000 chips (Agilent Technologies, Santa Clara, CA). Libraries were denatured with 0.2 N NaOH

and diluted to 6 pmol/L with a 20% PhiX control before loading onto the MiSeq flow cell. Sequencing was performed on an Illumina MiSeq platform using a 2 × 250 bp paired end protocol, as per manufacturer's instructions (Illumina), on multiple sequencing runs.
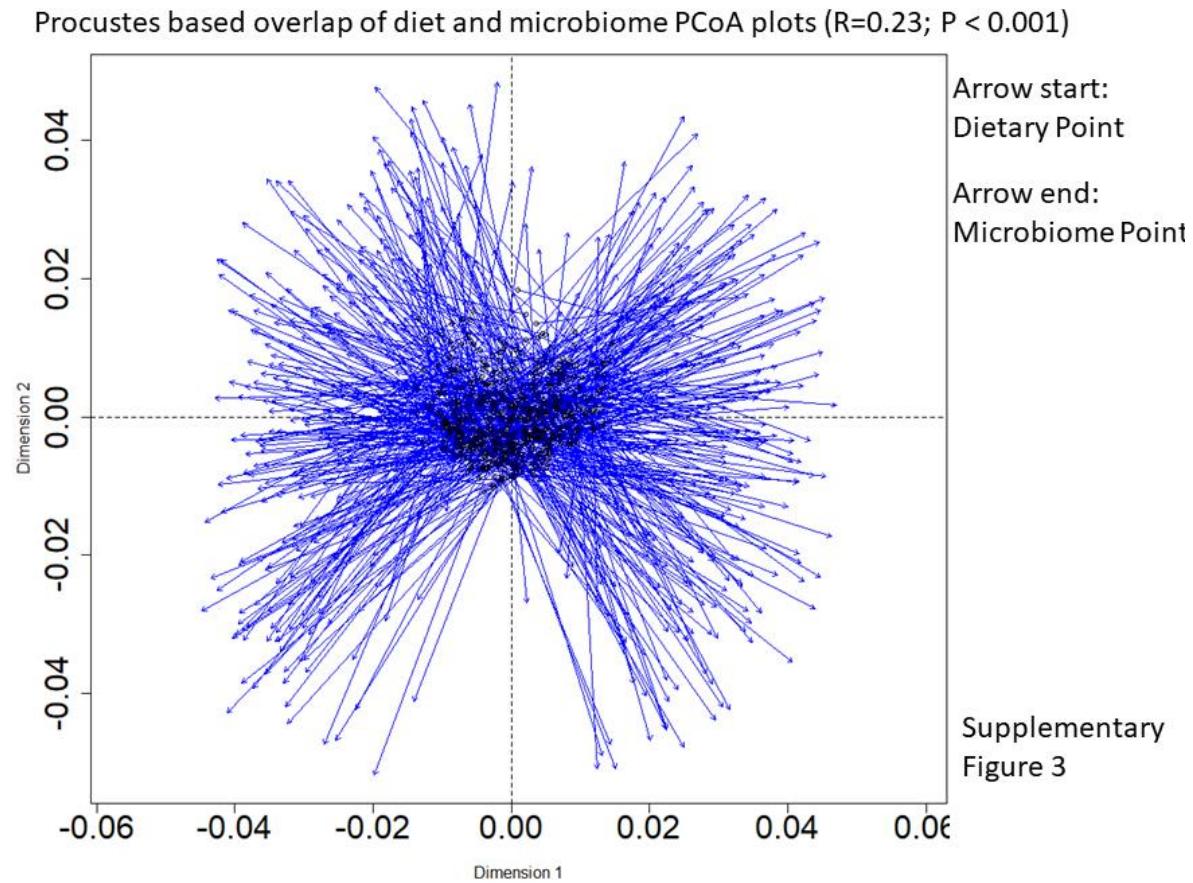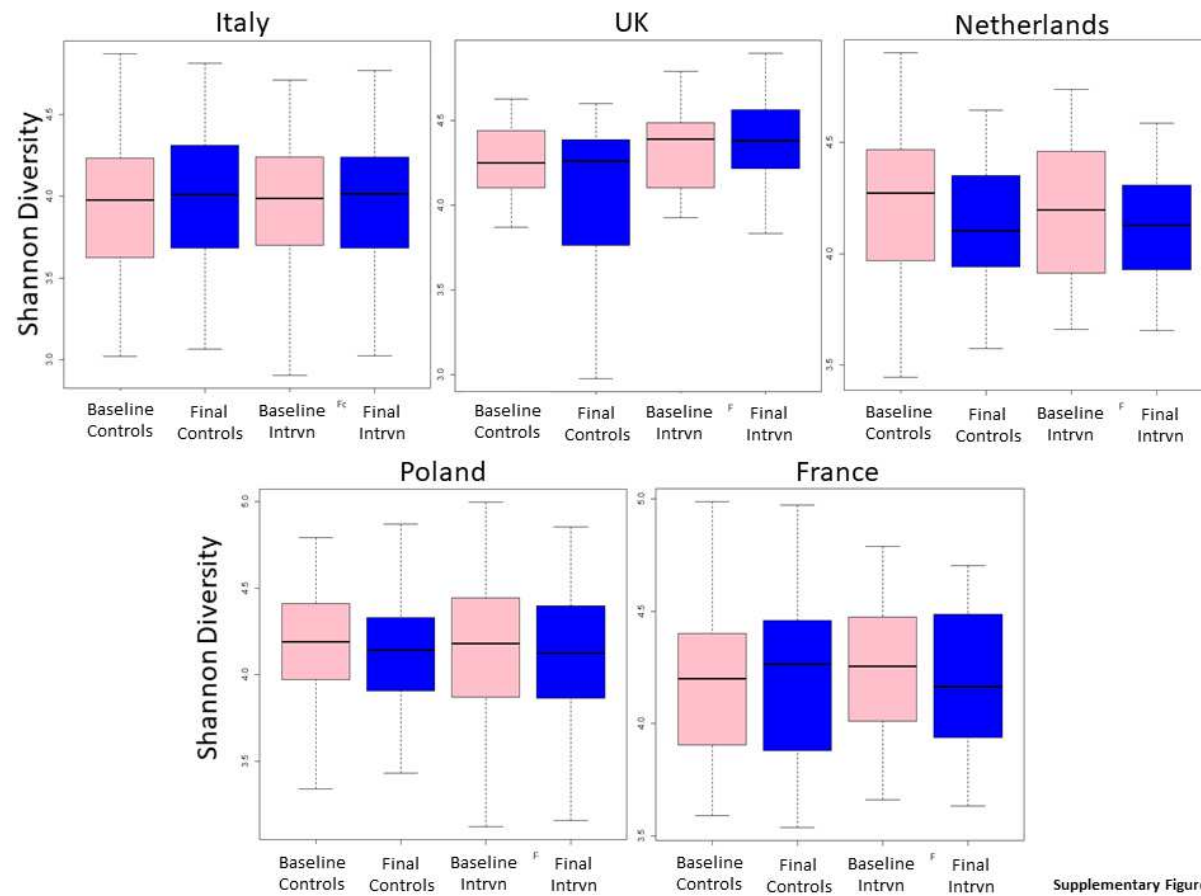
SUPPLEMENTARY

FIGURES

Supplementary Figure 1: Pictorial workflow describing the Random Forest based prediction of adherence scores from the microbiome abundance profiles and the identification of adherence score associated markers including the DietPositive and DietNegative markers.

Marker1 Marker2 Marker3 Marker4 Marker5 Marker6 MarkerN

Obtain the abundances of the Diet-Associated Markers in SampleX

Spearman Correlations between the markers OTUs and the adherence scores for the rest of the samples (except SampleX)

Diet Modulated Microbiome Index =

$$AG1_{SampleX\_Marker1} * Cor_{Marker1\_AdherenceScores} +$$
$$AG1_{SampleX\_Marker2} * Cor_{Marker2\_AdherenceScores} +$$
$$..$$
$$+ AG1_{SampleX\_MarkerN} * Cor_{MarkerN\_AdherenceScores}$$

This scoring scheme 'rewards' samples with higher abundances of Marker OTUs with increasingly positive association with adherence scores and taxes those which have higher abundances of Marker OTUs with negative associations with adherence scores

Adherence Scores

Marker1	$Cor_{Marker1\_AdherenceScores}$
Marker2	$Cor_{Marker2\_AdherenceScores}$
Marker3	$Cor_{Marker3\_AdherenceScores}$
Marker4	$Cor_{Marker4\_AdherenceScores}$
Marker5	$Cor_{Marker5\_AdherenceScores}$
Marker6	$Cor_{Marker6\_AdherenceScores}$
..	..
MarkerN	$Cor_{MarkerN\_AdherenceScores}$

Spearman Correlations

Supplementary Figure 2

Supplementary Figure 2: Pictorial representation of the methodology of computation of the microbiome indices using the leave-one-out strategy.

Procustes based overlap of diet and microbiome PCoA plots (R=0.23; P < 0.001)

Arrow start:
Dietary Point

Arrow end:
Microbiome Point

Supplementary
Figure 3

Supplementary Figure 3: Procustes plot showing the relative movement of the samples between the Principal Coordinate Analysis (PCoA) plots of the dietary and the microbiome profiles.

Supplementary Figure 4: Association of different genera with the specific nationalities shown as **a.** Plotted based on the weights of their association with the Principal Coordinate Analysis (PCoA) axes (as in Fig 1b). **b.** Plotted as heatmap showing the nationality-specific median abundances. Only those genera that are significantly over-abundant in at least one nationality as compared to the rest (Mann-Whitney tests using FDR corrected p-value < 0.15). Principal Component Analysis plots showing the changes in the microbiome profiles of subjects belonging to the various nationalities in **c.** control **d.** intervention cohorts. The p-values (obtained using envfit) of the association between the country and change of the microbiome in both the control and intervention cohorts are also indicated.

Supplementary Figure 5: Shannon diversity indices of the microbiota at baseline and Final time points for subjects in the intervention and control cohorts in the five different countries.
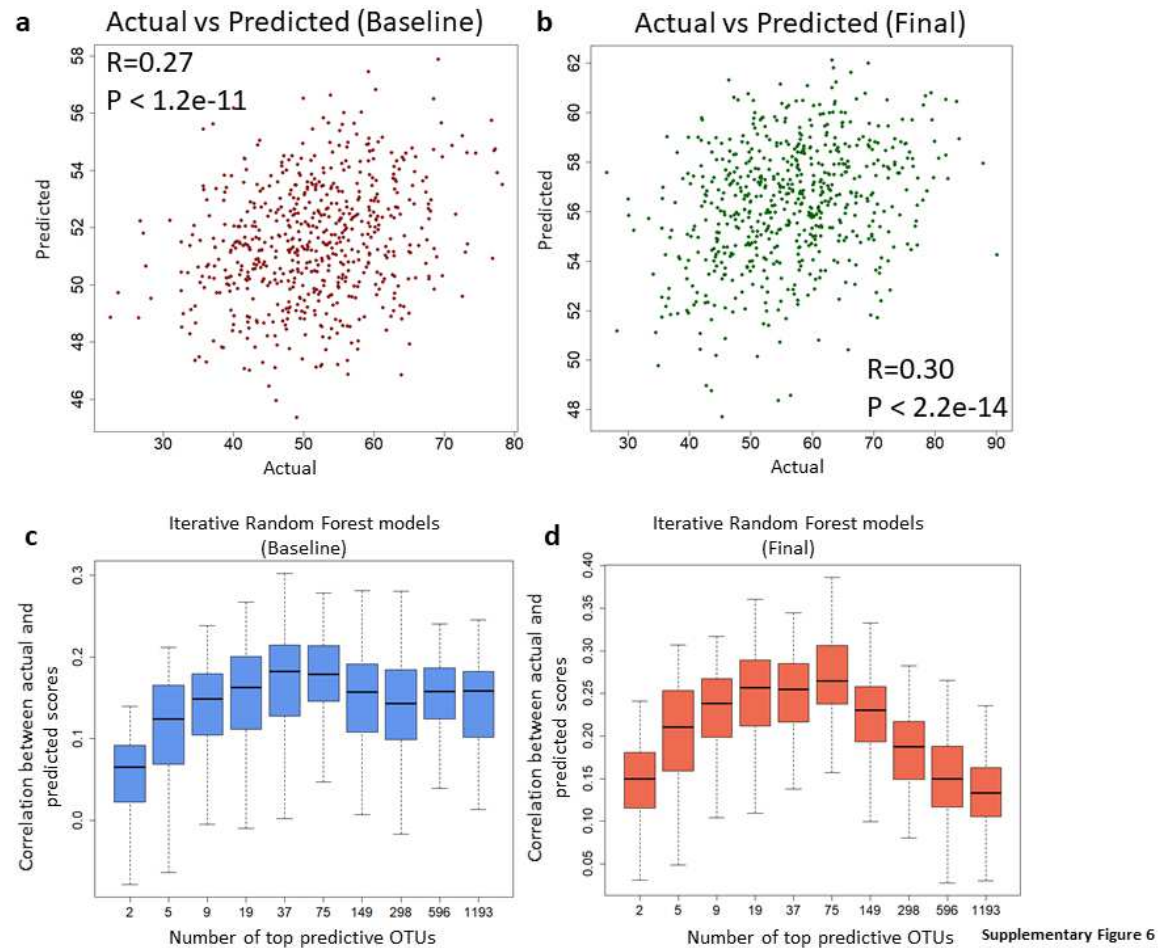
Supplementary Figure 6: Relationship between the Random Forest predicted and the actual dietary adherence scores for **a.** Baseline and **b.** Final time points. Boxplots showing the variation of the correlations between the actual and predicted dietary adherence scores obtained using iterative Random Forest prediction models with different number of top predictive features for **c.** Baseline and **d.** Final time-points. For both the time-points, the performance was observed to peak when the number of top features used was 75. Based on this, threshold of 75 top features was obtained for both time-points. The merger of the two lists of 75 features produced the final list of 129 features having optimal predictive ability across at least one of the time points.
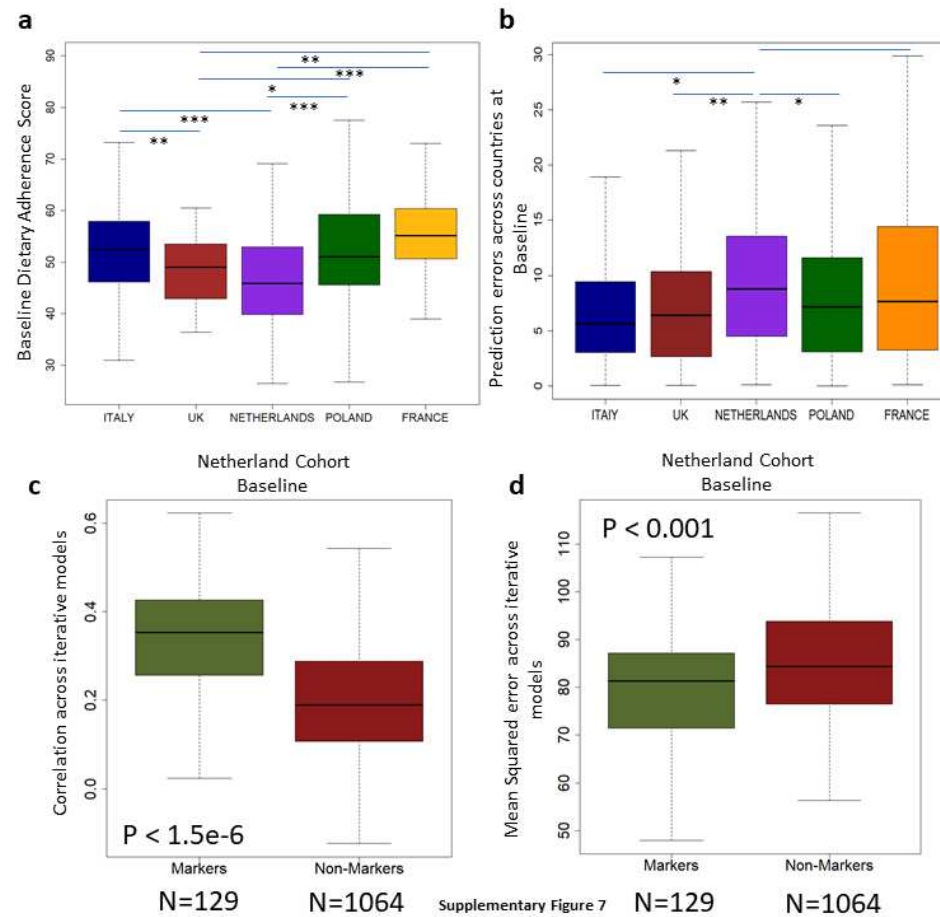
Supplementary Figure 7: a. Baseline Dietary Adherence Scores for the five different nationalities. **b.** Mean squared dietary adherence prediction errors for samples from five different nationalities **c.** Correlations and **d.** Mean squared errors between actual and predicted dietary adherence scores obtained using two different versions of iterative Random Forest models (two fold cross validation) separately for the samples from Netherlands (at baseline), While the first version was built using the set of 129 Diet-Associated Marker taxa as described in the previous figures (labelled as 'Markers') and the other using all 1064 OTUs besides the Diet-Associated Markers (labelled as 'Non-Markers').
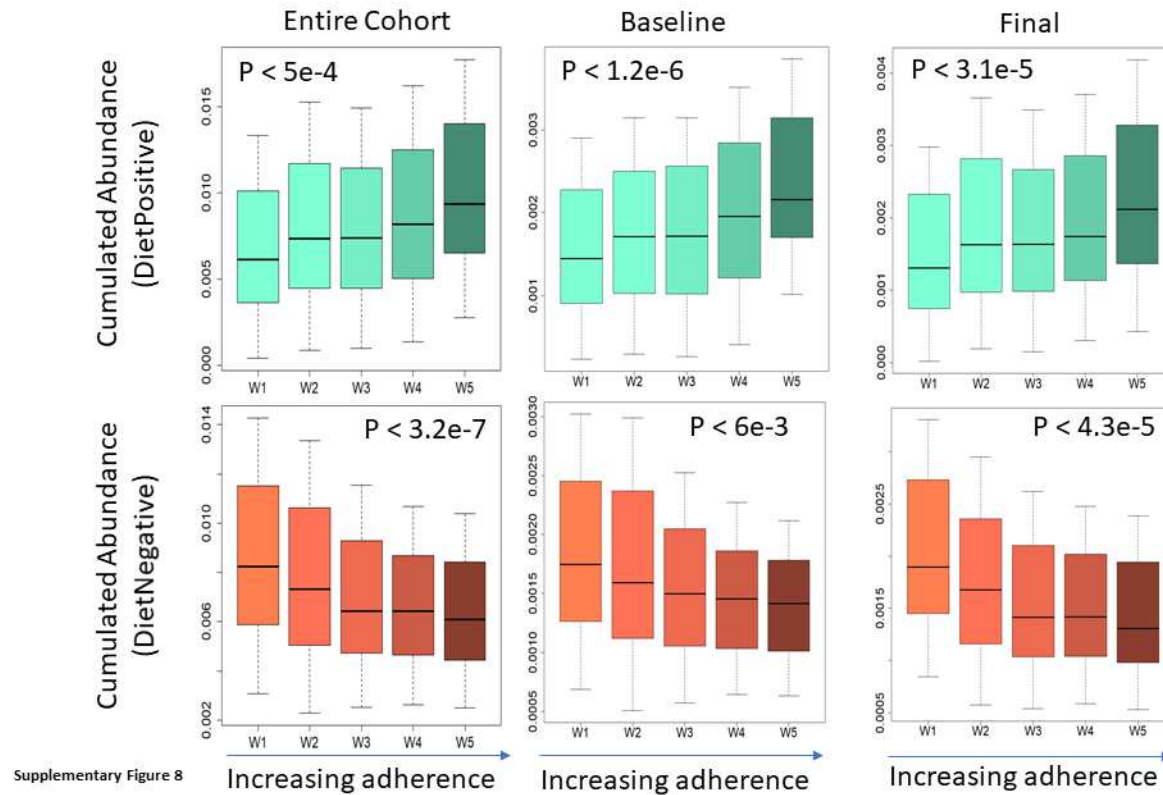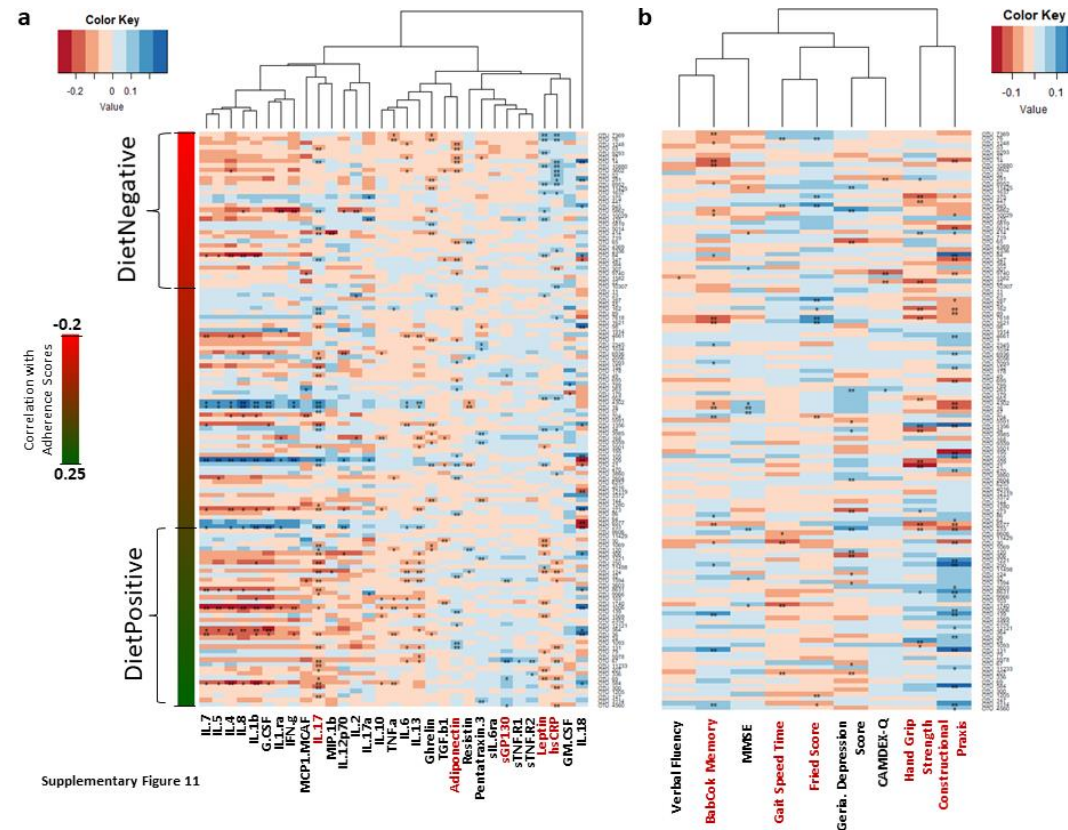
Supplementary Figure 8: Boxplots showing the variation of the cumulated abundances of the DietPositive and the DietNegative OTUs across overlapping windows of subjects with increasing adherence to the diet across the entire cohort as well as within the samples for the baseline and post-intervention time points (**See Methods**).
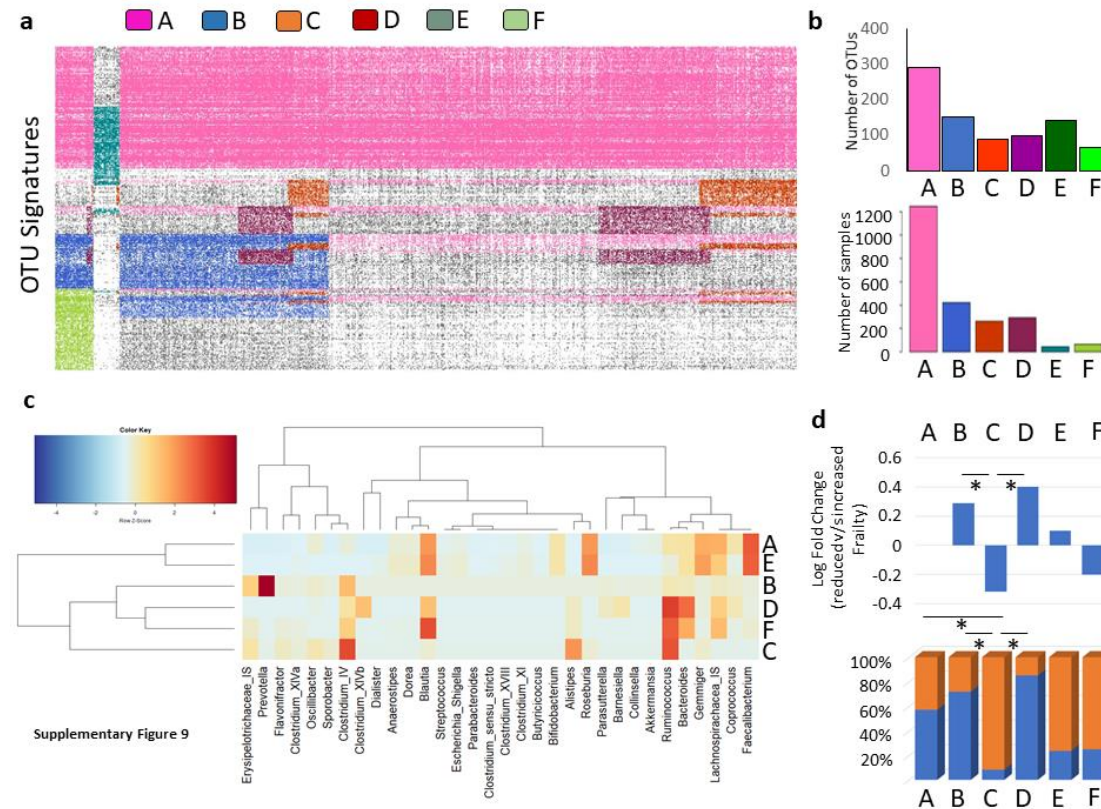
Supplementary Figure 9. Detection of specific taxonomic modules across the gut microbiomes using the iBBiG approach and association of specific modules with dietary adherence and reduced frailty. a. OTU detection profiles of the various modules obtained using the iBBiG approach. The color codes used for the various modules are: the primary core 'A' in pink; the Prevotella-associated 'B' in blue; the Alistipes-associated 'C' in orange; the Bacteroides-associated 'D' in maroon; the reduced core 'E' in darkgreen and; 'F' in light green. b. Bar-plot showing the number of samples containing each module (top) as well as the number of OTUs constituting each module (bottom). c. Heatmap showing the normalized abundances of the various genera within the OTUs constituting each module. d. Relative association of each of the modules with diet scores and frailty. The proportion bar-plots on the top right show the relative representation of the OTUs showing positive and negative association with diet scores within each module. The bar-plot on the bottom shows the log-fold increase in the number of samples containing each module in the individuals with reduced frailty (across time-points) as compared to those showing no change or an increase of frailty. Overall, these trends show the specific association of certain iBBiG modules with diet and frailty. While Modules B and D are associated positively with the Mediterranean diet and reduced frailty, Modules C shows the opposite trend.
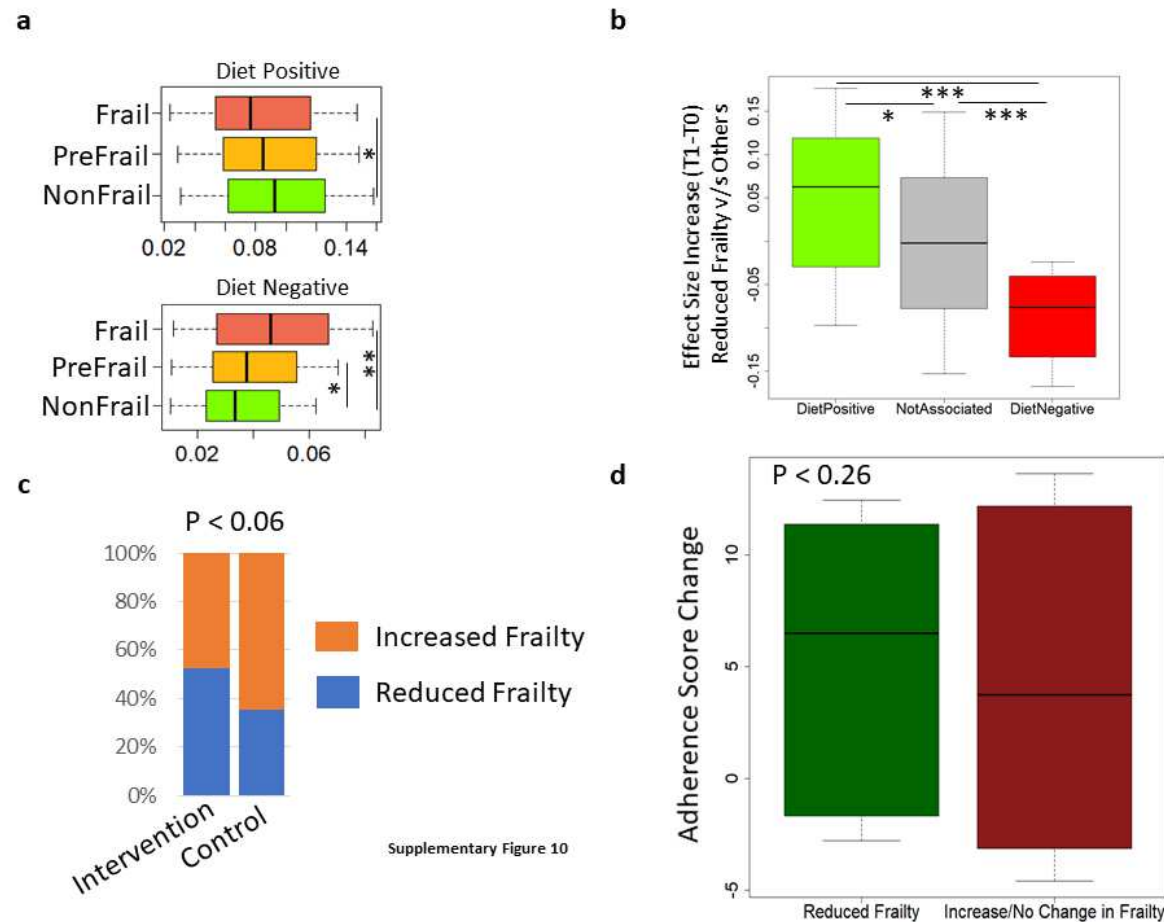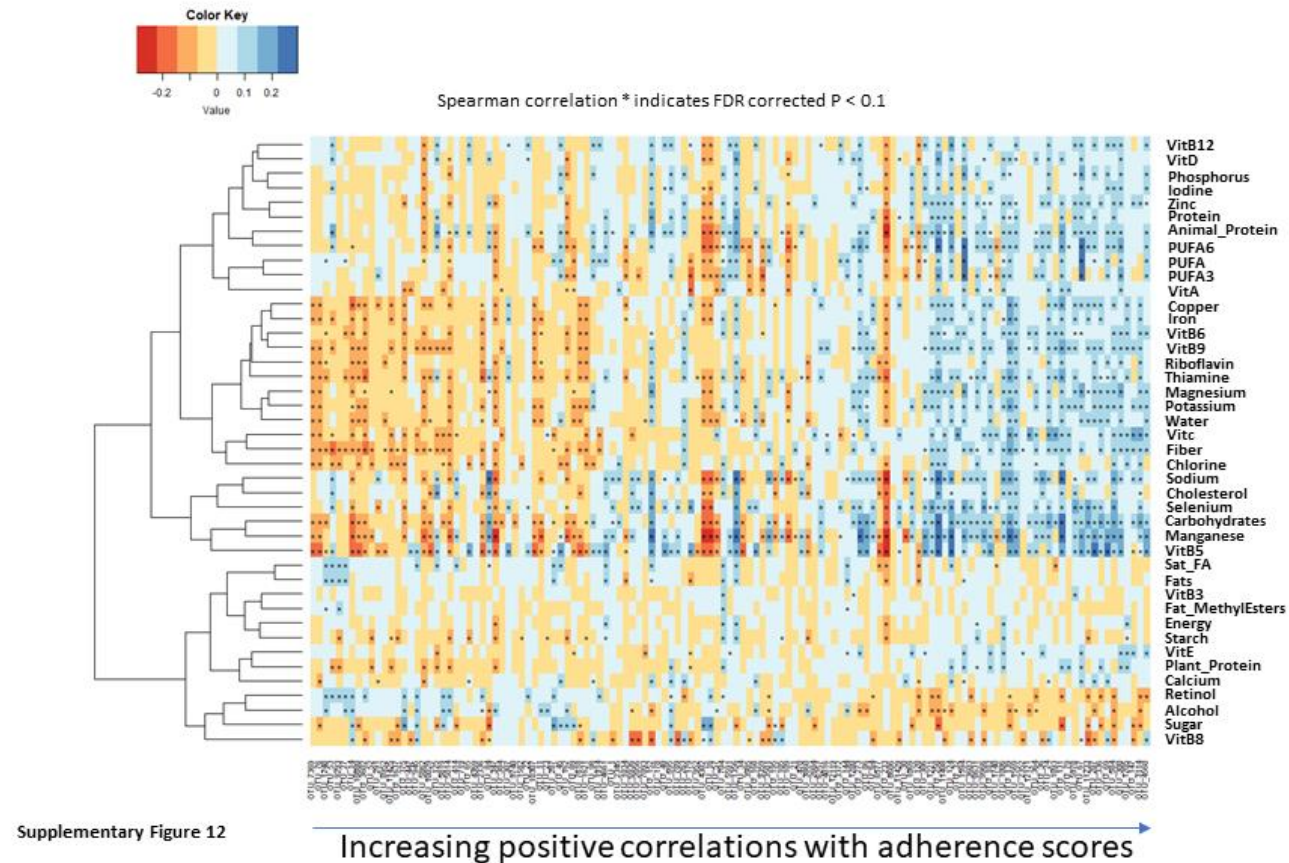
Supplementary Figure 10: **a.** Boxplots showing the variation of the cumulated abundance of the DietPositive and DietNegative OTUs in the Frail, Pre-Frail and the Non-Frail individuals. **b.** Variation of the effect-size differences (cohens' d) of the across time-point changes of the DietPositive OTUs, DietNegative OTUs and Not-Associated OTUs in Individuals with Reduced Frailty versus those with Increased or No Change in Frailty. **c.** Proportional representation of individuals with reduced and increased frailty in the control and intervention cohort. There was a marginally significant increase (Fishers' test P-value) in the representation of individuals with increased frailty in the control cohort. d. Boxplots showing the variation of adherence score changes in individuals with reduced frailty and those with increased or no change in frailty status.
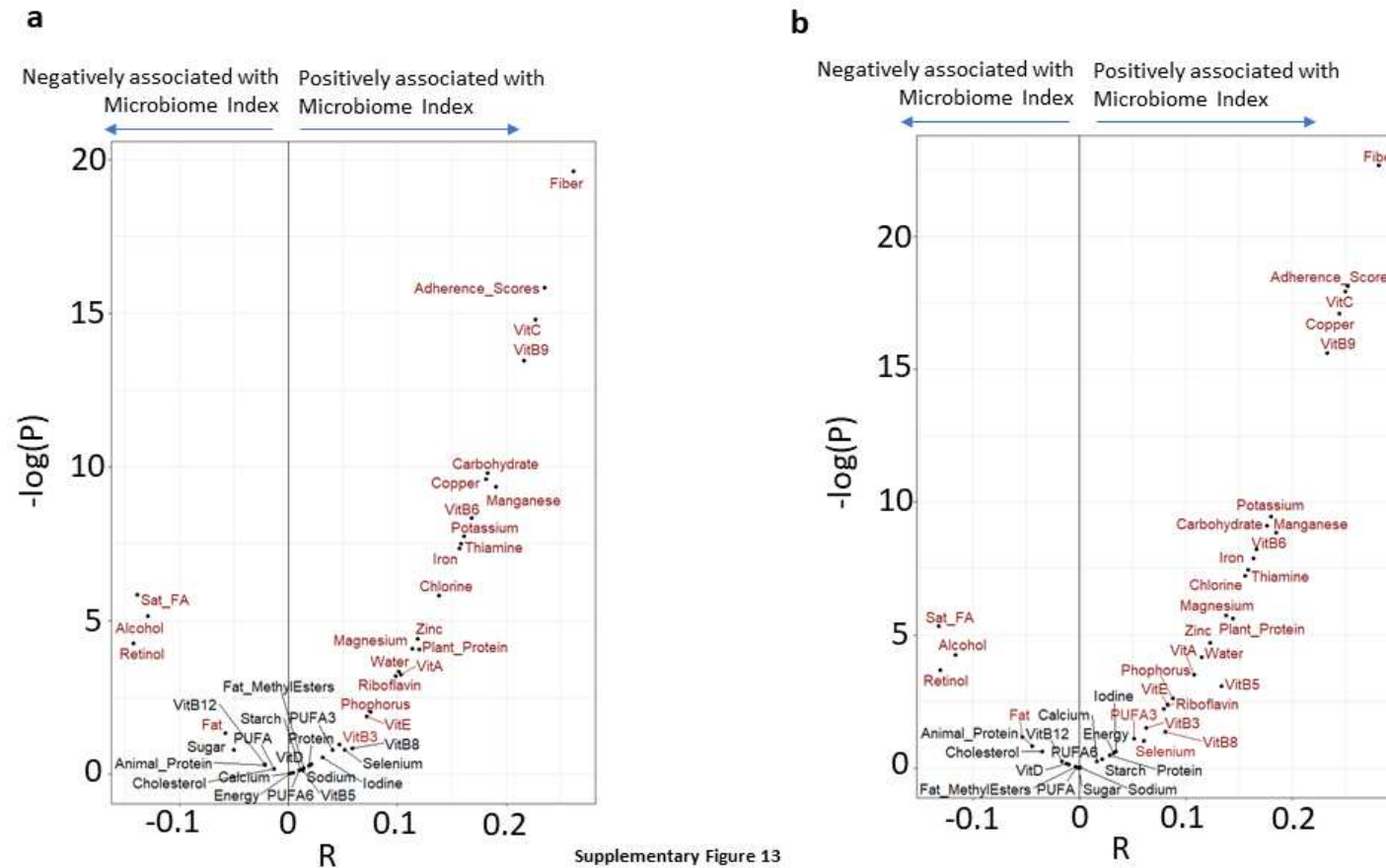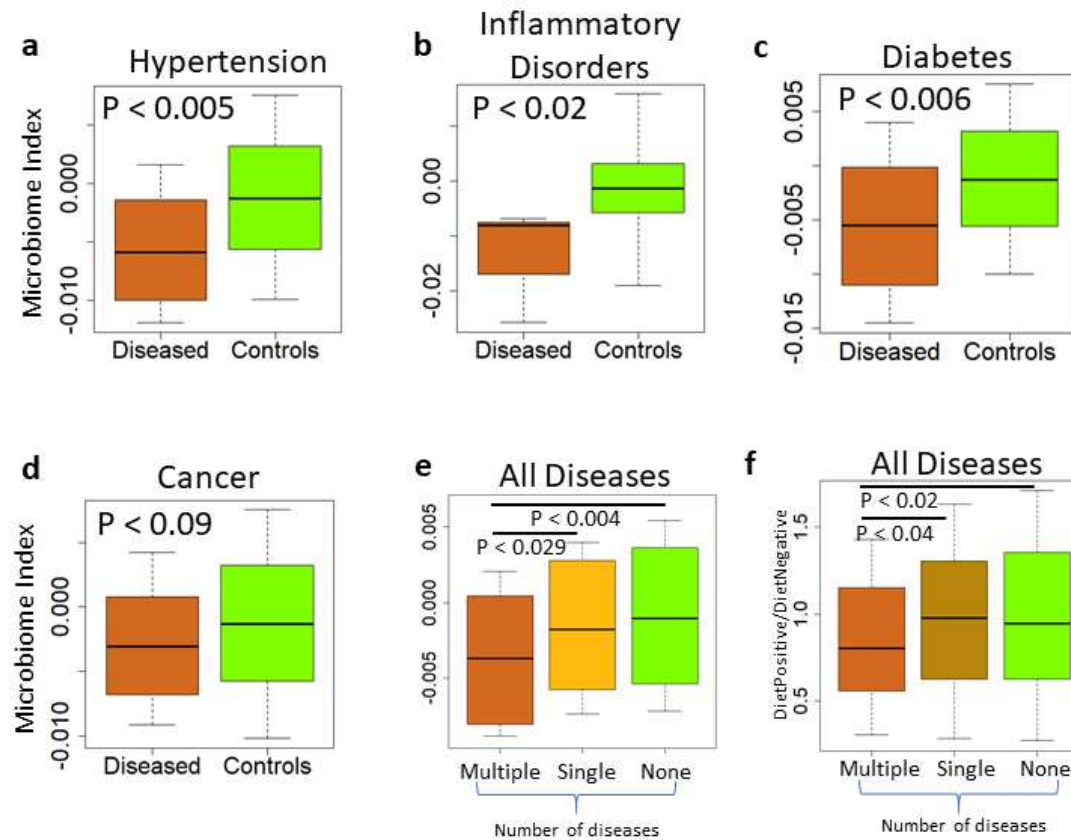
Supplementary Figure 11: Heatmap showing the variation of the association patterns (obtained using Spearman Rhos) of the adherence associated marker OTUs (arranged from top to bottom in increasing order of their correlations with the adherence scores) with a. each of the pro/anti-inflammatory cytokine levels and b. the different measures of frailty, cognitive function. For each cell, colors indicate the Spearman Rho values (as shown), ** indicates a significant association with FDR corrected P-value < 0.15, * indicates a marginal association with nominal P-value < 0.05. The DietPositive and the DietNegative OTUs are also demarcated. Measures highlighted in red are those, for which the association patterns with the individual OTUs were observed to exhibit significant positive or negative correlations (Spearman correlation FDR corrected P-value < 0.15) with the OTU-adherence score association values.

Supplementary Figure 12: Heatmap showing the partial spearman associations of the different dietary components with the marker OTUs arranged in increasing order of their association with the dietary adherence scores. For each marker OTU, partial spearman correlations were obtained after adjusting for the confounding effects of age, BMI, gender, country and poly-pharmacy.
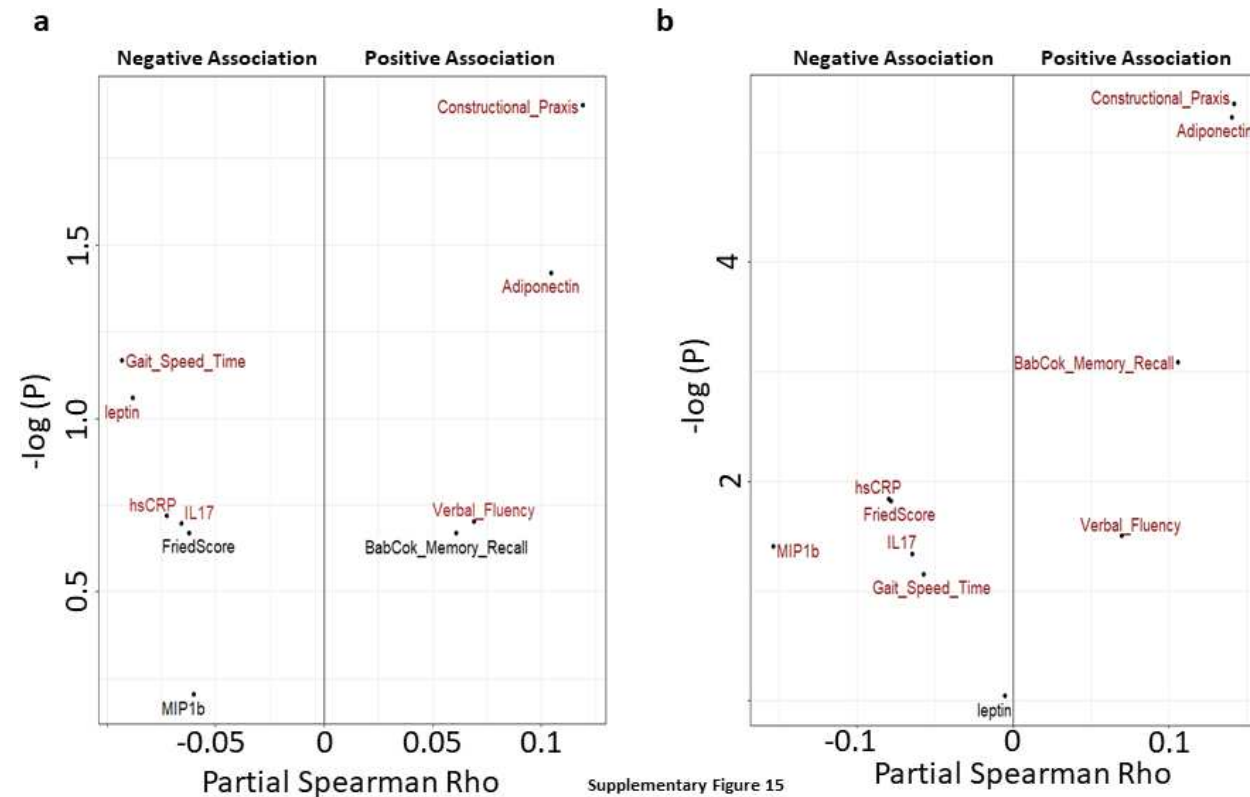
Supplementary Figure 13. Violin plots showing the a. Partial spearman correlations between the consumption of different dietary components and the microbiome index across all time-points taking into account age, bmi, gender, country and poly-pharmacy. b. Partial spearman correlations between the consumption of different dietary components and the microbiome index across all subjects at the baseline taking into account age, bmi, gender, country, disease-status and poly-pharmacy.
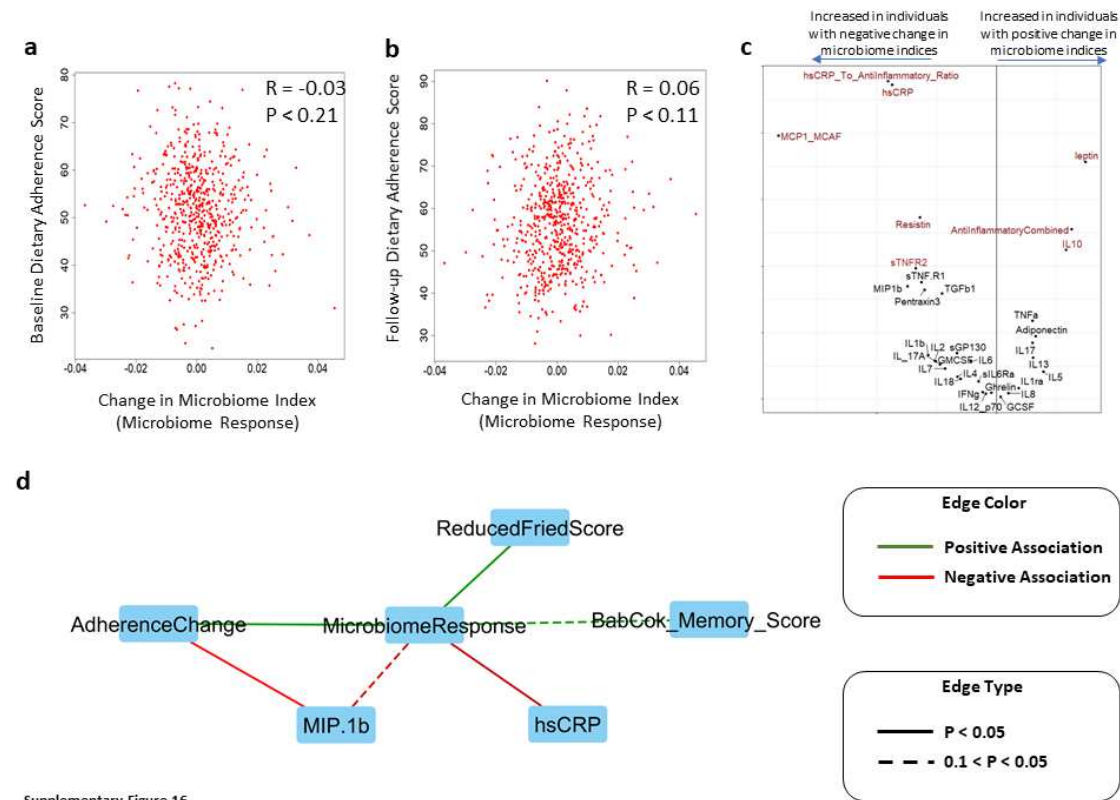
Supplementary Figure 14

Supplementary Figure 14. Boxplot showing the comparison of the MedDiet modulated microbiome index for individuals suffering from heart attack (**a**), inflammatory disorders (**b**), Type II Diabetes (**c**) and Cancer (**d**) with the control individuals (tagged as no-disease) at baseline. Boxplot showing the variation of microbiome index (**e**) and the abundance ratio of the DietPositive to DietNegative markers (**f**) for individuals with multiple, single and no-diseases at the baseline. The P-values of the Mann-Whitney U tests are also indicated for each pairwise-comparisons.
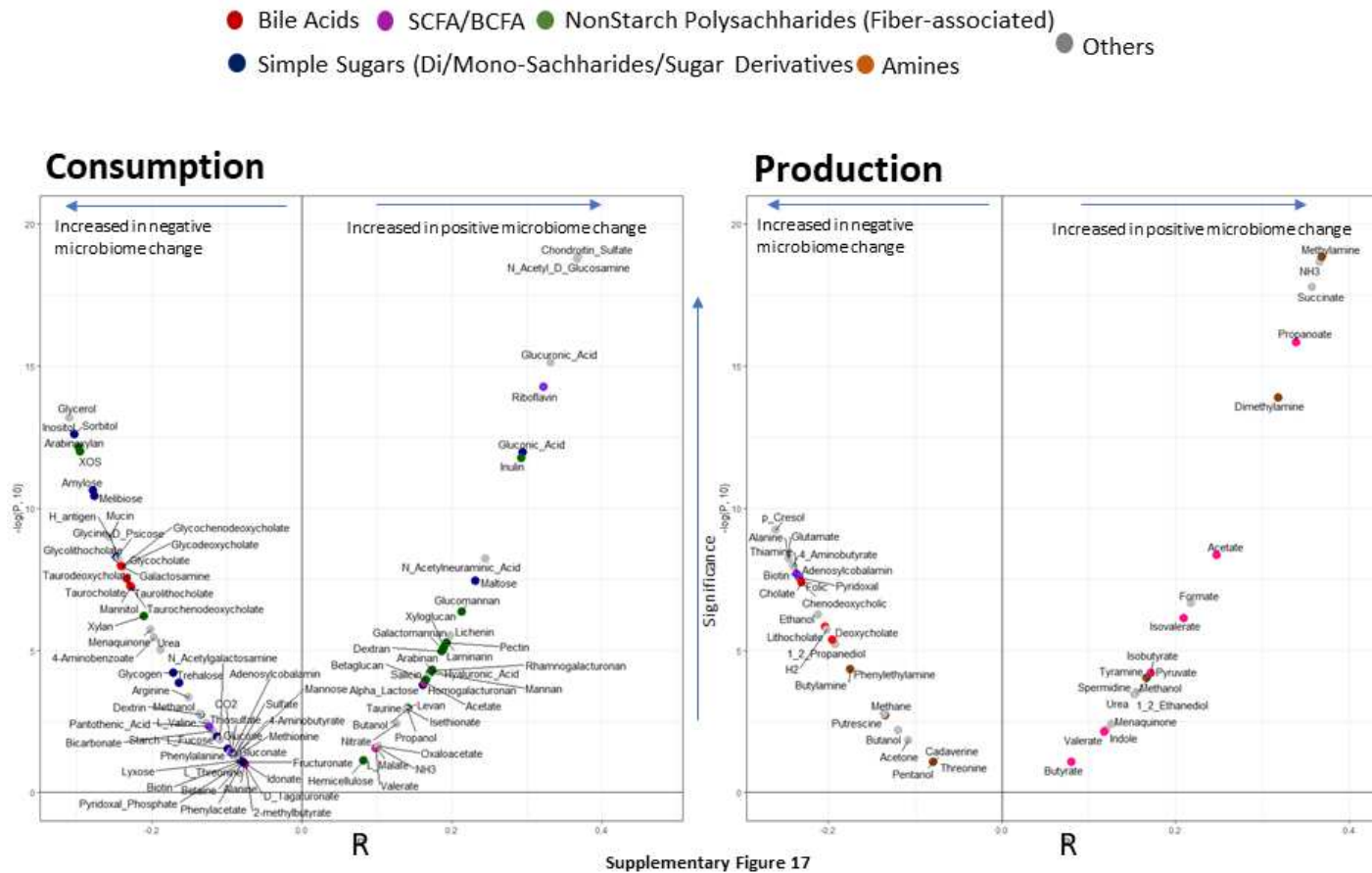
Supplementary Figure 15: a. Violin plot showing the association (partial Spearman correlations) of the different measures of frailty, cognitive function and inflammatory marker levels (identified in figure 4) with the MedDiet modulated Microbiome index at the baseline after taking into account the age, BMI, gender, nine different disease pathologies (with greater or equal to 10 subjects), polypharmacy and gender as a confounder. b. Violin plot showing the association (partial Spearman correlations) of the different measures of frailty, cognitive function and inflammatory marker levels (identified in figure 4) with the MedDiet modulated Microbiome index at both the baseline and follow-up time points after taking into account the age, BMI, gender, polypharmacy and gender as a confounder. X-axis contains the spearman Rho values, and Y-axis indicates the -log (base 10) of the P-values. While most negatively associated measures are expected to be extreme left of the plot, the most positively associated measures are expected to be extreme right of the plot. Points are colored based on the significance of the obtained associations (Red indicates associations with FDR corrected P-value < 0.1).

Supplementary Figure 16: Scatterplots showing the correlation between the microbiome response (that is the across time-point change in microbiome indices) and a. Baseline dietary adherence scores and b. final dietary adherence scores. c. Violin plots showing the spearman correlations of the changes in the different cytokine levels with the change in microbiome indices between the follow-up and baseline time points. X-axis contains the spearman Rho values, and Y-axis indicates the -log (base 10) of the P-values. While most negatively associated measures (that is those cytokines for which negative changes in levels are associated with positive changes in microbiome indices) are expected to be extreme left of the plot, the most positively associated measures (that is those cytokines for which positive changes in levels are associated with positive changes in microbiome indices) are expected to be extreme right of the plot. Points are colored based on the significance of the obtained associations (Red indicates associations with FDR corrected P-value < 0.1). Cumulated levels of anti-inflammatory cytokines were calculated as the mean ranked abundances of anti-inflammatory cytokines IL-10, IL-4, IL-5 and IL-1ra. Ratios of hsCRP to anti-inflammatory cytokines were calculated as ratios of the ranked abundance of hsCRP and the mean ranked abundances of anti-inflammatory cytokines IL-10, IL-4, IL-5 and IL-1ra. d. Graph showing the marginal or significant associations (dotted line indicating marginal associations with P < 0.1 and solid line indicating P < 0.05) between the across time-point changes of the various measures obtained using pairwise linear regressions.

Supplementary Figure 17: Violin plots showing the inferred metabolite **a.** consumption and **b.** production profiles showing significant (positive or negative) associations (Spearman correlation; FDR corrected P-value < 0.15) with microbiome responses. While most negatively associated measures are expected to be extreme left of the plot, the most positively associated measures are expected to be extreme right of the plot. The points are colored based on the metabolite groups as indicated on the top panel of the Figure.
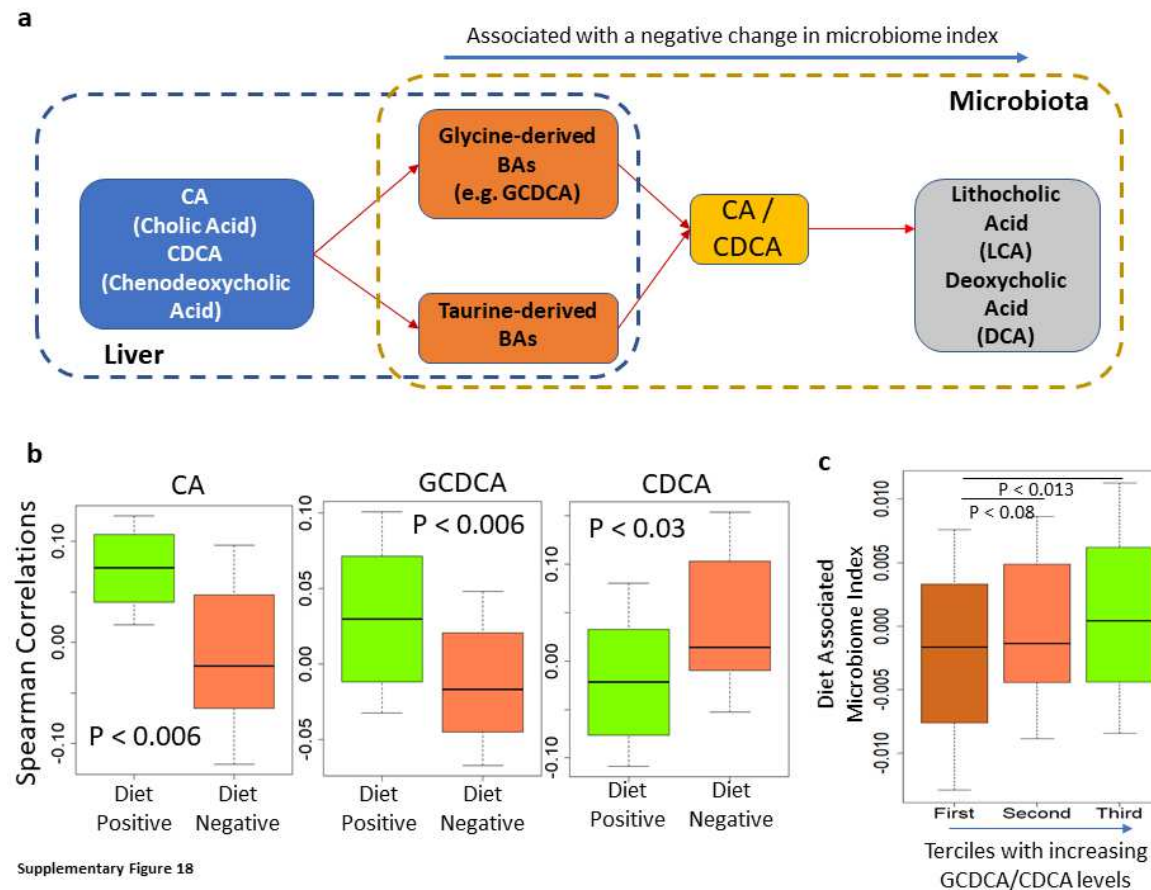
Supplementary Figure 18. a. Schematic representation of the bile acid conversion pathway, highlighting the specific sub-module converting glycine/taurine-conjugated bile acids (like TCA, GCDCA) to carcinogenic secondary bile acids (LCA/DCA), through CA and CDCA, that is associated with a negative change in microbiome index highlighted in red. b. Boxplots showing the spearman correlations of the abundances of the DietPositive and DietNegative OTUs with the measured plasma levels of CA, DCA and GCDCA for the subset of Italian and Polish individuals. c. Boxplot comparing the across the time-point changes in the microbiome index for individuals with increasing GCDCA/CA levels (grouped into three equal terciles with increasing GCDCA/CA ratios). P-values of pairwise Mann-Whitney U-tests are also indicated.
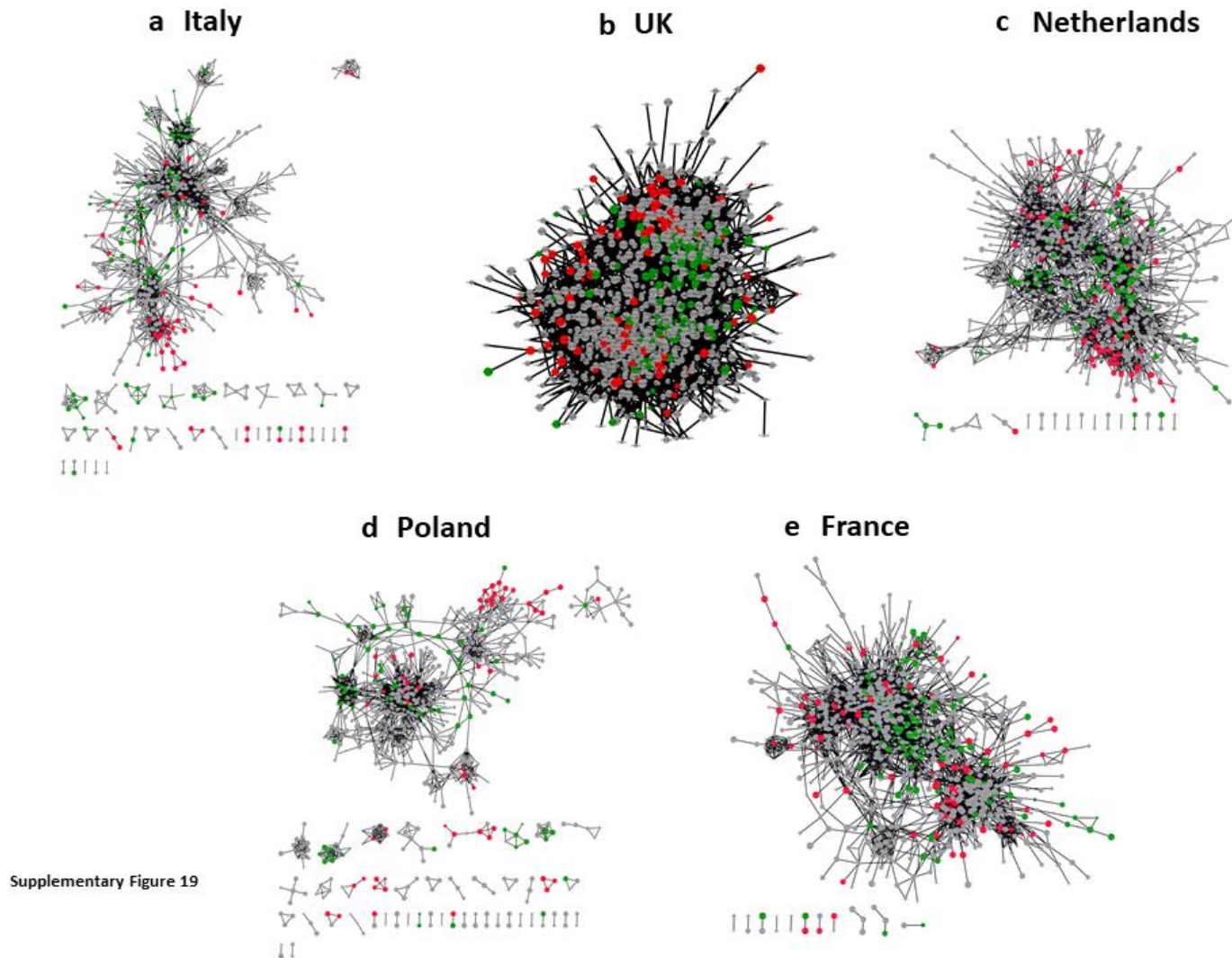
Supplementary Figure 19: Inter-OTU co-occurrence networks obtained for the different nationalities. DietPositive, DietNegative and the non-associated OTUs are shown in green, red and gray color, respectively. Despite variations in their overall structures, all networks gave a clear picture whereby in the DietPositive taxa were placed in the centre of the neworks, while the DietNegative taxa in the periphery.
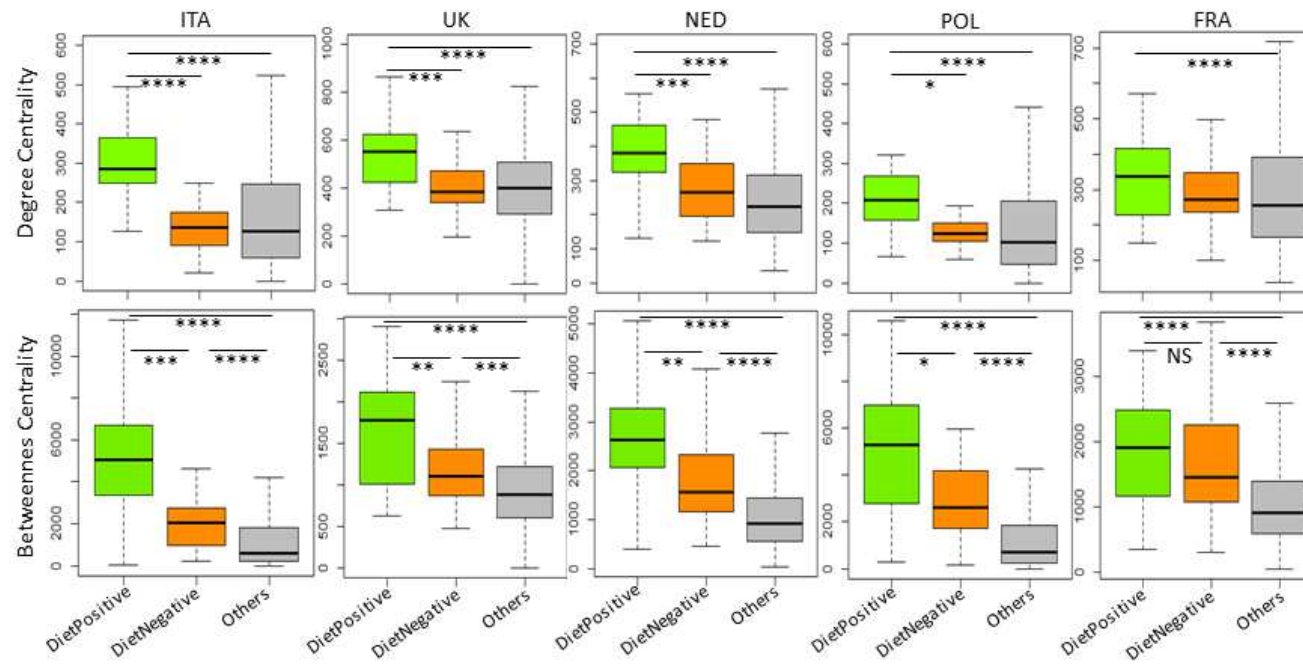
Supplementary Figure 20

Supplementary Figure 20: Variation of the degree and betweenness centrality of the different groups of taxa within the co-occurrence networks for the different nationalities.
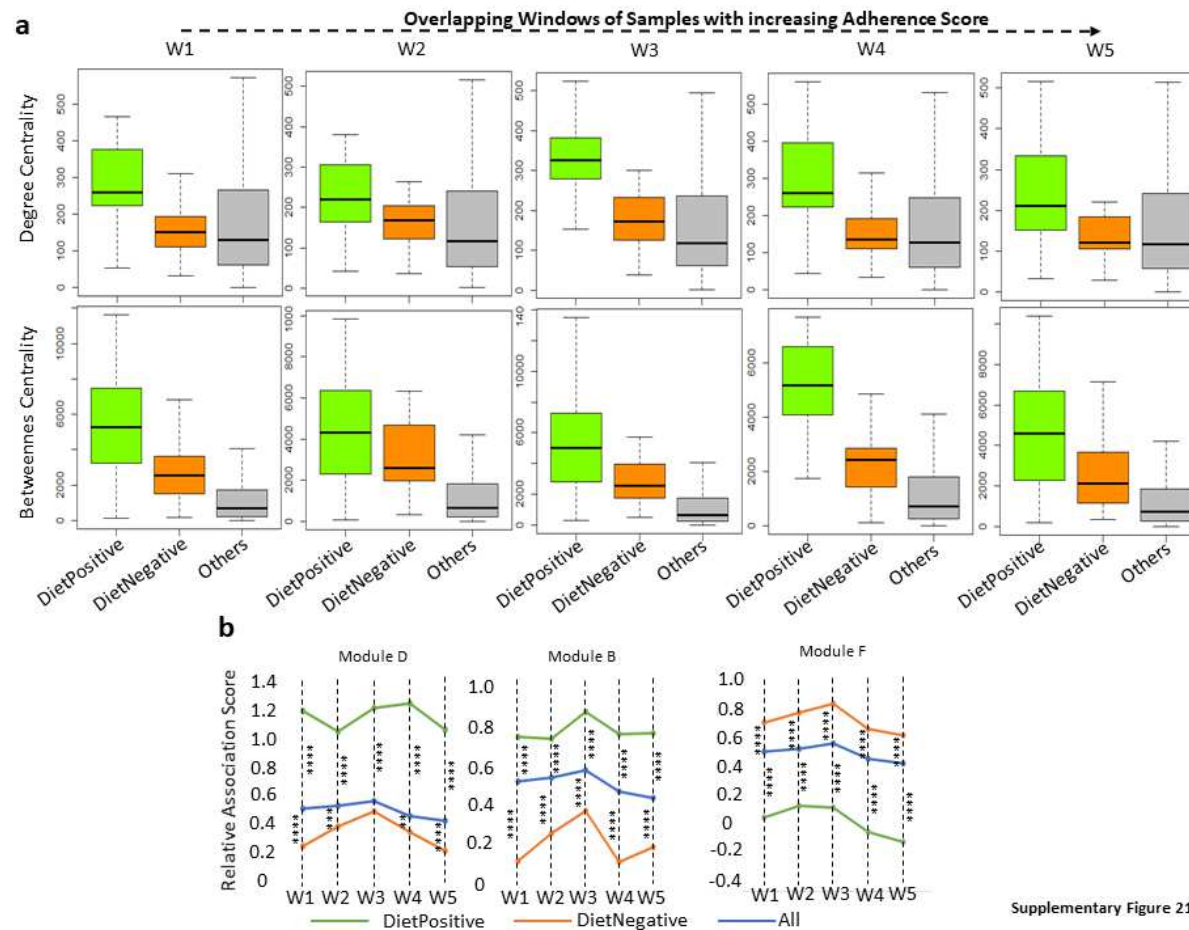
Supplementary Figure 21: a. Variation of the degree and betweenness centrality of the different groups of taxa across individuals belonging to the overlapping groups of increasing diet adherence scores. b. Relative co-occurrence propensities of the different groups of taxa with the iBBiG taxonomic modules. Relative co-occurrence propensities of the different groups with the frailty-associated Module C is provided in figure 5d.

SUPPLEMENTARY TEXT 3

COMPLETE DESCRIPTION OF METHODS USED FOR THE BIOINFORMATIC STATISTICAL ANALYSIS OF THE MICROBIOME DATA

Pre-processing of amplicon reads

The FLASH program was used to join the paired-end reads (25). The data was barcode-corrected and quality filtered using the QIIME package; followed by clustering of reads into Operational Taxonomic Units (OTUs) (97% identity threshold) using USEARCH Clustering algorithm; followed by chimeric removal (26, 27). The taxonomic classification of the representative sequences for each OTU was performed using both the RDP classifier (genus level: 0.8 confidence threshold) and the SPINGO classifier (species level: 0.7 confidence threshold) (28, 29).

Multi-variate analysis of dietary profiles and taxonomic profiles

Multivariate analyses using Principal Coordinate Analysis (PCoA) were performed using the ade4 package of the R programming interface, using Spearman distances of the individual sample profiles as well as the across time point changes (final-baseline).To test the significance of the between-country variation of the baseline dietary and microbiome profiles, Permutational Analysis of Variance (PERMANOVA) was performed on the PCoA objects using the adonis function of the vegan R package. Procrustes analysis was performed to quantify the relationships between the baseline diet and microbiome profiles using the procrustes function of the vegan package. The Shannon diversities of the samples were obtained using the diversity function of the vegan R package.

Machine Learning-based identification of microbiome taxa associated with the dietary intervention

The Machine learning based Random Forest (RF) approach (implemented in the randomForest package of R) was used to identify microbiome taxa significantly associated with NU-AGE FBDG adherence scores. We first divided individuals into three equal tertiles, namely 'High Adherence', 'Medium Adherence' and 'Low Adherence' in decreasing order of the change in adherence across time-points and the samples from each into two cohorts corresponding to the baseline and final time-points. Two separate models were created for the baseline and the final time points. The performance of the models was measured by calculating the correlation

between the actual and the predicted food scores obtained using the models. The RF approach provided the feature score importance scores for each microbiome component (OTUs) (indicating the extent of association of these with the dietary adherence scores). For identifying the most-predictive features, iterative random forest models (n=100, sample subset-size=100) with varying number of features (selected in decreasing order of their feature importance scores) were obtained using the randomForest package (two-fold cross validation) and their performances compared. Finally, to identify the OTUs associated with dietary adherence, a Reboot approach (using Spearman correlations) was used to identify OTUs that were significantly associated with adherence scores with an FDR corrected P-value < 1e-5 (30). OTUs positively and negatively associated with diet were classified as DietPositive and DietNegative, respectively. A pictorial representation of the workflow adopted for this entire step is provided in **Supplementary figure 1**.

Overview: iBBiG is based on the detection profile of the taxonomic units (in this case, the Operational Taxonomic Units (OTUs)). It then utilizes an iterative, heuristic, genetic-algorithm based methodology to identify modules of taxa within a microbial community that tend to show strong co-occurrence relationships across a given population of microbiomes. The primary advantage of this strategy is its flexibility, as it allows identification of over-lapping modules such that certain taxonomic units can be part of multiple modules. Such a partitioning strategy makes more biological sense as certain taxa (or species) can be part of multiple guilds because of their functional versatility or may be functionally specialized (i.e. belonging to specific guilds).

Method: For identifying modules within the gut microbiome, we used the **i**terative **B**inary **Bi**-clustering of **G**ene-sets (iBBiG) approach (38). Rather than profiling abundances or proportions, iBBiG investigates the detection profile of the taxonomic units or OTUs. Subsequently, an iterative, heuristic, genetic-algorithm based methodology is used to identify taxonomic modules that tend to show strong co-occurrence relationships across a given population of microbiomes. For performing the iBBiG based clustering, we used the iBBig function available within the Bioconductor package of R. While OTUs belonging to the different modules were then classified based on their clustering patterns, samples were classified based on the occurrence of the different iBBiG modules within them. The taxonomic compositional pattern of each module was then obtained by collating the RDP-based genus classification of each OTU and subsequently rank-normalizing these based on the abundance

of each genus (in terms of the number of OTUs) across a module. To associate the modules with frailty, we first obtained the frailty status of each individual at each time-point (0: Non Frail; 1: Pre Frail; 2: Frail). Subsequently based on the changes across time-points, individuals across the cohorts were classified as 'Reduced Frailty', 'No Change' and 'Increased Frailty'. The representation of each of the modules were obtained at both the time-points for each of three groups of individuals. The occurrence changes of each module (the number of samples in which a module is present at follow-up divided by the number of samples the module is present in at the baseline) were computed for each group. The log fold changes in these ratios in the Reduced Frailty with respect to the Increased frailty groups would provide the enrichment or depletion of the modules in individuals with reduced frailty as compared to those showing an increase in frailty across time-points. A positive change would indicate enrichment, and a negative value would indicate depletion. To compare the patterns across modules X and Y, Chi-square tests (using the chisq.test function of R) were then performed on the contingency tables containing four values, namely occurrence at baseline and follow-up of reduced frailty and occurrence at baseline and follow-up of increased frailty, corresponding to the two modules. To check for the significance of the differences of the occurrences across modules in terms of their diet association, we obtained number of times a module was present in the list of DietPositive and the DietNegative OTUs, and subsequently compared them using the Fishers' Exact test (fisher.test function of R).

Associating dietary adherence and microbiome changes with frailty and inflammation

For associating the abundances of the adherence associated marker OTUs with the different measures of frailty, cognitive function and cytokine profiles, we computed Spearman correlations using the corr.test function of the psych package in R (along with the Benjamini-Hochberg corrected p-values).

To account for various confounders, we used Partial Correlations (partial.r and the corr.p functions of the psych R package). Partial correlations measure the strength and the direction of the association between two variables considering the effect of confounding variable (s). Partial Correlations are like multiple regressions with confounders but not limited to specific distributions of the response and predictor variables. Further, one can compute rank-based non-parametric measures of association like the Spearman rho (which we have used in this study), after considering the confounding effect of other factors like adherence scores or age/BMI/gender.

Computation of Microbiome Indices

A pictorial representation of the methodology for this purpose is described in **Supplementary figure 2**. This scoring scheme 'rewards' samples with higher abundances of Marker OTUs with increasingly positive association with adherence scores and taxes those which have higher abundances of Marker OTUs with negative associations with adherence scores.

For each sample, the diet-modulated microbiome score was computed using the following formula:

$\sum_{\text{across all marker OTUs}}$ (OTU correlation with Diet adherence scores) * Abundance of the OTU

To avoid over-fitting, leave-one out strategy was applied where for computing the microbiome index for a given sample, the sample was not considered while calculating the OTU correlations (with Diet Adherence scores).

Obtaining Inferred Microbial Metabolite Profiles based on Species Abundance Profiles

Literature annotated Species-to-Metabolite consumption/production associations were already available as part of the Virtual Metabolic Human database as well as those obtained in a recent meta-analysis by Sung *et al* (32, 33). These were parsed to create a present/absence information map of around 300 metabolite production and consumption profiles in greater than 900 species in a 0 (absent) and 1 (present) notation. Given the SPINGO-based species abundance profile, from the 16S amplicon data, the inferred metabolite profile was then obtained as an inner product of the species abundance profile and the species-to-metabolite map.

Generation of co-occurrence networks and computation of centrality measures

We used the Reboot Approach for generating the inter-microbial co-occurrence/co-inhibition networks (30) (described in **Supplementary text 4**). The co-occurrence networks obtained were visualized using Cytoscape (34). For any network, two different centrality measures were calculated for the nodes, namely degree centrality and betweenness centrality using the igraph R package. The relative co-occurrence propensities between any two groups of taxa were calculated as the log of the number of positive edges divided by the number of negative edges.

Given any two features (in this case, the OTUs), the Reboot approach computes the association between the two features using two different distributions of association measures obtained using repeated iterations as described below(52). The association measure can be any score, like the Pearson correlation, Spearman correlation, the Regression coefficients, or even the effect size measures. The first distribution (bootstrap distribution) was obtained by taking the repeated sub-samples of randomly selected observations and then computing the

association between the two features. This profiled the association values across an entire observation landscape, thereby removing biases which could be present because of specific samples. The second distribution (null distribution) was obtained by performing an equal number of iterations, where in each iteration, a fixed set of values (which in this case was 50%) are swapped across samples for both the features. The profiles were then re-normalized and the associations computed for the two features. The distribution of the values obtained in the two distributions were then compared using any comparative tests (which in this case was Mann-Whitney). The p-values thus obtained were then False Discovery Rate (FDR) corrected (Benjamini-Hochberg) and those pairs of features having FDR-corrected associations of less than 1e-5 (threshold used in this study) were inferred to be significant and an edge drawn between them in the network. The directionality of the association was taken as the sign of the median value of the bootstrap distribution. While pairs of features with significant positive associations were used to create the co-occurrence network, those with negative associations were used to create the co-inhibition network.

*Please refer to the main document for the corresponding reference numbers.

Supplementary Table 2: List of top 129 OTU markers obtained using the Random Forest approach for the prediction of dietary compliance scores, along with their SPINGO classifications, association with food scores as well as the ranked feature importance scores for the baseline and followup time points

OTU	Ranked Feature Correlation		Genus	Species	Final
	Baseline	Follow-up			
OTU_4560	0.99412752	0.5041946	Unclassified	Unclassified	Unclassified
OTU_3114	0.98993289	0.9798658	Faecalibacterium	Faecalibacterium_prausnitzii	Faecalibacterium_prausnitzii
OTU_147	0.86996644	0.9941275	Unclassified	Unclassified	Unclassified
OTU_1205	0.95721477	0.9530201	Unclassified	Unclassified	Unclassified
OTU_300	0.96224832	0.9102349	Unclassified	Unclassified	Unclassified
OTU_584	0.98573826	0.9916107	Faecalibacterium	Faecalibacterium_prausnitzii	Faecalibacterium_prausnitzii
OTU_69	0.99832215	0.6526846	Unclassified	Unclassified	Unclassified
OTU_336	0.97986577	0.9781879	Roseburia	Unclassified	Roseburia_Unclassified
OTU_207	0.98238255	0.2332215	Unclassified	Unclassified	Unclassified
OTU_11233	0.2307047	0.9983221	Blautia	Ruminococcus_torques	Ruminococcus_torques
OTU_67	0.97651007	0.9974832	Unclassified	Unclassified	Unclassified
OTU_5978	0.98909396	0.658557	Unclassified	Unclassified	Unclassified
OTU_79	0.85151007	0.9899329	Eubacterium	Eubacterium_xylanophilum	Eubacterium_xylanophilum
OTU_131	0.90855705	0.9639262	Unclassified	Unclassified	Unclassified
OTU_1093	0.99496644	0.9211409	Anaerostipes	Unclassified	Anaerostipes_Unclassified
OTU_24	0.93875839	0.8557047	Unclassified	Unclassified	Unclassified
OTU_36	0.79026846	0.9630872	Prevotella	Prevotella_copri	Prevotella_copri
OTU_364	0.8909396	0.9513423	Clostridium	Clostridium_lactatifermentans	Clostridium_lactatifermentans
OTU_12121	0.95553691	0.8842282	Eubacterium	Eubacterium_eligens	Eubacterium_eligens
OTU_2352	0.95302013	0.1082215	Blautia	Unclassified	Blautia_Unclassified
OTU_1569	0.97902685	0.8213087	Clostridium	Unclassified	Clostridium_Unclassified
OTU_139	0.96979866	0.9991611	Clostridium	Unclassified	Clostridium_Unclassified
OTU_1006	0.41694631	0.9714765	Prevotella	Prevotella_copri	Prevotella_copri
OTU_1740	0.97147651	0.8708054	Unclassified	Unclassified	Unclassified
OTU_103	0.9295302	0.9437919	Unclassified	Unclassified	Unclassified
OTU_9966	0.97315436	0.7709732	Unclassified	Unclassified	Unclassified
OTU_8631	0.14177852	0.9446309	Prevotella	Prevotella_copri	Prevotella_copri
OTU_3603	0.68959732	0.9865772	Unclassified	Unclassified	Unclassified

OTU_1594	0.94211409	0.8011745	Eubacterium	Eubacterium_rectale	Eubacterium_rectale
OTU_32	0.98741611	0.9505034	Eubacterium	Eubacterium_eligens	Eubacterium_eligens
OTU_124	0.56459732	0.9731544	Unclassified	Unclassified	Unclassified
OTU_11498	0.23238255	0.989094	Anaerostipes	Anaerostipes_hadrus	Anaerostipes_hadrus
OTU_250	0.76090604	0.9790268	Unclassified	Unclassified	Unclassified
OTU_1221	0.95385906	0.9236577	Faecalibacterium	Faecalibacterium_prausnitzii	Faecalibacterium_prausnitzii
OTU_306	0.7533557	0.977349	Clostridium	Unclassified	Clostridium_Unclassified
OTU_120	0.9488255	0.4697987	Unclassified	Unclassified	Unclassified
OTU_1069	0.9647651	0.9966443	Clostridium	Clostridium_disporicum	Clostridium_disporicum
OTU_30	0.93791946	0.9127517	Clostridium	Clostridium_ruminantium	Clostridium_ruminantium
OTU_11429	0.97399329	0.7927852	Blautia	Unclassified	Blautia_Unclassified
OTU_8606	0.50838926	0.9572148	Unclassified	Unclassified	Unclassified
OTU_233	0.84563758	0.9463087	Eggerthella	Eggerthella_lenta	Eggerthella_lenta
OTU_6577	0.88255034	0.9488255	Blautia	Blautia_faecis	Blautia_faecis
OTU_64	0.77097315	0.9924497	Clostridium	Unclassified	Clostridium_Unclassified
OTU_86	0.92785235	0.9412752	Bacteroides	Bacteroides_thetaiotaomicron	Bacteroides_thetaiotaomicron
OTU_273	0.59647651	0.9689597	Clostridium	Unclassified	Clostridium_Unclassified
OTU_1280	0.94295302	0.2944631	Coprococcus	Coprococcus_catus	Coprococcus_catus
OTU_144	0.51677852	0.9521812	Parabacteroides	Unclassified	Unclassified
OTU_3372	0.70469799	0.9555369	Ruminococcus	Ruminococcus_bromii	Ruminococcus_bromii
OTU_12119	0.98322148	0.6417785	Unclassified	Unclassified	Unclassified
OTU_4016	0.94798658	0.3011745	Unclassified	Unclassified	Unclassified
OTU_6257	0.89597315	0.9379195	Roseburia	Roseburia_hominis	Roseburia_hominis
OTU_2604	0.96057047	0.8808725	Blautia	Blautia_faecis	Blautia_faecis
OTU_3860	0.94043624	0.817953	Blautia	Ruminococcus_obeum	Ruminococcus_obeum
OTU_470	0.97231544	0.4010067	Unclassified	Unclassified	Unclassified
OTU_21	0.96560403	0.965604	Anaerostipes	Anaerostipes_hadrus	Anaerostipes_hadrus
OTU_109	0.97483221	0.8372483	Clostridium	Clostridium_leptum	Clostridium_leptum
OTU_105	0.95637584	0.909396	Barnesiella	Barnesiella_intestinihominis	Barnesiella_intestinihominis
OTU_195	0.94379195	0.8833893	Unclassified	Unclassified	Unclassified
OTU_3501	0.55201342	0.9932886	Clostridium	Unclassified	Clostridium_Unclassified
OTU_9359	0.87248322	0.9387584	Faecalibacterium	Faecalibacterium_prausnitzii	Faecalibacterium_prausnitzii
OTU_168	0.87332215	0.9706376	Unclassified	Unclassified	Unclassified

OTU_3985	0.97063758	0.9110738	Faecalibacterium	Faecalibacterium_prausnitzii	Faecalibacterium_prausnitzii
OTU_34	0.9614094	0.8598993	Eubacterium	Eubacterium_hallii	Eubacterium_hallii
OTU_1356	0.73657718	0.9681208	Unclassified	Unclassified	Unclassified
OTU_5591	0.95050336	0.7063758	Unclassified	Unclassified	Unclassified
OTU_704	0.88422819	0.9404362	Unclassified	Unclassified	Unclassified
OTU_57	0.93959732	0.9244966	Eubacterium	Eubacterium_desmolans	Eubacterium_desmolans
OTU_16	0.96895973	0.9312081	Bifidobacterium	Bifidobacterium_longum	Bifidobacterium_longum
OTU_4302	0.96812081	0.8364094	Bifidobacterium	Bifidobacterium_longum	Bifidobacterium_longum
OTU_952	0.97567114	0.7206376	Eubacterium	Unclassified	Eubacterium_Unclassified
OTU_179	0.96644295	0.1635906	Unclassified	Unclassified	Unclassified
OTU_263	0.84647651	0.9614094	Veillonella	Veillonella_dispar	Veillonella_dispar
OTU_789	0.9840604	0.4572148	Parabacteroides	Parabacteroides_distasonis	Parabacteroides_distasonis
OTU_695	0.94714765	0.7986577	Blautia	Unclassified	Blautia_Unclassified
OTU_49	0.64765101	0.9395973	Eubacterium	Eubacterium_siraeum	Eubacterium_siraeum
OTU_178	0.9454698	0.9723154	Unclassified	Unclassified	Unclassified
OTU_152	0.9807047	0.738255	Unclassified	Unclassified	Unclassified
OTU_7093	0.55033557	0.9597315	Unclassified	Unclassified	Unclassified
OTU_2056	0.88842282	0.9823826	Unclassified	Unclassified	Unclassified
OTU_6936	0.01510067	0.9907718	Unclassified	Unclassified	Unclassified
OTU_1034	0.57969799	0.9496644	Unclassified	Unclassified	Unclassified
OTU_2345	0.65604027	0.9739933	Unclassified	Unclassified	Unclassified
OTU_1	0.85234899	0.942953	Eubacterium	Eubacterium_rectale	Eubacterium_rectale
OTU_4661	0.95889262	0.8473154	Clostridium	Unclassified	Clostridium_Unclassified
OTU_1914	0.36577181	0.9647651	Clostridium	Unclassified	Clostridium_Unclassified
OTU_98	0.86157718	0.9471477	Clostridium	Unclassified	Clostridium_Unclassified
OTU_1621	1	0.9169463	Unclassified	Unclassified	Unclassified
OTU_7618	0.96728188	0.7994966	Blautia	Blautia_faecis	Blautia_faecis
OTU_89	0.9135906	0.9538591	Eubacterium	Eubacterium_ventriosum	Eubacterium_ventriosum
OTU_162	0.89848993	0.9848993	Flavonifractor	Flavonifractor_plautii	Flavonifractor_plautii
OTU_45	0.94630872	0.9840604	Clostridium	Unclassified	Clostridium_Unclassified
OTU_287	0.94966443	0.7718121	Unclassified	Unclassified	Unclassified
OTU_23	0.90520134	0.9672819	Clostridium	Clostridium_ramosum	Clostridium_ramosum
OTU_11	0.87080537	0.9421141	Blautia	Blautia_luti	Blautia_luti

OTU_10307	0.96308725	0.7583893	Clostridium	Unclassified	Clostridium_Unclassified
OTU_22	0.89765101	0.9765101	Ruminococcus	Unclassified	Ruminococcus_Unclassified
OTU_1582	0.8716443	0.9454698	Unclassified	Unclassified	Unclassified
OTU_9740	0.38926174	0.954698	Unclassified	Unclassified	Unclassified
OTU_307	0.74412752	0.9479866	Clostridium	Unclassified	Clostridium_Unclassified
OTU_384	0.60067114	0.9815436	Unclassified	Unclassified	Unclassified
OTU_347	0.87416107	0.9857383	Clostridium	Unclassified	Clostridium_Unclassified
OTU_84	0.86493289	0.9832215	Unclassified	Unclassified	Unclassified
OTU_2336	0.95973154	0.7625839	Blautia	Unclassified	Blautia_Unclassified
OTU_4369	0.60151007	0.9563758	Clostridium	Clostridium_aldenense	Clostridium_aldenense
OTU_65	0.88674497	0.9580537	Unclassified	Unclassified	Unclassified
OTU_719	0.94127517	0.7365772	Clostridium	Unclassified	Clostridium_Unclassified
OTU_414	0.34312081	0.9605705	Mogibacterium	Unclassified	Mogibacterium_Unclassified
OTU_5014	0.98489933	0.9697987	Blautia	Ruminococcus_torques	Ruminococcus_torques
OTU_5819	0.94463087	0.7197987	Clostridium	Clostridium_sporosphaeroides	Clostridium_sporosphaeroides
OTU_125	0.9966443	0.454698	Unclassified	Unclassified	Unclassified
OTU_10029	0.76677852	0.9622483	Unclassified	Unclassified	Unclassified
OTU_5862	0.97734899	0.579698	Actinomyces	Actinomyces_lingnae	Actinomyces_lingnae
OTU_945	0.85067114	0.9807047	Unclassified	Unclassified	Unclassified
OTU_447	0.99580537	0.9152685	Unclassified	Unclassified	Unclassified
OTU_375	0.95469799	0.7919463	Clostridium	Clostridium_methylpentosum	Clostridium_methylpentosum
OTU_1637	0.99244966	0.8733221	Unclassified	Unclassified	Unclassified
OTU_11425	0.96392617	0.3833893	Blautia	Ruminococcus_torques	Ruminococcus_torques
OTU_8952	0.98825503	0.9874161	Coprococcus	Coprococcus_comes	Coprococcus_comes
OTU_291	0.95134228	0.9958054	Collinsella	Collinsella_aerofaciens	Collinsella_aerofaciens
OTU_52	0.98657718	0.966443	Eubacterium	Eubacterium_ramulus	Eubacterium_ramulus
OTU_3602	0.95218121	0.9161074	Unclassified	Unclassified	Unclassified
OTU_10880	0.95805369	0.8716443	Blautia	Unclassified	Blautia_Unclassified
OTU_14	0.98154362	1	Collinsella	Collinsella_aerofaciens	Collinsella_aerofaciens
OTU_27	0.97818792	0.9756711	Coprococcus	Coprococcus_comes	Coprococcus_comes
OTU_9293	0.99748322	0.9144295	Blautia	Ruminococcus_torques	Ruminococcus_torques
OTU_53	0.99077181	0.9949664	Blautia	Ruminococcus_torques	Ruminococcus_torques
OTU_1248	0.99161074	0.9588926	Unclassified	Unclassified	Unclassified

OTU_76	0.99916107	0.988255	Dorea	Dorea_formicigenerans	Dorea_formicigenerans
OTU_7369	0.99328859	0.9748322	Dorea	Dorea_formicigenerans	Dorea_formicigenerans

SUPPLEMENTARY TEXT 4

DIET-RESPONSIVE TAXA ARE NOT SPECIFIC TO NATIONALITY

We next checked if the diet-associated taxa differed across the nationalities. As noted above different nationalities were characterized by specific gut microbiome composition at baseline (**figure 1b; Supplementary figure 4a-b**) and different dietary adherence scores (Netherland and UK having significantly lower scores, followed by Italy with Poland and France having the highest) (**Supplementary figure 7a**) (as also reported by previous studies on this cohort) (18, 19). If nationality-specific differences in diet-associated taxa existed, the performance of the prediction models might vary across nationalities. However, despite differences in the baseline gut microbiome compositions, there was no significant difference in the performance of the model (mean squared errors) for the different nationalities except for Netherlands (where the error rate was significantly high) (**Supplementary figure 7b**). This indicates that the identified diet-associated taxa were similar across most of the nationalities. The higher error-rate for the Netherlands could either be a consequence of lower adherence scores or a different set of diet-responsive taxa in these individuals. We tested the latter possibility by creating two different versions of iterative Random Forest models (two-fold cross validation) for Dutch subjects at baseline. While one version was created by using only the 129 diet-associated markers, the second version was built upon all OTUs besides the diet-associated markers. We observed that the iterative models built using only the 129 diet-associated markers had significantly higher correlation and significantly lower mean squared errors as compared to those obtained using all OTUs besides the diet-associated taxa (**Supplementary figures 7c-d**). This indicated that the diet-responsive taxa were not specific to nationality.

SUPPLEMENTARY TEXT 5

VALIDATING THE ASSOCIATION OF THE DIET-RESPONSIVE TAXA WITHIN THE INTERVENTION AND CONTROL COHORTS AS WELL AS WITHIN INDIVIDUALS WITH VARYING ADHERENCE TO THE MED-DIET

There were 1224 microbiota datasets corresponding to 612 individuals having matched microbiome profiles for both the baseline and the follow-up time points. To further verify their association with the MedDiet adherence, we checked the variation of the relative abundances of these OTU-groups across an entire adherence landscape. For this we arranged the microbiota data (of the individuals) from the entire intervention study in increasing order of their adherence to the diet, and subsequently divided them into five equally sized overlapping windows (of increasing adherence scores; five overlapping windows of 204 samples with an overlap of 102 samples). Adopting such a window approach would illustrate the gradual transitions of specific changes across an entire adherence landscape (after eliminating variations caused due to specific samples). As expected, profiling the abundance variation of the two taxa groups across the windows identified a progressive increase of the DietPositive taxa (Kruskal Wallis H-test P-value < 5e-4) and a concomitant decrease of the DietNegative taxa (Kruskal Wallis H-test P-value < 3.2e-7) with increasing adherence to the Mediterranean diet (**Supplementary figure 8**). Performing this window-based analysis separately within the baseline and final time points also revealed the same pattern (**Supplementary figure 8**). We then checked whether the positive and negative associations of the DietPositive and DietNegative taxa in the intervention cohort were also reflected in the across time-point (final to baseline) changes in dietary adherence. For each of the diet-associated markers (i.e. the OTUs), we computed the log fold change in the gain/loss ratios (the number of individuals in whom an OTU is more abundant across the time-points divided by the number of individuals in whom it is decreased) in the intervention cohort with respect to the control cohort. We observed that for the DietPositive taxa, the intervention to control log fold difference of the gain/loss ratios were positive (indicating that the changes were more positive in the intervention cohort as compared to the controls) and significantly higher (Mann-Whitney U test P < 1.3e-4) than those obtained for the DietNegative taxa which were negative (indicating a decrease

across time-points in the intervention cohort as compared to the controls) (**figure 2c**). To further profile the changes in the abundance of the markers across individuals with varying degrees of changes in their adherence to the diet, we divided them into three equal tertiles, namely 'High Adherence', 'Medium Adherence' and 'Low Adherence' in decreasing order of their change in adherence across time-points. The abundance changes of the two groups of markers (DietPositive and DietNegative) were then profiled across the three groups separately. As expected, while the DietPositive OTUs had a significantly positive change in the High Adherence as compared to the Low Adherence individuals, an exactly opposite trend was observed for the DietNegative markers (**figure 2d**). These findings suggest that the associations of the specific taxa with diet are stable across cohorts as well as across the changes between time-points.

SupplementaryTable3: Number of OTUs belonging to each iBBiG module along with the major genera (relative abundance greater than 1% after removing the unclassified OTUs within each module

iBBiG_Module	Number of OTUs	Major Genera
a	291	Lachnospiracea_incertae_sedis, Blautia, Clostridium_IV, Coprococcus, Bacteroides, Faecalibacterium, Roseburia, Oscillibacter, Gemmiger, Ruminococcus, Alistipes, Bifidobacterium, Dorea, Flavonifractor, Clostridium_XlVa, Clostridium_XI, Parabacteroides, Streptococcus
b	153	Clostridium_IV, Coprococcus, Lachnospiracea_incertae_sedis, Oscillibacter, Faecalibacterium, Sporobacter, Flavonifractor, Gemmiger, Prevotella, Acetanaerobacterium, Bacteroides, Blautia, Butyricimonas, Clostridium_XlVa, Clostridium_XlVb, Enterorhabdus, Erysipelotrichaceae_incertae_sedis, Haemophilus, Methanobrevibacter, Parasutterella, Pseudobutyrivibrio, Ruminococcus, Slackia
c	90	Clostridium_IV, Alistipes, Oscillibacter, Erysipelotrichaceae_incertae_sedis, Flavonifractor, Sporobacter, Gemmiger, Ruminococcus, Acetanaerobacterium, Anaerofilum, Asaccharobacter, Blautia, Clostridium_XlVa, Eggerthella, Gordonibacter, Pseudoflavonifractor, Roseburia
d	99	Lachnospiracea_incertae_sedis, Bacteroides, Blautia, Coprococcus, Faecalibacterium, Ruminococcus, Alistipes, Clostridium_XlVb, Gemmiger, Anaerostipes, Barnesiella, Butyricimonas, Clostridium_IV, Dorea, Haemophilus, Lactococcus, Oscillibacter, Parasutterella, Sporacetigenium
e	142	Lachnospiracea_incertae_sedis, Blautia, Roseburia, Clostridium_XlVa, Bacteroides, Dorea, Bifidobacterium, Actinomyces, Clostridium_IV, Coprococcus, Gemmiger, Streptococcus, Anaerostipes, Clostridium_XI, Clostridium_XVIII, Faecalibacterium, Flavonifractor, Ruminococcus
f	66	Lachnospiracea_incertae_sedis, Blautia, Clostridium_IV, Faecalibacterium, Ruminococcus, Actinomyces, Alistipes, Anaerostipes, Bacteroides, Clostridium_XlVa, Coprococcus, Dorea, Eggerthella, Erysipelotrichaceae_incertae_sedis, Flavonifractor, Gemmiger, Gordonibacter, Granulicatella, Rothia

Supplementary Table 4: Number of samples classified to various iBBiG modules

iBBiG Modules	Number of Samples
a	1236
b	420
d	294
c	262
f	65
e	44

SUPPLEMENTARY TEXT 6

RESULTS OF THE APPLICATION OF THE iBBiG APPROACH ON THE NU-AGE DATASET

The application of the iBBiG approach on the NU-AGE data identified 6 overlapping taxonomic modules that had a high mutual co-occurrence (obtained by maximizing the internal entropy) within the dataset. These were referred to as modules 'A' to 'F' (**Supplementary figure 9a**). Based on their detection trends in the overlapping modules, the OTUs could be classified as belonging to either a single (e.g. module A) or a combination of any of the six modules. This resulted in 36 OTU classifications (including one 'not classified' group). In a similar manner, a given sample could be classified into one of 15 classifications (and one 'not classified' group) based on the detection of the various modules in that sample. The classifications of each OTU and sample obtained in the iBBiG approach is listed in **Supplementary Tables 3 and 4**. Each of the six modules were characterized by different number of OTUs, specific trends of prevalence across individuals, as well as distinct patterns of taxonomic composition (**Supplementary figures 9b-c**). We also identified differential associations of each of these modules with frailty, especially with modules B and D being significantly enriched in the individuals with reduced frailty from baseline to post intervention, as compared to the module C, which was enriched in those with increasing frailty (**Supplementary figures 9d**). This indicates module 'C' to be similar to the long-stay-like modules we identified in ELDERMET individuals using the iBBiG approach(16). However, module 'C was not only associated with a significant enrichment in individuals with increased frailty, but also an increase in representation of the set of DietNegative OTUs (**Supplementary figure 9d**). The observation that adherence to the diet could specifically select against taxa associated with frailty indicates the likelihood that the Mediterranean diet successfully modulated the gut microbiome in a manner predicted to be negatively associated with frailty. A major objective of the NU-AGE dietary intervention was the reduction of frailty and inflamm-ageing in the elderly. Therefore, we next investigated in detail the association of adherence-associated taxa with frailty as well as with the inflammation status of the individuals.

SUPPLEMENTARY TEXT 7

ACROSS TIME-POINT CHANGES OF THE DIET-RESPONSIVE TAXA IN INDIVIDUALS WITH VARIED CHANGES IN THEIR FRAILTY STATUS

Based on the changes in their frailty status across time-points, the individuals across the cohort could be divided into three groups, namely those with 'Reduced Frailty', 'No change in frailty' and 'Increased Frailty'. We then investigated the across time-point changes in these taxa. To measure whether the above trends were also reflected in the across time-point changes, for each OTU, we computed the effect-size of the time-point changes between the individuals with reduced frailty as compared to the other two groups (**See Methods**). A positive effect size change would indicate that the taxa show more positive change (that is either an increase or a relative lower decrease) in their abundance across time-points in individuals with reduced frailty (as compared to those with no change or increase in frailty), and vice-versa. In this regard, while the diet-enriched (that is the DietPositive) taxa showed significantly positive changes in the individuals with reduced frailty (as compared to the other two groups), the DietNegative group showed the opposite trend (**Supplementary figure 10b**). These findings further affirm our earlier observation of the depletion of the specific frailty-associated iBBiG module 'C' which was observed to have a negative association with diet as well as the notable increase of frail individuals in the control group. In line with these observations, in the control group, we observed a marginally significant increase (as compared to the intervention group) during the intervention period in the proportion of individuals with increased frailty (Fishers' Test $P < 0.06$; **Supplementary figure 10c**).

Supplementary Table 5: List of A) Frailty and Cognitive Function associated measures and B) Cytokines used for performing association analysis with the diet modulated microbiome components.

(A)

Category	Measure	Description	Directionality
Physical Frailty	Hand Grip Strength	Mean hand grip strength of the dominant hand (3 trials)	Negatively associated with frailty
	Gait Speed (Fastest time)	Gait Speed Fastest Time taken (2 trials)	Positively associated with frailty
	Fried Score	Fried Score for computing frailty	Positively associated with frailty
Cognitive Functioning	Geriatric Depression Score	Score for measuring geriatric depression (higher values indicate depression)	Negatively associated with cognitive function
	MMSE	Mini Mental State Examination (Scores range 0-30). Higher scores indicate better cognitive function	Positively associated with cognitive function
	BabCok Memory Score	Bab Cock Score for immediate recall	Positively associated with cognitive function
	CAMDEX-Q Scores	Cambridge Examination of Mental Disorders	Positively associated with cognitive function
	Constructional Praxis	CERAD Battery Total Score on Constructional Praxis	Positively associated with cognitive function
	Verbal Fluency	CERAD Battery Total Score on Verbal Fluency Categories	Positively associated with cognitive function

(B)

Cytokines

IL-13

Pentraxin-3

Adiponectin

TNF-A

G-CSF

IL-8

Ghrelin

IL-6

Resistin

hsCRP

IL-1b
TGF-b1
IL-4
IL-17
IL-1ra
IL-2
Leptin
sTNF-R1
IFN-g
IL-7
IL-12 p70
sGP130
GM-CSF
IL-5
sIL-6ra
MIP-1b
IL-10
MCP-1-MCAF
IL-18
sTNF-R2
IL-17a

SUPPLEMENTARY TEXT 8

ASSOCIATION OF MICROBIOME INDEX WITH THE DIFFERENT DIETARY COMPONENTS

Next, given that we calculated the microbiome index as a single value index providing a quantitative summary of the abundance patterns of the diet-associated markers (the higher the value, the higher the abundance of Diet-Positive taxa and the lower the abundance of Diet-Negative taxa, and vice-versa), as a sanity-check, it is important to validate that the calculated microbiome index captured the association patterns of the individual diet-associated marker OTUs. For this, we repeated the analysis performed earlier for the individual marker OTUs (**Supplementary figure 12**) on the overall microbiome index (after adjustment for confounders) (**Supplementary figure 13**).

SUPPLEMENTARY TEXT 9

ASSOCIATION OF DIET-ASSOCIATED MICROBIOME TAXA WITH DISEASE PATHOPHYSIOLOGIES, POLY-PHARMACY, AND OTHER HOST-FACTORS

The dataset included 11 diseases containing at least three (of 612) diseased subjects at the baseline. We first investigated the effect of these diseases on the diet-associated taxa at the baseline. Nine of the 11 diseases were associated with lower microbiome indices, significantly so for diabetes, heart attack and inflammatory disorders ($P < 0.05$) and marginally significant for Cancer ($P < 0.097$) (**Supplementary table 6**; **Supplementary figure 14 a-d**). Individuals with multiple diseases had significantly lower microbiome indices and significantly lower ratios of DietPositive to DietNegative taxa abundances compared to those with single or no disease, indicating that the diet-favoured microbiome components are negatively associated with disease at baseline (**Supplementary figure 14e-f**). However, when we examined partial spearman correlations at baseline, the pattern of association of microbiome index with seven of the 10 inflammatory markers and frailty indices (identified in **figure 4**) remained invariant even after taking into account all confounders including age, BMI, gender, poly-pharmacy and different disease pathophysiologies (**Supplementary figure 15a**). All the above associations were retained (except for leptin) even after considering age, gender, BMI and poly-pharmacy (as confounders) across both the baseline and follow-up time points, further supporting the hypothesis that it is the microbiome response that is linked to the above measures (even after adjusting for all host associated confounding factors) rather than dietary adherence alone (**Supplementary figure 15b**).

SUPPLEMENTARY TEXT 10

ASSOCIATION OF CHANGES (BETWEEN THE FINAL AND BASELINE TIME-POINTS) IN MICROBIOME INDEX WITH DIFFERENT MEASURES.

To further illustrate the links between diet, microbiome, and health, we investigated the associations of the across time-point changes of the various measures with the change in diet and microbiome. The change in microbiome index was not associated with either baseline dietary adherence scores (linear regression R=-0.034; P < 0.31) or the 12-month dietary adherence scores (linear regression R = 0.05; P < 0.12) (**Supplementary figure 15a, b**). We first performed an in-depth investigation of the association of the across-time-point (follow-up to baseline) changes in cytokine levels for each individual with the corresponding change in microbiome indices. Cumulated levels of anti-inflammatory cytokines were calculated as the summed ranked abundances of the anti-inflammatory cytokines (IL-10, IL-4, IL-5 and IL-1ra). Ratio of hsCRP levels to anti-inflammatory cytokine levels was calculated as the ratio of the ranked abundance of hsCRP to the cumulated levels of anti-inflammatory cytokines (calculated as above). Then for each cytokine (or cytokine ratio), the changes were calculated as the differences in levels between the follow-up and the baseline time-points. For inflammatory markers like hsCRP, MCP1-MCAF, Resistin, positive changes in microbiome indices were associated with significant negative changes in the levels of these cytokines (**Supplementary figure 15c**). For other inflammatory cytokines like IL-17, IL-6, MIP-1b, etc, the associations were still negative, although not significant. An exact opposite trend was observed for the anti-inflammatory cytokine IL-10, where positive changes in microbiome indices were associated with significant positive changes in the levels of this cytokine. As a consequence, positive changes in microbiome indices were associated with negative changes in the hsCRP to anti-inflammatory cytokine levels (**Supplementary figure 15c**).

Additionally, a pairwise regression approach was also used to identify associations between microbiome response, adherence score changes and the identified measures of frailty, cognitive function and inflammation. Given any two measures, we performed linear regressions of the measures with each of the scores using country and age as confounders. We did not use FDR correction at this stage as we were investigating associations with specific measures. Linear relationships with P-values less than 0.05 and between 0.1 and 0.05 were identified as being significant and trend, respectively.

The significant associations identified from this analysis are illustrated in **Supplementary figure 15d)** As expected, microbiome response was positively associated with dietary adherence changes. However, it was this increased microbiome response that displayed positive associations with reduced frailty (Reduced Fried Score), improved cognitive function (BabCock Memory Score) and negative associations with inflammation (hsCRP and another pro-inflammatory marker MIP-1b). The adherence score change, by itself, did not have any significant association (with exception of a negative association with MIP-1b). The above results clarify the relationships between diet, microbiome and improved life-status. Change in adherence (that is increasing adherence to a Mediterranean diet) is likely to modulate specific components of the microbiome. It is this microbiome response, when induced, that is associated with reduced frailty and reduced inflammation. However, at each interaction point, there may be exceptions.

Supplementary Table 6: A. Number of subjects with gut microbiome profiles at baseline that belong to the different disease categories (no_disease refers to those individuals who were not identified with any disease symptoms). B. Results of the Mann-Whitney test based comparative analysis of the diet-associated microbiome indices for the individuals with different diseases with control (no_disease type) individuals at baseline

(A)

Disease_Type	Number of Subjects
no_disease	324
high_cholesterol	183
thyroid	61
diabetes	30
cancer	27
hypertension	21
food_allergy	15
respiratory_disease	14
swollen_ankle	10
kidney_disease	6
heart_disease	5
inflammation	3

(B)

Disease Type	Diet-Associated Microbiome Index	
	Direction	Mann-Whitney P-value
hypertension	Lower In Disease	0.005346
diabetes	Lower In Disease	0.005646
inflammation	Lower In Disease	0.02727
cancer	Lower In Disease	0.097
respiratory_disease	Lower In Disease	0.1642
food_allergy	Lower In Disease	0.2569
swollen_ankle	Lower In Disease	0.3086
heart_disease	Lower In Disease	0.3212
high_cholesterol	Lower In Disease	0.4621
kidney_disease	Higher In Disease	0.7945
thyroid	Higher In Disease	0.8938