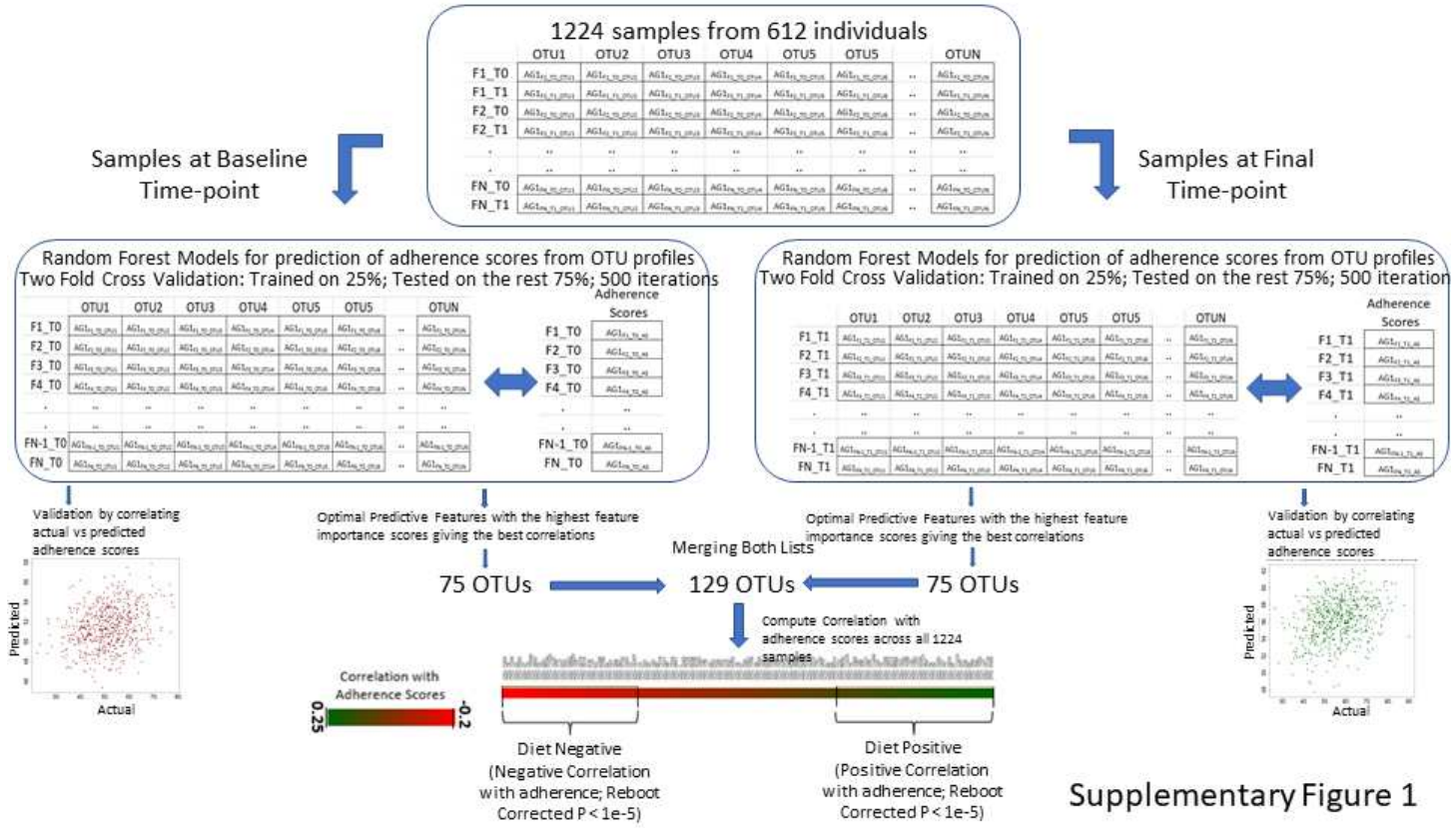
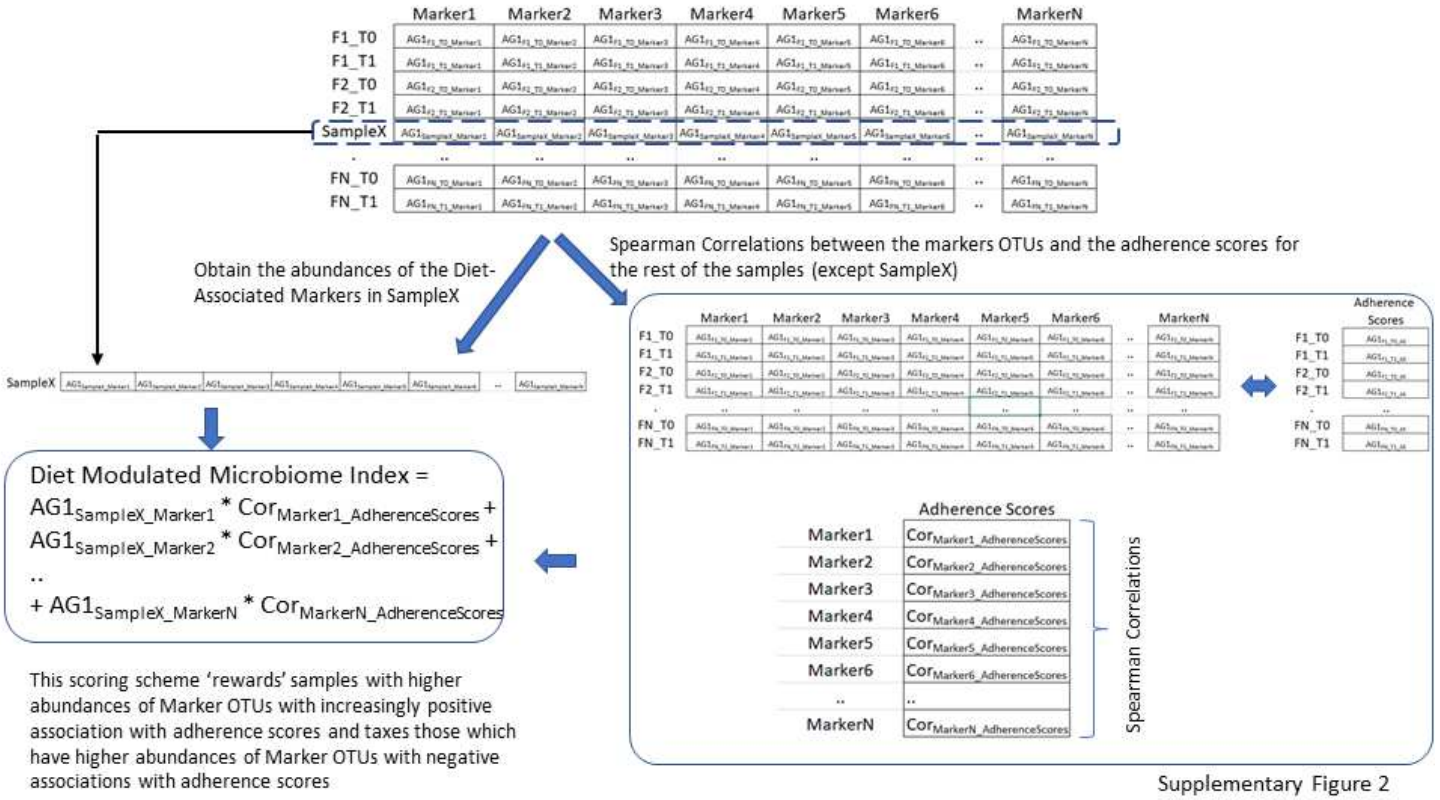
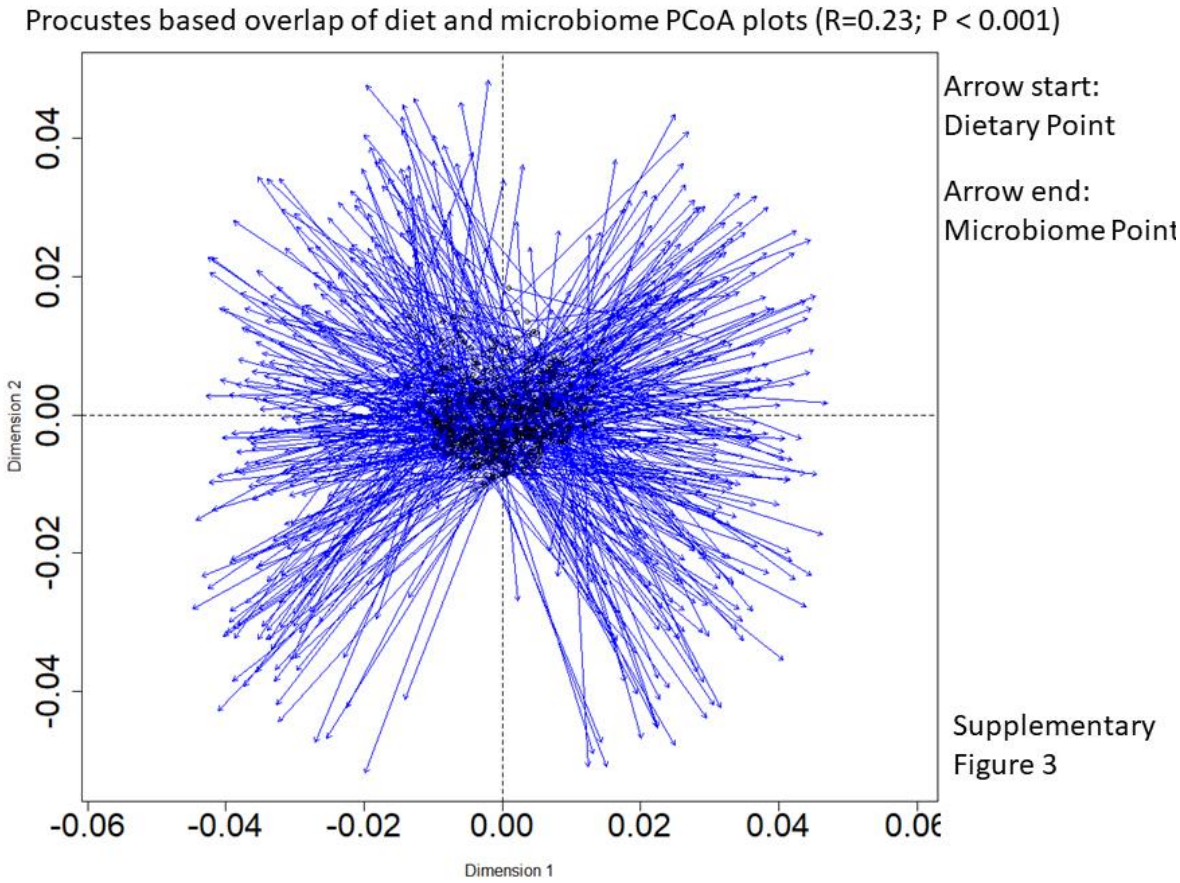
SUPPLEMENTARY FIGURES



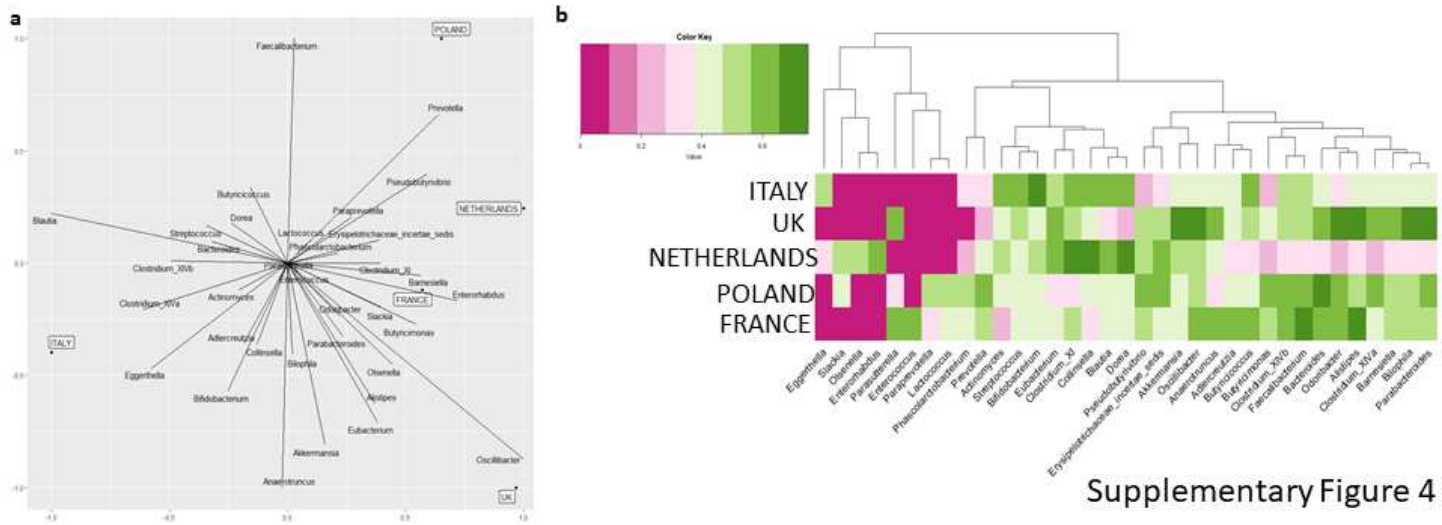
Supplementary Figure 1: Pictorial workflow describing the Random Forest based prediction of adherence scores from the microbiome abundance profiles and the identification of adherence score associated markers including the DietPositive and DietNegative markers.



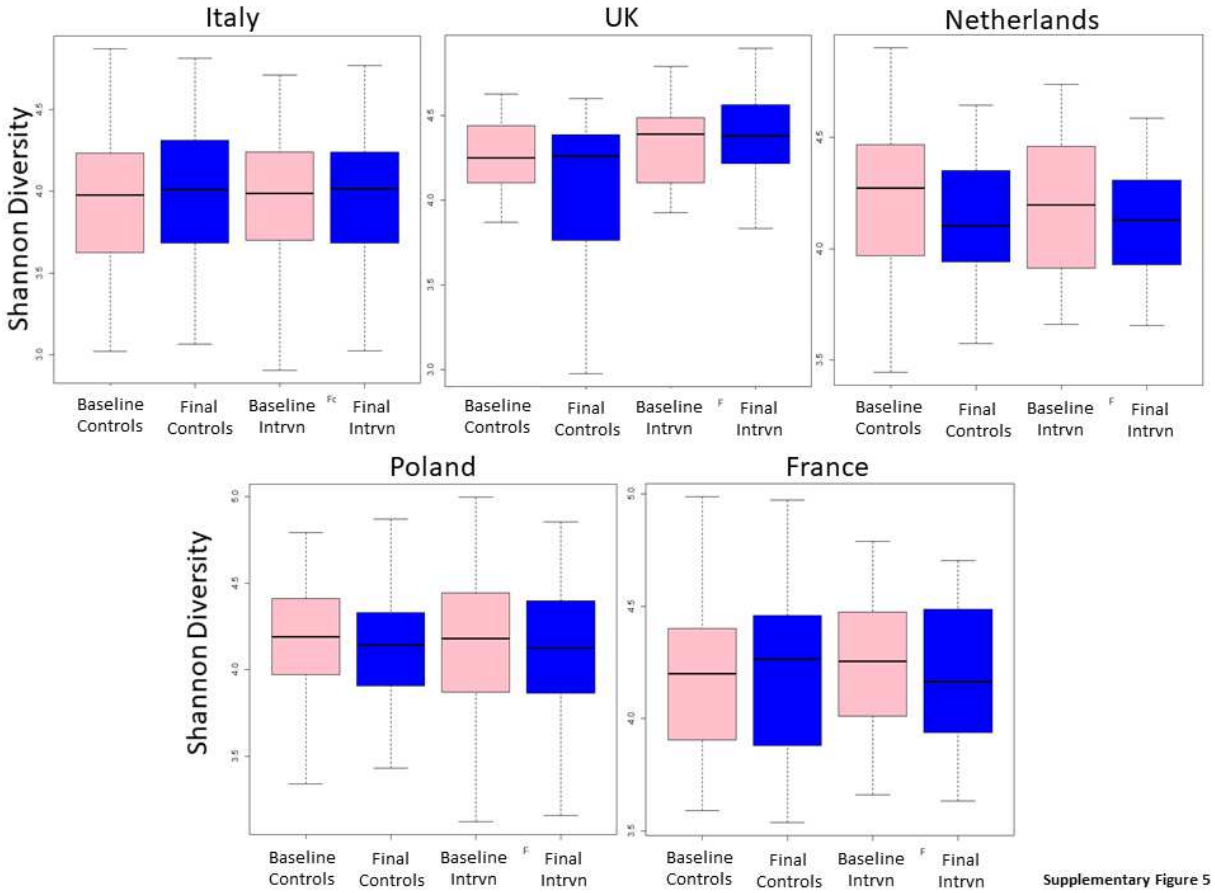
Supplementary Figure 2: Pictorial representation of the methodology of computation of the microbiome indices using the leave-one-out strategy.



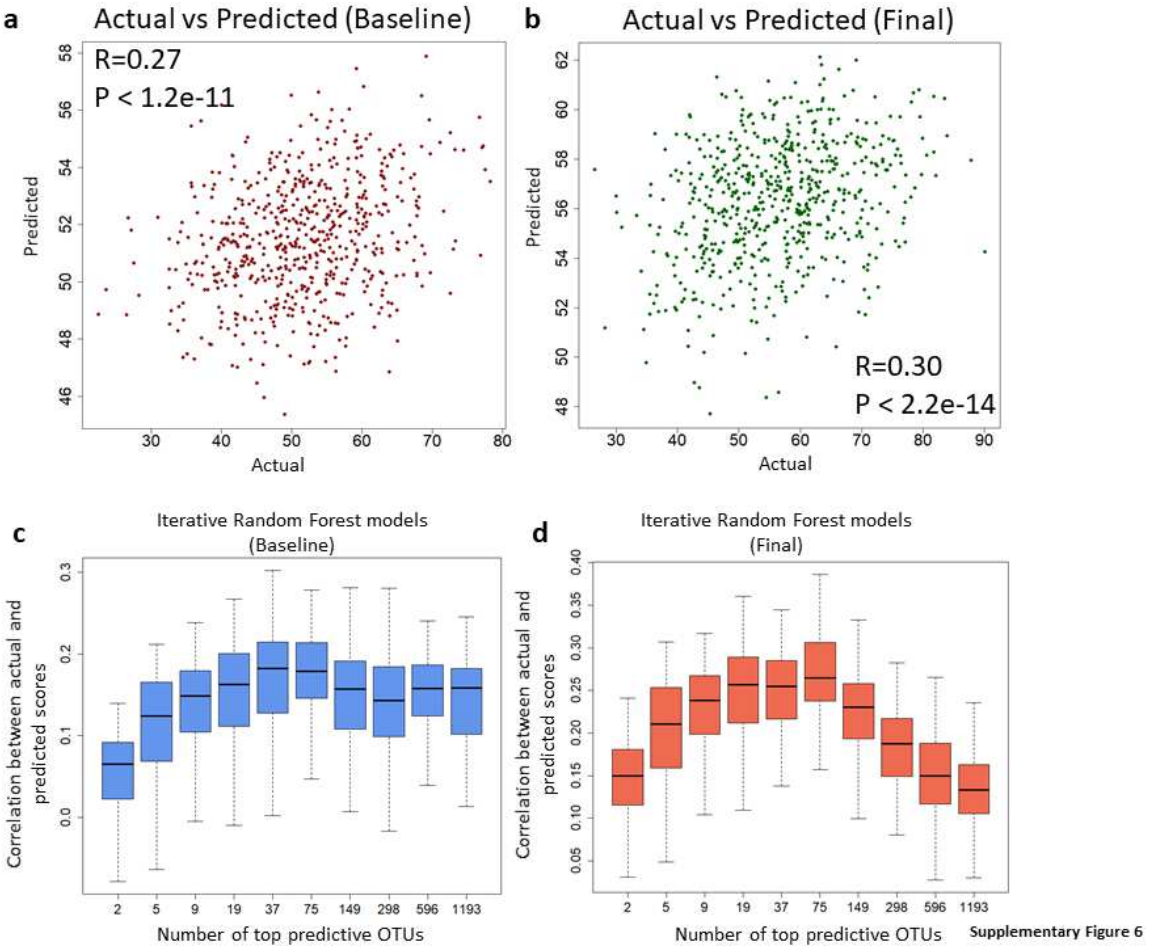
Supplementary Figure 3: Procustes plot showing the relative movement of the samples between the Principal Coordinate Analysis (PCoA) plots of the dietary and the microbiome profiles.



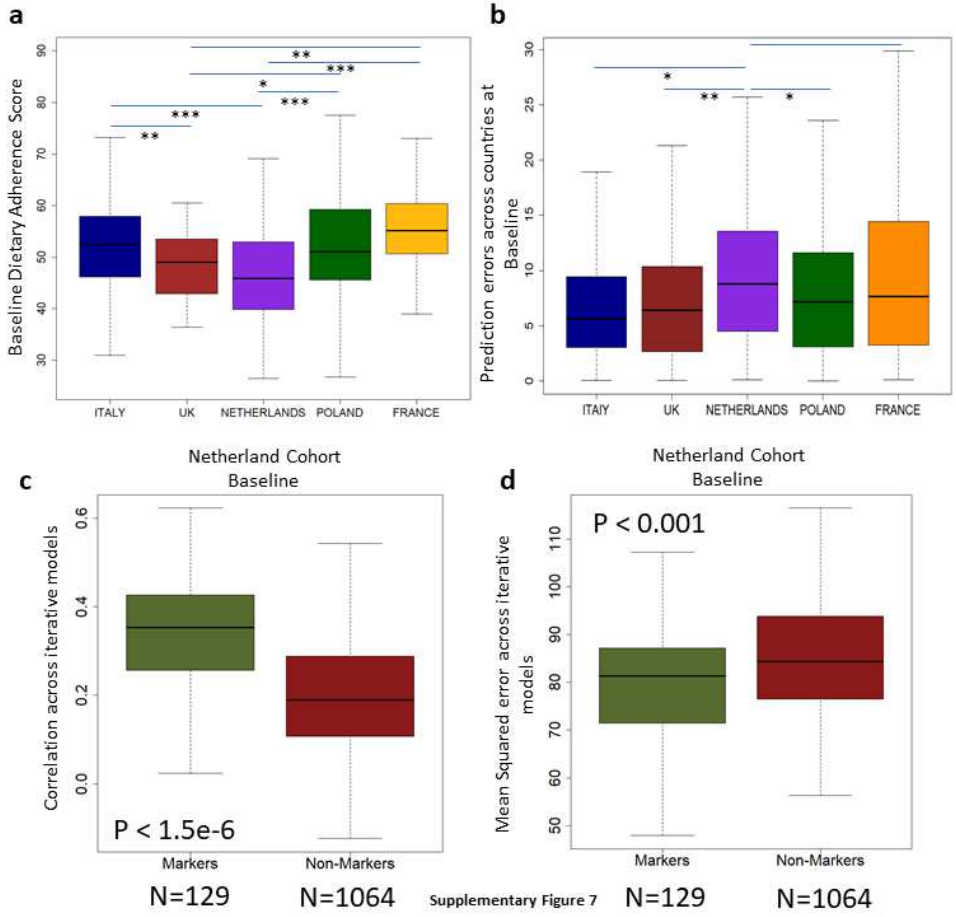
Supplementary Figure 4: Association of different genera with the specific nationalities shown as **a.** Plotted based on the weights of their association with the Principal Coordinate Analysis (PCoA) axes (as in Fig 1b). **b.** Plotted as heatmap showing the nationality-specific median abundances. Only 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.



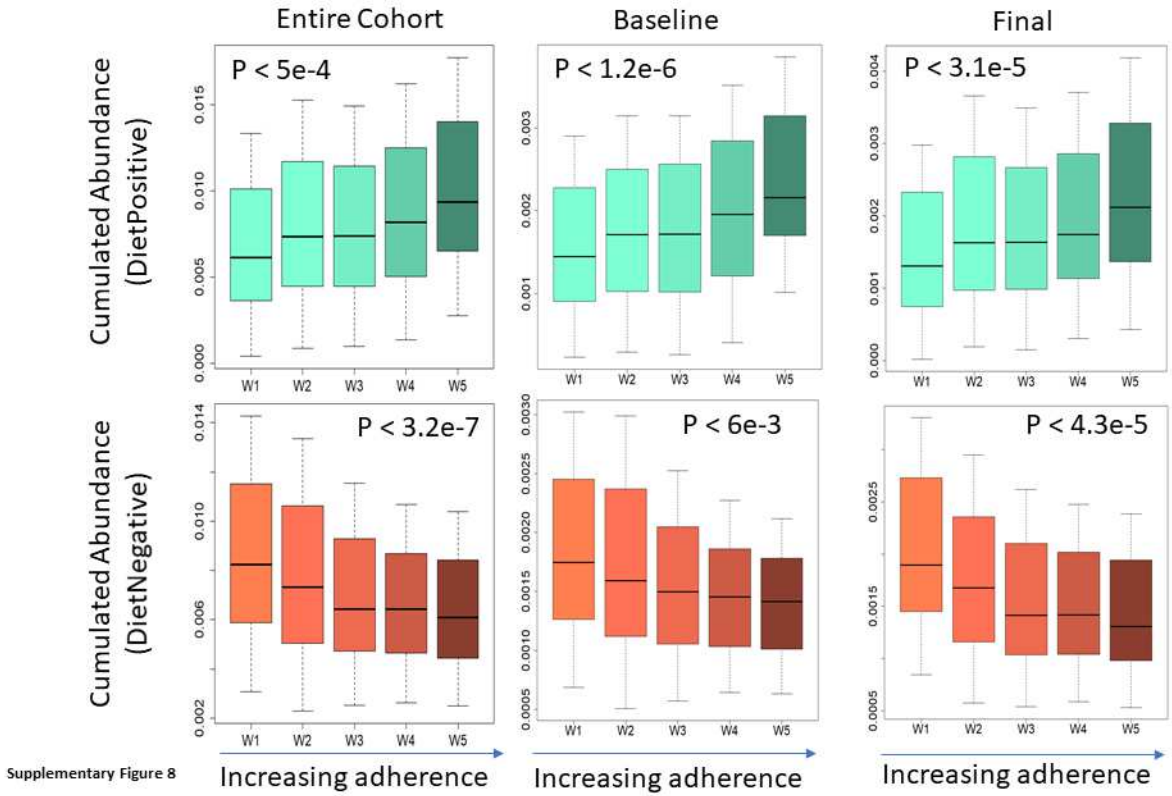
Supplementary Figure 5: Shannon diversity indices of the microbiota at baseline and Final time points for subjects in the intervention and control cohorts in the five different countries.



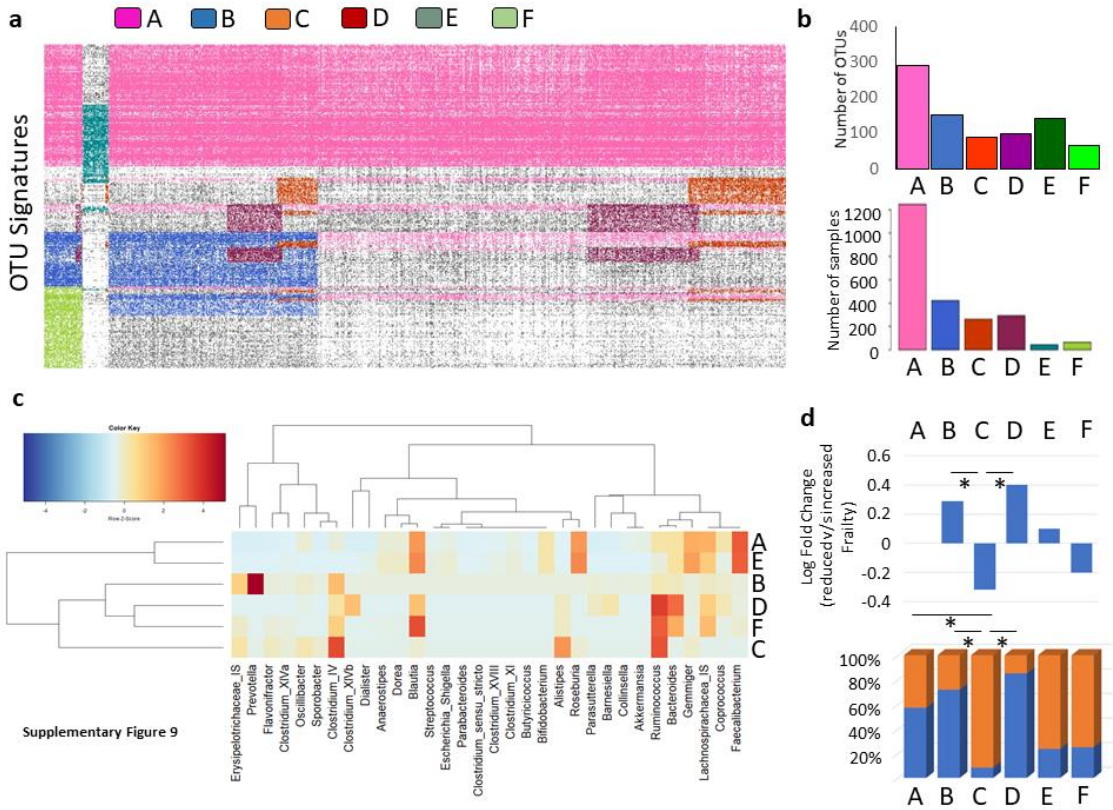
Supplementary Figure 6: Relationship between the Random Forest predicted and the actual dietary adherence scores for **a.** Baseline and **b.** Final time points. Boxplots showing the variation of the correlations between the actual and predicted dietary adherence scores obtained using iterative Random 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.



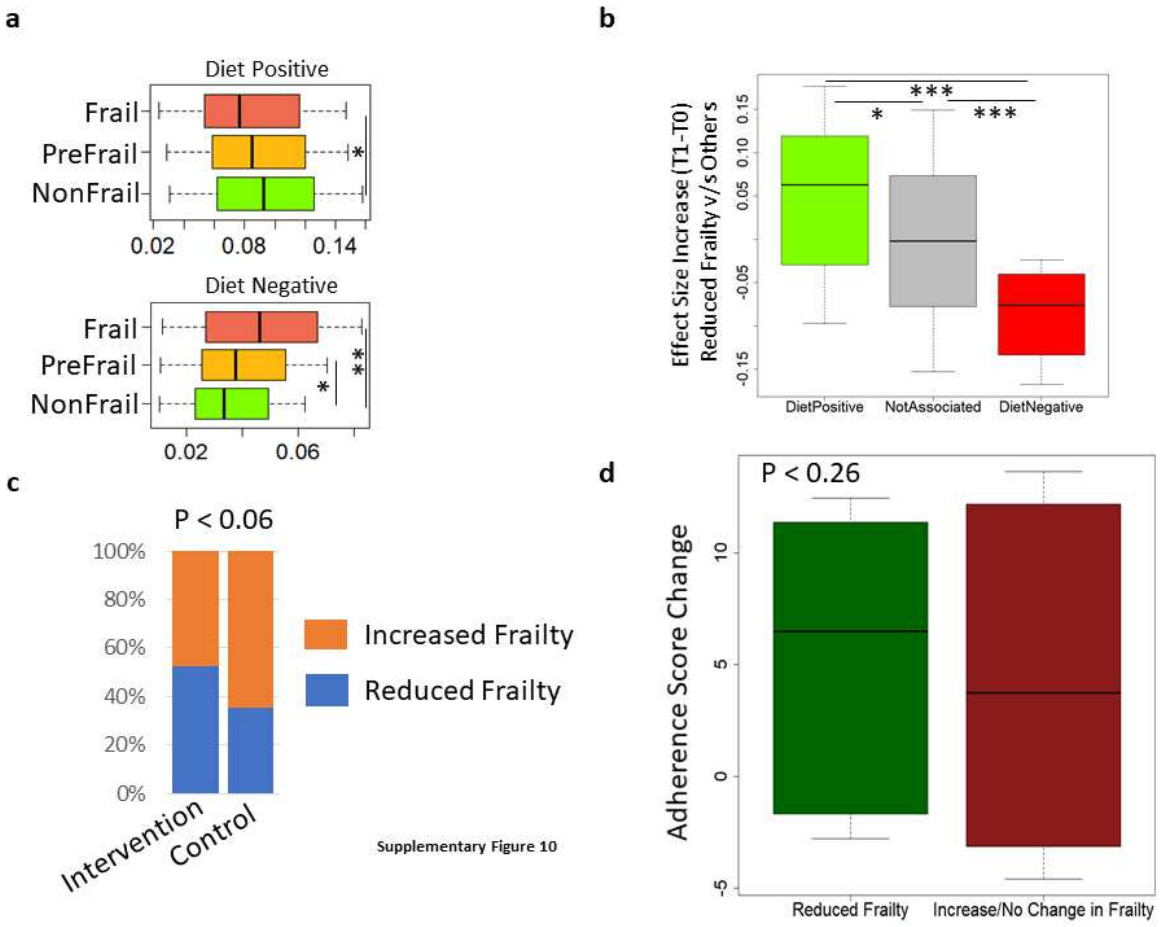
Supplementary Figure 7: a. Baseline Dietary Adherence Scores for the five different nationalities. **b.** Mean squared dietary adherence prediction errors for samples from five different nationalities **c.** Correlations and **d.** Mean squared errors between actual and predicted dietary adherence scores 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').



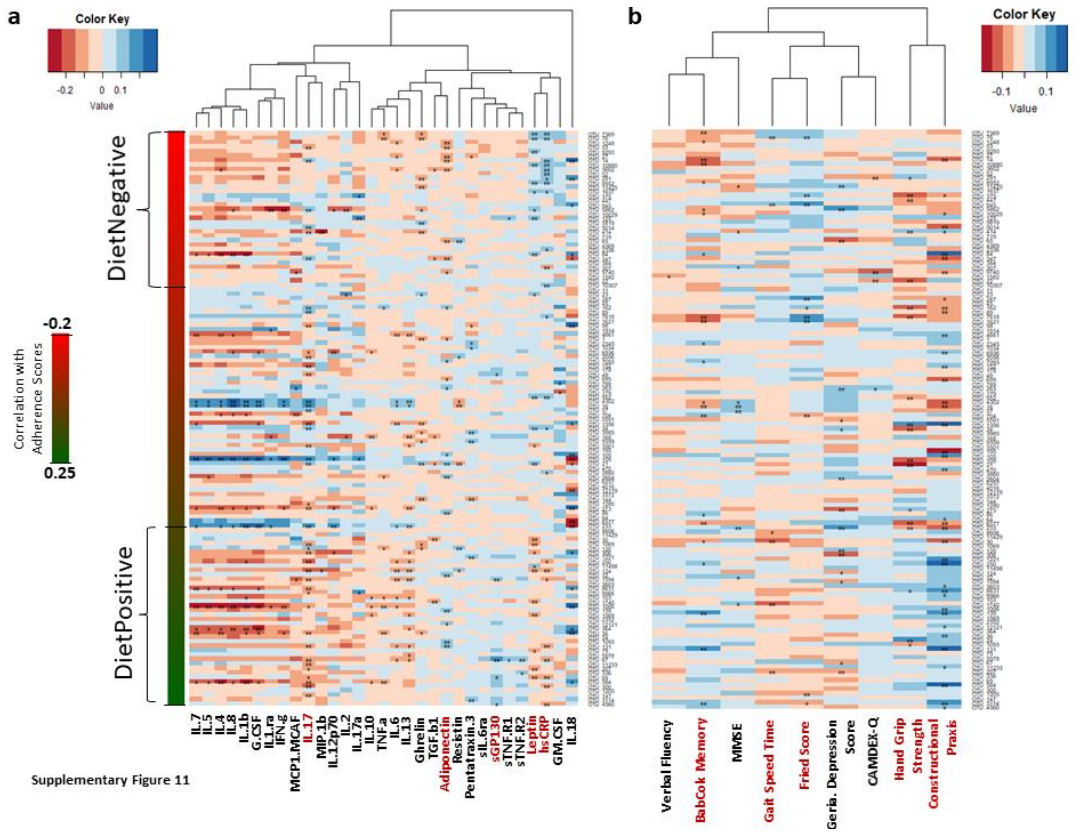
Supplementary Figure 8: Boxplots showing the variation of the cumulated abundances of the DietPositive and the DietNegative OTUs across overlapping windows of subjects with increasing adherence to the diet across the entire cohort as well as within the samples for the baseline and post-intervention time 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.

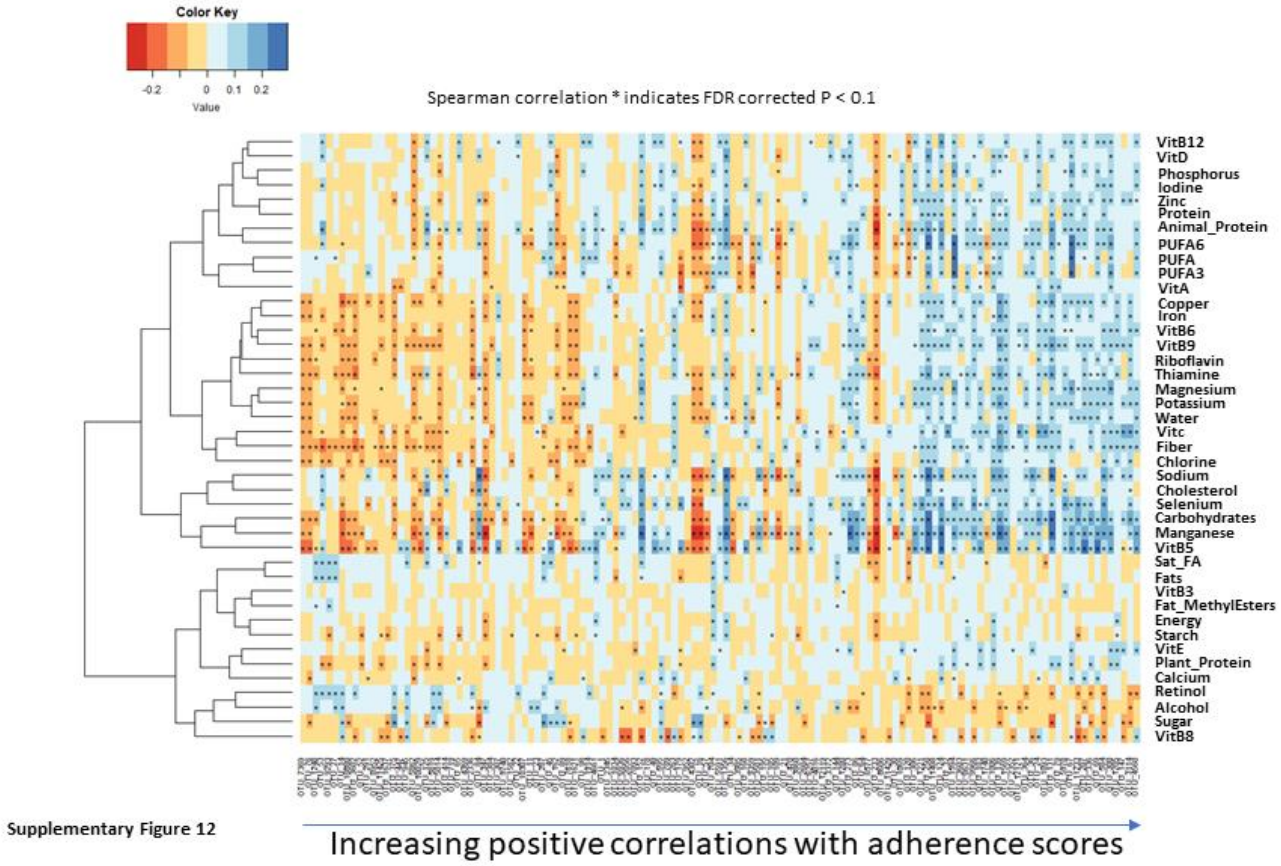


Supplementary Figure 10: **a.** Boxplots showing the variation of the cumulated abundance of the DietPositive and DietNegative OTUs in the Frail, Pre-Frail and the Non-Frail individuals. **b.** Variation of the effect-size differences (cohens' d) of the across time-point changes of the DietPositive OTUs, 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.

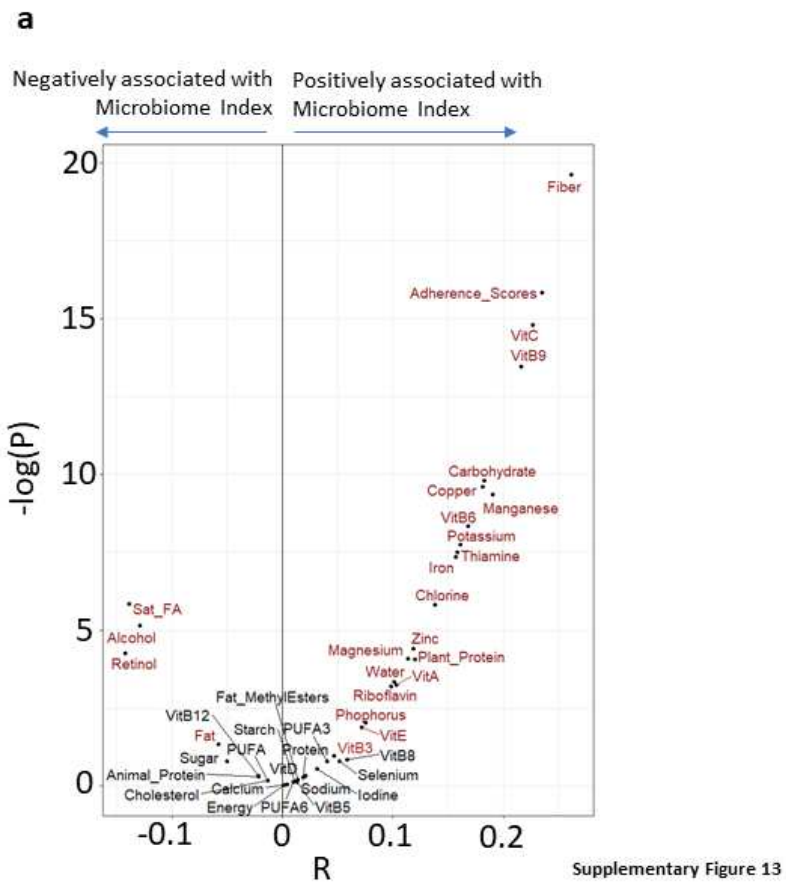


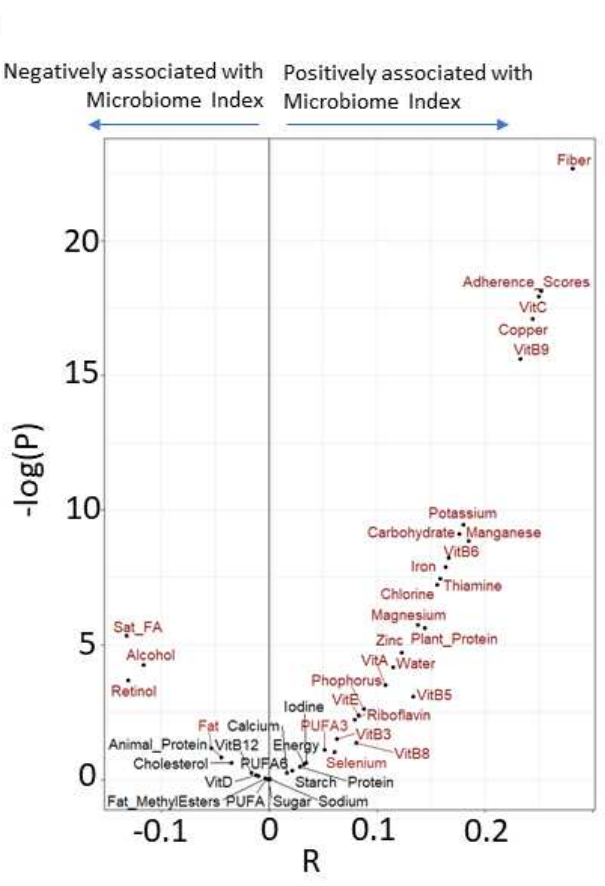
Supplementary Figure 11

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.



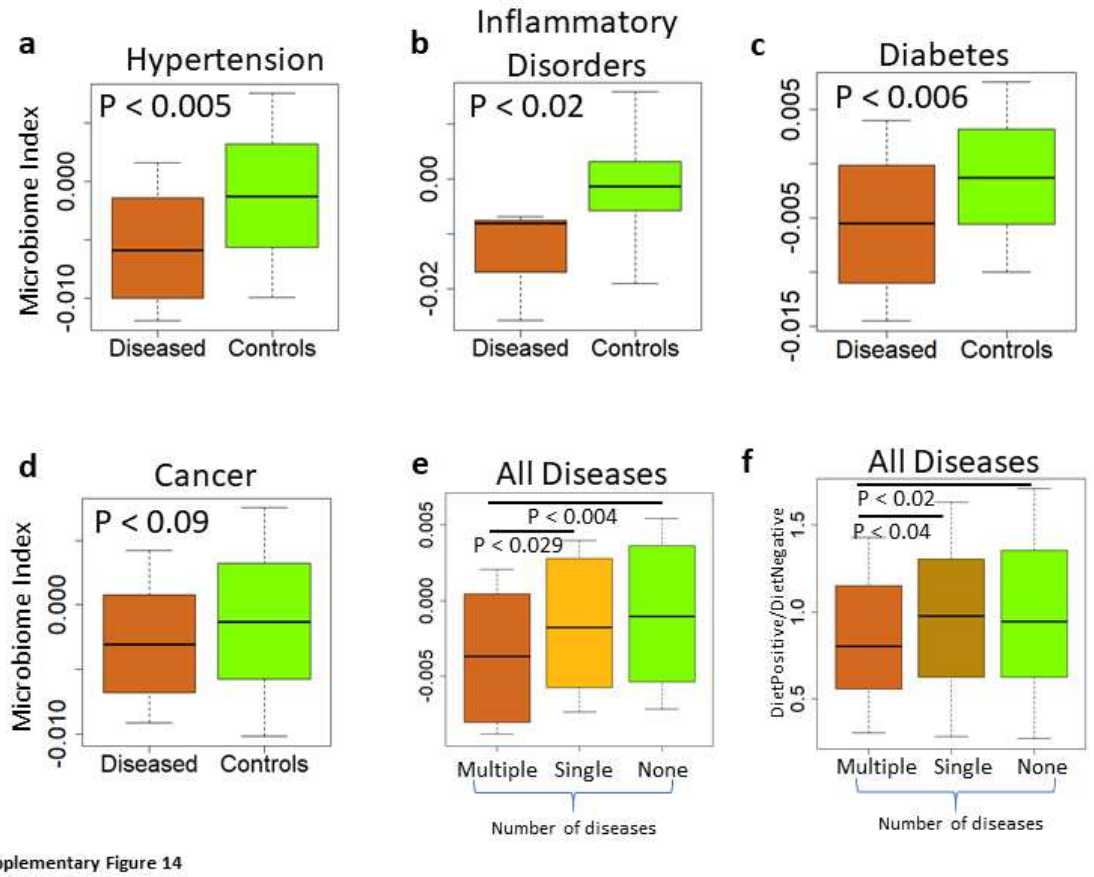
Supplementary Figure 12: Heatmap showing the partial spearman associations of the different dietary components with the marker OTUs arranged in increasing order of their association with the dietary adherence scores. For each marker OTU, partial spearman correlations were obtained after adjusting for the confounding effects of age, BMI, gender, country and poly-pharmacy.





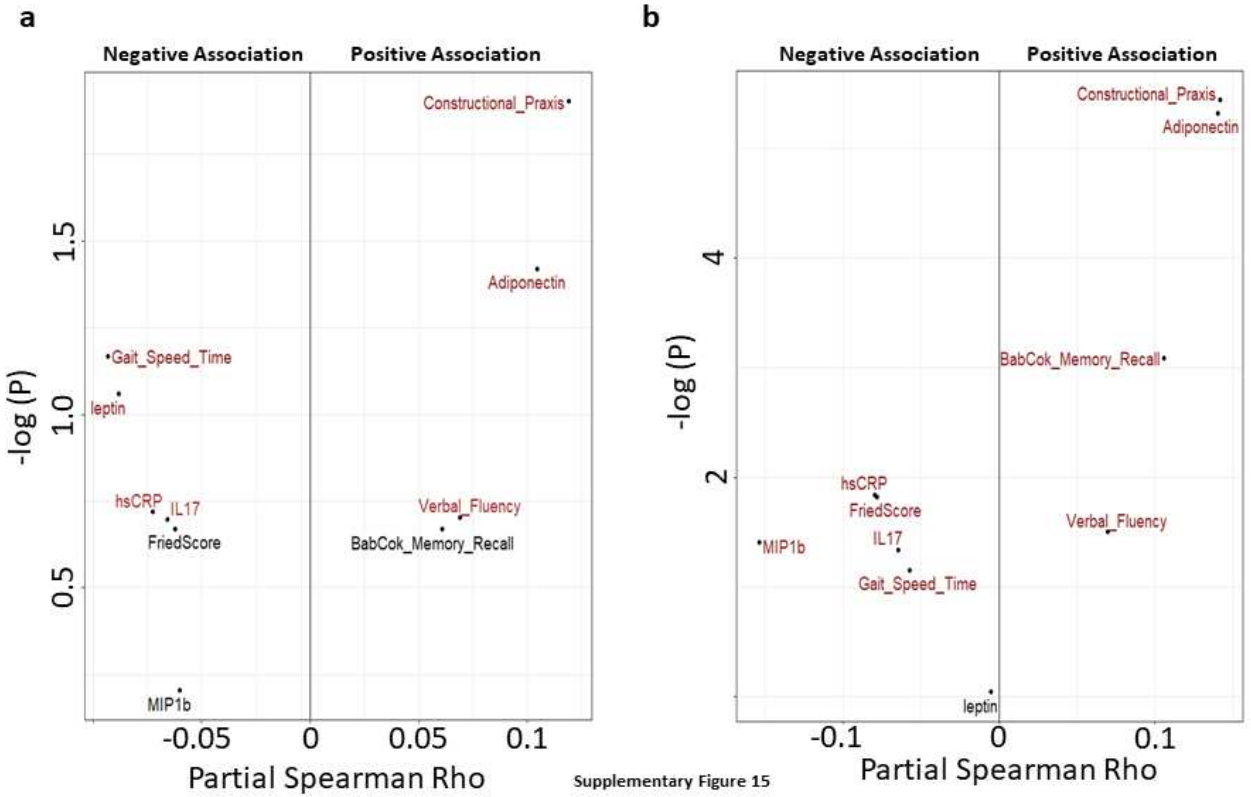
Supplementary Figure 13

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.



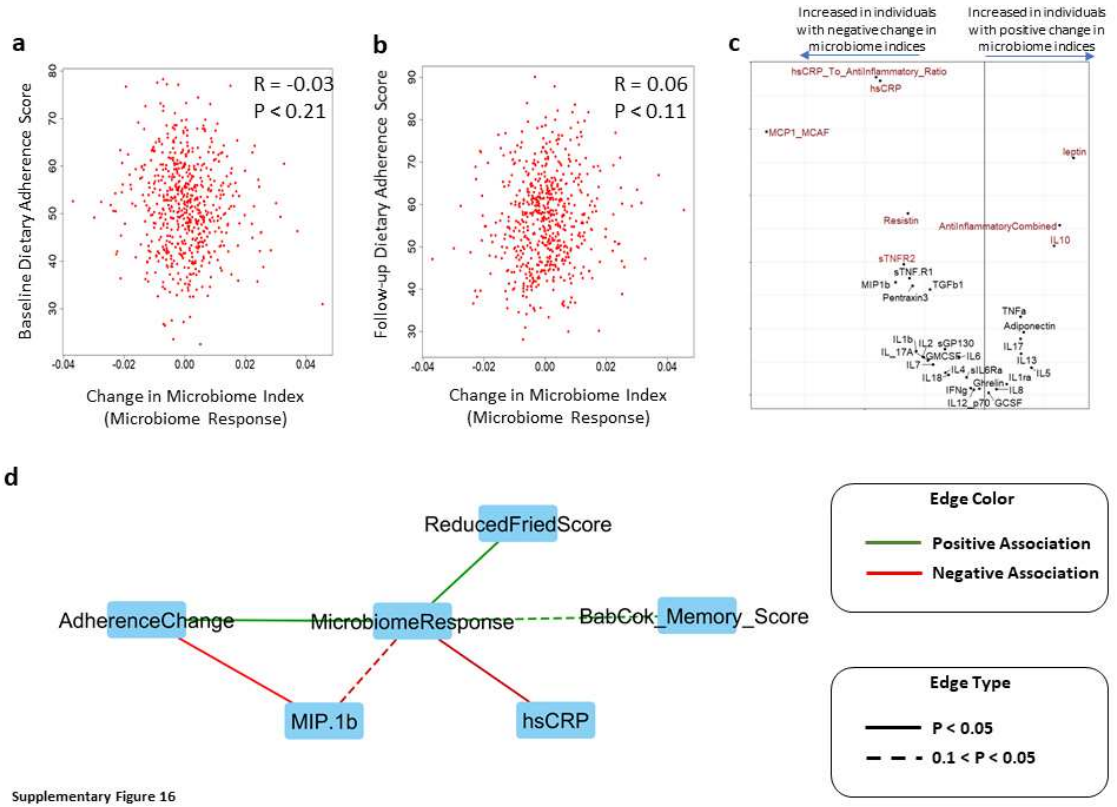
Supplementary Figure 14

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.

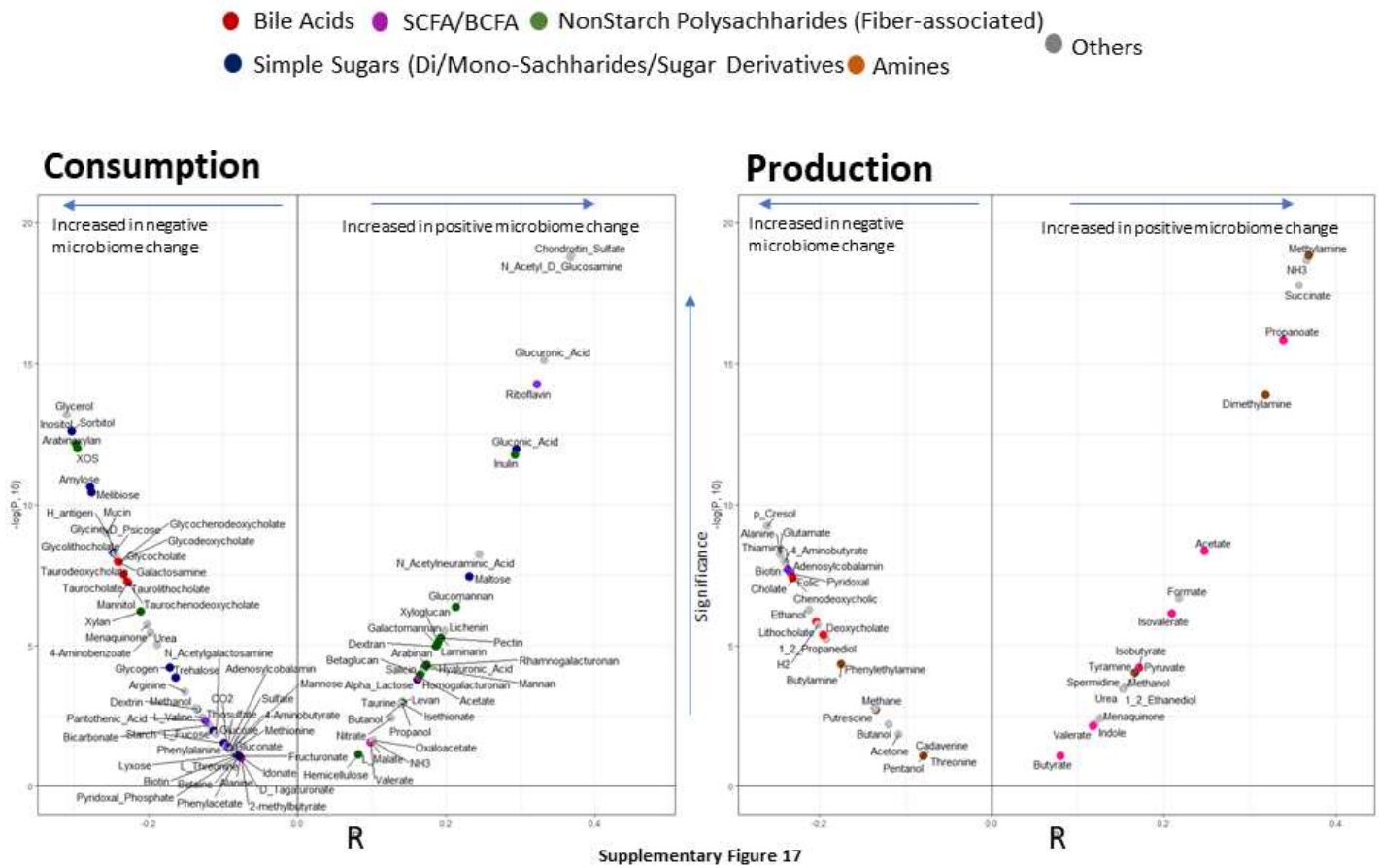


Supplementary Figure 15

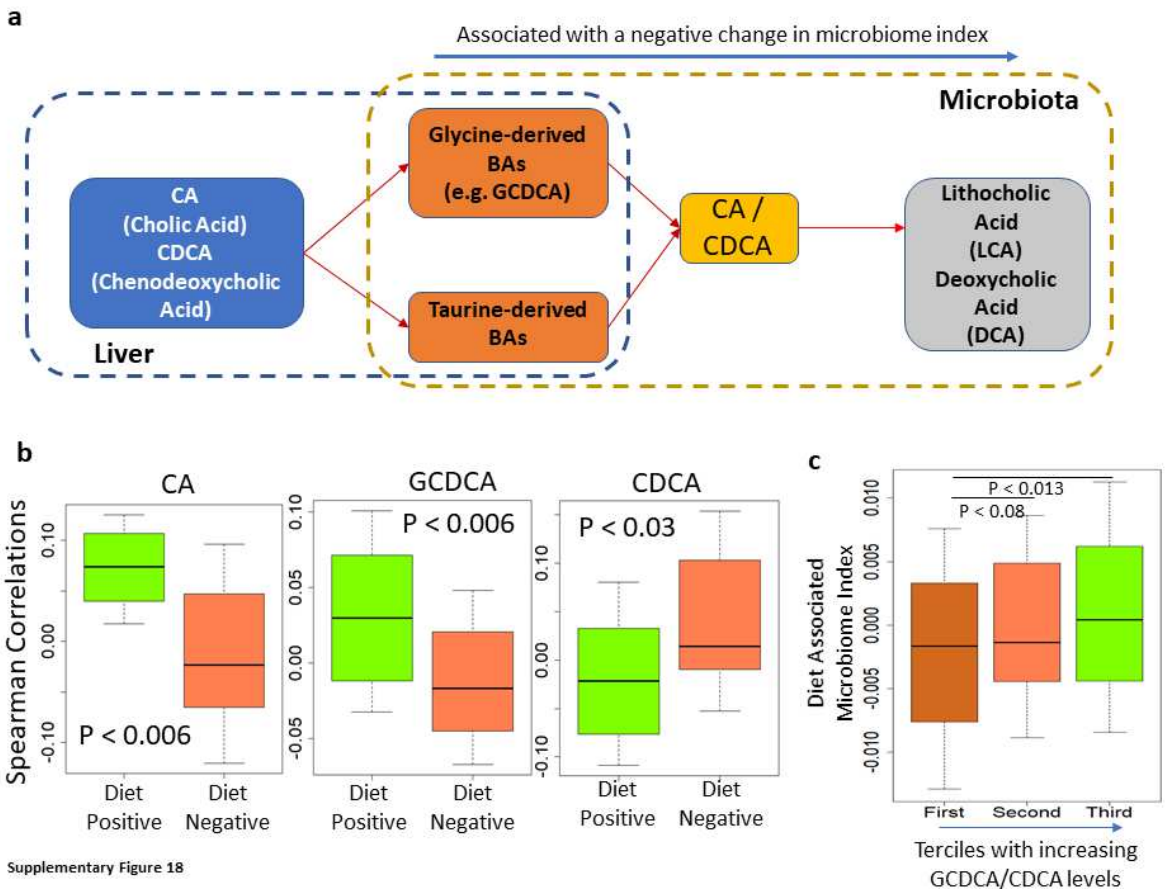
Supplementary Figure 15: a. Violin plot showing the association (partial Spearman correlations) of the different measures of frailty, cognitive function and inflammatory marker levels (identified in figure 4) with the MedDiet modulated Microbiome index at the baseline after taking into account the age, 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).



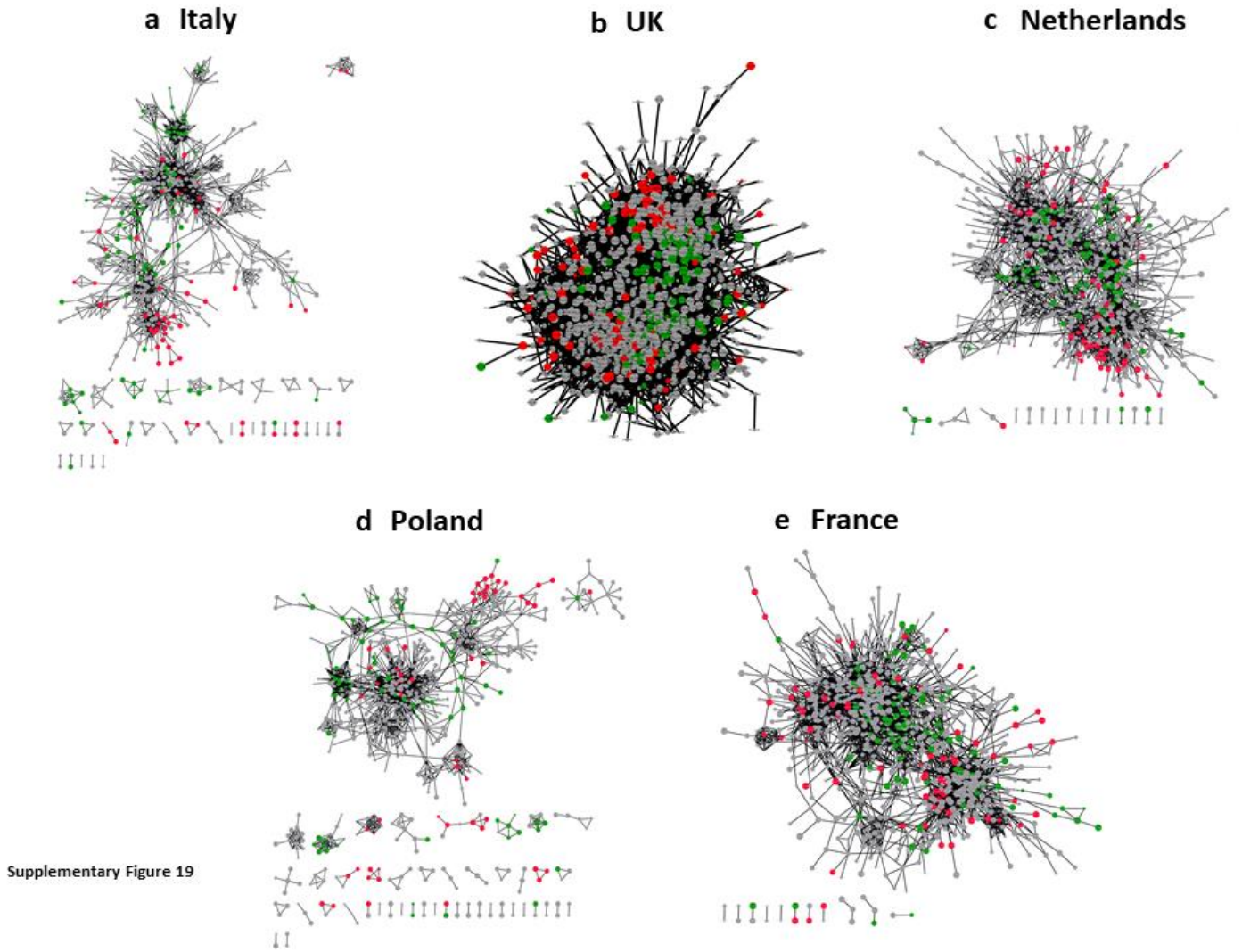
Supplementary Figure 16: Scatterplots showing the correlation between the microbiome response (that is the across time-point change in microbiome indices) and a. Baseline dietary adherence scores and b. final dietary adherence scores. c. Violin plots showing the spearman correlations of the changes in 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.



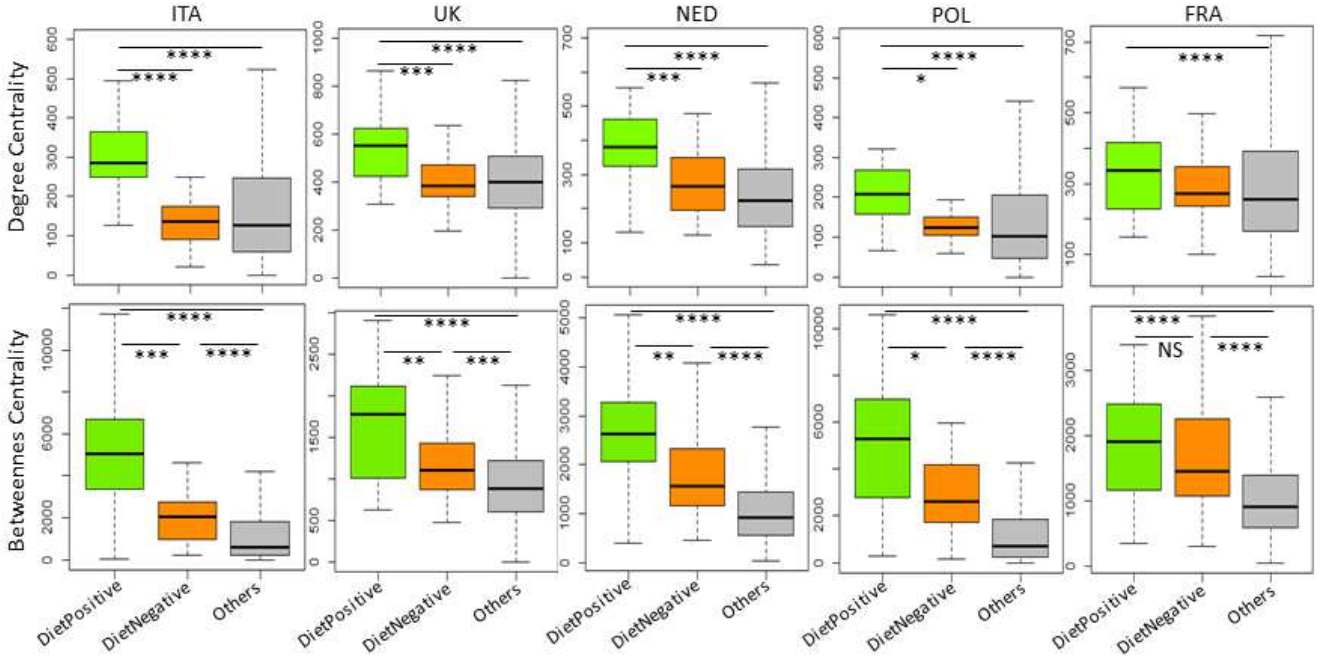
Supplementary Figure 17: Violin plots showing the inferred metabolite **a.** consumption and **b.** production profiles showing significant (positive or negative) associations (Spearman correlation; FDR corrected P-value < 0.15) with microbiome responses. While most negatively associated measures are 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.

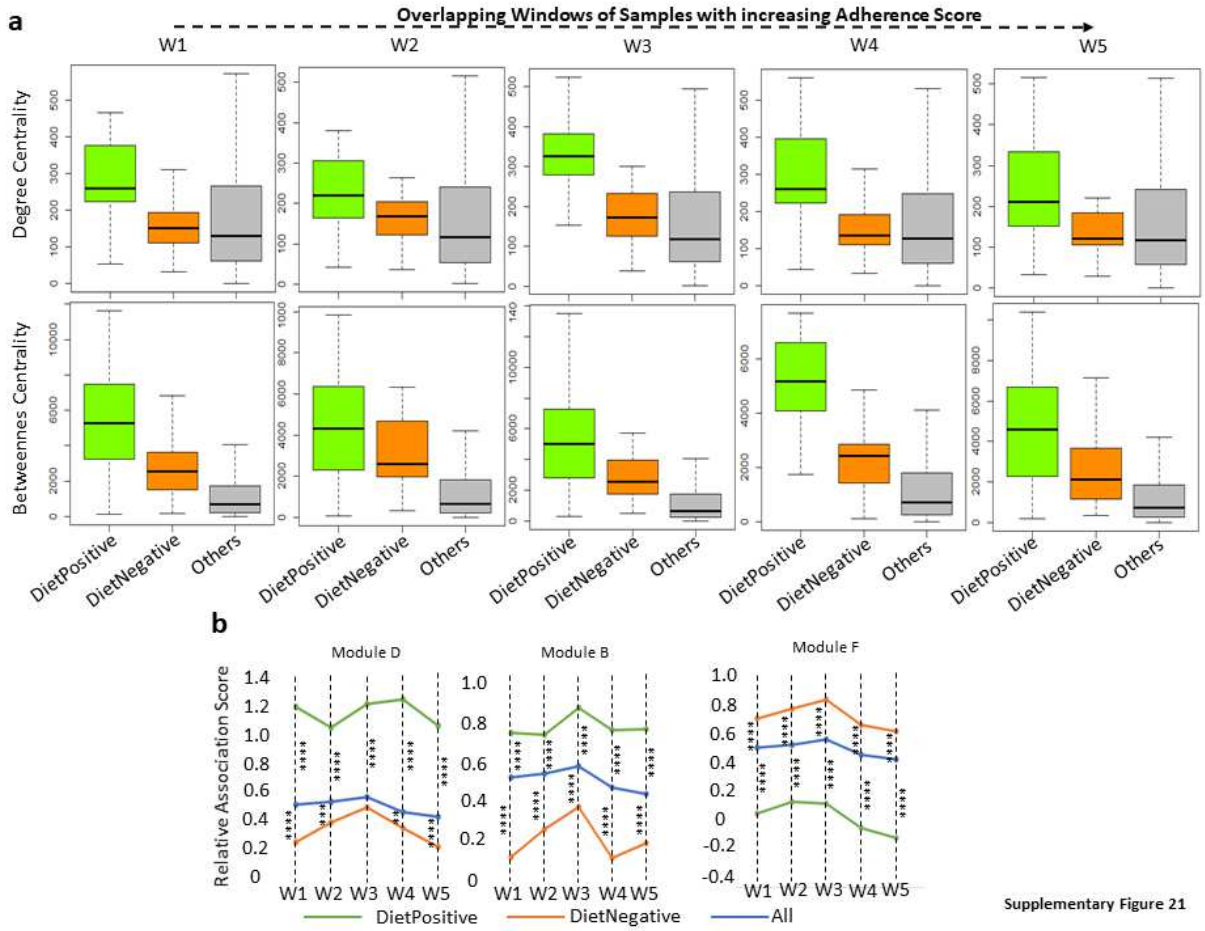


Supplementary Figure 19: Inter-OTU co-occurrence networks obtained for the different nationalities. DietPositive, DietNegative and the non-associated OTUs are shown in green, red and gray color, respectively. Despite variations in their overall structures, all networks gave a clear picture whereby in 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.