

A legume-enriched diet improves metabolic health in prediabetes mediated through gut microbiome: a randomized controlled trial

Received: 12 June 2024

Accepted: 7 January 2025

Published online: 22 January 2025

 Check for updates

Xiaorong Wu^{1,9}, Alvin Surya Tjahyo^{2,9}, Vera Sergeyevna Brok Volchanskaya¹, Long Hui Wong¹, Xianning Lai¹, Yi Ning Yong², Farhana Osman², Shia Lyn Tay², Priya Govindharajulu², Shalini Ponnalagu², Rachel Tso², Hwee Sze Teo², Kaijie Khoo¹, Huan Fan¹, Chew Chan Goh¹, Clara Poh Lian Yap¹, Melvin Khee-Shing Leow^{2,3,4,5,6,7}, Christiani Jeyakumar Henry², Sumanto Halder^{2,8,9} ✉ & Kevin Junliang Lim^{1,9} ✉

Healthy dietary patterns rich in legumes can improve metabolic health, although their additional benefits in conjunction with calorie restriction have not been well-established. We investigated effects of a calorie-restricted, legume-enriched, multicomponent intervention diet compared with a calorie-restricted control diet in 127 Chinese prediabetes participants, living in Singapore. The study was a 16-week, single-blind, parallel-design, randomized controlled trial (n = 63 intervention group (IG), n = 64 control group (CG); mean ± SD age 62.2 ± 6.3 years, BMI 23.8 ± 2.6 kg/m²). Primary outcomes were markers of glycemia and all measurements were taken at 2 or 4-weekly intervals. At the end of 16 weeks, both groups had significantly lower BMI ($q_{(Time)} = 1.92 \times 10^{-42}$, $\beta = -0.02$) compared with baseline, with minimal difference between groups. The IG had significantly greater reductions in LDL cholesterol ($q_{(Treatment \times Time)} = 0.01$, $\beta = -0.16$), total cholesterol ($q_{(Treatment \times Time)} = 0.02$, $\beta = -0.3$) and HbA1c ($q_{(Treatment \times Time)} = 0.04$, $\beta = -0.004$) compared with CG, alongside increases in fiber degrading species in IG, mediated through metabolites such as bile acids and amino acids. A legume-enriched, multicomponent intervention diet can improve metabolic health in a prediabetes population, in addition to benefits obtained from calorie restriction alone, partially mediated through changes in gut microbial composition and function. Trial registration: Clinical Trials NCT04745702.

The prevalence of type-2 diabetes (T2D) worldwide has increased by four-fold in the last four decades^{1,2} and it is estimated that more than half of the population with T2D live in Asia². Increasing evidence supports the utility of food-based interventions as promising approaches to improve metabolic health and reduce the risk of T2D and cardiovascular disease (CVD)³⁻⁵. Whole diet interventions are favored over single food or nutrient interventions due to their practical approach and potential

synergistic attributes⁶. For example, diet patterns such as the Mediterranean Diet and DASH (Dietary Approaches to Stop Hypertension) Diet have been frequently cited to improve glycemic control and reduce the risk of T2D⁶⁻¹¹. These diets contain a plethora of dietary components including fruits, vegetables, nuts, legumes and, olive oil among others that can favorably affect glucose homeostasis, to reduce cardiometabolic disease risk^{7,12,13}. As people with prediabetes are at much greater

A full list of affiliations appears at the end of the paper. ✉ e-mail: shaldar@bournemouth.ac.uk; kevin.limjunliang@sg.wilmar-intl.com

risk of developing T2D, particularly across Asia due to their greater predisposition to insulin resistance and β -cell dysfunction¹⁴, more evidence is required on the effectiveness of food-based dietary interventions in improving glycemic health in this population.

Recently, the transition from animal- to plant-based proteins and adoption of plant-based diets have taken prominence due to issues relating to environmental sustainability and health reasons. In plant-based diets, legumes are the main source of protein and are also high in dietary fiber. Consumption of legumes has been inversely associated with incidence of T2D^{15,16}. Beyond T2D prevention, plant protein intake has been associated with reduced CVD and all-cause mortality in epidemiological studies^{17–20}. In a randomized controlled trial (RCT), improvements in glycemic control in people with T2D through the incorporation of legumes as part of a low glycemic index (GI) diet was observed²¹. A more recent meta-analysis, however, did not find any significant association of legume consumption with T2D in several world regions and hence warrants further investigation²². Aside from legumes, other benefits associated with glycemic control have been attributed to consumption of low GI carbohydrates^{23,24}, antioxidant-rich polyphenol sources such as spices^{25,26}, and an overall high fiber diet²⁷. A previous meta-analysis has suggested that fiber consumption decreases the risk of prediabetes by lowering glycated hemoglobin A1c (HbA1c) with a mean difference of -2.0 mmol/mol²⁷. High-fiber whole grain diets have also been associated with increased glucagon-like peptide-1 (GLP-1) and reduced HbA1c^{28–30}.

Recent studies have increasingly highlighted the importance of gut microbiome metabolites in regulating host metabolic health. These metabolites can influence various metabolic processes in the host, including energy balance, lipid metabolism, and glucose homeostasis³¹. A large body of evidence has shown that fiber intake alters the structure of the gut microbiome and produces short chain fatty acids (SCFA)^{32–35}. SCFAs, such as acetate, propionate, and butyrate, are among the main products of microbial fermentation of dietary fibers in the gut. They have been shown to play key roles in regulating host energy metabolism, anti-inflammatory responses, and maintaining gut barrier integrity³⁶. Additionally, secondary bile acids, which are metabolic products of bile acids transformed by gut microbes, have been linked to the development of metabolic diseases such as obesity and diabetes³⁷. A study that included 1107 subjects found that obesity and diabetes were significantly associated with reduced serum hyocholic acid (HCA) species levels. Subsequent *in vitro* and mouse model experiments demonstrated that HCA acts on enteroendocrine L cells to improve glucose homeostasis by activating membrane receptor TGR5 signaling and inhibiting nuclear receptor FXR signaling, up-regulating glucagonogen gene expression, and promoting GLP-1 production and secretion^{38,39}. Amino acid metabolism of gut microbiome has also been shown to be closely related to host metabolism. One study showed that indole-3-propionic acid (IPA), a flora metabolite of tryptophan was significantly lower in patients with coronary artery disease (CAD); a combination of cellular experiments and experiments in IPA-supplemented/depleted mice demonstrated that IPA acted on macrophages to promote cholesterol efflux from macrophages by inhibiting miR-142-5p to induce ABCA1 expression, and therefore is protective against atherosclerosis⁴⁰. The gut microbiota can also metabolize choline and L-carnitine from dietary sources (eg, red meat, eggs and fish) to produce trimethylamine (TMA). This gut microbiota-derived TMA is then absorbed and reaches the liver where it is converted into TMAO, a compound associated with an increased risk of cardiovascular diseases⁴¹. This evidence prompts further investigations on how fecal or blood metabolites associated with gut microbiome may reduce prediabetes risk.

To the best of our knowledge, no study has explored whether on a background of a portion-controlled, low-calorie diet, differences in dietary quality can additionally modulate metabolic health in a prediabetic Asian population. Here, we present a randomized controlled

study to investigate the metabolic health effects of a food-based, portion-controlled intervention diet, consisting of legumes as the main protein and fiber source with low GI starches (i.e., rice, noodles etc.), prepared with a healthier vegetable oil blend as well as herbs and spices in comparison with an isocaloric, portion-controlled, control diet consisting of chicken as the main source of protein, with medium to high GI starches, prepared with corn oil and minimal herbs and spices. In addition to glycemic and cardiometabolic markers, this study also investigated changes in gut microbiome associated with adoption of either diet in a subset of the study population for a period of 16 weeks, with the hypothesis that our intervention can reduce the risk of T2D progression.

Results

Recruitment and study compliance

In this single-blinded, parallel-design study, participants with prediabetes were randomized to receive either a control diet or an intervention diet over 16 weeks. This was a partial feeding study where both the intervention and control groups had 2 of their 3 main meals replaced by the study intervention meals, which were provided to them. A CONSORT diagram detailing the flow of participants at different stages of the study is shown, resulting in 63 volunteers in the intervention group (IG) and 64 volunteers in the control group (CG) for intention-to-treat analysis; a subset of 54 volunteers in IG and 57 volunteers in CG for gut microbiome analysis (Fig. 1A). Throughout the study, plasma and stool samples were collected for various measurements at the indicated timepoints (Fig. 1B). At baseline, there were no significant differences in sex distribution, blood biomarkers, and anthropometric measurements between the two groups (Table S1). Both groups were on calorie-restricted diets and the nutritional composition of the diets were comparable except for a significantly higher amount of dietary fiber due to the inclusion of 100 g of cooked legumes per meal in the IG (Table S2). The reported mean compliance to the provided meals in completers based on the Intervention Food Records (IFR) were 98.7% and 99.3% for CG and IG, respectively (Table S3). No significant changes in physical activity and no adverse reactions in gastrointestinal symptoms reported, other than the expected increase in bloating from legumes consumption within the intervention group, was observed during the intervention period compared to the baseline (Table S4).

Dietary change during intervention

The dietary assessment data demonstrated a significant reduction in energy intake over the course of the intervention and improvements in the intakes of several other nutrients, while maintaining protein adequacy in both groups (Table 1, Figure S1). These nutritional intake improvements included approximately 40% reductions in total and saturated fat intakes, a greater than 33% reduction in sodium intake in both groups as well as an increase in dietary fiber intake in the IG by about 17 g/day. The increase in dietary fiber intake in the IG was likely due to the greater consumption of legumes and low GI starchy foods, which may also explain the relatively lower intake of carbohydrate in the total diet in the IG (by approximately 10%) compared with the CG throughout the course of the intervention.

Changes in anthropometric measures during intervention

A significant effect of time was observed in both intervention and control groups for BMI, weight, hip, waist and neck circumference, total, android, and gynoid fat mass, android to gynoid fat ratio and total lean mass with all values decreasing significantly by the end of the intervention compared to baseline levels (Tables 2 and 3). Both intervention and control group lost on average -3.23 ± 0.18 kg (-5.1%) and -2.7 ± 0.22 kg (-4.3%) weight from their baseline weight to week 16, respectively. While there were reductions in both fat mass and lean mass, the total lean mass to total fat mass ratio significantly increased

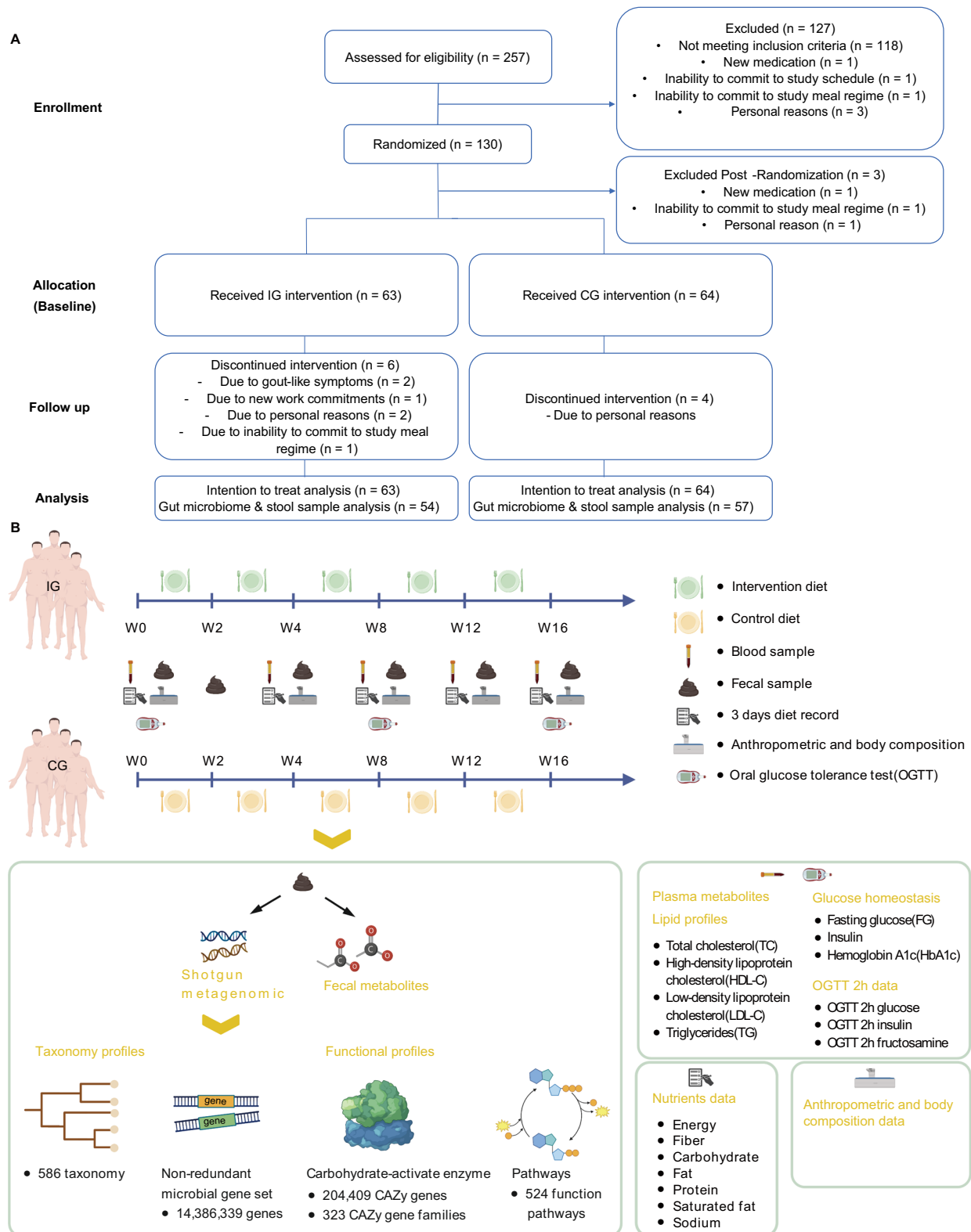


Fig. 1 | Research framework diagram. A Consort Flow Diagram of the study outlining the participant recruitment, assessment for eligibility, randomization, intervention allocation, and follow-up phases, concluding with the analysis stage. **B** Study design: The dietary intervention was administered to both groups over a period of up to 16 weeks. Blood samples were collected at baseline, week 4, week 8, week 12, and week 16. Additionally, dietary records were maintained, and both anthropometric measurements and body composition tests were conducted. We also conducted OGTT at baseline, week 8 and week 16. Fecal samples were collected at baseline, week 2, week 4, week 8, week 12, and week 16. We profiled the fecal

metagenome on all fecal samples and conducted targeted metabolomics of fecal and serum samples collected at baseline, week 4 and week 16. Lipid profile and glucose homeostasis biomarkers were assessed in all blood samples collected across five time points. For the metagenomic data, we undertook taxonomy annotation and functional characterization. The functional annotation encompassed the construction of non-redundant gene sets, annotation of carbohydrate-activate enzymes, and annotation of functional pathways. Created in BioRender. Latypov, O. (2024) <https://BioRender.com/w88p289>.

Table 1 | Daily intake of energy and nutrients during dietary intervention. Data is presented as mean ± SEM

	Week 0	Week 4	Week 8	Week 12	Week 16	q-value (β coefficients)		
						Time	Treatment	Interaction
Energy Intake (kcal/d)								
CG	1777.36 ± 64.34	1624.82 ± 47.83	1579.03 ± 38.29	1489.17 ± 37.44	1507.84 ± 35.60	2.49×10⁻¹³ (-0.05)	0.48 (-0.1)	0.38 (-0.01)
IG	1756.11 ± 39.83	1548.19 ± 41.54	1468.77 ± 38.80	1443.17 ± 34.72	1400.60 ± 31.62			
Carbohydrate (g/d)								
CG	222.40 ± 8.11	234.25 ± 6.45	230.32 ± 5.70	222.03 ± 5.73	224.82 ± 5.29	0.46 (-0.005)	0.15(-0.19)	0.03 (-0.03)
IG	223.15 ± 5.38	208.68 ± 5.24	201.62 ± 5.22	197.77 ± 4.60	196.35 ± 4.47			
Protein (g/d)								
CG	70.72 ± 2.74	72.45 ± 1.89	70.10 ± 1.40	66.23 ± 1.22	68.22 ± 1.21	8.03×10 ⁻³ (-0.02)	0.35 (-0.09)	0.21 (-0.01)
IG	68.76 ± 2.01	71.71 ± 1.72	68.35 ± 1.47	68.32 ± 1.56	67.91 ± 1.38			
Fat (g/d)								
CG	64.83 ± 3.42	43.10 ± 2.05	40.77 ± 1.65	36.08 ± 1.45	35.71 ± 1.35	2.52×10⁻³⁰ (-0.08)	0.67 (0.05)	0.19 (0.01)
IG	63.60 ± 2.63	46.31 ± 1.92	42.29 ± 1.79	41.41 ± 1.62	37.65 ± 1.32			
Saturated Fat (g/d)								
CG	22.57 ± 1.17	15.17 ± 0.82	14.65 ± 0.70	12.64 ± 0.62	13.63 ± 0.64	3.99×10⁻²¹ (-0.07)	0.85 (0.03)	0.42 (0.008)
IG	21.98 ± 0.94	15.97 ± 0.84	14.97 ± 0.89	14.17 ± 0.73	13.70 ± 0.61			
Dietary fiber (g/d)								
CG	17.39 ± 0.85	16.83 ± 0.70	16.25 ± 0.53	14.89 ± 0.58	14.00 ± 0.58	1.07×10⁻⁴ (-0.02)	3.44×10⁻¹³ (0.70)	9.49×10⁻³⁶ (0.10)
IG	18.62 ± 0.81	35.89 ± 0.73	36.26 ± 0.60	35.84 ± 0.70	35.44 ± 0.63			
Sodium (mg/d)								
CG	2478.50 ± 116.52	1807 ± 67.9	1909 ± 67.8	1604 ± 53.0	1632 ± 47.8	2.9×10⁻¹⁸ (-0.07)	0.18 (0.18)	0.84 (-0.002)
IG	2591 ± 96.5	2020 ± 89.1	1818 ± 71.6	1810 ± 68.1	1730 ± 56.5			

Analysis done using linear mixed effects model (two-sided) with week and treatment included as fixed factors, age and sex as covariates. q-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. Number of participants in IG over weeks: week 0, n = 63; week 4, n = 58; week 8, n = 57; week 12, n = 57; week 16, n = 57; number of participants in CG over weeks: week 0, n = 64; week 4, n = 60; week 8, n = 59; week 12, n = 59; week 16, n = 60. The numbers in brackets below the q-value are the corresponding β coefficients calculated from the scaled data. Bold values represent statistical significance with q < 0.05.

Table 2 | Anthropometric measurements during the dietary intervention. Data is presented as mean ± SEM

	Group	Week 0	Week 4	Week 8	Week 12	Week 16	q-value (β coefficients)		
							Time	Treatment	Interaction
BMI (kg/m²)	CG	24.17 ± 0.31	23.94 ± 0.32	23.81 ± 0.32	23.47 ± 0.32	23.33 ± 0.32	1.92×10⁻⁴² (-0.02)	0.94 (0.01)	0.06 (-0.004)
	IG	24.27 ± 0.33	23.97 ± 0.34	23.74 ± 0.33	23.4 ± 0.33	23.3 ± 0.32			
Weight (kg)	CG	63.31 ± 1.18	62.52 ± 1.19	62.20 ± 1.18	61.22 ± 1.16	60.95 ± 1.19	2.51×10⁻⁶³ (-0.02)	0.87 (0.15)	0.03 (-0.004)
	IG	64.53 ± 1.17	63.74 ± 1.16	63.06 ± 1.16	62.25 ± 1.14	61.76 ± 1.12			
Hip (cm)	CG	99.90 ± 0.77	98.84 ± 0.74	98.26 ± 0.71	97.78 ± 0.69	96.99 ± 0.79	3.67×10⁻¹³ (-0.03)	0.95 (0.07)	0.15 (-0.01)
	IG	100.32 ± 0.75	99.16 ± 0.74	98.18 ± 0.75	97.33 ± 0.73	96.4 ± 0.72			
Waist (cm)	CG	89.01 ± 1.12	87.80 ± 0.94	86.91 ± 1.00	88.42 ± 0.98	86.60 ± 1.02	3.25×10⁻⁴ (-0.02)	0.98 (0.03)	0.82 (-0.003)
	IG	89.61 ± 1.00	88.13 ± 1.06	86.56 ± 1.05	88.01 ± 1.01	87.04 ± 1.05			
Arm (cm)	CG	27.66 ± 0.33	28.38 ± 0.28	27.63 ± 0.31	28.37 ± 0.36	27.48 ± 0.24	0.57 (-0.003)	0.88 (0.11)	0.87 (0.002)
	IG	27.87 ± 0.38	28.69 ± 0.35	27.90 ± 0.33	28.68 ± 0.36	27.94 ± 0.34			
Neck (cm)	CG	36.43 ± 0.48	36.50 ± 0.39	35.96 ± 0.41	35.93 ± 0.44	35.86 ± 0.42	6.42×10⁻³ (-0.01)	0.97 (-0.04)	0.88 (0.002)
	IG	36.31 ± 0.48	36.16 ± 0.43	35.59 ± 0.44	35.72 ± 0.44	35.72 ± 0.40			
Systolic blood pressure (mmHg)	CG	125.13 ± 1.97	120.34 ± 1.98	123.85 ± 2.08	121.87 ± 1.75	126.50 ± 1.79	0.30 (0.007)	0.96 (-0.06)	0.65 (0.006)
	IG	127.32 ± 2.04	117.71 ± 1.74	120.79 ± 1.69	122.02 ± 1.99	128.76 ± 1.76			
Diastolic blood pressure (mmHg)	CG	76.59 ± 1.13	72.14 ± 1.14	73.48 ± 1.17	73.20 ± 1.02	74.06 ± 1.19	0.11 (-0.01)	0.91 (-0.05)	0.20 (0.01)
	IG	76.38 ± 1.17	71.81 ± 1.20	72.19 ± 1.17	74.48 ± 1.26	75.77 ± 1.02			
Pulse (bpm)	CG	71 ± 1.20	68 ± 1.10	68 ± 1.08	65 ± 1.14	67 ± 1.21	3.70×10⁻⁸ (-0.03)	0.68 (-0.24)	0.39 (0.01)
	IG	69 ± 1.16	65 ± 0.96	66 ± 1.21	64 ± 1.05	66 ± 1.10			

Analysis done using linear mixed effects model (two-sided) with week and treatment included as fixed factors, age and sex as covariates. q-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. Number of participants in IG over weeks: week 0, n = 63; week 4, n = 58; week 8, n = 57; week 12, n = 57; week 16, n = 57; number of participants in CG over weeks: week 0, n = 64; week 4, n = 61; week 8, n = 60; week 12, n = 59; week 16, n = 60. The numbers in brackets below the q-value are the corresponding β coefficients calculated from the scaled data. Bold values represent statistical significance with q < 0.05.

Table 3 | Body composition measurements during the dietary intervention measured using DEXA. Data is presented as mean \pm SEM

	Week 0	Week 16	q-value (β coefficients)		
			Time	Treatment	Interaction
Total fat mass (kg)					
CG	22.03 \pm 0.62	20.64 \pm 0.64	2.48$\times 10^{-5}$ (-0.008)	0.40 (0.19)	0.45 (-0.004)
IG	22.06 \pm 0.68	20.61 \pm 0.68			
Visceral fat mass (g)					
CG	645.92 \pm 31.38	582.61 \pm 29.52	0.1 (0.001)	0.99 (0.001)	0.86 (-0.0006)
IG	624.41 \pm 27.55	557.27 \pm 26.68			
Android fat mass (kg)					
CG	2.12 \pm 0.07	1.91 \pm 0.07	3.10$\times 10^{-6}$ (-0.007)	0.72 (-0.1)	0.87 (-0.002)
IG	2.05 \pm 0.08	1.84 \pm 0.08			
Gynoid fat mass (kg)					
CG	3.61 \pm 0.12	3.38 \pm 0.12	0.02 (-0.004)	0.72 (0.08)	0.86 (-0.002)
IG	3.59 \pm 0.12	3.31 \pm 0.12			
Android to gynoid fat ratio					
CG	1.13 \pm 0.02	1.10 \pm 0.02	1.59$\times 10^{-10}$ (-0.02)	0.72 (-0.21)	0.87 (-0.001)
IG	1.11 \pm 0.02	1.10 \pm 0.03			
Total lean mass (kg)					
CG	40.28 \pm 1.07	39.45 \pm 1.09	5.37$\times 10^{-11}$ (-0.007)	0.40 (0.17)	0.31 (-0.003)
IG	41.45 \pm 1.09	40.21 \pm 1.09			
Total lean mass/total fat mass ratio					
CG	1.94 \pm 0.09	2.05 \pm 0.010	1.28$\times 10^{-16}$ (-0.02)	0.99 (0.001)	0.87 (-0.002)
IG	2.03 \pm 0.10	2.11 \pm 0.11			

Analysis done using linear mixed effects model (two-sided) with week and treatment included as fixed factors, age and sex as covariates. q-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. Number of participants in IG over weeks: week 0, $n = 63$; week 16, $n = 57$. Number of participants in CG over weeks: week 0, $n = 64$; week 16, $n = 60$. The numbers in brackets below the q-value are the corresponding β coefficients calculated from the scaled data. Bold values represent statistical significance with $q < 0.05$.

in both groups over time indicating greater relative reductions in fat mass compared to reductions in lean mass following the intervention. Body composition parameters using BIA other than body weight is reported in Table S5.

Intervention diet improves blood lipids and glycemic indices in prediabetic participants

The temporal effect of glycated hemoglobin (HbA1c) differed across groups, with a statistically greater decrease in the IG at week 16 ($q_{(Treatment \times Time)} = 0.04$, $\beta = -0.004$, Fig. 2A). The change over time of fasting glucose (FG), insulin and HOMA-IR was comparable in both groups: insulin and HOMA-IR levels showed a decline over the study period while FG levels remained stable (Fig. 2B-D). The most significant changes in lipid profile were levels of blood cholesterol, including TC at week 4 ($q_{(Treatment \times Time)} = 0.02$, $\beta = -0.3$), high-density lipoprotein cholesterol (HDL-C) at week 4 ($q_{(Treatment \times Time)} = 0.02$, $\beta = -0.06$) and low-density lipoprotein cholesterol (LDL-C) at week 12 ($q_{(Treatment \times Time)} = 0.01$, $\beta = -0.16$), decreased significantly from baseline in both groups but there were greater reductions in the IG (Fig. 2E-G). LDL-C and TC plateaued following a sharp decrease after week 4 (Fig. 2E, G). In contrast, HDL-C increased in a time-dependent manner after an initial decrease after week 4 (Fig. 2F). TC to HDL-C ratio and TG were not significantly different between the two groups (Fig. 2H-I). The 2-hour oral glucose tolerance test (OGTT) revealed no notable alterations in blood glucose, insulin, or fructosamine levels over time for both groups (Fig. S2A). Leptin levels significantly declined in both groups ($q_{(Time)} = 2.12 \times 10^{-8}$, $\beta = -0.03$), whereas adiponectin remained relatively stable over time in both groups (Fig. S2B).

Intervention diet alters the species composition of gut microbiome

From shotgun metagenomic sequencing, we identified 586 distinct taxa, revealed 524 functional pathways (via MetaCyc), and established

a catalog comprising 14,386,339 unique microbial genes. Permutational multivariate analysis of variance (PERMANOVA) on gut microbiome composition revealed that the intervention diet and in particular dietary fiber intake were significant variables driving microbiome structure amongst others like LDL (Fig. 3A), these changes were visualized using a Principal Coordinate Analysis (PCoA) (Fig. 3B, C, Fig. S3A-B). A similar functional association was observed but dietary fiber was not significant (Figure S3C). Notably, both groups exhibited significant gut microbiota shifts from week 0 to week 2, with the IG showing a more pronounced change ($p_{(Treatment \times Time)} = 0.001$) and no further significant changes observed from week 2 to week 16 across the two groups ($p_{(Treatment \times Time)} = 0.365$) (Fig. 3D-E, Fig. S3D). However, we found no changes of the gut microbiota taxonomy diversity (Fig. S3F) and gene count (Fig. S3E) over time in either group.

The alterations in the structure of gut microbiota prompted us to investigate which species responded most strongly to the intervention diet. Linear Models for Differential Abundance (LinDA) analysis identified 30 species (using a cut-off $q_{(Treatment \times Time)} < 0.15$, $q_{(Time)} < 0.05$) and 15 genera (using a cut-off $q_{(Treatment \times Time)} < 0.05$, $q_{(Time)} < 0.05$) that varied between the groups over time. As expected, these species remained relatively stable over time in the CG (Fig. 3F, Figure S4A-B). Of the species that did change, 8 showed a significant increase compared to the CG, referred as positive responding species. The rest exhibited significant or marginal decreases, referred as negative responding species compared to the CG. Specifically, those with significant increases were species belonging to the class *Clostridia* (*Eubacterium rectale*, *Roseburia faecis*, *Roseburia hominis* and) as well as the *Bifidobacterium* genus under the phylum *Actinobacteria*. On the other hand, three *Ruminococcus* genera (*R. gnavus*, *R. torques* and *R. lactaris*), a number of species belonging to the genus *Bacteroides* (including *B. massiliensis*, *B. stercoris*; except *B. ovatus*) and *Bilophila* (*B. wadsworthia*) decreased in relative abundance (Fig. 3F, Figure S4B). The enrichment or depletion of these species peaked by week 2,

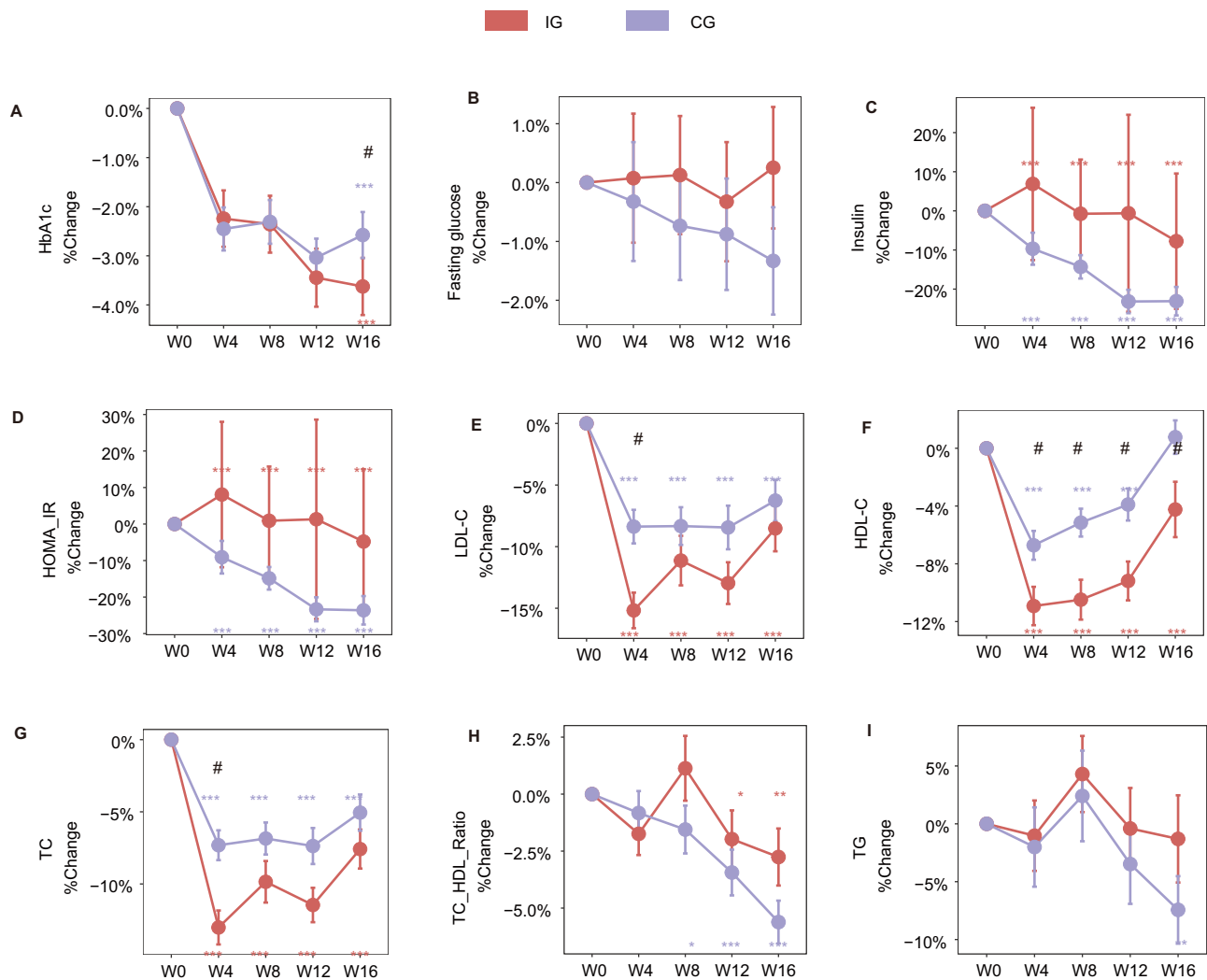


Fig. 2 | General response of the clinical outcomes during the study. Changes from baseline in (A) HbA1c, (B) fasting glucose, (C) insulin, (D) HOMA-IR, (E) LDL-C, (F) HDL-C, (G) TC, (H) TC/HDL-C ratio and (I) TG during the intervention are shown. Data presented as percent changes from week 0 (mean \pm standard error). # $q_{(Treatment \times Time)} < 0.05$ denote p-adjusted values for the Treatment \times Time interaction coefficient in the LMM model (two-sided; HbA1c: 0.04; LDL-C: W4 0.02; HDL-C: W4 0.02, W8 0.03, W12 0.01, W16 0.01; TC: W4 0.02); * $q_{(Time)} < 0.05$, ** $q_{(Time)} < 0.01$, and *** $q_{(Time)} < 0.001$ denote p-adjusted values for the time coefficient within the

same group. p -values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. Number of participants in IG over weeks: week 0, $n = 63$; week 4, $n = 58$; week 8, $n = 57$; week 12, $n = 57$; week 16, $n = 57$; number of participants in CG over weeks: week 0, $n = 64$; week 4, $n = 61$; week 8, $n = 60$; week 12, $n = 59$; week 16, $n = 60$. HbA1c: glycated hemoglobin A1c; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; LDL-C: Low-Density Lipoprotein Cholesterol; HDL-C: High-Density Lipoprotein Cholesterol; TC: Total Cholesterol; TG: Triglycerides.

maintaining stability thereafter (Fig. 3F), consistent with the pattern we observed in the overall gut microbiota structure (Fig. 3D-E).

LMM further revealed that 13 out of the 30 responding species and 7 out of the 15 responding genera were significantly correlated (using a cut-off $q < 0.05$) with dietary fiber intake. No other nutrient was significantly associated with these changes in microbiota (Fig. S4C). The direction of these nutritional associations also corresponded with the changes observed during the intervention. For example, *E. rectale* and genus *Bifidobacterium* increased over time in the IG and showed positive correlations with dietary fiber, while *R. torques*, *R. lactaris*, and *R. gnavus* decreased and demonstrated negative correlations (Fig. 3G-H, Fig. S4C-D).

Intervention diet alters gut bacterial fermentation of carbohydrates

We then conducted a gene-centric analysis to explore the functional changes in the gut microbiota that might contribute to improved host clinical outcomes. The presence of dietary fiber in the meal enhanced

the supply of fermentable carbohydrates, prompting us to concentrate on genes associated with carbohydrate utilization, specifically those encoding glycoside hydrolases (EC3.2.1). A total of 21 glycoside hydrolases were significantly different between the two groups over time; these trajectories were mainly changed in the IG (using cut-off $q_{(Treatment \times Time)} < 0.05$, $q_{(Time)} < 0.05$, Fig. 4A). Enzymes under the class of glycoside hydrolases that were elevated are mainly responsible for the degradation of plant structural carbohydrates such as pectic polysaccharides (EC3.2.1.89, EC3.2.1.99), arabinoxylans (EC3.2.1.37, EC3.2.1.55) and xyloglucan (EC3.2.1.151, EC3.2.1.156), amongst other polysaccharides. In contrast, enzymes involved in the digestion of trehalose, lactose or short fructooligosaccharides (EC3.2.1.28, EC3.2.1.122) were decreased in the IG (Fig. 4A, Figure S5A). To further explore the hypothesis on fermentable carbohydrates, we analyzed genes that code for carbohydrate-active enzymes (CAZymes). This analysis revealed a total of 204,409 CAZy genes grouped into 323 CAZy gene families, corroborating a similar pattern to glycoside hydrolases. Among the CAZy genes, we found 52 that were significantly

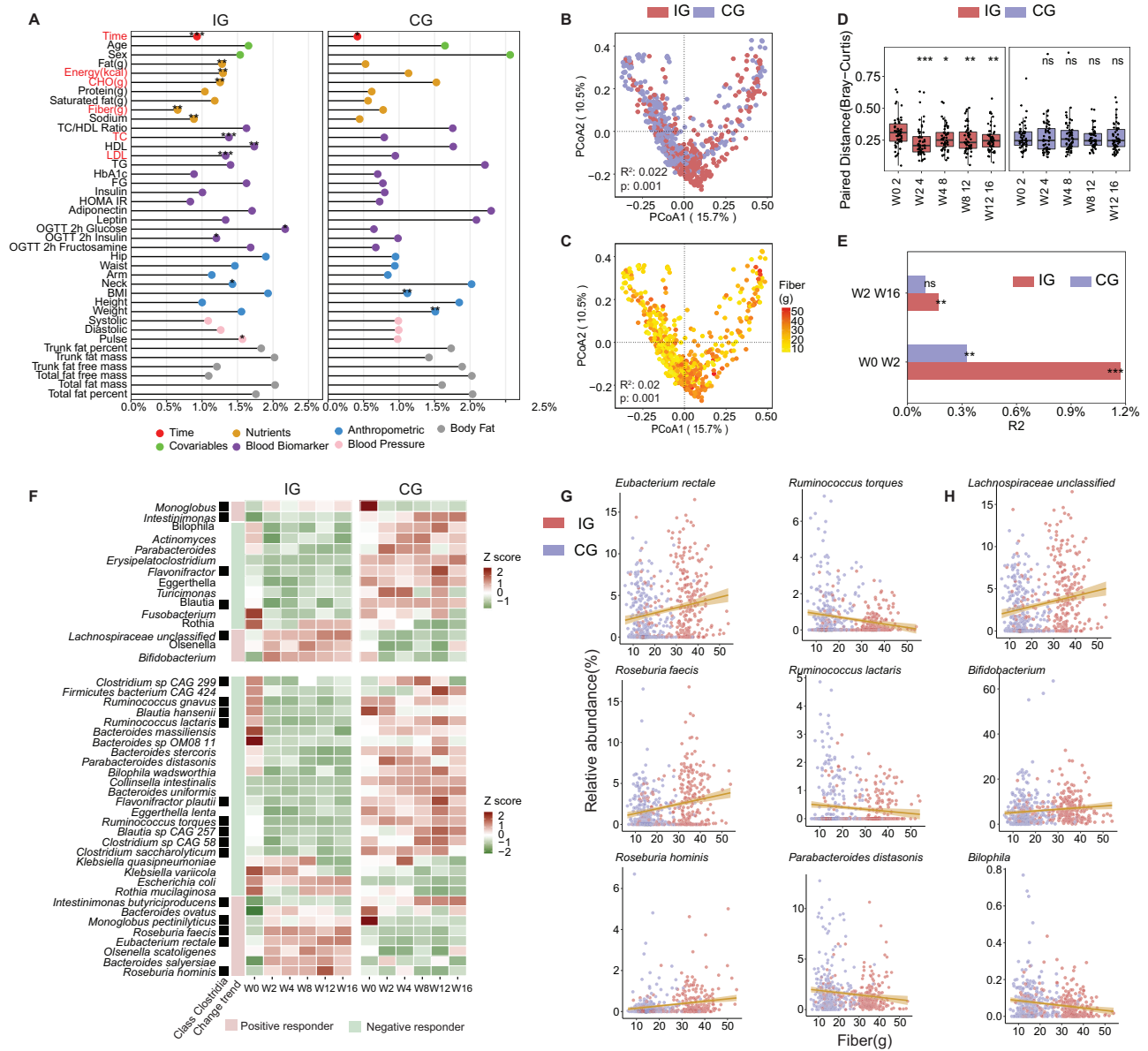


Fig. 3 | Taxonomic alterations of the human gut microbiome during dietary intervention. **A** Percentage of fecal microbiome variation explained by time, macronutrient, clinical indicators and anthropometric data by the PERMANOVA model (two-sided) based on Bray-Curtis distances in two groups separately. Red variable names represent significant Treatment × Time interaction effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (IG: Time 0.001, Fat 0.01, Energy 0.002, CHO 0.004, Fiber 0.002, Sodium 0.002, TC 0.001, HDL 0.005, LDL 0.001, OGTT 2 h glucose 0.027, OGTT 2 h glucose insulin 0.038, Neck 0.013, Pulse 0.049; CG: Time 0.025, BMI 0.005, Weight 0.007). CHO: Carbohydrate; HbA1c: glycated hemoglobin A1c; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; LDL-C: Low-Density Lipoprotein Cholesterol; HDL-C: High-Density Lipoprotein Cholesterol; TC: Total Cholesterol; TG: Triglycerides; FG: Fasting Glucose. **B–C** Principal coordinate analysis of Bray-Curtis distances. The axes are labeled with the percent variance explained. R^2 values and p -values were calculated from the two-sided PERMANOVA test. **B** Colors are filled according to different groups; **C** Colors are filled according to fiber intake. **D** Microbiome pairwise dissimilarity between weeks in each subject based on Bray-Curtis distances. The central line in each box represents the median. The bounds of the box indicate the interquartile range (IQR), with the lower bound at the 25th percentile (Q1) and the upper bound at the 75th percentile (Q3). The whiskers extend to the smallest and largest values within $1.5 \times$ IQR from the box. Data points beyond the whiskers are considered outliers. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ denote two-sided T-test p -values performed on every W_x to W_0 , where W_x represents the Bray-Curtis distance between

samples collected at W_x and W_y ($W_{2,4}$: 0.002; $W_{4,8}$: 0.021; $W_{8,12}$: 0.0097; $W_{12,16}$: 0.014). Number of participants in IG over W_0 : $n = 54$; $W_{2,4}$, $n = 54$; $W_{4,8}$, $n = 54$; $W_{8,12}$, $n = 53$; $W_{12,16}$, $n = 53$; number of participants in CG over W_0 : $n = 55$; $W_{2,4}$, $n = 55$; $W_{4,8}$, $n = 57$; $W_{8,12}$, $n = 55$; $W_{12,16}$, $n = 55$. **E** Percentage of fecal microbiome variation explained by different time periods in both groups. Two-sided PERMANOVA test, ** $p_{(Time)} < 0.01$, *** $p_{(Time)} < 0.001$. (W_0 to W_2 : IG 0.001, CG 0.009; W_2 to W_{16} : IG 0.004, CG 0.147). **F** The taxa that are significantly changed over time in intervention and control groups ($q_{(Treatment \times Time)} < 0.15$ and $q_{(Time)} < 0.05$). The values represent the Z scores of averages of study weeks and groups calculated per taxa. Change trend: positive responding species are annotated in pink and negative responding species are annotated in green. Upper panel: genus; lower panel: species. **G–H** Scatterplot of correlation between dietary fiber intake and significantly changed species (**G**) and genera (**H**) across all samples. Coefficients and adjusted p -values are derived from two-sided LMM (*Eubacterium rectale*: $q = 6.95 \times 10^{-5}$, $\beta = 0.17$; *Ruminococcus torques*: $q = 3.5 \times 10^{-4}$, $\beta = -0.15$; *Roseburia faecis*: $q = 2.7 \times 10^{-8}$, $\beta = 0.28$; *Ruminococcus lactaris*: $q = 4.7 \times 10^{-4}$, $\beta = -0.15$; *Roseburia hominis*: $q = 4.6 \times 10^{-3}$, $\beta = 0.17$; *Parabacteroides distasonis*: $q = 4.7 \times 10^{-4}$, $\beta = -0.13$; *Lachnospiraceae unclassified*: $q = 6.95 \times 10^{-5}$, $\beta = 0.17$; *Bifidobacterium*: $q = 4.0 \times 10^{-3}$, $\beta = 0.16$; *Bilophila*: $q = 4.0 \times 10^{-3}$, $\beta = -0.17$). p -values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. The smooth curve represents the trend line fitted using Linear Model, and the shaded area indicates the 95% Confidence Interval.

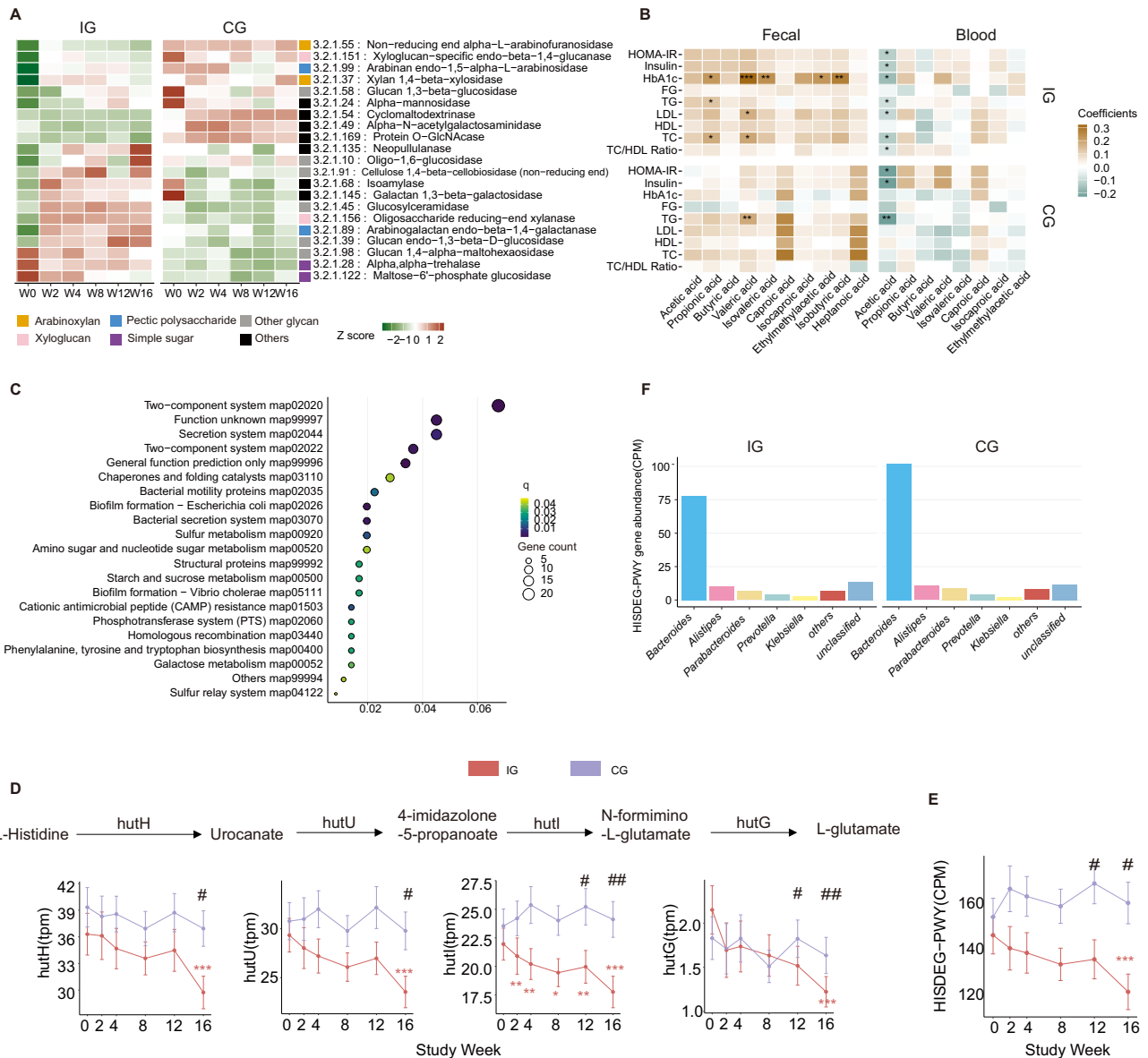


Fig. 4 | Alterations in gut bacterial fermentation of carbohydrates, SCFA profile and other functions. **A** The glycoside hydrolase genes that have significantly different progression profiles in the IG compared with the CG identified by LinDA ($q_{(Treatment \times Time)} < 0.05$ and $q_{(Time)} < 0.05$). The values represent the Z scores of averages of study weeks and groups calculated per enzyme. **B** Correlation of blood glucose and lipid levels with intestinal and plasma SCFAs in two groups respectively. * $q < 0.05$, ** $q < 0.01$, and *** $q < 0.001$. HbA1c: glycated hemoglobin A1c; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; LDL-C: Low-Density Lipoprotein Cholesterol; HDL-C: High-Density Lipoprotein Cholesterol; TC: Total Cholesterol; TG: Triglycerides; FG: Fasting Glucose. **C** Bubble chart of gene enrichment analyses for genes significantly changed in the IG. Size of circle denotes number of genes in the pathway, colored by adjusted p -values. **D** Changes in genes involved in each step of the histidine degradation pathway. Data presented as mean \pm standard error. # $q_{(Treatment \times Time)} < 0.05$, ## $q_{(Treatment \times Time)} < 0.01$ denote q -values for the Treatment \times Time interaction coefficient (hutH: W16 0.03; hutU:

W16 0.008; hutI: W12 0.01, W16 0.002; hutG: W12 0.01, W16 0.002); * $q_{(Time)} < 0.05$, ** $q_{(Time)} < 0.01$, and *** $q_{(Time)} < 0.001$ denote q -values for the time coefficient within the same group. hutH: histidine ammonia-lyase; hutU: urocanate hydratase; hutI: imidazolonepropionase; hutG: formiminoglutamase. Number of participants in IG over weeks: week 0, $n = 54$; week 2, $n = 54$; week 4, $n = 54$; week 8, $n = 54$; week 12, $n = 53$; week 16, $n = 54$; number of participants in CG over weeks: week 0, $n = 57$; week 2, $n = 55$; week 4, $n = 57$; week 8, $n = 57$; week 12, $n = 56$; week 16, $n = 57$. **E** Changes in histidine degradation pathway abundance annotated by HUMAnN3. Data presented as mean \pm standard error. # $q_{(Treatment \times Time)} < 0.05$ denote q -values for the Treatment \times Time interaction coefficient; *** $q_{(Time)} < 0.001$ denote q -values for the time coefficient within the same group. Number of participants in IG over weeks: week 0, $n = 54$; week 2, $n = 54$; week 4, $n = 54$; week 8, $n = 54$; week 12, $n = 53$; week 16, $n = 54$; number of participants in CG over weeks: week 0, $n = 57$; week 2, $n = 55$; week 4, $n = 57$; week 8, $n = 57$; week 12, $n = 56$; week 16, $n = 57$. **F** HISDEG-PWY gene abundance (CPM) for various bacterial taxa in IG and CG groups.

different between two groups (using a cut-off $q_{(Treatment \times Time)} < 0.05$). Those contributing to arabinoxylan, cellulose, pectin and starch degradation were significantly enriched, whereas those related to the use of mucin and agarose were reduced in the IG but not in the CG (Figure S5B).

In vitro validation showed larger increases in the populations of *Eubacterium rectale*, *Ruminococcus bromii*, and *Bacteroides ovatus*

(Fig. S6) in the presence of undigested red bean powder. Notably, immediate post-inoculation assessments did not indicate any growth of the responsive species (except for *Bacteroides ovatus*) on the first day, likely due to an initial adaptation period. Nonetheless, the fiber-degrading bacteria demonstrated noticeable growth between the first and second day. In contrast, we did not see these changes in the presence of in vitro-digested or undigested brown rice powder.

To explore whether enhanced carbohydrate fermentation by the gut microbiota led to elevated SCFAs levels, we analyzed the absolute quantities of 10 SCFAs in fecal samples and 8 SCFAs in plasma. There were no significant differences in the changes of absolute SCFA concentrations in feces between the intervention and control groups (Fig. S7A). Among 8 types of SCFAs detected in plasma, the absolute amount of acetic acid was significantly higher in the IG (q ($Treatment \times Time$) = 0.05, β = 20.61, Figure S7B).

Further correlation analysis reinforced the connection between these SCFA levels and biomarkers. In the fecal samples of the IG, the absolute abundance of a number of SCFAs were positively correlated to HbA1c (using a cut-off $q < 0.05$), while in the CG only valeric acid was positively correlated with TG (using a cut-off $q < 0.05$). The concentration of plasma acetic acid was mainly seen as significantly negatively correlated (using a cut-off $q < 0.05$) with HOMA-IR, insulin, and TG in both groups. However, this pattern was only significant in the IG for HbA1c, LDL-C, TC and TC/HDL-C ratio (Fig. 4E, Fig. S8B).

Gut microbiota changes may improve metabolic outcomes via amino acid metabolism

For the 14,386,339 microbial genes, which was mapped to 7568 KEGG orthology. LinDA identified 356 significant temporal changes between the two groups (using a cut-off q ($Treatment \times Time$) < 0.05) and had significant changes in the IG. Gene enrichment analysis on these 356 genes showed many pathways mainly comprising of carbohydrate and amino acid metabolism (Fig. 4C, Fig. S9B). MetaCyc aggregated pathway abundances also revealed numerous metabolic routes that had significant temporal shifts in the IG. Within these pathways, amino acid metabolic routes showed the most significant changes across time in the IG, in contrast to their stability in the CG (Fig. S9A). Intriguingly, these pathways also had notable correlations with HbA1c and certain lipids in the IG; these correlations were less evident in the CG (Fig. S10A).

Histidine metabolism is involved in producing 4-imidazolone-5-propanoate (IMP). Both the pathway and its corresponding genes were reduced in the IG (Fig. 4D-E), predominantly influenced by the genus *Bacteroides*, which were significantly reduced in the IG (Fig. 4F, Fig. 3F). Consistent to findings from the histidine pathway, both intestinal and plasma IMP concentrations showed modest declines in the IG (Wilcoxon $p = 0.03$), suggesting that diet-related improvements observed from the IG may be mediated through favourable modulation of IMP concentration (Figure S10B).

Intervention diet significantly alters fecal metabolome but has a minor effect on blood metabolome

We also explored the alterations in fecal and plasma metabolite profiles before and after dietary intervention, by performing targeted detection of a panel of 321 metabolites (Supplementary Data 1) in fecal and plasma samples. 195 and 193 metabolites were detected in the fecal and plasma samples respectively. Principal coordinate analysis revealed that the dietary intervention significantly impacted the fecal metabolite profile. In contrast, changes in plasma metabolite profiles remained comparatively subtle (Fig. 5A).

At a broad metabolite classification level, metabolites classified as bile acids, carnitines, furans and phenylpropanoic displayed consistent change patterns, decreasing over time in the fecal IG (Figure S11A). In plasma, there was a marked increase in metabolites identified as indoles and benzenoids, which are derivatives of aromatic amino acids (Figure S11B). LMM revealed significant alterations on 58 intestinal and 15 plasma metabolites respectively over time (using a cut-off q ($Treatment \times Time$) < 0.25, Fig. 5B-C). The 58 significantly altered intestinal metabolites, primarily consisting of amino acids and bile acids, aligned with the earlier observed shifts in metabolic functions predicted from gut microbiota genes. Specifically, intestinal amino acids and their

derivative metabolites decreased in the IG, while remaining relatively stable in the CG.

Amino acids with differential temporal trajectories between the groups can be classified into three categories. The amount of branched-chain amino acids (leucine, isoleucine, and valine) decreased in the feces of the IG (Fig. 5B). In contrast, plasma branched-chain amino acids were reduced in both groups (Figure S12A). Similarly, levels of fecal aromatic amino acids (phenylalanine, tryptophan, and tyrosine) diminished over time in the IG relative to the CG. Plasma levels of aromatic amino acids decreased in both groups (Fig. S12B). We also found significant interactions between group and time for derivatives of amino acids. Those increasing in the plasma of the IG while remaining stable in the CG, include (1) indole-3-propionic acid (IPA) (Figure S12C), a reductive product of tryptophan formed by bacteria in the gastrointestinal tract; (2) methylcysteine and pipercolic acid, which are known to be products of legume digestion⁴² (Fig. 5C). In contrast, 1-methylhistidine, a metabolite derived from meat, decreased in both the feces and plasma of the IG while remaining stable in the CG (Fig. 5B-C).

The temporal dynamics of bile acids were also significantly different between treatment groups with most intestinal bile acids downregulated over time in the IG. For example, intestinal bile acids of the lithocholic acid species decreased over time (12-ketoLCA, GLCA, isoLCA and LCA-3S) (Fig. 5B), but GLCA-3S was found to increase in the plasma in the IG (Fig. 5C). Intestinal bile acids of the hyocholic acid species (HCA and HDCA) decreased over time in the IG, but we did not detect any significant changes to hyocholic acid derivatives in plasma (Fig. 5B-C).

Metabolites act as mediators between alterations in microbiota and clinical outcomes

We subsequently investigated the significant correlations among microbiota alterations (species: $n = 25$, genera: $n = 15$), metabolite variations (fecal: $n = 58$, plasma: $n = 15$), and improvement in clinical outcomes ($n = 15$). Network analysis, performed separately within the two groups demonstrated that fecal metabolites exhibited more extensive associations with gut microbiome species and clinical indicators than plasma metabolites. Positively responding species *E. rectale*, *R. faecis*, *R. bromii* and *B. ovatus* were negatively correlated with lipid markers such as TC and LDL-C. HbA1c, however, was positively correlated with negative responding species (Fig. 6). This finding may imply that alterations in gut flora could be associated with enhanced lipid and blood glucose levels in individuals with pre-diabetes, potentially through diverse mechanisms. On the other hand, negatively responding species were positively correlated to many fecal metabolites. Within the network, *B. wadsworthia*, *R. torques* and genus *Fusobacterium* emerged as pivotal microorganisms with the highest degree of connections, while HbA1c, TC, and LDL-C stood out among the clinical indicators with the highest degree of connections, a correlation not observed in the CG (Fig. 6A, Figure S13A). In contrast, network analysis of plasma metabolites showed that only a few metabolites were associated with species and clinical indicators, including methylcysteine and pipercolic acid (Fig. 6B, Fig. S13B).

Bi-directional mediation analysis was performed to investigate the links between species, metabolites and clinical outcomes in both groups independently. We discovered 33 and 14 forward mediation linkages for the fecal and plasma metabolites respectively (Fig. 7A, 7C), and 15 and 9 reverse mediation linkages for fecal and plasma metabolites, respectively, specifically in the IG (Figure S14A-B). No significant mediation linkages were found in the CG. For fecal metabolites, most of these linkages were related to the impact of *B. wadsworthia*, *R. hominis* and *R. gnavus* on HbA1c and TC (Figure 7A). We observed that b-deoxycholic acid (bDCA) and norcholic acid (NorCA) could mediate the impact of *B. wadsworthia* and *R. hominis* on HbA1c, respectively. We also observed that 3-hydroxyisovaleric acid

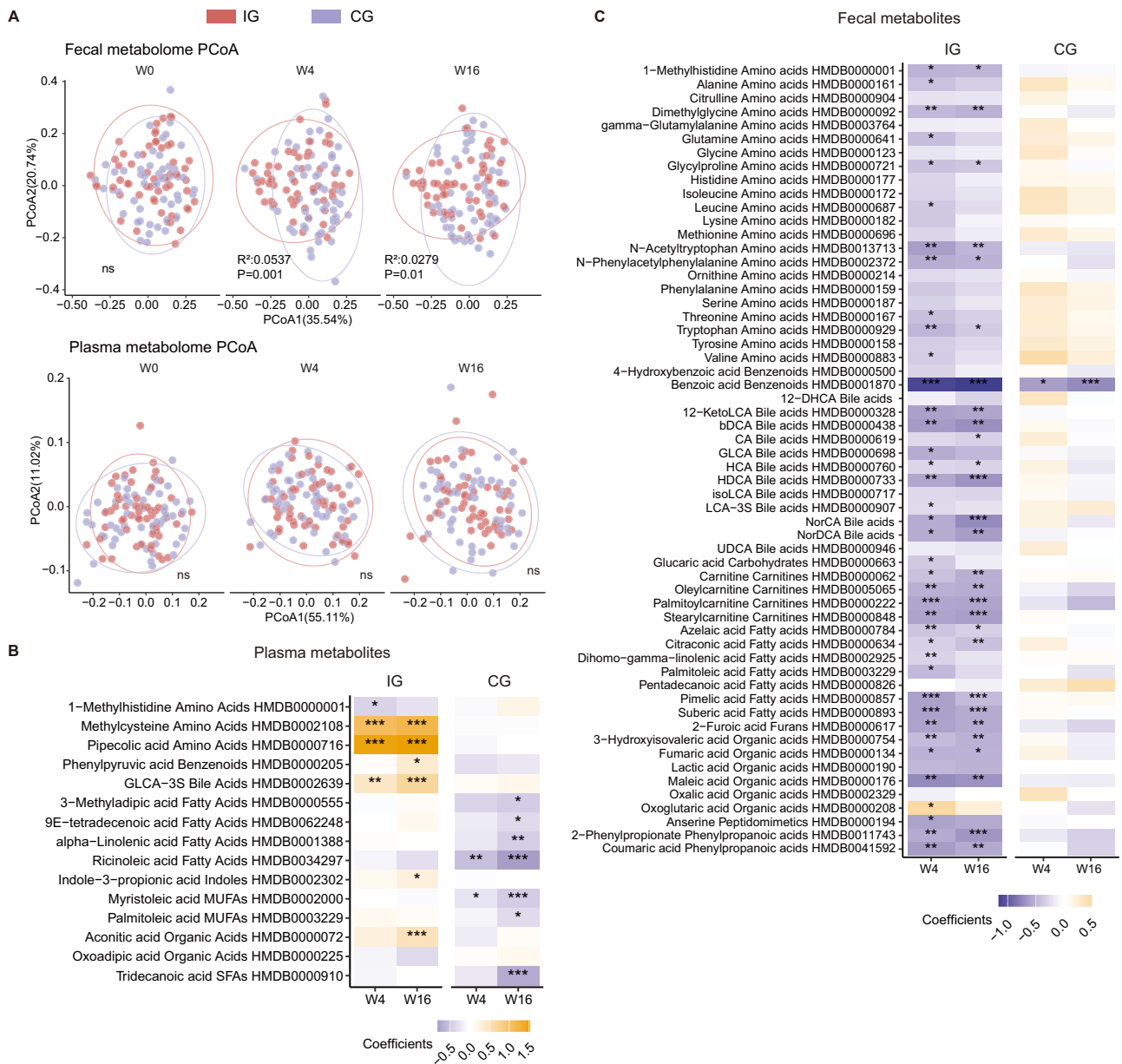


Fig. 5 | Alterations in fecal and blood metabolite levels. A Principal coordinate analysis of fecal and plasma metabolome based on Bray-Curtis distances. R^2 values and p -values were calculated from the two-sided PERMANOVA test. All three time points share the same axes, showing the first two principal coordinates and labeled with the percent variance explained. **B, C** Heatmap of q -values and beta coefficients of fecal metabolites (**B**) and plasma metabolites (**C**) across the whole study period.

* q ($Time$) < 0.05, ** q ($Time$) < 0.01, and *** q ($Time$) < 0.001 denote p -adjusted values for the time coefficient from the LMM model (two-sided) performed separately for each group. p -values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. See Supplementary Data 1 for full names of metabolite abbreviations.

and oleylcarnitine could mediate the impact of *B. wadsworthia* and *R. gnavus* on TC, respectively ($FDR_{forward\ mediation} = 3.24 \times 10^{-14}$, Fig. 7B). On the other hand, most plasma metabolite linkages were related to the impact of *R. hominis* and *R. faecis* on LDL-C, mediated by pipecolic acid, methylcysteine and GLCA-3S respectively (Fig. 7C-D).

Discussion

Several RCTs have reported dietary intervention for managing pre-diabetes. For example, adherence to Mediterranean diet lowered HbA1c^{43,44}; carbohydrate restriction led to improvement in several health parameters⁴⁵; low calorie, low fat diet along with increased physical activity reduced diabetes conversion risk from pre-diabetics^{46,47}; high protein diet led to remission of pre-diabetics⁴⁸; long-term almond consumption led to better insulin responses⁴⁹; etc.

Our RCT is one of the first studies in Asia exploring whether a multi-component dietary intervention on the background of calorie restriction can improve metabolic health in a population of pre-diabetes. Both hypocaloric intervention and control diets led to significant weight loss and improvements in body composition parameters over time. This, along with corresponding improvements in glycemic control and lipid profile, is consistent with prior observations on metabolic health benefits conferred by weight loss⁵⁰⁻⁵². Furthermore, we showed that the intervention diet did result in relatively greater improvements in TC, LDL-C and HbA1c compared to the control diet. Simultaneously, alterations in gut microbiome were accompanied by functional changes in carbohydrate fermentation and utilization in the gut, resulting in metabolite changes (SCFA, amino acids and bile acids). We observed inter-relationships between the gut

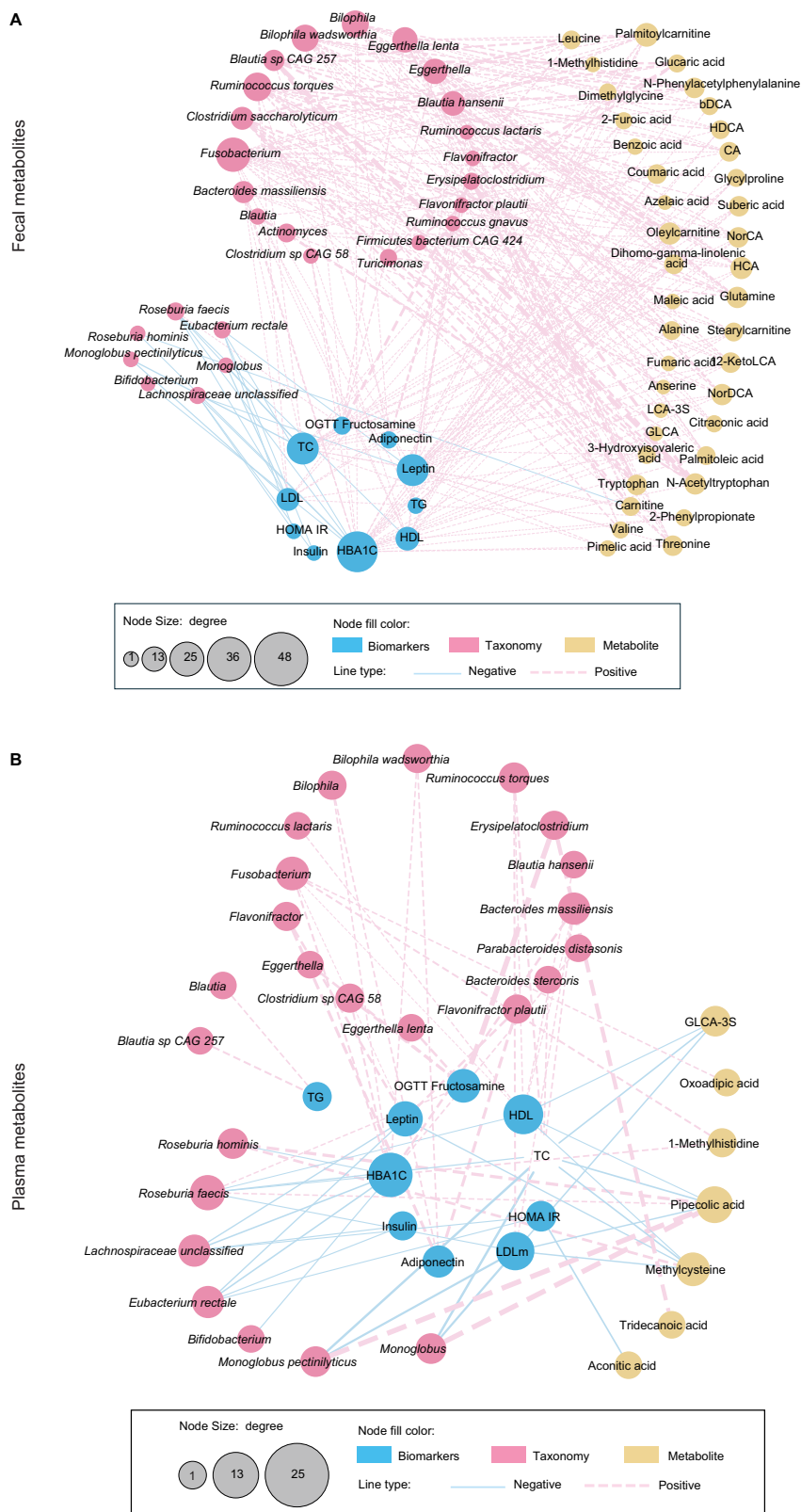


Fig. 6 | Interactions between microbiota, metabolites and blood biomarkers. Network diagram illustrating correlations among taxonomy, metabolites, and clinical indicators at all time points in the IG. **(A)** for fecal metabolites; **B** for plasma metabolites. Network analysis involving significantly altered taxonomy (species and genera) and metabolites: displaying links with $q_{(Treatment \times Time)} < 0.25$ and

$q_{(Time)} < 0.05$. Node size reflects connection count and colors represent different data types. Line width between nodes signifies correlation strength, and colors indicate correlation direction. See Supplementary Data 1 for full names of metabolite abbreviations and The Human Metabolome Database (HMDB) numbers.

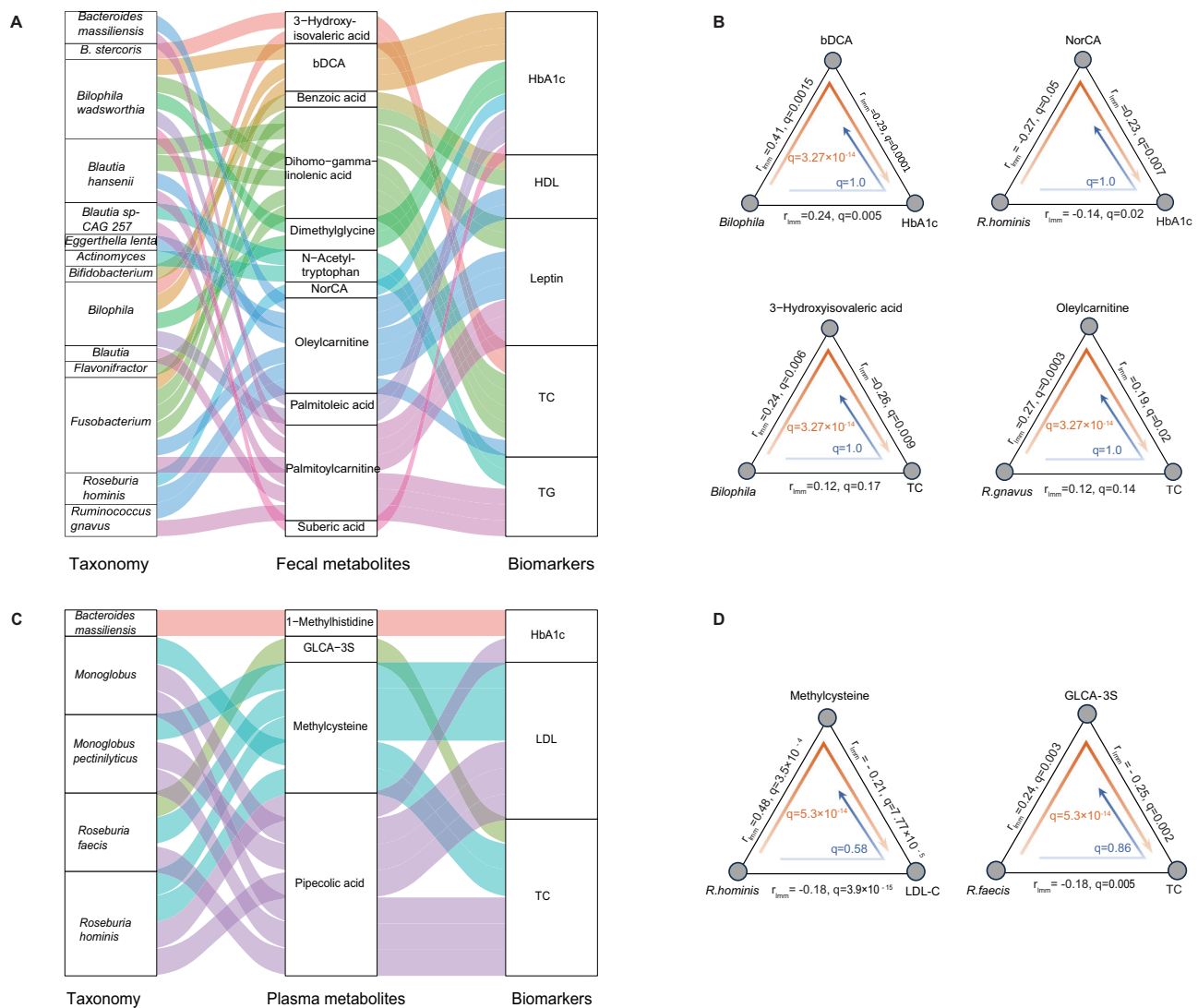


Fig. 7 | Mediation analysis identifies linkages between the species, metabolites and blood biomarkers. **A** Parallel coordinates chart showing the 33 mediation effects of the fecal metabolites that were significant at $q < 0.05$. Shown are taxonomy (left), fecal metabolites (middle) and clinical outcomes (right). The curved lines connecting the panels indicate the mediation effects, with colors corresponding to different metabolites. **B** Mediation effect of some highlighted fecal metabolites on the levels of HbA1c and TC. **C** Parallel coordinates chart showing the 14 mediation effects of the plasma metabolites that were significant at $q < 0.05$. Shown are taxonomy (left), plasma metabolites (middle) and clinical outcomes

(right). The curved lines connecting the panels indicate the mediation effects, with colors corresponding to different metabolites. **D** Mediation effect of some highlighted plasma metabolites on the levels of LDL-C and TC. For **(B)** and **(D)**, the grey lines indicate the associations between the two factors, with corresponding LMM coefficients and q-values. Direct mediation is shown by a red arrow and reverse mediation is shown by a blue arrow. Corresponding q-values from mediation analysis are shown. See Supplementary Data 1 for full names of metabolite abbreviations and HMDB numbers.

microbiome, metabolites and clinical outcomes, suggesting possible mechanisms for alleviating prediabetes.

Consumption of legumes has been shown to reduce TC, LDL-C and HbA1c, except in one study showing increased HbA1c⁵³. However, the latter was done over a short 8-week period without a control group⁵³. Meta-analysis⁵⁴ show beneficial effects with legume consumption were seen in RCTs in the context of weight loss^{55–58} and without weight loss^{59–64}. Beside soluble fiber, legumes also contain various phytonutrients known for their hypocholesterolemic effects⁶⁵. In agreement with our results, a previous study reported that participants who consumed ~190 g of legumes/day had similar changes in TC and HbA1c²¹. Other works also reported significant reductions in FG and HOMA-IR in type 2 diabetics following a legume-based DASH diet compared to control DASH diet^{66,67}. Notably, participants in both studies were on oral antihyperglycemic agents, which may have potentiated the benefits from the food-based interventions.

Whilst most studies highlight the importance of dietary changes in managing T2D risk, in general they did not elucidate the mechanistic underpinnings of how specific dietary modifications may impact the risk of progression to T2D in at-risk populations. More recently, it has been suggested that gut microbiome changes as a result of diet adaptation^{32–34,68–70} and may significantly influence glucose metabolism and insulin sensitivity, thereby playing a crucial role in the management of prediabetes and reducing the risk of T2D^{71–74}. Therefore, our RCT additionally attempted to elucidate the links between a legume-rich intervention diet with corresponding changes in the markers of T2D risk as well as changes related to gut microbiome and microbial-related metabolites.

Previous studies showed that bacterial species, such as *Eubacterium rectale*, *Roseburia faecis* and genus *Bifidobacterium*, are altered by fiber interventions [15]⁷⁵, promote the growth of SCFA producers²⁸ and confer health benefits such as reducing

diabetes risk^{76,77}. Similarly in our study, greater reductions were observed in IG in species belonging to the genera *Bacteroides* and *Ruminococcus*, such as *Ruminococcus torques*, *Ruminococcus gnavus* and *Ruminococcus lactaris*. These species were previously linked to diet lacking in fiber and inversely associated with Mediterranean diet^{72,75}. They have been reported to be related to intestinal health impairments and metabolic dysregulations^{78–81}.

Given that gut microbiota comprises a wide range of metabolic functions that aids in the breakdown of complex nutrients and non-nutrients, species perturbations observed within IG may have influenced metabolic activity as predicted from metagenomic functions. Analysis of genomic representation of glycoside hydrolases and CAZY showed that the intervention diet led to an increased ability to hydrolyse plant cell wall glycans with a decrease in the ability to digest simpler carbohydrates. Genes involved in the SCFA production pathways were pertinent to our understanding of fiber degradation. Among the 4 SCFA pathways, lysine degradation genes were downregulated in IG, suggesting that SCFA production under the influence of a high fiber diet may be via acetyl-CoA, aminobutryl and glutarate pathways. Intriguingly, legumes are high sources of lysine, methylcysteine and pipercolic acid⁸². Pipercolic acid is also a lysine degradation product via the piperidine intermediate, previously reported to be associated with reduction of inflammatory cytokines⁸³. Given that both methylcysteine and pipercolic acid concentrations were increased in the plasma metabolites but not lysine in IG, this may suggest that methylcysteine and pipercolic acid could be used as biomarkers for legume consumption. Conversely, 1-methylhistidine, which was downregulated in the feces and plasma of IG is recognized as a cardiovascular risk factor previously demonstrated to be substantially lower in vegetarians compared to non-vegetarians³⁴. Taken together, these amino acid changes suggest protective effects against metabolic disorders.

Multiple studies have highlighted the detrimental effects of branched-chain amino acids (BCAAs) in the blood on metabolism^{84,85}. One mouse study discovered that oral administration of *Parabacteroides merdae* (Pm) promoted the degradation of gut BCAAs into branched-chain short-chain fatty acids. This reduces blood levels of BCAAs, inhibits activation of mTORC1 pathway in arterial plaque macrophages, and exhibits anti-atherosclerotic effects⁸⁶. Previous research also indicated that elevated blood concentrations of IPA, a derivative of tryptophan, are associated with reduced risk of CAD and T2D. Tryptophan metabolism also leads to the production of kynurenin, which modulates insulin sensitivity⁸⁷. Furthermore, IPA has been shown to modulate intestinal barrier function through activation of the xenobiotic sensor, pregnane X receptor (PXR)^{40,88,89}. Lastly, imidazole propionate (IMP), a derivative of histidine, is known to hinder insulin signaling through the activation of mTORC1 and inhibits metformin^{90,91}. Reductions in both intestinal and plasma IMP observed in IG may be a mediator of the diet-related health outcomes.

Endogenous cholesterol is utilized to produce liver bile acids. 95% of bile acids are reabsorbed by intestinal villi and transported back via the hepatic portal vein. Reduced levels of intestinal lithocholic acid (LCA), deoxycholic acid (DCA), cholic acid (CA), together with decreased LDL-C and TC in IG may suggest reduced enterohepatic circulation of bile acids to the liver; inducing the liver to promote conversion of more cholesterol into bile acids. This results in the reduction of LDL-C synthesis. In the blood, only GLCA-3S was increased and may form an alternative explanation to the overall decrease in intestinal bile acids via gut microbiome modifications. GLCA-3S was previously shown to be correlated to dietary diversity⁹².

Network and causal mediation analysis further strengthened our findings by linking up the changes previously seen in the gut microbiome composition, metabolites with biomarker outcomes. Decreased *B. wadsworthia* in IG is correlated to a decrease in HbA1c, mediated by the decrease in intestinal bile acid bDCA. *B. wadsworthia* was reported to be enriched in T2D patients⁹³. Likewise, increased *R. bromii* in IG is

correlated to a decrease in LDL-C, mediated by an increase in plasma bile acid GLCA-3S. *R. bromii* was previously reported to be negatively correlated to apolipoprotein B-100, a carrier for LDL-C⁹⁴. These mediations in our study may further link secondary bile acids and warrant further research.

The strength of this study was that the intervention dietary regime, where 2 out of 3 main meals per day were provided (contributing to ~60% total energy and >66% total protein intake) and the remaining dietary intake, was controlled in terms of total calorie intake with regular 4-weekly assessment and dietary consultations with a research dietitian. This has enabled greater control of diet quality, quantity, and allowed some degree of flexibility for such studies to be undertaken in a 'free-living' environment and facilitating longer-term adherence. The excellent retention rate in our study population and high reported compliance support this.

While the multicomponent nature of the intervention diet involving a hypocaloric diet alongside the use of legumes, blended oil, low GI foods and spices makes it challenging to pinpoint the contribution of individual components to the overall effect, nonetheless several recent reports have highlighted that dietary patterns are more effective than individual foods and nutrients⁹⁵. One potential limitation in this partial feeding study design is the fixed dose of intervention meal components, which may lead to differences in the relative proportions of calories from non-intervention foods in different individuals, as their total caloric requirements can vary. However, this particular study design was deemed necessary to promote compliance to the demanding study protocol over an extended period (16 weeks) and to provide better external validity, since from a public health recommendation, this dietary advice will be easier to follow than to tailor specific intervention components according to individual body weight.

In summary, we demonstrated that a multi-component diet rich in legumes, low GI starches, healthy vegetable oils and spices can improve cardiometabolic health above and beyond effects that could be achieved through calorie restriction alone. This creates opportunities to design effective, longer-term food-based dietary interventions in Asia to reduce risks of T2D and CVD, especially in populations with greater predisposition to such diseases. We have also demonstrated that the favourable effects on the cardiometabolic biomarkers may be mediated through the composition and function of the gut microbiome, although some of these associations need to be further validated in other independent cohorts.

Methods

Study design and participants

This single-blinded, parallel-design clinical trial was registered at clinicaltrials.gov (identifier no. NCT04745702), approved by National Healthcare Group (NHG) Domain Specific Review Board ethics committee, Singapore (reference: C/2019/00997) and conducted in accordance with the Declaration of Helsinki 1983 and the Singapore Good Clinical Practice guidelines. The study was performed in prediabetic men and women of Chinese ethnicity in Singapore, aged between 45 to 75 years old with body mass index (BMI) between 19.5 to 32.0 kg/m² who gave written informed consent. Sex of the participants was determined based on self-reporting. According to the American Diabetes Association (ADA) Guideline, prediabetic was defined as having one or more of the following criteria: (1) fasting blood glucose >5.5 mmol/l and <7.0 mmol/l, (2) oral glucose tolerance test (OGTT) level ≥7.8 mmol/l and ≤11.0 mmol/l, (3) Hemoglobin A1c (HbA1c) level ≥5.7 and ≤6.4%. The main exclusion criteria included smoking, food allergies or intolerances to common food ingredients, dietary restrictions, sustained hypertension (>160/95 mmHg), history of metabolic, cardiac, liver or kidney disorders, thyroid dysfunction and recent weight loss.

Recruitment was conducted via newspaper advertisements, by word of mouth as well as using the recruitment database of the

research center [Clinical Nutrition Research Center (CNRC), Singapore]. Interested volunteers attended a Screening and Consent Visit following an overnight fast for assessment of study suitability using a Health and Lifestyle Questionnaire, anthropometric measurements as well as prediabetic screening based on the ADA criteria. The main measurements taken included volunteers' height (Seca 217; Seca GmbH, Germany), weight (Tanita BC-418; Tanita, Japan), FG/OGTT level (HemoCue® 201; Radiometer, Denmark) and HbA1c level (DCA Vantage Analyser; Siemens, Tarrytown, USA). The 2-hour OGTT was performed using 75-g glucose solution according to American Diabetes Association (ADA)⁹⁶. The trial was conducted from November 2020 to July 2021. All participants were reimbursed for their travel expenses and time.

Dietary intervention and assessment

Eligible volunteers were randomly assigned 2 intervention arms i.e. intervention and control. Randomization was done separately for each sex. The initial random assignment of one volunteer into each of the intervention arms was conducted using GraphPad prism and this is subsequently followed by minimization method in R. The randomization by minimization technique was applied in assigning volunteers to the treatment arms with a 1:1 allocation criteria. The BMI measured during the screening visit was used as the covariate to ensure the two intervention arms are balanced. The method used for randomization was developed in R using the methods suggested by Lin, Zhu⁹⁷.

Throughout the 16-week period, volunteers replaced 2 main meals/day with 2 specially formulated meals provided as frozen, pre-cooked low-calorie ready meals (approximately 400 kcal each). Volunteers in the intervention group (IG) received meals consisting of legumes (mixed beans, red kidney beans or chick peas, 100 g per meal), a portion of meat analogues (textured vegetable soy proteins), a portion of vegetables, low GI rice or noodles, with added mixed herbs and spices, and prepared with blended vegetable oil⁹⁸. Those in the control group (CG) were given meals consisting of a portion of meat (chicken or fish), a portion of vegetables, medium to high GI rice or noodles, with negligible herbs and spices, and prepared with refined corn oil. The meals in both groups were isocaloric. Additionally, male volunteers were provided with snacks containing either wholegrain cereals and nuts/seeds (IG) or refined cereal snack products (CG) to supplement their meals (~100 kcal/serving of snack). The total energy, macronutrient and sodium contents of the provided meals and snacks are shown in Table S2.

The remaining daily foods and snacks were *ad libitum* with individualized recommendations to limit caloric intake to induce a 5% weight loss in all volunteers (both CG and IG). These recommendations were given once every 4 weeks by a research dietitian. Advice given were non-biased but tailored to each volunteer's estimated energy requirements (calculated using Oxford Equation⁹⁹) with an estimated 500 kcal deficit. Only quantitative dietary advice on total caloric intake was provided to the volunteers with no qualitative recommendations provided. Volunteers were encouraged to retain their background habitual diet for the remainder of the daily foods and snacks and to maintain their habitual physical activity (PA) level throughout the study period. This was done to ensure no confounding factors from the involvement of research dietitian in the study.

Volunteers were required to attend measurement visits every 4 weeks with additional compliance checks every 2 weeks. Prior to each measurement visit, volunteers were required to complete a 3-day food diary for dietary assessment purposes. The food records were entered into the FoodWorks 10 Professional software system (version 10.0.4266, Xyris Software Pty Ltd, Queensland, Australia) and the volunteers' estimated energy, macronutrients and sodium intake were calculated. Food databases within Nutritionist Buddy (nBuddy)'s app¹⁰⁰ and individual foods and drinks' Nutrition Information Panel (NIP) were also referred to, especially for local foods with no available

data within databases in FoodWorks system. Random qualitative checks on the food record entry were performed by another researcher within the study team. To monitor volunteers' compliance to the study meals, they were also asked to complete the Intervention Food Record (IFR) daily throughout the study period. Intervention Food Record was kept by each volunteer to monitor reported compliance to the dietary intervention. Every time the volunteers finished an intervention meal, they were instructed to note down the specific meal code within the appropriate column. From the record, we analyzed whether the participants consumed 2 meals/day as per protocol. If participant missed a meal, they were asked to write in their reasons under the 'Remarks' column.

Physical activity assessments

The physical activity level of volunteers was assessed using the International Physical Activity Questionnaire (IPAQ; long-form) when they attended measurement visits. Volunteers were asked to complete 27 questions in the IPAQ by recalling their activities in the past seven days, according to 5 categories: 1) job-related PA; 2) transportation PA; 3) housework, house maintenance, and caring for family; 4) recreation, sport, and leisure-time PA; and 5) time spent sitting¹⁰¹. IPAQ data was analyzed using the standardized IPAQ scoring protocol¹⁰² as metabolic equivalent (MET)-minutes per week. This was calculated by multiplying MET intensity with minutes of each activity throughout the seven days followed by averaging MET minutes per day. Incomplete IPAQ data for specific visits or data with a total PA exceeding 960 min were excluded from analysis. Input data that were between 0 and 10 min were rounded up to 10 min.

Clinical assessments

The primary outcomes for this study include several markers of glucose homeostasis including HbA1c, FG, fasting insulin and OGTT test results. The secondary outcomes were measures of lipid profile, fructosamine, and adiponectin. We also measured other outcomes like body composition, gut microbiome, gut and plasma metabolites. Clinical assessments were carried out after an overnight fast (minimum of 10 h) at 4-week intervals on Weeks 0, 4, 8, 12 and 16. During measurement visits, HbA1c point-of-care testing (DCA Vantage Analyser; Siemens, Tarrytown, USA), blood pressure (Omron HEM907, Japan) and various anthropometric measurements were taken. These include height using a stadiometer (Seca 217; Seca GmbH, Germany), body weight and body fat percentage using bioelectrical impedance (BIA) (Tanita BC-418; Tanita, Japan) and whole body composition by Dual-Energy X-Ray Absorptiometry (DXA) scan (QDR 4500 A, fan-beam densitometer; Hologic, Waltham, MA, USA) according to manufacturer's instructions. Waist circumference was measured using a measurement tape at halfway between the lower ribs border and the iliac crest in a horizontal plane; whilst hip circumference was measured at the widest points just over the buttocks. Neck circumference was measured using the same measurement tape at mid-neck height, perpendicular to the neck's long axis, below the laryngeal prominence (for male volunteers). The mid-upper arm circumference was taken using a measurement tape at the mid-point between the acromion process and the olecranon. Volunteers were required to remove all metal items before measurement to prevent any interference during the scanning process. Manufacturer provided the software (version 8.21) to generate and analyse the data. The "android region" spans the area between the ribs and the pelvis, while the "gynoid region" overlaps the leg and trunk, covering the hips and upper thighs. The software also estimated the visceral fat mass within the L1-L5 regions. Overnight fasted blood was also drawn by venipuncture (using needle and syringe) during each measurement visit. We calculated the HOMA-IR using the below formula:

$$\text{HOMA} - \text{IR} = (\text{Fasting glucose} \times \text{Fasting Insulin}) / 22.5$$

Blood sampling and analyses

Blood samples were drawn into two different types of Vacutainer tubes (BD, USA) i.e. serum and EDTA tubes for primary and secondary outcomes analyses. Serum tubes were left to stand until the samples have clotted before centrifugation was carried out within 1 h of collection at $1500 \times g$ for 10 min at 4 °C. EDTA tubes were centrifuged within 45 min of collection at the same settings. The supernatant from the samples were aliquoted into cryotubes and stored at $-80\text{ }^{\circ}\text{C}$ until required for analyses. Serum glucose, insulin, and blood lipid panel were analyzed using the standardized clinical chemistry method of the National University Hospital Referral Laboratories, Singapore. Glucose and blood lipid panel [including TC, LDL-C (direct measurement), HDL-C, and TG] were measured with an AU 5800 clinical chemistry analyzer (Beckman Coulter, Inc., USA) using an enzymatic colorimetric assay, and insulin was measured using UniCel Dxl 800 Access Immunoassay System (Beckman Coulter, Inc.).

Adiponectin and leptin

Serum adiponectin and leptin levels were measured by enzyme-linked immunosorbent assay (ELISA), according to the instructions provided in commercially available kits (Human Total Adiponectin/Acrp30 Immunoassay, DRP300, Quantikine ELISA and Human Leptin ELISA Kit, RAB0333, Sigma-Aldrich).

Fecal sample collection

Volunteers who opted in for stool samples collection were instructed to do so at their own homes prior to arriving at the research center during each measurement visit at week 0, 2, 4, 8, 12 and 16. Collection of the stool samples was performed according to the manufacturer's instructions using the provided kits. This included a DNA/RNA Shield fecal collection tube (Zymo Research, California, USA), the OMNI-met.GUT kit (DNA Genotek, Ottawa, Canada) and BIOME-Preserve (Anaerobe Systems, California, USA) which preserves anaerobic microbiome communities at room temperature. The collected samples were transported in an insulated aluminium foil bag. Samples were aliquoted and stored in $-80\text{ }^{\circ}\text{C}$ when received in the laboratory.

DNA isolation and metagenomics sequencing

DNA was isolated from 450 μL of stool sample aliquots, preserved in DNA/RNA shield (Zymo Research). After treatment with 20 μL of 5 $\mu\text{g}/\text{mL}$ lysozyme (Sigma-Aldrich), 20 μL of 10 mg/mL Proteinase K (Promega) and 10 μL of 10 mg/mL RNaseA (Novagen) at 37 °C for 15 min then at 60 °C for 15 min, samples were homogenized twice using Lysing Matrix E tubes (MP Biomedicals) at 6 m/sec, for 40 sec. After centrifuging, the supernatant was treated by 10 μL of 10 mg/mL RNaseA (Novagen) at 60 °C for 30 min. The mixture was purified by phenol: chloroform extraction (1:1) with subsequent DNA isolation using Maxwell® 16 FFS Nucleic Acid Extraction Kit (Promega) according to manufacturer protocol. For sequencing, PCR-free genomic DNA library was used for shotgun Illumina sequencing using Novaseq 6000, delivering at least 7 GB of data per sample (Novogene) (Supplementary Data 2). The workflow was validated by using ZymoBIOMICS Gut Microbiome Standard (Zymo Research) before large scale DNA isolation and sequencing.

Fecal and blood metabolome

Fecal aliquots (sample number: 333), preserved in OMNI-met®GUT(ME-200) kit (DNA Genotek) and blood aliquots (sample number: 332) of three time points (W0, W4 and W16) were used for metabolome analysis.

Metabolites extraction

Samples were thawed on an ice-bath to minimize sample degradation. 20 μL of plasma or fecal sample was added to a 96-well plate. The plate was then transferred to the Eppendorf epMotion Workstation. 120 μL

ice-cold methanol with partial internal standards was automatically added to each sample and vortexed vigorously for 5 minutes. The plate was centrifuged at 4000 g for 30 minutes. 30 μL of supernatant was transferred to a clean 96-well plate, and 20 μL of freshly prepared derivative reagents was added to each well. The plate was sealed and the derivatization was carried out at 30 °C for 60 min. After derivatization, 330 μL of ice-cold 50% methanol solution was added to dilute the sample. The plate was then stored at $-20\text{ }^{\circ}\text{C}$ for 20 minutes followed by 4000 g centrifugation at 4 °C for 30 minutes. 135 μL of supernatant was transferred to a new 96-well plate with 10 μL internal standards in each well. Serial dilutions of derivatized stock standards were added to the left wells. Finally the plate was sealed for LC-MS analysis. To ensure quality control of the experiment, we prepared 31 QC samples each for fecal and plasma samples. The QC samples were equal-volume mixtures of the experimental samples.

HPLC-MS/MS analysis

An ultra-performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) system (ACQUITY UPLC-Xevo TQ-S, Waters Corp., Milford, MA, USA) was used to quantitate targeted metabolites using the N300 metabolomics kit in Novogene Co., Ltd. (Beijing, China). Samples were injected onto an ACQUITY UPLC BEH C18 1.7 μM VanGuard pre-column (2.1 \times 5 mm) and an ACQUITY UPLC BEH C18 1.7 μM analytical column (2.1 \times 100 mm) using a 18-min linear gradient at a flow rate of 0.4 mL/min for the positive/negative polarity mode. The eluents were eluent A (0.1% Formic acid-water) and eluent B (acetonitrile:IPA = 70:30). The solvent gradient was set as follows: 0-1 min (5% B), 1-11 min (5-78% B), 11-13.5 min (78-95% B), 13.5-14 min (95-100% B), 14-16 min (100% B), 16-16.1 min (100-5% B), 16.1-18 min (5% B). Xevo TQ-S mass spectrometer was operated in positive (negative) polarity mode with a Capillary of 1.5 (2.0) Kv, Source Temperature of 150 °C, Desolvation Temperature of 550 °C, Desolvation Gas Flow of 1000 (L/Hr).

Metabolites identification and quantification

The detection of the experimental samples using MRM (Multiple Reaction Monitoring) were based on Novogene self-built method. The Q1, Q3, RT (retention time), DP (declustering potential) and CE (collision energy) were used for the metabolite identification. The ratio of Q3 peak area of compound to internal standard was used for the standard curve. The concentration of the compound was calculated from the known internal standard concentration. The data files generated by UPLC-MS/MS were processed using the MassLynx Version 4.1 to integrate and correct the peak. For data below detection limits, k-nearest neighbors (kNN) was used for imputation. The list of metabolites is provided in Supplementary Data 1.

In vitro validation experiments

We performed in vitro experiments using fecal samples from six participants (three from the intervention group and three from the control group), which were introduced into a medium infused either with in vitro digested or undigested red kidney bean powder.

All preparation procedures and experiments were done under anaerobic conditions. To prepare inoculant, total six fecal samples, three from intervention and three from control groups, preserved in BIOME-Preserve Anaerobic Microbiome Collection Kit (Anaerobe Systems), were strained through the 100 μm nylon mesh (Fisherbrand). Amount of inoculant for each sample was adjusted to get final $\text{OD}_{600} = 0.02$ in the 10 mL of growth media (KH_2PO_4 10.9 g/L, K_2HPO_4 3.5 g/L, NaHCO_3 2 g/L, NaCl 1.242 g/L, NH_4Cl 2.5 g/L, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0.5 g/L, CaCl_2 3 mM, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 5 mg/L, MES 50 mM, mucin 4 g/L, Tween-80 0.25 mL/L, cholate 0.5 g/L, chenodeoxycholate 0.5 g/L, vitamin K_1 10 $\mu\text{L}/\text{L}$, folate 2 mg/L, vitamin B_{12} 0.1 mg/L, L-cysteine 0.3%). The media was degassed in an anaerobic chamber overnight before adding vitamins and L-cysteine; pH adjusted to 7.0. Each sample was

growing in media containing 10 g/L of fiber originating from cooked digested or undigested red kidney bean or brown rice powder at 37 °C for 48 hours under anaerobic conditions. An in vitro digestion of substrates was carried out based on the INFOGEST international consensus protocol¹⁰³. Changes in the microbiome community were analyzed at 0, 24 and 48 h post-inoculation. To assess reproducibility, some samples were done in duplicates.

Bioinformatics processing

From the raw metagenomics sequencing data, low-quality reads were discarded by the sequencing facility and reads belonging to the human genome were removed by mapping the data to the human reference genome (version NCBI37) using KneadData (v0.12.0)¹⁰⁴ and Bowtie2 (v2.3.5.1)¹⁰⁵. Microbial taxonomic profiles were generated using MetaPhlan3 (v3.0.14). Microbial general pathways were determined using HUMAnN3 (v3.6.1)¹⁰⁴, which maps DNA/RNA reads to a customized database of functionally annotated pan-genomes. HUMAnN3 reported the abundances of gene families from the UniProt Reference Clusters (UniRef90)¹⁰⁶, which were further mapped to microbial pathways from the MetaCyc metabolic pathway database¹⁰⁷. The abundance of gene families was then mapped to Enzyme Commission (EC) annotations already present in the HUMAnN database. The enzymes related to glycoside hydrolases were defined as all the enzymes that start with an EC number of EC:3.2.1 and analyzed further.

High-quality paired-end reads from each sample were used for de novo assembly with Megahit (v1.2.9)¹⁰⁸ into contigs of at least 500 bp. Genes were predicted using MetaGeneMark (v3.38)¹⁰⁹. A non-redundant gene catalog of 14,386,339 microbial genes was constructed with Cd-hit (v4.8.1)¹¹⁰ using the parameters “-c 0.95”. Salmon (v1.9.0)¹¹¹ was used to estimate the gene abundance for every sample. Non-redundant gene set were mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) protein database¹¹² to gene function annotation using KofamScan (v1.3.0)¹¹³, resulting in the annotation of 7,568 KEGG orthology, with genes annotated to the same orthology summed together. The reads were also aligned to curated databases of proteins involved in SCFAs (acetic acid, propionic acid and butyric acid) production¹¹⁴ using DIAMOND search (v0.9.30)¹¹⁵. Genes encoding carbohydrate-active enzyme (CAZs) were identified using dbCAN2¹¹⁶. Abundances of genes that belonged to the same CAZY family were summed together.

Statistical analysis

Data analysis was performed on an intention-to-treat basis without missing value imputation. Data was checked for normality visually using histograms, boxplots and quantile-quantile plots.

Sample size calculations. Power calculation for sample size estimation was based on a between group difference in HbA1c change of about 0.25% (conservative estimate) at the end of the trial from a previous study, which was also considered to be clinically meaningful⁴⁸. The variation was assumed to be similar in our pre-diabetic population. With 80% power and alpha of 0.05, a total of 110 volunteers were needed. Taking into consideration an attrition rate of 13%, the final sample size required was 125. Sample size calculation was done using G*Power 3.1.9.2.

PERMANOVA analysis. To further evaluate the intervention effect on the microbiome composition over time, we used the “adonis2” function from the vegan package (version 2.6.4) in R (version 4.3.1) to construct PERMANOVA models. Each variable was tested separately with the “group x variable” interaction terms as independent variables:

$$\text{Bray} - \text{Curtis distance} \sim \text{Treatment} * \text{Variable} \quad (1)$$

Additionally, we also constructed separate PERMANOVA models to assess each variable independently within each treatment:

$$\text{Bray} - \text{Curtis distance} \sim \text{Variable} \quad (2)$$

In the above tests, each variable was tested independently to avoid issues with variable ordering, so that the variances explained by individual terms sum to total variance. Permutation procedure was stratified by subjects to account for longitudinal structure of the data with 999 iterations. To test whether the means of the groups were changed in PERMANOVA, we additionally performed beta dispersion test to ensure that it is not significant. To test whether the variance was contributed by earlier timepoints or later timepoints, we further performed PERMANOVA on subsets of the data based on time periods using the same formula above.

To ascertain the beta-diversity difference across time, we computed Bray-Curtis distance between samples from adjacent timepoints, e.g. Wx-y represents the Bray-Curtis distance between samples collected at Wx and Wy.

Linear mixed-effects model. To investigate both longitudinal primary outcome changes, we used Linear mixed effect models (LMM), for microbiome outcomes we used Linear models for differential abundance analysis (LinDA) from the MicrobiomeStat (version 1.2) package in R (version 4.3.1) which applies centered log ratio (CLR) transformation and corrects for biases due to compositional effects. Zero values were handled by imputation to avoid issues with pseudo-counts inflating the rate of false positive discoveries. Time was treated as a categorical variable for microbiome data, blood lipids and glycemic indices except for HbA1c, as changes were not assumed to be linear. We test the following hypothesis based on the formulas below:

To detect differences response between groups over discrete time, we used fixed and random components using the following formula:

$$\text{Response} \sim \text{Treatment} * \text{Time} + \text{Sex} + \text{Age} + (1, |, \text{SubjectID}) \quad (3)$$

Within each group, to detect differences across time, we used the following formula:

$$\text{Response} \sim \text{Time} + \text{Sex} + \text{Age} + (1, |, \text{SubjectID}) \quad (4)$$

To detect correlative longitudinal trends, we used fixed and random components using the following formula:

$$\text{Response} \sim \text{Variable} + \text{Sex} + \text{Age} + (1, |, \text{SubjectID}) \quad (5)$$

p-values from LMM and LinDA were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure. For differential longitudinal trajectories across groups, we calculate the *q*-values for the “Group:Time” interaction coefficients. For correlative trends, we calculate the *q*-values for the variable coefficients.

Casual Mediation Analysis. To investigate causal mediation effects, we used the mediation package (version 4.5.0) in R (version 4.3.1) and filtered the effects such that (1) the magnitude of the mediation effect to be in the same direction as total effect (ACME/ADE > 0), (2) control false discovery rate of the total effects ($\text{FDR}_{\text{total effect}} < 0.25$), and (3) control false discovery rate of the forward or reverse mediation ($\text{FDR}_{\text{forward/reverse mediation}} < 0.05$).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw metagenomics sequencing data generated in this study have been deposited in the Sequence Read Archive (SRA) database under accession code [PRJNA1201249](https://www.ncbi.nlm.nih.gov/sra/PRJNA1201249). The processed clinical biomarkers data and metabolomics data are available at zenodo [<https://doi.org/10.5281/zenodo.14329282>].

Code availability

Source code for scripts used to analyze the data are available in a GitHub project at <https://github.com/xiaorongsmile/FBIP>.

References

- Khan, M. A. B. et al. Epidemiology of Type 2 Diabetes - Global Burden of Disease and Forecasted Trends. *J. Epidemiol. Glob. Health* **10**, 107–111 (2020).
- Saeedi, P. et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabet. Res. Clin. Pract.* **157**, 107843 (2019).
- Neuenschwander, M. et al. Role of diet in type 2 diabetes incidence: umbrella review of meta-analyses of prospective observational studies. *BMJ* **366**, l2368 (2019).
- Esposito, K. et al. Which diet for prevention of type 2 diabetes? A meta-analysis of prospective studies. *Endocrine* **47**, 107–116 (2014).
- Kris-Etherton, P. M., Etherton, T. D., Carlson, J. & Gardner, C. Recent discoveries in inclusive food-based approaches and dietary patterns for reduction in risk for cardiovascular disease. *Curr. Opin. Lipido.* **13**, 397–407 (2002).
- Shirani, F., Salehi-Abargouei, A. & Azadbakht, L. Effects of Dietary Approaches to Stop Hypertension (DASH) diet on some risk for developing type 2 diabetes: A systematic review and meta-analysis on controlled clinical trials. *Nutrition* **29**, 939–947 (2013).
- Koloverou, E., Esposito, K., Giugliano, D. & Panagiotakos, D. The effect of Mediterranean diet on the development of type 2 diabetes mellitus: A meta-analysis of 10 prospective studies and 136,846 participants. *Metabolism* **63**, 903–911 (2014).
- Esposito, K., Maiorino, M. I., Bellastella, G., Panagiotakos, D. B. & Giugliano, D. Mediterranean diet for type 2 diabetes: cardiometabolic benefits. *Endocrine* **56**, 27–32 (2017).
- Martín-Peláez S., Fito M. & Castaner O. Mediterranean Diet Effects on Type 2 Diabetes Prevention, Disease Progression, and Related Mechanisms. A Review. *Nutrients* **12**, 2236 (2020).
- Liese, A. D., Nichols, M., Sun, X., Agostino, R. B. & Haffner, S. M. Adherence to the DASH Diet Is Inversely Associated With Incidence of Type 2 Diabetes: The Insulin Resistance Atherosclerosis Study. *Diab. Care* **32**, 1434 (2009).
- Azadbakht, L. et al. Effects of the Dietary Approaches to Stop Hypertension (DASH) Eating Plan on Cardiovascular Risks Among Type 2 Diabetic Patients. *Diab. Care* **34**, 55 (2011).
- Campbell, A. P. DASH Eating Plan: An Eating Pattern for Diabetes Management. *Diab. Spectr.: a Publ. Am. Diab. Assoc.* **30**, 76–81 (2017).
- García-Arellano, A. et al. Dietary inflammatory index and all-cause mortality in large cohorts: The SUN and PREDIMED studies. *Clin. Nutr.* **38**, 1221–1231 (2019).
- Okamoto, K. et al. Identification of KCNJ15 as a susceptibility gene in Asian patients with type 2 diabetes mellitus. *Am. J. Hum. Genet.* **86**, 54–64 (2010).
- Villegas, R. et al. Legume and soy food intake and the incidence of type 2 diabetes in the Shanghai Women's Health Study. *Am. J. Clin. Nutr.* **87**, 162–167 (2008).
- Becerra-Tomás, N. et al. Legume consumption is inversely associated with type 2 diabetes incidence in adults: A prospective assessment from the PREDIMED study. *Clin. Nutr.* **37**, 906–913 (2018).
- Qian, F., Liu, G., Hu, F. B., Bhupathiraju, S. N. & Sun, Q. Association Between Plant-Based Dietary Patterns and Risk of Type 2 Diabetes: A Systematic Review and Meta-analysis. *JAMA Intern. Med.* **179**, 1335–1344 (2019).
- Toumpanakis, A., Turnbull, T. & Alba-Barba, I. Effectiveness of plant-based diets in promoting well-being in the management of type 2 diabetes: a systematic review. *BMJ Open Diab. Res. amp; Care* **6**, e000534 (2018).
- Chen, Z. et al. Plant versus animal based diets and insulin resistance, prediabetes and type 2 diabetes: the Rotterdam Study. *Eur. J. Epidemiol.* **33**, 883–893 (2018).
- Chen, Z. et al. Changes in Plant-Based Diet Indices and Subsequent Risk of Type 2 Diabetes in Women and Men: Three U.S. Prospective Cohorts. *Diab. Care* **44**, 663 (2021).
- Jenkins, D. J. A. et al. Effect of legumes as part of a low glycemic index diet on glycemic control and cardiovascular risk factors in type 2 diabetes mellitus: a randomized controlled trial. *Arch. Intern Med* **172**, 1653–1660 (2012).
- Pearce, M. et al. Associations of Total Legume, Pulse, and Soy Consumption with Incident Type 2 Diabetes: Federated Meta-Analysis of 27 Studies from Diverse World Regions. *J. Nutr.* **151**, 1231–1240 (2021).
- Barclay, A. W. et al. Glycemic index, glycemic load, and chronic disease risk—a meta-analysis of observational studies. *Am. J. Clin. Nutr.* **87**, 627–637 (2008).
- Zafar, M. I. et al. Low-glycemic index diets as an intervention for diabetes: a systematic review and meta-analysis. *Am. J. Clin. Nutr.* **110**, 891–902 (2019).
- Evans, J. L., Goldfine, I. D., Maddux, B. A. & Grodsky, G. M. Are oxidative stress-activated signaling pathways mediators of insulin resistance and beta-cell dysfunction? *Diabetes* **52**, 1–8 (2003).
- Bi, X., Lim, J. & Henry, C. J. Spices in the management of diabetes mellitus. *Food Chem.* **217**, 281–293 (2017).
- Reynolds, A. N., Akerman, A. P. & Mann, J. Dietary fibre and whole grains in diabetes management: Systematic review and meta-analyses. *PLoS Med* **17**, e1003053 (2020).
- Zhao, L. et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Sci. (N. Y., NY)* **359**, 1151–1156 (2018).
- Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
- Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**, 565–569 (2016).
- Lynch, S. V. & Pedersen, O. The Human Intestinal Microbiome in Health and Disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).
- Fragiadakis, G. K. et al. Long-term dietary intervention reveals resilience of the gut microbiota despite changes in diet and weight. *Am. J. Clin. Nutr.* **111**, 1127–1136 (2020).
- Von Schwartzberg, R. J. et al. Caloric restriction disrupts the microbiota and colonization resistance. *Nature* **595**, 272–277 (2021).
- Chen, L. et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat. Med.* **28**, 2333–2343 (2022).
- Liu, H. et al. Ecological dynamics of the gut microbiome in response to dietary fiber. *ISME J.* **16**, 2040–2055 (2022).
- Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332–1345 (2016).
- Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J. & Hylemon, P. B. Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* **7**, 22–39 (2016).

38. Zheng, X. et al. Hyocholic acid species improve glucose homeostasis through a distinct TGR5 and FXR signaling mechanism. *Cell Metab.* **33**, 791–803.e797 (2021).
39. Zheng X. et al. Hyocholic acid species as novel biomarkers for metabolic disorders. *Nat. Commun.* **12**, (2021).
40. Xue, H. et al. Gut Microbially Produced Indole-3-Propionic Acid Inhibits Atherosclerosis by Promoting Reverse Cholesterol Transport and Its Deficiency Is Causally Related to Atherosclerotic Cardiovascular Disease. *Circulat. Res.* **131**, 404–420 (2022).
41. Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
42. Perera, T. et al. Identification and monitoring of metabolite markers of dry bean consumption in parallel human and mouse studies. *Mol. Nutr. Food Res.* **59**, 795–806 (2015).
43. Ruiz-Canela, M. et al. Plasma branched chain/aromatic amino acids, enriched Mediterranean diet and risk of type 2 diabetes: case-cohort study within the PREDIMED Trial. *Diabetologia* **61**, 1560–1571 (2018).
44. Martínez-González, M. A. et al. Yearly attained adherence to Mediterranean diet and incidence of diabetes in a large randomized trial. *Cardiovasc. Diabetol.* **22**, 262 (2023).
45. Summers, C., Tobin, S. & Unwin, D. Evaluation of the Low Carb Program Digital Intervention for the Self-Management of Type 2 Diabetes and Prediabetes in an NHS England General Practice: Single-Arm Prospective Study. *JMIR Diab.* **6**, e25751 (2021).
46. Knowler, W. C. et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med* **346**, 393–403 (2002).
47. Lindström, J. et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet* **368**, 1673–1679 (2006).
48. Stentz, F. B. et al. Remission of pre-diabetes to normal glucose tolerance in obese adults with high protein versus high carbohydrate diet: randomized control trial. *BMJ Open Diab. Res Care* **4**, e000258 (2016).
49. Gravesteyn, E., Mensink, R. P. & Plat, J. The effects of long-term almond consumption on whole-body insulin sensitivity, postprandial glucose responses, and 48 h continuous glucose concentrations in males and females with prediabetes: a randomized controlled trial. *Eur. J. Nutr.* **62**, 2661–2672 (2023).
50. Dattilo, A. M. & Kris-Etherton, P. M. Effects of weight reduction on blood lipids and lipoproteins: a meta-analysis. *Am. J. Clin. Nutr.* **56**, 320–328 (1992).
51. Pasanisi, F., Contaldo, F., de Simone, G. & Mancini, M. Benefits of sustained moderate weight loss in obesity. *Nutr. Metab. Cardiovasc. Dis.: NMCD* **11**, 401–406 (2001).
52. Magkos, F. et al. Effects of Moderate and Subsequent Progressive Weight Loss on Metabolic Function and Adipose Tissue Biology in Humans with Obesity. *Cell Metab.* **23**, 591–601 (2016).
53. Ferreira H. et al. Impact of a daily legume-based meal on blood and anthropometric parameters in a group of omnivorous adults: A pilot study. *Nutr. Bull.* (2024).
54. Bazzano, L. A., Thompson, A. M., Tees, M. T., Nguyen, C. H. & Winham, D. M. Non-soy legume consumption lowers cholesterol levels: a meta-analysis of randomized controlled trials. *Nutr. Metab. Cardiovasc. Dis.: NMCD* **21**, 94–103 (2011).
55. Abete, I., Parra, D. & Martinez, J. A. Legume-, fish-, or high-protein-based hypocaloric diets: effects on weight loss and mitochondrial oxidation in obese men. *J. Med Food* **12**, 100–108 (2009).
56. Hermsdorff, H. H., Zulet, M., Abete, I. & Martínez, J. A. A legume-based hypocaloric diet reduces proinflammatory status and improves metabolic features in overweight/obese subjects. *Eur. J. Nutr.* **50**, 61–69 (2011).
57. Tonstad, S., Malik, N. & Haddad, E. A high-fibre bean-rich diet versus a low-carbohydrate diet for obesity. *J. Hum. Nutr. Diet.* **27**, 109–116 (2014).
58. Pittaway, J. K., Robertson, I. K. & Ball, M. J. Chickpeas may influence fatty acid and fiber intake in an ad libitum diet, leading to small improvements in serum lipid profile and glycemic control. *J. Am. Diet. Assoc.* **108**, 1009–1013 (2008).
59. Zhang, Z. et al. A high legume low glycemic index diet improves serum lipid profiles in men. *Lipids* **45**, 765–775 (2010).
60. Abeysekera, S., Chilibeck, P. D., Vatanparast, H. & Zello, G. A. A pulse-based diet is effective for reducing total and LDL-cholesterol in older adults. *Br. J. Nutr.* **108**, S103–S110 (2012).
61. Winham, D. M., Hutchins, A. M. & Johnston, C. S. Pinto bean consumption reduces biomarkers for heart disease risk. *J. Am. Coll. Nutr.* **26**, 243–249 (2007).
62. Trinidad, T. P., Mallillin, A. C., Loyola, A. S., Sagum, R. S. & Encabo, R. R. The potential health benefits of legumes as a good source of dietary fibre. *Br. J. Nutr.* **103**, 569–574 (2010).
63. Duane, W. C. Effects of legume consumption on serum cholesterol, biliary lipids, and sterol metabolism in humans. *J. Lipid Res.* **38**, 1120–1128 (1997).
64. Hosseinpour-Niazi, S., Mirmiran, P., Fallah-Ghohroudi, A. & Azizi, F. Non-soya legume-based therapeutic lifestyle change diet reduces inflammatory status in diabetic patients: a randomised cross-over clinical trial. *Br. J. Nutr.* **114**, 213–219 (2015).
65. Rochfort, S. & Panozzo, J. Phytochemicals for Health, the Role of Pulses. *J. Agric. Food Chem.* **55**, 7981–7994 (2007).
66. Hosseinpour-Niazi, S. et al. Improvement of glycemic indices by a hypocaloric legume-based DASH diet in adults with type 2 diabetes: a randomized controlled trial. *Eur. J. Nutr.* **61**, 3037–3049 (2022).
67. Hosseinpour-Niazi S. et al. Improvement of glycemic indices by a hypocaloric legume-based DASH diet in adults with type 2 diabetes: a randomized controlled trial.
68. Meslier, V. et al. Mediterranean diet intervention in overweight and obese subjects lowers plasma cholesterol and causes changes in the gut microbiome and metabolome independently of energy intake. *Gut* **69**, 1258–1268 (2020).
69. Rinott, E. et al. The effects of the Green-Mediterranean diet on cardiometabolic health are linked to gut microbiome modifications: a randomized controlled trial. *Genome Med* **14**, 29 (2022).
70. Sowah, S. A. et al. Calorie restriction improves metabolic state independently of gut microbiome composition: a randomized dietary intervention trial. *Genome Med* **14**, 30 (2022).
71. Deehan, E. C. et al. Elucidating the role of the gut microbiota in the physiological effects of dietary fiber. *Microbiome* **10**, 77 (2022).
72. Wang, D. D. et al. The gut microbiome modulates the protective association between a Mediterranean diet and cardiometabolic disease risk. *Nat. Med* **27**, 333–343 (2021).
73. Ben-Yacov, O. et al. Gut microbiome modulates the effects of a personalised postprandial-targeting (PPT) diet on cardiometabolic markers: a diet intervention in pre-diabetes. *Gut* **72**, 1486–1496 (2023).
74. Shoer, S. et al. Impact of dietary interventions on pre-diabetic oral and gut microbiome, metabolites and cytokines. *Nat. Commun.* **14**, 5384 (2023).
75. Tanes, C. et al. Role of dietary fiber in the recovery of the human gut microbiome and its metabolome. *Cell Host Microbe* **29**, 394–407.e395 (2021).
76. Hu, J. et al. Distinct signatures of gut microbiota and metabolites in different types of diabetes: a population-based cross-sectional study. *EClinicalMedicine* **62**, 102132 (2023).
77. Ren, Y. et al. Alterations in the Gut Microbiota in Pregnant Women with Pregestational Type 2 Diabetes Mellitus. *mSystems* **8**, e0114622 (2023).

78. Natividad J. M. et al. Bilophila wadsworthia aggravates high fat diet induced metabolic dysfunctions in mice. *Nat. Commun.* **9**, (2018).
79. Fu, T. et al. Paired microbiome and metabolome analyses associate bile acid changes with colorectal cancer progression. *Cell Rep.* **42**, 112997 (2023).
80. Raygoza Garay, J. A. et al. Gut Microbiome Composition Is Associated With Future Onset of Crohn's Disease in Healthy First-Degree Relatives. *Gastroenterology* **165**, 670–681 (2023).
81. Zhai, L. et al. Gut microbiota-derived tryptamine and phenethylamine impair insulin sensitivity in metabolic syndrome and irritable bowel syndrome. *Nat. Commun.* **14**, 4986 (2023).
82. Mookerjee, A. & Tanaka, T. Influence of enzymatic treatments on legume proteins for improved functional and nutritional properties: expansion of legume protein utilization as food ingredients. *Curr. Opin. Food Sci.* **49**, 100974 (2023).
83. Yuan, X. et al. Depression and anxiety in patients with active ulcerative colitis: crosstalk of gut microbiota, metabolomics and proteomics. *Gut Microbes* **13**, 1987779 (2021).
84. Newgard, C. B. et al. A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab.* **9**, 311–326 (2009).
85. Yu D. et al. The adverse metabolic effects of branched-chain amino acids are mediated by isoleucine and valine. *Cell Metab.* **33**, (2021).
86. Qiao, S. et al. Gut Parabacteroides merdae protects against cardiovascular damage by enhancing branched-chain amino acid catabolism. *Nat. Metab.* **4**, 1271–1286 (2022).
87. Koziet K., & Urbanska E. M. Kynurenine Pathway in Diabetes Mellitus—Novel Pharmacological Target? *Cells* **12**, (2023).
88. De Mello, V. D. et al. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Sci. Rep.* **7**, 46337 (2017).
89. Venkatesh, M. et al. Symbiotic Bacterial Metabolites Regulate Gastrointestinal Barrier Function via the Xenobiotic Sensor PXR and Toll-like Receptor 4. *Immunity* **41**, 296–310 (2014).
90. Koh, A. et al. Microbially Produced Imidazole Propionate Impairs Insulin Signaling through mTORC1. *Cell* **175**, 947–961.e917 (2018).
91. Koh, A. et al. Microbial Imidazole Propionate Affects Responses to Metformin through p38γ-Dependent Inhibitory AMPK Phosphorylation. *Cell Metab.* **32**, 643–653.e644 (2020).
92. Xiao, C. et al. Associations of dietary diversity with the gut microbiome, fecal metabolites, and host metabolism: results from 2 prospective Chinese cohorts. *Am. J. Clin. Nutr.* **116**, 1049–1058 (2022).
93. Gradisteanu Pircalabioru G. et al. Snapshot into the Type-2-Diabetes-Associated Microbiome of a Romanian Cohort. *Int. J. Mol. Sci.* **23**, (2022).
94. Park Y. E. et al. Effects of Lactobacillus plantarum Q180 on Postprandial Lipid Levels and Intestinal Environment: A Double-Blind, Randomized, Placebo-Controlled, Parallel Trial. *Nutrients* **12**, (2020).
95. Zheng, Y., Ley, S. H. & Hu, F. B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat. Rev. Endocrinol.* **14**, 88–98 (2018).
96. Elsayed, N. A. et al. 2. Diagnosis and Classification of Diabetes: Standards of Care in Diabetes—2024. *Diab. Care* **47**, S20–S42 (2024).
97. Lin, Y., Zhu, M. & Su, Z. The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemp. Clin. Trials* **45**, 21–25 (2015).
98. Haldar, S. et al. Two Blends of Refined Rice Bran, Flaxseed, and Sesame Seed Oils Affect the Blood Lipid Profile of Chinese Adults with Borderline Hypercholesterolemia to a Similar Extent as Refined Olive Oil. *J. Nutr.* **150**, 3141–3151 (2020).
99. Henry, C. J. Basal metabolic rate studies in humans: measurement and development of new equations. *Public Health Nutr.* **8**, 1133–1152 (2005).
100. Lim, S. L. et al. Lifestyle Intervention Enabled by Mobile Technology on Weight Loss in Patients With Nonalcoholic Fatty Liver Disease: Randomized Controlled Trial. *JMIR Mhealth Uhealth* **8**, e14802 (2020).
101. Hagströmer, M., Oja, P. & Sjöström, M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. *Public Health Nutr.* **9**, 755–762 (2006).
102. IPAQ Research Committee. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)—short and long forms. IPAQ <https://sites.google.com/view/ipaq/score>, accessed: June 2023
103. Minekus, M. et al. A standardised static in vitro digestion method suitable for food – an international consensus. *Food Funct.* **5**, 1113–1124 (2014).
104. Beghini F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, (2021).
105. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinforma. (Oxf., Engl.)* **35**, 421–432 (2019).
106. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
107. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2020).
108. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinforma. (Oxf., Engl.)* **31**, 1674–1676 (2015).
109. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**, 1107–1115 (1998).
110. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. (Oxf., Engl.)* **22**, 1658–1659 (2006).
111. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
112. Ogata, H. et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
113. Aramaki, T. et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinforma. (Oxf., Engl.)* **36**, 2251–2252 (2020).
114. Vital, M., Howe, A. C. & Tiedje, J. M. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. *MBio* **5**, e00889 (2014).
115. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
116. Zhang, H. et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).

Acknowledgements

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its IAF-ICP (Industrial Alignment Fund - Industry Collaboration Project, Award Number I1701E0011). Wilmar International Limited (WIL) was involved in producing the intervention meals for the study. The study design, analysis and interpretation were jointly performed by A*STAR and WIL. We thank Ms Susanna Lim Poh Suan and Ms Jun Wei Lim for their technical and/or administrative assistance. We also thank Prof Chua Nam Hai for his valuable advice given in the course of this clinical study. Finally, we would like to thank all volunteers for their participation in this research.

Author contributions

S.H., C.J.H., L.H.W. designed the study. X.W., K.J.L., A.S.T., Y.N.Y., F.O., S.P., S.H. analyzed the data. A.S.T., Y.N.Y., F.O., S.H., V.S.B.V., K.K., X.L. conducted experiments. C.C.G., C.P.L.Y., L.H.W., A.S.T., S.H. designed the meals. A.S.T., S.H., F.O., S.L.T., Y.N.Y., P.G., R.T., H.S.T. enrolled volunteers and conducted the clinical trial. M.K.S.L. was the consultant clinician for the clinical trial. S.P. generated the random allocation sequence. A.S.T., S.H., X.W., K.J.L., H.F., V.S.B.V., X.L., Y.N.Y., F.O., S.L.T. wrote the manuscript. All authors significantly contributed to interpreting the results, critically revised the manuscript for important intellectual content and approved the final manuscript.

Competing interests

X.W., K.J.L., V.S.B.V., K.K., X.L., L.H.W., C.P.L.Y. and H.F. are employees of Wilmar International Limited. The remaining authors declare no competing interests.

Inclusion & ethics

This study aligns by the guidelines of Nature Portfolio journals, with contributions of each author as disclosed in Author contributions.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56084-6>.

Correspondence and requests for materials should be addressed to Sumanto Haldar or Kevin Junliang Lim.

Peer review information *Nature Communications* thanks Sandi Yen, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

¹WIL@NUS Corporate Laboratory, National University of Singapore, Centre for Translational Medicine, Singapore, Singapore. ²Clinical Nutrition Research Centre (CNRC), Singapore Institute of Food and Biotechnology Innovation (SIFBI), Agency for Science Technology and Research (A*STAR), Singapore, Singapore. ³Institute for Human Development and Potential (IHDP), A*STAR, Singapore, Singapore. ⁴Division of Medicine, Department of Endocrinology, Tan Tock Seng Hospital, Singapore, Singapore. ⁵Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. ⁶Duke-NUS Medical School, Singapore, Singapore. ⁷Department of Obstetrics and Gynaecology and Human Potential Translational Research programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁸Faculty of Health and Social Sciences, Bournemouth University, Bournemouth Gateway Building, St. Paul's Lane, Bournemouth, United Kingdom. ⁹These authors contributed equally: Xiaorong Wu, Alvin Surya Tjahyo, Sumanto Haldar, Kevin Junliang Lim. ✉ e-mail: shaldar@bournemouth.ac.uk; kevin.limjunliang@sg.wilmar-intl.com