The plasma proteome of plant-based diets: analyses of 1463 proteins in 50,000 people.

Tammy Y.N. Tong[1], Karl Smith-Byrne[1], Keren Papier[1], Joshua R. Atkins[1], Mahboubeh Parsaeian[1], Timothy J. Key[1] and Ruth C. Travis[1]

1. Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

Corresponding author: Tammy Y.N. Tong

Postal address: Cancer Epidemiology Unit, University of Oxford, Richard Doll Building, Old Road Campus, Oxford, OX3 7LF

Tel: +44 (0) 1865 289653

Email: tammy.tong@ndph.ox.ac.uk

Abstract

Background and objectives: Circulating proteins are integral to many biological processes. We aimed to assess differences in the plasma proteome between people of different dietary groups defined by degree of animal food consumption.

Methods: The UK Biobank recruited middle-aged adults (mostly 40 to 69 years) throughout the UK between 2006-2010. Relative concentrations of 1463 plasma proteins were quantified using the Olink Proximity Extension Assay on blood samples from 49,326 participants, who were also asked to report their ethnicity and consumption of red and processed meat, poultry, fish, dairy and eggs. We defined six diet groups among the white British participants (23,116 regular meat eaters, 23,323 low meat eaters, 484 poultry eaters, 1074 fish eaters, 722 vegetarians, and 54 vegans), and two diet groups among the British Indians (390 meat eaters and 163 vegetarians). We used multivariable-adjusted linear regressions to assess differences in protein concentrations between diet groups, with correction for multiple testing.

Results: We observed significant differences in many plasma proteins by diet group (683 proteins in white British participants, 1 in British Indians), in particular many proteins that are majority expressed in the digestive system. Of the biggest differences, compared with regular meat eaters, the non-meat eaters had significantly higher FGF21 (e.g. +0.40 SD in vegetarians), GUCA2A (+0.33), FOLR1 (+0.32), IGFBP2 (+0.31) and DSG2 (+0.30); all groups except the vegans had lower HAVCR1 (-0.38 in vegetarians). The observed differences were generally similar in direction in both ethnicities.

Discussion: In this first comprehensive assessment of plasma proteins by diet group, we identified many differences in proteins between groups defined by animal food consumption; this variation in protein levels suggests differences in various biological activities, including

gastrointestinal tract and kidney function, which may relate to differences in future disease risk.

Keywords: vegetarians, vegans, proteomics

Introduction

Proteins are essential for many bodily functions including supporting cell and tissue growth and structural integrity, and for enabling a range of enzymatic, biochemical, signalling and transport functions across different systems in the human body. Proteomics describes the large-scale study of multiple proteins and their interconnected pathways, the investigation of which allows a more comprehensive understanding of disease mechanisms [1]. Protein expression can be altered by environmental factors [2], including diet, which is an important exogenous source of dietary protein, and which provides the nine essential amino acids that cannot be synthesised endogenously but are nonetheless necessary for protein synthesis in the human body. Previous studies have shown that differences in dietary habits can substantially impact dietary intakes of protein and both dietary intakes and circulating concentrations of amino acids [3]. However, there is scant evidence on how diet may affect the proteome, partly due to the limited availability of proteomics data in large scale studies, and no prior studies have investigated how vegetarian and vegan diets may influence the proteome. Nonetheless, protein levels have an established role in the aetiology of multiple diseases [4], and the examination of dietary influences on the proteome may offer unique insights into understanding how vegetarian and vegan diets may affect future disease risk.

The aim of this study is to provide a detailed description of circulating protein concentrations in people with varying degrees of animal food consumption, using data from the UK Biobank.

Methods

Study Population

The UK Biobank is a prospective cohort study of around 500,000 middle-aged people (recruitment target 40-69 years), recruited from across the United Kingdom between 2006 and 2010. The scientific rationale and design of the UK Biobank have been described in detail previously [5]. In brief, participants were identified from National Health Service registers, and were invited to join the study if they lived within travelling distance (~25 km) of one of the 22 assessment centres in England, Wales and Scotland. People who consented to participate in the study attended a baseline visit at the assessment centre where they completed a touchscreen questionnaire which asked about their lifestyle (including diet, alcohol consumption, smoking status, physical activity), socio-demographic characteristics and general health and medical history. All participants were also given a verbal interview, and had their physical measurements and blood samples taken by trained staff. Permission for access to patient records for recruitment was approved by the Patient Information Advisory Group (now the National Information Governance Board for Health and Social Care) in England and Wales and the Community Health Index Advisory Group in Scotland, and all participants gave informed consent to participate using a signature capture device at the baseline visit.

Ethnicity classification

On the touchscreen questionnaire, participants were asked to self-identify their ethnicity as ‘White’, ‘Mixed’, ‘Asian or Asian British’, ‘Black or Black British’, ‘Chinese’, ‘Other ethnic group’, ‘Do not know’, or ‘Prefer not to answer’, with further sub-categories under each option. Participants were included for our analyses if they self-identified as ‘White’, or as ‘Asian or Asian British’ and subsequently as ‘Indian’, hereafter referred to as ‘white British’ and ‘British Indian’. The white British population was included as it made up the majority of

the participants in UK Biobank, while the British Indian population was included because of the large proportion of vegetarians in this population group (25% compared to less than 2% in the overall population). The number of vegetarians in the other ethnic groups was too small to allow valid comparisons by diet groups, and therefore people of other ethnic groups were excluded from these analyses.

Diet group classification

Participants were classified into diet groups based on self-reported dietary data from the touchscreen questionnaire completed at recruitment. Participants were asked to report their frequency of consumption of processed meat (including processed poultry), unprocessed red meat (beef, lamb or mutton, pork), unprocessed poultry (such as chicken or turkey), oily fish, and other types of fish, with the possible responses ranging from “never” to “once or more daily”. Participants were also asked whether they never consumed dairy and eggs or foods containing eggs. Based on their responses to these questions, the white British participants were classified into six diet groups: regular meat eaters (red and processed meat consumption >3 times per week), low meat eaters (red and processed meat consumption ≤3 times per week), poultry eaters (participants who ate unprocessed poultry but no red and processed meat), fish eaters (participants who ate fish, but not red and processed meat, or poultry), vegetarians (participants who did not eat any meat or fish), and vegans (participants who did not eat any meat, fish, dairy products or eggs). The British Indian participants were classified into two diet groups: meat eaters (ate any combination of red or processed meat or poultry) and vegetarians; the numbers of fish eaters and vegans were low in the British Indians, and thus were not included in these analyses.

Assay for plasma proteins

Non-fasting blood samples were collected from all UK Biobank participants by trained personnel (either a phlebotomist or a nurse) except in a small proportion (0.3%) of participants who declined, were deemed unable to, or where the attempt was abandoned for technical or health reasons. Proteomic profiling was conducted using the Olink Proximity Extension Assay on blood plasma samples in 54,306 participants selected as part of the UK Biobank Pharma Proteomics Project (UKB-PPP). The samples were shipped on dry ice to Olink Analyses Service, Uppsala, Sweden for analysis; details of the selection procedures for inclusion in UKB-PPP and the technical details of the proteomics assays have been described elsewhere [6]. In brief, the Olink Explore 1536 platform used in this study is an antibody-based assay which measures the relative abundance of 1,472 protein analytes, including 1,463 unique proteins, distributed across four 384-plex panels: inflammation, oncology, cardiometabolic, and neurology. The Olink Explore platform is based on proximity extension assays that are highly sensitive and reproducible with low cross-reactivity. Relative concentrations of the 1,463 unique proteins were readout by next-generation sequencing. Measurements are expressed as normalized protein expression (NPX) values which are log-base-2 transformed. Protein values below the limit of detection (LOD) were replaced with the LOD divided by the square root of 2 [7]. Protein concentrations were subsequently inverse rank normal transformed. All results for differences in protein concentrations by diet groups may be interpreted as SD differences.

Inclusion and exclusion criteria

Of the 54,306 participants selected for proteomic profiling, 52,705 participants remained after the quality control procedures. Participants were further excluded if they were not of white

British or British Indian ethnicity (n=2964), could not be classified into one of the pre-specified diet groups (n=414), or had missing information for fasting time (n=1). After the exclusions, 49,326 participants (48,773 white British and 553 British Indian) were included in the analyses. A participant flow chart of the inclusion criteria is shown in **Supplementary figure 1**.

Statistical analyses

Baseline characteristics of UK Biobank participants included in these analyses were tabulated by six diet groups in the white British population and by two diet groups in the British Indian population, as mean (SD) for continuous variables and number (%) for categorical variables. We used multivariable-adjusted linear regressions to estimate differences in protein concentrations by diet group, separately by ethnicity, using regular meat eaters as a reference group in the white British participants, and meat eaters as a reference group in the British Indians. The model was adjusted for age at recruitment (5 year categories), sex, region (London, North-West England, North-East England, Yorkshire, West Midlands, East Midlands, South-East England, South-West England, Wales, Scotland), fasting status (0–1, 2, 3, 4, 5, 6–7, ≥8 hours), body mass index (BMI; <20, 20.0–22.4, 22.5–24.9, 25.0–27.4, 27.5–29.9, 30.0–32.4, 32.5–34.9, ≥35 kg/m^2, unknown), alcohol consumption (<1, 1–7, 8–15, ≥16 g/d, unknown), smoking status (never, previous, current <15 cigarettes/day, current ≥15 cigarettes/d, unknown) and physical activity (<10, 10-49.9, ≥50 excess metabolic equivalent of task, hr/wk, unknown). Wald tests were used to assess overall heterogeneity between diet group in each ethnicity, and for pairwise comparisons of each of the other diet groups compared to the reference group (regular meat eaters in white British participants/meat eaters in British Indians). Heterogeneity between vegetarians and vegans in the white British

population was assessed based on post-estimation linear combinations of parameters for all proteins.

To account for multiple testing while considering the high correlations between the circulating proteins, we conducted a principal component analysis of the circulating proteins in the complete proteins dataset before exclusions for these analyses, and determined that the first 639 principal components explained 95% of the total variation in the exposure data. Consequently, the effective number of independent tests [8] was determined to be 639, and the statistical significance level was set to be p-value 0.05/639 = 0.000078. For the top proteins identified (the top 10 significant proteins in each pairwise comparison against the reference group, based on ranking of p-values), we conducted additional sensitivity analyses to evaluate the extent to which the associations may be influenced by key covariates including BMI, smoking and alcohol consumption, by presenting models with and without adjustment of these variables. Additionally, we also restricted the analyses to people who self-reported to be in good and excellent health. All analyses were performed using Stata version 17.0 (StataCorp, TX, USA). All figures were generated using R version 4.2.1, the forest plots using "Jasper makes plots" package version 2-266 [9].

Integrating publicly available information on gene expression

To further understand the biological context for the proteins of interest, we extracted single cell mRNA expression from the Human Protein Atlas [10] to describe expression for any proteins that were significantly different in one or more comparisons. Normalized expression levels were extracted for genes in 30 different human tissues and 82 cell-types. Gene expression specificity at the cell or tissue type level was calculated as the ratio of each gene cell-type or tissue expression to the total expression of each gene across all cell or tissue

types. We subsequently identified genes for proteins that were majority expressed (more than 50% of total expression) in at least one cell or tissue type, and mapped these to their likely candidate cell and tissue of origin where possible.

Results

Baseline characteristics

The baseline characteristics of the study population are shown by ethnicity and diet group in Table 1. Compared with white British regular meat eaters, the non-meat eaters were on average slightly younger, more likely to be women, and to reside in London. They were less likely to be overweight, reported lower alcohol consumption, lower current smoking, and more physical activity. Compared with British Indian meat eaters, the vegetarians were more likely to be women and reported lower alcohol consumption, lower current smoking and slightly lower physical activity, but BMI was not noticeably different between the two groups. Fasting time was not meaningfully different by diet group in both ethnicities.

Differences in plasma proteins by diet group

The plasma proteins that were significantly different in white British vegetarians and vegans compared with regular meat eaters are shown as volcano plots in **Figure 1**. Significant differences in proteins in white British low meat eaters, poultry eaters and fish eaters compared with regular meat eaters, and in British Indian vegetarians compared with meat eaters are shown as volcano plots in **Supplementary figures 2-6**. Overall, 683 plasma proteins were significantly different by diet group (based on p-heterogeneity < 0.000078) in the white British participants, including 535 plasma proteins that were significantly different

in one or more pairwise comparisons between regular meat eaters and at least one of the other diet groups. This includes 296 proteins (23 higher, 273 lower) in low meat eaters compared with regular meat eaters, 59 (9 higher, 50 lower) in poultry eaters, 157 (104 higher, 53 lower) in fish eaters, 219 (194 higher, 25 lower) in vegetarians and 15 (12 higher, 3 lower) in vegans; and 2 proteins were significantly different between vegetarians and vegans (AHSP and FOLR2 were both higher in vegans). The proteins identified are involved in a range of different biological pathways. Of the 535 plasma proteins that were significantly different in one or more pairwise comparisons, 97 are majority expressed (protein mRNA >50% of total body mRNA expression) in one tissue type, while the other proteins have limited tissue specificity. In particular, these included 39 proteins that are majority expressed in the digestive system (including 16 in the liver, 8 in the pancreas, 8 in the small intestine), 30 in the nervous system (27 in the brain), and 12 in the respiratory system (7 in the lung) (**Supplementary figure 7**).

In the British Indian population, only 1 protein (SUMF2, which is higher in vegetarians) was significantly different when comparing vegetarians and meat eaters, likely owing to the much smaller number of British Indians in the study. However, the differences in many proteins were directly consistent between the two ethnicities, as illustrated by the forest plots of top proteins by diet group and ethnicity, with the six diet groups in the white British participants and the two diet groups in British Indians (**Figures 2-3, Supplementary figures 8-10**). The complete results of all proteins are shown in **Supplementary tables 1.**

Generally, in the multivariable-adjusted model, many proteins showed a gradient effect in magnitude of differences across diet groups by degree of animal food exclusion, from regular meat eaters to vegans. Figure 2 shows the top 10 proteins in white British vegetarians (based on the ranking of pairwise p-values in vegetarians compared with regular meat eaters), and Figure 3 shows the top 10 proteins in white British vegans. Of the top proteins in vegetarians

(Figure 2), compared with regular meat eaters, vegetarians and vegans both had significantly higher FGF21 (+0.40 SD in vegetarians; +0.74 SD in vegans), GUCA2A (+0.33; +0.79), FOLR1 (+0.32; +0.61) and IGFBP2 (+0.31; +0.55) after correction for multiple testing, whereas vegetarians but not vegans had significantly higher ART3 (+0.34), LAYN (+0.32), PCSK9 (+0.32), and lower CDHR2 (-0.26), HAVCR1 (-0.38) and CNDP1 (-0.39). In addition to the aforementioned proteins, vegans also had substantially higher FOLR2 (+0.69), GAS6 (+0.68), DSG2 (+0.68), THY1 (+0.62), MANSC1 (+0.58), but lower EGFR (-0.55) than regular meat eaters (Figure 3); the differences in these proteins were directionally consistent but less extreme in the vegetarians.

The top 10 proteins in white British low meat eaters, poultry eaters and fish eaters compared with regular meat eaters are shown in Supplementary figures 8, 9 and 10 respectively. Of these top proteins, none was uniquely different in one diet group only when compared with regular meat eaters. Similar to the observations in vegetarians, all three groups had substantially higher DSG2 and lower HAVCR1 compared with regular meat eaters; in both low meat eaters and poultry eaters HAVCR1 is the protein that exhibited the largest magnitude in difference across all proteins, while in fish eaters FGF23 showed the biggest difference. In addition, low meat eaters and poultry eaters both had lower OSM and CD99L2, the latter of which was higher in vegetarians. Low meat eaters, fish eaters and vegetarians all had lower ACP5 than regular meat eaters; this protein was lower but not significantly different in poultry eaters. In sensitivity analyses of the top proteins with different levels of covariate adjustment and limited to people of good and excellent health, results were similar (**Supplementary tables 2**).

Discussion

In this large British cohort, we observed differences in the levels of many proteins by diet group. This is the first comprehensive study on plant-based diets and plasma proteomics, and many of the substantial differences reported here are novel. In particular, we saw substantial differences in some proteins potentially associated with disease states such as HAVCR1 and DSG2 (associated with kidney disease [11] and inflammatory bowel disease, IBD [12], respectively), as well as proteins associated with nutritional status including IGFBP2 [13, 14], and possibly FGF21 [15] and the folate receptors FOLR1 and FOLR2 [16]. The magnitudes of differences in many proteins showed a gradient effect across the diet groups, and were directionally consistent between the white British and British Indian populations.

In terms of tissue mRNA expression, most of the proteins identified are not majority expressed in one specific tissue type, which suggests that they may be involved in a range of biochemical pathways and processes. Of the 97 proteins that were significantly different by diet group and also majority expressed in one tissue type, many were identified to be those expressed in the digestive system, which may signify dietary impact on digestive functions. We also see a number of proteins with high expression in the brain and lung, which may benefit from further research. Overall, the substantial differences in the protein concentrations observed likely reflect physiological differences between the different diet groups, as well as differences in underlying disease pathology and future disease risk.

No previous studies were found that examined plasma proteomics in vegetarians and vegans. A few previous studies have reported on proteomic profiles by other dietary patterns, including the Dietary Approaches to Stop Hypertension (DASH) diet, the Healthy Eating Index and the Mediterranean diet, as well as other data derived dietary patterns [17–20]. Whilst vegetarian and vegan diets share some common features with these other healthy dietary patterns, such as higher intakes of fruit, vegetables and whole grains, the diets are also inherently different in many other aspects. For example, while both the DASH diet and

Mediterranean diet recommend reducing red and processed meat consumption, they generally do not pose any restrictions on consumption of fish and fish products, or dairy products (particularly low-fat versions). In contrast, vegetarian and vegan diets are defined by the complete exclusion of animal food groups. Therefore, it is not surprising that while studies on the DASH diet also reported higher FOLR2 [17] and DSG2 [18] in people with higher adherence, we generally do not find similarities in the results.

FGF21

There are only a limited number of studies examining vegetarian and vegan diets with individual proteins. In support of our findings, higher concentrations of FGF21 in vegetarians and vegans have been shown in a small study of 36 each of omnivores, vegetarians and vegans [21]. FGF21 has been hypothesised to be high in vegans as a downstream effect of low intake of methionine [15], an essential amino acid that is known to be limiting in vegan diets [3]. Studies of FGF21 administration in rodents have found favourable effects on adiposity, lipid profiles, and non-alcohol fatty liver disease, but possible adverse effects on bone homeostasis [22]; randomised trials of FGF21 analogues also support the reduction in triglycerides [23] and improvement of fibrosis in patients with non-alcoholic steatohepatitis [24]. Consistent with these observations, we have previously reported lower BMI and body fat [25], more favourable lipid profiles [26] and lower heart disease risk [27], but also lower bone mineral density [25] and higher fracture risk [28] in vegetarians and vegans when compared with meat eaters; another study has reported lower odds of non-alcohol fatty liver disease in vegetarians than non-vegetarians [29].

GUCA2A

GUCA2A was substantially higher in all diet groups compared to regular meat eaters, with the exception of low meat eaters. GUCA2A activates the Guanylate Cyclase C receptor, which has a role in the maintenance of gut physiology including the increase of water movement into the intestinal lumen [30]. Previous studies have found that fasting plasma levels of GUCA2A are significantly lower in patients who have been diagnosed with Crohn's disease compared with healthy controls [31], and that loss of GUCA2A expression may be an important determinant of colorectal cancer [32]. Consistent with the observations in this study, a previous EPIC-Oxford study reported a higher frequency of bowel movements in vegetarians and vegans [33].

FOLR1 and FOLR2

We found both FOLR1 and FOLR2 to be higher in non-meat eaters, with vegans having particularly high levels. As folate receptors, FOLR1 and FOLR2 both bind to and import folic acid into cells, and are usually downregulated with repletion of folate, however, this process may also be mediated by homocysteine in a positive direction [16]. While vegetarians and vegans typically have high folate, they may also have high homocysteine due to low vitamin B12 intakes [34, 35], though no other studies have reported on folate receptor expression in vegetarians and vegans for comparison.

IGFBP2

Similar to our findings, a previous study in EPIC-Oxford has found that compared with meat-eaters, both vegetarians and vegans had higher concentrations of IGFBP2, a binding protein for IGF-1 [13]. The mechanism for higher IGFBP2 in vegetarians and vegans is not well established but has been suggested to be related to lower dietary intakes of essential amino

acids [13], protein (especially from dairy) [14], or lower total energy intake [36], which are all characteristic of vegetarian and vegan diets [3, 25, 37].

DSG2

Low and non-meat eaters, particularly vegans, had higher DSG2 in the current study. DSG2 belongs to the family of desmosomal cadherins, which includes both desmogleins (DSGs) and desmocollins (DSCs), with DSG2 and DSC2 being the major isoforms in humans, and the only isoforms expressed in the intestinal epithelium [12]. Consistently, the current study also showed that DSC2 was higher in non-meat eaters particularly vegans. Though the two proteins are structurally and functionally similar, laboratory evidence in mouse models suggest that DSG2 may have a more prominent role in the maintenance of the integrity of the intestinal epithelial barrier than DSC2 [38]. In humans, several small studies on IBD patients have found that lower intestinal protein levels of DSG2 in IBD patients, suggesting a role of the protein in IBD pathogenesis [12].

HAVCR1

HAVCR1 has been recognised as a biomarker of kidney injury and possibly an early diagnostic marker of kidney disease [11]. In mice, high-fat diets or the saturated fatty acid palmitate have been shown to upregulate expression of HAVCR1 in the proximal tubules of the kidney [39–41]. In humans, vegetarians and vegans tend to have lower intake of saturated fat, which is consistent with the observation of lower HAVCR1 in the non-meat eaters (though not significantly so in vegans) in the current study. Additionally, high protein diets, especially diets high in protein from animal sources, have been suggested to have adverse effects on kidney health [42]. A previous study suggested a lower risk of chronic kidney

disease in vegetarians after accounting for baseline risk of diabetes and hypertension [43]. Furthermore, as HAVCR1 is expressed on the surface of immune cells, it is also believed to have a role regulating immune responses via activation and proliferation of immune cells [11].

PCSK9

The observation of higher PCSK9 in vegetarians is of interest. PCSK9 is a well-established drug target for treating hypercholesterolaemia, whereby PCSK9 inhibitors lowers low-density-lipoprotein cholesterol (LDL-C) concentrations by preventing the binding of PCSK9 to LDL receptors, which subsequently enables the LDL receptors to bind and remove LDL-C from the bloodstream instead [44, 45]. As a result, higher PCSK9 is often accompanied by high LDL-C, whereas the vegetarians in the current [26] and other studies [46] typically have significantly lower LDL-C, likely partly due to vegetarians having lower saturated fat intake [47]. It may therefore be plausible that the relatively higher PCSK9 expression in people with low LDL-C may be related to maintaining the minimum level of LDL-C necessary to support cellular function and structural integrity, but further studies are needed to confirm this.

Other proteins

The functions of many other proteins identified are not well established. CNDP1 is a dipeptidase, and one previous study reported higher CNDP1 in people on high protein diets [48], which would be consistent with our observation of lower CNDP1 in non-meat eaters (not statistically significant in vegans) who generally have lower protein intake. Several of the proteins identified such as ART3, GAS6, THY1 and EGFR are involved in a range of signalling pathways and progression of tumour tissues [49–52], while other proteins

including LAYN and CDHR2 are believed to have roles in cell adhesion [53, 54]. AHSP, which was significantly higher in vegans than vegetarians, has a role in stabilising haemoglobin and its loss of function has been linked to inherited anaemia [55], but its relevance for nutritional anaemia is unclear. Overall, the reasons that these proteins appear to be influenced by diet and their implications require further research. As the first study on diet group and a large panel of circulating proteins, the current study focused on the biggest differences by diet groups, and many other differences were observed that could not be described in detail, but still deserve further investigation. The study demonstrates the utility of investigating protein differences by diet groups; future research should measure these and other proteins using multiple technologies in relation to dietary factors, to replicate and expand on our findings.

Strengths and limitations

The key strength of this study was that it showed for first time differences in circulating proteins between people of different habitual diet groups in a large population, and many of the findings were completely novel. We have defined six diet groups in the white British population and two diet groups in the British Indian population, which allowed a detailed comparison by both dietary habits and ethnicity. Of the limitations, as with all observational studies, some level of self-selection bias may be present, which limits the generalisability of the findings. While we have adjusted for several important confounders, residual confounding by other dietary and non-dietary factors may still be present, though sensitivity analyses showed that results were consistent across all adjustment models. We are also unable to infer causality due to the cross-sectional nature of the studies.

Conclusions

In this large population-based cohort in the United Kingdom, many differences in circulating protein concentrations were observed between different diet groups. These proteins are involved in a range of different biological pathways and processes, particularly related to gastrointestinal tract and kidney function. These differences likely reflect physiological differences between the different diet groups, and the implications of these differences for future disease risk require further investigation.

References

1. Cui M, Cheng C, Zhang L (2022) High-throughput proteomics: a methodological mini-review. Lab Invest 102:1170–1181. https://doi.org/10.1038/s41374-022-00830-7

2. Al-Amrani S, Al-Jabri Z, Al-Zaabi A, et al (2021) Proteomics: Concepts and applications in human medicine. World J Biol Chem 12:57–69. https://doi.org/10.4331/wjbc.v12.i5.57

3. Schmidt JA, Rinaldi S, Scalbert A, et al (2016) Plasma concentrations and intakes of amino acids in male meat-eaters, fish-eaters, vegetarians and vegans: a cross-sectional analysis in the EPIC-Oxford cohort. Eur J Clin Nutr 70:306–312. https://doi.org/10.1038/ejcn.2015.144

4. Pietzner M, Wheeler E, Carrasco-Zanini J, et al (2021) Mapping the proteo-genomic convergence of human diseases. Science (80-) 374:. https://doi.org/10.1126/science.abj1541

5. Fry A, Littlejohns TJ, Sudlow C, et al (2017) Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol 186:1026–1034. https://doi.org/10.1093/aje/kwx246

6. Sun B, Chiou J, Traylor M, et al (2022) Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. bioRxiv. https://doi.org/10.1101/2022.06.17.496443

7. Albanes D, Alcala K, Alcala N, et al (2023) The blood proteome of imminent lung cancer diagnosis. Nat Commun 14:3042. https://doi.org/10.1038/s41467-023-37979-8

8. Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genet Epidemiol 32:361–369. https://doi.org/10.1002/gepi.20310

9. Arnold M (2022) Jasper: Jasper makes plots. 2020

10. Thul PJ, Lindskog C (2018) The human protein atlas: A spatial map of the human proteome. Protein Sci 27:233–244. https://doi.org/10.1002/pro.3307

11. Liu S, Tang W, Cao J, et al (2022) A Comprehensive Analysis of HAVCR1 as a Prognostic and Diagnostic Marker for Pan-Cancer. Front Genet 13:1–11. https://doi.org/10.3389/fgene.2022.904114

12. Schlegel N, Boerner K, Waschke J (2021) Targeting desmosomal adhesion and signalling for intestinal barrier stabilization in inflammatory bowel diseases—Lessons from experimental models and patients. Acta Physiol 231:1–15. https://doi.org/10.1111/apha.13492

13. Allen NE, Appleby PN, Davey GK, et al (2002) The associations of diet with serum insulin-like growth factor I and its main binding proteins in 292 women meat-eaters, vegetarians, and vegans. Cancer Epidemiol Biomarkers Prev 11:1441–1448

14. Crowe FL, Key TJ, Allen NE, et al (2009) The association between diet and serum concentrations of IGF-I, IGFBP-1, IGFBP-2, and IGFBP-3 in the European

Prospective Investigation into Cancer and Nutrition. Cancer Epidemiol Biomarkers Prev 18:1333–1340. https://doi.org/10.1158/1055-9965.EPI-08-0781

15. McCarty MF (2017) The moderate essential amino acid restriction entailed by low-protein vegan diets may promote vascular health by stimulating FGF21 secretion. Horm Mol Biol Clin Investig 30:1–9. https://doi.org/10.1515/hmbci-2015-0056

16. Antony AC, Tang YS, Khan RA, et al (2004) Translational upregulation of folate receptors is mediated by homocysteine via RNA-heterogeneous nuclear ribonucleoprotein E1 interactions. J Clin Invest 113:285–301. https://doi.org/10.1172/JCI200411548

17. Kim H, Lichtenstein AH, Ganz P, et al (2023) Identification of Protein Biomarkers of the Dietary Approaches to Stop Hypertension Diet in Randomized Feeding Studies and Validation in an Observational Study. J Am Heart Assoc 12:e028821. https://doi.org/10.1161/JAHA.122.028821

18. Du S, Chen J, Kim H, et al (2023) Plasma Protein Biomarkers of Healthy Dietary Patterns: Results from the Atherosclerosis Risk in Communities Study and the Framingham Heart Study. J Nutr 153:34–46. https://doi.org/10.1016/j.tjnut.2022.11.008

19. Walker ME, Song RJ, Xu X, et al (2020) Proteomic and metabolomic correlates of healthy dietary patterns: The framingham heart study. Nutrients 12:1–20. https://doi.org/10.3390/nu12051476

20. Warensjö Lemming E, Byberg L, Stattin K, et al (2019) Dietary Pattern Specific Protein Biomarkers for Cardiovascular Disease: A Cross-Sectional Study in 2 Independent Cohorts. J Am Heart Assoc 8:. https://doi.org/10.1161/JAHA.118.011860

21. Castaño-Martinez T, Schumacher F, Schumacher S, et al (2019) Methionine restriction prevents onset of type 2 diabetes in NZO mice. FASEB J 33:7092–7102. https://doi.org/10.1096/fj.201900150R

22. Geng L, Lam KSL, Xu A (2020) The therapeutic potential of FGF21 in metabolic diseases: from bench to clinic. Nat Rev Endocrinol 16:654–667. https://doi.org/10.1038/s41574-020-0386-0

23. Bhatt DL, Bays HE, Miller M, et al (2023) The FGF21 analog pegozafermin in severe hypertriglyceridemia: a randomized phase 2 trial. Nat Med 29:. https://doi.org/10.1038/s41591-023-02427-z

24. Loomba R, Sanyal AJ, Kowdley K V., et al (2023) Randomized, Controlled Trial of the FGF21 Analogue Pegozafermin in NASH. N Engl J Med 1–11. https://doi.org/10.1056/nejmoa2304286

25. Tong TY, Key TJ, Sobiecki JG, Bradbury KE (2018) Anthropometric and physiologic characteristics in white and British Indian vegetarians and nonvegetarians in the UK Biobank. Am J Clin Nutr 107:909–920. https://doi.org/10.1093/ajcn/nqy042

26. Tong TYN, Perez-Cornago A, Bradbury KE, Key TJ (2021) Biomarker Concentrations in White and British Indian Vegetarians and Nonvegetarians in the UK Biobank. J Nutr 151:3168–3179. https://doi.org/10.1093/jn/nxab192

27. Tong TYN, Appleby PN, Bradbury KE, et al (2019) Risks of ischaemic heart disease and stroke in meat eaters, fish eaters, and vegetarians over 18 years of follow-up: results from the prospective EPIC-Oxford study. BMJ 366:l4897. https://doi.org/10.1136/bmj.l4897

28. Tong TYN, Appleby PN, Armstrong MEG, et al (2020) Vegetarian and vegan diets

and risks of total and site-specific fractures: results from the prospective EPIC-Oxford study. BMC Med 18:353. https://doi.org/10.1186/s12916-020-01815-3

29. Chiu T, Lin M-N, Pan W-H, et al (2018) Vegetarian diet, food substitution, and nonalcoholic fatty liver. Tzu Chi Med J 30:102. https://doi.org/10.4103/tcmj.tcmj_109_17

30. Dye FS, Larraufie P, Kay R, et al (2019) Characterisation of proguanylin expressing cells in the intestine – evidence for constitutive luminal secretion. Sci Rep 9:1–20. https://doi.org/10.1038/s41598-019-52049-0

31. von Volkmann HL, Brønstad I, Tronstad RR, et al (2020) Plasma levels of guanylins are reduced in patients with Crohn's disease. Scand J Gastroenterol 55:449–453. https://doi.org/10.1080/00365521.2020.1748224

32. Wilson C, Lin JE, Li P, et al (2014) The paracrine hormone for the GUCY2C tumor suppressor, guanylin, is universally lost in colorectal cancer. Cancer Epidemiol Biomarkers Prev 23:2328–2337. https://doi.org/10.1158/1055-9965.EPI-14-0440

33. Sanjoaquin MA, Appleby PN, Spencer EA, Key TJ (2004) Nutrition and lifestyle in relation to bowel movement frequency: a cross-sectional study of 20 630 men and women in EPIC–Oxford. Public Health Nutr 7:77–83. https://doi.org/10.1079/phn2003522

34. Gilsing AMJ, Crowe FL, Lloyd-Wright Z, et al (2010) Serum concentrations of vitamin B12 and folate in British male omnivores, vegetarians and vegans: results from a cross-sectional analysis of the EPIC-Oxford cohort study. Eur J Clin Nutr 64:933–939. https://doi.org/10.1038/ejcn.2010.142

35. Hung CJ, Huang PC, Lu SC, et al (2002) Plasma homocysteine levels in Taiwanese

vegetarians are higher than those of omnivores. J Nutr 132:152–158. https://doi.org/10.1093/jn/132.2.152

36. Kaaks R, Bellati C, Venturelli E, et al (2003) Effects of dietary intervention on IGF-I and IGF-binding proteins, and related alterations in sex steroid metabolism: The Diet and Androgens (DIANA) Randomised Trial. Eur J Clin Nutr 57:1079–1088. https://doi.org/10.1038/sj.ejcn.1601647

37. Sobiecki JG, Appleby PN, Bradbury KE, Key TJ (2016) High compliance with dietary recommendations in a cohort of meat eaters, fish eaters, vegetarians, and vegans: Results from the European Prospective Investigation into Cancer and Nutrition-Oxford study. Nutr Res 36:464–477. https://doi.org/10.1016/j.nutres.2015.12.016

38. Gross A, Pack LAP, Schacht GM, et al (2018) Desmoglein 2, but not desmocollin 2, protects intestinal epithelia from injury. Mucosal Immunol 11:1630–1639. https://doi.org/10.1038/s41385-018-0062-z

39. Yu Y, Mo H, Zhuo H, et al (2022) High Fat Diet Induces Kidney Injury via Stimulating Wnt/β-Catenin Signaling. Front Med 9:1–9. https://doi.org/10.3389/fmed.2022.851618

40. Zhao X, Chen X, Zhang Y, et al (2019) Kidney injury molecule-1 is upregulated in renal lipotoxicity and mediates palmitate-induced tubular cell injury and inflammatory response. Int J Mol Sci 20:. https://doi.org/10.3390/ijms20143406

41. Mori Y, Ajay AK, Chang JH, et al (2021) KIM-1 mediates fatty acid uptake by renal tubular cells to promote progressive diabetic kidney disease. Cell Metab 33:1042-1061.e7. https://doi.org/10.1016/j.cmet.2021.04.004

42. Ko GJ, Rhee CM, Kalantar-Zadeh K, Joshi S (2020) The effects of high-protein diets

on kidney health and longevity. J Am Soc Nephrol 31:1667–1679. https://doi.org/10.1681/ASN.2020010028

43. Liu H-W, Tsai W-H, Liu J-S, Kuo K-L (2019) Association of Vegetarian Diet with Chronic Kidney Disease. Nutrients 11:279. https://doi.org/10.3390/nu11020279

44. Schmidt AF, Holmes M V, Preiss D, et al (2019) Phenome-wide association analysis of LDL- cholesterol lowering genetic variants in. 1–10

45. Horton JD, Cohen JC, Hobbs HH (2007) Molecular biology of PCSK9: its role in LDL metabolism. Trends Biochem Sci 32:71–77. https://doi.org/10.1016/j.tibs.2006.12.008

46. Bradbury KE, Crowe FL, Appleby PN, et al (2014) Serum concentrations of cholesterol, apolipoprotein A-I and apolipoprotein B in a total of 1694 meat-eaters, fish-eaters, vegetarians and vegans. Eur J Clin Nutr 68:178–183. https://doi.org/10.1038/ejcn.2013.248

47. Mensink RP, Zock PL, Kester AD, Katan MB (2003) Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins□: a meta- - analysis of 60 controlled trials. Am J Clin Nutr 77:1146–1155. https://doi.org/11778166

48. Franzke B, Bileck A, Unterberger S, et al (2022) The plasma proteome is favorably modified by a high protein diet but not by additional resistance training in older adults: A 17-week randomized controlled trial. Front Nutr 9:. https://doi.org/10.3389/fnut.2022.925450

49. Tan L, Song X, Sun X, et al (2016) ART3 regulates triple-negative breast cancer cell function via activation of Akt and ERK pathways. Oncotarget 7:46589–46602. https://doi.org/10.18632/oncotarget.10306

50. Tanaka M, Siemann DW (2020) Gas6/Axl signaling pathway in the tumor immune microenvironment. Cancers (Basel) 12:1–14. https://doi.org/10.3390/cancers12071850

51. Hagood JS (2019) Thy-1 as an Integrator of Diverse Extracellular Signals. Front Cell Dev Biol 7:1–5. https://doi.org/10.3389/fcell.2019.00026

52. Wee P, Wang Z (2017) Epidermal growth factor receptor cell proliferation signaling pathways. Cancers (Basel) 9:1–45. https://doi.org/10.3390/cancers9050052

53. Pan J hua, Zhou H, Cooper L, et al (2019) LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers. Front Immunol 10:1–12. https://doi.org/10.3389/fimmu.2019.00006

54. Pinette JA, Mao S, Millis BA, et al (2019) Brush border protocadherin CDHR2 promotes the elongation and maximized packing of microvilli in vivo. Mol Biol Cell 30:108–118. https://doi.org/10.1091/mbc.E18-09-0558

55. Yu X, Kong Y, Dore LC, et al (2007) An erythroid chaperone that facilitates folding of α-globin subunits for hemoglobin synthesis. J Clin Invest 117:1856–1865. https://doi.org/10.1172/JCI31664

Funding: The work is supported by the UK Medical Research Council (MR/M012190/1), Wellcome Trust Our Planet Our Health (Livestock, Environment and People, LEAP 205212/Z/16/Z) and Cancer Research UK (C8221/A29017 and C8221/A29186). TYNT is support by a Nuffield Departmental of Population Health Fellowship. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Acknowledgements: This research has been conducted using UK Biobank Resource under application 24494.

Competing interests: The authors had no conflicts of interest.

Author's contributions: TYNT, TJK and RCT conceived and designed the research question. KSB conducted quality control of the proteomics data. TYNT analyzed the data and wrote the first draft of the manuscript. All authors provided input on data analysis and interpretation of results, revised the manuscript critically for important intellectual content, and read and approved the final manuscript.

Table 1: Baseline characteristics of white British and British Indian participants by diet groups in UK Biobank.

	White British participants						British Indian participants	
	Regular meat eaters N=23,116	Low meat eaters N=23,323	Poultry eaters N=484	Fish eaters N=1,074	Vegetarians N=722	Vegans N=54	Meat eaters N=390	Vegetarians N=163
Age at recruitment	57.2 (8.2)	57.2 (8.0)	56.6 (8.1)	54.4 (8.2)	52.9 (8.1)	55.0 (7.8)	53.9 (8.6)	54.7 (7.7)
Sex								
Women	9,810 (42.4%)	14,861 (63.7%)	389 (80.4%)	756 (70.4%)	483 (66.9%)	28 (51.9%)	170 (43.6%)	102 (62.6%)
Men	13,306 (57.6%)	8,462 (36.3%)	95 (19.6%)	318 (29.6%)	239 (33.1%)	26 (48.1%)	220 (56.4%)	61 (37.4%)
Region of assessment centre								
London	2,558 (11.1%)	2,797 (12.0%)	80 (16.5%)	191 (17.8%)	121 (16.8%)	8 (14.8%)	208 (53.3%)	88 (54.0%)
North-West England	3,957 (17.1%)	3,577 (15.3%)	77 (15.9%)	136 (12.7%)	108 (15.0%)	13 (24.1%)	24 (6.2%)	8 (4.9%)
North-East England	2,830 (12.2%)	3,011 (12.9%)	58 (12.0%)	93 (8.7%)	67 (9.3%)	3 (5.6%)	12 (3.1%)	4 (2.5%)
Yorkshire	3,792 (16.4%)	3,764 (16.1%)	68 (14.0%)	193 (18.0%)	111 (15.4%)	10 (18.5%)	35 (9.0%)	10 (6.1%)
West Midlands	1,957 (8.5%)	1,929 (8.3%)	38 (7.9%)	83 (7.7%)	60 (8.3%)	6 (11.1%)	59 (15.1%)	22 (13.5%)
East Midlands	1,863 (8.1%)	1,889 (8.1%)	38 (7.9%)	84 (7.8%)	60 (8.3%)	4 (7.4%)	15 (3.8%)	15 (9.2%)
South-East England	1,764 (7.6%)	1,790 (7.7%)	28 (5.8%)	69 (6.4%)	57 (7.9%)	5 (9.3%)	17 (4.4%)	9 (5.5%)
South-West England	1,698 (7.3%)	1,871 (8.0%)	38 (7.9%)	103 (9.6%)	60 (8.3%)	4 (7.4%)	10 (2.6%)	3 (1.8%)
Wales	882 (3.8%)	1,084 (4.6%)	23 (4.8%)	53 (4.9%)	35 (4.8%)	0 (0.0%)	5 (1.3%)	3 (1.8%)
Scotland	1,815 (7.9%)	1,611 (6.9%)	36 (7.4%)	69 (6.4%)	43 (6.0%)	1 (1.9%)	5 (1.3%)	1 (0.6%)
Fasting time (hours)	3.8 (2.5)	3.7 (2.3)	3.8 (2.2)	3.6 (2.3)	3.7 (2.4)	3.8 (3.0)	4.1 (2.4)	4.1 (2.5)
Body mass index (kg/m^2)								
<20	358 (1.5%)	548 (2.3%)	35 (7.2%)	65 (6.1%)	45 (6.2%)	5 (9.3%)	6 (1.5%)	3 (1.8%)
20.0–22.4	5,911 (25.6%)	7,659 (32.8%)	219 (45.2%)	503 (46.8%)	338 (46.8%)	29 (53.7%)	121 (31.0%)	52 (31.9%)
22.5-24.9	10,239	9,936	159 (32.9%)	379 (35.3%)	251 (34.8%)	13 (24.1%)	180 (46.2%)	71 (43.6%)

	(44.3%)	(42.6%)						
≥25	6,513 (28.2%)	5,105 (21.9%)	68 (14.0%)	123 (11.5%)	87 (12.0%)	7 (13.0%)	80 (20.5%)	35 (21.5%)
Unknown	95 (0.4%)	75 (0.3%)	3 (0.6%)	4 (0.4%)	1 (0.1%)	0 (0.0%)	3 (0.8%)	2 (1.2%)
Alcohol consumption								
<1g/d	3,600 (15.6%)	4,540 (19.5%)	151 (31.2%)	255 (23.7%)	169 (23.4%)	27 (50.0%)	156 (40.0%)	139 (85.3%)
1-7g/d	5,075 (22.0%)	6,721 (28.8%)	152 (31.4%)	293 (27.3%)	200 (27.7%)	9 (16.7%)	96 (24.6%)	18 (11.0%)
8-15g/d	4,986 (21.6%)	5,521 (23.7%)	76 (15.7%)	235 (21.9%)	166 (23.0%)	7 (13.0%)	52 (13.3%)	6 (3.7%)
≥16g/d	9,439 (40.8%)	6,525 (28.0%)	105 (21.7%)	291 (27.1%)	186 (25.8%)	11 (20.4%)	84 (21.5%)	0 (0.0%)
Unknown	16 (0.1%)	16 (0.1%)	0 (0.0%)	0 (0.0%)	1 (0.1%)	0 (0.0%)	2 (0.5%)	0 (0.0%)
Smoking status								
Never	11,729 (50.7%)	12,933 (55.5%)	268 (55.4%)	597 (55.6%)	422 (58.4%)	29 (53.7%)	291 (74.6%)	148 (90.8%)
Previous	8,557 (37.0%)	8,178 (35.1%)	179 (37.0%)	389 (36.2%)	237 (32.8%)	20 (37.0%)	42 (10.8%)	10 (6.1%)
Current < 15 cigs/d	707 (3.1%)	690 (3.0%)	14 (2.9%)	29 (2.7%)	21 (2.9%)	1 (1.9%)	21 (5.4%)	2 (1.2%)
Current 15 or more cigs/d	2,038 (8.8%)	1,438 (6.2%)	21 (4.3%)	55 (5.1%)	37 (5.1%)	3 (5.6%)	31 (7.9%)	3 (1.8%)
Unknown	85 (0.4%)	84 (0.4%)	2 (0.4%)	4 (0.4%)	5 (0.7%)	1 (1.9%)	5 (1.3%)	0 (0.0%)
Physical activity (MET hrs/wk)								
<10	4,735 (20.5%)	4,352 (18.7%)	78 (16.1%)	163 (15.2%)	128 (17.7%)	8 (14.8%)	95 (24.4%)	45 (27.6%)
10-49.9	9,066 (39.2%)	9,883 (42.4%)	183 (37.8%)	482 (44.9%)	324 (44.9%)	29 (53.7%)	127 (32.6%)	51 (31.3%)
≥50	4,089 (17.7%)	3,963 (17.0%)	124 (25.6%)	214 (19.9%)	137 (19.0%)	10 (18.5%)	53 (13.6%)	10 (6.1%)
Unknown	5,226 (22.6%)	5,125 (22.0%)	99 (20.5%)	215 (20.0%)	133 (18.4%)	7 (13.0%)	115 (29.5%)	57 (35.0%)

Numbers shown are mean (SD) or number (%).

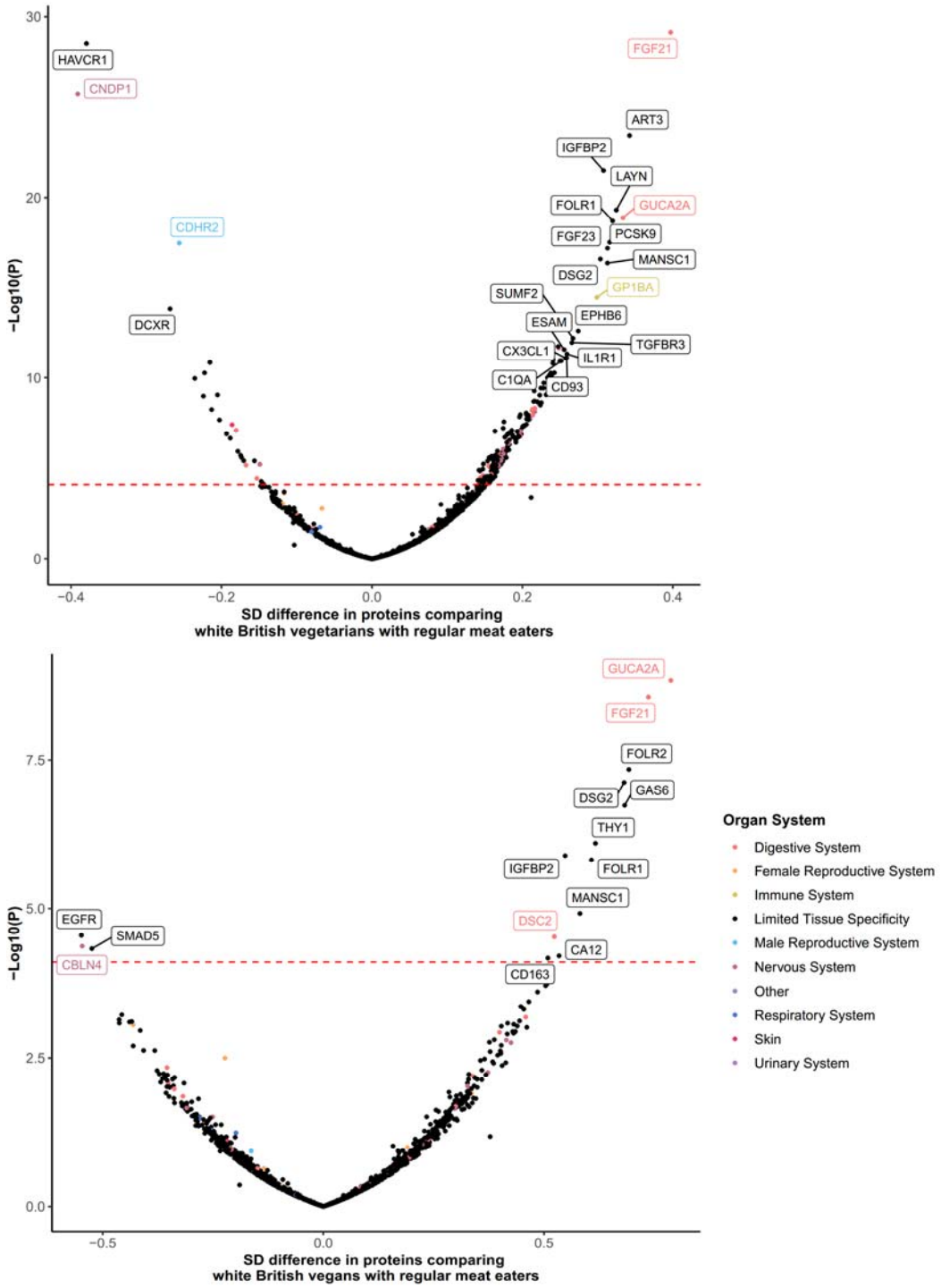

Figure 1: Volcano plots of proteins in white British vegetarians (top) and vegans (bottom) compared with regular meat eaters.

The red dotted line signifies *p*-value threshold for statistical significance. Each dot represents one protein, which were colour-coded by whether the protein is majority expressed (>50%) in one tissue type. Results were based on the multivariable model adjusted for age at recruitment, sex, region, fasting status, body mass index, alcohol consumption, smoking status and physical activity.

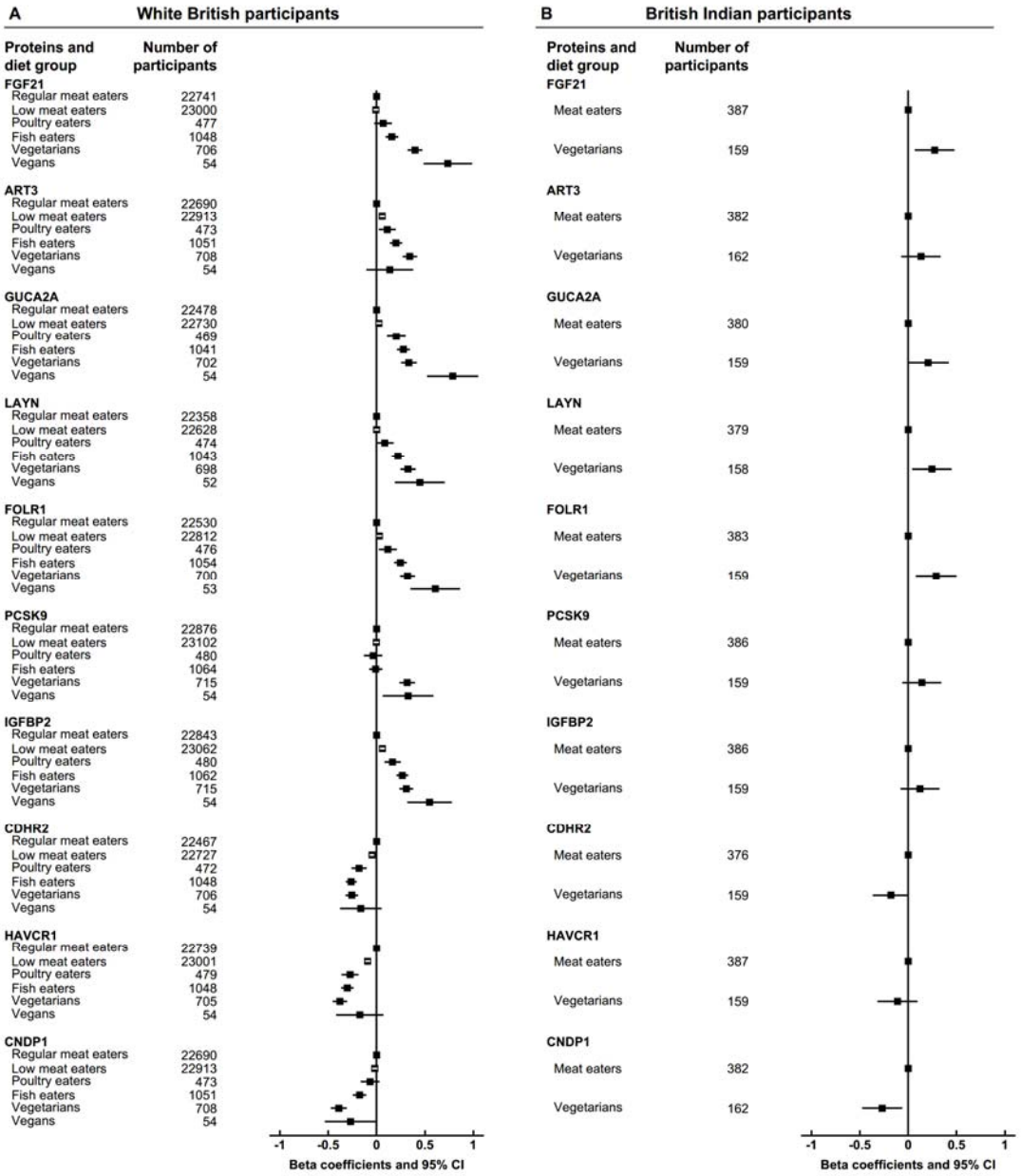

Figure 2: Top 10 proteins in vegetarians by diet group and ethnicity.

The top 10 proteins were selected by ranking the *p*-values of proteins comparing white British vegetarians with regular meat eaters, and sorted by betas in white British vegetarians, where the betas represent SD differences. Results were based on the multivariable model adjusted for age at recruitment, sex, region, fasting status, body mass index, alcohol consumption, smoking status and physical activity.

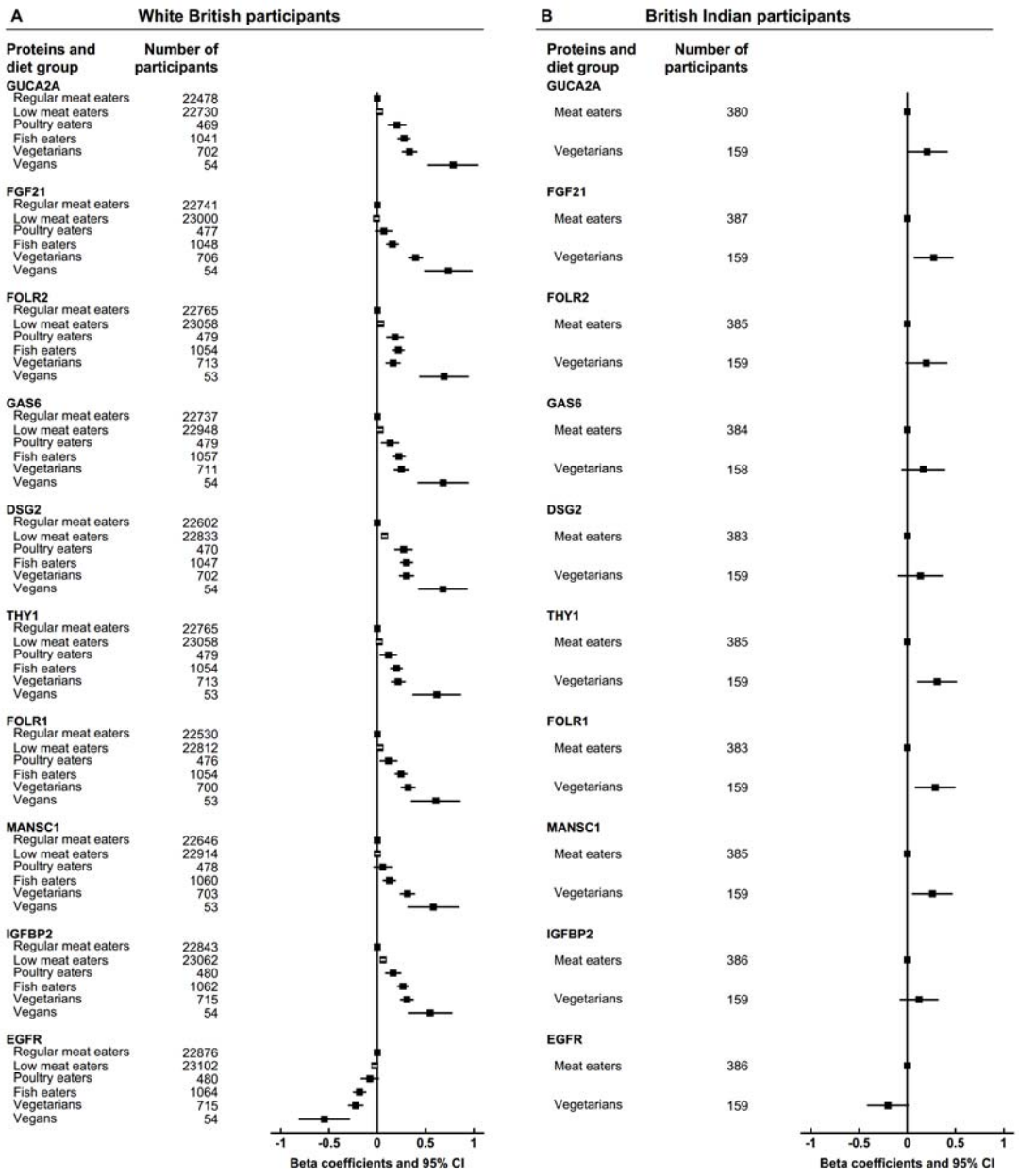

Figure 3: Top 10 proteins in vegans by diet group and ethnicity.

The top 10 proteins were selected by ranking the *p*-values of proteins comparing white British vegans with regular meat eaters, and sorted by betas in white British vegans, where the betas represent SD differences. Results were based on the multivariable model adjusted for age at recruitment, sex, region, fasting status, body mass index, alcohol consumption, smoking status and physical activity.