Faecal metaproteomics analysis reveals a high cardiovascular risk profile across healthy individuals and heart failure patients

Chaoran Yang, Leticia Camargo Tavares, Han-Chung Lee, Joel R. Steele, Rosilene V. Ribeiro, Anna L. Beale, Stephanie Yiallourou, Melinda J. Carrington, David M. Kaye, Geoffrey A. Head, Ralf B. Schittenhelm & Francine Z. Marques

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS | Check for updates

Faecal metaproteomics analysis reveals a high cardiovascular risk profile across healthy individuals and heart failure patients

Chaoran Yang [ID][a], Leticia Camargo Tavares [ID][a], Han-Chung Lee[b], Joel R. Steele[b], Rosilene V. Ribeiro[c], Anna L. Beale[d,e], Stephanie Yiallourou[f], Melinda J. Carrington[f], David M. Kaye[d,e,g], Geoffrey A. Head[h,i], Ralf B. Schittenhelm[b], and Francine Z. Marques [ID][a,d,j]

[a]Hypertension Research Laboratory, School of Biological Sciences, Faculty of Science, Monash, Clayton, Australia; [b]Monash Proteomics & Metabolomics Platform, Monash Biomedicine Discovery Institute & Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Australia; [c]Charles Perkins Centre, University of Sydney, Sydney, Australia; [d]Heart Failure Research Laboratory, Baker Heart and Diabetes Institute, Melbourne, Australia; [e]Department of Cardiology, Alfred Hospital, Melbourne, Australia; [f]Preclinical Disease and Prevention Unit, Baker Heart and Diabetes Institute, Melbourne, Australia; [g]School of Translational Medicine, Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Australia; [h]Neuropharmacology Laboratory, Baker Heart and Diabetes Institute, Melbourne, Australia; [i]Department of Pharmacology, Faculty of Medicine Nursing and Health Sciences, Monash University, Melbourne, Australia; [j]Victorian Heart Institute, Monash University, Clayton, Australia
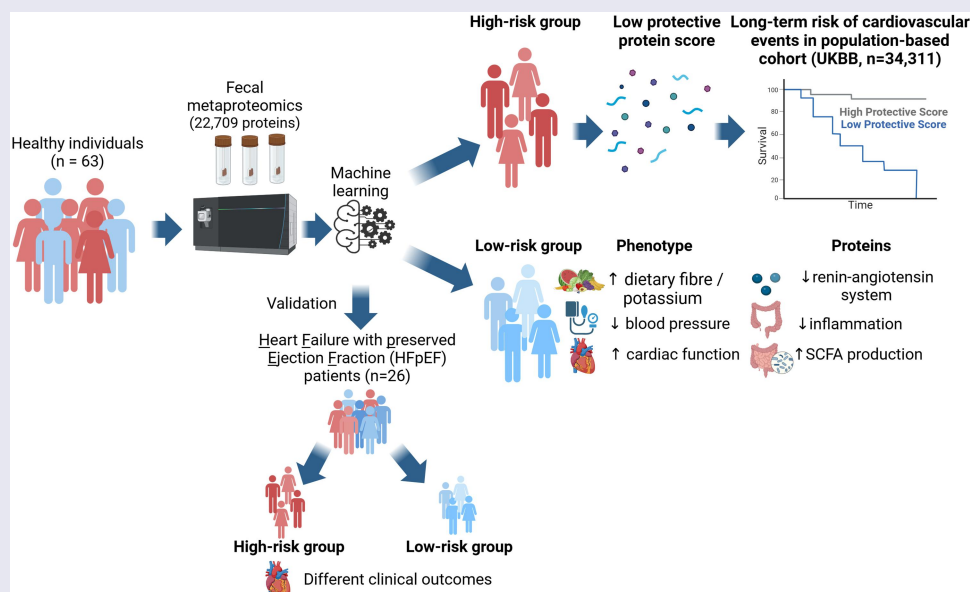
ABSTRACT

The gut microbiota is a crucial link between diet and cardiovascular disease (CVD). Using fecal metaproteomics, a method that concurrently captures human gut and microbiome proteins, we determined the crosstalk between gut microbiome, diet, gut health, and CVD. Traditional CVD risk factors (age, BMI, sex, blood pressure) explained < 10% of the proteome variance. However, unsupervised human protein-based clustering analysis revealed two distinct CVD risk clusters (low-risk and high-risk) with different blood pressure (by 9 mmHg) and sex-dependent dietary potassium and fiber intake. In the human proteome, the low-risk group had lower angiotensin-converting enzymes, inflammatory proteins associated with neutrophil extracellular trap formation and auto-immune diseases. In the microbial proteome, the low-risk group had higher expression of phosphate acetyltransferase that produces SCFAs, particularly in fiber-fermenting bacteria. This model identified severity across phenotypes in heart failure patients and long-term risk of cardiovascular events in a large population-based cohort. These findings underscore multifactorial gut-to-host mechanisms that may underlie risk factors for CVD.

Introduction

According to the latest Global Burden of Disease Study, both high blood pressure (BP) and dietary risks were among the leading global risk factors for death, as they are both recognized as risk factors for the development of cardiovascular diseases (CVDs).[1,2] The relationship between BP and diet is well-established,[3] with dietary changes listed as first-line therapy for hypertension treatment in international guidelines.[4] The Dietary Approaches to Stop Hypertension (DASH) and Mediterranean diets are examples of how diet can be successfully manipulated to reduce BP.[4] These diets are characterized by a high intake of potassium- and fiber-rich foods, such as fruits, vegetables, legumes, whole grains and nuts, and a low intake of red meat, saturated fat, refined sugar, processed foods, and sodium.

Mounting evidence in the past decade has underscored the pivotal role of gut dysbiosis (i.e., changes to the gut microbiome and gut barrier) in the development of hypertension[5] and CVDs.[6,7] Unhealthy diets directly impact human health through metabolic[8] and gene expression changes[9] and indirectly affect host health, including BP and cardiovascular health, by interacting with the gut microbiome and responding to downstream signaling.[10,11] However, the intricate mechanisms underlying the host and gut microbiome interplay in CVD development remain largely unknown.

Through the assessment of microbial DNA using amplicon (e.g., 16S) or shotgun sequencing, metagenomic analyses have revealed changes in gut microbiome taxonomy associated with CVDs.[5–7] Functional profiling and differential abundance analyses of gene families can indicate shifts in gut microbiome functions. However, metagenomic analyses are limited to estimating the presence of microbial genes, which may differ from the actual protein levels expressed by the microbial communities. Moreover, metagenome analyses cannot elucidate interactions between the host and gut microbiome. This is due to their inability to provide insights into human protein expression in the gut lumen. Thus, functional omics, such as fecal metaproteomics, have been suggested as an alternative technique to address the limitations of metagenomic studies. Fecal metaproteomics is an emerging technique capable of capturing all types of proteins in a fecal sample, including human, microbial, and food proteins,[12] along with the biomass of different types of gut microbiome.[13] This makes fecal metaproteomics suitable for elucidating the functional crosstalk between the host and gut microbiome, holding enormous potential. However, it remains to be explored in most disease settings, including in cardiovascular research.

In this study, we aimed to understand the functional crosstalk between the host and gut microbiome to CVD risk. We hypothesized that differences in human and microbial fecal proteins would be observed in healthy participants at higher risk of CVD and that these would influence CVD outcomes. We performed metaproteome analysis on fecal samples from healthy individuals whose BP was measured by ambulatory BP monitoring (ABPM) and patients with heart failure with preserved ejection fraction (HFpEF), diagnosed with cardiac catheterization. We further employed machine learning to establish connections between the expression profile of human proteins in the gut lumen, the composition and expression profile of gut microbial proteins, and cardiovascular risk factors. We then assessed these proteins in a subset of the UK Biobank cohort, where proteomics data is available, to validate that these proteins increased the long-term risk of cardiovascular events. Our results suggest that cardiovascular risk factors have a combined impact on the metaproteome, and that the resulting human and microbial protein changes may drive some mechanisms behind the development of CVD.

Materials and methods

Study population

This study complied with the Declaration of Helsinki and was approved by the human research ethics committee of the Alfred Hospital, Melbourne, Australia (approval 415/16 and 477/17). All participants provided informed consent and were recruited between October 2016 and January 2020. The study was registered in the Australian New Zealand Clinical Trials Registry under ACTRN12620000958987. The recruitment

and inclusion/exclusion criteria of the healthy cohort (VicGut) and HFpEF patients were described previously.[14,15] Briefly, for the VicGut, healthy participants were 40–70 years of age, either sex, had body mass index (BMI) 18.5–30 kg/m², were not using BP-lowering medication, and were recruited in two sites across Melbourne and Shepparton (Figure 1a). These participants were healthy, without diabetes or gastrointestinal disorders. The HFpEF patients were diagnosed by
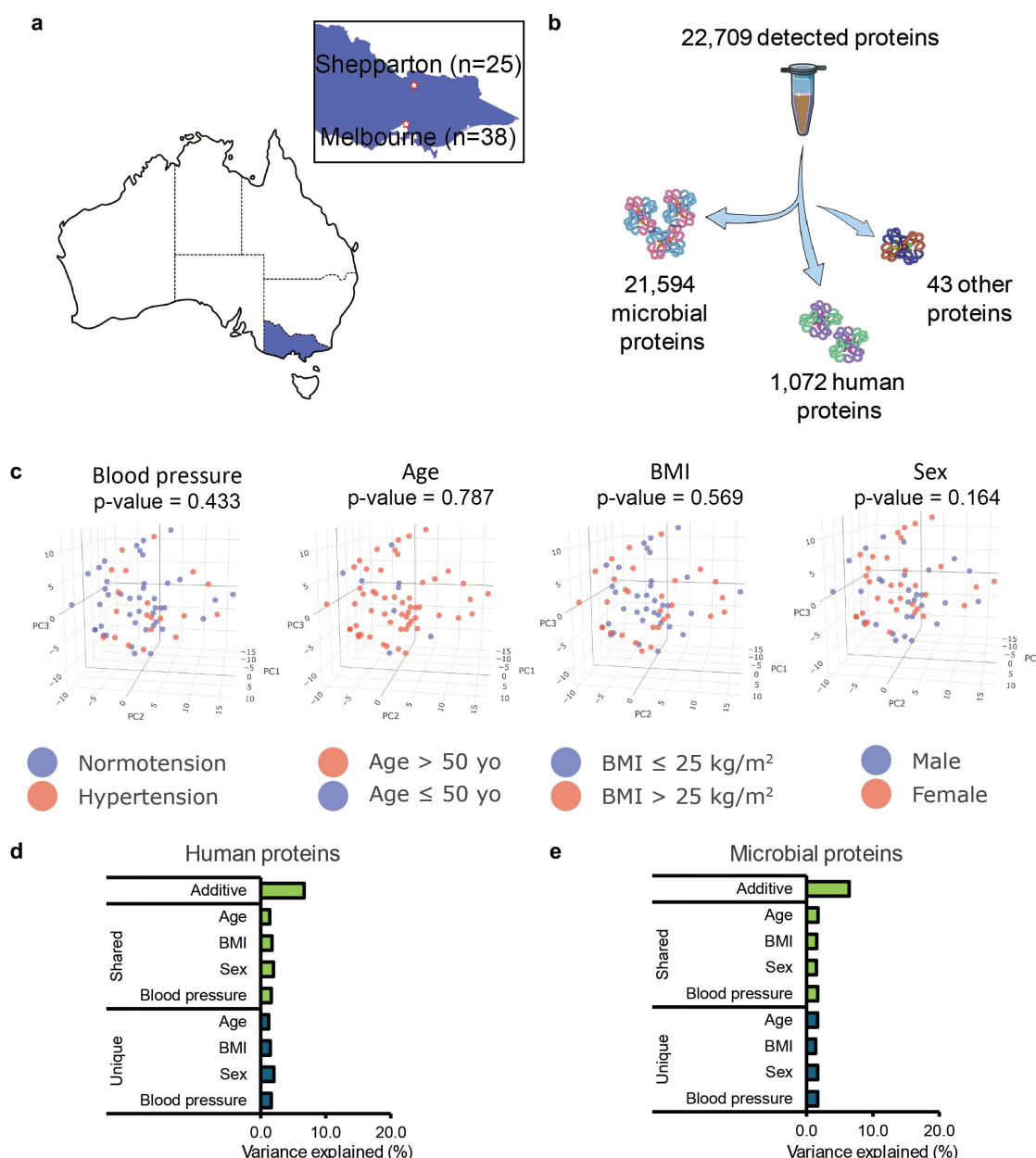
Figure 1. Overview of fecal metaproteome data from the VicGut cohort and the influence of conventional cardiovascular risk factors. (a). Geographic distribution of sample collection sites depicted on a map of Australia and Victoria. Fecal samples were collected from participants at the Baker Heart and Diabetes Institute in metropolitan Melbourne and Shepparton, a regional town in Victoria. (b). Graphical representation illustrating the distribution of detected proteins in the metaproteome dataset and their origins. Among the 22,709 detected proteins 21,594 were of microbial origin, 1,072 were human proteins, and 43 proteins originated from other sources such as food. Some icons are from SciDraw and Bioicons program (CC-BY). (c). PCA plot illustrating the influence of traditional cardiovascular risk factors on gut lumen human protein expression profiles. (d-e). Percentage of variation in d. human and (e). microbial proteins could be explained by four conventional cardiovascular disease risk factors: sex, BMI, age, and blood pressure. The 'shared model' represents an additive model incorporating all four risk factors, while the 'unique model' indicates separate analyses of individual risk factors. Sample size $n = 63$.

a right heart catheterization (RHC), which confirmed a resting pulmonary capillary wedge pressure (PCWP) ≥15 mmHg or exercise PCWP ≥25 mmHg and a left ventricular ejection fraction (LVEF) >50% according to recognized diagnostic criteria for HFpEF. Exclusion criteria included, amongst others, intake of antibiotics or probiotics in the last 6 months.

Blood pressure measurement and hypertension diagnosis

In the VicGut cohort, office BP was measured using an automatic BP monitor (Omron), and 2–3 measurements were taken.[15] Twenty-four-hour BP was measured using a calibrated ambulatory ABPM device (AND or SpaceLabs).[15] Hypertension was diagnosed based on 24-hour systolic BP >130 mmHg and/or diastolic BP >80 mmHg.[16]

Exercise right heart catheterisation protocol

Exercise RHC was performed using supine cycle ergometry, as described previously.[14] A 7-F Swan-Ganz catheter was inserted through the brachial or internal jugular vein with the patient under local anesthesia. End-expiratory measurements were taken from the right atrium, right ventricle, pulmonary artery, and pulmonary capillary wedge pressure (PCWP) was measured.

Food frequency questionnaire

Dietary intake over a period of 12 months was assessed using the Dietary Questionnaire for Epidemiological Studies (DQES) version 3.2, a self-administered and validated food frequency questionnaire (FFQ) developed by the Cancer Council Victoria that reflects the dietary intake of the Australian population.[17] Dietary intake estimates of 98 nutrients were derived from two Australian databases, AUSNUT 2007[18] and NUTTAB 2010.[19] Out of these, the following nutrients were assessed due to their relevance to BP and/or gut microbiome: sodium (mg/day), potassium (mg/day), fiber (g/day), fat (g/day), protein (g/day), fruit (g/day), vegetables (g/day), whole grains (g/day). Dietary fiber intake was classified as adequate or inadequate based on

a threshold of 25 g/day, defined according to a meta-analysis on fiber intake and CVD prevalence and mortality.[20]

Protein extraction from samples and enzymatic digestion

Metaproteomic analyses were performed on the fecal samples of 63 healthy participants and 26 HFpEF patients – these participants still had fecal samples available. Samples were cryopulverized and solubilized in 5% sodium dodecyl sulfate (SDS) 10 mm Tris HCL, with heat inactivation at 95°C for 10 minutes followed by centrifugation at 13,000 rcf for 5 minutes. Samples were cleaned by additional three times centrifugation at 13,000 rcf for 5 minutes then processed using the S-trap (Protifi) protocol as per the manufacturer's instructions.[21] Briefly, samples were reduced and alkylated using 10 mm TCEP (Thermo, #77720) and 40 mm chloroacetamide (Sigma, C0267-100 G) with incubation at 55°C for 15 minutes. Enzymatic digestion was performed using Trypsin (Promega, V528X) at a 1:50 wt:wt ratio alongside Lys-C at a 1:25 wt:wt ratio (Promega, VA1170) at 37°C for 16 hours. Digestion efficiency was greater than 89% for this analysis.

TMT labeling and fractionation

Labeling was performed using TMTpro 16plex reagent set (Lot:WD314806, Thermo Scientific) according to the manufacturer's instructions and utilized a singular reference channel (126) for each fecal sample set utilizing (two per tissue type). Pooled plexes were then fractionated utilizing high-pH RP-HPLC generating 36 fractions that were concatenated into 12 to ameliorate low complexity issues during acquisition. Additionally, a global pooled sample was fractionated into six fractions using the same method, specifically for generating a data-specific database. Each plex has been acquired individually by LC-MS/MS to maximize identifications. For each sample set, labeling efficiency was determined to be greater than 88%, utilizing data-dependent analysis of individually labeled samples.

Liquid chromatography mass spectrometry protocol

Liquid chromatography-mass spectrometric (LC-MS) analysis was conducted using the Mass Spectrometer and Nano LC system (Dionex Ultimate 3000 RSLCnano). The samples were loaded in an Acclaim PepMap RSLC (75 μm x 50 cm, nanoViper, C18, 2 μm, 100Å; Thermo Scientific) analytical column. The peptides were separated by increasing concentrations of buffer B (80% acetonitrile/0.1% formic acid) and analyzed via 2 kV nano-electrospray ionization with an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Scientific, Bremen, Germany) operated in data-dependent acquisition mode using in-house optimized parameters with 120 minutes of chromatographic separation used for each fraction. Briefly, the acquisition used three FAIMS compensation voltages (−40, −55, −70) operated under standard resolution with an ion transfer tube temperature of 300°C with a carrier gas flow rate of 4.6 L/min. Survey scans were performed at a resolution of 120,000 from 400–1,600 m/z, with a 250% AGC target and ion injection time set to auto. Fragmentation for peptide identification and reporter tag quantification were performed synchronously (10 per duty cycle per compensation voltage) with the fragmentation spectra generated in the ion trap using CID with turbo scan rate; MS3 reporter ion measurements were performed in the orbitrap with a resolution of 50,000. Dynamic exclusion was applied for 60 seconds across all compensation voltages with only one charge state per precursor selected for fragmentation.

Assembly of a sample-specific database

The sample-specific database was generated using MetaLab software (HGM1.0).[22] Initially, spectral clustering was applied to the MS/MS spectra from the pooled global fecal samples to reduce dataset redundancy by removing redundant or inferior spectra. This refined list was then searched against a gut microbial gene catalog database, tailored for human microbiomes by MetaLab, to create a candidate protein list for database construction.

Mass spectrometric data analysis

The raw data files were analyzed using Proteome Discoverer (v2.5.0.400, Thermo Scientific) to obtain protein identifications and their respective reporter ion intensities using in-house standard parameters with sequest. The sample-specific database generated by MetaLab was used for protein identification at a 1% false discovery rate (FDR) alongside a common contaminants database. Reporter ion quantifiers used a unique plus razor with analysis centered on protein groups for shared peptide sequences using all peptides for abundances determination, value output and quantitative values were corrected against stable isotope label impurities according to the manufacturer's values as per the lot number.

Metaproteomic downstream analysis

The abundance of proteins was scaled to the sum of the total protein abundance and multiplied by 10^5 to facilitate subsequent analysis. The abundance of all detected proteins under the corresponding taxonomy was summed and log-transformed before differential analysis to measure biomass of different microbial taxonomies. Only proteins present in > 50% of samples were kept for protein-level analysis. This resulted in 373 gut lumen human proteins and 7,519 high-quality microbial proteins detected in > 50% of samples that were considered as high-quality proteins and were used for the following analyses. Scaled abundance was then log-transformed, and missing values that remained were imputed by the K-nearest neighbor (KNN) method, widely used in current metaproteomic data analysis.[23,24] K-mean unsupervised clustering was performed using R 4.3 with the default setting.

Limma (v3.56.2)[25] was used to analyze the protein expression differences or the biomass of microbial taxonomies between different groups. Pathway overrepresentation analysis of human protein was performed using an online tool Enrichr.[26] The protein-protein interaction network was constructed using the online STRING database.[27] The co-expression network was constructed using WGCNA (v1.72.5),[28] and edges

with a weight > 0.05 were kept in an outputted network. A motif search on promoter regions was performed using Homer (v4.11).[29]

The random forest (RF) model was trained using the scikit-learn package (v1.1.3) run in a Python 3.10 environment with 100 estimators. Minimum samples per split and leaf were set to 3 to avoid model overfitting.

UK biobank data processing

The UK Biobank follows the principles of the Declaration of Helsinki and received ethical approval from a human ethics committee (11/NW/0382), with access approved under application number 86,879. The UK Biobank is a comprehensive longitudinal cohort study, comprising over 500,000 participants recruited between 2006 and 2010.[30] We assessed information from 53,014 participants who had plasma proteomics data available. To filter out low-quality samples and proteins, we excluded proteins detected in fewer than 50% of samples and samples with less than 20% proteins detected, resulting in a final dataset of 52,936 participants and 2,920 proteins. Missing values in the proteomics data were imputed using the KNN method, consistent with our approach for metaproteomic datasets.

We included 34,311 participants who were not taking BP-affecting medications in further analyses. Major adverse cardiovascular events (MACE) were defined following a method outlined by Zheng & Tavares et al. (2024).[31] Briefly, MACE included hospital admissions and death registers (ICD code I20.0, I21.*,I24.8, I24.9, I50.*, I63.*, I64.*) as well as surgical procedures (K40.*, K41.*, K42.*, K43.*, K44.*, K45.*, K46.*,K49.*, K50.*,K75.*) associated with acute coronary syndrome (ACS), ischemic stroke, or heart failure (HF). The circulating protein scores were calculated using the following method. First, we extracted circulating proteins also detected in the metaproteomic datasets. Proteins with zero importance in the human protein-based random forest model were excluded. The metaproteomic circulating protein scores were subsequently calculated as a weighted sum based on the random forest model's importance. Specifically, the score was computed as Σ (random forest importance × upregulated protein level) – Σ (random forest importance × downregulated protein level). High scores indicate a protein expression profile similar to the low-risk group we identified.

Statistical analysis

Statistical tests and data visualization were performed using R 4.3. Normally distributed data was analyzed using a two-tailed Welch two-sample t-test. Non-normally distributed data was analyzed using a two-tailed Wilcoxon-Mann-Whitney test or permutation test for a small sample size ($n < 30$). Binomial data, such as sex and living area among patients, were analyzed using the chi-square (X^2) test. In contrast, other binomial data were analyzed using Fisher's exact test due to the presence of low frequencies. Time-to-event analyses were conducted using a Cox proportional hazards model with the survival (version 3.6.4),[32] survminer (version 0.4.9), and adjustedCurves (version 0.11.2)[33] packages in R. Permutational analysis of variance (PERMANOVA) test based on Euclidean distance was conducted using the R package vegan (https://CRAN.R-project.org/package=vegan) to assess the equivalence of centroids among different groups. For high-dimensional data analysis, p-values were adjusted using the Benjamini – Hochberg FDR method with a cutoff of q < 0.05. Significance was determined as P or q < 0.05.

Results

Traditional cardiovascular risk factors poorly capture expression profiles of gut lumen human protein and microbial proteins

Healthy participants were recruited from Melbourne and Shepparton in Victoria, Australia – their baseline characteristics are detailed in Table 1 and Figure 1a. Fecal samples from these participants were analyzed using mass spectrometry to determine the fecal metaproteome. A total of 22,709 proteins were identified in the metaproteome data from these participants (Figure 1b). Among these 21,594 were classified as microbial proteins originating from seven bacterial phyla (Figure 1b and S1A),

Table 1. Characteristics of participants included in the healthy cohort and low- and high-risk groups identified.

Variable	Healthy participants ($n = 63$)	Cluster 1 (Low-risk) ($n = 26$)	Cluster 2 (High-risk) ($n = 37$)	P-value
Age	60[34,35]	60[34,36]	60[35,37]	0.695
Female	54.0% (34)	65.4% (17)	40.6% (17)	0.205
Regional patients	39.7% (25)	53.8% (14)	29.7% (11)	0.096
BMI (kg/cm^2)	25.036 ± 2.871	24.801 ± 2.655	25.202 ± 3.038	0.581
WHR	0.866 ± 0.088	0.856 ± 0.078	0.874 ± 0.094	0.419
Daytime SBP (mmHg)	128.992 ± 14.974	126.789 ± 15.769	130.539 ± 14.405	0.34
Daytime PP (mmHg)	49.926 ± 8.795	48.710 ± 9.584	50.781 ± 8.223	0.3754
Nighttime SBP (mmHg)	113.680 ± 16.523	108.597 ± 14.637	117.252 ± 17.018	0.035*
Nighttime PP (mmHg)	46.473 ± 8.978	43.393 ± 8.870	48.637 ± 8.516	0.023*
Overall HR (mmHg)	73.680 ± 12.066	77.673 ± 11.334	70.873 ± 11.913	0.025*
Fiber intake (g/day)	23.856 ± 8.796	24.915 ± 8.992	23.092 ± 8.699	0.428
Sodium (g/day)	1.978[1.643,2.539]	2.164[1.811,2.766]	1.950[1.53,2.24]	0.046*
Potassium (g/day)	3.622[2.967,4.348]	3.987[3.318,4.507]	3.408[2.760,3.956]	0.061
Sodium/Potassium	0.582 ± 0.171	0.596 ± 0.204	0.571 ± 0.143	0.605
CVD risk-associated dietary factor analysis by sex				

	Male			Female		
	Cluster 1 (Low-risk)	Cluster 2 (High-risk)	P-value	Cluster 1 (Low-risk)	Cluster 2 (High-risk)	P-value
Fiber intake (g/day)	30.581 ± 8.254	23.242 ± 6.943	0.022*	21.915 ± 8.040	22.925 ± 10.548	0.756
Sodium (g/day)	2.080[1.811,2.766]	1.924[1.573,2.324]	0.143	2.301[1.867,2.786]	1.919[1.486,2.242]	0.121
Potassium (g/day)	3.987[3.508,4.507]	3.128[2.805,3.700]	0.021*	3.896[3.124,4.527]	3.642[2.744,4.221]	0.769
Sodium/Potassium	0.565 ± 0.206	0.584 ± 0.124	0.772	0.614 ± 0.207	0.558 ± 0.164	0.4

Binomial data are presented as percentage (number of individual). Normally distributed data are presented as mean ± standard deviation. Non-normally distributed data are presented as median [Q25, Q75]. BMI: Body mass index; HR: Heart rate; SBP: Systolic blood pressure; PP: Pulse pressure; WHR: Waist-to-hip ratio. *$p < 0.05$.

with the majority originating from the Bacillota and Bacteroidota phyla, and a smaller proportion from Campylobacterota, Pseudomonadota, Thermodesulfobacteriota, and Verrucomicrobiota. This taxonomic distribution aligns with the well-defined structure of the human gut microbiota.[38] Additionally, 1,072 were identified as human proteins, representing gut lumen human proteins, and 43 proteins originated from other eukaryotic organisms, potentially linked to food intake (Figure 1b).

Based on previous findings linking gut microbiota to cardiovascular health,[39,40] we initially investigated whether the expression of gut lumen human proteins or microbial proteins differed between individual traditional risk factors for CVD, including hypertension, age (>50 years versus ≤50 years), BMI (>25 kg/m^2 or ≤25 kg/m^2), and sex (Figure 1c and S1B). None of these were significantly different. The variation in expression levels of both human and microbial proteins within the gut lumen could only be minimally explained by these traditional factors (less than 10%) (Figures 1d-e and S1B). BP, for example, explained only 1.6% of the variance for gut lumen human proteins ($p = 0.433$) and 1.7% for microbial proteins ($p = 0.397$) (Figure 1d). Even an additive model incorporating all these

factors could explain only 6.7% of the variance in gut lumen human proteins and 6.4% of the variance in microbial proteins (Figure 1e). Although dietary fiber intake has been previously associated with reduced CVD risk by promoting the production of short-chain fatty acids (SCFAs) by the gut microbiome,[11,41] we found a significant difference only in the human proteins ($p = 0.027$), with seven muscle-related proteins underrepresented in participants with inadequate dietary fiber intake (Figure S2C-E). These could be potentially related to gut dysmotility observed in constipation, an emerging risk factor for CVD.[31,42,43]

Unsupervised clustering identified distinct cardiovascular risk groups through metaproteome expression

We then used K-means, an unsupervised machine-learning algorithm renowned for its robustness, to cluster samples based on human and microbial protein expression. To select proper input PCs for cluster, we examined the top 50 human proteins contributing to different principal components. PC1 has a diverse function and may influence the individual differences. This included pathways related to insulin secretion and pathogenic

bacterial infection (Figures 2A and S3A). Conversely, proteins contributing to PC2 included those involved in the activation of the immune system such as lipocalin-2 (LCN2) and several immunoglobulins (Figures 2a and S3B, Table S1). Proteins contributing to PC3 included angiotensin converting enzyme (ACE, key for BP regulation) and mucins involved in the gut epithelial barrier (MUC12, MUC13), and predominantly played a role in pathways related to pancreatic secretion, protein digestion and absorption, and the renin-angiotensin system (RAS)[44] (Figures 2a and S3C, Table S1). Therefore, we did K-means analysis based on PC2 and PC3 for their relevance to CVD risk. This clustering method distinctly categorized samples into two clusters (Figure 2b-c), with a p-value <0.001 (under the detectable lower limit of PERMANOVA), indicating significant differences in human (Figure 2b) and microbial protein expression (Figure 2c) between the clusters.

A comparison of gut lumen human protein expression profiles between the two clusters revealed 25 overrepresented proteins in Cluster 1 (Figure 2d, Table S2). These proteins formed a protein-protein network (Figure S4A). Key over-represented pathways included pancreatic secretion, protein digestion and absorption, glycolysis/gluconeogenesis, and proteins involved in insulin signaling-mediated glucose transport proteins including YWHAE and YWHAH (Figure 2e, Table S4). Among these, we identified ATP2A2, a protein associated with systolic BP in a genome-wide association study,[45] which dysfunction leads to increased BP via oxidative stress;[46] CELA3B, an elastase that may be involved in cholesterol metabolism and transport in the intestine;[47] and MYO1D, a myosin that maintains gut epithelial integrity.[48]

Forty proteins were underrepresented in Cluster 1 (Figures 2d, Table S3). Relevant underrepresented pathways included neutrophil extracellular trap formation (a bacterial defense mechanism) and conversion of angiotensinogen to angiotensin II, and those associated with auto-immune regulation (including miRNA role in immune response in sepsis, asthma, systemic lupus erythematosus) (Figure 2f, Table S4). These proteins formed a tight protein-protein interaction network (Figure 2G). Central to this network were the

intestinal inflammation marker calprotectin (S100A8)[49] and an associated protein, S100A12.[50] LCN2 could interact with another MUC13, a negative regulator of tight junction.[51] Besides, another underrepresented protein, proteoglycan 2 (PRG2), also called eosinophil major basic protein (MBP), could suppress occludin expression in the gut epithelium and destabilize the gut epithelial barrier.[52] These findings indicate that gut epithelium barrier integrity may be compromised in Cluster 2. We also found two angiotensin-converting enzymes, cathepsin D (CTSD)[53] and cathepsin G (CTSG)[54] (Figure 2h), which may drive an exaggerated RAS observed in hypertension. CTSD is associated with cardiovascular risk,[55] worse CVD outcomes,[56] and pro-inflammatory factors.[34,37] Indeed, CTSD, but not CTSG, was significantly correlated to night pulse pressure in our dataset (Figure 2i). The small intestine and colon are among the tissues with the highest CTSD and CTSG expression levels (Figure S4B-C), suggesting a role of gut CTSD in BP regulation. LCN2 also serves as a neutrophil gelatinase-associated lipocalin previously shown to lead to cardiac hypertrophy and failure,[57] which also acts as part of the innate immune system, limiting bacterial growth[58] and preventing intestinal inflammation.[59] Together with LCN2, other key neutrophil-relevant proteins, including ELANE and MPO, were also underrepresented.

Fecal metaproteomics clusters are associated with cardiovascular risk factors

The combination of proteins and pathways related to CVD and inflammation prompted the hypothesis that these clusters may represent distinct CVD risk groups. To validate this hypothesis, we analyzed medical and demographic data of participants corresponding to the clusters identified in the metaproteome data (Table 1). Indeed, we found that participants in Cluster 1 had ~9 mmHg significantly lower night systolic BP and ~5 mmHg pulse pressure relative to Cluster 2, indicative of lower CVD risk.[60] These were independent of sex, BMI, and age. Accordingly, we classified Cluster 1 as a low cardiovascular risk group (referred below as 'low-risk group') and Cluster 2 as a high cardiovascular risk group
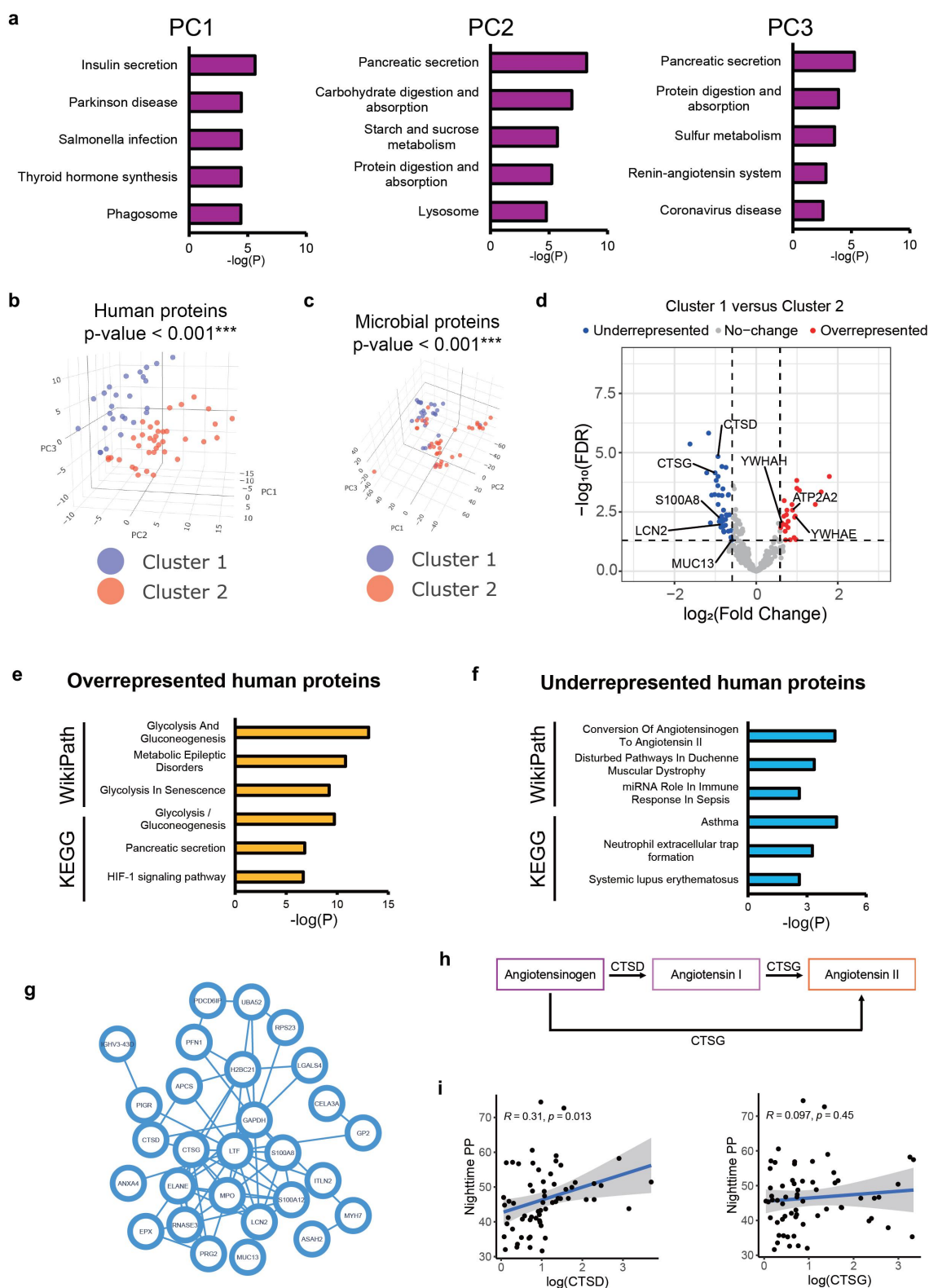
Figure 2. Differential expression of gut lumen human protein. A. Overrepresentation analysis of the top 50 proteins contributing to PC1-PC3 of human protein PCA using the KEGG database. The diverse functions observed in the top contributor proteins of PC1 suggest their association with individual differences, and thus its capacity to define groups with distinct cardiovascular disease risk is limited. In contrast, the top contributor proteins of PC2 and PC3 are primarily related to gut immune function and absorption, both of which are closely linked to human cardiovascular health. B. Principal component analysis (PCA) plot illustrating the outcome of unsupervised clustering. Two distinct clusters, cluster 1 and cluster 2, are clearly outlined by this supervised K-mean model (PERMANOVA p-value <0.001, based on Euclidean distances of all gut lumen human proteins). C. Two clusters based on gut lumen

('high-risk group'). Notably, the dietary patterns of the two clusters differed by sex. Low-risk males had a significantly higher dietary fiber and potassium intake. The low-risk group had a higher overall sodium intake (Table 1). However, after adjusting for sex, there was no significant difference in sodium intake between the high-risk and low-risk groups (Table 1).

Due to the contribution of diet to both the gut microbiome and CVD, we next assessed the variance explained by common dietary factors considered CVD risk factors across the two clusters. Collectively, dietary factors explained ~ 20% of the total variance in differentially expressed human proteins between the two groups (Figure S5A), underscoring the importance of dietary factors in mitigating CVD risk, albeit not the sole determinant. Notably, high potassium intake was observed in the low-risk group and could significantly explain ~ 5% variance in an additive model (Figure S5A).

Regulatory factors' contribution to cardiovascular risk groups

To identify potential key regulators of differentially expressed gut lumen human proteins between the two groups, we extracted the promoter sequences, defined as 1 kilobase upstream to 0.1 kilobase downstream of the transcription start site (−1kb to + 0.1kb of TSS), of coding genes associated with the differentially expressed proteins (Figures S5B-C). Subsequently, we conducted a motif search on the extracted promoter regions. Notably, the sequence motif of hypoxia-inducible factor (HIF) was significantly enriched in the promoters of genes encoding overrepresented proteins (Figures S5B-C). A moderate level of HIF1α expression is essential for the normal function of insulin-producing β cells,[61,62] which aligns with the observed association of upregulated proteins of HIF1 signaling, glycolysis/gluconeogenesis, and

pancreatic secretion in overrepresented proteins. Conversely, the sequence motif of myocyte enhancer factor-2 (MEF2) plays a role in the pathology of hypertension-induced cardiac hypertrophy,[63] was enriched in promoters of genes encoding underrepresented proteins.

Differential gut bacteria proteome and cardiovascular risk factors

Within the microbial proteins, we identified 430 significantly higher and 276 lower expressed proteins in the low-risk group (Figure 3a, Tables S5 and S6). Analysis of the source of these differentially expressed gut microbial proteins revealed that most underrepresented proteins in the low-risk group originated from the Bacillota (Firmicute) phylum. Conversely, underrepresented proteins in the high-risk group exhibited a significantly higher proportion of Bacteroidota-sourced proteins (Figure S6A). Classification based on Clusters of Orthologous Groups (COG) categories demonstrated that the most enriched functions of differentially expressed gut microbial proteins were associated with translation and metabolism (COG categories J, G, and C), indicating alterations in the bioactivity of the source gut microbiome. Notably, proteins related to translation (COG category J) were significantly overrepresented among overrepresented gut microbial proteins in the low-risk group (Figure S6B).

Considering that increased microbial SCFA production and/or sensing has been linked to reduced BP and CVD risk,[11,15,57,64] we hypothesized that SCFA production is elevated in the low-risk group. To test this hypothesis, we selected enzymes involved in the final production steps of SCFAs and examined their expression between the two groups. We observed a significant increase in the expression of acetate kinases (ackA) and phosphate acetyltransferase (pta) in several SCFA-producing bacteria, such as *Phocaeicola dorei* (*Bacteroides dorei*) and

---

human protein expression profiles also effectively distinguishes microbial protein expression profiles (PERMANOVA p-value <0.001, based on Euclidean distances of all microbial proteins). D. Volcano plot showing differentially expressed gut lumen human proteins between the low-risk and high-risk groups. Fold changes were calculated as low-risk group versus high-risk group. E-F. Overrepresentation analysis using the KEGG and Wikipathway databases for E. overrepresented and F. underrepresented proteins. G. A PPI network of underrepresented gut lumen human proteins in low-risk group. H. Schematics depicting the angiotensin converting by CTSD and CTSG protein. I. Correlation analysis between gut lumen CTSD and CTSG protein levels and nighttime PP. Sample size n = 63.
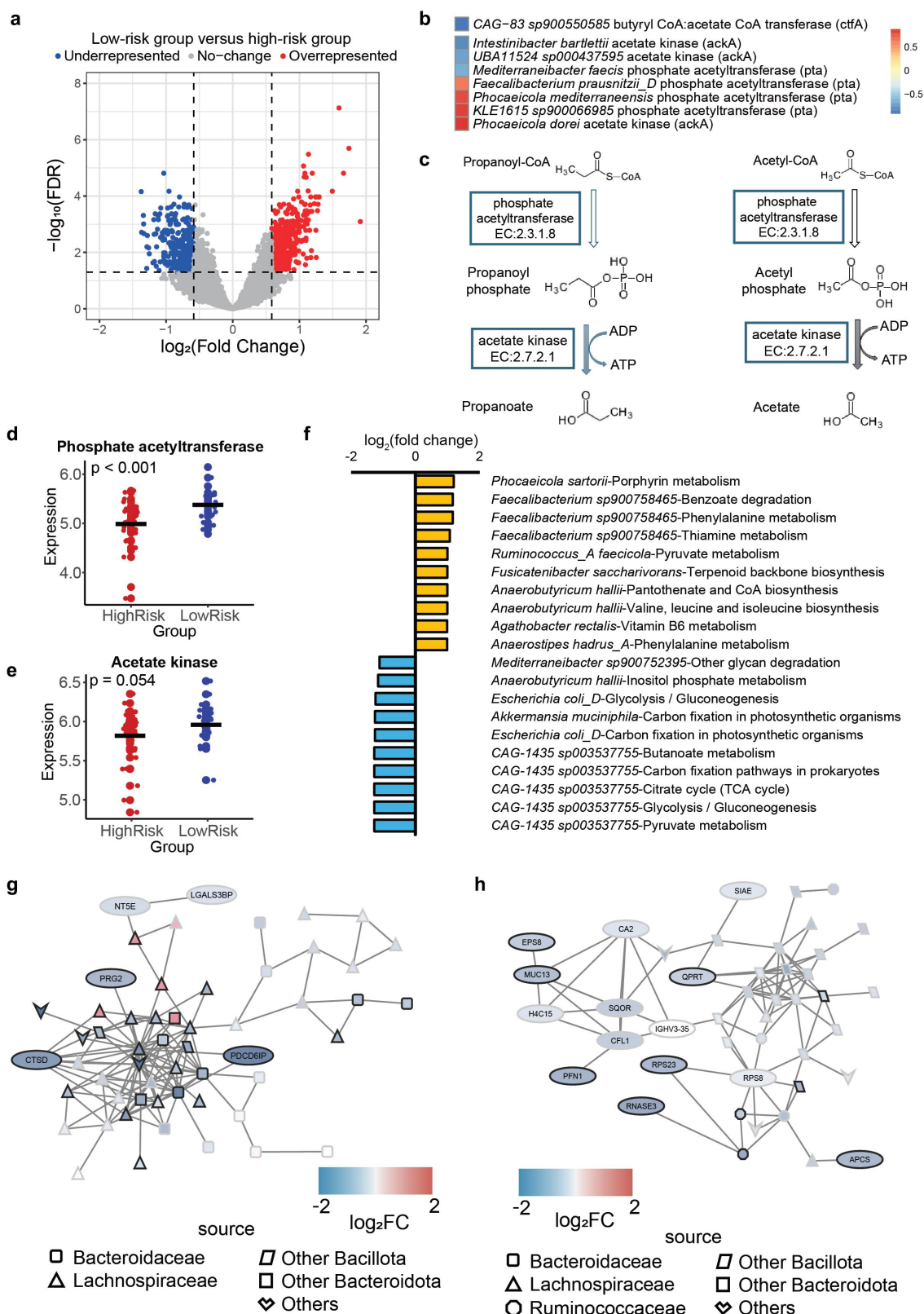
Figure 3. Differential expression of microbial proteins and pathways between two groups and crosstalk between gut lumen human proteins and microbial proteins. A. Volcano plot displaying differentially expressed gut microbial proteins between the low-risk and high-risk groups. Fold changes were calculated as the low-risk group versus the high-risk group. B. Identification of significantly differentially expressed microbial enzymes involved in directly producing short-chain fatty acids (SCFAs). C. Schematics depicting the catalytic process by which acetate kinase and phosphate acetyltransferase facilitates the chemical reaction leading to the production of acetate and propanoate. D-E. Comparison of the overall expression level of D. phosphate acetyltransferase and E. acetate kinase in

Faecalibacterium prausnitzii (Figure 3b-c), suggesting elevated levels of both acetate and propanoate production. Moreover, among four detected enzymes involved in the final production steps of SCFAs (Figure 3d-e; Figure S6C-D), the overall expression of pta in the gut microbiome population was significantly higher in the low-risk group (Figure 3d), while the overall expression of ackA was increased but did not reach statistical significance ($p = 0.054$, Figure 3e). There was no change in the levels of the other two enzymes, butyrate kinase and butyryl-CoA:acetate-CoA transferase (Figure S6C-D). Furthermore, analysis of metabolism-related KEGG pathways in each microbiota species revealed increased activity in SCFA-production pathways, such as pyruvate metabolism in *Ruminococcus_A faecicola*, and amino acid metabolism pathways contributing to SCFA production of species in *Faecalibacterium* and *Anaerobutyricum*, which are well-described SCFA-producing genera (Figure 3f).[36] Thus, the evidence above supports that the low-risk group has a gut microbiome with enhanced SCFA production capacity, which may contribute to the lower BP observed in these participants.

A previous randomized clinical trial found vitamin B1 supplementary could significantly lower BP.[65] KEGG pathways related to thiamine (vitamin B1) and pyridoxine (vitamin B6) metabolism are upregulated in the gut microbiome of the low-risk group. Vitamin B1,[35] B2,[66] B3[67] and B6[68] intake are all associated with lower BP. Moreover, vitamins B1, B2 and B3 were all significantly higher in the low-risk group, but not vitamin B6 (Table S7).

Crosstalk between the gut microbiome and human proteins

We hypothesized that there were also crosstalks between differentially presented microbiome proteins and gut lumen human proteins that may explain some CVD risk. To investigate this hypothesis, we explored potential co-expression networks between human proteins and gut microbiome proteins, which are strong indicators of crosstalk between microbiome-to-host communication. Indeed, we identified two co-expression networks significantly underrepresented in the low-risk group, likely associated with CVD risk. In the first co-expression networks, we found one downregulated network in the low-risk group that included CTSD (involved in the RAS) and PRG2, which is a proinflammatory factor destabilizing the intestinal barrier[52] (Figure 3g). In the second network (Figure 3h), several proteins known to play a relevant role in CVD development,[69] including S100A8, MUC13 and profilin 1 (PFN1), were identified.

Fecal metaproteome signature in heart failure

Because distinct expression patterns of gut lumen proteins were associated with clusters with varying risk for CVD (Figure 2c), we further investigated the differentially expressed microbial and human proteins in 26 patients diagnosed with HFpEF, an increasingly common form of heart failure in which high BP is a key risk factor. We hypothesized that the signature of protein expression in the gut lumen could serve not only as a predictor of CVD risk but also as an indicator of clinical phenotypes. HFpEF patients were, on average, 68 ± 7.5 years old, had $32.8 \pm 5.8 \, \text{kg/m}^2$ BMI, 113.2 ± 96 ng/L brain natriuretic peptide (BNP), 143.7 ± 19.1 mmHg office systolic BP, and 77% were female, as described previously.[14] These patients had a microbiome composition similar to healthy participants, with a higher abundance of the Bacillota and Bacteroideota phyla (Figure S7A). We trained two random forest models based on human proteins and microbial proteins, separately, to predict low- and high-risk individuals based on the healthy cohort data and compared their performance (Figures 4a-c, Figures S7B-D). Testing via 5-fold cross-validation yielded a high area under the curve (AUC) of 0.80 based on the microbial protein-

---

gut microbiome between low-risk group and high-risk group. F. Bar plot showing the top ten significantly overrepresented and underrepresented KEGG pathways by species in the low-risk group. Fold changes were calculated as the low-risk group versus the high-risk group. G-H. Presentation of two co-expression networks of gut lumen human proteins and microbial proteins, both found to be significantly downregulated in the low-risk group. Sample size $n = 63$.

based model (Figure S7C) and 0.91 for the human protein-based model (Figure 4a) on the Receiver Operating Characteristic plot, indicating both models were well-trained. We employed these two random forest models to classify HFpEF patients into low- and high-risk groups (Figure 4b, Figure S7D). In both models, the high-risk and low-risk groups within the HFpEf cohort did not show significant differences in age, BMI and sex (Table S8). In the human protein-based model, intestinal inflammation marker S100A8[50] and angiotensin-converting enzymes CTSD and CTSG show high evidence (Figure 4c), suggesting their pivotal associations with cardiovascular risks. In the microbial protein-based model, proteins from Bacillota species emerged as pivotal in distinguishing between the low- and high-risk groups, as evidenced by their high feature importance (Figure S7B).

Subsequently, we compared PCWP indexed to workload,[70] pulmonary compliance at exercise,[71] and workload-indexed exercise lactate[72] between high-risk and low-risk HFpEF patients identified by the two prediction models. These three indicators are well-established gold standards for evaluating clinical outcomes in HFpEF patients, associated with worse clinical outcomes.[70–72] In the human protein-based model, the high-risk group showed significantly higher PCWP (Figure 4d) and lactate (Figure 4f) after exercise indexed to workload and lower pulmonary compliance at exercise (Figures 4e). Conversely, the microbial protein-based model identified a high-risk group with significantly lower pulmonary compliance during exercise (Figure S7F), but no significant differences for the other two indicators between the high- and low-risk groups. These findings demonstrate that both the human and microbial protein-based models could effectively predict clinical outcomes in HFpEF patients, with the human protein-based model showing higher performance.

Population-level impact of differentially expressed human proteins on cardiovascular health detected in plasma

Given the higher performance of the human protein-based model, to further validate the impact of

differentially expressed human proteins between the high-risk and low-risk groups on cardiovascular health at the population level, we extracted human proteins that were differentially expressed between the high-risk and low-risk groups and overlapped with circulating proteins detected in the UK Biobank plasma proteomics dataset (Figure S8A). We calculated random forest importance-weighted circulating protein scores (see methods and Figure S8B). Higher circulating protein scores indicate a greater similarity to the protein signature identified in the low-risk group. We found that one standard deviation (SD) increase in this circulating protein score was associated with a reduction of 2.9 mmHg in systolic BP in an unadjusted model ($p < 0.001$). After adjusting for age, BMI, and sex, a one SD increase in the circulating protein score was associated with a 1.3 mmHg reduction in SBP ($p < 0.001$, Figure 4G). We then categorized the 34,311 UK Biobank participants based on whether their scores were above or below the population median. Two time-to-event models were developed to evaluate the occurrence of major CVD events within 10 years after the assessment: Model 1 was unadjusted, while Model 2 was adjusted for age, BMI, and sex (Figures 4H-I, Table S9). The results showed that participants with circulating protein scores above the median had a significantly lower incidence of CVD events within the next 10 years.

Discussion

Here, we report on fecal metaproteomics analysis in healthy and heart failure cohorts showing that human and gut microbial proteins can identify individuals at lower and higher risk of CVD. The low-risk group presented with lower night systolic BP (by 9 mmHg), angiotensin-converting enzymes and pro-inflammatory intestinal responses, and higher SCFA production capacity by the gut microbiome and sex-specific dietary potassium and fiber. We then used machine-learning analysis based on differential microbial and human protein signatures observed in the healthy low- and high-risk groups that identified HFpEF patients with worse clinical phenotypes (i.e., increased PCWP and lactate after exercise indexed to workload and reduced pulmonary compliance). Moreover, when evaluated in plasma samples, the low-risk protein
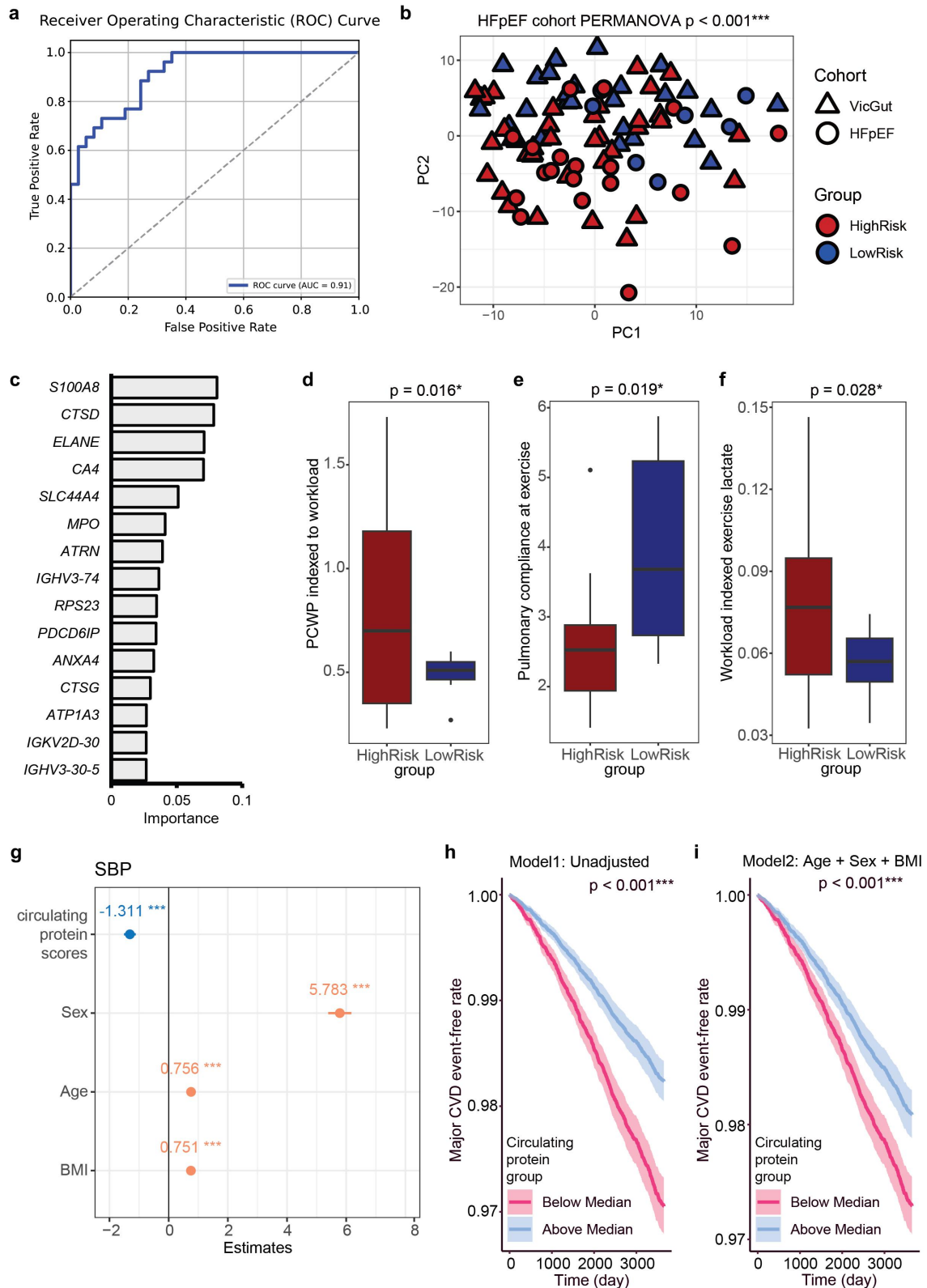
Figure 4. Machine learning model showing the relationship between human proteins and cardiovascular risk. A. Receiver operating characteristic (ROC) plot illustrating the results of 5-fold cross-validation of the constructed random forest model based on differentially expressed human proteins. B. PCA dimensionality reduction visualization of prediction results for heart failure with preserved ejection fraction (HFpEF) patients using the trained human protein-based random forest model. C. Presentation of the top 15 signature human proteins with the highest feature importance in the random forest model. D-F. Comparison of D. PCWP indexed to workload and E. Pulmonary compliance at exercise F. Workload indexed exercise lactate in HFpEF patients predicted as high-risk and

signature was associated with lower systolic BP and long-term cardiovascular risk at the population level. This suggests fecal metaproteomics data may be able to predict long-term cardiovascular outcomes and that several mechanisms that lead to these outcomes may develop in the gut. These could include, for example, a decrease in SCFA production leading to the breakdown of the gut epithelial barrier, passage of microbial metabolites and/or components to the intestinal tissue and activation of inflammatory pathways that contribute to the development of CVD.[15,73]

Unsupervised learning based on gut lumen human protein expression identified two distinct groups of individuals. Proteins highly expressed in the low-risk group primarily relate to pancreatic secretion and glycolysis/gluconeogenesis, which may reflect better glucose control. While all participants from the VicGut were healthy and a previous diagnosis of type 1 or type 2 diabetes was an exclusion criterion, we did not have fasting glucose information available to add to our model. Moreover, the low-risk group had an underrepresentation of proteins associated with inflammation (e.g., calprotectin, LCN2 and other neutrophil extracellular trap formation proteins) and gut epithelial barrier degradation (e.g., MUC13). These differences in protein expression profiles could be influenced by dietary patterns such as fiber or potassium intake, suggesting a link between diet-related metabolic alterations and subsequent effects on the gut microenvironment and cardiovascular health. This may provide novel insights into the contribution of diet via the gut microbiome to the development of CVD.

Consistently with their lower night systolic BP, which is a risk factor for cardiovascular mortality,[74] the low-risk group had an underrepresentation of proteins associated with the conversion of angiotensinogen to angiotensin I (CTSD, CTSG) and angiotensin I to angiotensin II (CTSG),[75] as part of the traditional arm of the RAS. The significant positive correlation between CTSD and nighttime pulse pressure suggests intestinal CTSD could play a role in systemic BP regulation. It remains unclear, however, whether these proteins are involved in the systemic or only intestinal RAS.

Studies in recent years showed that the host's dietary intake significantly influences the composition of the gut microbiome and their production of secondary metabolites.[10] Indeed, we observed that the gut microbiome in the low-risk group exhibited a significantly higher capacity for producing SCFAs, which reduces BP and associated end-organ damage in animal models and untreated hypertensive patients by driving a gut-cardiorenal axis.[11,64] This involved, in particular, higher levels of pta, a critical enzyme that converts the precursors of acetate and propionate into their final forms.

The metaproteomic data allowed us to predict the severity of phenotypes of patients with HFpEF. For instance, our prediction model based on both human and microbial proteins successfully categorized HFpEF patients into high- and low-risk groups, with significant differences in PCWP, lactate concentration and pulmonary compliance during exercise, which has implications for clinical outcomes associated with aerobic capacity and survival.[76,77] The same proteins, when detected in the systemic circulation, were associated with significantly higher BP and long-term risk of MACE in a population-based cohort from European ancestry. Collectively, this suggests that the metaproteomic changes observed in healthy individuals may predict long-term cardiovascular outcomes. Some of the mechanisms we identified, such as SCFA production, angiotensin production and inflammation, may be involved.

We acknowledge this study had limitations. Firstly, we did not establish causality in this study. However, there are no *in vitro* or *in vivo* models where we could emulate the combination

---

low-risk. Sample size $n = 26$.G. Age, BMI, sex adjusted regression model between SBP and standardized circulating protein scores derived from differentially expressed gut lumen human protein between high-risk and low-risk group. H.-I. Two 10-year follow-up time-to-event analysis H. unadjusted ($p < 0.001$***) and I. age, sex BMI adjusted ($p < 0.001$***) showing the incident rate of major CVD events in two groups based on circulating protein scores calculated. Sample size $n = 26$ for the HFpEF cohort and $n = 34,311$ for the UK biobank dataset.

of human and microbial proteins and their systemic effect on the host, as observed here. Secondly, the mapping of microbial proteins identified remains poorly understood, highlighting the need for further investigations and potential reanalysis of this dataset once better tools are available. This challenge partly arises because many gut microbial species have only been characterized recently by non-culture-dependent methods such as metagenomics. Thirdly, the sample size of our metaproteomics dataset, while unique in a CVD setting and larger than previous studies using metaproteome techniques to study human gut environments, was still relatively small. While we validated the findings using plasma proteomics, larger metaproteomics cohorts will be needed to validate the current findings, and other CVD risk factors, such as fasting glucose and cholesterol, could be included.

In conclusion, traditional cardiovascular risk factors could not clearly distinguish human gut lumen proteins nor gut microbial protein expression profiles in our dataset. By introducing machine learning models, we successfully identified two distinct groups with significantly different cardiovascular risk factors based on their unique protein expression profiles. Besides uncovering molecular mechanisms associated with these low- and high-risk groups, our findings unveil some of the intricate interplay between diet, gut microenvironment, gut microbiome, and the cardiovascular system that result in differential CVD risk, including long-term risk. Our results underscore the complexity of these interactions, as metaproteomic changes could not be solely attributed to a single factor. This highlights the necessity for applying machine learning models capable of capturing the complicated relationships between risk factors and phenotypes compared to conventional statistical methods.

Highlights

- Fecal metaproteomic analyses identified human intestinal and microbial proteins associated with low- and high-risk cardiovascular phenotypes in otherwise healthy individuals.

- Low-risk individuals had 9 mmHg lower night blood pressure and expression of angiotensin-converting enzymes (i.e. CTSD, CTSG) and inflammatory proteins (e.g. LCN2, S100A8/calprotectin), and higher intake of dietary potassium and fiber in males, expression of proteins associated with pancreatic secretion and glycolysis/gluconeogenesis, and a gut microbiome with a higher capacity for short-chain fatty acid production relative to high-risk individuals.

- Both gut microbial and human proteins within the high-risk healthy individuals could predict the severity of phenotypes in patients with heart failure with preserved ejection fraction (HFpEF) and long-term risk of a major cardiovascular event in a large population-based cohort.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Chaoran Yang http://orcid.org/0000-0001-6639-8136
Leticia Camargo Tavares http://orcid.org/0000-0001-7278-0999
Francine Z. Marques http://orcid.org/0000-0003-4920-9991

Author Contribution

Conceived and designed the analysis: C.Y., F.Z.M. Collected the data: H.C.L., J.R.S., S.Y., M.J.C., D.M.K., R.B.S. Contributed data or analysis tools: L.C.T. R.V.R., A.L.B., G. A.H. Performed the analysis: C.Y., L.C.T. Wrote the paper: C. Y., F.Z.M. All authors reviewed the results and approved the final version of the manuscript. Secured funding: F.Z.M.

Data availability statement

This research has been conducted using the UK Biobank Resource under Application Number 86,879. This work uses data provided by patients and collected by the NHS as part of their care and support. All original data used in this article is available upon request.

Nonstandard abbreviations and acronyms

Body mass index	BMI
Blood pressure	BP
False discovery rate	FDR
Heart failure with preserved ejection fraction	HFpEF
Kilobase	KB
Principal component	PC
Principal component analysis	PCA
Receiver Operating Characteristic	ROC
Short-chain fatty acids	SCFAs
Sodium dodecyl sulphate	SDS
Transcription start site	TSS

References

1. Afshin A, Sur PJ, Fay KA, Cornaby L, Ferrara G, Salama JS, Mullany EC, Abate KH, Abbafati C, Abebe Z, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2019;393 (10184):1958–1972. doi:10.1016/S0140-6736(19)30041-8

2. Murray CJL, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, Abd-Allah F, Abdelalim A, Abdollahi M, Abdollahpour I, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet. 2020;396 (10258):1223–1249. doi:10.1016/S0140-6736(20)30752-2

3. Ndanuko RN, Tapsell LC, Charlton KE, Neale EP, Batterham MJ. Dietary patterns and blood pressure in adults: a systematic review and meta-analysis of randomized controlled trials. Adv Nutr. 2016;7(1):76–89. doi:10.3945/an.115.009753.

4. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, Ramirez A, Schlaich M, Stergiou GS, Tomaszewski M, et al. 2020 international society of hypertension global hypertension practice guidelines. Hypertension. 2020;75(6):1334–1357. doi:10.1161/ HYPERTENSIONAHA.120.15026.

5. O'Donnell JA, Zheng T, Meric G, Marques FZ. The gut microbiome and hypertension. Nat Rev Nephrol. 2023;19(3):153–167. doi:10.1038/s41581-022-00654-0.

6. Tang WHW, Li DY, Hazen SL. Dietary metabolism, the gut microbiome, and heart failure. Nat Rev Cardiol. 2019;16(3):137–154. doi:10.1038/s41569-018-0108-7.

7. Witkowski M, Weeks TL, Hazen SL. Gut microbiota and cardiovascular disease. Circ Res. 2020;127 (4):553–570. doi:10.1161/CIRCRESAHA.120.316242.

8. Ferreira SR, Lerario DD, Gimeno SG, Sanudo A. Obesity and central adiposity in Japanese immigrants: role of the Western dietary pattern. J Epidemiol. 2002;12(6):431–438. doi:10.2188/jea.12.431.

9. Krautkramer KA, Kreznar JH, Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME, Keller MP, Attie AD, Rey FE, Denu JM. Diet-Microbiota interactions mediate global epigenetic programming in multiple host tissues. Mol Cell. 2016;64(5):982–992. doi:10.1016/j. molcel.2016.10.025.

10. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014;505 (7484):559–563. doi:10.1038/nature12820.

11. Jama HA, Rhys-Jones, D, Nakai, M, Yao, CK, Climie, RE, Sata, Y, Anderson, D, Creek, DJ, Head, GA, Kaye, DM, et al. Prebiotic intervention with HAMSAB in untreated essential hypertensive patients assessed in a phase II randomized trial. Nat Cardiovasc Res. 2023;2(1):35–43.

12. Xiong W, Abraham PE, Li Z, Pan C, Hettich RL. Microbial metaproteomics for characterizing the range of metabolic functions and activities of human gut microbiota. Proteomics. 2015;15(20):3424–3438. doi:10.1002/pmic.201400571.

13. Kleiner M, Thorson, E, Sharp, CE, Dong, X, Liu, D, Li, C, Strous, M. Assessing species biomass contributions in microbial communities via metaproteomics. Nat Commun. 2017;8(1).

14. Beale AL, O'Donnell JA, Nakai ME, Nanayakkara S, Vizi D, Carter K, Dean E, Ribeiro RV, Yiallourou S, Carrington MJ, et al. The gut microbiome of heart failure with preserved ejection fraction. J Am Heart Assoc. 2021;10(13). doi:10.1161/JAHA.120.020654.

15. Nakai M, Ribeiro RV, Stevens BR, Gill P, Muralitharan RR, Yiallourou S, Muir J, Carrington M, Head GA, Kaye DM, et al. Essential hypertension is associated with changes in gut microbial metabolic pathways: a multisite analysis of ambulatory blood pressure. Hypertension. 2021;78(3):804–815. doi:10.1161/HYPERTENSIONAHA.121.17288.

16. Parati G, Stergiou G, O'Brien E, Asmar R, Beilin L, Bilo G, Clement D, de la Sierra A, de Leeuw P, Dolan E, et al. European society of hypertension practice guidelines for ambulatory blood pressure monitoring. J Hypertens. 2014;32(7):1359–1366. doi:10.1097/HJH.0000000000000221.

17. Qiu C, Coughlin KB, Frederick IO, Sorensen TK, Williams MA. Dietary fiber intake in early pregnancy and risk of subsequent preeclampsia. Am J Hypertens. 2008;21(8):903–909. doi:10.1038/ajh.2008.209.

18. Zealand FSAN. AUSNUT 2007 - Australian food, supplement and nutrient database for estimation of population nutrient intakes. Canberra: food standards Australia New Zealand. Canberra: Food Standards Australia New Zealand; 2008.

19. Zealand FSAN. NUTTAB 2010 - Australian food composition tables. Canberra: Food Standards Australia New Zealand; 2011.

20. Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L. Carbohydrate quality and human health: a series of systematic reviews and meta-analyses. Lancet. 2019;393(10170):434–445. doi:10.1016/S0140-6736(18)31809-9.

21. Hailemariam M, Eguez RV, Singh H, Bekele S, Ameni G, Pieper R, Yu Y. S-Trap. an ultrafast sample-preparation approach for shotgun proteomics. J Proteome Res. 2018;17(9):2917–2924. doi:10.1021/acs.jproteome.8b00505.

22. Liao B, Ning Z, Cheng K, Zhang X, Li L, Mayne J, Figeys D. iMetalab 1.0: a web platform for metaproteomics data analysis. Bioinformatics. 2018;34(22):3954–3956. doi:10.1093/bioinformatics/bty466.

23. Tang J, Wang Y, Fu J, Zhou Y, Luo Y, Zhang Y, Li B, Yang Q, Xue W, Lou Y, et al. A critical assessment of the feature selection methods used for biomarker discovery in current metaproteomics studies. Briefings Bioinf. 2020;21(4):1378–1390. doi:10.1093/bib/bbz061.

24. Tang J, Mou M, Wang Y, Luo Y, Zhu F. MetaFS: performance assessment of biomarker discovery in metaproteomics. Briefings Bioinf. 2021;22(3). doi:10.1093/bib/bbaa105.

25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47–e47. doi:10.1093/nar/gkv007.

26. Kuleshov MV, Jones, MR, Rouillard, AD, Fernandez, NF, Duan, Q, Wang, Z, Koplev, S, Jenkins, SL, Jagodnik, KA, Lachmann, A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44(W1):W90–W97.

27. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607–D613. doi:10.1093/nar/gky1131.

28. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. 2008;9(1):559. doi:10.1186/1471-2105-9-559.

29. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576–589. doi:10.1016/j.molcel.2010.05.004.

30. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Med. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779.

31. Zheng T, Camargo Tavares L, D'Amato M, Marques FZ. Constipation is associated with an increased risk of major adverse cardiac events in a UK population. Am J Physiol-Heart Circulatory Physiol. 2024;327(4):H956–H964. doi:10.1152/ajpheart.00519.2024.

32. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox Model. Statistics for Biology and Health. New York (US); 2000.

33. Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust survival curves for confounders. Stat Med. 2023;42(10):1461–1479. doi:10.1002/sim.9681.

34. Ding L, Goossens GH, Oligschlaeger Y, Houben T, Blaak EE, Shiri-Sverdlov R. Plasma cathepsin D activity is negatively associated with hepatic insulin sensitivity in overweight and obese humans. Diabetologia. 2020;63(2):374–384. doi:10.1007/s00125-019-05025-2.

35. Duc HN, Oh H, Yoon IM, Kim M-S. Association between levels of thiamine intake, diabetes, cardiovascular diseases and depression in Korea: a national cross-sectional study. J Nutr Sci. 2021;10:10. doi:10.1017/jns.2021.23.

36. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. Dynamics of human gut microbiota and short-chain fatty acids in response to

dietary interventions with three fermentable fibers. mBio. 2019;10(1)p. e02566–18. doi:10.1128/mBio.02566-18.

37. Gao S. Cathepsin G and Its role in inflammation and autoimmune diseases. Archives of Rheumatology, Archiv Rheumatol. 2018;33(4):498–504. doi:10.5606/ArchRheumatol.2018.6595.

38. Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, Zhu D, Koya JB, Wei L, Li J, et al. Microbiota in health and diseases. Signal transduction and targeted therapy. Signal Transduct Targeted Ther. 2022;7(1):7(1. doi:10.1038/s41392-022-00974-4.

39. Kazemian N, Mahmoudi M, Halperin F, Wu JC, Pakpour S. Gut microbiota and cardiovascular disease: opportunities and challenges. Microbiome. 2020;8(1):8(1. doi:10.1186/s40168-020-00821-0.

40. Tang WHW, Kitai T, Hazen SL. Gut microbiota in cardiovascular health and disease. 2017;120 (7):1183–1196. doi:10.1161/CIRCRESAHA.117.309715.

41. Jama HA, Snelson M, Schutte AE, Muir J, Marques FZ. Recommendations for the use of dietary fiber to improve blood pressure control. Hypertension. 2024;81 (7):1450–1459. doi:10.1161/HYPERTENSIONAHA.123. 22575.

42. Sundboll J, Szépligeti SK, Adelborg K, Szentkúti P, Gregersen H, Sørensen HT. Constipation and risk of cardiovascular diseases: a Danish population-based matched cohort study. BMJ Open. 2020;10(9): e037080. doi:10.1136/bmjopen-2020-037080.

43. Judkins CP, Wang Y, Jelinic M, Bobik A, Vinh A, Sobey CG, Drummond GR. Association of constipation with increased risk of hypertension and cardiovascular events in elderly Australian patients. Sci Rep. 2023;13 (1):10943. doi:10.1038/s41598-023-38068-y.

44. Zhuo JL, Soleimani M, Li XC. New insights into the critical importance of intratubular Na+/H+ exchanger 3 and its potential therapeutic implications in hypertension. Curr Hypertens Rep. 2021;23(6). doi:10. 1007/s11906-021-01152-7.

45. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, Ntritsos G, Dimou N, Cabrera CP, Karaman I, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet. 2018;50 (10):1412–1425. doi:10.1038/s41588-018-0205-x.

46. Wang Y, Wang M, Su H, Song J, Ren M, Hu P, Liu G, Tong X. SERCA2 dysfunction triggers hypertension by interrupting mitochondrial homeostasis and provoking oxidative stress. Free Radical Biol And Med. 2024;212:284–294. doi:10.1016/j.freeradbiomed.2023. 12.044.

47. Tani T, Ohsumi J, Mita K, Takiguchi Y. Identification of a novel class of elastase isozyme, human pancreatic elastase III, by cDNA and genomic gene cloning. J Biol Chem. 1988;263(3):1231–1239. doi:10.1016/S0021-9258(19)57291-X.

48. McAlpine W, Wang K-W, Choi JH, San Miguel M, McAlpine SG, Russell J, Ludwig S, Li X, Tang M, Zhan X, et al. The class I myosin MYO1D binds to lipid and protects against colitis. Disease Model & Mechanisms. 2018;11(9):11(9. doi:10.1242/dmm.035923.

49. Jukic A, Bakiri L, Wagner EF, Tilg H, Adolph TE. Calprotectin: from biomarker to biological function. Gut, Gut. 2021;70(10):1978–1988. doi:10.1136/gutjnl-2021-324855.

50. Kaiser T, Langhorst J, Wittkowski H, Becker K, Friedrich AW, Rueffer A, Dobos GJ, Roth J, Foell D. Faecal S100A12 as a non-invasive marker distinguishing inflammatory bowel disease from irritable bowel syndrome. Gut, Gut. 2007;56(12):1706–1713. doi:10.1136/gut.2006.113431.

51. Segui-Perez C, Stapels DAC, Ma Z, Su J, Passchier E, Westendorp B, Wubbolts RW, Wu W, van Putten JPM, Strijbis K. MUC13 negatively regulates tight junction proteins and intestinal epithelial barrier integrity via protein kinase C. J Cell Sci. 2024;137(5):137(5. doi:10.1242/jcs.261468.

52. Furuta GT, Nieuwenhuis EES, Karhausen J, Gleich G, Blumberg RS, Lee JJ, Ackerman SJ. Eosinophils alter colonic epithelial barrier function: role for major basic protein. Am J Physiol-Gastrointestinal Liver Physiol. 2005;289(5):G890–G897. doi:10.1152/ajpgi.00015.2005 .

53. Fukuda N, Satoh C, Hu W-Y, Soma M, Kubo A, Kishioka H, Watanabe Y, Izumi Y, Kanmatsuse K. Production of angiotensin II by homogeneous cultures of vascular smooth muscle cells from spontaneously hypertensive rats. Arterioscler, Thromb, And Vasc Biol. 1999;19(5):1210–1217. doi:10.1161/01.ATV.19.5. 1210.

54. Tonnesen MG, Klempner MS, Austen KF, Wintroub BU. Identification of a human neutrophil angiotension ii-generating protease as cathepsin G. J Clin Investigation. 1982;69(1):25–30. doi:10.1172/JCI110437.

55. Hoes MF, Tromp J, Ouwerkerk W, Bomer N, Oberdorf-Maass SU, Samani NJ, Ng LL, Lang CC, van der Harst P, Hillege H, et al. The role of cathepsin D in the pathophysiology of heart failure and its potentially beneficial properties: a translational approach. Eur J Heart Fail. 2020;22(11):2102–2111. doi:10.1002/ejhf. 1674.

56. Jahanyar J, Youker KA, Loebe M, Assad-Kottner C, Koerner MM, Torre-Amione G, Noon GP. Mast cell-derived cathepsin g: a possible role in the adverse remodeling of the failing human heart. J Surg Res. 2007;140(2):199–203. doi:10.1016/j.jss.2007.02.040.

57. Marques FZ, Prestes PR, Byars SG, Ritchie SC, Würtz P, Patel SK, Booth SA, Rana I, Minoda Y, Berzins SP, et al. Experimental and human evidence for lipocalin-2 (neutrophil gelatinase-associated lipocalin [NGAL]) in the development of cardiac hypertrophy and heart failure.

J Am Heart Assoc. 2017;6(6):e005971. doi:10.1161/JAHA.117.005971.

58. Flo TH, Smith KD, Sato S, Rodriguez DJ, Holmes MA, Strong RK, Akira S, Aderem A. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestering iron. Nature. 2004;432(7019):917–921. doi:10.1038/nature03104.

59. Toyonaga T, Matsuura M, Mori K, Honzawa Y, Minami N, Yamada S, Kobayashi T, Hibi T, Nakase H. Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages. Sci Rep. 2016;6(1):35014. doi:10.1038/srep35014.

60. ProspectiveStudiesCollaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet. 2002;360 (9349):1903–1913. doi:10.1016/S0140-6736(02)11911-8

61. Lalwani A, Warren J, Liuwantara D, Hawthorne WJ, O'Connell PJ, Gonzalez FJ, Stokes RA, Chen J, Laybutt DR, Craig ME, et al. β cell hypoxia-inducible factor-1α is required for the prevention of type 1 diabetes. Cell Rep. 2019;27(8):2370–2384.e6. doi:10.1016/j.celrep.2019.04.086.

62. Cheng K, Ho K, Stokes R, Scott C, Lau SM, Hawthorne WJ, O'Connell PJ, Loudovaris T, Kay TW, Kulkarni RN, et al. Hypoxia-inducible factor-1α regulates β cell function in mouse and human islets. J Clin Investigation. 2010;120(6):2171–2183. doi:10.1172/JCI35846.

63. Cornwell JD, McDermott JC. MEF2 in cardiac hypertrophy in response to hypertension. Trends in Cardiovasc Med. 2023;33(4):204–212. doi:10.1016/j.tcm.2022.01.002

64. Kaye DM, Shihata WA, Jama HA, Tsyganov K, Ziemann M, Kiriazis H, Horlock D, Vijay A, Giam B, Vinh A, et al. Deficiency of prebiotic fiber and insufficient signaling through gut metabolite-sensing receptors leads to cardiovascular disease. Circulation, Circulation. 2020;141(17):1393–1403. doi:10.1161/CIRCULATIONAHA.119.043081.

65. Alaei-Shahmiri F, Soares MJ, Zhao Y, Sherriff J. The impact of thiamine supplementation on blood pressure, serum lipids and C-reactive protein in individuals with hyperglycemia: a randomised, double-blind cross-over trial. Diabetes & Metabolic Syndr: Clin Res & Rev. 2015;9(4):213–217. doi:10.1016/j.dsx.2015.04.014.

66. Liu M, Zhou C, Zhang Z, Li Q, He P, Zhang Y, Li H, Liu C, Qin X. Inverse association between riboflavin intake and new-onset hypertension. Hypertension, Hypertension. 2020;76(6):1709–1716. doi:10.1161/HYPERTENSIONAHA.120.16211.

67. Bays HE, Rader DJ. Does nicotinic acid (niacin) lower blood pressure? Int J Clin Pract. 2009;63(1):151–159. doi:10.1111/j.1742-1241.2008.01934.x.

68. Lal KJ, Dakshinamurti K, Thliveris J. The effect of vitamin B6 on the systolic blood pressure of rats in various animal models of hypertension. J Hypertens. 1996;14 (3):355–363. doi:10.1097/00004872-199603000-00013.

69. Allen A, Gau D, Roy P. The role of profilin-1 in cardiovascular diseases. J Cell Sci. 2021;134(9). doi:10.1242/jcs.249060.

70. Eisman AS, Shah, RV, Dhakal, BP, Pappagianopoulos, PP, Wooster, L, Bailey, C, Cunningham, TF, Hardin, KM, Baggish, BL, Ho, JE, et al. Pulmonary capillary wedge pressure patterns during exercise predict exercise capacity and incident heart failure. Circulation: Heart Failure; 2018;11(5).

71. Obokata M, Olson TP, Reddy YNV, Melenovsky V, Kane GC, Borlaug BA. Haemodynamics, dyspnoea, and pulmonary reserve in heart failure with preserved ejection fraction. Eur Heart J. 2018;39(30):2810–2821. doi:10.1093/eurheartj/ehy268.

72. Nayor M, Houstis NE, Namasivayam M, Rouvina J, Hardin C, Shah RV, Ho JE, Malhotra R, Lewis GD. Impaired exercise tolerance in heart failure with preserved ejection fraction. JACC: Heart Failure, JACC: Heart Fail. 2020;8(8):605–617. doi:10.1016/j.jchf.2020.03.008.

73. Snelson M, Vanuytsel T, Marques FZ. Breaking the barrier: the role of gut epithelial permeability in the pathogenesis of hypertension. Curr Hypertens Rep. 2024;26(9):369–380. doi:10.1007/s11906-024-01307-2.

74. Staplin N, de la Sierra A, Ruilope LM, Emberson JR, Vinyoles E, Gorostidi M, Ruiz-Hurtado G, Segura J, Baigent C, Williams B. Relationship between clinic and ambulatory blood pressure and mortality: an observational cohort study in 59 124 patients. Lancet. 2023;401(10393):2041–2050. doi:10.1016/S0140-6736(23)00733-X.

75. Nehme A, Zouein FA, Deris Zayeri Z, Zibara K. An update on the tissue renin angiotensin system and its role in physiology and pathology. J Cardiovasc Devel And Disease. 2019;6(2):14. doi:10.3390/jcdd6020014.

76. Lewis GD, Murphy RM, Shah RV, Pappagianopoulos PP, Malhotra R, Bloch KD, Systrom DM, Semigran MJ. Pulmonary vascular response patterns during exercise in left ventricular systolic dysfunction predict exercise capacity and outcomes. Circulation: Heart Failure, Circ: Heart Fail. 2011;4(3):276–285. doi:10.1161/CIRCHEARTFAILURE.110.959437.

77. Lalande S, Cross TJ, Keller-Ross ML, Morris NR, Johnson BD, Taylor BJ. Exercise intolerance in heart failure: central role for the pulmonary system. Exercise And Sport Sci Rev. 2020;48(1):11–19. doi:10.1249/JES.0000000000000208.