# The landscape of plasma proteomic links to human organ imaging

## Running title: Pan-organ imaging proteomics

Zirui Fan[1], Julio Chirinos[2,3], Xiaochen Yang[4], Juan Shu[1], Yujue Li[4], Joan M. O'Brien[5,6], Walter Witschey[7], Daniel J. Rader[8,9], Ruben Gur[10,11], and Bingxin Zhao[1,12-16]*

[1]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA 19104, USA.

[2]Division of Cardiovascular Medicine, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA.

[3]University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.

[4]Department of Statistics, Purdue University, West Lafayette, IN 47907, USA.

[5]Scheie Eye Institute, University of Pennsylvania, Philadelphia, PA 19104, USA.

[6]Penn Medicine Center for Ophthalmic Genetics in Complex Diseases, Philadelphia, PA 19104, USA.

[7]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[8]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[9]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[10]Lifespan Brain Institute (LiBI), Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA 19104, USA

[11]Brain Behavior Laboratory, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[12]Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania Philadelphia, PA 19104, USA.

[13]Center for AI and Data Science for Integrated Diagnostics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

[14]Population Aging Research Center, University of Pennsylvania, Philadelphia, PA 19104, USA.

[15]Institute for Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA.

[16]Penn Center for Eye-Brain Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

*Corresponding author:

Bingxin Zhao
413 Academic Research Building
265 South 37th Street, Philadelphia, PA 19104
E-mail: bxzhao@upenn.edu Phone: (215) 898-8222

## Abstract

Plasma protein levels provide important insights into human disease, yet a comprehensive assessment of plasma proteomics across organs is lacking. Using large-scale multimodal data from the UK Biobank, we integrated plasma proteomics with organ imaging to map their phenotypic and genetic links, analyzing 2,923 proteins and 1,051 imaging traits across multiple organs. We uncovered 5,067 phenotypic protein-imaging associations, identifying both organ-specific and organ-shared proteomic relations, along with their enriched protein-protein interaction networks and biological pathways. By integrating external gene expression data, we observed that plasma proteins associated with the brain, liver, lung, pancreas, and spleen tended to be primarily produced in the corresponding organs, while proteins associated with the heart, body fat, and skeletal muscle were predominantly expressed in the liver. We also mapped key protein predictors of organ structures and showed the effective stratification capability of plasma protein-based prediction models. Furthermore, we identified 8,116 genetic-root putative causal links between proteins and imaging traits across multiple organs. Our study presents the most comprehensive pan-organ imaging proteomics map, bridging molecular and structural biology and offering a valuable resource to contextualize the complex roles of molecular pathways underlying plasma proteomics in organ systems.

## Introduction

Proteins, as the functional outputs of genes, play essential roles in biological mechanisms. Advances in high-throughput proteomics technologies, such as the Olink[1] and SomaScan[2] platforms now enable the profiling of thousands of plasma proteins in biobank-scale cohorts[3-9], providing unprecedented insights into health[4,10], aging[11-15], disease[16,17], and drug discovery[18]. Plasma proteins, originating from tissues across the body[13], perform context-specific molecular roles within each organ[19]. Mapping the locations of these proteomic links is essential to understanding the organ-specific pathways and mechanisms through which plasma proteins relate to health and clinical outcomes. For example, brain structural phenotypes have been shown to mediate the effects of plasma levels of BCAN, NCAN, and MOG on cognitive ability[20]. Infection-related proteins (such as PIK3CG, PACSIN2, and PRKCB) may contribute to neurodegeneration and Alzheimer's disease through region-specific brain volume loss[21]. However, a comprehensive, integrated view of plasma protein links across organs is lacking, limiting our ability to fully contextualize their roles in health and disease.

Medical imaging, such as magnetic resonance imaging (MRI), provides non-invasive measures of organ structure and function[22-26]. Imaging-derived phenotypes (IDPs), which capture variations in tissue composition, organ morphology, and functional activity/connectivity, have been extensively linked to physiological and pathological processes in a wide range of diseases, such as heart failure[27], chronic liver diseases[28], Alzheimer's disease[29], and glaucoma[30]. Integrating IDPs with plasma proteomics holds significant potential for advancing our understanding of organ-specific proteomic biology[31]. However, few large-scale studies have collected both multi-organ imaging and plasma protein data within the same cohort. Similar to many omics mapping strategies of complex traits and diseases[32-34], such data constraints have led existing studies to rely on separate cohorts for imaging and (prote)omics data[35], limiting analyses solely to genetics-driven associations[36,37]. Genetic-based, separate-cohort approaches are known to face challenges[35], including demographic mismatches between cohorts, which can introduce bias, and limited power to detect proteomic effects influenced by non-genetic factor, which limits the insight into mechanisms. Moreover, existing separate-cohort studies have

typically focused on single-organ or single-modality IDPs[21,36,37]. These limitations underscore the need for large-scale, single-cohort datasets that directly integrate multi-organ imaging and proteomic data to comprehensively investigate the biological roles of proteins in each organ[31].

Leveraging multi-organ imaging[26] and plasma protein[4] data from the UK Biobank (UKB), we conducted the largest pan-organ imaging proteomics analysis to date. We mapped proteomic links across 2,923 Olink plasma proteins and 1,051 IDPs, encompassing a wide range of organs and tissues, including the brain, heart, aorta, liver, kidney, lung, pancreas, spleen, body fat, and muscle composition (average *n* = 4,896 participants with both imaging and protein data). We developed an atlas of phenotypic protein-imaging associations, revealing organ-specific proteomic networks and enriched biological pathways. To investigate the origins of these proteomic associations, we integrated external gene expression data[38] to examine whether proteins may be produced in other organs and act remotely, or originated from the same organ. Additionally, we generated a chart for the predictive ability of plasma proteins on organ structure and function, identifying key protein profiles that strongly predict specific IDPs for each organ. Together with UKB genetic data (average *n* = 40,682 participants for imaging and 34,566 for protein after removing overlapping imaging subjects), we further evaluated potential genetic causal associations across organs. An overview of the study design is presented in **Figure 1**.

## RESULTS

### Overview of phenotypic protein-imaging associations

We investigated the phenotypic associations between 2,923 plasma proteins from the UKB pharma proteomics project (UKB-PPP; **Table S1**) and a diverse set of brain and body IDPs. These included 258 brain structural MRI (sMRI) traits, 432 diffusion MRI (dMRI) traits, 82 brain resting-state functional MRI (fMRI) traits, 82 cardiac MRI traits, 41 abdominal MRI traits, as well as 46 optical coherence tomography (OCT) and 110 fundus imaging traits of the eye (**Table S2**). For discovery, we used data from unrelated white British participants (average *n* = 4,383), with effects of age, sex, age-sex interaction,

genetic principal components, height, weight, and body mass index removed. Associations were replicated in an independent hold-out sample of white non-British individuals (average $n$ = 513; **Methods**). We identified 5,067 associations that were significant in the discovery sample after Bonferroni correction ($P < 1.63\times 10^{-8}$) and remained nominally significant ($P < 0.05$) in the replication sample with concordant effect signs (**Figs. 2A** and **S1**, **Table S3**). To assess the impact of disease status on protein-imaging associations, we performed sensitivity analyses by additionally adjusting for organ-related diseases (**Methods**). The results showed that protein effect sizes remained highly consistent between models with and without disease status adjustment, with a correlation exceeding 0.99 (**Fig. S2** and **Table S4**).

We found that plasma proteins associated with organ structure and function tended to exhibit high levels of interaction, with significant enrichment observed for all organs using the STRING protein-protein interaction (PPI) database[39] ($P < 7.31 \times 10^{-3}$**, Fig. 2B** and **Table S5**). Since proteins often function collaboratively within biological pathways, these findings suggest the presence of proteomic modules and associated network biology within each organ. Further pathway enrichment analysis identified 1,429 enriched biological pathways[40] (**Fig. 2C** and **Table S6**) after multiple testing adjustments (**Methods**). Some protein clusters and pathways had highly organ-specific associations, while others suggested cross-organ links. In the following sections, we highlight key proteins associated with each organ, along with the proteomic interaction networks and biological pathways they are involved in.

**Plasma proteomic links and enriched biological pathways in the human heart**

We observed numerous connections between plasma proteins and body IDPs, spanning the heart, liver, body fat, and kidney (**Fig. S3**). Specifically, we identified 208 proteins associated with cardiac MRI traits of the heart, particularly end-diastolic volume, end-systolic volume, and stroke volume of left/right ventricles (|β| range = (0.07, 0.26), $P < 1.52\times 10^{-8}$). Most (over 94%) protein-heart associations were negative, except few proteins such as IGFBP1, IGFBP2, leptin (LEP), and NTPROBNP. These positively associated proteins often serve as markers of adverse cardiac conditions. For example,

NTPROBNP is a well-known biomarker associated with increased cardiac filling pressures, the presence of heart failure as well as outcomes in established heart failure[41]. Additionally, LEP promotes oxidative stress, inflammation, and atherogenesis, linking elevated levels to cardiovascular diseases such as coronary artery disease, stroke, and diabetes-related complications[42,43]. Many of the identified heart-associated plasma proteins have well-known biological functions on the heart. For example, GDF15 may protect the heart by activating SMAD2/3 signaling to attenuate hypertrophy and preserve ventricular function under pathological conditions[44]. Plasma levels of ICAM1 and SELE were elevated in individuals with coronary heart disease (CHD) and carotid artery atherosclerosis, independently predicting the risk of both conditions, making them potential molecular markers for atherosclerosis and CHD development[45].

Moreover, proteins associated with the heart were enriched in various biological pathways. For example, 19 heart-associated proteins were significantly enriched in regulation of angiogenesis and vasculature development (multiple testing-adjusted[40] *P* range = ($1.53\times 10^{-7}$, $1.14\times 10^{-7}$), **Fig. 2D**). Many proteins were growth factors (such as PGF, ADM, ANGPT2) or growth factor receptors (such as KDR, ERBB2, TGFBR2, ACVRL1), which played key roles in regulating vascular development[46-49]. There are other heart-associated proteins that were not growth factors but were also critical in angiogenesis. For example, EPHA1 promotes tumor angiogenesis by regulating endothelial tubulogenesis and recruiting endothelial progenitor cells, with its inhibition reducing tumor angiogenesis and growth[50]. Given the complex and multifaceted role of angiogenesis in cardiovascular diseases, these identified proteins enriched in the angiogenesis pathways could serve as biomarkers of disease state or context-specific therapeutic targets[51,52]. The associated cardiac MRI traits provide valuable insights into their underlying biologic pathways and mechanisms. Furthermore, heart-associated proteins were also enriched in the movement of cells or organisms in response to chemical stimuli, such as taxis, chemotaxis, and positive regulation of leukocyte migration (adjusted *P* range = ($5.89\times 10^{-7}$, $3.06\times 10^{-7}$), **Figs. S4-S5**). A large proportion of proteins exclusively enriched in taxis and chemotaxis were chemokines and cytokines, including CXCL10, CXCL13, CCL22, and CCL27. These proteins recruit immune cells and play critical roles

in inflammatory, infectious and immune responses[53-56] and are promising therapeutic targets for multiple cancer types.

We identified 12 proteins associated with cardiac MRI traits of the aorta (|β| range = (0.07, 0.14), $P < 1.49\times 10^{-8}$, **Fig. S6**), primarily linked to the descending aorta. Of these, nine proteins overlapped with those associated with the heart, and most of the aorta-protein associations were also negative. Positive associations were only found with two proteins IGFBP1 and Renin (REN) (β range = (0.07, 0.08), $P < 1.19\times 10^{-8}$). REN, encoded by *REN*, is secreted by the kidney and initiates the renin-angiotensin-aldosterone system cascade, regulating blood pressure and volume. Dysregulation of this system contributes to cardiovascular and renal disorders, making renin an important therapeutic target[57].

**Shared proteomic associations and systemic processes in abdominal organs**

In addition to the heart, we found many associations between plasma proteins and IDPs of the liver, body fat, and kidney, with these four organs/tissues showing the highest proportion of overlapping associated proteins among all the organs examined (**Fig. 2A**). Plasma proteins shared by these organs widely participate in systemic processes such as metabolic regulation, inflammation, and vascular health, highlighting potential targets for understanding systemic diseases and multi-organ interactions.

We identified 270 proteins associated with the liver (|β| range = (0.08, 0.25), *P* range = ($1.61\times 10^{-8}$, $6.62\times 10^{-67}$)), primarily linked to liver volume, fat fraction, and liver iron-corrected T1 (a marker of inflammation and fibrosis). Proteins associated with liver volume differed from those linked to fat fraction, inflammation, and fibrosis, with limited overlap, suggesting that these processes are driven by distinct molecular mechanisms. Liver-associated proteins enriched in many biological pathways. In addition to those shared with heart, such as chemotaxis and leukocyte migration, proteins associated with the liver were also enriched in the regulation of plasma lipoprotein particle levels (adjusted $P = 5.36\times 10^{-5}$, **Fig. 2E**) and cholesterol/sterol transport (adjusted *P* = 0.02), underscoring the liver's central role in lipid metabolism. Many of the involved proteins, such as LCAT, FGF21, APOA2, and FURIN, are produced by the liver and are key

regulators of lipid metabolism. For example, LCAT is essential for cholesterol transport, promotes the formation of high-density lipoprotein, potentially reducing atherosclerosis[58]. Reduced LCAT activity has also been observed in individuals with liver disease[59]. FGF21 is critical for maintaining energy balance, regulating glucose and lipid metabolism, and shows promise in treating obesity, type 2 diabetes, and non-alcoholic steatohepatitis[60]. Additionally, liver-associated proteins were uniquely enriched in pathways related to viral processes and the viral life cycle (**Fig. S7**). Many of these proteins act as viral entry receptors or facilitate viral invasion, including ICAM1, a receptor for human rhinovirus[61]; ACE2 and NRP1, which serve as receptors for SARS-CoV-2[62,63], and FURIN, which cleaves the SARS-CoV-2 spike protein to enable host cell entry[64]. CTSL further supports viral invasion by activating SARS-CoV spike protein-mediated membrane fusion under acidic conditions[65].

We identified 214 proteins associated with body fat (|β| range = (0.08, 0.37), *P* range = ($1.52\times10^{-8}$, $8.90\times10^{-160}$)), predominantly linked to visceral adipose tissue (VAT) volume and total trunk fat volume. Similar with the liver-associated proteins, body fat-associated proteins were also enriched in biological pathways related to lipid metabolism, including the regulation of plasma lipoprotein particle levels (adjusted $P = 3.68\times10^{-7}$), with substantial overlap with liver. Beyond the liver-derived proteins mentioned earlier, additional examples include adiponectin (ADIPOQ) and LPL. ADIPOQ, primarily produced by adipocytes, promotes fatty acid oxidation and improves insulin sensitivity[66]. LPL hydrolyzes triglycerides in chylomicrons and very low-density lipoprotein into fatty acids for cellular uptake, with reduced activity observed in poorly controlled diabetes[67]. These findings suggest the crucial roles of the liver and body fat in systemic lipid regulation and their metabolic interactions with other tissues and diseases. Furthermore, the kidney had the highest number of associations with plasma proteins among all organs, with 439 proteins linked to kidney parenchyma and kidney volume (|β| range = (0.08, 0.38), *P* range = ($1.57\times10^{-8}$, $4.33\times10^{-167}$)). Nearly all (98.98%, 1,751 out of 1,769 associated pairs) these associations were negative. In addition, 11 proteins were associated with the lung (|β| range = (0.08, 0.17), $P < 1.53\times10^{-8}$), many of which were shared with the heart, liver, and body fat as well (**Fig. 2A**).

## Organ-specific proteomic associations with the pancreas and spleen

In contrast to the abdominal organs and tissues with many shared proteomic links, the majority of proteins associated with pancreas and spleen were specific to these two organs (**Fig. 2A**). We observed 26 proteins associated with pancreas fat fraction and volume (|β| range = (0.08, 0.34), *P* range = ($8.95\times 10^{-9}$, $1.35\times 10^{-136}$), **Fig. S8**). These proteins had a high degree of interaction, with a mean PPI score of 0.22 (enrichment $P < 2.2\times 10^{-16}$), and were enriched in digestion-related biological pathways (**Fig. 2F**). Enriched proteins include pancreatic amylases AMY2A and AMY2B, which facilitate carbohydrate breakdown, and lipases such as PNLIP and PNLIPRP1, which digest triglycerides into free fatty acids and monoglycerides. Additionally, proteins such as CTRB1, CTRL, and CPA1 were proteases involved in protein digestion. In addition to digestion, pancreas-associated proteins also enriched in biological pathways related to insulin response. Proteins such as IGFBP2, GHR, and PLA2G1B have been linked to insulin sensitivity and implicated in the regulation of insulin signaling in preclinical studies and animal models[68-70]. These findings highlight the critical roles of pancreas in producing enzymes required for nutrient breakdown and regulating systemic insulin sensitivity and metabolic responses through these pancreas-associated proteins.

Moreover, 113 proteins associated with spleen volume (|β| range = (0.08, 0.39), *P* range = ($1.61\times 10^{-8}$, $4.64\times 10^{-145}$)). These proteins had enriched interaction (mean PPI score = 0.12, enrichment $P < 2.2\times 10^{-16}$) and pathway enrichment analysis underscored the spleen's role in modulating both innate and adaptive immunity. Spleen-associated proteins were significantly enriched in the regulation of leukocyte-mediated immunity (adjusted $P = 1.43\times 10^{-16}$) and lymphocyte-mediated immunity (adjusted $P = 6.67\times 10^{-15}$, **Fig. 2G**). These proteins play diverse roles in immune activation, migration, and function, particularly involving T cells, B cells, and natural killer (NK) cells. For example, LAG3 functions as an immune checkpoint to prevent T cell overactivation and maintain immune homeostasis[71]. IL12B and TNF were pro-inflammatory cytokines that regulate T cell and NK cell activity[72,73], while ICAM1 mediate leukocyte adhesion to endothelial cells[74]. In addition, CR1 and CR2 modulate B cell activation and are implicated

in autoimmune diseases such as systemic lupus erythematosus and rheumatoid arthritis[75,76].

**Plasma proteomic insights into the brain and eye**

We identified 33 plasma proteins that were associated with global and regional brain volumes measured by brain sMRI. These proteins were widely enriched in biological pathways that were crucial to neural development and nervous system function, such as axonogenesis, axon guidance, synapse maturation, and neuron projection guidance (adjusted $P < 0.046$). Most of these proteins were linked to total brain volume, white matter volume, and gray matter volume ($|\beta|$ range = (0.08, 0.24), $P < 1.55\times 10^{-8}$). Some of them were also linked to volumes of localized regions. For example, increased level of NCAN and BCAN were associated with increased volumes of the limbic regions, such as frontal pole, insular cortex, hippocampus, and amygdala ($\beta$ range = (0.11, 0.19), $P$ range = ($1.27\times 10^{-14}$, $6.75\times 10^{-40}$)). NCAN and BCAN were highly interacted proteins (PPI score = 0.96, **Fig. S9**). They are both chondroitin sulfate proteoglycans, key components of the extracellular matrix and are enriched in critical pathways such as the perineuronal net, perisynaptic extracellular matrix, and synapse-associated extracellular matrix (adjusted $P < 0.03$). The links between NCAN and BCAN and brain volumes were reported before and were associated with cognitive ability[20]. Additionally, elevated level of MOG and SLITRK1 were also associated with higher volumes of limbic system for emotion regulation, such as the frontal pole, thalamus, insular cortex, hippocampus, and amygdala ($\beta$ range = (0.11, 0.18), $P$ range = ($4.55\times 10^{-10}$, $4.18\times 10^{-32}$)). SLITRK1 is highly expressed in the brain and regulates excitatory synapse formation and neural connectivity, particularly in hippocampal neurons[77]. It was also implicated in the pathogenesis of Tourette syndrome[78,79] and has been found in schizophrenia patients[80]. MOG was expressed exclusively in central nervous system (CNS) myelin[81] and can cause myelin oligodendrocyte glycoprotein antibody-associated disease (MOGAD) when it was attacked by the immune system. Brain structural and functional changes have been reported in individuals with MOGAD, including gray matter atrophy of frontal and temporal lobe, insula, thalamus, and hippocampus[82]. Most of the observed associations between brain volumes and plasma proteins were positive, with only few exceptions such as SOST,

FABP3, GDF15, and LEP, whose increased levels in plasma associated with decreased volume of certain brain regions.

Six proteins were associated with white matter microstructural measures of brain dMRI, with two (BACN and NCAN) demonstrating overlap with sMRI. Additional proteins include AHSP, GFAP, OMG, and TF (|β| range = (0.08, 0.13), $P < 1.45\times 10^{-8}$, **Fig. S10**). Notably, over 70% of the dMRI-protein associations (26 out of 37) involved GFAP, which showed broad associations with multiple white matter tracts (|β| range = (0.08, 0.13), $P < 1.37\times 10^{-8}$). GFAP, a key astrocyte protein, regulates CNS homeostasis and astrocyte responses to stress and neurological disease[83]. Blood GFAP levels serve as a sensitive biomarker for CNS injuries and diseases, aiding in diagnosis, severity assessment, and prognostication in conditions such as traumatic brain injury and multiple sclerosis[84]. These dMRI-associated proteins were enriched in pathways that were critical for the formation, maintenance, and functional integrity of brain white matter, including glial cell differentiation, positive regulation of neurogenesis, nervous system development, and neuroblast proliferation (adjusted $P < 0.046$, **Fig. S11**). No significant associations were found between plasma proteins and brain resting state fMRI traits with the conservative Bonferroni correction. With a less stringent false discovery rate correction ($P < 9.33\times 10^{-4}$), several plasma proteins were identified as likely being associated with fMRI functional activity traits (|β| range = (0.04, 0.07), $P < 9.11\times 10^{-4}$, **Fig. S12**), with IGFBP2 emerging as a key protein linked to multiple networks (|β| range = (0.04, 0.06), $P < 5.89\times 10^{-4}$).

Eye OCT measures of the retina, a component of the CNS closely connected to the brain[85], showed substantial associations with plasma proteins. Specifically, 93 plasma proteins linked to the average thickness of ganglion cell-inner plexiform layer (GCIPL, |β| range = (0.08, 0.12), $P < 1.52\times 10^{-8}$) and the retinal nerve fiber layer (RNFL, |β| range = (0.08, 0.13), $P < 1.54\times 10^{-8}$). Some of these proteins have been highlighted in previous eye research. For example, SOD1 deficiency leads to oxidative stress, resulting in retinal pigment epithelial damage and features of age-related macular degeneration[86]. Similarly, CA1 regulates intraocular pressure by influencing aqueous humor production and has

been implicated in glaucoma[87]. Most of the eye OCT-associated proteins (87 of 93, over 94%) were specific to the eye, with minimal overlap observed with other organs. Most proteins were positively associated with GCIPL thickness and negatively associated with RNFL thickness, except for OMG, which showed the opposite effect and was also the only plasma protein shared between eye OCT traits and brain IDPs. In addition, 39 proteins are associated with eye fundus images features (|β| range = (0.07, 0.10), $P < 1.62\times 10^{-8}$, **Fig. S13**). The majority (27/39) of these proteins overlapped with those associated with eye OCT measures, emphasizing the shared proteomic links of these ocular imaging traits.

**Tracing the putative origins of proteomic effects on imaging traits**

Plasma proteins are produced by various organs and tissues throughout the body. Understanding the origins of the identified proteomic effects on IDPs can provide deeper insights into their biological roles and cross-organ interactions, particularly when proteins related to one organ imaging trait may originate from another. The putative origins and organ-specific plasma proteins can be inferred using external organ/tissue-specific RNA sequencing data[38]. Specifically, plasma proteins can be considered organ-specific[13-15] if their gene expression in a particular organ is at least four times higher than in any other organ[19]. Following this definition, 19% of plasma proteins (557 out of 2,923) could be traced back to a single organ, accounting for 28% of the identified phenotypic protein-imaging associations (1,415 out of 5,067). For each organ, we evaluated whether the identified imaging-associated proteins were enriched among these organ-specific plasma proteins (**Table S7**). We found that, in some organs, proteins linked to imaging traits were enriched in the set of highly expressed genes from the same organ. Conversely, in other cases, imaging-associated proteins in one organ were enriched in genes highly expressed in a different organ (**Figs. 3A-3B**). This raises the possibility of distant biologic effects, although alternative explanations exist. For example, these relationships may result from organ cross-talk through other mechanisms (neurogenic) or reflect overarching causative factors affecting imaging traits in one organ and protein expression in a different organ without a direct biologic effect the identified proteins across organs. The overall enrichment patterns are detailed below.

Proteins associated with IDPs of the brain (both sMRI and dMRI), liver, pancreas, spleen, body fat, and lung were significantly enriched among genes highly expressed in the same organ (enrichment $P < 6.95\times 10^{-4}$). For example, 4 of the 6 dMRI-associated proteins (BCAN, NCAN, GFAP, and OMG) had high gene expression in the brain (enrichment $P = 4.40\times 10^{-7}$), and 10 of the 33 sMRI-associated proteins were also defined to be brain-specific (enrichment $P = 1.52\times 10^{-8}$, **Fig. 3C**). These plasma proteins, likely originating from brain tissues, play specialized roles in brain structure and function[83,88]. Similarly, 25 of the 271 liver-associated proteins were highly expressed in the liver (enrichment $P = 2.54\times 10^{-9}$, **Fig. 3D**), almost all of which are known to be primarily produced in the liver (such as LCAT[58] and FGF21[60]) and involved in key liver functions such as detoxification, metabolism, lipid transport, and blood coagulation. Among pancreas-associated proteins, 15 of the 26 were pancreas-specific (such as AMY2A, AMY2B, PNLIP, and PNLIPRP1, enrichment $P = 0$, **Fig. 3E**), reflecting the pancreas's specialized role in digestion. Similar enrichment patterns were also observed on spleen, body fat, and lung (enrichment $P < 6.95 \times 10^{-4}$, **Figs. 3F-H** and **Supplementary Note**). Together, these findings demonstrate that proteins phenotypically associated with specific organs often have high organ-specific gene expression, highlighting their specialized roles in maintaining organ structure and function.

While many proteins associated with imaging data of specific organs are highly expressed in those same organs, some are predominantly expressed in other organs. Specifically, significant enrichments were observed among proteins produced in the liver, adipose tissue, and kidney (**Figs. 3A**), reflecting more interconnected, multi-organ regulatory mechanisms. Specifically, 19 heart-associated proteins (enrichment $P = 2.80\times 10^{-7}$) and 33 body fat-associated proteins (enrichment $P = 0$) were highly expressed in the liver, reflecting its systemic regulatory role across multiple physiological processes (**Figs. 3I-3J**). Many of these liver-produced heart-associated proteins (such as F7, F9, FGA, SERPIND1, and CFB) are crucial to thrombosis and maintaining heart health. For example, F7 is a coagulation factor that promotes thrombosis. It is also associated with coronary artery disease, which is strongly influenced by cholesterol and triglyceride

levels[89]. Specific genetic variants of *F7* were associated with lower F7 levels and lower risk of myocardial infarction[90]. Another coagulation factor, F9, was a risk factor of deep venous thrombosis[91], and F9 activation was showed in patients with acute coronary syndromes[92]. Higher levels of F9 have been observed in obese individuals[93]. SERPIND1 has protective effect on atherosclerosis and restenosis[94]. Additionally, liver-produced body fat-associated proteins, such as FGF21 and IGFBP2, are involved in energy balance, and lipid and glucose metabolism[60,95]. These findings suggest the connection between liver-produced proteins and their links to cardiovascular and metabolic regulation, highlighting the liver's pivotal role in heart health and systemic metabolism.

Moreover, many proteins associated with body muscle, liver, and heart were produced in the adipose tissue, such as CD300LG, FABP4, LEP, and ADIPOQ (enrichment $P < 1.18\times 10^{-4}$, **Figs. 3K-3L**). These proteins tend to have pleiotropic roles across various biological processes. FABP4, secreted by adipocytes, is implicated in cardiometabolic risk, atherosclerosis, coronary artery disease, and cardiac dysfunction, serving as both a potential biomarker and a contributor to cardiac metabolism[96]. It also serves as a predictive marker for the progression of nonalcoholic fatty liver disease to nonalcoholic steatohepatitis, indicating patients at higher risk of liver disease complications[97]. LEP, primarily produced by adipose tissue, plays a key role in regulating cardiac metabolism, preventing cardiac lipotoxicity, and is associated with various cardiovascular complications. Furthermore, it plays a crucial role in liver health by preventing lipid accumulation and facilitating lipid mobilization[98]. These findings highlight the systemic influence of adipose-derived proteins in coordinating metabolic and physiological processes across organs, emphasizing the interconnected roles of adipose tissue in maintaining whole-body homeostasis. Additionally, one aorta-associated protein, REN, was produced by the kidney. REN activates the renin-angiotensin system, with angiotensin II causing blood vessel constriction and increasing blood pressure, which places additional strain on the aorta. This highlights the kidney's indirect yet critical influence on aortic health via protein regulation[99].

**Organ-wise chart of proteomic prediction accuracy and key predictors**

Given the strong associations between plasma proteins and whole-body IDPs, we assessed their combined predictive power for organ structure and function and identified the top-ranked proteins for IDP prediction. Using an elastic-net model, we resampled half of the subjects for training and testing, repeated this process 200 times, evaluated the mean prediction *R*-squared ($r^2$) of plasma proteins for IDPs, and selected the top robust predictors[16] (**Methods**). Abdominal MRI traits showed the highest predictive power, particularly for body fat IDPs (median $r^2$ range = (0.45, 0.55), **Fig. 4A**). The full set of 2,923 proteins generally outperformed the subset of 557 organ-specific proteins (i.e., proteins whose genes are highly expressed in a specific organ) in predicting IDPs across all organs (**Fig. S14**). Nevertheless, we observed that the top-ranked predictors for IDPs of the brain, body fat, liver, pancreas, and lungs were predominantly organ-specific proteins corresponding to those respective organs. Below we highlight the IDPs for which plasma proteins demonstrated high predictive power. Complete results, including prediction performance and top five predictors for each IDP, are provided in **Table S8**.

The highest predictive power was observed for abdominal MRI traits (median $r^2$ range = (0.10, 0.55), **Fig. S15**). For some organs and tissues, including body fat, liver, pancreas, and lung, top protein predictors for related IDPs were primarily organ-specific. This makes the prediction power of the full set of plasma proteins similar to that of the subset primarily expressed in these organs. Among these, body fat traits showed the highest prediction accuracy, particularly for total trunk fat volume, total abdominal adipose tissue index, and abdominal fat ratio ($r^2$ range = (0.57, 0.61), **Figs. 4B** and **S16-S17**). Key predictors, such as FABP4 and LEP, which are primarily secreted by adipose tissue, consistently and strongly predicted all body fat imaging traits with high power. CLMP and CFH also demonstrated strong predictive power for body fat. CLMP is involved in adipocyte differentiation and linked to the progression of obesity[100]. Plasma proteins also showed strong prediction performance for liver IDPs. Among these, liver volume had the highest prediction power ($r^2$ = 0.34), with top predictors, such as SEZ6L and VWC2L, primarily expressed in the brain (**Fig. S18**). Additionally, plasma proteins effectively predicted liver inflammation and fibrosis ($r^2$ = 0.20, **Fig. S19**), with key predictors, such as APOF and SERPINA6, predominantly expressed in the liver. APOF regulates hepatic lipoprotein

metabolism by enhancing low-density lipoprotein triglyceride release and promoting the clearance of lipoprotein remnants and is associated with hepatic steatosis[101]. For pancreas IDPs, plasma proteins predicted pancreas volume and pancreas fat fraction with moderate power ($r^2$ range = (0.15, 0.17)). Top predictors for pancreas volume included PLA2G1B, CELA2A, and KIRREL2 (**Fig. 4C**), all primarily produced in the pancreas. PLA2G1B and CELA2A, produced by pancreatic acinar cells, are involved in digesting dietary proteins and phospholipids and are linked to insulin sensitivity[70,102]. KIRREL2, primarily expressed in pancreatic beta cells, may play a role in beta cell function and pancreas development[103]. For pancreas fat fraction, PLA2G1B was also the strongest predictor. Additionally, LEP and FABP4, which were key predictors for body and visceral fat, showed strong predictive power for pancreas fat fraction as well (**Fig. S20**). In the lung, the strongest predictor of lung volume was AGER, which is highly expressed in lung tissue. Together with other key proteins such as NCAN, LEP, ENG, and CD93, plasma proteins predicted lung volume with a prediction $r^2$ of 0.19 (**Fig. S21**).

Plasma proteins also show strong predictive power for IDPs of body muscle, spleen, and kidney, despite the genes of top predictors not being highly expressed in these corresponding organs. Among these, spleen volume showed the highest predictive power ($r^2$ = 0.45), with top predictors including VCAM1, CRLF1, SEMA7A, PTPRH, and WFIKKN1 (**Fig. S22**). Many of these proteins are key players in inflammation and immune responses. For example, VCAM1 regulates leukocyte adhesion to blood vessel walls and their passage through the endothelial layer[104], while SEMA7A, expressed on activated T cells, stimulates macrophages to produce proinflammatory cytokines, driving inflammatory immune responses[105]. For body muscle, weight-to-muscle ratio and muscle fat infiltration had the highest predictive power ($r^2$ range = (0.35, 0.44), **Figs. S23-S25**). Key proteins such as CLMP, ART3, and RGMA consistently predicted multiple muscle-related traits. Notably, CLMP also demonstrated strong predictive power for body fat traits, highlighting its dual role in muscle and fat-related processes. Similarly, LEP and FABP4, both key predictors of body fat traits, also strongly predicted muscle traits like muscle fat infiltration in the posterior thigh and weight-to-muscle ratio, further emphasizing the overlap in proteins driving fat and muscle composition. Additionally, plasma proteins also

showed robust predictive ability for kidney IDPs, particularly kidney parenchyma volume and kidney volume ($r^2$ range = (0.24, 0.35), **Figs. S26-S27**). Top predictors for kidney traits included SPOCK1, VWC2L, and LRTM2, which are highly expressed in the brain, and CYTL1, which is highly expressed in the aorta.

Similar to the kidney, specific proteins highly expressed in the brain also dominate the prediction of MRI traits of the heart and aorta. Heart IDPs with the highest predictive performance include myocardial mass and ventricular end-diastolic volumes (both left and right) as well as global left ventricular myocardial-wall thickness at end-diastole ($r^2$ range = (0.18, 0.24), **Fig. S28**). For the aorta, IDPs with high predictive power were related to the descending aorta, particularly descending aorta minimum and maximum areas ($r^2$ = 0.11). SEZ6L and VWC2L, highly expressed in the brain, consistently predicted IDPs of multiple organs, including left ventricular myocardial mass and end-diastolic volumes, as well as liver volume, emphasizing their significance across multiple systems (**Figs. 4D** and **S29-S32**). Notably, predictors for left ventricular myocardial-wall thickness differed from those of other IDPs. Top predictors for left ventricular myocardial-wall thickness include CCN3, REN, and CHGB (**Fig. S33**), many of which are involved in cardiac remodeling and hypertrophy. For example, REN activates the renin-angiotensin system and contributes to left ventricular hypertrophy[99]. Similarly, CCN3-deficient mice exhibit septal thickening, hypertrophic cardiomyopathy, and ventricular dilation, highlighting CCN3's role in regulating myocardial-wall thickness and maintaining overall heart health[106]. Additionally, CHGB plays a critical role in catecholamine secretion, and genetic variations in *CHGB* have been associated with increased risks of hypertension[107], which likely contribute to the relationship between CHGB and left ventricular myocardial-wall thickness.

Among brain IDPs, relatively high predictive performance was observed for global measures, including total brain volume, white matter volume, and grey matter volume ($r^2$ range = (0.08, 0.10)). Many top predictors (such as MOG, PTPRR, and NCAN) were highly expressed in the brain (**Figs. 4E** and **S34-S36**). These proteins, along with other strong predictors such as KLK6, GFRA3, and SLITRK1, were associated with structural

brain IDPs and contributed to brain structure, function, or development. Regional brain volumes generally had smaller prediction power than global ones. For example, plasma proteins predicted volumes of the cerebellum VIIIa/VIIIb, insular cortex, thalamus, amygdala, and hippocampus with moderate accuracy ($r^2$ range = (0.03, 0.05)). NCAN, MOG, and BCAN, highly expressed in the brain, were the strongest predictors for these regions, except for the cerebellum (**Figs. 4F** and **S37-S39**). In contrast, protein predictors for cerebellum volumes differed and varied slightly across subregions. GDF15 and LEP consistently predicted the volumes of both left and right cerebellum VIIIa/VIIIb (**Fig. S40**). LEP influences neural activity in the posterior cerebellum, modulating brain responses to food-related stimuli and food intake-related plasticity[108]. More results for brain dMRI and eye are summarized in the **Supplementary Note** and **Figures S41-S47**. For example, one of the top predictors of brain dMRI was APCS (**Fig. S46**), which produced exclusively in the liver and has been linked to neurodegenerative diseases (such as Alzheimer's disease) when it appears in the brain due to compromised blood-brain barrier integrity. APCS is cytotoxic to cerebral neurons, promotes Aβ amyloid formation, highlighting an inter-organ proteomic relationship in disease pathology[109,110].

**Stratification capability of plasma protein prediction models**

The developed plasma protein-based prediction models for IDPs can be used to stratify individuals with abnormal organ structure and function without requiring actual imaging data. Such applications take advantage of the greater accessibility of blood-based measurements compared to imaging modalities like whole-body MRIs. In this section, we evaluated the stratification capability by calculating the ratio of true IDP values between the top and bottom 10% of protein-predicted IDP deciles[111]. Consistent with the cohort-level prediction accuracy, IDPs related to body organ fat and visceral fat exhibited the highest individual-level stratification ratios. The largest difference was observed for VAT, where the true VAT volume in the top 10% of protein-predicted VAT was over four times higher than in the bottom 10% (**Fig. 4G**). This demonstrates that plasma proteins-trained computational models can effectively stratify individuals with high VAT volumes. Other body fat traits, such as total trunk fat volume and total abdominal adipose tissue index, also showed substantial stratification (median ratio > 3.5, **Fig. S48**). Stratifications were

similarly effective for organ-associated fat volumes, such as liver fat fraction and pancreas fat fraction (median ratio > 2.8, **Fig. S49**).

Plasma protein prediction models also effectively stratify volumetric traits of multiple organs, with spleen volume exhibiting the highest stratification ratio (median ratio > 2.6, **Fig. S50**). Stratification for other organ volumes, including the kidney, liver, lung, and pancreas, showed consistent patterns (median ratio > 1.4, **Fig. S51**). For cardiac MRI traits, left ventricle end-diastolic volume, left ventricular myocardial mass, and right ventricle end-systolic volume had modest stratification (median ratio > 1.3, **Fig. S52**). For brain IDPs, both global volumes such as cerebral white matter (median ratio > 1.12, **Fig. S53**) and regional volumes such as the amygdala and insular cortex (median ratio > 1.07, **Fig. S54**) demonstrated stratification capability. Importantly, we found that the stratification performance is generally robust over time. Due to the UKB data collection procedures, plasma samples and imaging scans were collected at different visits, with a gap of 3 to 17 years between collections depending on the individual. We found that this time gap has only a minor impact on stratification performance, as illustrated by body and visceral fat traits (**Fig. S55**) and kidney and lung volumes (**Fig. S56**). Overall, plasma protein-based stratification is particularly valuable for whole body images that are not routinely accessible in healthy large-scale cohorts. Computational models with plasma protein levels from blood samples provide a promising tool for evaluating the health of organs and their specific regions.

**Genetic-root putative causal links between plasma proteins and imaging traits**

Mendelian randomization (MR) analysis was performed to identify plasma proteins with genetic-driven putative causal links to IDPs (**Methods**). We tested 2,856 proteins using their *cis*-protein quantitative trait loci (pQTL) variants with inverse variance weighted[112] and Wald-ratio methods[113,114] (**Table S9**). To ensure robustness, several sensitivity tests[115] were performed to validate the MR assumptions, and colocalization analysis[33] were further conducted to identify protein-imaging pairs sharing causal variants (**Table S10**). After Bonferroni correction ($P < 2.98\times 10^{-8}$) and excluding results that did not pass sensitivity tests, we identified 8,116 significant protein-IDP genetic causal pairs involving

318 proteins and 1,041 IDPs, with 448 of these pairs (98 proteins and 334 IDPs) showing PPH4 > 80% in the colocalization analysis (**Fig. 5A**). The MR associations of these 318 proteins were broadly shared across organs, with 250 influencing more than one IDP (**Fig. 5B**). These findings indicate that genetic-driven proteomic links are often pleiotropic, affecting multiple organs. Below we report the overall patterns and highlight key genetically causal plasma proteins for each organ.

Plasma proteins demonstrated extensive MR associations with brain IDPs, including regional brain volumes, white matter microstructure, and fMRI traits. Such strong MR association power resulted in the identification of numerous enriched biological pathways[40] in the brain (**Fig. 5C** and **Table S11**). Approximately 60% of the identified proteins (over 190) exhibited genetic causal links to brain sMRI and dMRI IDPs (|β| range = (0.01, 0.67), $P < 2.98\times 10^{-8}$), with 140 proteins affecting both modalities. Two examples of overlapping proteins are GRN and CTSB. Elevated GRN level was linked to increased thalamus volumes (β range = (0.12, 0.13), $P < 3.12\times 10^{-10}$, PPH4 range = (98.52%, 99.21%), **Fig. S57**) and reduced mean orientation dispersion index in the fornix cres and stria terminalis (β = -0.15, $P = 6.32\times 10^{-12}$, PPH4 = 97.53%, **Fig. S58**). Elevated CTSB level was associated with reduced grey matter volume in the left inferior frontal gyrus pars opercularis (β = -0.09, $P = 3.70\times 10^{-39}$, PPH4 = 80.00%, **Fig. S59**) and impaired white matter integrity in uncinate fasciculus, anterior corona radiata, and cingulum hippocampus tracks (β range = (-0.10, -0.08), $P$ range = ($3.09\times 10^{-32}$, $1.20\times 10^{-37}$)). Both GRN and CTSB are known therapeutic targets for neurodegenerative diseases such as Alzheimer's disease and frontotemporal dementia[116,117]. We provide more examples regarding BCAN, OMG, FOXO1, PCDH7 in the **Supplementary Note** and **Figures 6A** and **S60-S66**. In addition, although plasma proteins showed no phenotypic associations with fMRI traits, genetic causal relationships were observed for both functional activity (94 proteins, |β| range = (0.01, 0.57), $P < 2.95\times 10^{-8}$) and functional connectivity measures (38 proteins, |β| range = (0.02, 0.60), $P < 2.73\times 10^{-8}$). For example, APOE had a genetic causal link with functional activity (amplitude) in the default mode and central executive networks (β = 0.06, $P = 5.40\times 10^{-9}$, PPH4 = 99.71%, **Fig. 6B**). APOE is a well-

documented genetic risk factor for Alzheimer's disease, with the ε4 allele increasing risk by impairing Aβ clearance and promoting aggregation, contributing to Aβ pathology[118].

Among non-brain organs, plasma proteins had the strongest causal relationships with the heart and aorta (|β| range = (0.01, 0.59), $P < 2.94\times 10^{-8}$). A total of 114 plasma proteins (104 for heart and 37 for aorta) were identified to have MR associations with cardiac MRI traits, primarily involving end-diastolic/systolic volumes and ascending/descending aorta features. For example, increased level of FGF5 was causally associated with larger ascending aorta maximum area and minimum area (β = 0.05, $P < 3.81\times 10^{-11}$, PPH4 range = (98.72%, 99.26%), **Figs. 6C** and **S67**). FGF5 promotes angiogenesis in human aortic endothelial cells by enhancing vascular sprouting[119] and has been linked to hypertension risk across multiple ethnicities in genetic studies[120,121]. Elevated levels of CILP were causally linked to a reduced left atrium ejection fraction (β = -0.11, $P = 9.10\times 10^{-10}$, PPH4 = 98.26%, **Fig. 6D**). CILP plays a key role in extracellular matrix remodeling and serves as a marker for cardiac fibrosis[122,123]. EFEMP1, another protein critical for maintaining ECM structure, was positively associated with right atrium stroke volume (β = 0.14, $P = 2.41\times 10^{-10}$, PPH4 = 93.57%, **Fig. S68**). EFEMP1 promotes proper scar formation after myocardial infarction, preventing cardiac rupture[124]. Additionally, elevated level of PDE5A was causally associated with increased right ventricular end-diastolic volume and stroke volume (β range = (0.12, 0.15), $P$ range = ($1.84\times 10^{-33}$, $2.83\times 10^{-74}$), PPH4 range = (89.37%, 92.62%), **Fig. S69**). PDE5A is an approved therapeutic target for hypertension[125], and its inhibition provides protective effects against cardiac stresses such as ischemia-reperfusion injury, drug toxicity, pressure-induced hypertrophy, and acute stress responses in the heart muscle[126].

Genetic causal links were identified between plasma proteins and eye IDPs, including both OCT measures (84 proteins, |β| range = (0.01, 0.69), $P < 2.79\times 10^{-8}$) and fundus images (93 proteins, |β| range = (0.01, 0.48), $P < 2.87\times 10^{-8}$). Many of these linked proteins play critical roles in eye health, and some also have causal relationships with other organs, particularly the brain and heart. For example, GRN was causally associated with the average thickness between the external limiting membrane and the inner and

outer photoreceptor segments layers (β = -0.11, $P = 5.28\times 10^{-9}$, PPH4 = 99.89%, **Fig. 6E**). GRN promotes the differentiation of retinal precursor cells into photoreceptor cells[127] and supports retinal ganglion cell survival during development by regulating astrocyte activation[128], highlighting its essential role in retinal regeneration and development. Similarly, EFEMP1 was causally linked to the vertical cup-to-disc ratio (β = -0.25, $P = 5.00\times 10^{-19}$, **Fig. S70**), a trait associated with primary open-angle glaucoma. The Arg345Trp mutation in *EFEMP1* has been linked to drusen formation in Malattia Leventinese, closely resembling age-related macular degeneration pathology[129,130]. Moreover, CFH had a MR association with increased thickness between the inner nuclear layer and retinal pigment epithelium (β = 0.10, $P = 1.10\times 10^{-54}$, PPH4 = 93.71%). CFH plays a protective role in the eye by regulating complement activation and preventing inflammatory damage. Notably, the Y402H variant in *CFH* has been linked to increased risk of age-related macular degeneration[131,132].

A total of 79 proteins showed MR associations with abdominal MRI IDPs, with the strongest relationships observed in the liver, spleen, and pancreas. Specifically, in the liver, elevated levels of RNF43 and APOH were causally associated with lower liver iron-corrected T1 (β range = (-0.06, -0.31), $P$ range = ($3.46\times 10^{-10}$, $3.84\times 10^{-32}$), PPH4 range = (98.87%, 99.84%), **Figs. S71-S72**), a measure of liver inflammation and fibrosis. RNF43 regulates WNT signaling, maintaining hepatocyte proliferation and differentiation balance, ensuring liver homeostasis and preventing liver cancer[133,134]. *APOH*, predominantly expressed in the liver[135], has been linked to nonalcoholic fatty liver disease in genetic studies[136]. Elevated level of ADH4 was causally associated with increased liver volume (β = 0.25, $P = 4.81\times 10^{-12}$, PPH4 = 95.08%, **Fig. S73**). ADH4, which plays a role in alcohol metabolism, is downregulated in hepatocellular carcinoma and serves as a potential prognostic biomarker and therapeutic target[137,138]. Additionally, strong MR relationships were identified for spleen volume, with 114 plasma proteins showing significant associations (|β| range = (0.03, 0.42), $P < 1.90\times 10^{-9}$)). Among these, TNFSF13B and EGF demonstrated the strongest links, with increased levels of both proteins linked to spleen enlargement (β range = (0.26, 0.42), $P$ range = ($7.70\times 10^{-18}$, $3.99\times 10^{-23}$), PPH4 range = (94.52%, 99.90%), **Figs. 6F** and **S74**). TNFSF13B, primarily

expressed in the spleen, regulates B-cell survival, proliferation, and differentiation[139], highlighting the spleen as a key site for its immune-regulatory function. EGF is essential for suppressing cellular senescence and sustaining growth in mammals, and inhibition of its receptor has been shown to cause spleen atrophy in mice[140]. MR associations were also observed for pancreas IDPs. For example, ADAM15 was linked to pancreatic iron levels (β = 0.03, $P = 1.62\times 10^{-26}$, PPH4 = 81.22%, **Fig. S75**). ADAM15 is overexpressed in pancreatic cancer cells, suggesting its role in tumor progression[141]. For the lung, IL4R was causally associated with lung volume (β = -0.05, $P = 3.52\times 10^{-11}$) and is a known drug target for chronic lung diseases such as asthma and pulmonary fibrosis. IL4R is critical for airway inflammation[142], and polymorphisms in the IL4 promoter (C-589T) are associated with asthma[143].

## Discussion

This study represents the largest imaging proteomics analysis to date. Using plasma protein and multi-organ imaging data from the UKB study, we identified novel phenotypic protein-imaging associations, shedding light on the roles plasma proteins play in organ-specific mechanisms and related biological pathways. These associations were largely robust to disease status, with a correlation over 99.8%, which is likely due to the predominantly healthy composition of the UKB cohort. Integration with external gene expression data revealed that some proteins associated with specific organs were secreted or highly expressed in the same organ, while others exhibited cross-organ relationships, notably involving proteins derived from adipose tissue and the liver. This finding underscores the liver's central role in regulating systemic and metabolic processes. The predictive power of plasma proteins varied across organs, with abdominal tissues and organs showing the highest predictive accuracy. Genetic causal associations between plasma proteins and IDPs were identified using genetic instrumental variables, with approximately 5.5% further passing colocalization tests, indicating strong evidence of shared causal variant. As the most comprehensive pan-organ imaging proteomics analysis conducted within a single large-scale cohort, this study offers novel insights into the interplay between plasma proteins and organ-specific biology.

Building on these novel findings, our study bridges molecular-level biological processes with structural biology observed through whole-body imaging, revealing mechanisms underlying structural and functional changes in organs. For example, in the brain, consistent links were identified between specific proteins, such as BCAN, NCAN, MOG, and SLITRK1 and key brain regions, including the frontal pole, hippocampus, and amygdala, which are critical hubs of the limbic system involved in emotion regulation and the fight-or-flight response. The thalamus also emerged consistently, indicating the involvement in the reward systems. These proteins not only associate with these brain regions but also demonstrate strong predictive abilities, highlighting their potential to reveal molecular pathways underlying emotion regulation and motivational behavior. Furthermore, our results suggest that plasma proteins serve as robust non-invasive biomarkers for the heart and various abnormal organs, enabling early detection of organ dysfunction, large-cohort subject stratification, and the development of protein-targeted therapies for organ-specific and systemic diseases. Together, these findings enhance our understanding of molecular and structural biology and provide a foundation for future proteomic applications in precision medicine.

The phenotypic and genetic associations between proteins and imaging traits reveal distinct yet complementary patterns across different organs. Phenotypic association analysis identified more signals in non-brain organs, with particularly strong and extensive associations observed between plasma proteins and abdominal MRI traits. Notably, the set of proteins identified through phenotypic associations displayed higher levels of interaction (**Table S5**), greater enrichment in biological pathways (**Table S6**), and elevated gene expression within the same or other specific organs (**Table S7**). These findings provide deep insights into the proteomic networks and mechanisms underlying protein-imaging associations, highlighting how phenotypic analyses capture both organ-specific and organ-shared biological influences, potentially driven by environmental and non-genetic factors. These findings are consistent with previous phenotypic analyses of omics data, which have identified co-regulated clusters and protein modules enriched in network biology[7,144].

In contrast, genetic causal proteins identified through the MR approach showed minimal overlap with phenotypically associated proteins. Notably, MR demonstrated greater power in the brain, identifying more putative causal proteins for brain IDPs, including fMRI traits that lacked power of identifying phenotypic associations. Of note, the set of proteins identified by MR consistently exhibited lower protein interaction scores (**Table S12**) compared to phenotypically associated proteins across most organs (**Fig. 5D**). Similarly, except for the brain and eye, MR-identified proteins had lower number of enriched biological pathways than phenotypically associated proteins (**Fig. 5E**). Specifically, there was minimal overlap between pathways linked to phenotypic and MR associations, except for the heart, body fat, liver, and kidney, where pathways enriched by phenotypically associated proteins largely encompassed those enriched by MR-associated proteins. Moreover, MR-identified proteins in most organs did not show significant enrichment in organ-specific gene expression, with the only exception of the lung (**Table S13**). These findings suggest that, compared to phenotypic analyses, MR-based genetic mapping may prioritize isolated proteins rather than interconnected protein co-regulated clusters and may be less connected to organ-specific biological mechanisms reflected in gene expression.

Notably, on the other hand, many MR-identified proteins are either approved or in-development drug targets[145], with the associated diseases aligning with the organ of the imaging data. For example, APOE, GRN, BCHE, CTSB, and APCS, which have genetic causal associations with brain imaging traits, are drug targets for neurodegenerative diseases such as Alzheimer's disease, dementia, cognitive impairment, and Parkinson's disease. Similarly, proteins genetically linked to heart imaging traits serve as drug targets for cardiovascular diseases, including PLAU for myocardial infarction, AOC3, PDE5A, and SELE for hypertension, and LCAT for coronary artery disease. It is well established that genetic evidence is valuable for drug development and clinical trial success[146]. Our MR analysis provides key genetic insights into potential drug targets and highlights opportunities for leveraging imaging data in therapeutic target discovery process. Together, the differences between phenotypic and genetic results underscore the unique strengths of each approach, which may be helpful in different downstream applications.

By combining phenotypic and genetic analyses, the present study provides a more comprehensive resource for understanding protein-imaging relationships.

**Acknowledgements**

Research reported in this publication was supported by the National Institute On Aging under Award Numbers RF1AG082938 and 1R01AG085581. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The study has also been partially supported by funding from the Purdue University Statistics Department, Department of Statistics and Data Science at the University of Pennsylvania, Wharton Dean's Research Fund, Analytics at Wharton, Wharton AI & Analytics Initiative, Perelman School of Medicine CCEB Innovation Center Grant, and the University Research Foundation Grant. The individual-level data UK Biobank used in this study were obtained under application 76139 subject to a data transfer agreement. We would like to thank the individuals who represented themselves in the UK Biobank for their participation and the research teams for their efforts in collecting, processing, and disseminating these datasets. We would like to thank the research computing and IT groups at the Wharton School of the University of Pennsylvania and the Rosen Center for Advanced Computing at the Purdue University for providing computational resources and support that have contributed to these research results.

**Author Contributions Statement**

Z.F. and B.Z. designed the study. Z.F. analyzed the data and generated the results. X.Y., J.S., and Y.L. helped with data preprocessing and analysis. J.C., J.M.O., W.W., D.J.R., and R.G. provided feedback on study design and helped results interpretations. Z.F. and B.Z. wrote the manuscript with feedback from all authors.

**Competing Interests Statement**

The authors declare no competing interests.

**Reference**

1. Assarsson, E. *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PloS one* **9**, e95192 (2014).
2. Gold, L. *et al.* Aptamer-based multiplexed proteomic technology for biomarker discovery. *Nature Precedings*, 1-1 (2010).
3. Suhre, K., McCarthy, M.I. & Schwenk, J.M. Genetics meets proteomics: perspectives for large population-based studies. *Nature Reviews Genetics* **22**, 19-37 (2021).
4. Sun, B.B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329-338 (2023).
5. Eldjarn, G.H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348-358 (2023).
6. Sun, B.B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79 (2018).
7. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769-773 (2018).
8. Carrasco-Zanini, J. *et al.* Mapping biological influences on the human plasma proteome beyond the genome. *Nature Metabolism*, 1-14 (2024).
9. Pietzner, M. *et al.* Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
10. Williams, S.A. *et al.* Plasma protein patterns as comprehensive indicators of health. *Nature medicine* **25**, 1851-1857 (2019).
11. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nature Reviews Genetics* **23**, 715-727 (2022).
12. Argentieri, M.A. *et al.* Proteomic aging clock predicts mortality and risk of common age-related diseases in diverse populations. *Nature medicine* **30**, 2450-2460 (2024).
13. Oh, H.S.-H. *et al.* Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164-172 (2023).
14. Goeminne, L.J. *et al.* Plasma-based organ-specific aging and mortality models unveil diseases as accelerated aging of organismal systems. *medRxiv*, 2024.04. 08.24305469 (2024).
15. Oh, H.S.-H. *et al.* Plasma proteomics in the UK Biobank reveals youthful brains and immune systems promote healthspan and longevity. *bioRxiv*, 2024.06. 07.597771 (2024).
16. Carrasco-Zanini, J. *et al.* Proteomic signatures improve risk prediction for common and rare diseases. *Nature medicine* **30**, 2489-2498 (2024).
17. Deng, Y.-T. *et al.* Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* (2024).
18. Meissner, F., Geddes-McAlister, J., Mann, M. & Bantscheff, M. The emerging role of mass spectrometry-based proteomics in drug discovery. *Nature Reviews Drug Discovery* **21**, 637-654 (2022).
19. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
20. Harris, S.E. *et al.* Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nature Communications* **11**, 800 (2020).
21. Duggan, M.R. *et al.* Proteomics identifies potential immunological drivers of postinfection brain atrophy and cognitive decline. *Nature Aging* **4**, 1263-1278 (2024).
22. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
23. Miller, K.L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience* **19**, 1523-1536 (2016).

24. Bai, W. *et al.* A population-based phenome-wide association study of cardiac and aortic structure and function. *Nature Medicine* **26**, 1654-1662 (2020).
25. Rajiah, P.S., François, C.J. & Leiner, T. Cardiac MRI: state of the art. *Radiology* **307**, e223008 (2023).
26. Littlejohns, T.J. *et al.* The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nature communications* **11**, 1-12 (2020).
27. Karamitsos, T.D., Francis, J.M., Myerson, S., Selvanayagam, J.B. & Neubauer, S. The role of cardiovascular magnetic resonance imaging in heart failure. *Journal of the American College of Cardiology* **54**, 1407-1424 (2009).
28. Taouli, B., Ehman, R.L. & Reeder, S.B. Advanced MRI methods for assessment of chronic liver disease. *American journal of roentgenology* **193**, 14-27 (2009).
29. Jack Jr, C.R. *et al.* The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **27**, 685-691 (2008).
30. Geevarghese, A., Wollstein, G., Ishikawa, H. & Schuman, J.S. Optical coherence tomography and glaucoma. *Annual review of vision science* **7**, 693-726 (2021).
31. Yoon, J.H. *et al.* Integrative approach of omics and imaging data to discover new insights for understanding brain diseases. *Brain Communications* **6**, fcae265 (2024).
32. Wu, C., Zhang, Z., Yang, X. & Zhao, B. Large-scale imputation models for multi-ancestry proteome-wide association analysis. *bioRxiv*, 2023.10. 05.561120 (2023).
33. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383 (2014).
34. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS genetics* **17**, e1009440 (2021).
35. Konopka, G. & Bhaduri, A. Functional genomics and systems biology in human neuroscience. *Nature* **623**, 274-282 (2023).
36. Ayubcha, C., Sun, B.B., Ge, T. & Chen, C.-Y. Large-scale proteomic and genomic analysis identify plasma proteins influencing human brain structure and Alzheimer's disease risk. in *Alzheimer's Association International Conference* (ALZ, 2024).
37. Zhao, B. *et al.* Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. *Nature communications* **12**, 2878 (2021).
38. Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
39. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638-D646 (2023).
40. Kolberg, L. *et al.* g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res* **51**, W207-W212 (2023).
41. Panagopoulou, V. *et al.* NTproBNP: an important biomarker in cardiac diseases. *Curr Top Med Chem* **13**, 82-94 (2013).
42. Katsiki, N., Mikhailidis, D.P. & Banach, M. Leptin, cardiovascular diseases and type 2 diabetes mellitus. *Acta Pharmacol Sin* **39**, 1176-1188 (2018).
43. Poetsch, M.S., Strano, A. & Guan, K. Role of Leptin in Cardiovascular Diseases. *Front Endocrinol (Lausanne)* **11**, 354 (2020).

44. Xu, J. *et al.* GDF15/MIC-1 functions as a protective and antihypertrophic factor released from the myocardium in association with SMAD protein activation. *Circ Res* **98**, 342-50 (2006).
45. Hwang, S.J. *et al.* Circulating adhesion molecules VCAM-1, ICAM-1, and E-selectin in carotid atherosclerosis and incident coronary heart disease cases: the Atherosclerosis Risk In Communities (ARIC) study. *Circulation* **96**, 4219-25 (1997).
46. Yancopoulos, G.D. *et al.* Vascular-specific growth factors and blood vessel formation. *Nature* **407**, 242-8 (2000).
47. Gale, N.W. & Yancopoulos, G.D. Growth factors acting via endothelial cell-specific receptor tyrosine kinases: VEGFs, angiopoietins, and ephrins in vascular development. *Genes Dev* **13**, 1055-66 (1999).
48. van Cruijsen, H., Giaccone, G. & Hoekman, K. Epidermal growth factor receptor and angiogenesis: Opportunities for combined anticancer strategies. *Int J Cancer* **117**, 883-8 (2005).
49. Kim, W. *et al.* Angiogenic role of adrenomedullin through activation of Akt, mitogen-activated protein kinase, and focal adhesion kinase in endothelial cells. *FASEB J* **17**, 1937-9 (2003).
50. Ieguchi, K. & Maru, Y. Roles of EphA1/A2 and ephrin-A1 in cancer. *Cancer Sci* **110**, 841-848 (2019).
51. Khurana, R., Simons, M., Martin, J.F. & Zachary, I.C. Role of angiogenesis in cardiovascular disease: a critical appraisal. *Circulation* **112**, 1813-24 (2005).
52. Deveza, L., Choi, J. & Yang, F. Therapeutic angiogenesis for treating cardiovascular diseases. *Theranostics* **2**, 801-14 (2012).
53. Hsieh, C.H. *et al.* Potential Role of CXCL13/CXCR5 Signaling in Immune Checkpoint Inhibitor Treatment in Cancer. *Cancers (Basel)* **14**(2022).
54. House, I.G. *et al.* Macrophage-Derived CXCL9 and CXCL10 Are Required for Antitumor Immune Responses Following Immune Checkpoint Blockade. *Clin Cancer Res* **26**, 487-504 (2020).
55. Rapp, M. *et al.* CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. *J Exp Med* **216**, 1170-1181 (2019).
56. Pivarcsi, A. *et al.* Tumor immune escape by the loss of homeostatic chemokine expression. *Proc Natl Acad Sci U S A* **104**, 19055-60 (2007).
57. Atlas, S.A. The renin-angiotensin aldosterone system: pathophysiological role and pharmacologic inhibition. *J Manag Care Pharm* **13**, 9-20 (2007).
58. Rousset, X., Vaisman, B., Amar, M., Sethi, A.A. & Remaley, A.T. Lecithin: cholesterol acyltransferase--from biochemistry to role in cardiovascular disease. *Curr Opin Endocrinol Diabetes Obes* **16**, 163-71 (2009).
59. Floren, C.H., Chen, C.H., Franzen, J. & Albers, J.J. Lecithin: cholesterol acyltransferase in liver disease. *Scand J Clin Lab Invest* **47**, 613-7 (1987).
60. Geng, L., Lam, K.S.L. & Xu, A. The therapeutic potential of FGF21 in metabolic diseases: from bench to clinic. *Nat Rev Endocrinol* **16**, 654-667 (2020).
61. Greve, J.M. *et al.* The major human rhinovirus receptor is ICAM-1. *Cell* **56**, 839-47 (1989).
62. Scialo, F. *et al.* ACE2: The Major Cell Entry Receptor for SARS-CoV-2. *Lung* **198**, 867-877 (2020).

63. Daly, J.L. *et al.* Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* **370**, 861-865 (2020).
64. Walls, A.C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292 e6 (2020).
65. Simmons, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci U S A* **102**, 11876-81 (2005).
66. Lihn, A.S., Pedersen, S.B. & Richelsen, B. Adiponectin: action, regulation and association to insulin sensitivity. *Obes Rev* **6**, 13-21 (2005).
67. Feingold, K.R. Introduction to Lipids and Lipoproteins. in *Endotext* (eds. Feingold, K.R. *et al.*) (South Dartmouth (MA), 2000).
68. Haywood, N.J., Slater, T.A., Matthews, C.J. & Wheatcroft, S.B. The insulin like growth factor and binding protein family: Novel therapeutic targets in obesity & diabetes. *Mol Metab* **19**, 86-96 (2019).
69. Sharma, R., Kopchick, J.J., Puri, V. & Sharma, V.M. Effect of growth hormone on insulin signaling. *Mol Cell Endocrinol* **518**, 111038 (2020).
70. Huggins, K.W., Boileau, A.C. & Hui, D.Y. Protection against diet-induced obesity and obesity- related insulin resistance in Group 1B PLA2-deficient mice. *Am J Physiol Endocrinol Metab* **283**, E994-E1001 (2002).
71. Durham, N.M. *et al.* Lymphocyte Activation Gene 3 (LAG-3) modulates the ability of CD4 T-cells to be suppressed in vivo. *PLoS One* **9**, e109080 (2014).
72. Trinchieri, G. Interleukin-12 and the regulation of innate resistance and adaptive immunity. *Nat Rev Immunol* **3**, 133-46 (2003).
73. Khan, A.U.H. *et al.* The TNFalpha/TNFR2 axis mediates natural killer cell proliferation by promoting aerobic glycolysis. *Cell Mol Immunol* **20**, 1140-1155 (2023).
74. Sans, M. *et al.* VCAM-1 and ICAM-1 mediate leukocyte-endothelial cell adhesion in rat experimental colitis. *Gastroenterology* **116**, 874-83 (1999).
75. Tuveson, D.A., Ahearn, J.M., Matsumoto, A.K. & Fearon, D.T. Molecular interactions of complement receptors on B lymphocytes: a CR1/CR2 complex distinct from the CR2/CD19 complex. *J Exp Med* **173**, 1083-9 (1991).
76. Erdei, A. *et al.* Expression and role of CR1 and CR2 on B and T lymphocytes under physiological and autoimmune conditions. *Mol Immunol* **46**, 2767-73 (2009).
77. Beaubien, F., Raja, R., Kennedy, T.E., Fournier, A.E. & Cloutier, J.F. Slitrk1 is localized to excitatory synapses and promotes their development. *Sci Rep* **6**, 27343 (2016).
78. Abelson, J.F. *et al.* Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* **310**, 317-20 (2005).
79. Miranda, D.M. *et al.* Association of SLITRK1 to Gilles de la Tourette Syndrome. *Am J Med Genet B Neuropsychiatr Genet* **150B**, 483-6 (2009).
80. Piton, A. *et al.* Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol Psychiatry* **16**, 867-80 (2011).
81. Bernard, C.C. *et al.* Myelin oligodendrocyte glycoprotein: a novel candidate autoantigen in multiple sclerosis. *J Mol Med (Berl)* **75**, 77-88 (1997).
82. Zhuo, Z. *et al.* Brain structural and functional alterations in MOG antibody disease. *Mult Scler* **27**, 1350-1363 (2021).
83. Hol, E.M. & Pekny, M. Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system. *Curr Opin Cell Biol* **32**, 121-30 (2015).

84. Abdelhak, A. *et al.* Blood GFAP as an emerging biomarker in brain and spinal cord disorders. *Nat Rev Neurol* **18**, 158-172 (2022).
85. Zhao, B. *et al.* Eye-brain connections revealed by multimodal retinal and brain imaging genetics. *Nat Commun* **15**, 6064 (2024).
86. Imamura, Y. *et al.* Drusen, choroidal neovascularization, and retinal pigment epithelium dysfunction in SOD1-deficient mice: a model of age-related macular degeneration. *Proc Natl Acad Sci U S A* **103**, 11282-7 (2006).
87. Stoner, A. *et al.* Topical carbonic anhydrase inhibitors and glaucoma in 2021: where do we stand? *Br J Ophthalmol* **106**, 1332-1337 (2022).
88. Vourc'h, P. & Andres, C. Oligodendrocyte myelin glycoprotein (OMgp): evolution, structure and function. *Brain Res Brain Res Rev* **45**, 115-24 (2004).
89. Junker, R., Heinrich, J., Schulte, H., van de Loo, J. & Assmann, G. Coagulation factor VII and the risk of coronary heart disease in healthy men. *Arterioscler Thromb Vasc Biol* **17**, 1539-44 (1997).
90. Iacoviello, L. *et al.* Polymorphisms in the coagulation factor VII gene and the risk of myocardial infarction. *N Engl J Med* **338**, 79-85 (1998).
91. Meijers, J.C., Tekelenburg, W.L., Bouma, B.N., Bertina, R.M. & Rosendaal, F.R. High levels of coagulation factor XI as a risk factor for venous thrombosis. *N Engl J Med* **342**, 696-701 (2000).
92. Minnema, M.C. *et al.* Activation of clotting factors XI and IX in patients with acute myocardial infarction. *Arterioscler Thromb Vasc Biol* **20**, 2489-93 (2000).
93. Abdollahi, M., Cushman, M. & Rosendaal, F.R. Obesity: risk of venous thrombosis and the interaction with coagulation factor levels and oral contraceptive use. *Thromb Haemost* **89**, 493-8 (2003).
94. Rau, J.C., Beaulieu, L.M., Huntington, J.A. & Church, F.C. Serpins in thrombosis, hemostasis and fibrinolysis. *J Thromb Haemost* **5 Suppl 1**, 102-15 (2007).
95. Kaaks, R. & Lukanova, A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. *Proc Nutr Soc* **60**, 91-106 (2001).
96. Rodriguez-Calvo, R. *et al.* Role of the fatty acid-binding protein 4 in heart failure and cardiovascular disease. *J Endocrinol* **233**, R173-R184 (2017).
97. Coilly, A., Desterke, C., Guettier, C., Samuel, D. & Chiappini, F. FABP4 and MMP9 levels identified as predictive factors for poor prognosis in patients with nonalcoholic fatty liver using data mining approaches and gene expression analysis. *Sci Rep* **9**, 19785 (2019).
98. Martinez-Una, M., Lopez-Mancheno, Y., Dieguez, C., Fernandez-Rojo, M.A. & Novelle, M.G. Unraveling the Role of Leptin in Liver Function and Its Relationship with Liver Diseases. *Int J Mol Sci* **21**(2020).
99. Baker, K.M., Chernin, M.I., Wixson, S.K. & Aceto, J.F. Renin-angiotensin system involvement in pressure-overload cardiac hypertrophy in rats. *Am J Physiol* **259**, H324-32 (1990).
100. Eguchi, J. *et al.* Identification of adipocyte adhesion molecule (ACAM), a novel CTX gene family, implicated in adipocyte maturation and development of obesity. *Biochem J* **387**, 343-53 (2005).
101. Deprince, A. *et al.* Apolipoprotein F is reduced in humans with steatosis and controls plasma triglyceride-rich lipoprotein metabolism. *Hepatology* **77**, 1287-1302 (2023).

102. Esteghamat, F. *et al.* CELA2A mutations predispose to early-onset atherosclerosis and metabolic syndrome and affect plasma insulin and platelet activation. *Nat Genet* **51**, 1233-1243 (2019).
103. Sun, C. *et al.* Kirrel2, a novel immunoglobulin superfamily gene expressed primarily in beta cells of the pancreatic islets. *Genomics* **82**, 130-42 (2003).
104. Kong, D.H., Kim, Y.K., Kim, M.R., Jang, J.H. & Lee, S. Emerging Roles of Vascular Cell Adhesion Molecule-1 (VCAM-1) in Immunological Disorders and Cancer. *Int J Mol Sci* **19**(2018).
105. Suzuki, K., Kumanogoh, A. & Kikutani, H. Semaphorins and their receptors in immune cell interactions. *Nat Immunol* **9**, 17-23 (2008).
106. Heath, E. *et al.* Abnormal skeletal and cardiac development, cardiomyopathy, muscle atrophy and cataracts in mice with a targeted disruption of the Nov (Ccn3) gene. *BMC Dev Biol* **8**, 18 (2008).
107. Currie, G., Freel, E.M., Perry, C.G. & Dominiczak, A.F. Disorders of blood pressure regulation-role of catecholamine biosynthesis, release, and metabolism. *Curr Hypertens Rep* **14**, 38-45 (2012).
108. Berman, S.M. *et al.* Effects of leptin deficiency and replacement on cerebellar response to food-related cues. *Cerebellum* **12**, 59-67 (2013).
109. Schmidt, A.F. *et al.* Genetic evidence for serum amyloid P component as a drug target in neurodegenerative disorders. *Open Biol* **14**, 230419 (2024).
110. Tennent, G.A., Lovat, L.B. & Pepys, M.B. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. *Proc Natl Acad Sci U S A* **92**, 4299-303 (1995).
111. Yang, X. *et al.* Multi-organ imaging-derived polygenic indexes for brain and body health. *medRxiv* (2024).
112. Yavorska, O.O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* **46**, 1734-1739 (2017).
113. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).
114. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13**, e1007081 (2017).
115. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* **52**, 1122-1131 (2020).
116. Hook, V. *et al.* Cathepsin B in neurodegeneration of Alzheimer's disease, traumatic brain injury, and related brain disorders. *Biochim Biophys Acta Proteins Proteom* **1868**, 140428 (2020).
117. Rhinn, H., Tatton, N., McCaughey, S., Kurnellas, M. & Rosenthal, A. Progranulin as a therapeutic target in neurodegenerative diseases. *Trends Pharmacol Sci* **43**, 641-652 (2022).
118. Yamazaki, Y., Zhao, N., Caulfield, T.R., Liu, C.C. & Bu, G. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat Rev Neurol* **15**, 501-518 (2019).
119. Seo, H.R. *et al.* Intrinsic FGF2 and FGF5 promotes angiogenesis of human aortic endothelial cells in 3D microfluidic angiogenesis system. *Sci Rep* **6**, 28832 (2016).

120. Takeuchi, F. *et al.* Blood pressure and hypertension are associated with 7 loci in the Japanese population. *Circulation* **121**, 2302-9 (2010).
121. Liu, C. *et al.* Common variants in or near FGF5, CYP17A1 and MTHFR genes are associated with blood pressure and hypertension in Chinese Hans. *J Hypertens* **29**, 70-5 (2011).
122. van Nieuwenhoven, F.A. *et al.* Cartilage intermediate layer protein 1 (CILP1): A novel mediator of cardiac extracellular matrix remodelling. *Sci Rep* **7**, 16042 (2017).
123. Park, S., Ranjbarvaziri, S., Zhao, P. & Ardehali, R. Cardiac Fibrosis Is Associated With Decreased Circulating Levels of Full-Length CILP in Heart Failure. *JACC Basic Transl Sci* **5**, 432-443 (2020).
124. Murtha, L.A. *et al.* Fibulin-3 is necessary to prevent cardiac rupture following myocardial infarction. *Sci Rep* **13**, 14995 (2023).
125. Tedford, R.J. *et al.* PDE5A inhibitor treatment of persistent pulmonary hypertension after mechanical circulatory support. *Circ Heart Fail* **1**, 213-9 (2008).
126. Kass, D.A., Champion, H.C. & Beavo, J.A. Phosphodiesterase type 5: expanding roles in cardiovascular regulation. *Circ Res* **101**, 1084-95 (2007).
127. Kuse, Y. *et al.* Progranulin promotes the retinal precursor cell proliferation and the photoreceptor differentiation in the mouse retina. *Sci Rep* **6**, 23811 (2016).
128. Kuse, Y., Tsuruma, K., Mizoguchi, T., Shimazawa, M. & Hara, H. Progranulin deficiency causes the retinal ganglion cell loss during development. *Sci Rep* **7**, 1679 (2017).
129. Marmorstein, L.Y. *et al.* Aberrant accumulation of EFEMP1 underlies drusen formation in Malattia Leventinese and age-related macular degeneration. *Proc Natl Acad Sci U S A* **99**, 13067-72 (2002).
130. Stone, E.M. *et al.* A single EFEMP1 mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. *Nat Genet* **22**, 199-202 (1999).
131. Klein, R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* **308**, 385-9 (2005).
132. Haines, J.L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419-21 (2005).
133. Sun, T. *et al.* ZNRF3 and RNF43 cooperate to safeguard metabolic liver zonation and hepatocyte proliferation. *Cell Stem Cell* **28**, 1822-1837 e10 (2021).
134. Belenguer, G. *et al.* RNF43/ZNRF3 loss predisposes to hepatocellular-carcinoma by impairing liver regeneration and altering the liver lipid metabolic ground-state. *Nat Commun* **13**, 334 (2022).
135. Consortium, G.T. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
136. Sveinbjornsson, G. *et al.* Multiomics study of nonalcoholic fatty liver disease. *Nat Genet* **54**, 1652-1663 (2022).
137. Liu, X. *et al.* Prognostic implications of alcohol dehydrogenases in hepatocellular carcinoma. *BMC Cancer* **20**, 1204 (2020).
138. Li, L. *et al.* ADH4-a potential prognostic marker for hepatocellular carcinoma with possible immune-related implications. *BMC Cancer* **24**, 927 (2024).
139. Yang, L. *et al.* Characterization of the molecular structure, expression and bioactivity of the TNFSF13B (BAFF) gene of the South African clawed frog, Xenopus laevis. *Int Immunopharmacol* **15**, 478-87 (2013).

140. Alexander, P.B. *et al.* EGF promotes mammalian cell growth by suppressing cellular senescence. *Cell Res* **25**, 135-8 (2015).
141. Yamada, D. *et al.* Increased expression of ADAM 9 and ADAM 15 mRNA in pancreatic cancer. *Anticancer Res* **27**, 793-9 (2007).
142. Oh, C.K., Geba, G.P. & Molfino, N. Investigational therapeutics targeting the IL-4/IL-13/STAT-6 pathway for the treatment of asthma. *Eur Respir Rev* **19**, 46-54 (2010).
143. Burchard, E.G. *et al.* Association between a sequence variant in the IL-4 gene promoter and FEV(1) in asthma. *Am J Respir Crit Care Med* **160**, 919-22 (1999).
144. Hartl, C.L. *et al.* Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nature neuroscience* **24**, 1313-1323 (2021).
145. Zhou, Y. *et al.* TTD: Therapeutic Target Database describing target druggability information. *Nucleic Acids Res* **52**, D1465-D1477 (2024).
146. Minikel, E.V., Painter, J.L., Dong, C.C. & Nelson, M.R. Refining the impact of genetic evidence on clinical success. *Nature*, 1-6 (2024).
147. Sudlow, C. *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* **12**, e1001779 (2015).
148. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* **166**, 400-424 (2018).
149. Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* **20**, 45-57 (2001).
150. Patenaude, B., Smith, S.M., Kennedy, D.N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907-22 (2011).
151. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-55 (2002).
152. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A. & Alexander, D.C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000-16 (2012).
153. Daducci, A. *et al.* Accelerated Microstructure Imaging via Convex Optimization (AMICO) from diffusion MRI data. *Neuroimage* **105**, 32-44 (2015).
154. Smith, S.M. *et al.* Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* **31**, 1487-505 (2006).
155. Smith, S.M. *et al.* An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci* **24**, 737-745 (2021).
156. Zhao, B. *et al.* Common variants contribute to intrinsic human brain functional networks. *Nat Genet* **54**, 508-517 (2022).
157. Zhao, B. *et al.* Heart-brain connections: Phenotypic and genetic insights from magnetic resonance images. *Science* **380**, abn6598 (2023).
158. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
159. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191-W198 (2019).
160. Fan, Z. *et al.* The role of sleep in the human brain and body: insights from multi-organ imaging genetics. *medRxiv*, 2022.09.08.22279719 (2024).
161. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).

## Methods

### Image-derived phenotypes of the brain and body.

Image-derived phenotypes (IDPs) used in our study were based on data from the UKB study, which recruited approximately half a million participants aged 40 to 69 between 2006 and 2010[147]. The ethics approval of the UKB study was from the North West Multicentre Research Ethics Committee (approval number: 11/NW/0382) and informed consent was obtained by participants. Overall, we used 258 structural MRI (sMRI) traits capturing regional and total brain volumes derived from T1-weighted structural images, 432 diffusion MRI (dMRI) traits capturing white matter integrity through microstructural and tract-specific measures derived using tract-based spatial statistics (TBSS), 76 resting functional MRI (fMRI)-derived node amplitude traits and 6 global functional connectivity traits summarizing the pairwise coactivity of nodes, 82 cardiac and aortic MRI traits capturing global and regional metrics of four heart chambers and two aortic sections, and 46 derived OCT measures of retinal structure and 110 fundus image features extracted using transfer learning models, as well as 41 abdominal MRI traits capturing body fat and muscle composition, along with kidney, liver, lung, spleen, and pancreas characteristics. These IDPs captured the structural and functional characteristics of multiple human organs and tissues, including the brain, eye, heart, aorta, body fat, body muscle, liver, kidney, lung, pancreas, and spleen.

Brain IDPs used in this study were processed and generated by Alfaro-Almagro, et al. [148]. Briefly, we used sMRI traits extracted by three pipelines, FMRIB's automated segmentation tool[149] (FAST), FMRIB's integrated registration and segmentation tool[150] (FIRST), and FreeSurfer's aseg tool[151]. The FAST pipeline generates 139 regional grey matter volumes by segmenting T1-weighted brain images and combining these results with predefined regions of interest that cover both cortical and subcortical regions. The FIRST pipeline models and measures the shape and volume of 15 subcortical structures and outputs volumes of 15 subcortical structures. The aseg tool segments and measures the volumes of subcortical structures using T1-weighted images. In total, we used 248 sMRI traits generated from the three pipelines, including 139 IDPs processed by the

FAST pipeline, 14 IDPs processed by the FIRST pipeline, and 95 IDPs processed by the aseg pipeline. We additionally included 10 global brain volume measures, both normalized for head size and non-normalized, including combined grey and white matter volume, total white matter volume, total grey matter volume, peripheral cortical grey matter volume, and ventricular cerebrospinal fluid volume. For brain dMRI, diffusion-tensor imaging (DTI) fitting was performed using the DTIFIT tool, generating DTI outputs including fractional anisotropy (FA), mean diffusivity (MD), mode of anisotropy (MO), axial diffusivity (L1), and radial diffusivities (L2 and L3). In addition to DTIFIT, voxelwise microstructural parameters, such as intra-cellular volume fraction (ICVF), isotropic volume fraction (ISOVF), and orientation dispersion index (OD), were derived using Neurite Orientation Dispersion and Density Imaging (NODDI) modelling with the AMICO tool[152,153]. These DTI and NODDI outputs were then processed using TBSS[154], which aligns and skeletonizes the data to generate measures for 48 different white matter tracts for each DTI/NODDI output, resulting in 432 dMRI IDPs. For fMRI traits, we used 76 node amplitude traits reflecting neuronal activity and 6 global functional connectivity measures summarizing the coactivity between node pairs. These global measures were derived using a combined principal component analysis and independent component analysis approach applied to all pairwise functional connectivity traits of the 76 nodes[22,155,156].

We analyzed three main categories of non-brain IDPs: heart and aorta, eye, and abdominal organs/tissues. For the heart and aorta, we used 76 cardiac and 6 aortic MRI traits derived from short-axis, long-axis, and aortic cine images[24,157]. The heart IDPs include global and regional metrics for four cardiac chambers: 64 traits for the left ventricle, 4 for the right ventricle, 4 for the left atrium, and 4 for the right atrium. The aortic traits consist of 3 traits for the ascending aorta and 3 for the descending aorta. For the eye, we analyzed 156 retinal imaging traits[85], comprising 46 measures derived from optical coherence tomography (OCT) images and 110 features from fundus photographs. The derived OCT measures include retinal thickness across specific layers, vertical cup-to-disc ratio, and disc diameter. The fundus image traits were generated using 11 pre-trained transfer learning models built on ImageNet, extracting the top 10 principal components (PCs) from each model's final layer, resulting in 110 fundus image features. Abdominal

IDPs comprised 41 MRI measures, covering body fat, body muscle, and organ-specific features. These included 8 body fat traits assessing abdominal fat distribution and visceral/subcutaneous fat volumes, and 12 body muscle traits measuring lean tissue volumes, fat-free muscle volumes in the thighs, and muscle fat infiltration as an indicator of muscle quality. Organ-specific measures included 6 kidney traits (volume and parenchyma), 10 liver traits (fat and iron content, with liver iron corrected T1 reflecting inflammation and fibrosis), 1 lung volume trait, 1 spleen volume trait, and 3 pancreas traits (volume, fat, and iron content). In total, there were 1051 IDPs across 8 imaging modalities (**Table S2**).

**UK Biobank pharma proteomics project.**

UK Biobank pharma proteomics project (UKB-PPP) profiled the plasma proteomes of 54,219 UKB participants using Olink Explore 3072, measuring 2,941 protein analytes representing 2,923 unique proteins across eight panels focused on inflammation, cardiometabolic health, neurology, and oncology. The project included 46,595 randomly selected participants at baseline, 6,376 consortium-selected participants, and 1,268 from the COVID-19 repeat imaging study. Protein signals were quantified as normalized protein expression values, processed using Olink's MyData Cloud Software. Details on data collection, processing, and quality control are documented in previous study[4].

**Phenotypic plasma protein-imaging associations.**

To examine the phenotypic associations between plasma proteins and IDPs, we fitted linear regression models using unrelated white British subjects[158] as discovery samples (average $n$ = 4,383). The analysis was adjusted for covariates including age (at protein assessment and imaging visit), age squared, sex, age-sex interaction, age-squared-sex interaction, top ten genetic PCs, height, weight, and body mass index. For fMRI IDPs, we additionally adjusted for volumetric scaling, head motion, and brain position[22,155]. For regional brain volumes, we additionally adjusted for total brain volume to account for global effects. $P$-values were derived from two-sided $t$-tests (R version 4.1.0) and multiple testing was corrected using Bonferroni adjustment. The analysis was replicated in an independent hold-out set of white non-British subjects (average $n$ = 513). Associations

were considered significant if they passed the Bonferroni threshold in the discovery sample ($P < 1.63\times 10^{-8}$), had a $P$-value < 0.05 in the replication sample, and exhibited concordant effect directions between the two models.

**Protein-protein interaction enrichment analysis.**

Protein-protein interaction (PPI) was obtained from the STRING database[39], which integrates evidence from various interaction sources, including textmining, experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence, to assign interaction scores. For a given list of proteins, the mean PPI score was calculated by averaging the pairwise PPI scores between all protein pairs. Additionally, to assess whether the observed interactions among the proteins in the list were significantly enriched, a statistical test was performed using a one-sided Wilcoxon rank-sum test. The test compared the PPI scores of the proteins in the query list against the PPI scores of all 2,923 proteins in our study, serving as the background. This analysis determined whether the query protein set exhibited significantly higher interaction compared to the overall background set of proteins.

**Pathway enrichment analysis.**

To identify biological pathways and functional annotations enriched among the identified proteins, we performed a pathway enrichment analysis using g:Profiler[159]. The protein list was uploaded to the g:Profiler web tool, which applies its built-in multiple testing correction method ‘g:SCS threshold’, ensuring rigorous control of false discovery rates. We selected data resources including Gene Ontology molecular function, cellular component, and biological process, as well as KEGG and Reactome pathways, to evaluate potential functional enrichment. To focus on biologically relevant terms, we restricted enriched terms to a size range of 10 to 500, ensuring that the enriched terms were neither too broad nor too narrow. This approach allowed us to identify meaningful biological pathways and processes in which the associated proteins with each organ were involved, offering deeper insights into the molecular mechanisms underlying the protein-imaging trait associations.

**Sensitivity analysis.**

To evaluate the potential impact of disease status on the associations between plasma proteins and IDPs, we performed a sensitivity analysis by including relevant disease statuses as covariates in the regression models. Diseases for each organ were defined based on the following ICD-10 categories: brain-related diseases (F and G), including mental and behavioral disorders, as well as diseases of the nervous system; heart and aorta-related diseases (I), encompassing ischemic heart disease, heart failure, and arrhythmias; body fat-related disorders (E65-E78), covering obesity, hyperlipidemia, and metabolic syndromes; kidney-related diseases (N00-N29), including nephrotic syndromes and chronic kidney diseases; liver-related diseases (K70-K77), such as alcoholic liver disease, nonalcoholic fatty liver disease, and cirrhosis; pancreas-related diseases (K85-K87), including acute and chronic pancreatitis; eye-related diseases (H00-H59), such as glaucoma, cataracts, and retinal disorders; spleen-related diseases (D73), including splenomegaly and other splenic disorders; and muscle-related diseases (M60-M62 and M79), including myopathies, muscle inflammation, and fibromyalgia. To ensure robust statistical power, diseases with fewer than 500 cases were excluded from the analysis. Beta coefficients and *P*-values for the associations were reported after adjusting for disease status.

**Organ-specific plasma proteins.**

We assigned organ-specific labels to plasma proteins following the approach described by Oh, et al. [13]. Briefly, gene expression data were obtained from the GTEx[38], and tissue gene expression levels were normalized. A gene was defined as organ-specific, which served as the organ label of the corresponding protein, if its expression level in one organ was at least four times higher than in any other organ, consistent with the definition proposed by the Human Protein Atlas[19]. To ensure consistency, tissues from the same organ were grouped, and the maximum expression level among sub tissues was used to represent the expression level for that organ. Following Oh, et al. [13], the immune organ was defined as combined gene expression in blood and spleen tissues. Using this approach, 557 out of 2,923 proteins were assigned organ-specific labels based on their gene expression profiles. To evaluate whether a list of proteins was enriched for proteins

labeled with that organ based on gene expression, we performed a hypergeometric test across all organs. The test compared the number of proteins in the list with matching organ labels to the total proteins in the list, the total proteins labeled with that organ, and the total proteins analyzed, using the hypergeometric distribution to calculate the significance of enrichment.

**Prediction model using proteomic data.**

To predict IDPs using plasma proteins, we used an elastic-net regression model for each imaging trait. Following Carrasco-Zanini et al.[16], we randomly sampled half of the data for training and testing, repeating this process across 200 iterations to ensure robust results. Before analysis, IDPs were normalized to allow comparability across different traits, while protein predictors data were also standardized within the elastic-net model. During each iteration, elastic-net regression was performed, and coefficients for each protein were estimated using the optimal lambda determined through cross-validation. Predicted IDP values for the testing data were then calculated based on these estimated coefficients, and prediction performance was assessed using the *R*-squared metric. After the 200 repetitions, absolute values of coefficients across the 200 repetitions were summed up, as an importance measure of each protein for IDPs. GLIPR1 was excluded from the prediction model due to its low data quality as described in Sun, et al. [4]. To determine the top five predictors for each IDP, we calculated a combined score for each protein by integrating the absolute sum of coefficients and the frequency of selection across iterations. Both metrics were first normalized to a [0, 1] scale by dividing each value by the maximum value in its respective column. A weighted sum of the normalized metrics was then computed for each protein, with equal weights (0.5) assigned to the normalized coefficient sum and frequency. Proteins were ranked in descending order of their combined scores, and the top five with the highest scores were selected as the strongest predictors for the IDP.

**Mendelian randomization and colocalization.**

We conducted Mendelian randomization (MR) analysis to investigate the genetic causal links between plasma proteins and IDPs. The list of proteins and IDPs used in the MR

analysis was largely the same as in previous sections, with minor adjustments due to data availability or data filtering criteria. For IDPs, genome-wide association study (GWAS) summary statistics were obtained from previous studies[85,111,157,160]. Ten global brain volume measures were excluded due to data unavailability, while the remaining 1,041 IDPs were included in the MR analysis (average *n* = 40,682 participants). Detailed information on these GWAS, including data preprocessing and quality control procedures, is available in the respective studies. For plasma proteins, 67 out of the 2,923 plasma proteins were excluded based on the following three criteria: First, fusion proteins which represent combined entities or closely related paralogs rather than distinct proteins, were excluded to avoid ambiguity in biological interpretation. Second, proteins encoded by genes located on the sex chromosomes were removed to account for sex-specific differences in expression. Finally, proteins that could not be mapped to genomic locations in the GRCh37/hg19 genome build were excluded. Detailed description of these exclusions is available in the previous study[32]. Additionally, 36 proteins located within the MHC region were excluded to minimize bias from high linkage disequilibrium (LD) in this genomic region, leaving 2,820 proteins for the MR analysis. To prevent overlap with subjects in the imaging GWAS, we generated protein quantitative trait loci (pQTL) specifically for this MR analysis, ensuring that participants (and their relatives[158]) from the imaging GWAS were excluded from the pQTL analysis (average *n* = 34,566 participants). We removed the effects of the same set of covariates as in the previous study[32]. In addition, only *cis*-pQTLs, defined as genetic variants located within a gene's region and extending one megabase on both sides, were used in the MR analysis. To generate independent instrumental variables, *cis*-pQTLs were clumped using PLINK 2.0[161], with a *P*-value threshold of $5\times 10^{-8}$, a window size of 250 kb, and an $r^2$ threshold of 0.001 for LD. Additionally, variants with an *F*-statistic less than 10 were excluded to avoid weak instrument bias.

The MR methods used in the study were the inverse variance weighted (IVW) and the Wald-ratio (for single instrument variant). The Wald-ratio was conducted with the TwoSampleMR[113,114] package, while the IVW method was implemented with the MendelianRandomization[112] package. Following the pipeline described by Zheng, et al.

[115], we used a range of sensitivity tests to ensure the robustness of our results. Specifically, the Steiger *P*-value evaluated whether the genetic instruments explained more variance in the exposure (plasma proteins) than in the outcome (IDPs), confirming the correct causal direction. Additionally, the heterogeneity *P*-value assessed the consistency of causal estimates across genetic instruments, with significant heterogeneity indicating potential violations of MR assumptions. To account for multiple testing, a Bonferroni correction was applied after filtering out results that failed sensitivity tests, resulting in an adjusted *P*-value threshold of $P < 2.98\times 10^{-8}$ and 8,116 significant causal protein-imaging trait associations.

To further assess whether the significant MR pairs shared a causal variant, we performed colocalization analysis on the 8,116 pairs with significant causal relationships. We applied the coloc.abf() method from the coloc package[33], a Bayesian framework that calculates five posterior probabilities: PPH0 (no association), PPH1 (association with the protein only), PPH2 (association with the imaging trait only), PPH3 (independent associations with both traits), and PPH4 (a shared causal variant). Pairs with PPH4 > 80% were considered colocalized, providing strong evidence for a shared genetic basis between the plasma protein and IDP.

## Code availability

We made use of publicly available software and tools. The code used in this study will be deposited in Zenodo upon publication.

## Data availability

The individual-level data used in this study can be obtained from UK Biobank (https://www.ukbiobank.ac.uk/). The GWAS summary statistics for imaging phenotypes across different organs are available from previous studies[85,111,157,160]. The GWAS summary statistics of Olink plasma proteins (removing imaging subjects) produced in this study will be deposited in Zenodo upon publication. Other datasets in this paper include: the STRING database (https://string-db.org/), the Therapeutic Target Database (https://idrblab.net/ttd/), and the GTEx dataset v8 (https://gtexportal.org/).

## Figure legends

### Fig. 1 Study overview.

**(A)** Our study aimed to explore the phenotypic and genetic connections between plasma proteome and organ structure and function, using 2,923 proteins and 1,051 imaging traits spanning multiple organs from the UK Biobank study. Building on a diverse range of imaging modalities, the study included brain imaging traits such as structural MRI, diffusion MRI, and functional MRI; cardiac MRI traits derived from short-axis, long-axis, and aortic cine images; abdominal MRI traits capturing measurements related to the liver, kidney, lung, pancreas, spleen, and body fat and muscle composition; and eye imaging traits such as derived retinal optical coherence tomography and fundus photography images**. (B)-(E)** Overview of the major analyses, data resources, and key scientific questions addressed in this study.

### Fig. 2 Phenotypic protein-imaging associations and selected enriched pathways.

**(A)** The number of significant proteins associated with each organ after the Bonferroni correction ($P < 1.63\times 10^{-8}$) and the overlapping pattern across organs. **(B)** Enrichment of protein-protein interaction (PPI) scores using the STRING database was assessed. Enrichment scores were calculated for the set of associated proteins for each organ, and statistical significance was tested using a one-sided Wilcoxon rank-sum test. **(C)** Overview of the number of enriched biological pathways in each organ. **(D)** Selected biological pathways enriched among proteins associated with heart imaging traits, including the regulation of angiogenesis and vasculature development (multiple testing-adjusted $P < 1.53\times 10^{-7}$). **(E)** Selected biological pathway enriched among proteins associated with liver imaging traits, highlighting the regulation of plasma lipoprotein particle levels pathway (adjusted $P = 5.36\times 10^{-5}$). **(F)** Selected biological pathways enriched among proteins associated with pancreas imaging traits, including digestion and protein digestion and absorption (adjusted $P < 1.03\times 10^{-7}$). **(G)** Selected biological pathways enriched among proteins associated with spleen volume, including lymphocyte mediated immunity and regulation of leukocyte mediated immunity (adjusted $P < 6.67\times 10^{-15}$).

**Fig. 3 Tracing putative origins of proteomic effects on imaging traits.**

**(A)** Bubble plot that highlights whether proteins associated with imaging traits of an organ (*y*-axis) were significantly enriched among genes highly expressed in the same or another organ (*x*-axis). Bubble size reflects the enrichment significance level and bubble colors differentiate between within-organ or cross-organ enrichments. Only significant enrichments were highlighted in colors. For example, proteins associated with imaging data from four organs/tissues were significantly enriched among those with high gene expression in the liver. This included the liver itself (within-organ) as well as adipose tissue, muscle, and heart (cross-organ). **(B)** Overview of putative origins of proteomic effects on imaging traits. The left axis represents the organs where proteins exhibit phenotypic associations with imaging traits, while the right axis indicates the proteins' original organs mapped by gene expression data. The thickness of the connections between the axes reflects the relative number of proteins contributing to each connection. For example, proteins with high gene expression in the liver (right axis) were associated with IDPs of the liver, heart, body muscle, and body fat (left axis). **(C)-(H)** Example proteins demonstrating putative within-organ proteomic associations in each of the six organs/tissues showing significant within-organ enrichment in **(A)**. For example, in **(C)**, BCAN, NCAN, OMG, and GFAP were associated with brain sMRI and dMRI traits and exhibited high gene expression in the brain. **(I)-(J)** Example of putative cross-organ associations. In **(I)**, we show correlation coefficients between plasma proteins highly expressed in the liver (*x*-axis) and imaging traits of the heart, liver, body muscle, and body fat with which these proteins were phenotypically associated (*y*-axis). The color represents correlation estimates. Coefficients that passed Bonferroni correction ($P < 1.63\times 10^{-8}$) and replicated were marked with asterisk, while those that passed Bonferroni correction but not replicated were marked with a plus sign. In **(J)**, we illustrate example proteins (F9, SERPIND1, and FGF21) that showed significant associations with the heart, body fat, body muscle, and liver and were highly expressed in the liver. **(K)-(L)** Example of putative cross-organ associations. In **(K)**, we show correlation coefficients between plasma proteins highly expressed in the body fat (*x*-axis) and imaging traits of the lung, liver, body fat, body muscle, and heart with which these proteins were phenotypically associated (*y*-axis). The color represents correlation estimates. Coefficients that passed

Bonferroni correction ($P < 1.63\times 10^{-8}$) and replicated were marked with asterisk, while those that passed Bonferroni correction but not replicated were marked with a plus sign. In **(L)**, we illustrate example proteins (FABP4 and LEP) that showed significant associations with the heart, body muscle, lung, liver, and body fat and were highly expressed in body fat.

**Fig. 4 Predictive power of plasma proteins on organ structure and function.**

**(A)** Predictive accuracy of selected imaging traits across all organs, including the brain, eye, heart, spleen, lung, liver, kidney, pancreas, and body fat, along with the top three plasma protein predictors for each trait. ICVF refers to the intra-cellular volume fraction and RNFL refers to retinal nerve fiber layer. LVM stands for left ventricular myocardial mass, while LVEDV and RVEDV represent left and right ventricular end-diastolic volumes, respectively. **(B)-(F)** Top-ranked plasma proteins for selected imaging traits. The *x*-axis represents the frequency of proteins appearing across 200 resampling iterations, while the *y*-axis represents the average absolute value of coefficients across the 200 resampling iterations. Top proteins are labeled with their names and are further highlighted in color if this protein is highly expressed in certain specific organ (at least 4 times higher than any other organs). We illustrate the results for total trunk fat volume (**B**), pancreas volume (**C**), left ventricular end-diastolic volume (LVEDV, **D**), volume of cerebral white matter (**E**), and volume of the insular cortex (**F**). **(G)** The top-to-bottom ratios of average visceral fat volume for percentile groups representing the top and bottom 10%, 20%, and 30% of predicted imaging traits based on plasma protein-based prediction model. The *x*-axis represents the percentile groups, and the *y*-axis represents the top-to-bottom ratios. Sex groups, including all individuals, females, and males, are distinguished by color.

**Fig. 5 Genetic-root putative causal protein-imaging links.**

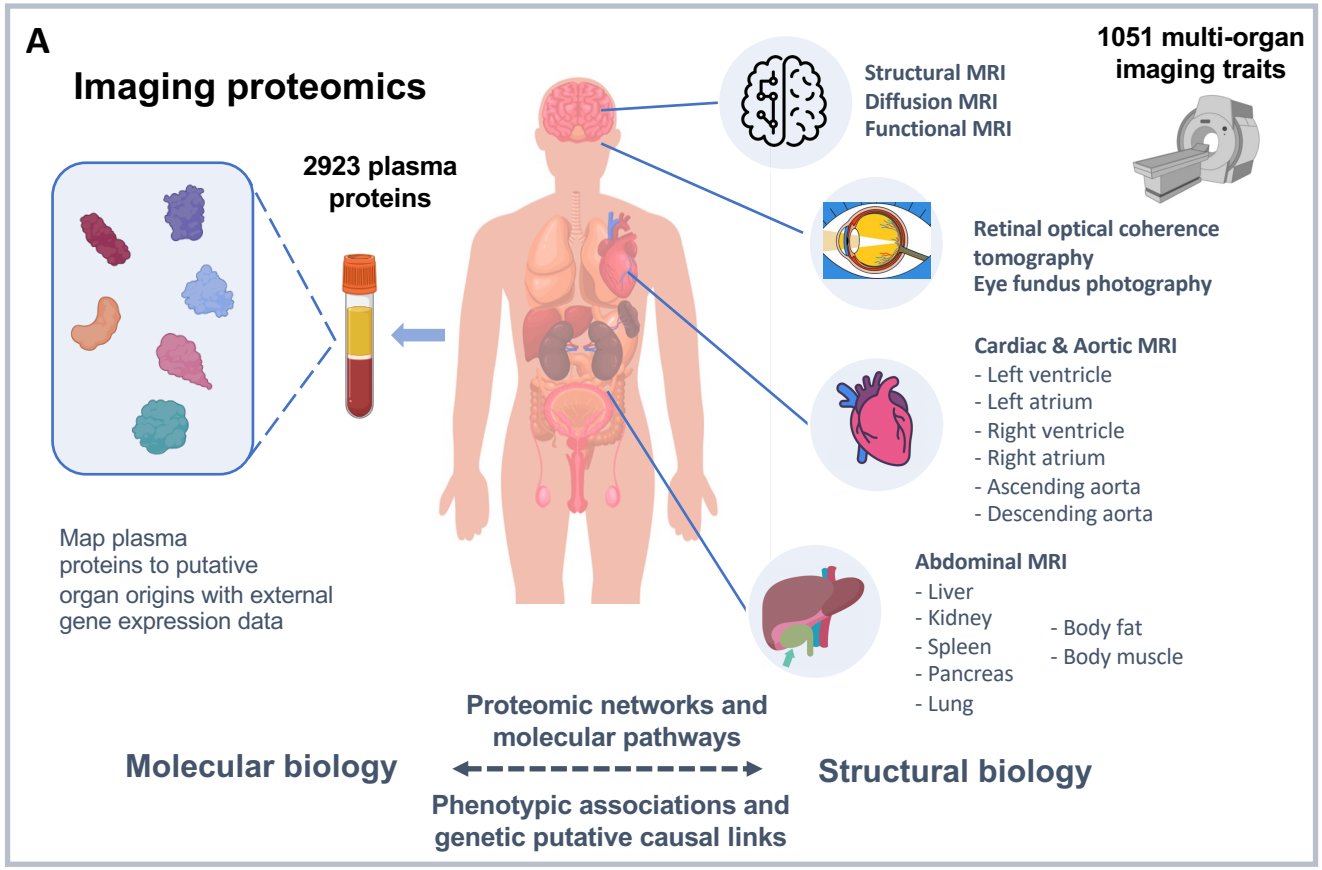
**(A)** Putative Causal associations between plasma proteins and imaging traits identified by Mendelian randomization (MR). The *x*-axis represents the chromosomes where the proteins are located, the *y*-axis represents the -log10(P-value) of protein-imaging associations. Only protein-imaging pairs significant under Bonferroni correction from the

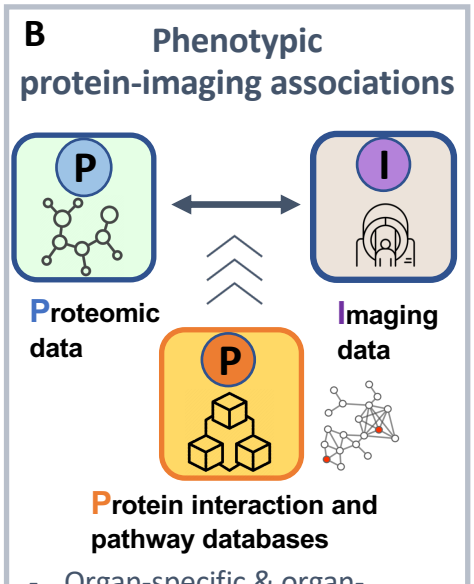
MR analysis were plotted ($P < 2.98\times 10^{-8}$). Pairs with strong evidence of sharing causal variants in Bayesian colocalization analysis (PPH4 > 80%) were highlighted in organ-specific colors. Brain sMRI stands for brain structural MRI, brain dMRI stands for brain diffusion MRI, ICA edge and ICA nodes are brain functional MRI traits generated by independent component analysis, OCT stands for optical coherence tomography. **(B)** The number of significant proteins associated with each organ after the Bonferroni correction ($P < 2.98\times 10^{-8}$) in MR analysis and the overlapping pattern across organs. **(C)** Overview of the number of MR-enriched biological pathways in each organ. **(D)** Comparison of protein-protein interaction (PPI) enrichment scores between phenotypically and causally associated proteins across organs. PPI scores were calculated using the STRING database for proteins associated with each organ. The *x*-axis shows PPI enrichment scores for phenotypically associated proteins, while the *y*-axis shows scores for MR-associated proteins. Points are colored by organ. Dashed lines mark an enrichment score of one, representing the baseline level of including all proteins. **(E)** Number of enriched pathways for phenotypically associated and MR-associated proteins, as well as the number of pathways shared between the two approaches.

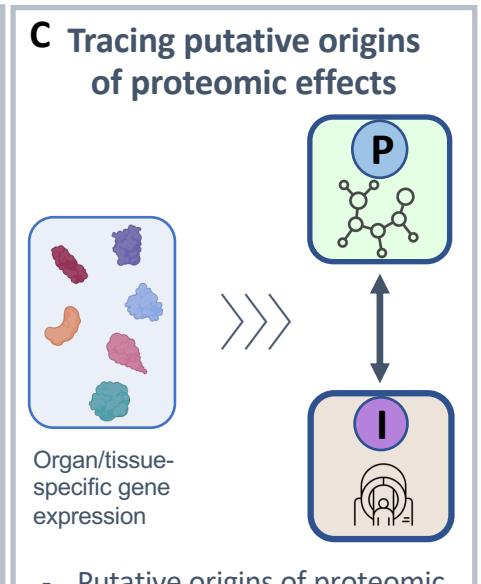
**Fig. 6 Selected associations identified by Mendelian randomization (MR) with shared causal variants.**

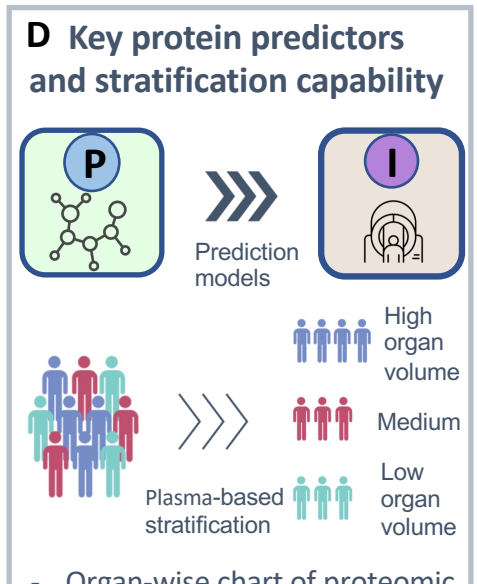
**(A)** In MR analysis, BCAN was causally associated with mean intracellular volume fraction (ICVF) in the genu of the corpus callosum. The posterior probability of Bayesian colocalization analysis for the shared causal variant hypothesis (PPH4) is 100.00%. **(B)** APOE was causally associated with functional activity (amplitude) trait in the default mode and central executive networks. The PPH4 of Bayesian colocalization analysis is 99.71%. **(C)** FGF5 was causally associated with ascending aorta maximum area (AAo max area). The PPH4 of Bayesian colocalization analysis is 98.72%. **(D)** CILP was causally associated with left atrium ejection fraction (LAEF). The PPH4 of Bayesian colocalization analysis is 98.26%. **(E)** GRN was causally associated with the average thickness between the external limiting membrane (ELM) and the inner and outer photoreceptor segments (ISOS) layers. The PPH4 of Bayesian colocalization analysis is 99.89%. **(F)** TNFSF13B

was causally associated with spleen volume. The PPH4 of Bayesian colocalization analysis is 99.90%.


A
Imaging proteomics
2923 plasma proteins
Map plasma proteins to putative organ origins with external gene expression data
1051 multi-organ imaging traits
Structural MRI
Diffusion MRI
Functional MRI
Retinal optical coherence tomography
Eye fundus photography
Cardiac & Aortic MRI
- Left ventricle
- Left atrium
- Right ventricle
- Right atrium
- Ascending aorta
- Descending aorta
Abdominal MRI
- Liver
- Kidney
- Spleen
- Pancreas
- Lung
- Body fat
- Body muscle
Proteomic networks and molecular pathways
Molecular biology
Structural biology
Phenotypic associations and genetic putative causal links


B
Phenotypic protein-imaging associations
Proteomic data
Imaging data
Protein interaction and pathway databases
- Organ-specific & organ-shared proteomic links
- Enriched protein-protein interaction networks
- Enriched biological pathways


C
Tracing putative origins of proteomic effects
Organ/tissue-specific gene expression
- Putative origins of proteomic effects on imaging traits
- Insights into organ-specific biological mechanisms and cross-organ interactions


D
Key protein predictors and stratification capability
Prediction models
High organ volume
Medium
Low organ volume
Plasma-based stratification
- Organ-wise chart of proteomic prediction accuracy and key predictors
- Plasma-based model for organ abnormality stratification

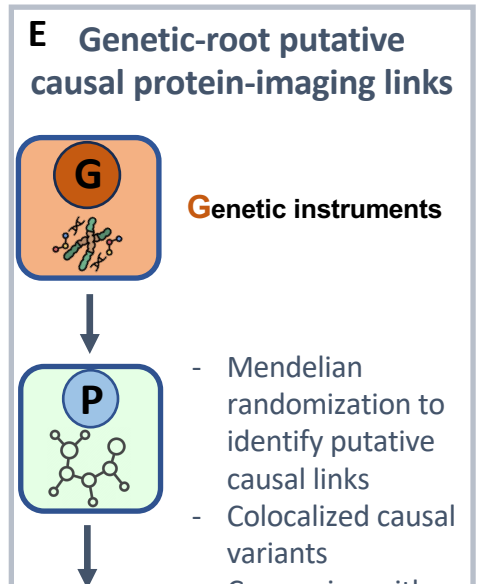
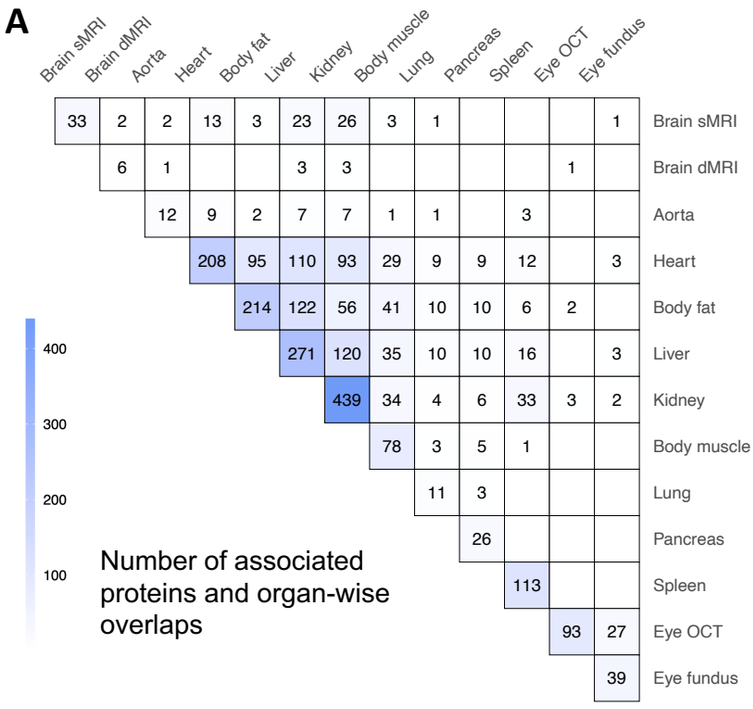
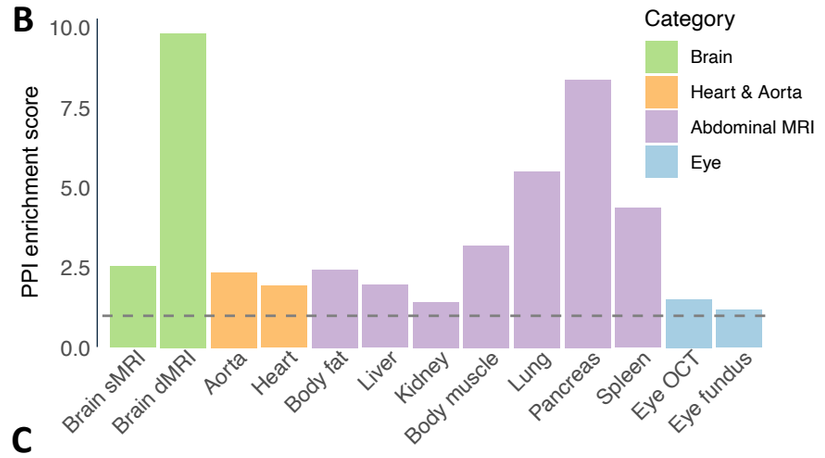
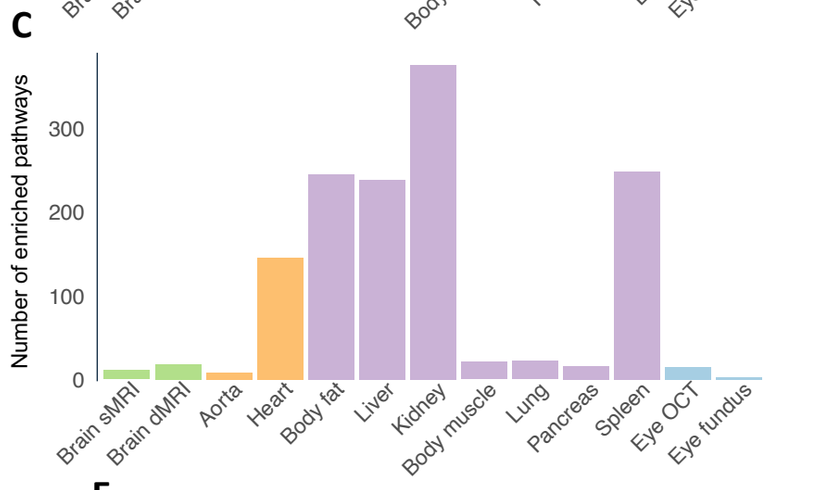
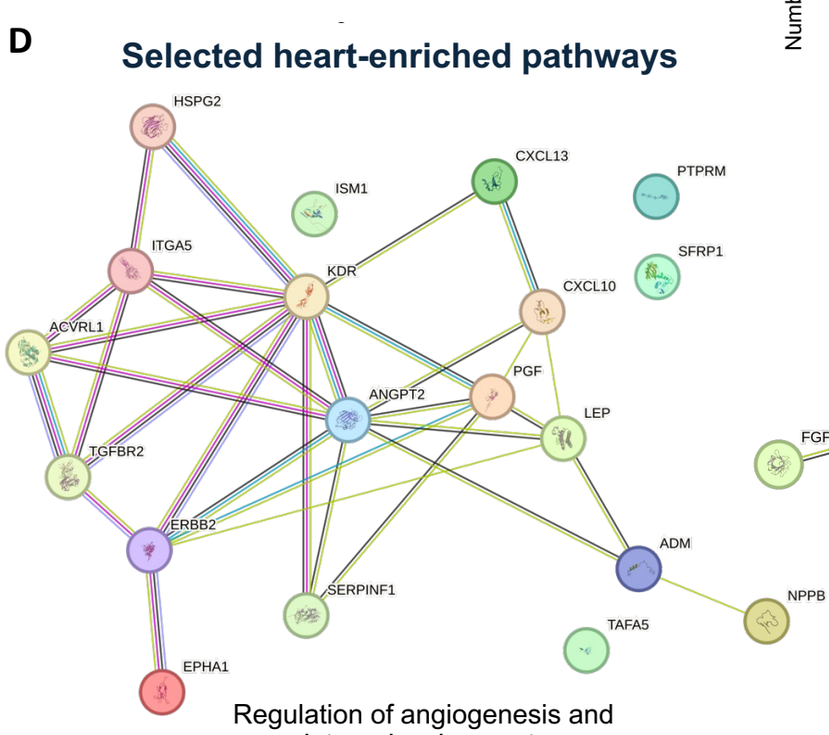

E
Genetic-root putative causal protein-imaging links
Genetic instruments
- Mendelian randomization to identify putative causal links
- Colocalized causal variants
- Comparing with phenotypic findings

**Figure 1**

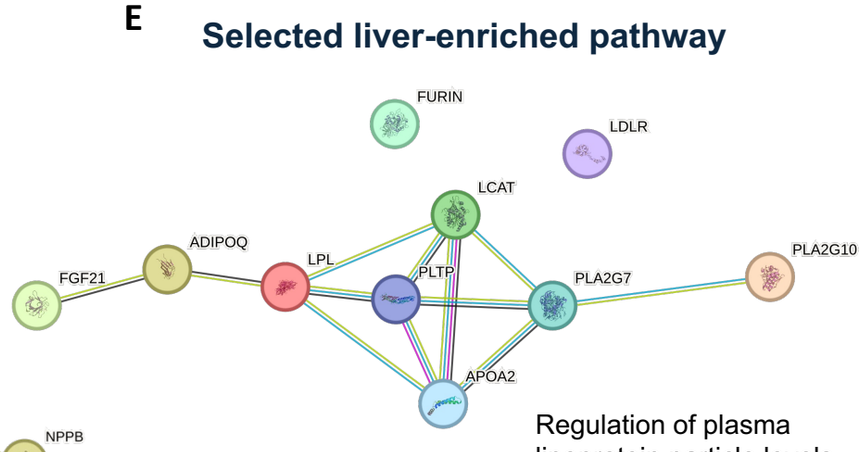


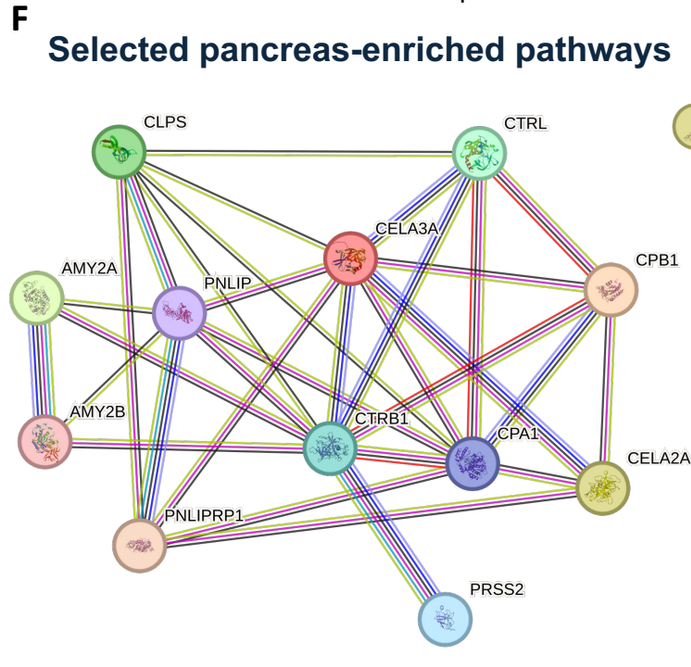



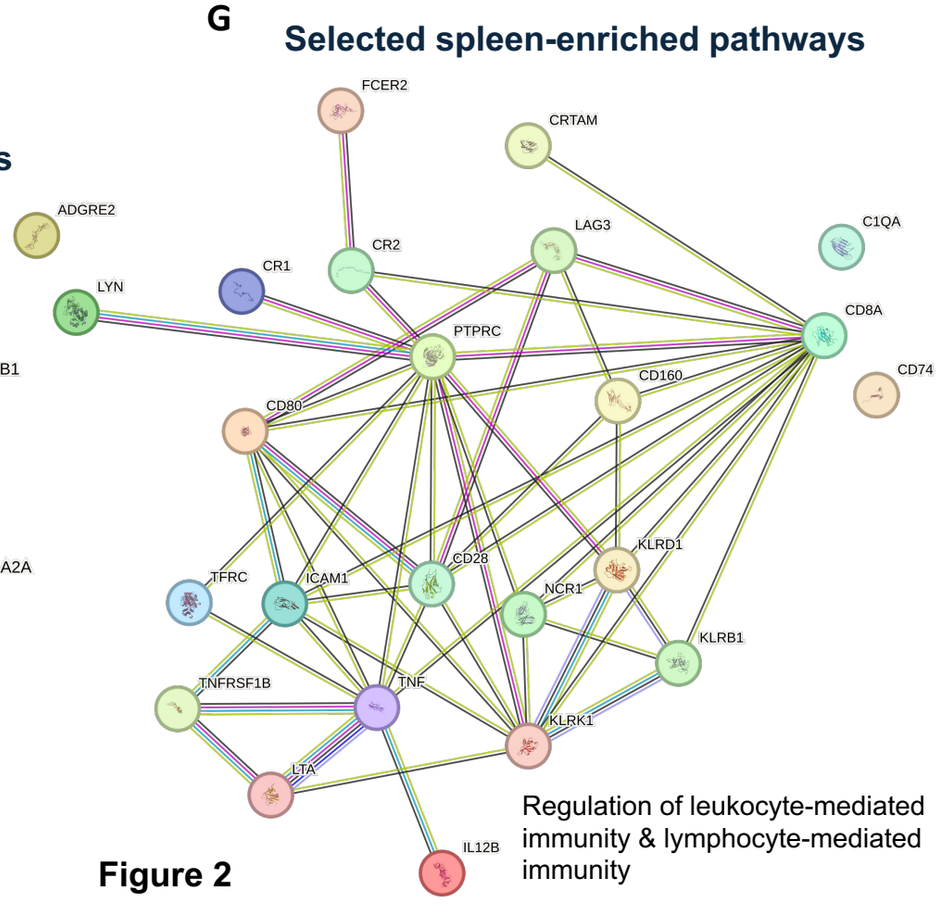

Regulation of angiogenesis and vasculature development




Regulation of plasma lipoprotein particle levels




Digestion (left part) and protein digestion and absorption (right part) pathways




Regulation of leukocyte-mediated immunity & lymphocyte-mediated immunity

**Figure 2**

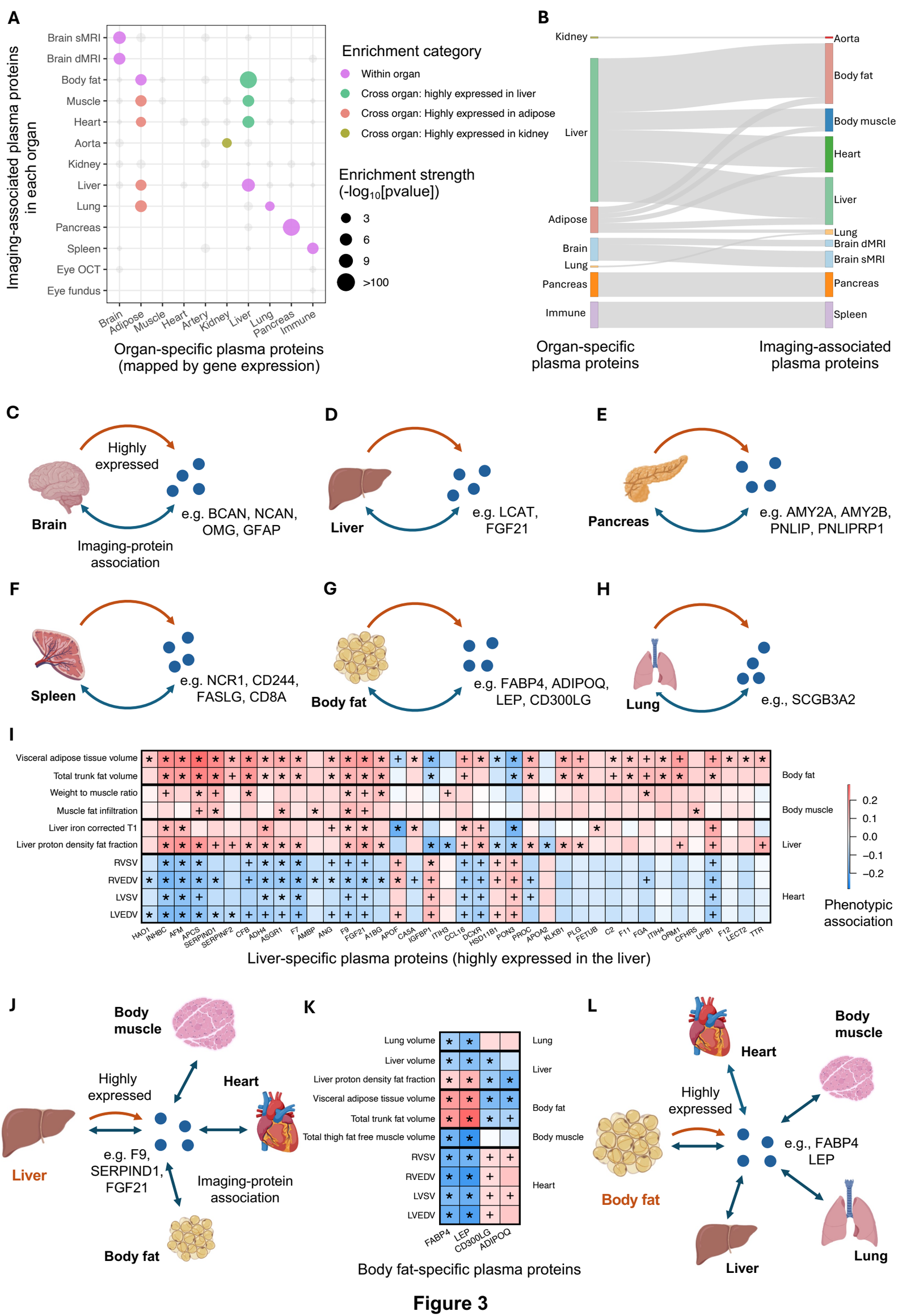

A
Imaging-associated plasma proteins in each organ
Brain sMRI
Brain dMRI
Body fat
Muscle
Heart
Aorta
Kidney
Liver
Lung
Pancreas
Spleen
Eye OCT
Eye fundus
Brain
Adipose
Muscle
Heart
Artery
Kidney
Liver
Lung
Pancreas
Immune
Organ-specific plasma proteins (mapped by gene expression)
Enrichment category
Within organ
Cross organ: highly expressed in liver
Cross organ: Highly expressed in adipose
Cross organ: Highly expressed in kidney
Enrichment strength (-log10[pvalue])
3
6
9
>100
B
Kidney
Liver
Adipose
Brain
Lung
Pancreas
Immune
Aorta
Body fat
Body muscle
Heart
Liver
Lung
Brain dMRI
Brain sMRI
Pancreas
Spleen
Organ-specific plasma proteins
Imaging-associated plasma proteins
C
Highly expressed
Brain
e.g. BCAN, NCAN, OMG, GFAP
Imaging-protein association
D
Liver
e.g. LCAT, FGF21
E
Pancreas
e.g. AMY2A, AMY2B, PNLIP, PNLIPRP1
F
Spleen
e.g. NCR1, CD244, FASLG, CD8A
G
Body fat
e.g. FABP4, ADIPOQ, LEP, CD300LG
H
Lung
e.g., SCGB3A2
I
Visceral adipose tissue volume
Total trunk fat volume
Weight to muscle ratio
Muscle fat infiltration
Liver iron corrected T1
Liver proton density fat fraction
RVSV
RVEDV
LVSV
LVEDV
HAO1 INHBC AFM APCS SERPIND1 SERPINF2 CFB ADH4 ASGR1 F7 AMBP ANG F9 FGF21 A1BG APOF CA5A IGFBP1 ITIH3 CCL16 DCXR HSD11B1 PON3 PROC APOA2 KLKB1 PLG FETUB C2 F11 FGA ITIH4 ORM1 CFHR5 UPB1 F12 LECT2 TTR
Body fat
Body muscle
Liver
Heart
0.2
0.1
0.0
−0.1
−0.2
Phenotypic association
Liver-specific plasma proteins (highly expressed in the liver)
J
Body muscle
Highly expressed
Heart
Liver
e.g. F9, SERPIND1, FGF21
Imaging-protein association
Body fat
K
Lung volume
Liver volume
Liver proton density fat fraction
Visceral adipose tissue volume
Total trunk fat volume
Total thigh fat free muscle volume
RVSV
RVEDV
LVSV
LVEDV
Lung
Liver
Body fat
Body muscle
Heart
FABP4 LEP CD300LG ADIPOQ
Body fat-specific plasma proteins
L
Heart
Body muscle
Highly expressed
e.g., FABP4 LEP
Body fat
Liver
Lung

**Figure 3**

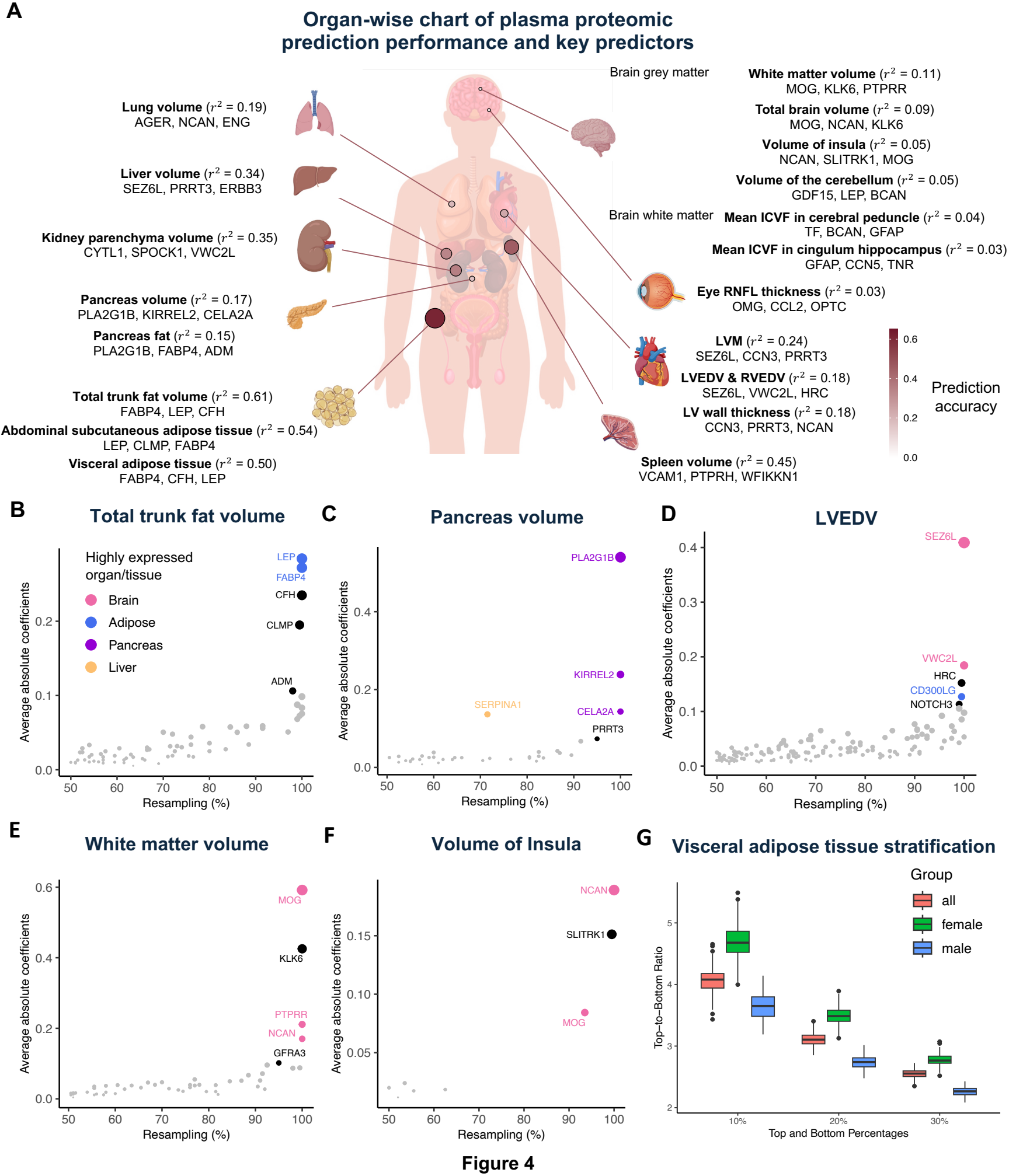

A
Organ-wise chart of plasma proteomic prediction performance and key predictors
Lung volume ($r^2$ = 0.19)
AGER, NCAN, ENG
Liver volume ($r^2$ = 0.34)
SEZ6L, PRRT3, ERBB3
Kidney parenchyma volume ($r^2$ = 0.35)
CYTL1, SPOCK1, VWC2L
Pancreas volume ($r^2$ = 0.17)
PLA2G1B, KIRREL2, CELA2A
Pancreas fat ($r^2$ = 0.15)
PLA2G1B, FABP4, ADM
Total trunk fat volume ($r^2$ = 0.61)
FABP4, LEP, CFH
Abdominal subcutaneous adipose tissue ($r^2$ = 0.54)
LEP, CLMP, FABP4
Visceral adipose tissue ($r^2$ = 0.50)
FABP4, CFH, LEP
Brain grey matter
Brain white matter
White matter volume ($r^2$ = 0.11)
MOG, KLK6, PTPRR
Total brain volume ($r^2$ = 0.09)
MOG, NCAN, KLK6
Volume of insula ($r^2$ = 0.05)
NCAN, SLITRK1, MOG
Volume of the cerebellum ($r^2$ = 0.05)
GDF15, LEP, BCAN
Mean ICVF in cerebral peduncle ($r^2$ = 0.04)
TF, BCAN, GFAP
Mean ICVF in cingulum hippocampus ($r^2$ = 0.03)
GFAP, CCN5, TNR
Eye RNFL thickness ($r^2$ = 0.03)
OMG, CCL2, OPTC
LVM ($r^2$ = 0.24)
SEZ6L, CCN3, PRRT3
LVEDV & RVEDV ($r^2$ = 0.18)
SEZ6L, VWC2L, HRC
LV wall thickness ($r^2$ = 0.18)
CCN3, PRRT3, NCAN
Spleen volume ($r^2$ = 0.45)
VCAM1, PTPRH, WFIKKN1
0.6
0.4
0.2
0.0
Prediction accuracy
B
Total trunk fat volume
Highly expressed organ/tissue
Brain
Adipose
Pancreas
Liver
LEP
FABP4
CFH
CLMP
ADM
Average absolute coefficients
Resampling (%)
C
Pancreas volume
PLA2G1B
KIRREL2
SERPINA1
CELA2A
PRRT3
D
LVEDV
SEZ6L
VWC2L
HRC
CD300LG
NOTCH3
E
White matter volume
MOG
KLK6
PTPRR
NCAN
GFRA3
F
Volume of Insula
NCAN
SLITRK1
MOG
G
Visceral adipose tissue stratification
Group
all
female
male
Top−to−Bottom Ratio
Top and Bottom Percentages
10%
20%
30%

**Figure 4**

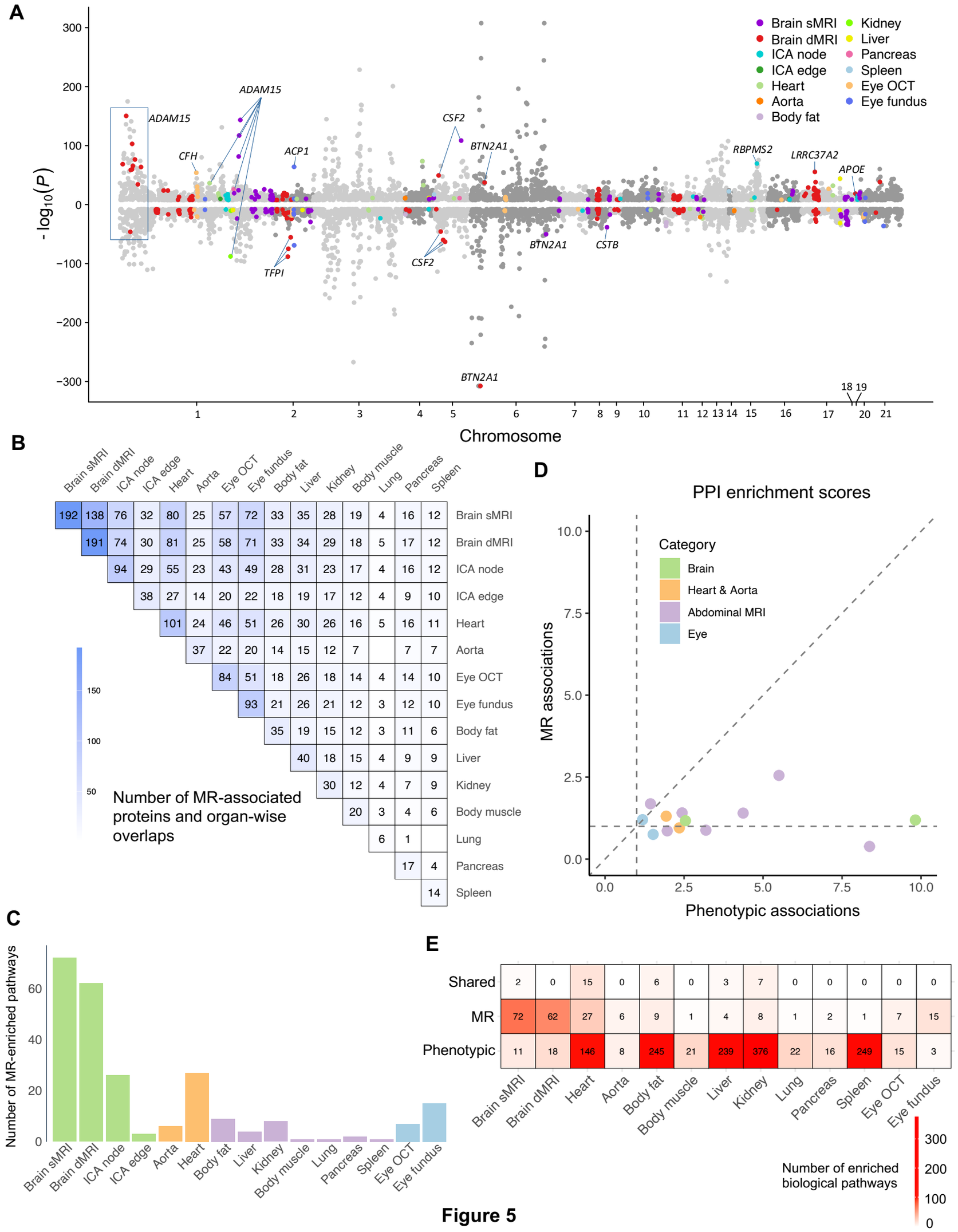

A
Brain sMRI
Brain dMRI
ICA node
ICA edge
Heart
Aorta
Body fat
Kidney
Liver
Pancreas
Spleen
Eye OCT
Eye fundus
ADAM15
CFH
ACP1
TFPI
CSF2
BTN2A1
RBPMS2
LRRC37A2
APOE
CSTB
- log10(P)
Chromosome
B
Number of MR-associated proteins and organ-wise overlaps
C
Number of MR-enriched pathways
D
PPI enrichment scores
Category
Brain
Heart & Aorta
Abdominal MRI
Eye
MR associations
Phenotypic associations
E
Shared
MR
Phenotypic
Number of enriched biological pathways

**Figure 5**

A
chr1, Region: 1q23.1
156.4 mb
156.6 mb
156.8 mb
156.5 mb
156.7 mb
BCAN
rs3795736
-log10(p)
ICVF of the genu of the corpus callosum
rs3795736
Gene Mode
IQGAP3
BCAN
RRNAD1
SH2D2A
LOC101928177
NES
HDGF
NTRK1
MEF2D
GPATCH4
CRABP2
PRCC
TTC24
ISG20L2
INSRR
NAXE
MRPL24
HAPLN2
B
chr19, Region: 19q13.32
45.2 mb
45.4 mb
45.6 mb
45.3 mb
45.5 mb
rs769449
APOE
-log10(p)
Functional activity (amplitude) in the default mode and central executive networks
rs429358
rs769449
Gene Model
MIR8085
NECTIN2
RELB
NKPD1
BCL3
TOMM40
CLASRP
CAM16
CBLC
APOC4-APOC2
GEMIN7-AS1
BCAM
APOC1P1
ZNF296
APOE
GEMIN7
APOC1
PPP1R37
APOC4
APOC2
CLPTM1
C
chr4, Region: 4q21.21
81 mb
81.2 mb
81.1 mb
81.3 mb
FGF5
rs58418317
4:81170475_CT_C
-log10(p)
Ascending aorta maximum area (AAo max area)
rs58418317
4:81170475_CT_C
Gene Model
PRDM8
FGF5
CFAP299
D
chr15, Region: 15q22.31
65.3 mb
65.5 mb
65.7 mb
65.4 mb
65.6 mb
rs2585035
rs2681038
CILP
-log10(p)
Left atrium ejection fraction (LAEF)
rs2681038
rs2585035
Gene Model
SLC51B
PDCD7
CILP
RNU5A-1
IGDCC4
KBTBD13
CLPX
PARP16
DPP8
MTFMT
SNORA24B
RASL12
RNU5B-1
UBAP1L
IGDCC3
E
chr17, Region: 17q21.31
42.3 mb
42.5 mb
42.4 mb
42.6 mb
GRN
rs5848
-log10(p)
Thickness between the external limiting membrane (ELM) and the inner and outer photoreceptor segments (ISOS) layers
rs5848
Gene Model
ATXN7L3
RUNDC3A
GPATCH8
FZD2
TMUB2
RUNDC3A-AS1
ITGA2B
HROB
SLC4A1
FAM171A2
ASB16-AS1
SLC25A39
ASB16
GRN
MIR6782
UBTF
Colocalized
index SNPs
LD to index SNP
r² ≥ 0.8
0.8 > r² ≥ 0.6
0.6 > r² ≥ 0.4
0.4 > r² ≥ 0.2
0.2 > r² ≥ 0
F
chr13, Region: 13q33.3
108.8 mb
109 mb
108.9 mb
109.1 mb
TNFSF13B
rs374039502
13:108960380_TGCTG_T
-log10(p)
Spleen volume
rs374039502
13:108960380_TGCTG_T
Gene Model
LIG4
TNFSF13B
ABHD13

**Figure 6**